

UNIVERSIDAD NACIONAL DEL LITORAL  
FACULTAD DE INGENIERÍA QUÍMICA

*Tesis presentada como parte de los requisitos de la Universidad Nacional del Litoral, para la obtención del Grado Académico de Doctor en Ingeniería Química*

OPTIMIZACIÓN EN TIEMPO REAL (RTO) DE  
PROCESOS INDUSTRIALES.  
APLICACIONES A SISTEMAS DE ENERGÍA

Autor: Fernán J. Serralunga

Director : Miguel C. Mussati  
Codirector: Pio A. Aguirre



Jurados: Díaz, María Soledad (PLAPIQUI, UNS-CONICET)

Marchetti, Alejandro (CIFASIS, UNR-CONICET)

Vecchietti, Aldo (INGAR, UTN-CONICET)

Santa Fe, Octubre de 2014



## AGRADECIMIENTOS

A mi director, Miguel, y a mi codirector, Pío, por el acompañamiento y los consejos a lo largo de este trabajo, y por la libertad que me dieron en todo momento para tomar la iniciativa y proponer ideas.

A mis amigos del Ingar, sobre todo a mi compañero de oficina Carlos, por las discusiones sobre el trabajo y la vida.

A mi familia, sobre todo, por la paciencia en los días de más trabajo y por el apoyo constante. ¡Gracias!



## RESUMEN

La operación eficiente de los procesos industriales es determinante para reducir el consumo de energía, minimizar el impacto ambiental y maximizar los beneficios de la producción. La *optimización en tiempo real* (RTO) es una herramienta de la informática aplicada a la ingeniería de procesos, que utiliza las mediciones existentes en la planta para *adaptar* los modelos disponibles y determinar la forma óptima de operar un proceso. Las *estrategias de adaptación* caracterizan a una implementación de RTO; entre ellas, la estrategia de *adaptación por modificadores* se utiliza como base para este trabajo.

Esta tesis desarrolla formulaciones alternativas de la optimización en tiempo real basada en modelos de estado estacionario. A lo largo de este trabajo se propone una evolución partiendo de la estructura tradicional con un modelo no lineal, continuo y de un solo período, a través de la incorporación progresiva de decisiones discretas (con modelos de programación disyuntiva) y de formulaciones multiperíodo que facilitan la integración de las técnicas de RTO con el *scheduling* óptimo.

Cada una de las formulaciones se aplica a modelos de sistemas de generación de calor y potencia. Estos sistemas son apropiados para la optimización en tiempo real con modelos estacionarios por la frecuente presencia de grados de libertad, la dinámica rápida y la necesidad de responder a cambios en la demanda de potencia y calor, precios y condiciones ambientales, teniendo en cuenta restricciones de capacidad, operativas y ambientales.

La estructura particular de los sistemas de energía se utiliza para obtener formulaciones alternativas de la estrategia de adaptación por modificadores. Estas formulaciones (que reciben el nombre de *adaptación por modificadores basada en ecuaciones de eficiencia*) permiten reducir la cantidad de datos necesarios para la estimación del gradiente y aplicar la estrategia a sistemas con un gran número de variables de entrada.

La inclusión de decisiones discretas en la optimización en tiempo real se aborda a través de una formulación basada en programación disyuntiva. Este enfoque incluye la reformulación de las estrategias de adaptación (incluyendo la *adaptación basada en ecuaciones de eficiencia*) para obtener una técnica apropiada para la *RTO con disyunciones*, y la revisión y corrección de las prácticas utilizadas en la RTO con modelos continuos.

La integración de la RTO con los demás sistemas de toma de decisión (*scheduling*, control) es clave para la operación eficiente de un proceso. En esta tesis se aborda la integración con el *scheduling óptimo*, a través de una formulación de RTO multiperíodo, y una resolución en dos niveles (*SO+RTO*) que establece una jerarquía de variables discretas: algunas se fijan en un problema de *scheduling* óptimo inicial (resuelto para un horizonte dado), y el resto, junto con las variables continuas, en subproblemas de RTO (resueltos en tiempo real).

Finalmente, se aborda la problemática del tiempo en la RTO. Un requerimiento de esta técnica es que los resultados óptimos deben obtenerse en un tiempo acotado (en el orden de segundos o minutos). Se propone una metodología para aprovechar el *tiempo máximo disponible* para generar *soluciones candidatas* a ser aplicadas en la planta, a través del uso de más de un algoritmo de solución y de modelos de diferente complejidad. Por otra parte, se plantean estrategias para el uso del *tiempo entre ejecuciones de RTO*, que puede ser mucho mayor al tiempo efectivamente usado para la adaptación del modelo y la optimización.

# ÍNDICE

AGRADECIMIENTOS .....	I
RESUMEN.....	III
ÍNDICE .....	V
ÍNDICE DE FIGURAS.....	XIII
ÍNDICE DE TABLAS.....	XV
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>Resumen.....</b>	<b>1</b>
<b>1.1. Optimización en tiempo real.....</b>	<b>1</b>
1.1.1. Necesidad de la optimización en tiempo real .....	1
1.1.2. Características generales de la optimización en tiempo real (RTO).....	3
1.1.3. Estructura jerárquica de optimización industrial. Integración con otras capas. ....	5
<b>1.2. Problemas actuales de la RTO.....</b>	<b>6</b>
1.2.1. Complejidad de los métodos de estimación del gradiente .....	6
1.2.2. Ausencia de un planteo formal para problemas MINLP o GDP .....	7
1.2.3. Integración de RTO con <i>scheduling</i> de procesos – Formulaciones multiperíodo.....	9
1.2.4. Infactibilidades y resultados subóptimos .....	10
<b>1.3. Objetivos de la tesis .....</b>	<b>11</b>
1.3.1. Reducir el problema de estimación del gradiente .....	11
1.3.2. Extender el problema de RTO a formulaciones GDP.....	11
1.3.3. Integrar RTO y <i>scheduling</i> con formulaciones multiperíodo .....	12
1.3.4. Aprovechar el tiempo y los recursos computacionales disponibles .....	12
1.3.5. Aplicar las estrategias a sistemas de calor y potencia .....	12
<b>1.4. Organización de la tesis .....</b>	<b>13</b>
<b>1.5. Publicaciones de esta tesis .....</b>	<b>14</b>
1.5.1. Artículos en publicaciones científicas.....	14
1.5.2. Trabajos presentados en reuniones científicas .....	14

---

<b>2. ESTADO DEL ARTE EN RTO.....</b>	<b>17</b>
<b>Resumen .....</b>	<b>17</b>
<b>2.1. Planteo general del problema .....</b>	<b>17</b>
<b>2.2. Errores estructurales y paramétricos .....</b>	<b>18</b>
<b>2.3. Estrategias de adaptación .....</b>	<b>21</b>
2.3.1. Estrategia de dos pasos .....	21
2.3.2. Adaptación por modificadores .....	23
2.3.3. Estrategias de adaptación directa de las entradas .....	27
2.3.4. Otras estrategias .....	28
<b>2.4. Técnicas de estimación del gradiente.....</b>	<b>29</b>
2.4.1. Perturbaciones en la planta.....	29
2.4.2. Múltiples unidades .....	30
2.4.3. Neighboring Extremals .....	30
2.4.4. Estimación dinámica.....	31
2.4.5. RTO dual .....	32
2.4.6. Regresión con estructura fija .....	35
<b>2.5. Detección de estado estacionario .....</b>	<b>37</b>
<b>2.6. Validación de datos .....</b>	<b>39</b>
2.6.1. Validación simple.....	39
2.6.2. Reconciliación de datos .....	39
2.6.3. Detección de errores gruesos.....	41
<b>2.7. Validación de los resultados .....</b>	<b>42</b>
2.7.1. Verificación estadística.....	42
2.7.2. Filtrado de las entradas óptimas.....	43
2.7.3. Optimización de <i>targets</i> .....	44
<b>2.8. Métricas para evaluar el desempeño de un sistema de RTO .....</b>	<b>45</b>
<b>2.9. RTO con incertidumbre .....</b>	<b>45</b>
2.9.1. Intervalos de confianza .....	46
2.9.2. Margen de seguridad en las restricciones.....	46
2.9.3. Formulaciones con incertidumbre .....	47
<b>2.10. RTO multiperíodo.....</b>	<b>48</b>

<b>2.11. RTO con modelos dinámicos .....</b>	<b>48</b>
2.11.1. Integración con control predictivo multivariable (MPC).....	48
2.11.2. RTO dinámica (DRTO) .....	49
<b>2.12. Aplicaciones industriales.....</b>	<b>50</b>
<b>2.13. Otros usos de las técnicas de RTO .....</b>	<b>51</b>
<b>2.14. Scheduling en presencia de incertidumbre .....</b>	<b>51</b>
<b>Conclusiones parciales .....</b>	<b>52</b>
<b>3. ADAPTACIÓN POR MODIFICADORES BASADA EN ECUACIONES DE EFICIENCIA.....</b>	<b>53</b>
<b>Resumen.....</b>	<b>53</b>
<b>3.1. Descripción del método.....</b>	<b>53</b>
<b>3.2. Implementación de las estrategias de adaptación.....</b>	<b>60</b>
3.2.1. Alternativa 1: Corrección + Filtro exponencial (MA-E1).....	60
3.2.2. Alternativa 2: Regresión lineal ponderada (MA-E2) .....	61
3.2.3. Estructura propuesta.....	64
<b>3.3. Casos de estudio .....</b>	<b>64</b>
3.3.1. Caso de estudio 1: Sistema genérico .....	65
3.3.1.1. Descripción del caso.....	65
3.3.1.2. Pasos a seguir en el sistema de RTO .....	67
3.3.1.3. Adaptación por modificadores con ecuaciones de eficiencia – Alternativa 1 (MA-E1).....	68
3.3.1.4. Adaptación por modificadores con ecuaciones de eficiencia – Alternativa 2 (MA-E2).....	70
3.3.1.5. Adaptación de las restricciones (CA).....	71
3.3.1.6. Adaptación por modificadores con actualización de Broyden (MA-Broyden) .....	71
3.3.1.7. Resultados y discusión.....	73
3.3.2. Caso de estudio 2: Sistema de calor y potencia de una planta de azúcar y etanol.....	82
3.3.2.1. Descripción del caso.....	82
3.3.2.2. Implementación de la estrategia de RTO .....	88
3.3.2.3. Resultados y discusión.....	90
<b>Conclusiones parciales .....</b>	<b>96</b>
<b>4. RTO CON PROGRAMACIÓN DISYUNTIVA.....</b>	<b>99</b>

<b>Resumen .....</b>	<b>99</b>
<b>4.1. Necesidad del modelado con decisiones discretas .....</b>	<b>99</b>
<b>4.2. Programación disyuntiva generalizada (GDP).....</b>	<b>100</b>
<b>4.3. Disyunciones típicas en RTO .....</b>	<b>101</b>
4.3.1. Funciones definidas a trozos .....	101
4.3.2. Contratos de compra y venta – Penalidades en la función objetivo .....	102
4.3.3. Regiones operativas desconectadas.....	102
4.3.4. RTO dual .....	102
4.3.5. Cambios mínimos .....	103
<b>4.4. Optimización en tiempo real incluyendo disyunciones.....</b>	<b>105</b>
4.4.1. Definición del problema.....	105
4.4.2. Estrategias de adaptación .....	106
4.4.2.1. Estrategia de dos pasos .....	106
4.4.2.2. Adaptación por modificadores .....	106
4.4.2.3. Adaptación por modificadores con ecuaciones de eficiencia .....	109
4.4.3. Adaptación por modificadores de funciones definidas a trozos.....	112
<b>4.5. Revisión y reformulación de prácticas de RTO continua.....</b>	<b>114</b>
4.5.1. Restricción de máximo cambio o penalización del cambio en las entradas .....	115
4.5.2. Filtrado de las entradas óptimas .....	118
<b>4.6. Casos de estudio.....</b>	<b>125</b>
4.6.1. Caso de estudio 1: Sistema genérico.....	126
4.6.1.1. Descripción y modelado del sistema .....	126
4.6.1.2. Resultados y discusión.....	132
4.6.2. Caso de estudio 2: Sistema de calor y potencia.....	137
4.6.2.1. Descripción y modelado del sistema .....	137
4.6.2.2. Resultados y discusión.....	142
<b>Conclusiones parciales.....</b>	<b>147</b>
<b>5. INTEGRACIÓN DE RTO Y SCHEDULING ÓPTIMO EN SISTEMAS DE ENERGÍA .....</b>	<b>149</b>
<b>Resumen .....</b>	<b>149</b>
<b>5.1. Introducción.....</b>	<b>149</b>
<b>5.2. Planteo del problema .....</b>	<b>151</b>

<b>5.3. Fenómenos que requieren una formulación multiperíodo en RTO .....</b>	<b>152</b>
5.3.1. Restricciones de transición .....	152
5.3.2. Variables de acumulación .....	154
<b>5.4. Integración de <i>scheduling</i> óptimo y optimización en tiempo real (SO+RTO) .....</b>	<b>155</b>
5.4.1. Descripción de la metodología propuesta .....	155
5.4.2. Formulación de la metodología .....	156
5.4.2.1. Nivel 1: Optimización del schedule completo (SO) .....	156
5.4.2.2. Nivel 2: RTO multiperíodo (RTO) .....	156
5.4.3. Reacción a las soluciones no factibles .....	158
5.4.4. Estructura completa propuesta para el sistema integrado SO+RTO.....	160
<b>5.5. Caso de estudio: sistema de calor y potencia .....</b>	<b>160</b>
5.5.1. Descripción del caso .....	160
5.5.2. Resultados y discusión.....	164
<b>Conclusiones parciales .....</b>	<b>170</b>
<b>6. USO EFICIENTE DEL TIEMPO EN RTO .....</b>	<b>171</b>
<b>Resumen .....</b>	<b>171</b>
<b>6.1. Introducción.....</b>	<b>171</b>
<b>6.2. Problemas frecuentes en la optimización en sistemas de RTO .....</b>	<b>172</b>
6.2.1. Soluciones no factibles .....	172
6.2.2. Límite de tiempo.....	173
6.2.3. Óptimos locales.....	173
6.2.4. Algoritmos de optimización.....	174
<b>6.3. Estrategias para el aprovechamiento del tiempo y los recursos disponibles .....</b>	<b>175</b>
6.3.1. Uso del tiempo máximo disponible para obtener la solución.....	175
6.3.1.1. Generación de soluciones candidatas .....	175
6.3.1.2. Resolución en serie .....	177
6.3.1.3. Resolución en paralelo .....	179
6.3.1.4. Combinación serie-paralelo .....	179
6.3.2. Uso del tiempo entre ejecuciones .....	180
<b>Conclusiones parciales .....</b>	<b>182</b>

<b>7. CONCLUSIONES Y LÍNEAS DE TRABAJO A FUTURO .....</b>	<b>185</b>
<b>7.1. Conclusiones.....</b>	<b>185</b>
<b>7.2. Líneas de trabajo futuras .....</b>	<b>189</b>
7.2.1. Estrategias de adaptación .....	189
7.2.1.1. Análisis de convergencia .....	189
7.2.1.2. Formulación de restricciones de RTO dual .....	189
7.2.1.3. Adaptación en base a variables discretas .....	189
7.2.1.4. Adaptación en base a perturbaciones medidas u observables .....	189
7.2.2. Filtrado de las entradas óptimas para RTO con disyunciones .....	190
7.2.3. Simulaciones y trabajo experimental.....	190
7.2.3.1. Procesos de mayor escala .....	190
7.2.3.2. Modelado dinámico del proceso .....	190
7.2.3.3. Trabajo experimental.....	190
7.2.4. Integración de sistemas .....	191
7.2.4.1. Integración entre métodos de RTO .....	191
7.2.4.2. Integración SO+RTO+control .....	191
7.2.5. Uso del tiempo disponible .....	191
7.2.6. Extensión a otros procesos.....	191
7.2.6.1. Procesos continuos – modelos de estado estacionario.....	191
7.2.6.2. Optimización dinámica en tiempo real – DRTO .....	191
7.2.6.3. Run-to-run optimization.....	192
<b>BIBLIOGRAFÍA .....</b>	<b>193</b>
<b>A. APÉNDICE A: MODELO DEL SISTEMA DE ENERGÍAS DE LA PLANTA DE AZÚCAR Y ETANOL .....</b>	<b>203</b>
<b>A.1. Nomenclatura .....</b>	<b>203</b>
<b>A.2. Límites operativos.....</b>	<b>204</b>
<b>A.3. Modelos de los equipos .....</b>	<b>205</b>
A.3.1. Calderas ( $cb = \{1,2\}$ ).....	205
A.3.2. Etapas del turbogenerador ( $tg = \{1,2\}$ ).....	205
A.3.3. Válvulas reductoras de presión ( $vl = \{1,2\}$ ) .....	206
A.3.4. Turbinas de vapor de contrapresión ( $tb = \{1..4\}$ ).....	206
A.3.5. Efectos del evaporador ( $e = \{0..5\}$ (el índice 0 se utiliza para el jugo clarificado a la entrada y el vapor al primer efecto).....	206

A.3.6. Desaireador .....	207
<b>A.4. Balances .....</b>	<b>208</b>
A.4.1. Matriz de incidencias .....	208
A.4.2. Balances de masa, energía y entropía: .....	209
<b>A.5. Conexiones equipos-corrientes .....</b>	<b>210</b>
<b>B. APÉNDICE B: MODELO DEL SISTEMA DE ENERGÍAS (CASO DE ESTUDIO 2, CAPÍTULO 4).....</b>	<b>211</b>
<b>B.1. Nomenclatura .....</b>	<b>211</b>
<b>B.2. Límites operativos.....</b>	<b>212</b>
<b>B.3. Modelos de los equipos .....</b>	<b>213</b>
B.3.1. Calderas ( $cb = \{1,2\}$ ) .....	213
B.3.2. Etapas del turbogenerador ( $tg = \{1,2\}$ ) .....	213
B.3.3. Válvula reductora de presión ( $vl$ ) .....	214
B.3.4. Turbinas de vapor de contrapresión ( $tb = \{1..3\}$ ) .....	214
B.3.5. Turbina de gas .....	215
B.3.6. Desaireador .....	215
B.3.7. Penalidad de compra de electricidad .....	215
<b>B.4. Balances .....</b>	<b>216</b>
B.4.1. Matriz de incidencias.....	216
B.4.2. Balances de masa, energía y entropía:.....	216
<b>B.5. Perfil de temperaturas, precios y demandas .....</b>	<b>217</b>
<b>C. APÉNDICE C: MODELO DEL SISTEMA DE ENERGÍAS (CASO DE ESTUDIO, CAPÍTULO 5).....</b>	<b>219</b>
<b>C.1. Modelo base.....</b>	<b>219</b>
<b>C.2. Nomenclatura .....</b>	<b>219</b>
<b>C.3. Modelos de los equipos.....</b>	<b>220</b>
C.3.1. Calderas ( $cb = \{1,2\}$ ).....	220
C.3.2. Válvula reductora de presión ( $vl$ ).....	220
C.3.3. Turbinas de vapor de contrapresión ( $tb = \{1..3\}$ ).....	220
C.3.4. Producción de NOx .....	221
<b>C.4. Perfil de temperaturas, precios y demandas.....</b>	<b>222</b>



## ÍNDICE DE FIGURAS

<b>Figura 1-1.</b> Estructura de un ciclo de optimización en tiempo real.....	5
<b>Figura 1-2.</b> Estructura jerárquica de la toma de decisiones en la industria.....	6
<b>Figura 2-1.</b> Evolución de un sistema de RTO con errores estructurales en el modelo. .....	21
<b>Figura 3.1.</b> Esquema de la estructura de RTO propuesta. ....	64
<b>Figura 3.2.</b> Sistema modelado en el caso de estudio I. ....	65
<b>Figura 3.3.</b> Caso de estudio 1. Evolución del sistema – Estrategia MA-E1. ....	75
<b>Figura 3.4.</b> Caso de estudio 1. Evolución del sistema – Estrategia MA-E2. ....	76
<b>Figura 3.5.</b> Caso de estudio 1. Evolución del sistema – Estrategia MA-Broyden....	77
<b>Figura 3.6.</b> Caso de estudio 1. Evolución de la función objetivo para todas las estrategias analizadas.....	78
<b>Figura 3.7.</b> Influencia del filtro exponencial (MA-E1, Broyden-MA) y del peso inicial (MA-E2) en el Costo de Diseño Extendido. Nivel de ruido: desviación estándar 0.1.....	81
<b>Figura 3.8.</b> Influencia del nivel de ruido (desviación estándar del ruido gaussiano) en el Costo de Diseño Extendido. Parámetros de filtrado según <b>Tabla 3-1</b> .....	81
<b>Figura 3.9.</b> Caso de estudio 2. Sistema de calor y potencia de la planta de azúcar y etanol. ....	82
<b>Figura 3.10.</b> Caso de estudio 2 – Escenario 1. Evolución de la función objetivo. ....	92
<b>Figura 3.11.</b> Caso de estudio 2 – Escenario 2. Evolución de la función objetivo. ....	93
<b>Figura 3.12.</b> Caso de estudio 2 – Escenario 3. Evolución de la función objetivo. ....	94
<b>Figura 3.13.</b> Caso de estudio 2 – Escenario 4. Evolución de la función objetivo. ....	95
<b>Figura 4-1.</b> Convergencia utilizando restricciones de cambio máximo y términos de convexificación. ....	117
<b>Figura 4-2 .</b> Filtrado de las entradas óptimas y regiones factibles desconectadas.	121
<b>Figura 4-3.</b> Problemas de convergencia y estrategia corregida para el filtrado de las entradas incluyendo costos fijos.....	124
<b>Figura 4-4.</b> Caso de estudio 1. Diagrama del sistema estudiado. ....	126
<b>Figura 4-5.</b> Caso de estudio 1. Evolución de las entradas para adaptación de restricciones (CA) y adaptación por modificadores (MA-Broyden). ....	134
<b>Figura 4-6.</b> Caso de estudio 1. Evolución de las entradas vs. ciclo RTO. ....	135
<b>Figura 4-7.</b> Caso de estudio 1. Evolución de la función objetivo vs. ciclo RTO....	135
<b>Figura 4-8 .</b> Caso de estudio 2. Diagrama del sistema de calor y potencia. ....	138
<b>Figura 4-9.</b> Caso de estudio 2, escenario 1. (a) Evolución del costo operativo. (b) Evolución del uso de la postcombustión.....	144

<b>Figura 4-10.</b> Caso de estudio 2, escenario 2. (a) Evolución del costo operativo. (b) Perfiles de precio eléctrico, demanda de vapor, temperatura ambiente y demanda de potencia. ....	145
<b>Figura 4-11.</b> Caso de estudio 2, escenario 2. (a) Evolución de la compra de electricidad. (b) Uso de la postcombustión. ....	145
<b>Figura 5-1.</b> Estructura completa propuesta para el sistema integrado SO+RTO. .	160
<b>Figura 5-2.</b> Diagrama del sistema de calor y potencia estudiado. ....	161
<b>Figura 5-3.</b> Pronósticos y valores reales para precio de la electricidad, demandas y temperatura ambiente. ....	163
<b>Figura 5-4.</b> Costo operativo por período. ....	166
<b>Figura 5-5.</b> Producción acumulada de NOx y límite diario. ....	166
<b>Figura 5-6.</b> Exceso de vapor venteadado a la atmósfera. ....	167
<b>Figura 5-7.</b> Número de turbinas de contrapresión en marcha .....	167
<b>Figura 5-8.</b> Turbinas de contrapresión en operación por período en la implementación <i>SO-RTO</i> . ....	168
<b>Figura 5-9.</b> Calderas en operación por período para las dos estrategias. ....	168
<b>Figura 5-10.</b> Producción de electricidad en la turbina de gas y potencia máxima. ....	169
<b>Figura 5-11.</b> Uso de la postcombustión por período para cada estrategia. ....	169
<b>Figura 6-1.</b> Resolución en serie – Inicialización. ....	178
<b>Figura 6-2.</b> Resolución en serie – Reacción a las infactibilidades (ejemplo). ....	178
<b>Figura 6-3.</b> Resolución en paralelo. ....	179
<b>Figura 6-4.</b> Combinación serie-paralelo. ....	180
<b>Figura 6-5.</b> Estructura <i>SO+RTO</i> con aprovechamiento del tiempo ocioso entre ejecuciones de <i>RTO</i> . ....	182

## ÍNDICE DE TABLAS

<b>Tabla 2-1.</b> Evolución de un caso de ejemplo con errores estructurales de modelado. ....	21
<b>Tabla 3-1.</b> Parámetros utilizados en las alternativas de RTO del Caso de Estudio 1. ....	74
<b>Tabla 3-2.</b> Costo de Diseño Extendido ( $C_E$ ) para las estrategias de adaptación implementadas en el caso de estudio 1. ....	80
<b>Tabla 3-3.</b> Nomenclatura utilizada en el caso de estudio 2. ....	83
<b>Tabla 3-4.</b> Caso de estudio 2. Parámetros operativos fijos. ....	84
<b>Tabla 3-5.</b> Caso de estudio 2. Condiciones operativas iniciales. ....	84
<b>Tabla 3-6.</b> Caso de estudio 2. Variables medidas y variancia de las mediciones. ....	88
<b>Tabla 3-7.</b> Caso de estudio 2. Costo de diseño extendido ( $C_E$ ) ....	96
<b>Tabla 4-1.</b> Caso de estudio 1. Funcionalidad de las salidas y para la planta real y el modelo RTO. ....	127
<b>Tabla 4-2.</b> Caso de estudio 1. Costo de diseño extendido ( $C_E$ ) ....	136
<b>Tabla 4-3.</b> Caso de estudio 2. Ecuaciones de eficiencia para la planta real y el modelo RTO. ....	139
<b>Tabla 4-4.</b> Caso de estudio 2. Variables de entrada iniciales. ....	141
<b>Tabla 4-5.</b> Caso de estudio 2. Especificaciones para el escenario 1. ....	142
<b>Tabla 4-6.</b> Caso de estudio 2. Desviación estándar de las mediciones en el escenario 2. ....	142
<b>Tabla 4-7.</b> Caso de estudio 2. Costo de diseño extendido ( $C_E$ ) y Costo de diseño extendido relativo ( $C_E^{rel}$ ) ....	146
<b>Tabla 5-1.</b> Diferencias entre el modelo aproximado (RTO) y la <i>planta real</i> . $F$ : caudal de vapor generado en las calderas (t/h); $Q$ : uso de combustible en calderas, turbina de gas y postcombustión (Gcal/h); $\theta$ : período del horizonte de optimización. ....	162
<b>Tabla A- 1.</b> Nomenclatura completa del modelo. ....	203
<b>Tabla A- 2.</b> Límites en las variables operativas (LI: Límite inferior; LS: Límite superior) ....	204
<b>Tabla A- 3.</b> Matriz de incidencias (Elementos $a_{ij}$ : fila i: corriente de vapor (según <b>Figura 3.9</b> ) columna j: nodo; 1: entra al nodo, -1: sale del nodo). ....	208
<b>Tabla A- 4.</b> Conexiones (fila: corriente de vapor; columna: equipo; 1: corriente de entrada, -1: corriente de salida)(ver <b>Figura 3.9</b> ) ....	210
<b>Tabla B- 1.</b> Nomenclatura completa del modelo. ....	211

---

<b>Tabla B- 2.</b> Límites en las variables operativas (LI: Límite inferior; LS: Límite superior) .....	212
<b>Tabla B- 3.</b> Matriz de incidencias (Elementos $a_{ij}$ : fila $i$ : colector de vapor (según <b>Figura 4-8</b> ), columna $j$ : equipo; 1: entra al colector, -1: sale del colector).....	216
<b>Tabla B- 4.</b> Temperatura ambiente (T), precio de compra de electricidad ( $C_{ee}$ ), demanda de vapor ( $D_v$ ) y demanda de potencia ( $D_{ee}$ ) para cada ciclo de RTO. ..	217
<b>Tabla C- 1.</b> Nomenclatura del del modelo. ....	219
<b>Tabla C- 2.</b> Temperatura ambiente ( $^{\circ}\text{C}$ ). Pronóstico y valor real para cada período $\theta$ . ....	222
<b>Tabla C- 3.</b> Precio de compra de electricidad ( $\$/\text{kW}$ ). Pronóstico y valor real para cada período $\theta$ . ....	223
<b>Tabla C- 4.</b> Demanda de vapor (t/h). Pronóstico y valor real para cada período $\theta$ . ....	224
<b>Tabla C- 5.</b> Demanda de potencia eléctrica (kW). Pronóstico y valor real para cada período $\theta$ . ....	225

# INTRODUCCIÓN

## RESUMEN

Este capítulo presenta la optimización en tiempo real (RTO), justificando su uso en la toma de decisiones en la industria e introduciendo brevemente sus características principales (sección 1.1). Se mencionan algunas de las principales problemáticas y aspectos sin resolver en esta disciplina (sección 1.2). Se establecen los principales objetivos de este trabajo de tesis en respuesta a los problemas planteados (sección 1.3). Se presenta la organización de los restantes capítulos del trabajo (sección 1.4). Finalmente, se mencionan los trabajos y publicaciones en los que se adelantaron los resultados y conclusiones presentados en esta tesis (sección 1.5)

## 1.1. OPTIMIZACIÓN EN TIEMPO REAL

### 1.1.1. Necesidad de la optimización en tiempo real

La operación eficiente de los procesos industriales es un requerimiento cada vez más importante en la actualidad. El costo de la energía y las materias primas, la incertidumbre de los mercados y la reciente crisis de la economía global obligan a la industria a optimizar sus procesos para maximizar los márgenes de beneficio y mantener la rentabilidad.

La optimización de procesos debe hacer frente a diferentes restricciones. Las tecnologías disponibles y los materiales utilizados permiten trabajar dentro de ciertos márgenes operativos; el tamaño de los equipos y las cañerías limitan la capacidad máxima y mínima de producción. La conciencia social sobre el cuidado del medio ambiente ha crecido en las últimas décadas, y esto se ve reflejado en una legislación ambiental más exigente. Cada vez son más requeridas las certificaciones de calidad relativas a la eficiencia energética y el respeto por el ambiente; para alcanzarlas, la industria se impone restricciones adicionales. Por otra parte, con el objetivo de facilitar el mantenimiento de los equipos y el manejo de los procesos por parte del personal, se suelen fijar reglas operativas que limitan aún más la región factible de operación del proceso.

Con frecuencia, el valor óptimo de las condiciones de operación varía a lo largo del tiempo, ya que el proceso se enfrenta a escenarios cambiantes. La composición y calidad de la materia prima no es la misma en todo momento; los caudales y el nivel de producción pueden cambiar varias veces en el día; el costo

de la energía eléctrica puede variar a lo largo de la jornada, desde casos simples con precios en horas pico, llano y valle a mercados totalmente desregulados donde el precio eléctrico no puede conocerse *a priori*, y es actualizado en base a la oferta y demanda con una frecuencia de pocos minutos. Las condiciones ambientales (temperatura, presión atmosférica, humedad) pueden afectar el desempeño de los equipos y las condiciones de entrada. Otros eventos, planeados o inesperados, pueden modificar el modo de operación de la planta.

En ciertos procesos, el control automático convencional o avanzado es suficiente para reaccionar a los cambios inesperados en las condiciones de operación, garantizando la factibilidad de la operación. Pero en otros procesos industriales, además de las variables controladas necesarias para la operación estable y segura, existen más variables independientes que pueden modificarse dentro de una región factible. En estos casos, estos grados de libertad adicionales representan una oportunidad para minimizar costos, maximizar la producción u operar en forma eficiente y amigable con el medio ambiente. Más aún, pueden determinar la factibilidad económica del proceso a optimizar. En períodos de recesión económica, con frecuencia aparecen más grados de libertad ya que muchas instalaciones industriales reducen la producción, operando en condiciones alejadas de la capacidad máxima de los equipos (Mitra et al., 2013).

En algunos casos, es posible prever con anticipación los escenarios que pueden presentarse, de manera tal de planificar la operación para cada instante. Sin embargo, otras situaciones solo pueden pronosticarse en forma aproximada, por lo que las condiciones de trabajo seleccionadas pueden estar lejos de las realmente óptimas.

Por otra parte, los modelos utilizados para estimar la operación óptima no pueden predecir con total exactitud el comportamiento del proceso. A menudo los modelos de planificación requieren simplificaciones para disminuir el costo computacional de los cálculos. Incluso cuando se cuenta con modelos más sofisticados, los parámetros de las ecuaciones podrían variar en forma inesperada debido al ensuciamiento de los equipos, la inactivación de catalizadores, el desgaste de los materiales o las condiciones ambientales.

En la actualidad, los sistemas de información de planta proveen valores en línea de las mediciones existentes y acceso a bases de datos con valores históricos de estas mismas mediciones. Esta información puede utilizarse para actualizar los modelos disponibles, con el fin de mejorar las predicciones de estos modelos para la situación actual de la planta.

Aprovechando este recurso, y respondiendo a la problemática que se describe en los párrafos anteriores, surge una técnica de la informática aplicada a la ingeniería de procesos. Se trata de la *optimización en tiempo real* (RTO por sus siglas en inglés).

### 1.1.2. Características generales de la optimización en tiempo real (RTO)

Las estrategias de optimización en tiempo real (RTO) comparten una característica distintiva: usar las mediciones disponibles para optimizar la situación actual de la planta (Chachuat et al., 2009). Como consecuencia, para que el resultado de la optimización sea válido, el tiempo de cálculo requerido debe estar en el orden de los segundos o pocos minutos. Los resultados de la optimización se aplican en la planta, y luego de un determinado tiempo, se vuelven a tomar mediciones para un nuevo ciclo de optimización.

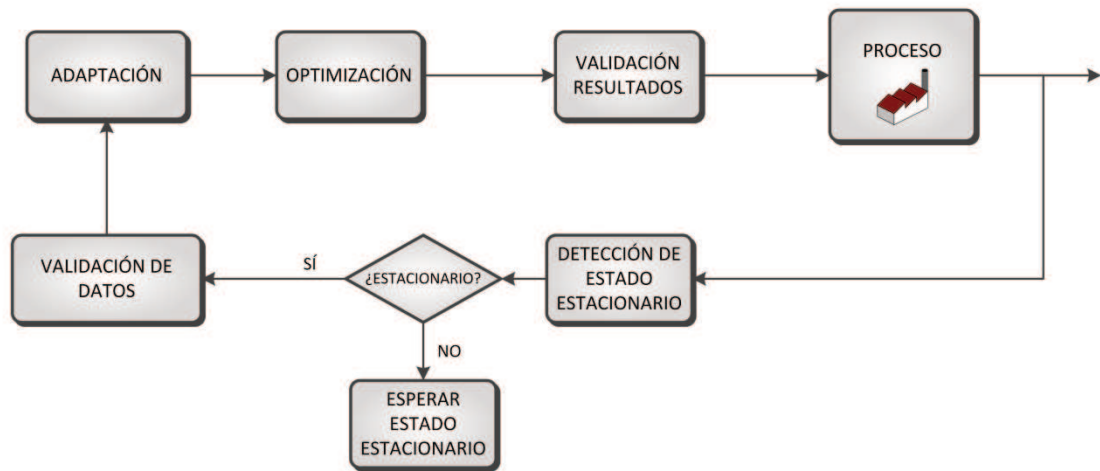
De acuerdo con publicaciones recientes (Darby et al., 2011; Mansour y Ellis, 2008), una implementación típica de RTO consta de los siguientes pasos:

1. *Detección de estado estacionario*. La RTO generalmente utiliza modelos de estado estacionario. Por este motivo, el primer paso de la técnica consiste en verificar si, en el instante actual, las variables medidas en la planta presentan valores aproximadamente constantes que pueden asimilarse a un estado estacionario. Las técnicas están basadas en análisis estadístico de las mediciones (Bhat y Saraf, 2004; Cao y Rhinehart, 1995; Roux et al., 2008).
2. *Validación de datos*. Esta etapa puede incluir simplemente en fijar límites aceptables de los valores medidos (Darby et al., 2011). Otra alternativa más compleja, factible cuando se cuenta con redundancia en las mediciones, consiste en utilizar técnicas de reconciliación de datos y detección de errores gruesos (Narasimhan y Jordache, 1999; Romagnoli y Sanchez, 1999). Estas técnicas proveen el valor más probable de las variables de la planta de acuerdo con las mediciones disponibles.
3. *Adaptación del modelo*. Aun cuando se utilizan modelos rigurosos, es inevitable la presencia de *errores paramétricos* (los parámetros utilizados en el modelo no son los correspondientes al comportamiento actual de la planta) y *estructurales* (las ecuaciones del modelo no pueden representar fielmente el comportamiento de la planta en toda la región factible con ningún conjunto de parámetros). La presencia de estos errores puede determinar que el modelo prediga resultados que en la práctica resultan subóptimos o infactibles. La etapa de adaptación intenta resolver o minimizar estos errores, y es la que caracteriza a las distintas

implementaciones de RTO (Chachuat et al., 2009). Entre las estrategias utilizadas, se pueden mencionar las siguientes:

- *Estrategia de dos pasos* (Chen y Joseph, 1987), que consiste en ajustar los parámetros del modelo minimizando el error de las mediciones. Este método asegura la factibilidad, pero rara vez garantiza que se alcance el óptimo real del sistema.
  - *Adaptación por modificadores* (Marchetti et al., 2009), que consiste en corregir el error en el valor actual y en el gradiente del costo y de las restricciones. Esta estrategia presenta una propiedad interesante: si el sistema converge a un punto luego de un cierto número de ciclos de RTO, el punto alcanzado cumple las condiciones de mínimo de Karush-Kuhn-Tucker del sistema real.
  - *Adaptación directa de las entradas*: Esta serie de técnicas no resuelve un problema de optimización en cada ciclo de RTO, sino que implementa una estrategia de control *feedback* donde la consigna (*set-point*) de las variables controladas garantiza una operación óptima. Entre las técnicas existentes se pueden mencionar las de *NCO tracking* (donde *NCO* hace referencia a las condiciones necesarias de óptimo, *necessary conditions of optimality*) (François et al., 2014; Srinivasan et al., 2008) y *extremum-seeking control* (Adetola y Guay, 2007; Guay y Zhang, 2003). En ambas estrategias, las variables *controladas* son las restricciones activas y el gradiente de la función objetivo aumentada (Lagrangiano); la consigna es mantener todas estas variables en 0 para cumplir con las condiciones necesarias de óptimo de Karush-Kuhn-Tucker.
4. *Optimización de la función objetivo*. Excepto en las estrategias de adaptación directa de las entradas, debe resolverse un problema numérico de optimización con restricciones. En general, se formula un problema de programación no lineal (NLP) con restricciones. La función objetivo a optimizar puede ser el costo operativo (minimizar), la producción, el beneficio o algún indicador de eficiencia o de calidad (maximizar).
  5. *Validación de los resultados de la optimización*. Antes de aplicar los resultados en la planta, puede realizarse una validación de los resultados, analizando la magnitud de los cambios propuestos y el efecto en la función objetivo (Zhang et al., 2001). Por otra parte, es común aplicar parcialmente los resultados obtenidos, aplicando una técnica conocida como filtrado de las entradas (*input filtering*) (Brdys y Tatjewski, 2005; Bunin et al., 2011, 2013a).

La **Figura 1-1** muestra un esquema de la estructura tradicional de RTO.



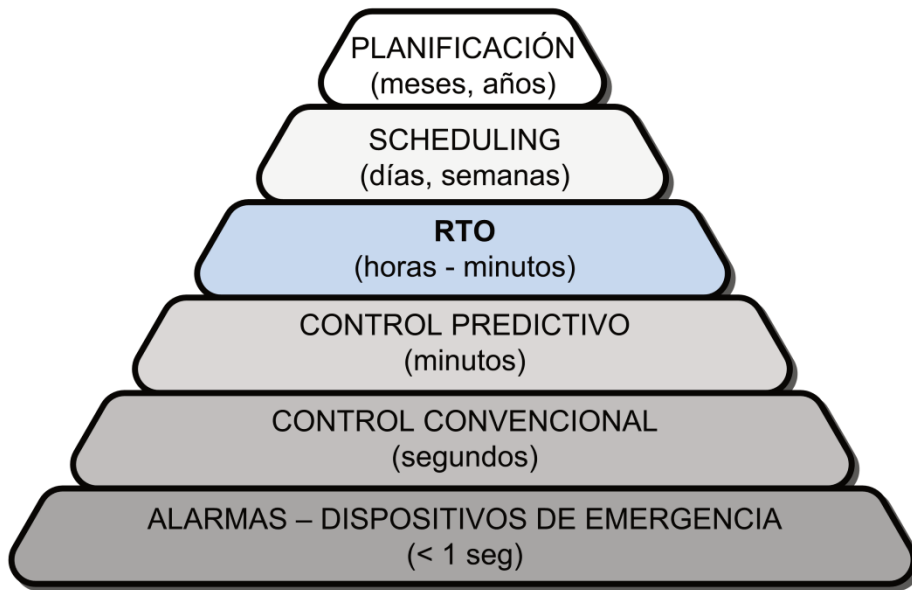
**Figura 1-1.** Estructura de un ciclo de optimización en tiempo real.

El estado del arte y las alternativas existentes para la optimización en tiempo real se analizan con mayor detalle en el capítulo 2 de esta tesis.

### 1.1.3. Estructura jerárquica de optimización industrial. Integración con otras capas.

La toma de decisiones en la operación de una planta industrial pueden organizarse en forma jerárquica (Harjunkoski et al., 2014; Seborg et al., 2010), tal como se muestra en la **Figura 1-2**. A medida que se asciende en la estructura, la frecuencia con la que se ejecuta cada etapa disminuye. Las decisiones tomadas en una etapa superior son adoptadas por las etapas inferiores.

La planificación de la producción, ejecutada con una frecuencia del orden de meses o años, define estrategias y niveles de actividad en la planta. La programación de la producción (*scheduling*), en el orden de las semanas o días, organiza las tareas necesarias para llevar a cabo el plan. La optimización en tiempo real aprovecha los grados de libertad que quedan disponibles para minimizar el costo operativo o maximizar otro indicador. Los valores óptimos de las variables independientes se envían como objetivos a la capa de control predictivo multivariable (MPC), o, en caso de no existir esta capa, como consignas al sistema de control convencional. El control, en el orden de los segundos, es el encargado de mantener la estabilidad y la seguridad de los procesos. Finalmente, las alarmas y los dispositivos de emergencia funcionan como un respaldo para garantizar la seguridad en caso de eventos accidentales.



**Figura 1-2.** Estructura jerárquica de la toma de decisiones en la industria.

Las distintas capas de la estructura suelen utilizar modelos diferentes en cuanto al grado de detalle (más o menos rigurosos) y a la estructura (lineales, no lineales). Esto puede llevar a contradicciones en los resultados de la optimización, ya que el óptimo predicho por una capa podría resultar infactible o subóptimo para la capa inferior. En muchos casos, la capa de RTO es la que resuelve un modelo más riguroso, puesto que es frecuente que tanto el scheduling como el control predictivo recurran a modelos lineales. Para salvar estas contradicciones, y aprovechando la evolución en los algoritmos y la capacidad de procesamiento, se han propuesto estrategias que integran la capa de RTO estacionario con la de MPC (Adetola y Guay, 2010; Alamo et al., 2014; Souza et al., 2010; Zanin et al., 2002).

## 1.2. PROBLEMAS ACTUALES DE LA RTO

### 1.2.1. Complejidad de los métodos de estimación del gradiente

Cuando las estrategias de optimización en tiempo real tienen como objetivo explícito alcanzar el óptimo real de la planta a pesar de los inevitables errores estructurales del modelo, surge la necesidad de estimar en forma experimental (es decir, utilizando mediciones de la planta) los gradientes del costo y de las restricciones. De esta manera, si el sistema converge a un punto operativo y se predicen con exactitud los gradientes y las restricciones activas, se cumplirán las condiciones de óptimo del sistema real. En las estrategias de adaptación por modificadores, el gradiente experimental se utiliza para corregir el error del modelo. En las de adaptación directa de las entradas, en cambio, se evalúa para calcular la diferencia entre los gradientes actuales y la consigna, y de esta manera

aplicar la ley de control *feedback* que intenta mantener el sistema en condiciones óptimas de operación.

La estimación experimental de los gradientes no es un problema trivial. Existe una gran variedad de técnicas para estimar gradientes (Bunin et al., 2013b; Bunin y François, 2012; François y Bonvin, 2014; Mansour y Ellis, 2003; Srinivasan et al., 2011). Todas estas técnicas requieren de datos actuales y pasados de la operación de la planta; estos datos deben contener información suficiente para evaluar las derivadas parciales en todas las direcciones del espacio de las *entradas* (variables independientes del sistema). El error en las mediciones debido al ruido (variancia inherente a los sensores), se amplifica al calcular las derivadas parciales en forma numérica.

Para la optimización en tiempo real, varios autores han propuesto una estrategia *dual* (Brdys y Tatjewski, 2005; Marchetti et al., 2010; Marchetti, 2013). En este caso, los gradientes se estiman a partir de datos de las mediciones actuales y de los estados estacionarios alcanzados en ciclos anteriores de RTO. Esta estrategia agrega restricciones adicionales al problema de RTO, con el fin de mantener acotado el error del cálculo del gradiente utilizando el punto óptimo en el siguiente ciclo de RTO. El calificativo *dual* proviene de estas restricciones adicionales, que implican que la optimización en tiempo real tiene dos objetivos: maximizar el indicador de desempeño de la planta y generar información útil para continuar con la RTO en ciclos siguientes.

La estrategia dual presenta una desventaja que dificulta su aplicación a gran parte de los sistemas reales: a medida que aumenta el número de entradas, mayor es la cantidad de ciclos de RTO pasados necesarios para estimar los gradientes. Aunque en teoría esto no significa un inconveniente, en la práctica es posible que el sistema se vea afectado por perturbaciones no medidas, que afectan las variables de salida del sistema y que invalidan los gradientes estimados. Cuanto mayor es el número de ciclos pasados utilizados, mayor es la probabilidad de que estas perturbaciones estén presentes (Marchetti et al., 2010).

Por este motivo, es necesario modificar esta estrategia, con el fin de resolver el problema de la cantidad de ciclos de RTO necesarios para la estimación del gradiente. Es deseable también desarrollar herramientas para reducir el impacto del ruido en las mediciones.

### 1.2.2. Ausencia de un planteo formal para problemas MINLP o GDP

Tradicionalmente, la optimización en tiempo real ha utilizado formulaciones de programación no lineal (NLP) (Bunin et al., 2012b; Chachuat et al., 2009; Chen y Joseph, 1987; Forbes y Marlin, 1996). Esto excluye la utilización

de variables discretas o restricciones lógicas, que requieren de una formulación de programación no lineal mixta entera (MINLP) o programación disyuntiva generalizada (GDP) (Raman y Grossmann, 1994).

En primer lugar, la ausencia de variables discretas en los modelos de RTO se debe simplemente a que en muchos casos no son necesarias. Gran parte de las posibles variables binarias están relacionadas con la puesta en marcha y parada de equipos o unidades. En general no es conveniente tomar estas decisiones en la etapa de RTO, que suele utilizar una formulación de un solo período para resolver la situación actual. Por este motivo, la puesta en marcha y parada de equipos se reserva para la etapa de *scheduling*.

Sin embargo, existen otros fenómenos, situaciones y eventos que requieren un modelado con disyunciones lógicas y que pueden abordarse en la etapa de optimización en tiempo real. Entre otros, pueden mencionarse algunos contratos complejos de provisión de electricidad o combustible, la presencia de regiones factibles desconectadas entre sí, el modelado con funciones a trozos y las restricciones adicionales de RTO dual (Serralunga et al., 2014).

Por otra parte, el uso de formulaciones continuas obedece a una razón práctica. La RTO requiere la obtención de una solución en un tiempo computacional reducido. La mayor complejidad de los problemas MINLP o GDP puede transformarlos en impracticables en la escala de tiempo de la RTO.

La capacidad de procesamiento de la información ha evolucionado rápidamente en los últimos años, tanto por el incremento en la velocidad de los procesadores como por los avances en los algoritmos de procesamiento paralelo. Al mismo tiempo, los resolvers MINLP y GDP han alcanzado una madurez y se mejoran día a día (GAMS Development Corporation, 2008; Vecchietti y Grossmann, 1999). Por este motivo, problemas con cientos o miles de variables continuas y restricciones, y con decenas de disyunciones lógicas, podrían resolverse dentro de la escala de los minutos, lo que es aceptable para una implementación de RTO.

La industria, de hecho, ha implementado sistemas de RTO con variables discretas para sistemas de servicios auxiliares (vapor y electricidad) (Puranik et al., 2013; Ruiz et al., 2005). Estas implementaciones no tienen como objetivo explícito lograr la convergencia al óptimo real de la planta; incluyen el arranque/parada de equipos, y no están directamente conectadas al sistema de control, sino que requieren de la aplicación manual por parte del operador (este tipo de implementaciones se conoce como *RTO en lazo abierto*).

Por todas estas razones, un planteo formal de la optimización en tiempo real con disyunciones lógicas orientado a la convergencia al óptimo real de la planta ampliaría el campo de aplicación de la RTO, y agregaría grados de libertad a las aplicaciones ya existentes.

### **1.2.3. Integración de RTO con *scheduling* de procesos – Formulaciones multiperíodo**

Como se ha mencionado en la sección 1.1.3, existen propuestas para la integración de la RTO con el control (capa inmediata inferior en la jerarquía de toma de decisiones). Sin embargo, durante el trabajo de revisión bibliográfica para esta tesis, no se han encontrado intentos de integración con el *scheduling* (capa inmediata superior en la estructura de la **Figura 1-2**).

Algunas aplicaciones de *scheduling* utilizan estrategias propias de la RTO. En algunos casos, ante la presencia de eventos disruptivos, se utiliza la información disponible en la planta para recalcular el schedule, intentando garantizar ante todo la factibilidad del programa de producción (Henning y Cerdá, 2000). En otros casos, también se busca alcanzar el óptimo, actualizando el schedule después de cada período usando una estrategia de horizonte deslizante (Castro et al., 2011). Sin embargo, tanto por los modelos utilizados (mixto-enteros lineales) como por los tiempos de solución informados, esta estrategia está fuera de la órbita de la RTO.

Por otro lado, más allá de una aplicación de RTO continua para *blending* de la industria petroquímica (Singh et al., 2000), la optimización en tiempo real se aplica para un solo período (sólo un estado estacionario). El uso de formulaciones multiperíodo ampliaría el campo de aplicación de la RTO, ya que permitiría incluir términos de acumulación, además de permitir una optimización adecuada de la puesta en marcha/parada de equipos y unidades, teniendo en cuenta las *restricciones de transición* (Mitra et al., 2013) que imponen tiempos máximos y mínimos para los tiempos de parada, operación y arranque de los equipos.

Una formulación GDP multiperíodo para la optimización en tiempo real abriría las puertas a una posible integración con el *scheduling* de la operación de procesos continuos. La utilización del mismo modelo en las etapas de RTO y *scheduling* permitiría eliminar los conflictos de factibilidad y optimalidad que surgen al intentar aplicar la solución del *scheduling* en los modelos más rigurosos de la optimización en tiempo real.

#### 1.2.4. Infactibilidades y resultados subóptimos

La optimización en tiempo real requiere, en un tiempo limitado, obtener una solución factible para implementarse a través del sistema de control o, en los sistemas que trabajan en modo de recomendación (*lazo abierto*), a través de las acciones manuales de los operadores. Es posible que esto no se consiga en algunos ciclos de RTO, bien porque el algoritmo fracasa (no converge a una solución factible), o bien porque debido a la complejidad del modelo y a los datos que inicializan el algoritmo en esos ciclos, el modelo no alcanza una solución en el tiempo máximo aceptable para el cálculo.

Si se implementaran formulaciones GDP y/o multiperíodo, como se menciona en las secciones 1.2.2 y 1.2.3, la frecuencia de estas infactibilidades aumentaría en consonancia con el incremento en la dificultad para resolver los problemas planteados. Una implementación robusta de RTO debe incluir alguna estrategia para *reaccionar* a esta problemática y proponer una solución factible en la mayor cantidad de casos posibles.

Por otro lado, es frecuente la utilización de modelos no lineales que pueden presentar no convexidades. La mayoría de los algoritmos de optimización buscan óptimos locales, por lo que en principio no hay garantía de que la solución alcanzada corresponda al óptimo global del problema.

El tiempo, visto en general como una limitación en el ámbito de la RTO, puede a su vez visualizarse como una oportunidad. En primer lugar, el tiempo requerido para alcanzar una solución podría ser suficiente para obtener más de una solución. Ya sea procesando diferentes algoritmos en paralelo o en serie, en muchos casos se podría contar con diferentes conjuntos *óptimos* de entradas para aplicar en la planta (por ejemplo, una solución de un problema MINLP, otra NLP y otra obtenida a través de un heurístico). Utilizando el mismo modelo para evaluar el resultado obtenido, se podría seleccionar la solución más apropiada de este conjunto de candidatos. Si alguno de los algoritmos fracasa, simplemente se descarta esta solución.

En segundo lugar, se dispone de una gran cantidad de tiempo ocioso entre los ciclos de RTO. Una vez que se ha resuelto el problema de optimización y se lo ha aplicado en la planta, pueden transcurrir varios minutos, o incluso varias horas, hasta el inicio del siguiente ciclo de RTO. Si se cuenta con un ordenador dedicado al sistema de RTO, este tiempo puede aprovecharse para generar puntos de inicialización de los algoritmos en el próximo ciclo; por ejemplo utilizando algoritmos de optimización global, o finalizando la ejecución de los problemas con variables discretas en los casos en los que no hayan terminado durante la etapa de

optimización del ciclo anterior. También podría utilizarse para generar reglas o puntos de inicialización en base a resultados obtenidos en ciclos pasados.

### 1.3. OBJETIVOS DE LA TESIS

Esta tesis abordará las problemáticas presentadas en la sección 1.2. Para ello se plantean los siguientes objetivos generales y específicos:

#### 1.3.1. Reducir el problema de estimación del gradiente

- Generar una implementación alternativa de la adaptación por modificadores utilizando información de ciclos de RTO anteriores, que requiera de una menor cantidad de conjuntos de datos para estimar los gradientes.
- Proponer una estructura de modelado que permita concentrar los errores estructurales y paramétricos en un subconjunto de ecuaciones del modelo, mientras el resto se expresa como balances rigurosos de materia, energía y entropía.
- Aprovechar el conocimiento que se tiene de la planta o proceso para identificar las direcciones del espacio de las entradas en las que el gradiente se hace nulo, con el fin de evitar el cálculo experimental en esas direcciones. Esto permitirá, por un lado, alcanzar el objetivo de reducir el número de conjuntos de datos necesarios para estimar el gradiente; por otro lado, disminuirá el impacto negativo del ruido en la estimación.
- Simplificar o relajar las restricciones necesarias para una implementación dual de RTO.
- Evaluar la estrategia propuesta a través de casos de estudio, comparándola con otras implementaciones existentes.

#### 1.3.2. Extender el problema de RTO a formulaciones GDP

- Generar una formulación de programación disyuntiva generalizada (GDP) para el problema de RTO.
- Describir las potenciales aplicaciones de RTO con disyunciones para problemas que involucren un solo período.
- Reformular las estrategias de adaptación disponibles para un sistema de RTO con variables discretas.

- Revisar las prácticas usuales en la RTO de problemas continuos. Redefinirlas en los casos en que sea necesario, para que no afecten el desempeño de sistemas que involucran disyunciones.
- Evaluar las estrategias propuestas a través de casos de estudio.

### **1.3.3. Integrar RTO y *scheduling* con formulaciones multiperíodo**

- Generar una estrategia para integrar la etapa de scheduling para procesos continuos y la de optimización en tiempo real, utilizando el mismo modelo.
- Establecer una jerarquía de variables de decisión para asignar al scheduling y a la RTO, con el fin de reducir el tamaño del problema de optimización que se resuelve en la escala de tiempo de la RTO.
- Identificar sistemas reales en los que la integración propuesta presenta utilidad.
- Proponer una formulación multiperíodo con disyunciones para la optimización en tiempo real.
- Evaluar las estrategias propuestas a través de casos de estudio.

### **1.3.4. Aprovechar el tiempo y los recursos computacionales disponibles**

- Proponer estrategias para disminuir el efecto de las soluciones no factibles en los sistemas de RTO.
- Plantear esquemas de resolución de algoritmos en serie o en paralelo para aprovechar el tiempo máximo que puede destinarse a la búsqueda del óptimo en RTO.
- Enumerar alternativas para la utilización del tiempo ocioso entre ciclos de RTO, incluyendo algoritmos de optimización global y generación de heurísticos para la optimización de ciclos posteriores.
- Especificar los pasos necesarios para implementar las estrategias propuestas.

### **1.3.5. Aplicar las estrategias a sistemas de calor y potencia**

- Justificar la utilidad de la aplicación de la optimización en tiempo real a sistemas de calor y potencia, en base a criterios económicos y a la factibilidad de aplicar la estrategia.

- Desarrollar e implementar modelos de sistemas de calor y potencia para la generación de casos de estudio que evalúan las estrategias de RTO propuestas en la tesis.
- Identificar las fuentes de error estructural en el modelado de sistemas de calor y potencia para obtener una estrategia de adaptación adecuada.

#### 1.4. ORGANIZACIÓN DE LA TESIS

Esta tesis está organizada en siete capítulos, incluyendo este capítulo introductorio.

El **capítulo 2** revisa el estado del arte de la optimización en tiempo real, con especial énfasis en las estrategias de adaptación y las estrategias de filtrado de los resultados. Además, analiza los antecedentes en el uso de las mediciones para actualizar el programa obtenido a través de la optimización del scheduling.

El **capítulo 3** presenta la *adaptación por modificadores con ecuaciones de eficiencia*. Esta estrategia, aplicada a ciertos procesos y con un modelado adecuado, permite alcanzar los objetivos propuestos en la sección 1.3.1. El desempeño de la estrategia se evalúa a través de dos casos de estudio: una planta genérica y el sistema de vapor y generación de electricidad de una planta de producción de azúcar y etanol.

El **capítulo 4** presenta una formulación GDP para la optimización en tiempo real (*RTO con disyunciones*). Se enumeran posibles casos reales que pueden modelarse con disyunciones lógicas. Las estrategias de adaptación más comunes se reformulan para tratar problemas con disyunciones. Algunas prácticas comunes en RTO continua se redefinen para la inclusión de variables discretas. Los casos de estudio incluyen una planta genérica y un sistema de vapor y potencia incluyendo una turbina de gas con caldera de recuperación.

El **capítulo 5** propone un marco conceptual y metodológico que integra el scheduling óptimo y la optimización en tiempo real. Se descompone la jerarquía de decisiones asignando un subconjunto de las variables discretas a la etapa de scheduling y el resto a la RTO. Se propone una formulación multiperíodo con horizonte fijo para el problema de RTO. El caso de estudio analizado corresponde nuevamente a un sistema de vapor y potencia, incluyendo esta vez restricciones a las emisiones.

El **capítulo 6** analiza la problemática de las soluciones no factibles y subóptimas. Se proponen alternativas para aprovechar el tiempo ocioso entre ciclos de RTO y el tiempo máximo asignado a la optimización.

El **capítulo 7** analiza los resultados alcanzados en esta tesis, extrayendo las principales conclusiones y definiendo líneas de trabajo para el futuro.

## 1.5. PUBLICACIONES DE ESTA TESIS

Durante el desarrollo de esta tesis, los resultados parciales alcanzados se dieron a conocer a través de artículos en revistas científicas y se presentaron en congresos y reuniones científicas nacionales e internacionales.

### 1.5.1. Artículos en publicaciones científicas

Serralunga, F.J., Aguirre, P.A., Mussati, M.C., 2014. Including disjunctions in real-time optimization. *Industrial & Engineering Chemistry Research* (en prensa) doi: 10.1021/ie5004619.

Serralunga, F.J., Mussati, M.C., Aguirre, P.A., 2013. Optimización en tiempo real con disyunciones lógicas. Aplicación a sistemas de calor y potencia. *Iberoamerican Journal of Industrial Engineering* 5 (10), pp. 188-202.

Serralunga, F.J., Mussati, M.C., Aguirre, P.A., 2013. Model Adaptation for Real-Time Optimization in Energy Systems. *Industrial & Engineering Chemistry Research* 52, 16795-16810.

### 1.5.2. Trabajos presentados en reuniones científicas

Serralunga, F.J., Aguirre, P.A., Mussati, M.C., 2014. A novel framework for integrating real-time optimization and optimal scheduling. Application to heat and power systems. *43 Jornadas Argentinas de Informática en Investigación Operativa - 3° Simposio de Informática Industrial*. Buenos Aires, Argentina, 3 de septiembre de 2014.

Serralunga, F.J., Mussati, M.C., Aguirre, P.A., 2013. Optimización en tiempo real con disyunciones lógicas. Aplicación a sistemas de calor y potencia. *42 Jornadas Argentinas de Informática en Investigación Operativa - 2° Simposio de Informática Industrial*. Córdoba, Argentina 16-17 de septiembre de 2012; pp. 237-248. ISSN: 2313-9102

Serralunga, F., Mussati, M.C., Aguirre, P.A., 2012. Real-time optimization of energy systems in sugar and ethanol facilities: a modifier adaptation approach. *11th International Symposium on Process Systems Engineering*. Singapur, 15 al 19 de julio de 2012. En: *Computer Aided Chemical Engineering* 31, pp. 375 - 379. ISBN: 978-0-444-59505-8.

Serralunga, F., Mussati, M.C., Aguirre, P.A., 2012. An alternative real-time optimization algorithm with modifier adaptation: Application to heat and power

---

systems. *22nd European Symposium on Computer Aided Process Engineering*. Londres, Reino Unido, 17 al 20 de junio de 2012. En: *Computer Aided Chemical Engineering* 30, pp. 367 - 371. ISBN: 978-0-444-59431-0.



## 2. ESTADO DEL ARTE EN RTO

### RESUMEN

Este capítulo consiste principalmente en una revisión de las técnicas de optimización en tiempo real (RTO), especialmente aquellas basadas en modelos. Se analizan las variantes existentes para la validación de datos, la adaptación del modelo y el tratamiento de los resultados. Se presentan extensiones de la técnica para el caso multiperíodo y para modelos dinámicos. La sección concluye con una descripción de las implementaciones industriales y de usos alternativos de las técnicas de RTO. Finalmente, se enumeran brevemente los métodos de *scheduling* con incertidumbre y las propuestas existentes para la integración del *scheduling* con control de procesos y RTO.

### 2.1. PLANTEO GENERAL DEL PROBLEMA

El objetivo de la optimización en tiempo real es encontrar un vector de entradas  $\mathbf{u}^*$  que resuelva el siguiente problema de optimización (B. Chachuat et al., 2008):

$$\begin{aligned} \mathbf{u}^* \in \underset{\mathbf{u}}{\operatorname{argmin}} Q_p(\mathbf{y}_p, \mathbf{u}) \\ \text{s.a. } \mathbf{y}_p = \mathbf{f}_p(\mathbf{u}) \\ \mathbf{g}_p(\mathbf{y}_p, \mathbf{u}) \leq 0 \end{aligned} \quad (2-1)$$

donde  $\mathbf{u}$  e  $\mathbf{y}_p$  son las entradas (variables independientes) y salidas (variables controladas) del proceso, respectivamente;  $Q_p$  es una función escalar que calcula el costo operativo real del proceso;  $\mathbf{f}_p$  es un conjunto de ecuaciones que representa el comportamiento *real* de la planta, y  $\mathbf{g}_p$  es el vector de restricciones de desigualdad del proceso.

Como la funcionalidad real  $\mathbf{f}_p(\mathbf{u})$  nunca puede ser conocida con total certeza en la práctica, las salidas se estiman a través de un modelo del proceso (Marchetti et al., 2009):

$$\begin{aligned} \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \mathbf{0} \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \end{aligned} \quad (2-2)$$

donde  $y$  es el vector de salidas estimadas;  $x$  son las variables de estado;  $f$  es el conjunto de ecuaciones que conforman el modelo del proceso;  $\alpha$  y  $\beta$  son conjuntos de parámetros fijos y ajustables, respectivamente.

## 2.2. ERRORES ESTRUCTURALES Y PARAMÉTRICOS

Cuando se utiliza un modelo para representar el comportamiento de un proceso, la predicción de las variables de salida estará inevitablemente sujeta a error. Es de esperar que el uso de modelos más rigurosos genere predicciones con mayor exactitud (Yip y Marlin, 2004), pero de cualquier manera el desempeño del proceso estará afectado por perturbaciones no medidas y cambios en la eficiencia de los equipos por ensuciamiento o desgaste, que afectan la fidelidad del modelo a la operación real.

La literatura clasifica los errores del modelo en dos tipos: *paramétricos* y *estructurales* (Chachuat et al., 2009). Ambos tipos de errores están presentes en cualquier implementación de RTO y han llevado al desarrollo de distintas estrategias de adaptación para corregirlos o para alcanzar el óptimo de la planta a pesar de ellos.

Los *errores paramétricos* están relacionados con la identificación del modelo. Si la estructura del modelo es la correcta, las ecuaciones que componen el modelo representan exactamente el comportamiento real de la planta para algún conjunto de valores de los parámetros ajustables. Sin embargo, en la práctica los valores óptimos de los parámetros no pueden identificarse con exactitud, debido a la variabilidad y las desviaciones en las mediciones, la falta de *observabilidad* de algunos parámetros (es decir, de información suficiente para obtenerlos) y la frecuente no convexidad de los problemas de optimización para estimación (Marchetti et al., 2008).

De cualquier manera, los modelos matemáticos siempre representan de manera más o menos aproximada el comportamiento real de la planta, dando lugar a los *errores estructurales*, que están inevitablemente presentes aun en los modelos más detallados, e incluso si los valores óptimos de los parámetros del modelo se pudieran determinar con exactitud.

En el ámbito de la optimización en tiempo real, los errores estructurales pueden causar que, aun conociendo con exactitud los parámetros del modelo, el sistema converja a puntos de operación subóptimos, o incluso no factibles. Cuando se estiman los parámetros del modelo, en general se intenta obtener una predicción exacta de las salidas del sistema, pero no de las condiciones necesarias

de óptimo (gradiente de la función objetivo, valor y gradiente de las restricciones). (Bonvin y François, 2013).

Con el objetivo de eliminar o reducir el impacto del error estructural y paramétrico, los sistemas de optimización en tiempo real implementan diferentes tipos de *estrategias de adaptación* (Benoît Chachuat et al., 2008), que se analizan en la sección 2.3.

Un modelo de optimización en tiempo real es *adecuado* (Forbes, 1994) si, al converger a un punto luego de sucesivas ejecuciones de optimización en tiempo real, este punto cumple con las condiciones necesarias de óptimo de la *planta real*, que, a partir del problema (2-1) pueden expresarse como:

$$\begin{aligned}
 \nabla_{\mathbf{u}} Q_p(\mathbf{u}^*) + (\mathbf{v}^*)^T \cdot \nabla_{\mathbf{u}} g_p(\mathbf{u}^*) &= \mathbf{0} \\
 \mathbf{g}_p(\mathbf{u}^*) &\leq \mathbf{0} \\
 \mathbf{v}^* &\geq \mathbf{0} \\
 (\mathbf{v}^*)^T \cdot \mathbf{g}_p(\mathbf{u}^*) &= 0 \\
 \mathbf{d}^T \cdot \left( \nabla_{\mathbf{u}}^2 Q_p(\mathbf{u}^*) + (\mathbf{v}^*)^T \cdot \nabla_{\mathbf{u}}^2 g_p(\mathbf{u}^*) \right) \cdot \mathbf{d} &\geq 0, \\
 \forall \mathbf{d} \neq \mathbf{0} / \mathbf{d}^T \cdot \nabla_{\mathbf{u}} g_{p,i}(\mathbf{u}^*) &\leq 0, i \in I
 \end{aligned} \tag{2-3}$$

donde el conjunto  $I$  contiene a las restricciones activas:

$$I = \{i / g_{p,i}(\mathbf{u}^*) = 0\} \tag{2-4}$$

y las salidas se expresan implícitamente como funciones de  $\mathbf{u}$ :

$$\begin{aligned}
 Q_p(\mathbf{u}) &= Q_p(\mathbf{f}_p(\mathbf{u}), \mathbf{u}) = Q_p(\mathbf{y}_p, \mathbf{u}) \\
 \mathbf{g}_p(\mathbf{u}) &= \mathbf{g}_p(\mathbf{f}_p(\mathbf{u}), \mathbf{u}) = \mathbf{g}_p(\mathbf{y}_p, \mathbf{u})
 \end{aligned} \tag{2-5}$$

y los vectores  $\mathbf{v}$  son los multiplicadores de Lagrange asociados a las restricciones (Avriel, 2012).

El siguiente ejemplo ilustra el problema de los errores estructurales, que pueden determinar que un modelo no sea *adecuado* para la optimización en tiempo real. Se desea optimizar en tiempo real un proceso de una entrada y una salida. El comportamiento *real* del proceso, desconocido para el sistema de RTO, es el siguiente:

$$y = \left( -1 + 0.5 \cdot u + \frac{3}{u} \right) \cdot u \tag{2-6}$$

y se aproxima a través del modelo

$$y = \theta \cdot u \quad (2-7)$$

donde  $\theta$  es un parámetro ajustable, que se calcula para cada ciclo de optimización  $k$  como:

$$\theta^k = \frac{y^k}{u^k} \quad (2-8)$$

donde  $u^k$  e  $y^k$  son la entrada y la salida observadas en dicho ciclo, respectivamente.

Se toma como punto inicial  $u^0 = 2.7$ . El problema de optimización que se resuelve en cada ciclo  $k$  es:

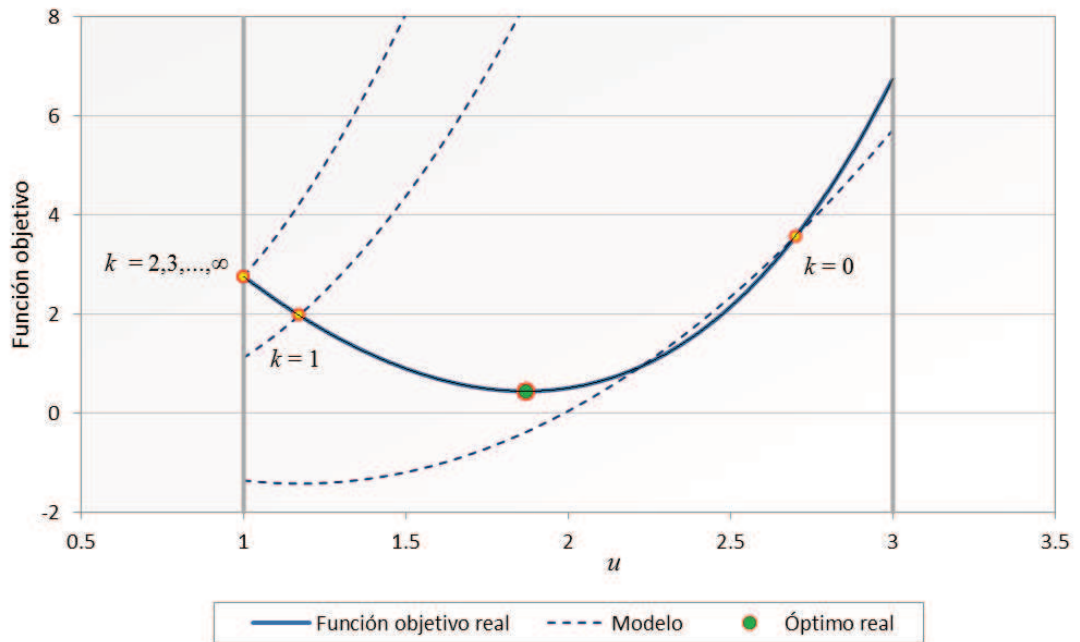
$$\begin{aligned} u^{k+1} = \operatorname{argmin} Q &= 1.5 + y^2 - 5 \cdot u \\ \text{s.a.} \quad y &= \theta^k \cdot u \\ 1 &\leq u \leq 3 \end{aligned} \quad (2-9)$$

Cada ciclo de RTO comprende la solución de las ecuaciones (2-8) y (2-9). El óptimo real del sistema, teniendo en cuenta la funcionalidad de  $y$  a través de la ecuación (2-6) es  $(u^* = 1.870, Q^* = 0.435)$ . Sin embargo, utilizando el modelo aproximado, el sistema converge luego de dos ejecuciones de RTO al punto  $(u^\infty = 1.000, Q^\infty = 2.750)$ . Los resultados obtenidos se muestran en la **Tabla 2-1**.

Puede verificarse que en el punto de convergencia, la salida  $y$  se predice sin error, por lo que a partir del ciclo 3, el valor de la función objetivo predicha para el óptimo alcanzado según la ecuación (2-9) es la función objetivo real. Sin embargo, este punto no es un mínimo para el proceso real; por el contrario, es un máximo local, como puede observarse en la Figura 2-1. La figura muestra la función objetivo real, y las funciones objetivo aproximadas que se obtienen al ajustar el modelo.

De todo esto, puede concluirse que el modelo *no es adecuado* para la optimización de la planta, ya que converge a un punto que no es un óptimo local del proceso real.

El comportamiento mostrado en este ejemplo está presente en mayor o menor medida en las aplicaciones de RTO. Algunas estrategias de adaptación, analizadas en la sección 2.3, tienen en cuenta este fenómeno. Otras, como la *estrategia de dos pasos*, la más utilizada en aplicaciones industriales, no tienen en cuenta este comportamiento y generalmente no pueden garantizar la convergencia al óptimo real del sistema.



**Figura 2-1.** Evolución de un sistema de RTO con errores estructurales en el modelo.

**Tabla 2-1.** Evolución de un caso de ejemplo con errores estructurales de modelado.

$k$	$u^k$	$y^k$	$\theta^k$	$Q^k$
0	2.700	3.945	1.461	3.563
1	1.171	2.515	2.147	1.968
2,3,...,∞	1.000	2.500	2.500	2.750

## 2.3. ESTRATEGIAS DE ADAPTACIÓN

### 2.3.1. Estrategia de dos pasos

La estrategia más utilizada en la optimización en tiempo real es la estrategia de dos pasos (Chen y Joseph, 1987). Su nombre se debe a que en cada ciclo de RTO propone resolver dos problemas de optimización: el primero es un ajuste de parámetros del modelo, y el segundo es la minimización del costo o la maximización de un índice de desempeño utilizando el modelo con los parámetros actualizados en el primer paso.

El problema de adaptación de parámetros (Yip y Marlin, 2002) suele formularse en simultáneo con una reconciliación de datos:

$$\begin{aligned} \beta^k &= \underset{\beta, y, u, x}{\operatorname{argmin}} \left( \begin{matrix} y^{m,k} - y \\ u^{m,k} - u \end{matrix} \right)^T \cdot V^{-1} \cdot \begin{pmatrix} y^{m,k} - y \\ u^{m,k} - u \end{pmatrix} \\ \text{s.a.} \quad & \mathbf{f}(\mathbf{x}, \mathbf{u}, \alpha, \beta) = \mathbf{0} \\ & \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \alpha, \beta) \end{aligned} \quad (2-10)$$

donde  $\mathbf{u}^{m,k}$  e  $\mathbf{y}^{m,k}$  son las entradas y salidas medidas, respectivamente, y  $\mathbf{V}$  es la matriz de variancia-covariancia de las mediciones. En algunos casos, se supone que las entradas están libres de error (Navia et al., 2012); en este caso, el problema (2-10) puede completarse agregando la restricción  $\mathbf{u} = \mathbf{u}^{m,k}$ .

El problema de minimización del costo con los parámetros actualizados resulta:

$$\begin{aligned} \mathbf{u}^{k+1} &= \underset{y, u, x}{\operatorname{argmin}} Q(\mathbf{x}, \mathbf{u}, \mathbf{y}, \alpha, \beta^k) \\ \text{s.a.} \quad & \mathbf{f}(\mathbf{x}, \mathbf{u}, \alpha, \beta^k) = \mathbf{0} \\ & \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \alpha, \beta^k) \\ & \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \alpha, \beta^k) \leq 0 \end{aligned} \quad (2-11)$$

Miletic y Marlin (1998a) sugieren, antes de ejecutar la optimización, realizar un análisis de la calidad de los parámetros estimados, basado en la matriz de variancia-covariancia de estos parámetros con respecto a las variables medidas. Si el análisis indica que la información presente en los datos es insuficiente, los parámetros no se actualizan.

Forbes (1994) ha planteado condiciones para que un modelo adaptado por el método de dos pasos sea adecuado. Chachuat et al. (2009) muestran que, para asegurar la factibilidad de los resultados, el número de parámetros ajustables debe ser mayor o igual al número de salidas  $ny$ . Para alcanzar la optimalidad, en cambio, el sistema debe poder estimar acertadamente tanto las salidas del proceso como las derivadas de estas salidas con respecto a las entradas,  $\frac{\partial y}{\partial u}$ . Por este motivo, en este caso es necesario un número de parámetros ajustables igual a  $ny \cdot (nu + 1)$  (donde  $nu$  es el número de entradas). En la práctica, es raro que un sistema de RTO con la estrategia de dos pasos cumpla con estos requerimientos. Un sistema relativamente pequeño, con 10 entradas y 10 salidas, requeriría de 110 parámetros ajustables para poder predecir exactamente las salidas y las derivadas parciales con respecto a todas las entradas. Además, al aumentar el número de parámetros ajustables, aumenta la incertidumbre con respecto al valor de estos parámetros.

Si bien el problema (2-10) está planteado originalmente para un solo conjunto de variables medidas (es decir, para mediciones tomadas en el mismo instante de tiempo), hay algunos parámetros que requieren de más de un conjunto de datos para ser estimados. Por ejemplo, para estimar la energía de activación y el factor preexponencial de la expresión de una constante de reacción, se requieren al menos dos observaciones a temperaturas diferentes. Por este motivo, se han propuesto implementaciones de RTO con múltiples conjuntos de datos para la estimación de parámetros (Yip, 2002). El trabajo de Yip sugiere dividir los parámetros en tres subconjuntos. El primer conjunto,  $\beta_1$ , incluye los parámetros que podrían ser estimados con un solo conjunto de datos, pero se sabe que cambian lentamente, como la inactivación de un catalizador o un factor de ensuciamiento. En el segundo,  $\beta_2$ , aquellos que requieren más de un conjunto, tales como la pendiente de una recta o la energía de activación. En el tercero,  $\beta_3$ , se agrupan los parámetros que cambian rápidamente y que pueden estimarse con un solo conjunto de datos. El problema (2-10) modificado resulta:

$$\begin{aligned} \beta_1^k, \beta_2^k, \beta_3^k = \arg \min_{\beta, y, u, x} \quad & \sum_{i=k-N-1}^k \begin{pmatrix} \mathbf{y}^{m,i} - \mathbf{y}^i \\ \mathbf{u}^{m,i} - \mathbf{u}^i \end{pmatrix}^T \cdot \mathbf{V}_{k-i}^{-1} \cdot \begin{pmatrix} \mathbf{y}^{m,i} - \mathbf{y}^i \\ \mathbf{u}^{m,i} - \mathbf{u}^i \end{pmatrix} \\ \text{s.a.} \quad & \left. \begin{aligned} \mathbf{f}(\mathbf{x}, \mathbf{u}, \alpha, \beta_1, \beta_2, \beta_3^i) &= \mathbf{0} \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{u}, \alpha, \beta_1, \beta_2, \beta_3^i) \end{aligned} \right\} i = k - N - 1, \dots, k \end{aligned} \quad (2-12)$$

Mientras los parámetros  $\beta_1$  y  $\beta_2$  son los mismos para los  $N$  conjuntos de datos, el parámetro  $\beta_3$  puede tomar un valor diferente para cada conjunto  $i$ . Sin embargo, sólo el valor del ciclo  $k$  se utiliza en el modelo actualizado.

Los datos disponibles podrían no contener información suficiente para estimar todos los parámetros. Por ejemplo, si la variación de la temperatura es pequeña en los datos disponibles, la calidad de la estimación de una energía de activación sería pobre. Por este motivo, en algunos casos es conveniente perturbar algunas variables de la planta para forzar “experimentos” que provean información para estimar los parámetros en forma apropiada (Yip y Marlin, 2003).

### 2.3.2. Adaptación por modificadores

En lugar de ajustar los parámetros del modelo, esta estrategia propone añadir términos de corrección (llamados *modificadores*) para la función objetivo y las restricciones (Gao y Engell, 2005; Marchetti et al., 2009). Los modificadores tienen como objetivo eliminar el error en las restricciones (para que una solución factible en el modelo sea factible en la práctica) y en los gradientes (para que las condiciones de óptimo del modelo coincidan con las de la planta).

Expresando el modelo (2-2) como  $\mathbf{y} = \mathbf{h}(\mathbf{u}, \boldsymbol{\alpha})$  (con los estados  $\mathbf{x}$  y las ecuaciones  $\mathbf{f}$  incluidas en forma implícita, y sin parámetros ajustables  $\boldsymbol{\beta}$ ), el problema de optimización en tiempo real (sin adaptación) puede formularse como:

$$\begin{aligned} \min_{\mathbf{u}} \quad & Q(\mathbf{h}(\mathbf{u}, \boldsymbol{\alpha}), \mathbf{u}) = Q(\mathbf{u}, \boldsymbol{\alpha}) \\ \text{s.a.} \quad & \mathbf{g}(\mathbf{h}(\mathbf{u}, \boldsymbol{\alpha}), \mathbf{u}) = \mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}) \leq 0 \end{aligned} \quad (2-13)$$

En un ciclo de RTO  $k$ , este problema puede ser adaptado como (Marchetti et al., 2010):

$$\begin{aligned} \min_{\mathbf{u}} \quad & Q(\mathbf{u}, \boldsymbol{\alpha}) + \varepsilon^{Q,k} + \left( \boldsymbol{\lambda}^{Q,k} \right)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\ \text{s.a.} \quad & \mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}) + \varepsilon^{g,k} + \left( \boldsymbol{\lambda}^{g,k} \right)^T \cdot (\mathbf{u} - \mathbf{u}^k) \leq 0 \end{aligned} \quad (2-14)$$

donde  $\varepsilon^{Q,k}$  y  $\varepsilon^{g,k}$  son los modificadores del *costo* (función objetivo) y las restricciones, respectivamente, y  $\boldsymbol{\lambda}^{Q,k}$  y  $\boldsymbol{\lambda}^{g,k}$  son los *modificadores del gradiente* del costo y de las restricciones. Estos términos, constantes para cada ciclo de RTO, se definen como:

$$\begin{aligned} \varepsilon^{Q,k} &= Q_p(\mathbf{y}_p^k, \mathbf{u}^k) - Q(\mathbf{u}^k, \boldsymbol{\alpha}) \\ \varepsilon^{g,k} &= \mathbf{g}_p(\mathbf{y}_p^k, \mathbf{u}^k) - \mathbf{g}(\mathbf{u}^k, \boldsymbol{\alpha}) \\ \boldsymbol{\lambda}^{Q,k} &= \nabla_{\mathbf{u}} Q_p \Big|_{\mathbf{u}=\mathbf{u}^k} - \nabla_{\mathbf{u}} Q(\mathbf{u}, \boldsymbol{\alpha}) \Big|_{\mathbf{u}=\mathbf{u}^k} \\ \boldsymbol{\lambda}^{g,k} &= \nabla_{\mathbf{u}} \mathbf{g}_p \Big|_{\mathbf{u}=\mathbf{u}^k} - \nabla_{\mathbf{u}} \mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}) \Big|_{\mathbf{u}=\mathbf{u}^k} \end{aligned} \quad (2-15)$$

El modificador del costo  $\varepsilon^{Q,k}$  no tiene influencia en el cálculo de las entradas óptimas, y generalmente se omite en el problema de optimización, a menos que sea de utilidad conocer el valor exacto de la función objetivo, o que este término se utilice en algún cálculo auxiliar (Marchetti y Basualdo, 2012).

Cuando el sistema converge, es posible ver que el punto alcanzado cumple con las condiciones necesarias de óptimo (de primer orden) de la planta (Chachuat et al., 2009). Las condiciones de óptimo para el modelo son:

$$\begin{aligned}
\nabla Q(\mathbf{u}^*, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{Q, \infty})^T + \mathbf{v}^T \cdot \left( \nabla \mathbf{g}(\mathbf{u}^*, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{g, \infty})^T \right) &= \mathbf{0} \\
\mathbf{g}(\mathbf{u}^*, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}^{g, \infty} + (\boldsymbol{\lambda}^{g, \infty})^T \cdot (\mathbf{u}^* - \mathbf{u}^\infty) &\leq \mathbf{0} \\
\mathbf{v}^T \cdot \left( \mathbf{g}(\mathbf{u}^*, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}^{g, \infty} + (\boldsymbol{\lambda}^{g, \infty})^T \cdot (\mathbf{u}^* - \mathbf{u}^\infty) \right) &= 0 \\
\mathbf{v} &\geq \mathbf{0}
\end{aligned} \tag{2-16}$$

Cuando el sistema ha convergido, el nuevo óptimo es el mismo del ciclo anterior, es decir,  $\mathbf{u}^* = \mathbf{u}^\infty$ . Por lo tanto, teniendo en cuenta (2-15),

$$\begin{aligned}
\nabla Q(\mathbf{u}^*, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{Q, \infty})^T &= \nabla Q(\mathbf{u}^\infty, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{Q, \infty})^T = \nabla Q_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) \\
\nabla \mathbf{g}_i(\mathbf{u}^*, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{g, \infty})^T &= \nabla \mathbf{g}_i(\mathbf{u}^\infty, \boldsymbol{\alpha}) + (\boldsymbol{\lambda}^{g, \infty})^T = \nabla \mathbf{g}_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) \\
\mathbf{g}(\mathbf{u}^*, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}^{g, \infty} + (\boldsymbol{\lambda}^{g, \infty})^T \cdot (\mathbf{u}^* - \mathbf{u}^\infty) &= \mathbf{g}(\mathbf{u}^\infty, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}^{g, \infty} = \mathbf{g}_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty)
\end{aligned} \tag{2-17}$$

y las condiciones de (2-16) pueden reescribirse como

$$\begin{aligned}
\nabla Q_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) + \mathbf{v}^T \cdot \nabla \mathbf{g}_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) &= 0 \\
\mathbf{g}_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) &\leq 0 \\
\mathbf{v}^T \cdot \mathbf{g}_p(\mathbf{y}_p^\infty, \mathbf{u}^\infty) &= 0 \\
\mathbf{v} &\geq \mathbf{0}
\end{aligned} \tag{2-18}$$

que son, precisamente, las condiciones necesarias de óptimo de primer orden de la planta real.

Como las condiciones de óptimo de primer orden de la planta están aseguradas en este método (si el sistema converge), para que el modelo sea adecuado sólo hace falta que se cumplan las condiciones de segundo orden. Marchetti (2013) ha demostrado que la única condición necesaria para que el modelo sea adecuado es:

$$\begin{aligned}
\mathbf{d}^T \cdot \left( \nabla^2 Q(\mathbf{u}^*, \boldsymbol{\alpha}) + \mathbf{v}^T \cdot \nabla^2 \mathbf{g}(\mathbf{u}^*, \boldsymbol{\alpha}) \right) \cdot \mathbf{d} &> 0, \\
\forall \mathbf{d} \neq \mathbf{0} / \mathbf{d}^T \cdot \nabla_{\mathbf{u}} \mathbf{g}_{p,i}(\mathbf{u}^*, \boldsymbol{\alpha}) &\leq 0, i \in I
\end{aligned} \tag{2-19}$$

donde  $I$  es el conjunto de restricciones activas indicado en (2-4).

François et al. (2013) muestra que un modelo estrictamente convexo es siempre adecuado para la adaptación por modificadores, ya que en todos los puntos de la región factible cumple las condiciones dadas por (2-19).

Recientemente, Faulwasser y Bonvin (2014) han extendido la técnica de adaptación incluyendo modificadores de *segundo orden* (de manera de corregir el

error en el valor actual, en el gradiente y en la matriz Hessiana), de manera de replicar las condiciones (2-19) de la planta en el modelo.

Si los modificadores calculados según (2-15) se utilizan directamente, la corrección puede volverse muy sensible al ruido en las mediciones, especialmente en el cálculo de los gradientes. Además, los términos obtenidos serían válidos para el último punto, pero no se tendrían en cuenta los modificadores calculados en puntos anteriores, dando lugar a una corrección agresiva y, probablemente, a resultados de optimización muy diferentes para los distintos ciclos de RTO, aun sin perturbaciones externas. Por este motivo, es conveniente actualizar los parámetros utilizando un *filtro exponencial*:

$$\begin{bmatrix} \varepsilon^{Q,k} \\ \varepsilon^{g,k} \\ \lambda^{Q,k} \\ \lambda^{g_1,k} \\ \vdots \\ \lambda^{g_{ng},k} \end{bmatrix} := \begin{bmatrix} \varepsilon^{Q,k-1} \\ \varepsilon^{g,k-1} \\ \lambda^{Q,k-1} \\ \lambda^{g_1,k-1} \\ \vdots \\ \lambda^{g_{ng},k-1} \end{bmatrix} + \mathbf{K} \cdot \left( \begin{bmatrix} \varepsilon^{Q,k} \\ \varepsilon^{g,k} \\ \lambda^{Q,k} \\ \lambda^{g_1,k} \\ \vdots \\ \lambda^{g_{ng},k} \end{bmatrix} - \begin{bmatrix} \varepsilon^{Q,k-1} \\ \varepsilon^{g,k-1} \\ \lambda^{Q,k-1} \\ \lambda^{g_1,k-1} \\ \vdots \\ \lambda^{g_{ng},k-1} \end{bmatrix} \right) \quad (2-20)$$

donde  $ng$  es el número de restricciones, y  $\mathbf{K}$  es la *matriz de filtrado* (generalmente una matriz diagonal con elementos positivos y menores que 1). Se ha demostrado que el filtro exponencial es de utilidad para la convergencia de un sistema de RTO (Marchetti et al., 2009).

Una formulación alternativa para la adaptación por modificadores corrige el error en las variables de salida  $\mathbf{y}$  (Rodger y Chachuat, 2011):

$$\begin{aligned} \min_{\mathbf{u}} \quad & Q(\mathbf{y}, \mathbf{u}) \\ \text{s.a.} \quad & \mathbf{y} = \mathbf{h}(\mathbf{u}, \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}^k + \left(\boldsymbol{\lambda}^k\right)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\ & \mathbf{g}(\mathbf{y}, \mathbf{u}) \leq 0 \end{aligned} \quad (2-21)$$

donde los modificadores se calculan como:

$$\begin{aligned} \varepsilon^{g,k} &= \mathbf{y}_p^k - \mathbf{h}(\mathbf{u}^k, \boldsymbol{\alpha}) \\ \boldsymbol{\lambda}^k &= \nabla_{\mathbf{u}} \mathbf{y}_p \Big|_{\mathbf{u}=\mathbf{u}^k} - \nabla_{\mathbf{u}} \mathbf{h}(\mathbf{u}, \boldsymbol{\alpha}) \Big|_{\mathbf{u}=\mathbf{u}^k} \end{aligned} \quad (2-22)$$

Si bien en la evolución en un sistema que implementa (2-21) no necesariamente es la misma que al adoptar la estrategia de (2-14), en ambos casos se verifica que, cuando el modelo converge, lo hace a un punto que cumple las condiciones necesarias de óptimo de Karush-Kuhn-Tucker.

En algunas implementaciones de RTO no se incluye la corrección de los gradientes, privilegiando la estabilidad y la sencillez de la técnica a costa de una posible suboptimalidad. Esta estrategia recibe el nombre de *adaptación de las restricciones* (Bunin et al., 2012b; B. Chachuat et al., 2008; Marchetti et al., 2008), y había sido anticipada por Forbes y Marlin (1994).

El uso de términos de corrección del gradiente para permitir que las condiciones de óptimo del modelo y la planta coincidan ya había sido presentado hace más de tres décadas. El algoritmo *ISOPE (Integrated System Optimization and Parameter Estimation – Optimización de sistemas y estimación de parámetros integradas)* (Roberts, 1995, 1979) propone la corrección del gradiente para la función objetivo. Este algoritmo requiere una estimación de parámetros idéntica a la utilizada en la estrategia de dos pasos (Tatjewski, 2002). Algunas versiones más modernas del algoritmo incluyen el agregado de términos cuadráticos para *convexificar* la función objetivo (Brdys y Tatjewski, 2005). En este caso, el problema adaptado podría formularse como:

$$\begin{aligned} \min_{\mathbf{u}} \quad & Q(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}^k) - (\boldsymbol{\lambda}^{Q,k})^T \cdot \mathbf{u} + \rho \cdot (\mathbf{u} - \mathbf{u}^k)^2 \\ \text{s.a.} \quad & \mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}^k) \leq 0 \end{aligned} \quad (2-23)$$

donde  $\boldsymbol{\beta}^k$  son los parámetros ajustables obtenidos de la misma manera que en la estrategia de dos pasos (2-10),  $\boldsymbol{\lambda}^k$  el modificador del gradiente del costo (calculado según (2-15)), y  $\rho$  es un parámetro, seleccionado al momento de diseñar cada implementación particular, que caracteriza al *término de convexificación*  $\rho \cdot (\mathbf{u} - \mathbf{u}^k)^2$ .

### 2.3.3. Estrategias de adaptación directa de las entradas

Algunas estrategias de RTO no resuelven explícitamente un problema de optimización en cada ciclo. En cambio, utilizan un sistema de control (o emulan el funcionamiento de un sistema de control) utilizando un conjunto de variables controladas que, si toman un determinado valor de consigna, aseguran que el sistema está operando en condiciones óptimas (Chachuat et al., 2009).

La estrategia de *Self-Optimizing Control (SOC)* (Jäschke y Skogestad, 2011; Skogestad, 2000a, 2000b) consiste en encontrar combinaciones lineales de las variables de salida del sistema, que estén directamente relacionadas con el óptimo de la planta. Utilizando un modelo, se resuelve una optimización de la planta, fuera de línea, para diseñar el sistema de control. A partir de estos resultados, se buscan combinaciones de las variables de salida que aseguren que la función objetivo tomará el valor óptimo o se mantendrá en valores cercanos. Estas

combinaciones son adoptadas como variables controladas para el sistema de control.

Otras estrategias, tales como *Extremum-Seeking Control* (Adetola y Guay, 2007; Guay y Zhang, 2003), *Neighboring Extremals* (Gros et al., 2009) y *NCO Tracking* (François et al., 2012, 2005) utilizan como “variables controladas” las condiciones necesarias de óptimo: las restricciones activas y el gradiente del Lagrangiano son las *variables*, y la consigna es mantenerlas en 0. Una ley de control integral (discreta) puede ser suficiente:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{K} \cdot (\mathbf{Z}^{ref} - \mathbf{Z}^k) \quad (2-24)$$

donde  $\mathbf{K}$  es una matriz de ganancias,  $\mathbf{Z}^k$  es el valor de las variables controladas en el ciclo  $k$  y  $\mathbf{Z}^{ref}$  la consigna.

Un problema a resolver en estas estrategias es que las variables a controlar son válidas mientras el conjunto de restricciones activas no cambie. Para completar la técnica, es necesario detectar en forma separada cuál es el conjunto de restricciones activas para actualizar el sistema (Woodward et al., 2010).

#### 2.3.4. Otras estrategias

Las estrategias *basadas en datos* (sin modelo) (Srinivasan y Bonvin, 2003) buscan el óptimo del sistema a través de datos experimentales obtenidos mediante perturbaciones de la planta. Cuando los datos indican que una dirección del espacio de las entradas produce una mejora en la función objetivo, se implementa; en caso contrario, se aplica un cambio en la dirección opuesta. Esta estrategia converge siempre a un óptimo local de la planta, aunque la evolución puede ser muy lenta en comparación con las estrategias mencionadas en secciones anteriores.

La *evolución en tiempo real (RTE)* (Sequeira et al., 2004) es una estrategia similar a la anterior pero en lugar de realizar perturbaciones en la planta, selecciona una dirección descendente utilizando un modelo y realiza un movimiento incremental en este sentido. Los datos obtenidos luego del cambio son utilizados para actualizar el modelo y proseguir con la evolución. Al realizar movimientos pequeños, los autores indican que no es necesario esperar al estado estacionario para ejecutar un paso de la evolución (aunque sí para actualizar el modelo).

## 2.4. TÉCNICAS DE ESTIMACIÓN DEL GRADIENTE

La adaptación por modificadores, el algoritmo ISOPE y algunas estrategias de adaptación directa de las entradas requieren la determinación experimental de los gradientes de la función objetivo y las restricciones, utilizando mediciones de la planta. Se han desarrollado diversas técnicas para la estimación del gradiente basada en datos (Mansour y Ellis, 2003; Srinivasan et al., 2011). Las principales técnicas para obtener experimentalmente, en tiempo real, los gradientes de la planta en estado estacionario, se mencionan a continuación.

### 2.4.1. Perturbaciones en la planta

El algoritmo ISOPE original propone utilizar una aproximación por *diferencias finitas adelantadas* (Brdys et al., 1987). Para estimar las derivadas parciales que componen el gradiente, se generan perturbaciones en cada una de las entradas, y el gradiente se estima como:

$$\nabla_{\mathbf{u}} \mathbf{y} \approx \frac{1}{\delta} \cdot \begin{pmatrix} \mathbf{y}(u_1^0 + \delta, \dots, u_i^0, \dots, u_{nu}^0) - \mathbf{y}(\mathbf{u}^0) \\ \vdots \\ \mathbf{y}(u_1^0, \dots, u_i^0 + \delta, \dots, u_{nu}^0) - \mathbf{y}(\mathbf{u}^0) \\ \vdots \\ \mathbf{y}(u_1^0, \dots, u_i^0, \dots, u_{nu}^0 + \delta) - \mathbf{y}(\mathbf{u}^0) \end{pmatrix} \quad (2-25)$$

donde  $\delta$  es la perturbación en torno al punto inicial  $\mathbf{u}^0$ . En la ecuación (2-25),  $nu$  perturbaciones son necesarias, una para cada una de las entradas  $u_i$ .

En procesos con un gran número de entradas, o en procesos con una dinámica lenta, esta técnica puede ser difícil de implementar debido al tiempo necesario para realizar todas las perturbaciones y esperar que el sistema alcance el estado estacionario después de cada una de ellas. Sin embargo, la técnica puede ser útil para inicializar un sistema de RTO, cuando no se dispone de información experimental para utilizar otro método (Marchetti et al., 2010).

En la técnica de *Extremum-Seeking Control* (Ariyur y Krstic, 2003), se propone aplicar la perturbación:

$$u_i(t) = u_i + \delta \cdot \text{sen}(\omega_i t) \quad (2-26)$$

donde  $\delta$  es la amplitud de la perturbación y las frecuencias  $\omega_i$  deben ser diferentes entre sí y bajas (con un período mayor que el tiempo necesario para que la planta alcance el estado estacionario). Aplicando filtros *pasa-altos* y *pasa-bajos*, se obtiene la derivada del gradiente con respecto al tiempo (François et al., 2012), la

que, a partir de un valor inicial del gradiente, permite calcular el gradiente en cada instante.

### 2.4.2. Múltiples unidades

Cuando el proceso consiste en un conjunto de unidades idénticas operando en paralelo, los gradientes pueden obtenerse perturbando cada una de las unidades con un valor diferente de las variables de entrada (Woodward et al., 2009). El número de unidades en paralelo necesarias para obtener el gradiente es  $nu+1$ , con  $nu$  el número de entradas del sistema. Las variables de entradas para cada unidad, con respecto a la unidad de referencia (con índice 0 en la ecuación) son:

$$\mathbf{u}^i = \mathbf{u}^0 + \delta \cdot \mathbf{e}_i, \quad i = 1, \dots, nu \quad (2-27)$$

donde  $\mathbf{e}_i$  es un vector con todos sus elementos nulos excepto el de la fila  $i$  que vale 1. Las derivadas parciales para las salidas se calculan como:

$$\frac{\partial \mathbf{y}}{\partial u_i} = \frac{1}{\delta} (\mathbf{y}^i - \mathbf{y}^0) \quad (2-28)$$

Este método es similar al de diferencias finitas, con la ventaja de que las perturbaciones se realizan todas al mismo tiempo, pero sólo es aplicable en los casos especiales en que el proceso utiliza unidades idénticas operando en paralelo.

### 2.4.3. Neighboring Extremals

Esta técnica (Gros et al., 2009) está basada en un modelo de la planta, por lo que en teoría sólo es útil cuando el error es sólo de naturaleza paramétrica. Sin embargo, ha probado comportarse en forma satisfactoria en presencia de errores estructurales (François y Bonvin, 2014).

El modelo estacionario puede expresarse como:

$$\mathbf{y} = H(\mathbf{u}, \boldsymbol{\theta}) \quad (2-29)$$

donde  $\boldsymbol{\theta}$  son los parámetros del modelo.

Para una optimización sin restricciones, el gradiente se estima a través de un análisis variacional en torno al óptimo nominal  $\mathbf{u}^0$  (donde el gradiente de la función objetivo para el modelo es cero) (François et al., 2012). Definiendo  $\delta \mathbf{y} = \mathbf{y} - \mathbf{y}^0$ ,  $\delta \mathbf{u} = \mathbf{u} - \mathbf{u}^0$  y  $\delta \boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^0$ , la variación de  $\mathbf{y}$  es:

$$\delta \mathbf{y} = \frac{\partial H}{\partial \mathbf{u}} \cdot \delta \mathbf{u} + \frac{\partial H}{\partial \boldsymbol{\theta}} \cdot \delta \boldsymbol{\theta} \quad (2-30)$$

donde las derivadas se evalúan en el óptimo nominal. Entonces,

$$\delta \boldsymbol{\theta} = \left( \frac{\partial H}{\partial \boldsymbol{\theta}} \right)^+ \cdot \left( \delta \mathbf{y} - \frac{\partial H}{\partial \mathbf{u}} \cdot \delta \mathbf{u} \right) \quad (2-31)$$

donde  $\left( \frac{\partial H}{\partial \boldsymbol{\theta}} \right)^+$  es la pseudoinversa del gradiente de  $H$  con respecto a  $\boldsymbol{\theta}$ . El gradiente del costo (función objetivo)  $\phi$  puede obtenerse en forma similar:

$$\frac{\partial \phi}{\partial \mathbf{u}} = \frac{\partial^2 \phi}{\partial \mathbf{u}^T \partial \mathbf{u}} \cdot \delta \mathbf{u} + \frac{\partial^2 \phi}{\partial \mathbf{u}^T \partial \boldsymbol{\theta}} \cdot \delta \boldsymbol{\theta} \quad (2-32)$$

Las ecuaciones (2-31) y (2-32) permiten obtener el gradiente del costo en términos de las entradas y salidas medidas:

$$\frac{\partial \phi}{\partial \mathbf{u}} = \frac{\partial^2 \phi}{\partial \mathbf{u}^T \partial \mathbf{u}} \cdot \delta \mathbf{u} + \frac{\partial^2 \phi}{\partial \mathbf{u}^T \partial \boldsymbol{\theta}} \cdot \left( \frac{\partial H}{\partial \boldsymbol{\theta}} \right)^+ \cdot \left( \delta \mathbf{y} - \frac{\partial H}{\partial \mathbf{u}} \cdot \delta \mathbf{u} \right) \quad (2-33)$$

Esta técnica se utiliza en general para la adaptación directa de las entradas, como se ha mencionado en la sección 2.3.3. Marchetti y Zumoffen (2013) han mostrado que cuando sólo hay incertidumbre paramétrica y no cambia el conjunto de restricciones activas, esta técnica equivale a una aproximación de primer orden de la estrategia de dos pasos. Otras técnicas de adaptación directa de las entradas presentadas en la sección 2.3.3 también proponen métodos para calcular el gradiente, diferentes al de *Neighboring Extremals* (Adetola y Guay, 2007; Srinivasan et al., 2011).

#### 2.4.4. Estimación dinámica

Los gradientes de estado estacionario pueden obtenerse utilizando datos dinámicos y un modelo no estacionario, lineal o no lineal (Mansour y Ellis, 2003). En ambos casos, se realiza una identificación de los parámetros dinámicos del modelo utilizando información de las entradas y las salidas medidas.

Para el caso lineal con una entrada y una salida, por ejemplo, puede obtenerse un modelo del tipo:

$$y^k = a_0 + a_1 \cdot y^{k-1} + \dots + a_n \cdot y^{k-n} + b_0 \cdot u^k + \dots + b_n \cdot u^{k-n} \quad (2-34)$$

Obteniendo por regresión los parámetros, el modelo correspondiente para estado estacionario es:

$$y \cdot (1 - a_1 - \dots - a_n) = a_0 + (b_0 + \dots + b_n) \cdot u \quad (2-35)$$

y, por lo tanto, la derivada estimada es

$$\frac{dy}{du} = \frac{(b_0 + \dots + b_n)}{(1 - a_1 - \dots - a_n)} \quad (2-36)$$

Si bien esta alternativa presenta la ventaja de no requerir el estado estacionario para obtener la estimación del gradiente, los datos dinámicos obtenidos de la operación normal de la planta podrían no contener información suficiente para la estimación de los parámetros del modelo, lo que haría necesario agregar perturbaciones (Brdys y Tatjewski, 2005).

#### 2.4.5. RTO dual

La información de ciclos anteriores de RTO puede utilizarse para estimar los gradientes de la planta (Mansour y Ellis, 2003). En el ciclo  $k$ , el gradiente de la variable  $y$  puede obtenerse en forma aproximada resolviendo la ecuación lineal:

$$\nabla y \approx (\mathbf{U}^k)^{-1} \begin{bmatrix} y^k - y^{k-1} \\ y^k - y^{k-2} \\ \dots \\ y^k - y^{k-nu} \end{bmatrix} \quad (2-37)$$

donde

$$\mathbf{U}^k = \left[ (\mathbf{u}^k - \mathbf{u}^{k-1})(\mathbf{u}^k - \mathbf{u}^{k-2}) \dots (\mathbf{u}^k - \mathbf{u}^{k-nu}) \right]^T \quad (2-38)$$

Esta técnica tiene la ventaja de no requerir perturbaciones para el problema de estimación. Sin embargo, de acuerdo con los cambios en el vector de entradas, la matriz  $\mathbf{U}^k$  podría no ser de rango  $nu$ , haciendo que el problema no tenga solución. Por otra parte, hay dos fuentes de error en la estimación (Marchetti et al., 2010). El *error por ruido*, a causa de la inevitable incertidumbre presente en las mediciones de  $y$ , provoca estimaciones inexactas si el problema (2-37) no está bien condicionado. El *error por truncamiento*, debido a la aproximación de primer orden de (2-37), hará que la estimación no sea válida si los puntos  $\mathbf{u}^k, \mathbf{u}^{k-1}, \dots, \mathbf{u}^{k-nu}$  están muy alejados entre sí.

Por estos motivos, los métodos que estiman el gradiente en base a datos de ciclos pasados de RTO necesitan de una estrategia que asegure que el próximo vector de entradas, sumado a los datos anteriores, contendrá información suficiente para la estimación. Estas estrategias incluyen restricciones adicionales en el problema de optimización, y reciben el nombre de *RTO dual*, porque cada ciclo persigue dos objetivos: minimizar la función objetivo y generar información para el cálculo del gradiente.

Para desarrollar las restricciones de RTO dual, se define la matriz

$$\mathbf{U} = \left[ (\mathbf{u} - \mathbf{u}^k) \ (\mathbf{u}^k - \mathbf{u}^{k-1}) \ \dots \ (\mathbf{u}^k - \mathbf{u}^{k-nu+1}) \right]^T \quad (2-39)$$

Brdys y Tatjewski (2005) proponen la siguiente restricción:

$$\frac{1}{\kappa(\mathbf{U})} \geq \delta \quad (2-40)$$

donde  $\kappa(\mathbf{U})$  es el número de condición de la matriz  $\mathbf{U}$ , que en el óptimo del problema se convertirá en  $\mathbf{U}^{k+1}$ , es decir la matriz utilizada en el próximo ciclo para resolver (2-37). El parámetro  $\delta$  se selecciona para la aplicación y debe estar entre 0 y 1.

Marchetti et al. (2010), en cambio, abordan explícitamente el error por truncamiento y por ruido, calculando cotas superiores para los dos tipos de error. El error total  $\mathbf{e}(\mathbf{u})$  se define como:

$$\mathbf{e}(\mathbf{u}) = \mathbf{e}^t(\mathbf{u}) + \mathbf{e}^r(\mathbf{u}) = \mathbf{U}^{-1} \cdot \begin{bmatrix} y - y^k \\ y - y^{k-1} \\ \dots \\ y - y^{k-nu+1} \end{bmatrix} - \nabla y \quad (2-41)$$

donde  $\mathbf{e}^t(\mathbf{u})$  es el error por truncamiento y  $\mathbf{e}^r(\mathbf{u})$  el error por ruido.

Los autores demuestran que la cota para el error por truncamiento es:

$$\begin{aligned} \|\mathbf{e}^t(\mathbf{u})\| \leq \xi^t(\mathbf{u}) = \\ \frac{\sigma_{\max}}{2} \cdot \left\| \left[ (\mathbf{u} - \mathbf{u}^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \ \dots \ (\mathbf{u} - \mathbf{u}^{k-nu+1})^T \cdot (\mathbf{u} - \mathbf{u}^{k-nu+1}) \right] \cdot \mathbf{U}^{-1} \right\| \end{aligned} \quad (2-42)$$

donde  $\sigma_{\max}$  es una cota superior para el radio espectral de la Hessiana de  $y(\mathbf{u})$ .

La cota para el error por ruido, mientras tanto, es:

$$\|\mathbf{e}^r(\mathbf{u})\| \leq \xi^r(\mathbf{u}) = \frac{\zeta}{l_{\min}(\mathbf{u})} \quad (2-43)$$

donde  $\zeta$  es un intervalo que contiene al valor máximo y al valor mínimo del ruido, y  $l_{\min}(\mathbf{u})$  es la menor distancia entre todos los subespacios afines complementarios que pueden generarse a partir del conjunto  $\{\mathbf{u}, \mathbf{u}^k, \dots, \mathbf{u}^{k-nu+1}\}$ . El trabajo de Marchetti et al. (2010) incluye una técnica para el cálculo de  $l_{\min}$ .

La restricción adicional incorporada por esta estrategia es:

$$\xi^r(\mathbf{u}) + \xi^r(\mathbf{u}) \leq \delta \quad (2-44)$$

donde  $\delta$  es el error máximo tolerado en la aplicación.

En la práctica, la restricción (2-44) genera regiones factibles no convexas, por lo que se agregan restricciones adicionales que garanticen la convexidad del problema y eviten la presencia de óptimos locales (si el problema de optimización original es convexo).

Tanto la restricción (2-40) como (2-44) generan dos regiones factibles desconectadas. En la práctica, en ambos casos, puede resolverse el problema sin incluir las restricciones para luego verificar si se cumplen. Si no se cumplen, será necesario resolver dos problemas de optimización, incluyendo cada una de las regiones factibles generadas. De los dos resultados, se seleccionará el que genere una solución (factible) con menor valor de la función objetivo (en el caso de un problema de minimización).

Marchetti (2013) propone, para la estrategia de adaptación por modificadores, calcular directamente el término modificador del gradiente:

$$\lambda^k = (\mathbf{U}^k)^{-1} \cdot \begin{bmatrix} \varepsilon^k - \varepsilon^{k-1} \\ \varepsilon^k - \varepsilon^{k-2} \\ \dots \\ \varepsilon^k - \varepsilon^{k-nu} \end{bmatrix} \quad (2-45)$$

Esta alternativa reduce el error por truncamiento, por lo que las regiones factibles generadas por las restricciones de RTO dual son de mayor tamaño que cuando el cálculo de los gradientes utiliza la ecuación (2-37) (Marchetti y Basualdo, 2012).

En lugar de actualizar los gradientes a través de una ecuación lineal, es posible estimarlos con un algoritmo del tipo Broyden (Rodger y Chachuat, 2011):

$$\mathbf{BR}^k = \mathbf{BR}^{k-1} + \frac{y^k - y^{k-1} - (\mathbf{BR}^{k-1})^T (\mathbf{u}^k - \mathbf{u}^{k-1})}{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|^2} (\mathbf{u}^k - \mathbf{u}^{k-1}) \quad (2-46)$$

donde  $\mathbf{BR}$  es el estimador del gradiente. El vector  $\mathbf{BR}$  debe inicializarse al inicio de la operación, por ejemplo con el gradiente del modelo sin adaptación. Este método evita los problemas de mal condicionamiento de las ecuaciones (2-37) y (2-45), ya que sólo actualiza el gradiente en la dirección del vector  $(\mathbf{u}^k - \mathbf{u}^{k-1})$ . Sin embargo, también requiere de restricciones adicionales para limitar los errores por

truncamiento y por ruido, para forzar al sistema a actualizar el gradiente en todas las direcciones del espacio de las entradas (Rodger, 2010).

Las técnicas de RTO dual presentan dificultades para ser aplicadas en sistemas con un gran número de variables de entrada, debido a que se necesitan al menos  $nu+1$  ciclos de RTO para generar la información necesaria para la estimación. En la práctica, esto puede requerir varias horas, y es probable la aparición de *disturbancias* (perturbaciones no medidas) no incluidas en el modelo, que alteren los valores de las variables de salida e invaliden los gradientes estimados (Marchetti et al., 2010). Además, las restricciones generadas pueden generar un problema no factible, por lo que puede ser necesario ajustar iterativamente el nivel de error o de mala condición tolerado, complicando excesivamente la estrategia.

#### 2.4.6. Regresión con estructura fija

Esta técnica también utiliza datos de ciclos de RTO pasados, desde el inicio de la operación del sistema de optimización en tiempo real.

En primer lugar, esta técnica tiene en cuenta que con frecuencia es posible poner cotas a los gradientes estimados. Para esto, se recurre a las constantes de Lipschitz para cada componente del gradiente,  $\boldsymbol{\kappa} = [\kappa_1 \dots \kappa_{nu}]^T$  (Bunin y François, 2012). Estas constantes se definen como

$$\kappa_j \in [0, \infty): y(\mathbf{u}^a) - y(\mathbf{u}^b) \leq \sum_{j=1}^{nu} \kappa_j |u_j^a - u_j^b|, \quad \forall \mathbf{u}^a, \mathbf{u}^b \quad (2-47)$$

El valor de las constantes  $\kappa_j$  es normalmente una cota superior en la derivada parcial  $\frac{\partial y}{\partial u_j}$ .

En segundo lugar, se selecciona una estructura para la función  $y(\mathbf{u})$ . Bunin et al. (2013b) proponen las siguientes estructuras para las salidas estimadas  $\hat{y}$  y  $k$  conjuntos de datos disponibles:

- *Lineal*: con  $nu+1$  parámetros,  $a$  y  $\mathbf{b}$ :

$$\left. \begin{aligned} \hat{y}^i &= a + \mathbf{b}^T \cdot \mathbf{u}^i \\ \nabla \hat{y}(\mathbf{u}^i) &= \mathbf{b} \end{aligned} \right\}, \quad i = 1 \dots k \quad (2-48)$$

- *Cuadrática*: con  $2nu+1$  parámetros,  $a$ ,  $\mathbf{b}$  y  $\mathbf{Q}$  ( $\mathbf{Q}$  es una matriz diagonal):

$$\left. \begin{aligned} \hat{y}^i &= a + \mathbf{b}^T \cdot \mathbf{u}^i + (\mathbf{u}^i)^T \cdot \mathbf{Q} \cdot \mathbf{u}^i \\ \nabla \hat{y}(\mathbf{u}^i) &= \mathbf{b} + 2 \cdot \mathbf{Q} \cdot \mathbf{u}^i \end{aligned} \right\}, \quad i = 1 \dots k \quad (2-49)$$

- *Convexa*: sin parámetros, sólo restricciones de desigualdad:

$$\hat{y}^i + (\nabla \hat{y}(\mathbf{u}^i))^T \cdot (\mathbf{u}^j - \mathbf{u}^i) \leq \hat{y}^j \quad i, j = 1 \dots k \quad (2-50)$$

- *Cóncava*:

$$\hat{y}^i + (\nabla \hat{y}(\mathbf{u}^i))^T \cdot (\mathbf{u}^j - \mathbf{u}^i) \geq \hat{y}^j \quad i, j = 1 \dots k \quad (2-51)$$

- *Suma de funciones convexas y cóncavas*: La salida es una suma de dos funciones, una cóncava y una convexa, cada una limitada por las constantes de Lipschitz.

Una vez definida la estructura, pueden obtenerse los gradientes a través de una regresión por cuadrados mínimos con restricciones. Para los casos paramétricos, el problema es una simple estimación de parámetros, que incluye la restricción

$$-\boldsymbol{\kappa} \leq \nabla y(\mathbf{u}^i) \leq \boldsymbol{\kappa} \quad i = 1 \dots k \quad (2-52)$$

Para los casos con restricciones, puede implementarse un método similar al propuesto por Bunin y François (2012). Por ejemplo, para la estructura convexa, la estimación del gradiente puede resolverse como:

$$\begin{aligned} \min_{\mathbf{z}^k} \quad & \sum_{i=1}^{k-1} w_i \left( \frac{(\mathbf{z}^k)^T \cdot (\mathbf{u}^k - \mathbf{u}^i) - (y^k - y^i)}{\|\mathbf{u}^k - \mathbf{u}^i\|_2} \right) \\ \text{s.a.} \quad & y^k + (\mathbf{z}^k)^T \cdot (\mathbf{u}^i - \mathbf{u}^k) \leq y^i \\ & -\boldsymbol{\kappa} \leq \mathbf{z}^k \leq \boldsymbol{\kappa} \quad i = 1 \dots k \end{aligned} \quad (2-53)$$

donde  $\mathbf{z}^k$  es la estimación del gradiente en el ciclo  $k$ , y los pesos  $w_i$  deben seleccionarse de acuerdo con algún criterio y con la experiencia del proceso. En el trabajo de Bunin y François (2012) se componen en base al ruido en la medición y a la distancia entre el punto  $\mathbf{u}^k$  y el punto  $\mathbf{u}^i$  (ya que los puntos más cercanos a  $\mathbf{u}^k$  serán más útiles en la estimación del gradiente local, y los puntos con mayor nivel de ruido serán menos útiles):

$$w_i = w_i^r \cdot w_i^d \quad (2-54)$$

donde el peso  $w_i^r$  debido al ruido se estima en base al cociente entre el nivel de ruido  $\sigma$  y la señal, a través de la función

$$w_i^r = \min \left( \frac{0.5 \cdot \sigma}{\sum_{j=1}^{nu} \kappa_j \cdot |u_j^k - u_j^i|}, 1 \right) \quad (2-55)$$

y el peso  $w_i^d$  propuesto para la distancia entre los puntos es

$$w_i^d = 1 - \frac{\|\mathbf{u}^k - \mathbf{u}^i\|}{\max_j (\|\mathbf{u}^k - \mathbf{u}^j\|)} \quad (2-56)$$

En el capítulo 3 de esta tesis también se propone una regresión utilizando pesos, y se fijan límites a las estimaciones. En lugar de estimar los gradientes o las variables de salida, el problema del capítulo 3 está orientado directamente a obtener los términos constantes en una estrategia de adaptación por modificadores. Los resultados fueron publicados prácticamente en simultáneo con los trabajos citados en esta sección (Serralunga et al., 2012a, 2012b, 2013).

## 2.5. DETECCIÓN DE ESTADO ESTACIONARIO

En general, las estrategias de dos pasos y de adaptación por modificadores utilizan modelos de estado estacionario. Por este motivo, antes de comenzar un ciclo de RTO, es necesario verificar que el proceso a optimizar no esté atravesando un estado transiente que invalide los modelos empleados.

La mayoría de las técnicas para detección de estado estacionario requieren la selección de una ventana de tiempo, para analizar la variancia y la pendiente por regresión a través de tests estadísticos (Bhat y Saraf, 2004). Esto puede ser apropiado para la construcción de un modelo fuera de línea o para el control estadístico de procesos, pero no es apropiado para la optimización en tiempo real y otras aplicaciones en línea, ya que suele existir un retardo importante para conocer el estado estacionario correspondiente a un determinado instante.

Cao y Rhinehart (1995) proponen un método sencillo para la detección de estado estacionario en línea, utilizando un filtrado exponencial. El método requiere 3 parámetros de filtrado entre 0 y 1,  $L_1$ ,  $L_2$  y  $L_3$ . En el instante  $k$ , para una variable medida  $y^k$ , se obtiene el valor filtrado  $y_f^k$ :

$$y_f^k = L_1 \cdot y^k + (1 - L_1) \cdot y_f^{k-1} \quad (2-57)$$

Se calcula la desviación cuadrática con respecto al valor filtrado anterior,  $(v_f^k)^2$ , como:

$$(v_f^k)^2 = L_2 \cdot (y^k - y_f^{k-1})^2 + (1 - L_2) \cdot (v_f^{k-1})^2 \quad (2-58)$$

Si la variancia  $(s_1^k)^2$  se estima a partir del error cuadrático  $(v_f^k)^2$ , se obtiene como:

$$(s_1^k)^2 = \frac{2 - L_1}{2} \cdot (v_f^k)^2 \quad (2-59)$$

Por otra parte, la diferencia cuadrática filtrada entre dos datos sucesivos,  $(d_f^k)^2$ , se calcula como:

$$(d_f^k)^2 = L_3 \cdot (y^k - y^{k-1})^2 + (1 - L_3) \cdot (d_f^{k-1})^2 \quad (2-60)$$

y la variancia correspondiente,  $(s_2^k)^2$ , es:

$$(s_2^k)^2 = \frac{(d_f^k)^2}{2} \quad (2-61)$$

Se toma el cociente  $R^k$  de las dos estimaciones de la variancia,

$$R^k = \frac{(s_1^k)^2}{(s_2^k)^2} = \frac{(2 - L_1) \cdot (v_f^k)^2}{(d_f^k)^2} \quad (2-62)$$

Cuando el proceso está en estado estacionario,  $R^k$  permanece aproximadamente constante y en torno a 1. Cuando el proceso comienza un transiente, el estimado  $(v_f^k)^2$  aumenta en mayor magnitud que  $(d_f^k)^2$ , por lo que  $R^k$  aumenta. Cuando el cociente supera un valor crítico,  $R_{crit}$ , se considera que el sistema no está en estado estacionario.

La técnica mencionada corresponde a una variable. Si bien una variable correctamente escogida puede ser suficiente para determinar el estado estacionario de un proceso completo, se ha extendido el método para más de una variable (Mansour y Ellis, 2008).

Le Roux et al. (2008) revisan el desempeño de este método y otros tests estadísticos alternativos para la detección de estado estacionario, a través de un caso industrial. Los autores proponen el uso de un método de interpolación polinómica y sugieren que podría ser aplicable en RTO.

## 2.6. VALIDACIÓN DE DATOS

### 2.6.1. Validación simple

Las mediciones utilizadas para la adaptación del modelo contienen inevitablemente errores, debido a la variancia (ruido) inherente a los instrumentos de medición y a desviaciones con respecto al valor real originadas por fallas en la instrumentación, falta de calibración o pérdidas no medidas en el proceso. Las técnicas de *reconciliación de datos* reducen al mínimo posible el error por ruido, mientras que la *detección de errores gruesos* se encarga de encontrar aquellas mediciones con desviaciones significativas del valor esperado.

Si bien desde el ámbito académico la reconciliación de datos con detección de errores gruesos se propone como un paso clave en un ciclo de RTO (Mansour y Ellis, 2008; Sequeira et al., 2004; Zhang et al., 2002), en la práctica industrial es común realizar una simple validación que consiste en fijar límites superior e inferior para las variables y verificar si los valores medidos están dentro del rango aceptable (Darby et al., 2011). Los límites se fijan de acuerdo con la experiencia del ingeniero y con datos históricos de las mediciones. Cuando una medición viola los límites, se reemplaza su valor por una estimación o un valor por defecto; en caso de mediciones críticas, se puede cancelar la ejecución del ciclo de RTO hasta que los valores obtenidos sean válidos.

### 2.6.2. Reconciliación de datos

Los datos obtenidos de instrumentos de medición en general, no satisfacen los balances de materia y energía del proceso. La *reconciliación de datos* consiste en encontrar el valor más probable de las variables de un proceso, tanto medidas como no medidas, que cumpla con los balances (Romagnoli y Sanchez, 1999). Los resultados pueden aplicarse a aplicaciones de contabilidad de la producción, control de procesos, simulación y optimización (Poulin et al., 2010).

Existen dos conceptos claves en la reconciliación de datos: la *observabilidad* y la *redundancia* (Benqlilou, 2004). Una variable es *observable* si puede ser estimada a partir de las mediciones y el modelo del proceso. Una variable medida es *redundante* si continúa siendo observable aun si se remueve su medición.

El problema de reconciliación de en estado estacionario puede plantearse como la obtención de los valores más probables  $(\mathbf{y}^r, \mathbf{x}^r)$  de las variables observables:

$$\begin{aligned} \mathbf{y}^r, \mathbf{x}^r = \arg \min_{\mathbf{y}, \mathbf{x}} \sum_{i=1}^{ny} \left( \frac{y_i^m - y_i}{\sigma_i} \right)^2 \\ \text{s.a.} \quad \mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \\ \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \end{aligned} \quad (2-63)$$

donde  $\mathbf{y}$  es el vector de variables medidas,  $\mathbf{x}$  el vector de variables no medidas pero observables,  $\sigma_i$  la desviación estándar de la variable medida  $y_i$ , e  $y_i^m$  el valor medido de  $y_i$ ;  $\mathbf{f}$  y  $\mathbf{g}$  son las restricciones de igualdad y desigualdad que modelan el proceso. La formulación con cuadrados mínimos pesados considerada en (2-63) está basada en una distribución normal para el error. Para distribuciones distintas de la normal, pueden formularse otras funciones objetivo basadas en funciones de máxima verosimilitud (Özyurt y Pike, 2004).

En muchos casos, la reconciliación de datos incluye sólo balances de masa, lo que genera un problema lineal. Por este motivo, se han desarrollado técnicas de descomposición matricial, que resuelven en forma eficiente el problema calculando en primer lugar las variables medidas redundantes, y a continuación el resto de las variables, eliminando aquellas no observables. (Narasimhan y Jordache, 1999). Estos métodos pueden extenderse a procesos bilineales, tales como balances de energía y de materia por componentes (Sánchez y Romagnoli, 1996).

En algunos casos, la reconciliación de datos puede resolverse en conjunto con un problema de estimación de parámetros (Arora y Biegler, 2001). En este caso, el problema de reconciliación-estimación simultáneo es el indicado en la ecuación (2-10) para la estrategia de dos pasos. El agregado de límites máximos y mínimos para los parámetros estimados es de utilidad para la obtención de valores válidos de las variables reconciliadas (Narasimhan y Jordache, 1999).

Cuando la reconciliación de datos está orientada a procesos dinámicos, se usan tradicionalmente dos estrategias (Romagnoli y Sanchez, 1999). La estrategia de *filtrado* utiliza generalmente el llamado *filtro de Kalman*, que estima recursivamente los valores de las variables a partir de estimaciones anteriores y de las mediciones actuales. Su versión original asume un proceso lineal, pero se ha extendido a procesos no lineales. La estrategia de *optimización*, en cambio, se asemeja al problema propuesto en (2-63), con un modelo dinámico en lugar del estacionario.

### 2.6.3. Detección de errores gruesos

Cuando el proceso tiene pérdidas no medidas o instrumentos que fallan, las estimaciones provistas por (2-63) no son válidas. Por este motivo, es necesario detectar estos problemas en la medición. Las técnicas de detección de errores gruesos deben alcanzar los siguientes objetivos (Narasimhan y Jordache, 1999):

- Detectar la presencia de errores gruesos (*problema de detección*).
- Identificar el origen del error y el tipo de error (*problema de identificación*).
- Identificar y detectar múltiples errores gruesos presentes simultáneamente (*problema de identificación múltiple de errores gruesos*).
- Estimar la magnitud de los errores (*problema de estimación*).

En general, la detección de errores gruesos se realiza mediante análisis estadístico. Mansour y Ellis (2008) dividen las técnicas de detección en dos grupos: en el primero están los métodos que usan la distribución de los *residuos de las restricciones* (errores de balance); en el segundo, los que analizan la distribución de los valores estimados de las variables medidas. Este segundo método requiere resolver el problema de reconciliación para luego examinar el resultado obtenido, por eso también se los denomina como *reconciliación de datos combinada con detección de errores gruesos*.

Una vez identificados los errores, el problema de estimación puede plantearse, para el caso lineal, como:

$$\begin{aligned}
 & \min_{\mathbf{y}, \mathbf{x}, \mathbf{d}, \mathbf{p}} \sum_{i=1}^{ny} \left( \frac{y_i^m - y_i - d_i}{\sigma_i} \right)^2 \\
 & \text{s.a.} \quad \mathbf{f}(\mathbf{x}, \mathbf{y} + \mathbf{d}, \mathbf{p}) = \mathbf{0} \\
 & \quad \quad \mathbf{g}(\mathbf{x}, \mathbf{y} + \mathbf{d}, \mathbf{p}) \leq \mathbf{0} \\
 & \quad \quad \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U \\
 & \quad \quad \mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U
 \end{aligned} \tag{2-64}$$

donde  $\mathbf{d}$  es el vector de errores en las mediciones, y  $\mathbf{p}$  las pérdidas no medidas (Romagnoli y Sanchez, 1999).

Algunas formulaciones *robustas* de reconciliación incluyen una función objetivo que asigna un valor bajo a los valores que difieren mucho del valor estimado, tales como la *distribución normal contaminada* (Özyurt y Pike, 2004). De esta manera, los errores gruesos no afectan la estimación.

Al igual que la reconciliación de datos, si bien se ha formulado originalmente para balances en estado estacionario, la detección de errores gruesos puede aplicarse a procesos dinámicos (Bagajewicz y Jiang, 2000; Chen y Romagnoli, 1998).

## 2.7. VALIDACIÓN DE LOS RESULTADOS

### 2.7.1. Verificación estadística

Antes de aplicar los resultados de la optimización del modelo adaptado, es conveniente verificar si los cambios propuestos son significativos para implementarlos. Se ha propuesto un análisis estadístico de los cambios propuestos en las variables de entrada (Miletic y Marlin, 1996, 1998b).

El análisis consiste en un test de hipótesis:

$$\begin{aligned} H_0 : \boldsymbol{\mu}^* &= \boldsymbol{\mu}^k \\ H_1 : \boldsymbol{\mu}^* &\neq \boldsymbol{\mu}^k \end{aligned} \quad (2-65)$$

donde  $\boldsymbol{\mu}^*$  y  $\boldsymbol{\mu}^k$  son los valores esperados de las entradas óptimas y las entradas en el ciclo actual  $k$ , respectivamente. Estos valores son estimados por los valores conocidos  $\mathbf{u}^*$  (resultado del problema de optimización) y  $\mathbf{u}^k$  (entradas conocidas actuales). Se supone que la desviación entre  $\mathbf{u}^*$  y  $\mathbf{u}^k$  sigue una distribución  $T^2$ :

$$T^2 = (\mathbf{u}^* - \mathbf{u}^k)^T \cdot \mathbf{Q}^{-1} \cdot (\mathbf{u}^* - \mathbf{u}^k) \sim \mathcal{G} \cdot F \quad (2-66)$$

donde  $\mathbf{Q}$  es la matriz de variancia-covariancia de las entradas, calculada a través de un análisis de sensibilidad lineal con respecto a los parámetros estimados del modelo.  $\mathcal{G}$  es una constante que depende de la forma del test de hipótesis, y que multiplica a la distribución estadística  $F$ .

Miletic y Marlin (1998b) proponen varias alternativas para calcular el límite de control del test de hipótesis. La más simple es:

$$LS = \frac{p \cdot (N+1) \cdot (N-1)}{N \cdot (N-p)} F_{\eta}(p, N-p) \quad (2-67)$$

donde  $p$  es la dimensión del vector  $\mathbf{u}$  y  $N$  el número de datos utilizados para estimar la variancia de los datos. Si el valor de  $T^2$  supera el límite de control  $LS$ , se rechaza la hipótesis nula, y se considera que el nuevo óptimo es significativamente diferente del resultado anterior. Por lo tanto, los resultados de la optimización deben aplicarse en el proceso.

### 2.7.2. Filtrado de las entradas óptimas

Debido a la incertidumbre en el modelado, es posible que los cambios propuestos por el modelo adaptado sean infactibles o subóptimas. Una práctica común para aplicar los cambios parcialmente, reduciendo la probabilidad de infactibilidades y favoreciendo la convergencia del sistema, es el *filtrado de las entradas óptimas* (Brdys et al., 1987; Mansour y Ellis, 2008). En el ciclo  $k$ , en lugar de aplicar directamente el resultado óptimo  $\mathbf{u}^{*,k+1}$ , el nuevo conjunto de entradas se implementa como:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + K^k \cdot (\mathbf{u}^{*,k+1} - \mathbf{u}^k) \quad (2-68)$$

con  $K^k$  una constante entre 0 y 1, que normalmente es un parámetro de diseño del sistema de RTO, que se define en base a la experiencia y toma el mismo valor en todos los ciclos.

Bunin et al. (2011) analizaron los valores que puede tomar  $K^k$  para garantizar la factibilidad de un sistema de RTO. Asumiendo que todas las restricciones de la planta  $\mathbf{g}_p$  (expresadas como en la ecuación (2-5)) son Lipschitz-continuas, es decir:

$$\exists \kappa_j \in [0, \infty): g_{p,j}(\mathbf{u}^a) - g_{p,j}(\mathbf{u}^b) \leq \kappa_j \cdot \|\mathbf{u}^a - \mathbf{u}^b\|, \quad \forall \mathbf{u}^a, \mathbf{u}^b \quad (2-69)$$

la condición suficiente para que el punto  $\mathbf{u}^{k+1}$  sea factible es:

$$K^k \leq \min_i \left( \frac{-g_{p,i}(\mathbf{u}^k)}{\kappa_i \cdot \|\mathbf{u}^{*,k+1} - \mathbf{u}^k\|} \right) \quad (2-70)$$

En otro trabajo reciente, Bunin et al. (2013a) también establecen las condiciones para que el nuevo punto  $\mathbf{u}^{k+1}$  genere una mejora en la función objetivo con respecto al punto actual. Si la función objetivo  $Q_p(\mathbf{u})$  es continua diferenciable dos veces, pueden definirse en la región factible de  $\mathbf{u}$  las cotas

$$-M_{ij} \leq \frac{\partial^2 Q_p}{\partial u_i \partial u_j} \leq M_{ij}, \quad \forall \mathbf{u}, \quad i, j = 1, \dots, nu \quad (2-71)$$

Estas cotas se utilizan para construir la matriz  $R$ , diagonal con sus elementos calculados como

$$R_{ii} = \sum_{j=1}^{nu} M_{ij}, \quad i = 1, \dots, nu \quad (2-72)$$

A partir de esta definición, Bunin et al. (2013a) prueban que se garantiza un descenso en la función objetivo si

$$0 \leq K^k \leq -2 \frac{\nabla Q_p(\mathbf{u}^k)^T \cdot (\mathbf{u}^{*,k+1} - \mathbf{u}^k)}{(\mathbf{u}^{*,k+1} - \mathbf{u}^k)^T \cdot R \cdot (\mathbf{u}^{*,k+1} - \mathbf{u}^k)} \quad (2-73)$$

Si la restricción (2-70) se implementa directamente, puede causar una convergencia prematura o demasiado lenta del sistema de RTO, ya que si una restricción está activa o muy próxima a estarlo, los valores de  $K^k$  serán muy bajos o nulos. Por este motivo, se han propuesto alternativas prácticas que incluyen tolerar violaciones en algunas restricciones (Bunin et al., 2011) o aplicar el resultado utilizando una proyección del vector  $(\mathbf{u}^{*,k+1} - \mathbf{u}^k)$  hacia una dirección de descenso que permita valores de  $K^k$  más elevados (Bunin et al., 2013a, 2013c).

### 2.7.3. Optimización de *targets*

Cuando los resultados de la optimización en tiempo real se implementan a través de un controlador predictivo multivariable (MPC), es posible que el resultado óptimo de la RTO no sea factible para el modelo del MPC. Por este motivo, es frecuente definir los *targets* (objetivos) del MPC a través de una optimización que minimice el error con respecto a los resultados propuestos por la RTO pero garantizando la factibilidad de los resultados para el modelo del MPC (Darby et al., 2011; Marchetti et al., 2014). Esta optimización puede tomar la siguiente forma:

$$\begin{aligned} & \min_{\mathbf{y}^t, \mathbf{u}^t} (\mathbf{y}^t - \mathbf{y}^{k+1})^T \cdot Q \cdot (\mathbf{y}^t - \mathbf{y}^{k+1}) + (\mathbf{u}^t - \mathbf{u}^{k+1})^T \cdot R \cdot (\mathbf{u}^t - \mathbf{u}^{k+1}) + \mathbf{q}^T \cdot \mathbf{y}^t + \mathbf{r}^T \cdot \mathbf{u}^t \\ & s.a. \quad \mathbf{y}^t = \mathbf{G}_{ee} \cdot \mathbf{u}^t + \mathbf{b} \\ & \quad \mathbf{y}^L \leq \mathbf{y}^t \leq \mathbf{y}^U \\ & \quad \mathbf{u}^L \leq \mathbf{u}^t \leq \mathbf{u}^U \end{aligned} \quad (2-74)$$

donde  $\mathbf{y}^t$  y  $\mathbf{u}^t$  son los valores de los objetivos para las salidas y las entradas, respectivamente;  $\mathbf{y}^{k+1}$  y  $\mathbf{u}^{k+1}$  son los valores obtenidos de la RTO en el ciclo  $k$ ;  $Q$  y  $R$  son matrices que pesan las diferencia entre el objetivo y el resultado de la RTO;  $q$  y  $r$  son costos lineales;  $\mathbf{G}_{ee}$  es una matriz ganancias en estado estacionario (obtenida a partir del modelo dinámico del MPC),  $\mathbf{b}$  es un vector de *bias* (para

corregir el error de la ganancia  $\mathbf{G}_{ee}$ ); y los límites en las salidas y las entradas están dados por los vectores  $\mathbf{y}^L$ ,  $\mathbf{y}^U$ ,  $\mathbf{u}^L$  y  $\mathbf{u}^U$ .

## 2.8. MÉTRICAS PARA EVALUAR EL DESEMPEÑO DE UN SISTEMA DE RTO

Al diseñar un sistema de RTO, es necesario definir las características del sistema: la estructura del modelo, el método de adaptación, los parámetros ajustables y las mediciones a utilizar, entre otras. Forbes y Marlin (1996) han propuesto un método para evaluar el desempeño de un sistema de RTO, basado en el cálculo de un parámetro llamado *costo de diseño*, que para un problema de minimización se expresa como:

$$C = E[Q_p(\mathbf{u}^{k_f})] - Q_p(\mathbf{u}^{*,P}) \quad (2-75)$$

donde el costo de diseño  $C$  se calcula como la diferencia entre el valor esperado del costo operativo luego de aplicar el método de RTO durante  $k_f$  ciclos,  $Q_p(\mathbf{u}^{k_f})$ , y el costo operativo óptimo de la planta,  $Q_p(\mathbf{u}^{*,P})$ . El valor esperado del costo alcanzado puede obtenerse a partir de una formulación probabilística, o a partir de experimentos. Para evaluar el costo operativo óptimo en tiempo de diseño puede utilizarse un modelo más riguroso que el del sistema de RTO, o buscarlo a través de un método sin modelo como los descritos en la sección 2.3.4.

El *costo de diseño extendido* (Zhang y Forbes, 2000) tiene en cuenta no sólo el costo operativo alcanzado en el tiempo final de la evaluación, sino también la pérdida de beneficio durante la evolución del sistema de RTO. Se define como:

$$C_E = \int_{t_0}^{t_f} E[Q_p(\mathbf{u}(t))] - Q_p(\mathbf{u}^{*,P}) dt \quad (2-76)$$

donde los tiempos inicial y final de la evaluación,  $t_0$  y  $t_f$  son los extremos de integración. De esta manera, se tienen en cuenta los costos debidos al error estructural del modelo, a la variabilidad en las mediciones y al transiente necesario hasta alcanzar la convergencia del sistema.

## 2.9. RTO CON INCERTIDUMBRE

En la optimización en tiempo real hay cuatro fuentes inevitables de incertidumbre, que pueden conducir a la violación de restricciones o a suboptimalidades al aplicar las soluciones resultantes de la optimización (Zhang et al., 2001):

- *Incertidumbre en el proceso*: fluctuaciones inesperadas en caudales, temperaturas o niveles, cambio en las condiciones ambientales, variación de las propiedades de las corrientes de entrada.
- *Incertidumbre en las mediciones*: variancia en las mediciones, fallas en los sensores, pérdidas no medidas en las corrientes de proceso. Esta incertidumbre es mitigada por las técnicas de reconciliación y detección de errores gruesos descritas en la sección 2.6, pero no desaparece totalmente.
- *Incertidumbre en el modelo*: errores estructurales y paramétricos. Los métodos de adaptación de la sección 2.3 reducen los errores pero sólo en forma local: los modelos adaptados no están libres de error en la totalidad de la región factible del problema, sólo en el entorno de los últimos puntos de operación.
- *Incertidumbre del mercado*: cambios en precios, disponibilidad de materia prima y demanda de productos.

Una de las estrategias para operar un sistema de RTO ante incertidumbre en el modelado es el filtrado de las entradas, desarrollado en la sección 2.7.2. Bunin et al. (2013c) proponen una implementación robusta del filtrado de las entradas, que tiene en cuenta la incertidumbre en las mediciones.

A continuación, se describen otras estrategias para responder a la incertidumbre presente, inevitablemente, en un sistema de RTO:

### 2.9.1. Intervalos de confianza

Una estrategia común en la optimización es limitar los movimientos con respecto al punto actual (Darby et al., 2011; Mansour y Ellis, 2008; Zhang et al., 2001). Usualmente el límite se define para cada variable de entrada, agregando al problema de optimización en tiempo real las siguientes restricciones:

$$\max(u_i^L, u_i^k - \delta_i) \leq u_i \leq \min(u_i^U, u_i^k + \delta_i), \quad i = 1, \dots, nu \quad (2-77)$$

donde  $u_i^L$  y  $u_i^U$  son los límites inferior y superior para la entrada  $u_i$ , y  $\delta_i$  es el cambio máximo permitido para esta variable.

### 2.9.2. Margen de seguridad en las restricciones

Otra estrategia para evitar la violación de restricciones frente a la incertidumbre es reducir la región factible agregando un margen de seguridad en las restricciones (Chachuat et al., 2009). De esta manera, las restricciones de desigualdad se reformulan como:

$$\mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}) + \mathbf{m} \leq 0 \quad (2-78)$$

donde el margen de seguridad  $\mathbf{m}$  es un vector con todos sus elementos positivos.

### 2.9.3. Formulaciones con incertidumbre

Zhang et al. (2002) abordan el problema de RTO con incertidumbre en forma explícita. Expresando implícitamente las variables de salida como función de las entradas, y diferenciando los parámetros económicos  $\mathbf{c}$  de los parámetros del modelo  $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ , la función objetivo puede formularse como:

$$\min E[Q(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{c})] \quad (2-79)$$

Como los parámetros económicos suelen aparecer en forma lineal en el modelo, y el valor de los parámetros ajustables se asume constante en el valor estimado  $\boldsymbol{\beta}^k$ , la función objetivo dada por (2-79) se reduce a

$$\min Q(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}^k, \bar{\mathbf{c}}) \quad (2-80)$$

donde  $\bar{\mathbf{c}}$  es el valor esperado de los parámetros económicos.

Para las restricciones, la forma propuesta para incluir la incertidumbre es una formulación en términos probabilísticos. Hay dos posibles formulaciones: las restricciones con *probabilidad individual* y las restricciones con *probabilidad conjunta*.

La probabilidad individual se expresa como:

$$P(g_i(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq 0) \geq \rho_i, \quad i = 1, \dots, ng \quad (2-81)$$

donde  $\rho_i$  es el valor deseado de probabilidad de que la restricción  $g_i$  se cumpla.

La probabilidad conjunta, en cambio, se formula como:

$$P(\mathbf{g}(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq 0) \geq \rho_0 \quad (2-82)$$

es decir, se trabaja con la probabilidad de que la solución obtenida sea factible en la práctica (no viole ninguna restricción).

En ambos casos, las probabilidades buscadas deben seleccionarse con atención, ya que una probabilidad muy elevada podría hacer que el problema de optimización sea infactible, o que el valor óptimo de la función objetivo sea muy elevado (para un problema de minimización)

En el trabajo de Zhang et al. (2002) se proponen métodos de solución para RTO con probabilidad individual y conjunta de cumplimiento de las restricciones, basados en la linealización del problema en el punto actual. Si bien la formulación

(2-82) es más apropiada desde el punto de vista teórico, los métodos de resolución son más complejos, y requieren de simulaciones del tipo Monte Carlo para obtener las distribuciones de probabilidad en cada ciclo de RTO, lo que podría ser impracticable en el tiempo disponible para resolver la optimización. Por otra parte, los resultados informados para la formulación (2-81) son satisfactorios y no presentan desventajas apreciables con respecto al uso de la probabilidad conjunta.

## 2.10. RTO MULTIPERÍODO

La gran mayoría de las aplicaciones de RTO en estado estacionario están basadas en un solo período. Singh et al. (2000) proponen una formulación multiperíodo de RTO para el mezclado (*blending*) de gasolinas. Esta formulación tiene un horizonte fijo, ya que lo que se busca es una mezcla dentro de especificación al final del tiempo de operación. La formulación propuesta puede expresarse en forma simplificada como

$$\begin{aligned} \max_{\mathbf{u}} \quad & \sum_{t=t^k}^T Q(\mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\alpha}, \boldsymbol{\beta}^k) \cdot \Delta t \\ \text{s.a.} \quad & \sum_{t=0}^{t^k-1} \mathbf{g}_f(\hat{\mathbf{y}}(t), \mathbf{u}(t)) \cdot \Delta t + \sum_{t=t^k}^T \mathbf{g}_f(\mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\alpha}, \boldsymbol{\beta}^k) \cdot \Delta t \leq 0 \\ & \mathbf{g}_t(\mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\alpha}, \boldsymbol{\beta}^k) \leq 0, \quad t = k, \dots, T \end{aligned} \quad (2-83)$$

donde las restricciones  $\mathbf{g}_f$  se evalúan en los valores acumulados al final del horizonte, y  $\mathbf{g}_t$  en cada período desde el actual ( $k$ ) hasta el final. El modelo se adapta al comienzo de cada período, utilizando el valor experimental  $\hat{\mathbf{y}}(t)$  para los datos de tiempos anteriores, y actualizando los parámetros  $\boldsymbol{\beta}^k$  (en caso de que la estrategia de adaptación sea la de dos pasos, aunque puede utilizarse cualquier otra estrategia de las descritas en la sección 2.3).

## 2.11. RTO CON MODELOS DINÁMICOS

En general, la optimización en tiempo real de procesos continuos utiliza modelos en estado estacionario. Sin embargo, existen formulaciones que utilizan modelos dinámicos, para optimizar la operación de procesos continuos, discontinuos o semicontinuos.

### 2.11.1. Integración con control predictivo multivariable (MPC)

Como se ha mencionado en la sección 1.1.3, los resultados de la optimización en tiempo real suelen aplicarse en la planta a través de un *controlador*

*predictivo multivariable* (MPC) (Souza et al., 2010). Se han propuesto estrategias para integrar las etapas de RTO y MPC en un solo paso (Adetola y Guay, 2010; Zanin et al., 2002).

La estrategia propuesta con más frecuencia resuelve un problema dinámico con horizonte deslizante para el MPC, pero incluye en la función objetivo el costo alcanzado al final del horizonte (Alamo et al., 2014, 2012).

Este problema puede formularse en forma simplificada como:

$$\begin{aligned}
 \min_{\mathbf{u}} \quad & \sum_{j=k}^{k+N-1} \left( \mathbf{x}^j - \mathbf{x}^{k+N} \right)^T \cdot Q \cdot \left( \mathbf{x}^j - \mathbf{x}^{k+N} \right) + \left( \mathbf{u}^j - \mathbf{u}^{k+N} \right)^T \cdot R \cdot \left( \mathbf{u}^j - \mathbf{u}^{k+N} \right) \\
 & + f_{eco} \left( \mathbf{x}^{k+N}, \mathbf{u}^{k+N} \right) \\
 \text{s.t.} \quad & \mathbf{x}^k = \hat{\mathbf{x}}^k \\
 & \left. \begin{aligned} \mathbf{x}^{j+1} &= A \cdot \mathbf{x}^j + B \cdot \mathbf{u}^j \\ \mathbf{x}^j &\in X, \mathbf{u}^j \in U \end{aligned} \right\}, \quad j = k, \dots, k+N-1 \\
 & \left( \mathbf{x}^{k+N}, \mathbf{u}^{k+N} \right) \in Z
 \end{aligned} \tag{2-84}$$

donde la evolución está dada por el modelo lineal discreto  $(A, B)$ ,  $X$  y  $U$  son las regiones factibles durante las trayectorias dinámicas, y  $Z$  es la región factible para el modelo estacionario, incluyendo el modelo no lineal de la planta. La función  $f_{eco}$  es la función objetivo de RTO, mientras que los términos cuadráticos basados en las matrices  $Q$  y  $R$  contienen la parte dinámica de la función objetivo.

En cada tiempo  $k$ , el sistema se inicializa con las variables medidas  $\hat{\mathbf{x}}^k$ . El sistema se aplica con una estrategia de *horizonte deslizante*: Al resolver (2-84), del conjunto de entradas óptimas  $(\mathbf{u}^k, \mathbf{u}^{k+1}, \dots, \mathbf{u}^{k+N})$  se aplica sólo el valor óptimo de  $\mathbf{u}^k$ ; al avanzar un período, se define  $k := k+1$  y vuelve a resolverse (2-84).

En la práctica, el costo computacional de evaluar la función  $f_{eco}$  puede ser prohibitivo para la frecuencia de cálculo del MPC, por lo que se recurre a linealizaciones (Alamo et al., 2012; Souza et al., 2009).

### 2.11.2. RTO dinámica (DRTO)

El problema de optimización en tiempo real puede formularse directamente como un problema dinámico (Adetola y Guay, 2010; Biegler, 2009; Ochoa et al., 2010; Tosukhowong et al., 2004).

El problema en este caso puede formularse como (Peters et al., 2007):

$$\begin{aligned}
\min_{\mathbf{u}} J(\mathbf{u}) &= \int_{t^k}^{t^{k+T}} q(\mathbf{x}(t), \mathbf{u}(t)) dt \\
s.a. \quad \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{u}) \\
\mathbf{x}(t^k) &= \mathbf{x}^k \\
\mathbf{g}^f(\mathbf{x}(t^{k+T}), \mathbf{u}(t^{k+T})) &\leq 0 \\
\mathbf{g}^t(\mathbf{x}(t), \mathbf{u}(t)) &\leq 0
\end{aligned} \tag{2-85}$$

donde  $\mathbf{g}^f$  son restricciones para el tiempo final, y  $\mathbf{g}^t$  para la trayectoria. El tiempo final en (2-85) está expresado según un horizonte deslizante (el tiempo de integración es siempre  $T$ ), que puede ser útil para procesos continuos. Para procesos discontinuos, puede utilizarse un horizonte fijo (con el límite superior de integración constante).

Las estrategias de adaptación del modelo pueden ser similares a la RTO tradicional. Bonvin y Srinivasan (2013) mencionan la estrategia de dos pasos, la adaptación por modificadores y el *NCO tracking* como posibles técnicas a utilizar en DRTO.

## 2.12. APLICACIONES INDUSTRIALES

La optimización en tiempo real se utiliza ampliamente en la industria, sobre todo en el área de refinación y petroquímica. Darby et al. (2011) estiman que, a la fecha de publicación de su trabajo, se reportaban 250-300 implementaciones exitosas de RTO utilizando herramientas de software comerciales con modelado riguroso.

Entre las aplicaciones existentes, pueden mencionarse plantas de etileno y olefinas (Rejowski et al., 2009; Valleru et al., 2012); unidades de destilación de aire (Puranik et al., 2013); en refinerías, unidades de craqueo catalítico (FCC) (Zanin et al., 2002), unidades de crudo (Liporace et al., 2009) y sistemas de servicios auxiliares (Lorenz y Roberto, 2009; Majchrowicz et al., 2011; Ruiz et al., 2005).

La gran mayoría de las aplicaciones industriales utilizan modelos continuos. Sin embargo, algunas de las aplicaciones mencionadas incluyen variables discretas en una formulación no lineal, mixta-entera (MINLP) (Puranik et al., 2013; Ruiz et al., 2005). Las publicaciones académicas, mientras tanto, utilizan una formulación de programación no lineal (NLP). No se han encontrado trabajos publicados que traten formalmente las estrategias de adaptación, estimación de gradientes o filtrado de los resultados óptimos, entre otros aspectos, que incluyan formulaciones basadas en decisiones discretas.

### 2.13. OTROS USOS DE LAS TÉCNICAS DE RTO

Las técnicas de RTO pueden extenderse a otras aplicaciones industriales o científicas, en las que el comportamiento del proceso no se pueda conocer perfectamente. Entre ellas, se pueden mencionar (Bunin et al., 2013a):

- *run-to-run optimization* (François et al., 2004; Srinivasan y Bonvin, 2002): optimización del perfil dinámico de un proceso *batch*. Antes de comenzar cada ciclo del proceso, se determina la política de operación óptima (el perfil de las entradas  $\mathbf{u}(t)$ ), adaptando el modelo con datos de las corridas anteriores. Las técnicas de adaptación son las mismas que las de RTO para estado estacionario (Chachuat et al., 2009).
- Ajuste de controladores (Bunin et al., 2012a): las entradas óptimas a determinar son los parámetros de ajuste de los controladores.
- Resolución de problemas de optimización numérica donde el costo computacional de evaluar las funciones es elevado, y pueden aproximarse estas funciones con un modelo más sencillo (Bunin, 2013). Los ciclos “de RTO” consisten en adaptar el modelo aproximado evaluando las funciones del modelo original, para luego resolver un problema de optimización con el modelo adaptado; el ciclo se repite hasta que se alcanza la convergencia del sistema.

### 2.14. SCHEDULING EN PRESENCIA DE INCERTIDUMBRE

El *scheduling* es una herramienta clave en la operación de un gran número de procesos industriales. En los últimos años se han desarrollado una gran variedad de técnicas de modelado y de resolución de problemas para la búsqueda del *schedule* óptimo, que están fuera del alcance de este trabajo. Harjunoski et al. (2014) han presentado recientemente una extensa revisión del estado del arte y las aplicaciones industriales en esta área.

La determinación del programa de producción óptimo está sujeta a los mismos problemas de incertidumbre que la optimización en tiempo real. Como el período de tiempo abarcado por el *scheduling* es mayor (días, semanas), el problema se acentúa, ya que los precios, demandas y disponibilidad de recursos en el futuro no pueden ser determinados con exactitud. Por este motivo, se han desarrollado estrategias para ejecutar los programas de producción haciendo frente a la incertidumbre (Aytug et al., 2005).

Entre estas estrategias, puede mencionarse el *scheduling reactivo* (Méndez y Cerdá, 2003; Novas y Henning, 2010; Palombarini y Martínez, 2012). Cuando

ocurre algún evento inesperado que hace infactible el programa original, estas técnicas intentan *reparar* el schedule, buscando soluciones factibles que permitan cumplir con la producción prevista.

El *scheduling robusto* (Kouvelis et al., 2000) considera los parámetros inciertos del problema, y resuelve la optimización de manera tal que cumpla con las restricciones en el peor escenario posible. Si bien asegura la factibilidad del programa propuesto, en muchos casos puede ser un enfoque muy conservador que impide alcanzar el máximo beneficio posible del proceso.

El *scheduling multi-etapa* (Cui y Engell, 2009) divide el horizonte de cálculo en  $N$  etapas. Cada una contiene un conjunto de parámetros inciertos, que se asumen conocidos antes de que la etapa comience. Utilizando una distribución discreta de escenarios posibles para cada etapa, se genera un árbol de posibles realizaciones de los parámetros para todo el horizonte, y se resuelve el schedule teniendo en cuenta todos los escenarios posibles. Como al aumentar el número de etapas la complejidad computacional aumenta, el problema se suele abordar a través de dos etapas.

En lugar de aplicar la totalidad del schedule de acuerdo con la solución inicial, se han planteado alternativas en tiempo real, que actualizan el schedule con una frecuencia preestablecida, utilizando un esquema de horizonte móvil y un problema de dos etapas (Cui y Engell, 2010; Sand y Engell, 2004).

Engell y Harjunoski (2012) han analizado las posibilidades de integración del scheduling con el control de procesos, incluyendo en este último la optimización estacionaria (RTO). Se proponen varias estructuras: una *monolítica*, que intenta resolver todos los problemas en simultáneo; una *jerárquica*, que es la que se utiliza corrientemente (ver Figura 1.1); y una *colaborativa*, donde los resultados de la optimización y el control sean utilizados para refinar el schedule original.

## CONCLUSIONES PARCIALES

La optimización en tiempo real (RTO) es una metodología que ha alcanzado una madurez tanto desde el punto de vista del desarrollo conceptual como desde su aplicación industrial. Sin embargo, todavía presenta desafíos para la investigación y posibilidades de mejora, en campos como las estrategias de adaptación con corrección del gradiente, el uso de formulaciones con variables discretas para la optimización, y la integración con los métodos de scheduling óptimo. Estos aspectos serán abordados en los próximos capítulos de esta tesis.

### 3. ADAPTACIÓN POR MODIFICADORES BASADA EN ECUACIONES DE EFICIENCIA

#### RESUMEN

El capítulo 3 propone una variante al método tradicional de adaptación por modificadores, que concentra la adaptación en un subconjunto de ecuaciones del modelo, llamadas ecuaciones de eficiencia. Se presentan los fundamentos de la estrategia, que incluye la definición de variables auxiliares para la corrección de cada gradiente (3.1). Se proponen dos estrategias de implementación; la primera se basa en una estimación numérica del gradiente combinada con un filtro exponencial, y la segunda en una regresión lineal de datos de ciclos previos de RTO (3.2). El desempeño de las dos estrategias se ilustra en dos casos de estudio, el primero analiza un sistema genérico formado por tres operaciones unitarias (3.3.1) y el segundo un sistema de calor y potencia de una planta de azúcar y etanol (3.3.2).

#### 3.1. DESCRIPCIÓN DEL MÉTODO

En ciertos tipos de procesos, todos los errores paramétricos y estructurales pueden asignarse a un subconjunto de las ecuaciones del modelo, que se llamarán de aquí en adelante *ecuaciones de eficiencia*. Estas ecuaciones pueden utilizarse, entre otros casos, para predecir velocidades de reacción, eficiencias de calderas o turbinas, o coeficientes de transferencia de calor. Las ecuaciones de eficiencia se obtienen en forma empírica o semiempírica, y pueden corregirse a través de las mediciones disponibles en línea. El resto de las ecuaciones del modelo, incluyendo balances de masa, energía y entropía, pueden ser consideradas como libres de errores estructurales, aunque podrían incluir, como parámetro, uno o más términos obtenidos de las ecuaciones de eficiencia (por ejemplo, el balance de energía de una caldera puede incluir la eficiencia, calculada a través de una ecuación empírica).

El problema de optimización en tiempo real a resolver puede formularse como:

$$\begin{aligned}
& \min_{\mathbf{y}, \mathbf{x}, \mathbf{u}, \boldsymbol{\eta}} Q(\mathbf{y}, \mathbf{u}) \\
& s.t. \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\eta}) = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}, \mathbf{u}) \\ \mathbf{f}_2(\mathbf{x}, \mathbf{u}, \boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \\
& \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}) \\
& \boldsymbol{\eta} = \mathbf{p}_m(\mathbf{u}) \\
& \mathbf{g}(\mathbf{y}, \mathbf{x}, \mathbf{u}) \leq \mathbf{0} \\
& \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^L
\end{aligned} \tag{3-1}$$

donde  $\mathbf{y}$  son las salidas del proceso (medidas),  $\mathbf{x}$  las variables de estado,  $\mathbf{u}$  las entradas del proceso (que también pueden estar medidas),  $\mathbf{f}$  el modelo del proceso (balances de masa, energía y entropía),  $\mathbf{h}$  la función que relaciona las salidas  $\mathbf{y}$  con respecto a las entradas  $\mathbf{u}$  y los estados  $\mathbf{x}$ , y  $\mathbf{g}$  las restricciones de desigualdad del problema. La funcionalidad de los factores de eficiencia  $\boldsymbol{\eta}$  con respecto a las entradas se conoce sólo en forma aproximada a través del modelo  $\mathbf{p}_m(\mathbf{u})$ . Toda la incertidumbre paramétrica y estructural está concentrada en este modelo  $\mathbf{p}_m$ . Se asume que el submodelo  $\mathbf{f}_2$  permite obtener las eficiencias  $\boldsymbol{\eta}$  si se conocen los valores de  $\mathbf{u}$  e  $\mathbf{y}$ .

Por otra parte, el modelo para cada eficiencia  $\eta_j$  puede expresarse como una función de nuevas variables  $\mathbf{v}_j$ , que pueden obtenerse a su vez como función de las entradas  $\mathbf{u}$ :

$$\eta_j = p_{mj}(\mathbf{v}_j(\mathbf{u})) \quad j = 1, 2, \dots, nj \tag{3-2}$$

donde  $nj$  es el número de ecuaciones de eficiencia en el modelo.

Los siguientes ejemplos ilustran algunas de las posibles estructuras que pueden presentar las variables  $\mathbf{v}_j$  en función de un vector de entradas

$$\mathbf{u} = (u_1, u_2, u_3)^T :$$

- $\mathbf{v}$  es el vector de entradas  $\mathbf{u}$ :

$$\eta_1 = p_{m1}(u_1, u_2, u_3) \quad \mathbf{v}_1 = \mathbf{u} \tag{3-3}$$

- $\mathbf{v}$  es un subconjunto del vector  $\mathbf{u}$ :

$$\begin{aligned}
\eta_2 &= p_{m2}(u_1, u_3) & \mathbf{v}_2 &= (u_1, u_3)^T \\
\eta_3 &= p_{m3}(u_2) & \mathbf{v}_3 &= u_2
\end{aligned} \tag{3-4}$$

- $\mathbf{v}$  es una combinación lineal de los elementos de  $\mathbf{u}$ :  $\mathbf{v}(\mathbf{u}) = \mathbf{P} \cdot \mathbf{u}$ , donde  $\mathbf{P}$  es una matriz de dimensión  $nv \times nu$  (con  $nv$  el número de elementos de  $\mathbf{v}$ ):

$$\eta_4 = p_{m4}(u_1 - u_2, u_2 + u_3) \quad \mathbf{v}_4 = (u_1 - u_2, u_2 + u_3)^T \quad (3-5)$$

- $\mathbf{v}$  es una función continua de  $\mathbf{u}$ :

$$\eta_5 = p_{m5}(u_1 \cdot u_3) \quad v_5 = u_1 \cdot u_3 \quad (3-6)$$

En los sistemas estudiados en este capítulo, la dimensión de los vectores  $\mathbf{v}_j$  que modifican cada ecuación de eficiencia es menor o igual a la del vector de entradas  $\mathbf{u}$ . Esta situación es común en procesos formados por redes de operaciones unitarias (por ejemplo, sistemas de calor y potencia): en estos casos, la eficiencia de cada operación unitaria no es afectada por otros procesos que se ejecutan en paralelo o aguas abajo, sino sólo por las condiciones de entrada de la operación unitaria propiamente dicha.

Cuando esto sucede, los términos de corrección del gradiente pueden calcularse en función de  $\mathbf{v}_j$  en lugar de  $\mathbf{u}$ . Por lo tanto, la dimensión de los problemas de estimación del gradiente se reduce con respecto a las ecuaciones (2-37)-(2-38) ya que se requiere una menor cantidad de conjuntos de datos (es decir, información de menos ciclos de RTO) para estimar el término de corrección del gradiente en cada ecuación de eficiencia. Por otra parte, al utilizar el conocimiento de la planta para determinar los gradientes, se evita el cálculo experimental del gradiente en direcciones del espacio de las entradas en las que se sabe que la derivada direccional es nula.

Consideremos como ejemplo el caso dado por la ecuación (3-5), donde  $\mathbf{v}_j$  es una combinación lineal de las variables de  $\mathbf{u}$ . Si se sabe, por conocimiento de la planta, que la eficiencia  $\eta_j$  es función de  $\mathbf{v}_j = \begin{pmatrix} v_{j1} \\ v_{j2} \end{pmatrix} = \begin{pmatrix} u_1 - u_2 \\ u_2 + u_3 \end{pmatrix}$ , a la hora de estimar los gradientes con respecto a  $\mathbf{u}$ , se tiene:

$$\frac{\partial \eta}{\partial u_1} = \frac{\partial \eta}{\partial v_{j1}} \quad (3-7)$$

$$\frac{\partial \eta}{\partial u_2} = -\frac{\partial \eta}{\partial v_{j1}} + \frac{\partial \eta}{\partial v_{j2}} \quad (3-8)$$

$$\frac{\partial \eta}{\partial u_3} = \frac{\partial \eta}{\partial v_{j2}} \quad (3-9)$$

entonces puede deducirse, reemplazando (3-7) y (3-9) en (3-8), que

$$\frac{\partial \eta}{\partial u_2} = -\frac{\partial \eta}{\partial u_1} + \frac{\partial \eta}{\partial u_3} \quad (3-10)$$

Esta información es valiosa a la hora de estimar los gradientes en forma experimental, ya que si no se utiliza se añade un grado de libertad al problema de estimación (lo que requiere un dato más para el cálculo), y conjuntamente se genera una fuente de error, ya que los valores calculados para las derivadas parciales no cumplirán con (3-10).

El problema de adaptación del modelo utiliza las mediciones disponibles para realizar las correcciones necesarias en las ecuaciones de eficiencia. Contando con mediciones para las salidas  $\mathbf{y}$ , y con una estructura apropiada del modelo (la propuesta en la ecuación (3-1)), la adaptación puede descomponerse en tres etapas: (1) reconciliación de datos, (2) cálculo de las eficiencias, y (3) adaptación de las ecuaciones de eficiencia.

#### *Etapa 1: Reconciliación de datos*

Los valores reconciliados  $(\mathbf{y}^k, \mathbf{u}^k)$  para el ciclo de RTO  $k$  se obtienen de los valores medidos  $(\mathbf{y}^{k,m}, \mathbf{u}^{k,m})$  como:

$$\begin{aligned} (\mathbf{y}^k, \mathbf{x}^k, \mathbf{u}^k) &= \underset{\mathbf{y}, \mathbf{x}, \mathbf{u}}{\operatorname{arg\,min}} \left( \mathbf{y}^{m,k} - \mathbf{y} \right)^T \cdot A^{-1} \cdot \left( \mathbf{y}^{m,k} - \mathbf{y} \right) + \left( \mathbf{u}^{m,k} - \mathbf{u} \right)^T \cdot B^{-1} \cdot \left( \mathbf{u}^{m,k} - \mathbf{u} \right) \\ \text{s.t. } \mathbf{f}_1(\mathbf{x}, \mathbf{u}) &= 0 \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{u}) \end{aligned} \quad (3-11)$$

donde  $A$  y  $B$  son matrices de variancia-covariancia para las salidas  $\mathbf{y}$  y para las entradas  $\mathbf{u}$ , respectivamente. (Chen et al., 1997; Narasimhan y Jordache, 1999; Romagnoli y Sanchez, 1999).

#### *Etapa 2: Cálculo de las eficiencias*

Usando los valores reconciliados obtenidos de (3-11), el sistema de ecuaciones

$$\mathbf{f}_2(\mathbf{x}^k, \mathbf{u}^k, \boldsymbol{\eta}^k) = 0 \quad (3-12)$$

se resuelve para obtener las eficiencias actuales  $\boldsymbol{\eta}^k$ .

Si la estructura del modelo no corresponde exactamente a (3-1), el cálculo de las eficiencias puede resolverse en forma simultánea con la etapa de reconciliación de datos, incluyendo la ecuación (3-12) como restricción del problema (3-11). Esta estrategia también puede ser útil cuando se cuenta con

valores máximos y mínimos para las eficiencias: en este caso, estos límites también se incluyen como restricciones en el problema de reconciliación, y aportan información para obtener valores reconciliados confiables (Narasimhan y Jordache, 1999).

*Etapa 3: Adaptación de las ecuaciones de eficiencia*

Las ecuaciones de eficiencia pueden adaptarse para cada ciclo de RTO  $k$  a través de una estructura de adaptación por modificadores. En lugar de corregir la función objetivo y las restricciones de desigualdad (como se describe en la sección 2.3.2), en este caso la estrategia se aplica a las ecuaciones de eficiencia:

$$\eta_j = p_{mj}(\mathbf{v}_j) + \beta_j^k + (\boldsymbol{\lambda}_j^k)^T \cdot (\mathbf{v}_j - \mathbf{v}_j^k) \quad j = 1, 2, \dots, nj \quad (3-13)$$

donde el modificador de restricción  $\beta_j^k$  se calcula para cada ecuación  $j$  como

$$\beta_j^k = \eta_j^k - p_{mj}(\mathbf{v}_j^k) \quad (3-14)$$

y el modificador del gradiente para la misma ecuación se obtiene como

$$\boldsymbol{\lambda}_j^k = \nabla_{\mathbf{v}} \eta_j \Big|_{\mathbf{v}_j = \mathbf{v}_j^k} - \nabla_{\mathbf{v}} p_{mj} \Big|_{\mathbf{v}_j = \mathbf{v}_j^k} \quad (3-15)$$

El gradiente real de  $\eta_j$  con respecto a  $\mathbf{v}_j$  no se conoce en la práctica (ya que sólo está disponible el modelo aproximado  $p_{mj}(\mathbf{v}_j)$ ). Por lo tanto, debe obtenerse en forma experimental, a través de datos de planta. Una de las formas de estimar el gradiente es utilizando información del estado estacionario actual y de los estados estacionarios alcanzados al aplicar la RTO en ciclos anteriores:

$$\nabla_{\mathbf{v}} \eta_j^k \approx [\eta_j^k - \eta_j^{k-1} \quad \eta_j^k - \eta_j^{k-2} \quad \dots \quad \eta_j^k - \eta_j^{k-ni_j}] \cdot (\mathbf{V}_j^k)^{-1} \quad j = 1, 2, \dots, nj \quad (3-16)$$

donde

$$\mathbf{V}_j^k = [\mathbf{v}_j^k - \mathbf{v}_j^{k-1} \quad \mathbf{v}_j^k - \mathbf{v}_j^{k-2} \quad \dots \quad \mathbf{v}_j^k - \mathbf{v}_j^{k-ni_j}] \quad (3-17)$$

y  $ni_j$  es la dimensión de los vectores  $\mathbf{v}_j$  y  $\nabla_{\mathbf{v}} \eta_j$ . Como se mencionó antes, esta estrategia es útil si  $ni_j$  es menor que  $nu$ , ya que esto permite una reducción en la dimensión del problema de estimación del gradiente. Cuanto menor sea  $ni_j$ , más factible será una implementación de RTO estimando gradientes con datos de ciclos anteriores, ya que al aumentar el número de ciclos pasados, aumenta la posibilidad de que cambie el valor de variables no controladas ni manipuladas

(*disturbancias*) en el proceso, y que invaliden la estimación del gradiente (Marchetti et al., 2010).

Si se quieren obtener los gradientes con respecto a las variables independientes  $\mathbf{u}$ , pueden calcularse siguiendo la regla de la cadena:

$$\frac{d\eta_j}{d\mathbf{u}} = \left( \frac{d\eta_j}{d\mathbf{v}_j} \right)^T \cdot \frac{d\mathbf{v}_j}{d\mathbf{u}} = \frac{dp_{mj}}{d\mathbf{u}} + \left( \boldsymbol{\lambda}_j^k \right)^T \cdot \frac{d\mathbf{v}_j}{d\mathbf{u}}, \quad j = 1 \dots nj \quad (3-18)$$

Esto permitiría corregir en la forma “tradicional” la ecuación de eficiencia como

$$\eta_j = p_{mj}(\mathbf{u}) + \beta_j^k + \left( \boldsymbol{\lambda}_j^k \right)^T \cdot \frac{d\mathbf{v}_j}{d\mathbf{u}} \cdot (\mathbf{u} - \mathbf{u}^k), \quad j = 1 \dots nj \quad (3-19)$$

Si todos los errores estructurales y paramétricos pueden incluirse en las ecuaciones de eficiencia, y la corrección del gradiente puede obtenerse en función de las variables propuestas  $\mathbf{v}$ , todos los gradientes de la función objetivo y de las restricciones pueden calcularse correctamente una vez calculados los modificadores  $\boldsymbol{\lambda}_j^k$ . Por lo tanto, si el sistema de RTO converge, lo hará a un punto que cumple las condiciones de Karush-Kuhn-Tucker para la planta real. Las condiciones (necesarias) de KKT para la planta pueden expresarse como:

$$\begin{aligned}
 \nabla_{\mathbf{u}} Q + \boldsymbol{\mu}_1^T \cdot \nabla_{\mathbf{u}} \mathbf{f}_1 + \boldsymbol{\mu}_2^T \cdot \nabla_{\mathbf{u}} \mathbf{f}_2 - \boldsymbol{\mu}_h^T \cdot \nabla_{\mathbf{u}} \mathbf{h} + \boldsymbol{\mu}_g^T \cdot \nabla_{\mathbf{u}} \mathbf{g} - \boldsymbol{\pi}^L + \boldsymbol{\pi}^U &= \mathbf{0} \\
 \boldsymbol{\mu}_1^T \cdot \nabla_{\mathbf{x}} \mathbf{f}_1 + \boldsymbol{\mu}_2^T \cdot \nabla_{\mathbf{x}} \mathbf{f}_2 - \boldsymbol{\mu}_h^T \cdot \nabla_{\mathbf{x}} \mathbf{h} + \boldsymbol{\mu}_g^T \cdot \nabla_{\mathbf{x}} \mathbf{g} &= \mathbf{0} \\
 \nabla_{\mathbf{y}} Q &+ \boldsymbol{\mu}_h &+ \boldsymbol{\mu}_g^T \cdot \nabla_{\mathbf{y}} \mathbf{g} &= \mathbf{0} \\
 \mathbf{f}_1(\mathbf{x}^*, \mathbf{u}^*) &= \mathbf{0} \\
 \mathbf{f}_2(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\eta}(\mathbf{u}^*)) &= \mathbf{0} \\
 \mathbf{y}^* - \mathbf{h}(\mathbf{x}^*, \mathbf{u}^*) &= \mathbf{0} \\
 \mathbf{g}(\mathbf{y}^*, \mathbf{x}^*, \mathbf{u}^*) &\leq \mathbf{0} \\
 \mathbf{u}^L - \mathbf{u}^* &\leq \mathbf{0} \\
 \mathbf{u}^* - \mathbf{u}^U &\leq \mathbf{0} \\
 \boldsymbol{\mu}_g^T \cdot \mathbf{g}(\mathbf{y}^*, \mathbf{x}^*, \mathbf{u}^*) &= 0 \\
 (\boldsymbol{\pi}^L)^T \cdot (\mathbf{u}^L - \mathbf{u}^*) &= 0 \\
 (\boldsymbol{\pi}^U)^T \cdot (\mathbf{u}^* - \mathbf{u}^U) &= 0 \\
 \boldsymbol{\mu}_g &\geq \mathbf{0} \\
 \boldsymbol{\pi}^L &\geq \mathbf{0} \\
 \boldsymbol{\pi}^U &\geq \mathbf{0}
 \end{aligned} \tag{3-20}$$

donde  $\boldsymbol{\mu}_1$ ,  $\boldsymbol{\mu}_2$ ,  $\boldsymbol{\mu}_h$  y  $\boldsymbol{\mu}_g$  son los multiplicadores de Lagrange correspondientes a las restricciones  $\mathbf{f}_1$ ,  $\mathbf{f}_2$ ,  $\mathbf{h}$  y  $\mathbf{g}$ .  $\boldsymbol{\pi}^L$  y  $\boldsymbol{\pi}^U$  son los multiplicadores correspondientes a los límites superior e inferior de  $\mathbf{u}$ . En la ecuación (3-20), los únicos términos que contienen error estructural son  $\mathbf{f}_2(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\eta}(\mathbf{u}^*))$  y  $\nabla_{\mathbf{u}} \mathbf{f}_2(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\eta}(\mathbf{u}^*))$ , ya que se desconoce la funcionalidad exacta de  $\boldsymbol{\eta}(\mathbf{u})$ . Cuando el sistema ha convergido, los valores estimados son:

$$\mathbf{f}_2 \left( \mathbf{x}^*, \mathbf{u}^*, \mathbf{p}_m(\mathbf{u}^*) + \boldsymbol{\beta}^k + \begin{pmatrix} (\boldsymbol{\lambda}_1^k)^T (\mathbf{v}_1^* - \mathbf{v}_1^k) \\ \vdots \\ (\boldsymbol{\lambda}_{nj}^k)^T (\mathbf{v}_{nj}^* - \mathbf{v}_{nj}^k) \end{pmatrix} \right) = 0 \tag{3-21}$$

$$\nabla_{\mathbf{u}} \mathbf{f}_2 \left( \mathbf{x}^*, \mathbf{u}^*, \mathbf{p}_m(\mathbf{u}^*) + \boldsymbol{\beta}^k + \begin{pmatrix} (\boldsymbol{\lambda}_1^k)^T (\mathbf{v}_1^* - \mathbf{v}_1^k) \\ \vdots \\ (\boldsymbol{\lambda}_{nj}^k)^T (\mathbf{v}_{nj}^* - \mathbf{v}_{nj}^k) \end{pmatrix} \right) = \frac{\partial \mathbf{f}_2}{\partial \mathbf{u}} + \frac{\partial \mathbf{f}_2}{\partial \boldsymbol{\eta}} \cdot \left( \nabla_{\mathbf{u}} \mathbf{p}_m + \sum_{j=1}^{nj} (\boldsymbol{\lambda}_j^k)^T \cdot \frac{\partial \mathbf{v}_j}{\partial \mathbf{u}} \right) \tag{3-22}$$

Como en la convergencia  $\mathbf{v}_1^* = \mathbf{v}_1^k$  y  $\mathbf{u}_1^* = \mathbf{u}_1^k$ , y teniendo en cuenta que  $\boldsymbol{\beta}^k = \boldsymbol{\eta}(\mathbf{u}^*) - \mathbf{p}_m(\mathbf{u}^*)$  (de (3-14)), la ecuación (3-21) equivale a

$$\mathbf{f}_2(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\eta}(\mathbf{u}^*)) = 0 \quad (3-23)$$

y utilizando además (3-18), la ecuación

(3-22) equivale a

$$\nabla_{\mathbf{u}} \mathbf{f}_2(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\eta}(\mathbf{u}^*)) = \frac{\partial \mathbf{f}_2}{\partial \mathbf{u}} + \frac{\partial \mathbf{f}_2}{\partial \boldsymbol{\eta}} \cdot \frac{\partial \boldsymbol{\eta}}{\partial \mathbf{u}} \quad (3-24)$$

Como las ecuaciones (3-23) y (3-24) son válidas tanto para el modelo adaptado como para la planta, al reemplazarlas en (3-20) las condiciones de óptimo se cumplen tanto para el modelo como para la planta real.

## 3.2. IMPLEMENTACIÓN DE LAS ESTRATEGIAS DE ADAPTACIÓN

### 3.2.1. Alternativa 1: Corrección + Filtro exponencial (MA-E1)

El modelo puede adaptarse usando las ecuaciones (3-13) a (3-17). Para reducir el efecto del ruido en las mediciones y para favorecer la convergencia de la estrategia, se aplica un filtro exponencial a los modificadores calculados (Marchetti et al., 2009):

$$\begin{aligned} \boldsymbol{\beta}^k &:= K_{\beta} \cdot (\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}) + \boldsymbol{\beta}^{k-1} \\ \boldsymbol{\lambda}_j^k &:= K_{\lambda_j} \cdot (\boldsymbol{\lambda}_j^k - \boldsymbol{\lambda}_j^{k-1}) + \boldsymbol{\lambda}_j^{k-1}, \quad j = 1..nj \end{aligned} \quad (3-25)$$

donde  $K_{\beta}$  y  $K_{\lambda_j}$  ( $j = 1..nj$ ) son matrices diagonales con sus elementos entre 0 y 1. Como alternativa, para aquellos parámetros que cambian lentamente con el tiempo, puede utilizarse una media móvil en lugar del filtro exponencial.

Para limitar el efecto del ruido en las mediciones, se pueden definir valores máximos y mínimos para los modificadores  $\boldsymbol{\beta}^k$  y  $\boldsymbol{\lambda}_j^k$ . Sin embargo, si las ecuaciones de eficiencia tienen un error paramétrico y estructural significativo con respecto al desempeño real de la planta, el uso de valores máximos y mínimos implica que ya no está garantizado que se alcance un punto KKT cuando el sistema converge. Esto es así porque si en el óptimo de la planta el error en una ecuación  $j$  está fuera de los límites permitidos para el parámetro  $\beta_j$ , o el error en

el cálculo del gradiente no está incluido en el rango permitido para  $\lambda_j$ , el sistema de RTO no podrá replicar en el modelo las condiciones de KKT para este punto.

Para generar puntos útiles para la estimación del gradiente, puede aplicarse una estrategia dual. Por ejemplo, la propuesta de Marchetti et al. (2010) agrega restricciones al problema de optimización en el ciclo de RTO  $k$  generando dos regiones factibles desconectadas. Dentro de estas regiones, el error de estimación del gradiente por truncamiento y por ruido en el ciclo  $k+1$  está acotado. Otras estrategias generan restricciones adicionales basadas en el número de condición de la matriz  $\mathbf{V}_j^k$  (Brdys y Tatjewski, 2005).

Los casos de estudio presentados en la Sección 3.3.1 implementan una estrategia simplificada, que definen un valor máximo  $\mathbf{m}$  para los cambios en las entradas del proceso:

$$\|\mathbf{u} - \mathbf{u}_k\| \leq \mathbf{m} \quad (3-26)$$

Como esta restricción no garantiza que el problema de estimación del gradiente (ecuación (3-16)) estará bien condicionado, se define un criterio para actualizar cada modificador de gradiente  $\lambda_j$ : el modificador sólo será actualizado si el cambio en las variables asociadas  $\mathbf{v}_j$  es mayor a un valor mínimo. Como se verá en la sección 3.3.1.3, este criterio se aplica para el caso en que  $\mathbf{v}_j$  tiene dimensión 1; cuando la dimensión de  $\mathbf{v}_j$  es mayor a 1, deberán implementarse otros criterios que eviten que el problema (3-16) esté mal condicionado, o que limiten el error de estimación del gradiente.

### 3.2.2. Alternativa 2: Regresión lineal ponderada (MA-E2)

Las eficiencias  $\eta_j$  pueden adaptarse de acuerdo con una función genérica

$\gamma_j^k$ :

$$\eta_j = p_{mj}(\mathbf{u}) + \gamma_j^k(\mathbf{v}_j(\mathbf{u})) \quad j = 1, 2, \dots, nj \quad (3-27)$$

El término de adaptación  $\gamma_j^k$  se obtiene por regresión utilizando un horizonte móvil. Por ejemplo, puede proponerse una función de adaptación cuadrática:

$$\eta_j = p_{mj}(\mathbf{u}) + a_j^k + (\mathbf{b}_j^k)^T \cdot \mathbf{v}_j + \mathbf{v}_j^T \cdot \mathbf{C}_j^k \cdot \mathbf{v}_j \quad j = 1, 2, \dots, nj \quad (3-28)$$

Los coeficientes pueden obtenerse a través de una regresión lineal ponderada:

$$\begin{aligned}
 \min_{a^k, \mathbf{b}^k, C^k} \sum_{j=1}^{nj} \sum_{i=k-N}^k w^i (\beta_j^i - \gamma_j^i)^2 &= \sum_{j=1}^{nj} \left[ \min_{a_j^k, \mathbf{b}_j^k, C_j^k} \sum_{i=k-N}^k w^i (\beta_j^i - \gamma_j^i)^2 \right] \\
 \left. \begin{aligned}
 \beta_j^i &= \eta_j^i - p_{mj}(\mathbf{v}_j^i) \\
 \gamma_j^i &= a_j^k + (\mathbf{b}_j^k)^T \cdot \mathbf{v}_j^i + (\mathbf{v}_j^i)^T C_j^k \mathbf{v}_j^i \\
 \mathbf{v}_j^i &= \mathbf{v}_j(\mathbf{u}^i) \\
 a_j^L &\leq a_j^k \leq a_j^U \\
 \mathbf{b}_j^L &\leq \mathbf{b}_j^k \leq \mathbf{b}_j^U \\
 C_j^L &\leq C_j^k \leq C_j^U
 \end{aligned} \right\} \begin{aligned}
 &j = 1..nj \\
 &i = (k - N), \dots, k
 \end{aligned} \quad (3-29)
 \end{aligned}$$

donde  $N$  es el número de datos utilizados para la regresión, y  $w^i$  es el peso asignado a cada dato.

La inclusión de las matrices  $C_j^k$  agrega un término cuadrático a la estrategia de adaptación, lo que provee más grados de libertad con respecto a la corrección lineal de la estrategia de la sección 3.2.1 y de la adaptación por modificadores tradicional. Estos términos cuadráticos no son necesarios para obtener la propiedad deseada de alcanzar un punto KKT de la planta cuando el sistema converge. Sin embargo, pueden ser útiles para generar una mejor estimación del gradiente cuando las ecuaciones de eficiencia no son lineales. Por otra parte, estos términos pueden aproximar mejor la funcionalidad real de los términos de eficiencia  $\eta_j$ , aumentando el rango de validez del modelo adaptado.

Sin embargo, la inclusión de términos de adaptación cuadráticos requiere de mayor cantidad de información que una adaptación lineal. Además, estos términos no deben incluirse en el problema de optimización (3-29) en los primeros ciclos de RTO, cuando el sistema se inicializa, ya que el problema no estaría completamente especificado y los parámetros de la función cuadrática podrían tomar cualquier valor. Por otra parte, la predicción de valores extrapolados (fuera del rango definido por los  $N$  conjuntos de datos de problema (3-29)) podría no ser precisa si la funcionalidad real del término de corrección no es cuadrática. Para resolver este problema, se sugiere utilizar valores conservadores para los límites superior e inferior  $C_j^U$  y  $C_j^L$  (ver ejemplos en la sección 3.3).

Los pesos  $w^i$  pueden adoptarse siguiendo diferentes criterios. En los casos de estudio de la sección 3.3, el criterio seleccionado es asignar un peso de 1 a la información del ciclo actual, y los restantes pesos decayendo en forma sucesiva

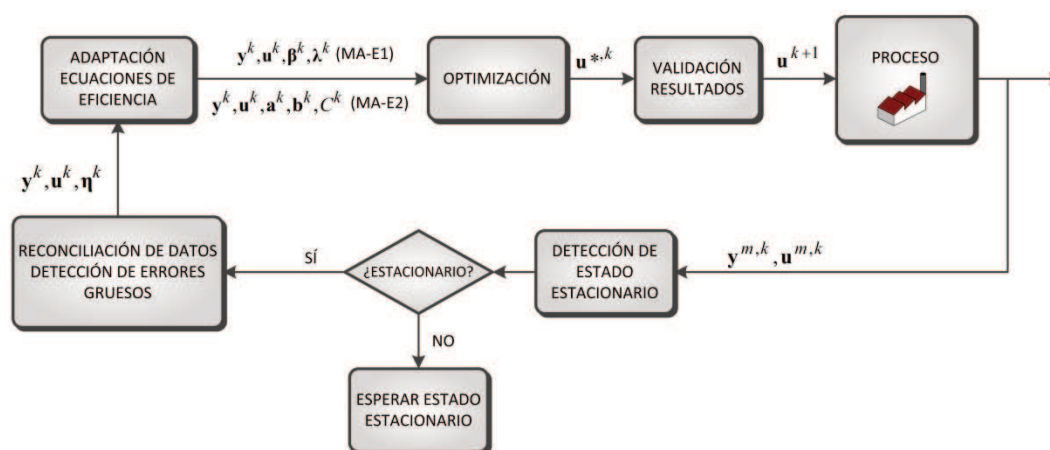
para los conjuntos de datos más antiguos. Esta estrategia penaliza la probabilidad de *disturbancias* de baja frecuencia que pueden hacer que los datos más antiguos aporten información de inferior calidad para la estimación del gradiente actual. Al mismo tiempo, si el sistema está convergiendo al óptimo de la planta, los datos de los últimos ciclos probablemente estarán a una menor distancia del punto actual. Por lo tanto, reciben un peso mayor (en comparación con los datos más antiguos), ya que aportan más información para la estimación local del gradiente. Otros autores sugieren un peso basado en la distancia con respecto al punto operativo actual y en el nivel de ruido (Bunin et al., 2013b; Bunin y François, 2012).

Aunque la estrategia de regresión lineal presentada en esta sección no requiere de restricciones adicionales para asegurar el buen condicionamiento del problema de estimación del gradiente (Bunin y François, 2012), el conjunto de datos utilizado para resolver el problema (3-29) podría contener información insuficiente para una estimación precisa de los parámetros de ajuste propuestos. Esta situación podría ser particularmente problemática cuando el sistema de RTO converge, ya que si no aparecen *disturbancias*, el sistema permanecerá en el entorno de este punto, y la calidad de la estimación de los parámetros lineales y cuadráticos disminuirá. En caso de ser necesario, podría realizarse un chequeo de la calidad de la estimación, y forzar experimentos (perturbaciones en los valores de las entradas) para aumentar esta calidad (Yip y Marlin, 2003). En los casos de estudio de la sección 3.3, no se implementó este tipo de validación; sin embargo, la estrategia propuesta mostró un buen desempeño.

Yip y Marlin (2002) sugieren una estructura que tiene analogías con la estimación propuesta en (3-29). El trabajo de Yip y Marlin propone una estructura de dos pasos: en el primero, se emplean múltiples conjuntos de datos para resolver una reconciliación de datos en simultáneo con un problema de estimación de parámetros; en el segundo, se minimiza la función objetivo de RTO. El objetivo de dicha propuesta es ajustar parámetros que no pueden obtenerse con un solo conjunto de datos. En cambio, el método planteado en esta sección, que también utiliza múltiples conjuntos de datos, tiene como objetivo alcanzar el óptimo real de la planta (convergiendo a un punto KKT de la planta real), implementando una estrategia de adaptación por modificadores. El problema de reconciliación, por otra parte, se resuelve en forma separada del de adaptación. Además, se incluye el uso de las variables auxiliares  $v$ , que reducen el problema de estimación del gradiente y son útiles para ciertos tipos de procesos, como el sistema de calor y potencia analizado en la sección 3.3.2.

### 3.2.3. Estructura propuesta

La **Figura 3.1** muestra un diagrama ilustrando la estructura de RTO propuesta. Como se mencionó en el capítulo 2, se incluye un paso inicial de detección de estado estacionario, que no se aborda en detalle en este capítulo. La literatura incluye abundantes referencias sobre esta técnica (Bhat y Saraf, 2004; Cao y Rhinehart, 1995; Roux et al., 2008).



**Figura 3.1.** Esquema de la estructura de RTO propuesta.

### 3.3. CASOS DE ESTUDIO

En esta sección se presentan y analizan dos casos de estudio para mostrar el comportamiento y el desempeño de las estructuras propuestas. En primer lugar, las principales propiedades del esquema presentado en la sección 3.2 se desarrollan y ejemplifican a través de un proceso genérico (Caso de estudio 1). A continuación, se considera un caso de estudio más complejo y realista, que consiste en la optimización del sistema de calor y potencia de una planta de producción de azúcar y etanol (Caso de estudio 2).

La metodología aplicada para analizar el desempeño del sistema de RTO se utiliza ampliamente en las publicaciones del área (Yip y Marlin, 2004). Consiste en emplear dos modelos: el primero, llamado *planta real*, representa exactamente el sistema a optimizar. Se utiliza para simular el estado estacionario alcanzado por la planta y para evaluar el valor real de la función objetivo. El segundo, llamado *modelo RTO*, difiere de la *planta real* en forma estructural y paramétrica. Un conjunto de los resultados de la *planta real*, llamados *mediciones*, se utiliza para la reconciliación de datos y la adaptación del *modelo RTO*. La optimización de la planta se realiza con el *modelo RTO* adaptado; las entradas óptimas calculadas son las variables independientes para la simulación de la *planta real*, lo que permite evaluar el costo real obtenido y obtener nuevas *mediciones* para el ciclo siguiente.

Para evaluar el desempeño de las estrategias de adaptación propuestas en la sección 3.2, y para compararlas con otros esquemas de RTO, la métrica que se emplea es el Costo de Diseño Extendido ( $C_E$ ) (Zhang y Forbes, 2000). Esta métrica calcula el beneficio potencial que no es aprovechado por un dado esquema de RTO, tomando como referencia el mejor valor posible de la función objetivo (es decir, el óptimo real de la planta, que se obtiene resolviendo el problema (3-1) con el modelo *planta real*).

### 3.3.1. Caso de estudio 1: Sistema genérico

#### 3.3.1.1. Descripción del caso

La Figura 3.2 muestra el proceso analizado en este caso de estudio. El sistema estudiado consiste en tres operaciones unitarias, con dos grados de libertad. Las variables de entrada seleccionadas son  $\mathbf{u} = [F_1, F_2]^T$ . Cada proceso (operación unitaria) está caracterizado por una eficiencia  $\eta_j$ . Las variables de salida son  $\mathbf{y} = [y_1, y_2, y_3]^T$ . Se conoce que cada eficiencia  $\eta_j$  es función del correspondiente variable  $F_j$ .

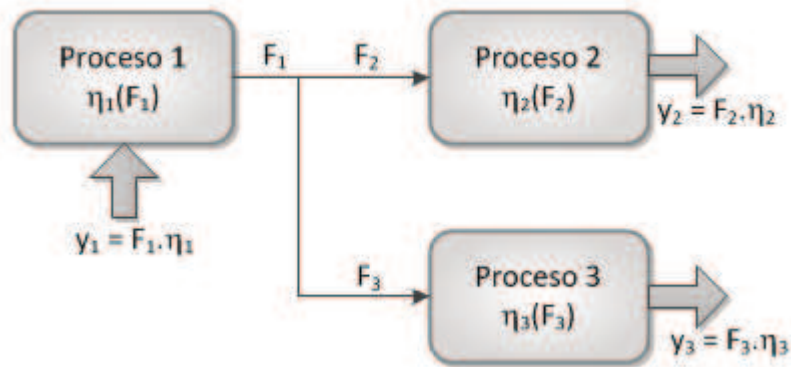


Figura 3.2. Sistema modelado en el caso de estudio I.

El proceso se modeló de la siguiente manera:

$$F_3 = F_1 - F_2 \quad (3-30)$$

$$x_j = \eta_j \cdot F_j, j = 1, 2, 3 \quad (3-31)$$

$$y_j = x_j, j = 1, 2, 3 \quad (3-32)$$

$$\eta_j = p_{mj}(F_j), j = 1, 2, 3 \quad (3-33)$$

$$F_j^L \leq F_j \leq F_j^U \quad (3-34)$$

$$F^L = \begin{pmatrix} 80 \\ 55 \\ 20 \end{pmatrix} ; F^U = \begin{pmatrix} 120 \\ 90 \\ 50 \end{pmatrix}$$

y la función objetivo se expresa como

$$Q_p = y_1 - y_2 - y_3 \quad (3-35)$$

Comparando con la estructura propuesta en (3-1), la ecuación (3-30) corresponde al vector  $\mathbf{f}_1$ ; (3-31) a  $\mathbf{f}_2$ ; (3-32) a  $\mathbf{h}$ ; y (3-33) al modelo  $\mathbf{p}_m(\mathbf{u})$ . Las restricciones de desigualdad y los límites superior e inferior de las entradas están contenidos en (3-34). Los estados  $x_j$  en (3-31) y (3-32) se incluyen para conservar la notación de (3-1); podrían omitirse reemplazando  $y_j$  en (3-31).

Las variables medidas son las entradas  $\mathbf{u} = [F_1, F_2]^T$ , y las salidas  $\mathbf{y} = [y_1, y_2, y_3]^T$ . La variable restante,  $F_3$ , se calcula por *balance* mediante (3-30).

Como cada variable de salida  $y_j$  es función de la variable  $F_j$ , toda la incertidumbre estructural y paramétrica está incluida en los modelos de la eficiencia  $p_{mj}(F_j)$  (3-33). Las restantes ecuaciones (3-30) y (3-35) se cumplen rigurosamente. Con el propósito de evaluar el desempeño del sistema de RTO propuesto, como se indicó al comienzo de la sección 3.3, se construyeron dos modelos: la *planta real* y el *modelo RTO*.

El comportamiento de la planta real se describe en forma exacta por las siguientes ecuaciones:

$$\begin{aligned} \eta_1 &= 10000 / (85 - 20 \cdot \exp(-(F_1 - 80)/190)) \\ \eta_2 &= 65 - 0.006 \cdot (F_2 - 80)^2 - \frac{24}{F_2} \\ \eta_3 &= 80 - 0.006 \cdot (F_3 - 35)^2 - \frac{12}{F_3} \end{aligned} \quad (3-36)$$

El modelo RTO sin adaptación se expresa como:

$$\begin{aligned} \eta_1 &= p_{m1}(F_1) = 10000 / (58 - 0.08 \cdot F_1) \\ \eta_2 &= p_{m2}(F_2) = 65 - 0.06 \cdot F_2 \\ \eta_3 &= p_{m3}(F_1 - F_2) = 50 + 0.04 \cdot (F_1 - F_2) \end{aligned} \quad (3-37)$$

El conocimiento del proceso permite identificar las variables  $v_j(\mathbf{u})$  presentadas en la ecuación (3-2). En este caso de estudio, las tres variables utilizadas  $[v_1, v_2, v_3]^T$  son unidimensionales, y cada una de ellas es una combinación lineal del vector de entradas  $[F_1, F_2]^T$ :

$$\begin{aligned} v_1 &= F_1 & \eta_1 &= \eta_1(v_1) \\ v_2 &= F_2 & \eta_2 &= \eta_2(v_2) \\ v_3 &= F_1 - F_2 = F_3 & \eta_3 &= \eta_3(v_3) \end{aligned} \quad (3-38)$$

En este caso de estudio, se asume que los datos están en estado estacionario, y que las mediciones no contienen errores gruesos. Aunque los valores medidos pueden contener ruido, no hay redundancia en las mediciones, por lo que no es posible realizar una reconciliación de datos.

### 3.3.1.2. Pasos a seguir en el sistema de RTO

Se analizaron cuatro implementaciones de RTO: La alternativa MA-E1 (sección 3.2.1), la alternativa MA-E2 (sección 3.2.2) y, como casos de comparación, una implementación de adaptación de las restricciones (B. Chachuat et al., 2008) y una implementación de adaptación por modificadores con actualización del gradiente a través del método de Broyden (Rodger y Chachuat, 2011).

Los pasos a seguir en todas las estrategias para cada ciclo  $k$  se describen a continuación:

1. Inicio:  $k = 0$
2. Mediciones: obtener los valores de  $F_1^k, F_2^k, y_1^k, y_2^k$  e  $y_3^k$  del modelo de la planta real.

3. Balances:

$$F_3^k = F_1^k - F_2^k \quad (3-39)$$

4. Cálculo de los factores de eficiencia:

$$\eta_j^k = \frac{y_j^k}{F_j^k}, \quad j = 1, 2, 3 \quad (3-40)$$

5. Adaptación del modelo: dependiente de la estrategia

6. Minimización del costo: el problema de optimización a resolver en cada ciclo (llamado  $P^k$  en las páginas subsiguientes) depende de la estrategia de adaptación. El resultado son las entradas óptimas  $(F_1^{*,k}, F_2^{*,k})$ . Los dos primeros ciclos se utilizan para obtener valores útiles para la estimación del gradiente, perturbando cada una de las entradas con un cambio  $\Delta$ . Por este motivo, la optimización se resuelve a partir del tercer ciclo, es decir, para  $k \geq 2$ .

$$\begin{aligned}
 &\text{IF } k \geq 2 \text{ THEN} \\
 &\quad (F_1^{*,k}, F_2^{*,k}) = \text{argmin}(P^k) \\
 &\text{ELSE} \\
 &\quad \text{IF } k = 0 \text{ THEN } (F_1^{*,k}, F_2^{*,k}) = (F_1^k + \Delta, F_2^k) \\
 &\quad \text{IF } k = 1 \text{ THEN } (F_1^{*,k}, F_2^{*,k}) = (F_1^k, F_2^k + \Delta) \\
 &\text{END IF}
 \end{aligned} \tag{3-41}$$

7. Implementación de la solución óptima: se aplica directamente el resultado, sin utilizar estrategias de filtrado de las entradas óptimas o validación estadística de los cambios.

$$\begin{aligned}
 F_1^{k+1} &= F_1^{*,k} \\
 F_2^{k+1} &= F_2^{*,k} \\
 k &:= k + 1
 \end{aligned} \tag{3-42}$$

8. Volver al paso 2.

Los pasos 5 y 6 se analizan por separado para cada estrategia de adaptación en las subsecciones siguientes.

### 3.3.1.3. Adaptación por modificadores con ecuaciones de eficiencia – Alternativa 1 (MA-E1)

Los pasos característicos de esta estrategia aplicados al caso de estudio son:

5. Adaptación del modelo:

- a. Cálculo del término modificador  $\beta_j^k$ , que corrige la función de eficiencia  $p_{mj}(F_j^k)$ :

$$\beta_j^k = \eta_j^k - p_{mj}(F_j^k) \tag{3-43}$$

- b. Actualización del término de corrección del gradiente  $\lambda_j^k$ :

$$\left. \begin{array}{l}
 \text{IF} \left( \left| F_j^k - F_j^{k-1} \right| \geq \delta \text{ AND } k \geq 1 \right) \text{ THEN} \\
 \lambda_j^k = \frac{\beta_j^k - \beta_j^{k-1}}{F_j^k - F_j^{k-1}} - p'_{mj}(F_j^k) \\
 \text{Exp}_j = 0 \\
 \text{ELSE} \\
 \lambda_j^k = \lambda_j^{k-1} \\
 \text{Exp}_j = 1 \\
 \text{END IF}
 \end{array} \right\} j = 1, 2, 3 \quad (3-44)$$

La variable  $\text{Exp}_j$  se utiliza para alertar que el modificador del gradiente no se ha actualizado. Cuando esto sucede, se incluye una restricción adicional al problema de optimización para forzar un cambio en la variable asociada al modificador  $\lambda_j$  (en este caso de estudio,  $F_j$ ). De esta manera, se asegura que en el ciclo siguiente se contará con información para la estimación del modificador.

c. Filtro exponencial y selección del valor máximo:

$$\left. \begin{array}{l}
 \lambda_j^k := f_j \cdot \lambda_j^k + (1 - f_j) \cdot \lambda_j^{k-1} \\
 \lambda_j^k := \max(\lambda_j^L, \min(\lambda_j^U, \lambda_j^k))
 \end{array} \right\} j = 1, 2, 3 \quad (3-45)$$

6. Minimización del costo: El problema  $P^k$  se formula como:

$$\left. \begin{array}{l}
 \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\
 \text{s.t.} \\
 F_1 - F_2 - F_3 = 0 \\
 y_j - F_j \cdot \eta_j = 0 \\
 \eta_j = p_{mj}(F_j) + \beta_j^k + \lambda_j^k \cdot (F_j - F_j^k) \\
 F_j^L \leq F_j \leq F_j^U \\
 |F_j - F_j^k| \leq m_j \\
 F_j \geq \min(F_j^k + \zeta, F_j^U) \quad \text{si } \text{Exp}_j = 1 \text{ y } \sigma_j \leq 0 \\
 F_j \leq \max(F_j^k - \zeta, F_j^L) \quad \text{si } \text{Exp}_j = 1 \text{ y } \sigma_j > 0
 \end{array} \right\} j = 1, 2, 3 \quad (3-46)$$

Se observa que la función objetivo contiene un término de convexificación (Brdys y Tatjewski, 2005), caracterizado por el parámetro  $\rho$ . Los cambios en las variables están limitados por los límites  $F_j^L$  y  $F_j^U$ , y por el máximo cambio permitido por ciclo  $m_j$ . Además, cuando el parámetro  $\text{Exp}_j$  toma el valor 1, se

fuerza un movimiento de magnitud  $\zeta$  en la variable  $F_j$ . El sentido del cambio está dado por el parámetro  $\sigma_j$ , calculado como:

$$\sigma_j = \text{signo} \left( \frac{dQ_p}{dF_j} \right)_{\mathbf{F}=\mathbf{F}^k}, \quad j=1,2,3 \quad (3-47)$$

donde la función costo  $Q_p$  se calcula según (3-35), y el gradiente se estima utilizando el modelo adaptado para el ciclo  $k$ .

Los valores seleccionados para los parámetros  $\delta$ ,  $\lambda^L$ ,  $\lambda^U$ ,  $\mathbf{f}$ ,  $\rho$ ,  $\mathbf{m}$  y  $\zeta$  se resumen en la **Tabla 3-1**.

#### 3.3.1.4. Adaptación por modificadores con ecuaciones de eficiencia – Alternativa 2 (MA-E2)

Los pasos característicos de esta estrategia son:

##### 5. Adaptación del modelo:

- a. Cálculo del error de la ecuación de eficiencia  $p_{mj}(F_j^k)$ ,  $\beta_j^k$ :

$$\beta_j^k = \eta_j^k - p_{mj}(F_j^k) \quad (3-48)$$

- b. Actualización de los términos de corrección:

$$\begin{aligned} (a_j^k, b_j^k, c_j^k) &= \arg \min \left( \sum_{i=k-N}^k w^i \cdot (\beta_j^i - a_j - b_j \cdot F_j^i - c_j \cdot (F_j^i)^2) \right) \\ \text{s.t. } \mathbf{a}^L &\leq \mathbf{a} \leq \mathbf{a}^U \\ \mathbf{b}^L &\leq \mathbf{b} \leq \mathbf{b}^U \\ \mathbf{c}^L &\leq \mathbf{c} \leq \mathbf{c}^U \end{aligned} \quad (3-49)$$

Los términos cuadráticos  $c_j^k$  sólo se utilizan a partir del ciclo  $k=7$ , cuando se considera que se cuenta con información suficiente para utilizar estos parámetros. Los términos lineales  $b_j^k$  se estiman desde el ciclo  $k=2$ , que es el primero en el que se resuelve el problema de optimización.

6. Minimización del costo: El problema  $P^k$  se formula como:

$$\begin{aligned}
 & \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\
 & \text{s.t.} \\
 & F_1 - F_2 - F_3 = 0 \\
 & y_j - F_j \cdot \eta_j = 0 \\
 & \eta_j = p_{mj}(F_j) + a_j^k + b_j^k \cdot F_j + c_j^k \cdot F_j^2 \\
 & F_j^L \leq F_j \leq F_j^U \\
 & |F_j - F_j^k| \leq m_j
 \end{aligned} \quad \left. \vphantom{\begin{aligned} & \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\ & \text{s.t.} \\ & F_1 - F_2 - F_3 = 0 \\ & y_j - F_j \cdot \eta_j = 0 \\ & \eta_j = p_{mj}(F_j) + a_j^k + b_j^k \cdot F_j + c_j^k \cdot F_j^2 \\ & F_j^L \leq F_j \leq F_j^U \\ & |F_j - F_j^k| \leq m_j \end{aligned}} \right\} j = 1, 2, 3 \quad (3-50)$$

Los valores seleccionados para los parámetros  $w^i$ ,  $N$ ,  $\mathbf{a}^L$ ,  $\mathbf{a}^U$ ,  $\mathbf{b}^L$ ,  $\mathbf{b}^U$ ,  $\mathbf{c}^L$ ,  $\mathbf{c}^U$ ,  $\rho$  y  $\mathbf{m}$  se resumen en la **Tabla 3-1** **Error! Reference source not found.**

### 3.3.1.5. Adaptación de las restricciones (CA)

5. Adaptación del modelo: se realiza a través de la ecuación (3-48). Al igual que en la sección 3.3.1.3, no se utiliza un filtro exponencial para corregir los modificadores de las restricciones (en 3.3.1.3, sí se filtran los modificadores del gradiente calculados).
6. Minimización del costo: sólo incluye los términos modificadores  $\beta_j^k$ . El problema  $P^k$  se formula como

$$\begin{aligned}
 & \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\
 & \text{s.t.} \\
 & F_1 - F_2 - F_3 = 0 \\
 & y_j - F_j \cdot \eta_j = 0 \\
 & \eta_j = p_{mj}(F_j) + \beta_j^k \\
 & F_j^L \leq F_j \leq F_j^U \\
 & |F_j - F_j^k| \leq m_j
 \end{aligned} \quad \left. \vphantom{\begin{aligned} & \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\ & \text{s.t.} \\ & F_1 - F_2 - F_3 = 0 \\ & y_j - F_j \cdot \eta_j = 0 \\ & \eta_j = p_{mj}(F_j) + \beta_j^k \\ & F_j^L \leq F_j \leq F_j^U \\ & |F_j - F_j^k| \leq m_j \end{aligned}} \right\} j = 1, 2, 3 \quad (3-51)$$

### 3.3.1.6. Adaptación por modificadores con actualización de Broyden (MA-Broyden)

5. Adaptación del modelo:
  - a. Cálculo del término modificador  $\beta_j^k$ : se estima para el error en las variables de salida  $y_j^k$ .

$$\beta_j^k = y_j^k - p_{mj}(F_j^k) \cdot F_j^k \quad (3-52)$$

- b. Estimación de los gradientes por el método de Broyden, con respecto a las entradas  $\mathbf{u} = [F_1, F_2]^T$ :

$$\nabla y_j^k = \nabla y_j^{k-1} + \frac{y_j^k - y_j^{k-1} - (\nabla y_j^{k-1})^T (\mathbf{u}^k - \mathbf{u}^{k-1})}{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|^2} \cdot (\mathbf{u}^k - \mathbf{u}^{k-1}) \quad (3-53)$$

- c. Actualización de los términos de corrección del gradiente:

$$\lambda_j^k = \nabla y_j^k - \nabla y_j^{\text{mod}}(\mathbf{u}^k) = \nabla y_j^k - \begin{pmatrix} \partial(p_{mj}(F_j) \cdot F_j) / \partial F_1 \\ \partial(p_{mj}(F_j) \cdot F_j) / \partial F_2 \end{pmatrix}_{\mathbf{u}=\mathbf{u}^k} \quad (3-54)$$

- d. Filtro exponencial: como simplificación, se usa la misma constante de filtrado  $k_j$  para los dos elementos del vector  $\lambda_j^k$ .

$$\lambda_j^k = \lambda_j^{k-1} + k_j \cdot (\lambda_j^k - \lambda_j^{k-1}) \quad (3-55)$$

- e. Generación de restricciones de RTO dual: para generar puntos útiles en la estimación del gradiente, que limiten el error por truncamiento y exploren el comportamiento del proceso en las dos direcciones del espacio de entradas, se generan las siguientes restricciones (Rodger y Chachuat, 2011):

$$\begin{aligned} & (F_1 - F_1^k)^2 + (F_2 - F_2^k)^2 \leq 1 \\ & \text{abs} \left( \mathbf{a}_k^T \cdot \begin{pmatrix} F_1 - F_1^k \\ F_2 - F_2^k \end{pmatrix} \right) \geq \sqrt{0.5 \cdot \mathbf{a}_k^T \cdot \mathbf{a}_k} \end{aligned} \quad (3-56)$$

donde  $\mathbf{a}_k^T = \left( (-F_2^k + F_2^{k-1}), (F_1^k - F_1^{k-1}) \right)^T$ , es decir, un vector normal al generado por las dos últimas entradas,  $\left( (F_1^k - F_1^{k-1}), (F_2^k - F_2^{k-1}) \right)^T$ .

6. Minimización del costo: el problema  $P^k$  se formula como:

$$\begin{aligned}
 & \min_{\mathbf{F}, \mathbf{y}, \boldsymbol{\eta}} y_1 - y_2 - y_3 + \rho \sum_{j=1}^3 (F_j - F_j^k)^2 \\
 & \text{s.t.} \\
 & F_1 - F_2 - F_3 = 0 \\
 & y_j = F_j \cdot \eta_j + \beta_j^k + (\boldsymbol{\lambda}_j^k)^T \cdot \begin{pmatrix} F_1 - F_1^k \\ F_2 - F_2^k \end{pmatrix} \\
 & \eta_j = p_{mj}(F_j) \\
 & F_j^L \leq F_j \leq F_j^U \\
 & |F_j - F_j^k| \leq m_j \\
 & (F_1 - F_1^k)^2 + (F_2 - F_2^k)^2 \leq 1 \\
 & \text{abs} \left( \boldsymbol{\alpha}_k^T \begin{pmatrix} F_1 - F_1^k \\ F_2 - F_2^k \end{pmatrix} \right) \geq \sqrt{0.5 \cdot \boldsymbol{\alpha}_k^T \cdot \boldsymbol{\alpha}_k}
 \end{aligned} \quad \left. \vphantom{\begin{aligned} \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \end{aligned}} \right\} j = 1, 2, 3 \quad (3-57)$$

Como se ha mencionado en la sección 2.4.5, las restricciones de RTO dual generan dos regiones factibles desconectadas. Por este motivo, el problema (3-57) se resuelve dos veces: en el primer caso, se asigna un valor positivo al argumento de la función valor absoluto de (3-56); en el segundo caso, se asigna un valor negativo a este argumento. De las dos soluciones, se selecciona aquella solución factible con menor valor de la función objetivo.

La **Tabla 3-1** presenta los valores seleccionados para los parámetros  $\mathbf{k}$ ,  $\rho$  y  $\mathbf{m}$ , junto con los valores adoptados para los parámetros en las restantes estrategias de adaptación.

### 3.3.1.7. Resultados y discusión

El modelo y las cuatro estrategias de adaptación presentadas en las secciones anteriores fueron implementadas en GAMS 24.1 (GAMS Development Corporation, 1997). Los problemas de optimización fueron resueltos con el resolutor NLP CONOPT 3 (Drud, 2012).

El punto inicial elegido para el estudio es  $[F_1^0 = 90, F_2^0 = 60]$ . El valor de la función objetivo (real) en este punto es  $Q_p^0 = 5907.75$ . El punto óptimo real se obtiene minimizando el costo con el modelo de la planta real; la solución calculada es  $[F_1 = 105.59, F_2 = 71.09]$ , que corresponde a un costo  $Q_p = 5520.81$ .

**Tabla 3-1.** Parámetros utilizados en las alternativas de RTO del Caso de Estudio 1.

Parámetro	Valor	Parámetro	Valor
$\delta$	0.3	$N$	$\min(k, 10)$
$\lambda^L$	$[-0.9, -0.9, -0.9]^T$	$\mathbf{a}^L$	$[-500, -500, -500]^T$
$\lambda^U$	$[0.9, 0.9, 0.9]^T$	$\mathbf{a}^U$	$[500, 500, 500]^T$
$\mathbf{f}$	$[0.7, 0.7, 0.7]^T$	$\mathbf{b}^L$	$[-7, -7, -7]^T$
$\rho$	5	$\mathbf{b}^U$	$[7, 7, 7]^T$
$\mathbf{m}$	$[3, 3, 3]^T$	$\mathbf{c}^L$	$[-0.1, -0.1, -0.1]^T$
$\zeta$	1	$\mathbf{c}^U$	$[0.1, 0.1, 0.1]^T$
$w^i$	$0.9^{k-i}$	$\mathbf{k}$	$[0.6, 0.6, 0.6]^T$

A partir del punto inicial, se estudió el comportamiento del sistema al ejecutar 40 ciclos sucesivos de RTO. Se analizaron dos escenarios: en el primero, las mediciones se obtienen sin error; en el segundo, tanto las salidas  $y_j$  como las entradas  $F_1$  y  $F_2$  son modificadas con la adición de ruido, de distribución normal, con media 0 y desviación estándar 0.1.

La **Figura 3.3**, la **Figura 3.4** y la **Figura 3.5** muestran la evolución para las estrategias MA-E1, MA-E2 y MA-Broyden, respectivamente. Todos los casos se comparan con la estrategia CA sin ruido.

Se observa que la estrategia CA lleva la planta a un punto operativo diferente del óptimo real. Este comportamiento es esperable, ya que al no incluir un término de corrección del gradiente, la estrategia no corrige el error estructural presente en el modelo.

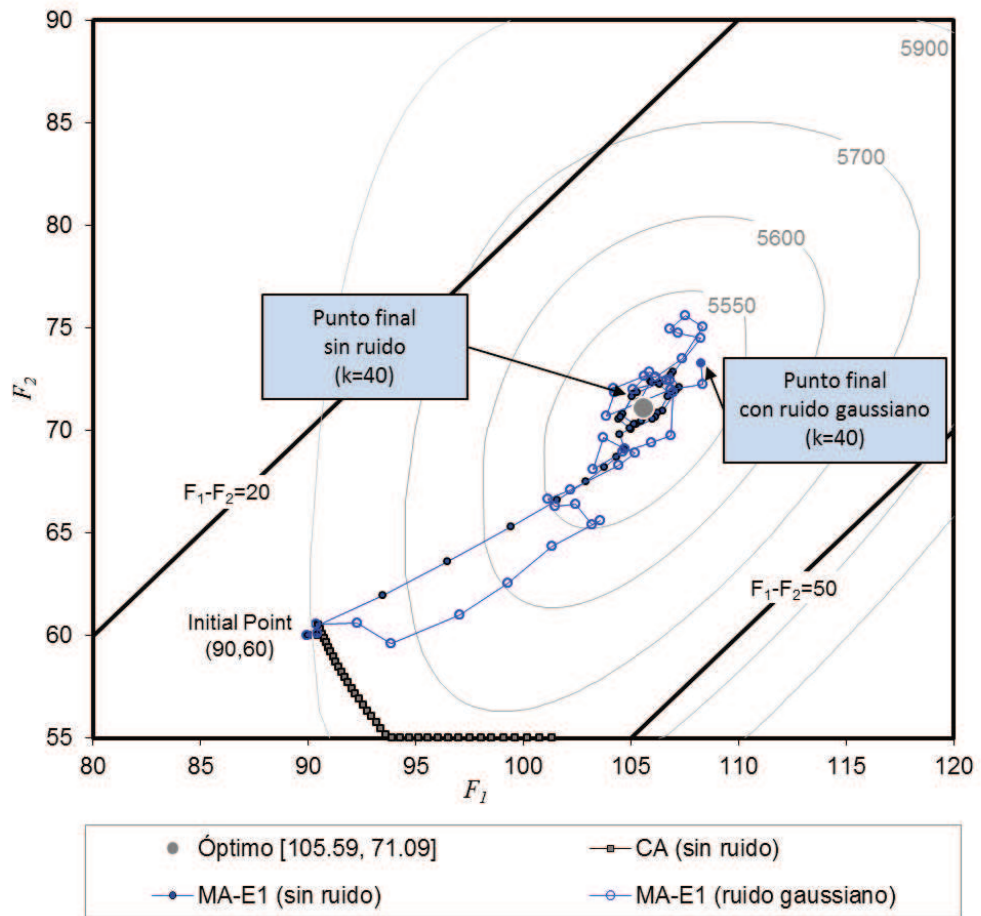


Figura 3.3. Caso de estudio 1. Evolución del sistema – Estrategia MA-E1.

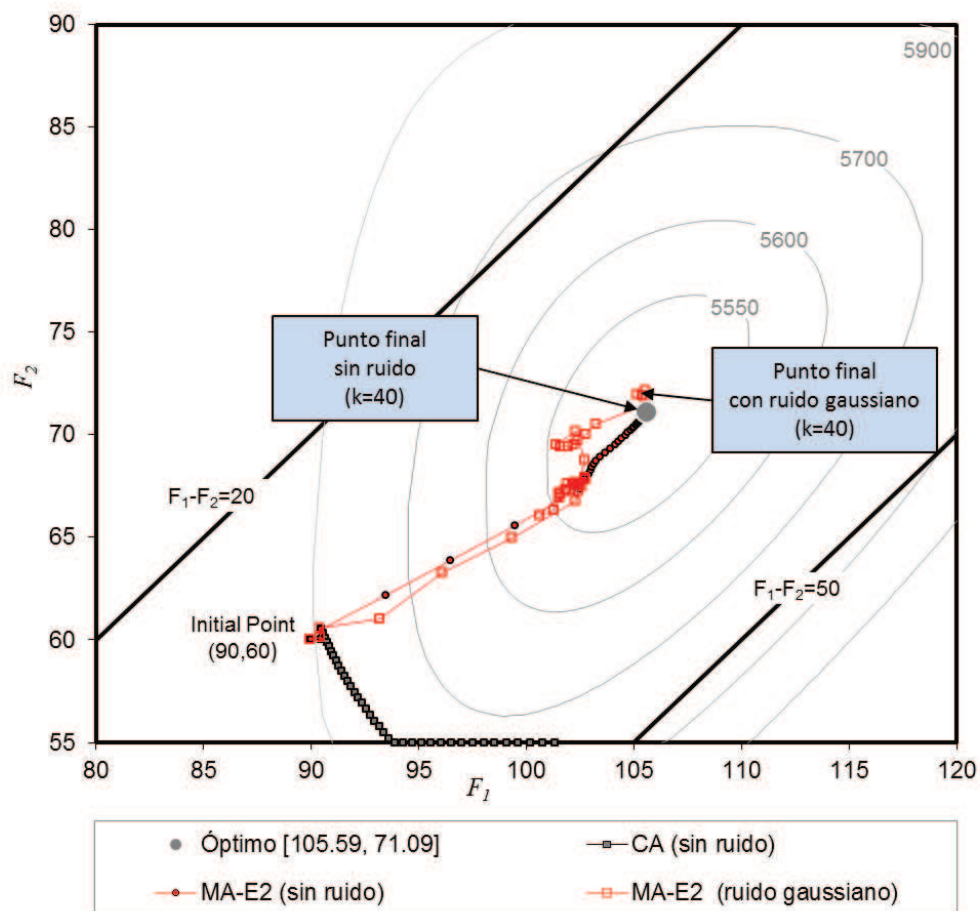


Figura 3.4. Caso de estudio 1. Evolución del sistema – Estrategia MA-E2.

Para los casos sin ruido, MA-E1 y MA-E2 convergen al entorno del óptimo real de la planta, mientras que la estrategia MA-Broyden no puede acercarse suficientemente al costo mínimo posible en el número de ciclos analizados. Las estrategias propuestas en esta tesis también muestran un desempeño superior al de MA-Broyden cuando las mediciones incluyen ruido gaussiano. Como puede observarse en las figuras, MA-E2 presenta una convergencia más lenta hacia el óptimo que MA-E1, pero la variabilidad en la operación es mucho menor una vez que la convergencia se ha alcanzado. Esta es una propiedad deseada en un sistema de RTO, ya que los cambios en las proximidades del óptimo generan beneficios muy pequeños (o incluso un aumento en el costo, debido al error en el modelo).

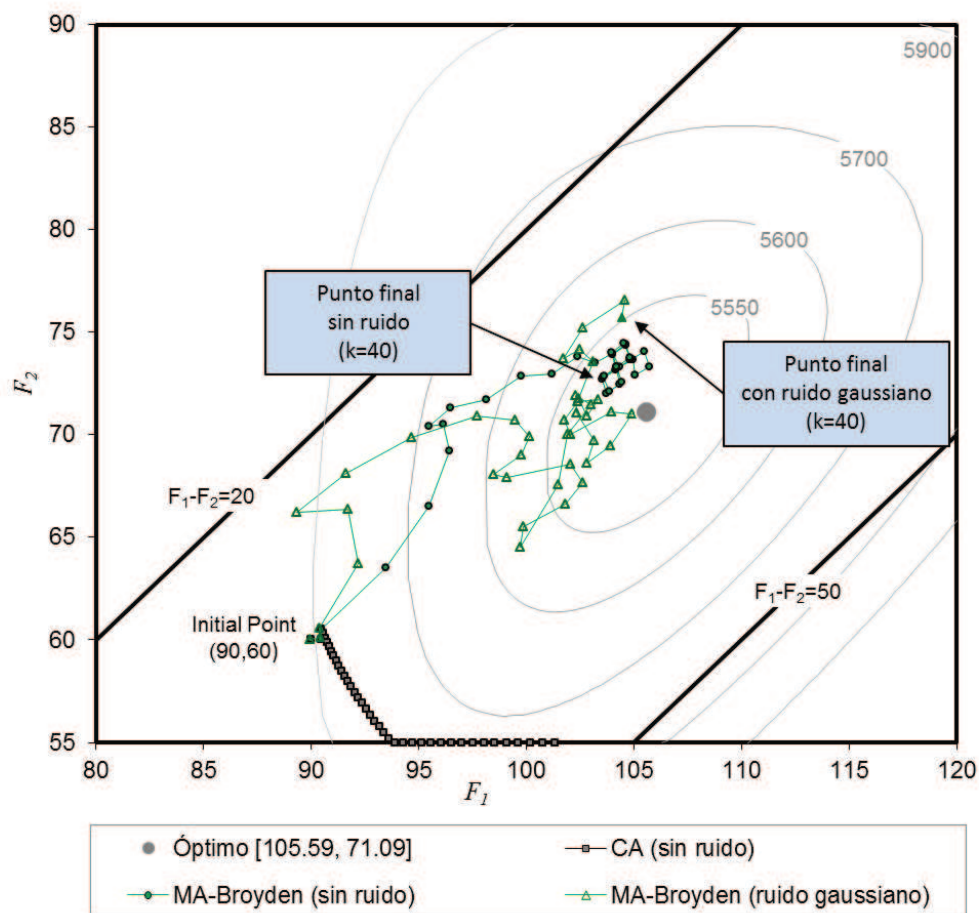
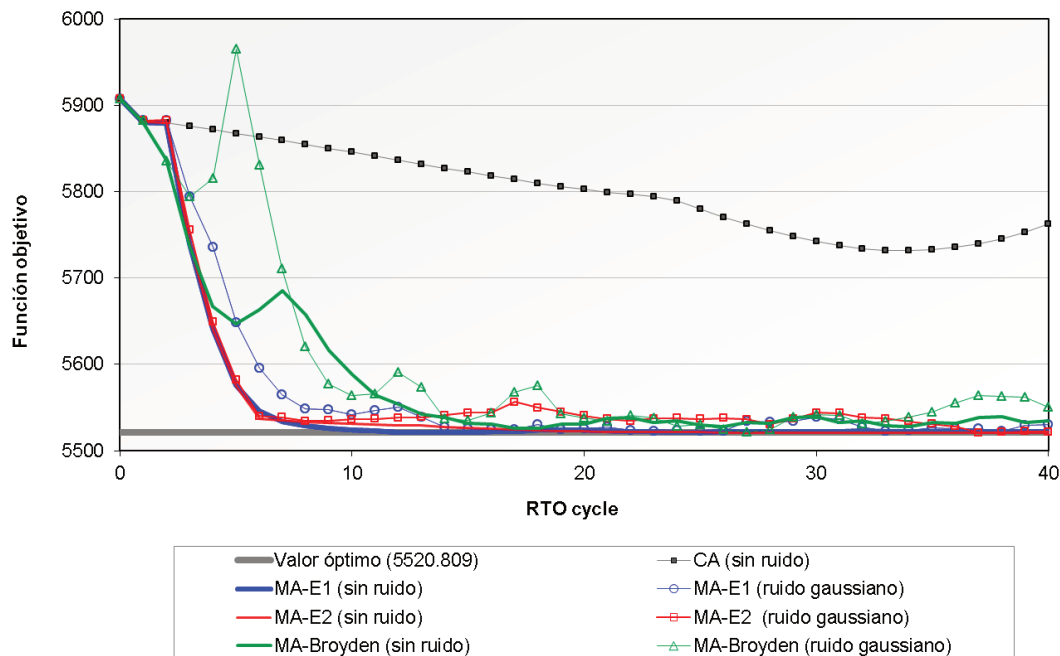


Figura 3.5. Caso de estudio 1. Evolución del sistema – Estrategia MA-Broyden.

La corrección del gradiente hace que la estrategia MA-Broyden tenga un desempeño mucho mejor que la adaptación de restricciones (CA). Sin embargo, no consigue converger al óptimo real en los 40 ciclos estudiados. El motivo principal de estos problemas de convergencia es que los gradientes se estiman en forma numérica en todas las direcciones del espacio, incluidas aquellas en las que la derivada direccional es nula. Por ejemplo, la salida  $y_3$  es función de la diferencia entre  $F_1$  y  $F_2$  (ver (3-30)-(3-32) y (3-36)), por lo que la derivada direccional de  $y_3$  en la dirección perpendicular (dada por el vector  $(1,1)^T$ ) es nula, ya que  $(F_1 - F_2)$  es constante en esta dirección. El método de adaptación MA-Broyden no tiene en cuenta esto, calculando cada derivada parcial independientemente. Esto genera errores adicionales con respecto a las estrategias MA-E1 y MA-E2, que causan las dificultades en la convergencia y el mayor costo operativo.

La Figura 3.6 presenta la evolución de la función objetivo (real) para todas las estrategias de adaptación analizadas. Puede observarse que en este caso de estudio, las alternativas propuestas (MA-E1 y MA-E2) presentan un mejor

desempeño que las estrategias implementadas como comparación (CA y MA-Broyden).



**Figura 3.6.** Caso de estudio 1. Evolución de la función objetivo para todas las estrategias analizadas.

Los resultados observados en la **Figura 3.6** pueden utilizarse para evaluar el Costo de Diseño Extendido ( $C_E$ ). La fórmula (2-76) puede aproximarse como:

$$C_E = \int_{t_0}^{t_f} E[Q_p(\mathbf{u}(t)) - Q_p(\mathbf{u}^{*,p})] dt \approx \Delta t \cdot \left( \frac{1}{2} (Q_p^{k_0} - Q_p^*) + \sum_{k=k_0+1}^{k_f-1} (Q_p^k - Q_p^*) + \frac{1}{2} (Q_p^{k_f} - Q_p^*) \right) \quad (3-58)$$

donde  $Q_p^* = Q_p(\mathbf{u}^{*,p})$ , y los tiempos  $t_0$  y  $t_f$  se reemplazan por los ciclos  $k_0$  y  $k_f$ . Los costos  $Q_p^k$  se calculan con el modelo de la planta real. Se asume que el tiempo entre ciclos  $\Delta t$  se mantiene constante, y a los fines del cálculo, igual a 1. De todos modos, otros valores de  $\Delta t$  modificarían proporcionalmente todos los valores de  $C_E$ , conduciendo a las mismas conclusiones a la hora de comparar el desempeño de las estrategias de adaptación.

Los resultados del cálculo del  $C_E$  se resumen en la **Tabla 3-2**. El  $C_E$  se evalúa para el total de los ciclos del caso ( $k_0 = 0, k_f = 40$ ) y también para el

período ( $k_0 = 5, k_f = 40$ ). El primer valor permite evaluar la pérdida total de beneficio de la estrategia de RTO, mientras que el segundo permite determinar cómo se comporta el sistema cuando la planta está en cercanías del óptimo en las estrategias con corrección del gradiente (la estrategia adaptación de restricciones lleva la planta a un punto diferente). Para seleccionar  $k_0 = 5$  como inicio de la segunda evaluación, el criterio utilizado es que las alternativas MA-E1 y MA-E2, sin ruido, alcanzan un costo que difiere en menos del 1% del costo óptimo real de la planta. De acuerdo con este criterio, la evolución del sistema puede dividirse en dos etapas: en la primera, el costo real alcanzado difiere significativamente del costo óptimo; en la segunda, el costo alcanzado es cercano al valor deseado de la función objetivo, y las variaciones del costo entre ciclo y ciclo son pequeñas. En esta segunda etapa, la diferencia entre el costo alcanzado y el costo ideal se deben a errores en la estimación del gradiente y a la variabilidad en los datos.

La **Tabla 3-2** también presenta un  $C_E$  relativo, calculado como:

$$C_E^{rel} = \frac{C_E}{C_E^{NA}} \cdot 100 \quad (3-59)$$

donde  $C_E^{NA}$  es el Costo de Diseño Extendido obtenido si no se implementa ninguna estrategia de RTO en el sistema (es decir, las entradas se mantienen en su valor original  $(F_1^0, F_2^0)$ ).

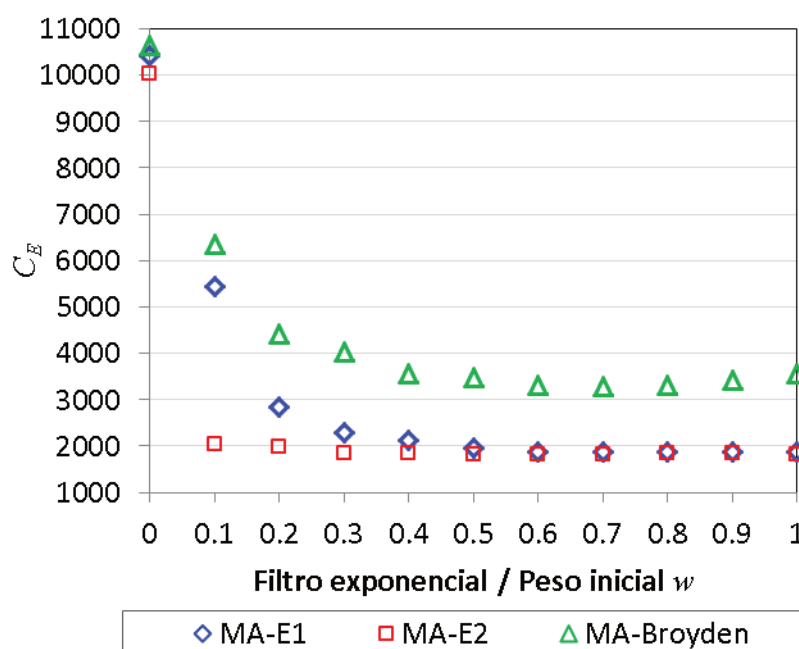
El Costo de Diseño Extendido muestra claramente que las alternativas MA-E1 y MA-E2 tienen un mejor desempeño que la implementación de adaptación por modificadores MA-Broyden, y que la adaptación de las restricciones (CA). La diferencia es más notoria para  $k_0 > 5$ , es decir, cuando las estrategias propuestas convergen a las cercanías del óptimo. Como se mencionó antes, la estrategia CA es la que muestra el  $C_E$  más elevado, debido a que no corrige las diferencias estructurales entre la planta y el modelo. Nuevamente, se observa que la estrategia MA-Broyden aporta beneficios significativos por corregir los gradientes, si bien no logra alcanzar el desempeño de las estrategias propuestas en este capítulo.

La selección del parámetro de filtrado exponencial ( $\mathbf{f}$  en MA-E1,  $\mathbf{k}$  en MA-Broyden) puede ser crítica para el desempeño de la estrategia de adaptación. Lo mismo puede decirse de los pesos  $w^i$  en MA-E2. La **Figura 3.7** muestra la variación del  $C_E$  con respecto al filtrado exponencial de los modificadores del gradiente, y con respecto al valor inicial de filtrado de la estrategia MA-E2 (0.9 en el caso de estudio original, de acuerdo con la **Tabla 3-1**. Parámetros utilizados en las alternativas de RTO del Caso de Estudio 1.). El nivel de ruido es el utilizado en

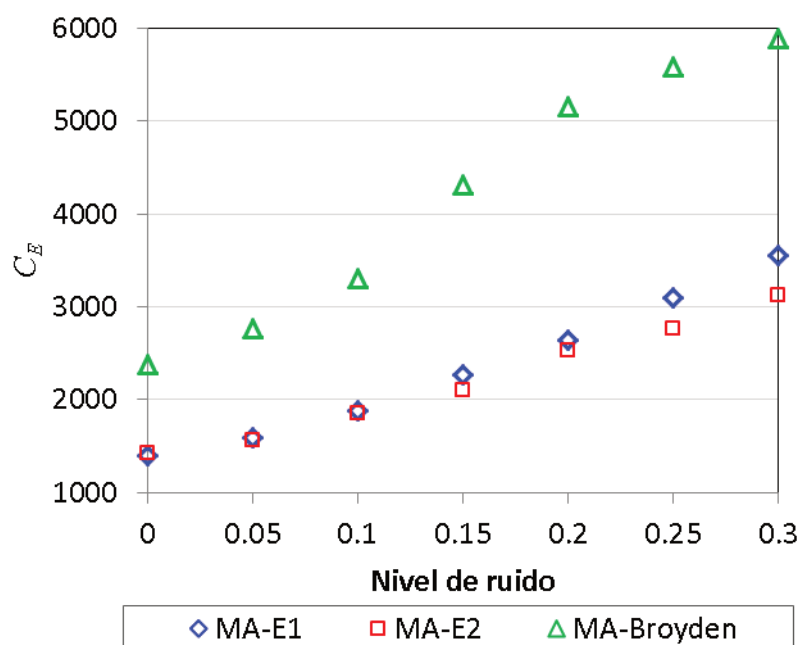
el caso de estudio (distribución normal de media 0 y desviación estándar 0.1). Cada punto de la **Figura 3.7** corresponde al valor promedio de 50 ejecuciones del caso de estudio completo (40 ciclos de RTO en cada una, con el mismo punto inicial), con generación aleatoria del ruido. Puede observarse que las alternativas MA-E1 y MA-E2 no muestran una sensibilidad significativa a los parámetros para valores superiores a 0.6, Para MA-Broyden, los resultados sugieren una constante de filtrado óptima entre 0.6 y 0.8. De todos modos, las estrategias propuestas alcanzan un  $C_E$  menor que MA-Broyden para todo el rango de valores de los parámetros.

**Tabla 3-2.** Costo de Diseño Extendido ( $C_E$ ) para las estrategias de adaptación implementadas en el caso de estudio 1.

Escenarios	k=0 a 40		k=5 a 40	
	$C_E$	$C_E^{rel}$	$C_E$	$C_E^{rel}$
Sin RTO	15477.6	100	13542.9	100
<b>Sin ruido</b>				
CA	11244.4	72.65	9453.1	69.80
MA-Broyden	2384.7	15.41	1090.1	8.05
MA-E1	1402.4	9.06	123.7	0.91
MA-E2	1426.3	9.22	150	1.11
<b>Ruido gaussiano</b>				
MA-Broyden	3320.6	21.45	1661.5	12.27
MA-E1	1969.1	12.72	499.9	3.69
MA-E2	1898.8	12.27	587.3	4.34



**Figura 3.7.** Influencia del filtro exponencial (MA-E1, Broyden-MA) y del peso inicial (MA-E2) en el Costo de Diseño Extendido. Nivel de ruido: desviación estándar 0.1.



**Figura 3.8.** Influencia del nivel de ruido (desviación estándar del ruido gaussiano) en el Costo de Diseño Extendido. Parámetros de filtrado según **Tabla 3-1**.

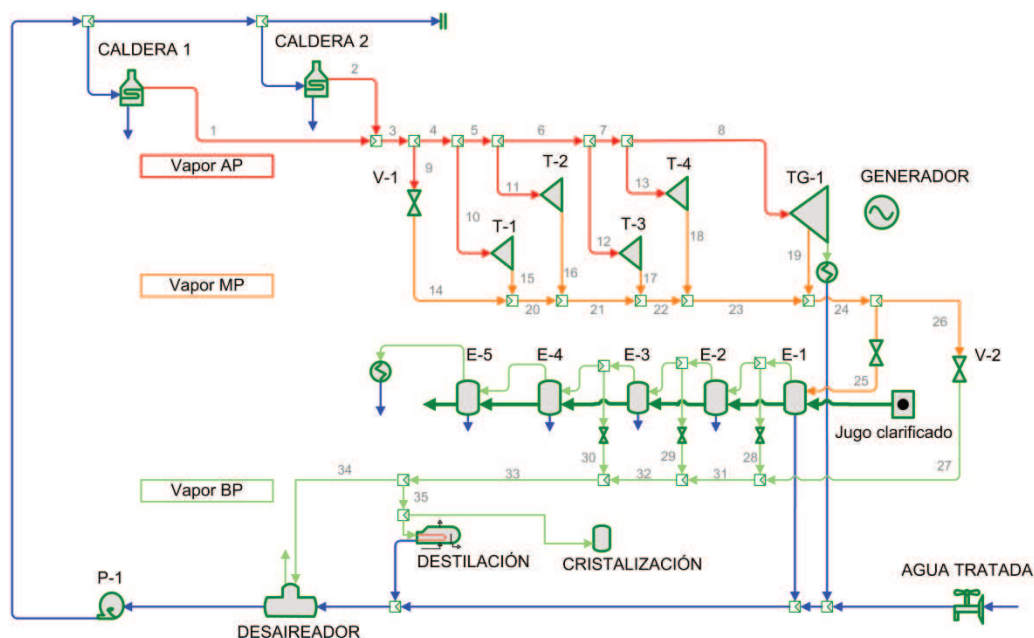
El nivel de ruido también puede afectar el desempeño de la optimización en tiempo real. La **Figura 3.8** muestra la variación del Costo de Diseño Extendido para cada estrategia cuando cambia el nivel de ruido, es decir, la desviación estándar del ruido normal adicionado a las mediciones. Nuevamente, cada punto es el promedio de 50 ejecuciones del caso de estudio con ruido normal aleatorio;

los parámetros de filtrado y de peso son los listados en la **Tabla 3-1** *Error! Reference source not found.* Como es de esperar, cuando aumenta el nivel de ruido, todas las implementaciones muestran un  $C_E$  más elevado. Para niveles de ruido más altos, MA-E2 tiene el mejor desempeño; cuando el ruido es bajo, MA-E1 y MA-E2 tienen un  $C_E$  promedio similar. MA-Broyden muestra un costo más elevado en todos los casos.

### 3.3.2. Caso de estudio 2: Sistema de calor y potencia de una planta de azúcar y etanol

#### 3.3.2.1. Descripción del caso

El caso de estudio 2 estudia el sistema de vapor y energía eléctrica de una planta de azúcar y etanol. El sistema incluye dos calderas alimentadas con bagazo de caña, cuatro turbinas de contrapresión para la molienda de caña, una turbina de extracción-condensación para generación de energía eléctrica, un sistema de evaporación de cinco efectos, y otros consumidores de vapor para calentamiento y destilación. La **Figura 3.9** muestra un diagrama del sistema estudiado. La **Tabla 3-3** presenta la nomenclatura utilizada en este caso de estudio.



**Figura 3.9.** Caso de estudio 2. Sistema de calor y potencia de la planta de azúcar y etanol.

**Tabla 3-3.** Nomenclatura utilizada en el caso de estudio 2.

Símbolo	Descripción	Símbolo	Descripción
$F$	Caudal (kg/h)	<b>Subíndices</b>	
$P$	Presión (Pa)	$cb$	Caldera
$T$	Temperatura (K)	$tb$	Turbina de contrapresión
$X^{med}$	Fracción de sólidos másica	$tg$	Etapa de la turbina de extracción-condensación
$Cp$	Capacidad calorífica (kJ/kg.K)	$e$	Efecto del evaporador
$d$	Demanda de vapor de la planta (t/h)	<b>Superíndices</b>	
$W$	Potencia de la turbina (kW)	$i$	Corriente de entrada
$Q$	Combustible quemado en la caldera (kJ/h)	$o$	Corriente de salida
$c$	Coefficiente de evaporación (kg/m <sup>2</sup> .°C)	$sat$	Saturado (agua o vapor)
$V$	Vapor del evaporador utilizado en el efecto siguiente (t/h)		
$VB$	Extracción de vapor en un efecto (para uso en otros procesos) (t/h)		

El modelo detallado del sistema estudiado se incluye en el Apéndice A de esta tesis. La **Tabla 3-4** muestra los valores operativos que están fijos en este caso de estudio. La **Tabla 3-5** lista las condiciones operativas en el punto inicial para todos los escenarios analizados.

**Tabla 3-4.** Caso de estudio 2. Parámetros operativos fijos.

Parámetro	Valor	Parámetro	Valor
Presión de vapor alta (bar)	80	Caudal de entrada jugo clarificado (t/h)	500
Presión de vapor media (bar)	10	Brix de entrada (°Bx)	15
Presión de vapor baja (bar)	0.8	Brix de salida (°Bx)	65
Presión condensador TG (bar)	0.35	Potencia turbinas, $tb = 1.4$ (kW)	500
Temperatura vapor calderas (°K)	650	Eficiencia turbinas, $tb = 1.4$ (%)	50
Área de intercambio evaporadores, $e = 1...5$ (m <sup>2</sup> )	2000		

**Tabla 3-5.** Caso de estudio 2. Condiciones operativas iniciales.

Variable	Valor	Variable	Valor
$F_{cb1}^o$ (t/h)	70	$VB_{e2}$ (t/h)	12.4
$F_{cb2}^o$ (t/h)	158.9	$VB_{e3}$ (t/h)	15.8
$F_{tg1}^i$ (t/h)	100	Demanda de vapor $d^0$ (t/h)	200
$F_{tg1}^o$ (t/h)	70	Precio de venta electricidad $c_w^0$ (\$/kWh)	0.018
$VB_{e1}$ (t/h)	139.0		

Si las presiones en los colectores de vapor y la concentración final del jugo están fijas, el sistema tiene 6 grados de libertad. Las 6 entradas seleccionadas para controlar la planta son el caudal de extracción de la turbina TG-1 ( $F_{tg1}^o$ ); el caudal de condensado de la misma turbina ( $F_{tg2}^i$ ); los caudales de vapor extraído para el proceso de las efectos 1, 2 y 3 del evaporador ( $VB_{e1}$ ,  $VB_{e2}$  y  $VB_{e3}$ ); y la diferencia en producción de vapor entre las calderas 1 y 2 ( $F_{cb1}^o - F_{cb2}^o$ ). Si se implementara una estrategia de adaptación por modificadores (o ISOPE) con RTO dual (sección 2.4.5 de esta tesis), se necesitarían al menos 6 conjuntos de datos para fijar las

restricciones adicionales que evitan el mal condicionamiento del problema de estimación del gradiente o limitan el error en la estimación. Este número puede ser suficientemente alto como para desalentar una implementación dual, ya que las restricciones adicionales pueden limitar en exceso la región de operación factible en cada ciclo de RTO, y podrían presentarse perturbaciones no medidas con una frecuencia mayor a 6 ciclos de RTO, lo que invalidaría los conjuntos de datos utilizados antes de la perturbación (Marchetti et al., 2010). Por este motivo, en este caso se implementa la estrategia de adaptación basada en ecuaciones de eficiencia.

Los valores de entalpía y entropía de las corrientes de agua y vapor, así como las temperaturas y presiones de saturación, se estiman según las correlaciones propuestas por Irvine y Liley (1984). La elevación del punto de ebullición (BPE) y la capacidad calorífica ( $C_p$ ) para el jugo de caña se calculan de acuerdo con Higa (2003). Todas estas expresiones, así como los balances de masa, energía y entropía, se consideran como libres de error para el caso analizado. Los errores en el modelo se concentran en la predicción de las eficiencias de calderas, las eficiencias isoentrópicas en las etapas del turbogenerador y los coeficientes de evaporación de cada efecto del evaporador.

Los factores de eficiencia de las calderas se calculan como:

*Planta real:*

$$\eta_1 = \eta_{cb1} = 89.975 - 0.001 \cdot (140 - F_{cb1}^o)^2 - 10^{-6} \cdot (F_{cb1}^o)^3 \quad (3-60)$$

$$\eta_2 = \eta_{cb2} = 91 - 0.001 \cdot (155 - F_{cb2}^o)^2 - 1.2 \times 10^{-6} \cdot (F_{cb2}^o)^3 \quad (3-61)$$

*Modelo RTO:*

$$\eta_1 = \eta_{cb1} = 92 - 0.005 \cdot (150 - F_{cb1}^o)^2 \quad (3-62)$$

$$\eta_2 = \eta_{cb2} = 93 - 0.005(150 - F_{cb2}^o)^2 \quad (3-63)$$

Los factores de eficiencia de las turbinas se calculan como:

*Planta real:*

$$\eta_3 = \eta_{tg1} = 53 - 0.0023 \cdot (120 - F_{tg1}^o)^2 + 1 \times 10^{-6} \cdot (F_{tg1}^o)^3 \quad (3-64)$$

$$\eta_4 = \eta_{tg2} = 66 - 0.0018 \cdot (50 - F_{tg2}^o)^2 + 2.2 \times 10^{-6} \cdot (F_{tg2}^o)^3 + 10^{-2} \cdot T_{tg1}^o \quad (3-65)$$

Modelo RTO:

$$\eta_3 = \eta_{tg1} = 70 - 0.005 \cdot (130 - F_{tg1}^o)^2 \quad (3-66)$$

$$\eta_4 = \eta_{tg2} = 70 - 0.004 \cdot (41 - F_{tg2}^o)^2 \quad (3-67)$$

Los coeficientes de evaporación se calculan como:

Planta real:

$$\eta_{4+e} = c_e = (0.0012 - 0.0001 \cdot e) \cdot (96 - X_e^{med}) \cdot (T_{e-1}^{sat} - 54) - 0.2 \cdot \left( \frac{X_e^{med}}{100} / 100 \right)^2 - 0.5 \cdot \exp\left( - \frac{(T_{e-1}^{sat} - 50)}{100} \right), \quad e = 1, \dots, 5 \quad (3-68)$$

Modelo RTO:

$$\eta_{4+e} = c_e = 0.001 \cdot (100 - X_e^{med}) \cdot (T_{e-1}^{sat} - 54), \quad e = 1, \dots, 5 \quad (3-69)$$

Los coeficientes de evaporación son equivalentes a coeficientes de transferencia de calor: predicen la cantidad de agua evaporada por unidad de área de transferencia de calor y por grado de diferencia de temperatura entre el vapor de calefacción y el jugo en ebullición. Las ecuaciones utilizadas para el modelo de RTO sin adaptación son del tipo de la ecuación de Dessin (Hugot, 1986). La expresión para calcular los coeficientes de evaporación de la planta real es intencionalmente diferente de la ecuación de Dessin, como lo muestran (3-68) y (3-69). Mientras cada coeficiente se expresa en forma explícita como una función de las temperaturas y las concentraciones en grados Brix del jugo, la corrección del gradiente se realiza con respecto al caudal de vapor *vivo* utilizado para calentar el primer efecto ( $V_0$ ). Se asume que la extracción de vapor en los efectos será generalmente maximizada, y por lo tanto, las temperaturas y los °Brix en la configuración óptima son función del vapor vivo de entrada.

En (3-65), la eficiencia real de la segunda etapa del turbogenerador es una función de dos variables: el caudal de vapor en la etapa y la temperatura de entrada (que depende de la eficiencia de la primera etapa y la temperatura y presión del vapor de admisión a la entrada de la turbina). Sin embargo, el modelo de RTO y la estrategia de corrección del gradiente consideran despreciable esta funcionalidad con la temperatura, por lo que el gradiente se corrige sólo con respecto al caudal de la etapa. Como se verá al analizar los resultados, el desempeño del sistema de RTO es correcto a pesar de esta simplificación.

De acuerdo con (3-60)-(3-69) y con lo expresado en párrafos anteriores, las variables auxiliares  $v_j$  de la ecuación (3-2) son variables unidimensionales:

$$\begin{aligned}
 v_1 &= F_{cb1}^o \\
 v_2 &= F_{cb1}^o \\
 v_3 &= F_{tg1}^o \\
 v_4 &= F_{tg2}^o \\
 v_i &= V_0, \quad i = 5, \dots, 9
 \end{aligned}
 \tag{3-70}$$

La función objetivo a minimizar tiene en cuenta un costo asignado al bagazo quemado, y un ingreso debido a la venta de electricidad, y su expresión es la siguiente:

$$\min Q_1 + Q_2 - c_W^k \cdot W_{TG}
 \tag{3-71}$$

Se estudiaron cuatro escenarios. El escenario 1, utilizado para chequear la convergencia de la propuesta al óptimo real de la planta, considera una demanda constante de vapor para destilación y otros usos. En el escenario 2, la demanda de vapor puede variar entre dos ciclos de RTO. En el período entre dos ciclos, el cambio en la demanda se resuelve ajustando la producción de vapor en las calderas 1 y 2, manteniendo constante la diferencia óptima entre los dos caudales de vapor. En el escenario 3, el precio de venta de la electricidad puede cambiar cada cuatro ciclos. El escenario 4 combina los cambios en la demanda y el precio de venta de electricidad de los escenarios 2 y 3.

Las variables *medidas* y la variancia  $\sigma^2$  de cada medición se resumen en la **Tabla 3-6**. Los subíndices en los caudales y las temperaturas indican el número de corriente, de acuerdo con la **Figura 3.9**. Los subíndices en las temperaturas de saturación indican el número de efecto. Se incluyen como variables medidas las eficiencias de las calderas. En la práctica, como es difícil medir con precisión los caudales de bagazo (Golato et al., 2008), la eficiencia se calcula a través de un método indirecto, que tiene en cuenta las temperaturas del aire y los humos, el contenido de CO<sub>2</sub> y O<sub>2</sub> en los humos, y una estimación de las pérdidas por radiación y por bagazo no quemado. En este caso de estudio, se asume que la variancia indicada en la tabla está relacionada al impacto conjunto de las mediciones inciertas de temperaturas, composiciones y poderes caloríficos para el cálculo indirecto de la eficiencia.

**Tabla 3-6.** Caso de estudio 2. Variables medidas y variancia de las mediciones.

Variable	Puntos de medición	$\sigma^2$
Caudal de vapor	$F_1, F_2, F_8, F_9, F_{19}, F_{25}, F_{26}, F_{28}, F_{29}, F_{30}, F_{35}$	0.09 (t/h) <sup>2</sup>
Eficiencia	$\eta_{cb1}, \eta_{cb2}$	0.36 (%) <sup>2</sup>
Temperatura	$T_1, T_2, T_8$	0.25 (°C) <sup>2</sup>
Potencia generada en el turbogenerador	$W_{TG}$	625 (kW) <sup>2</sup>
Temperatura del vapor en cada efecto	$T_1^{sat}, T_2^{sat}, T_3^{sat}, T_4^{sat}, T_5^{sat}$	0.25 (°C) <sup>2</sup>

### 3.3.2.2. Implementación de la estrategia de RTO

Como existe redundancia en las mediciones, los valores más probables de las variables se obtienen mediante reconciliación de datos, resuelta en simultáneo con la obtención de los factores de eficiencia  $\eta_1$  a  $\eta_9$ . El problema de optimización resuelto en este paso es una combinación de (3-11) y (3-12):

$$\begin{aligned}
 (\mathbf{y}^k, \mathbf{x}^k, \mathbf{u}^k, \boldsymbol{\eta}^k) = \arg \min_{\mathbf{y}, \mathbf{x}, \mathbf{u}, \boldsymbol{\eta}} & (\mathbf{y}^{m,k} - \mathbf{y})^T \cdot A^{-1} \cdot (\mathbf{y}^{m,k} - \mathbf{y}) + (\mathbf{u}^{m,k} - \mathbf{u})^T \cdot B^{-1} \cdot (\mathbf{u}^{m,k} - \mathbf{u}) \\
 & + (\boldsymbol{\eta}^{m,k} - \boldsymbol{\eta})^T \cdot D^{-1} \cdot (\boldsymbol{\eta}^{m,k} - \boldsymbol{\eta}) \quad (3-72) \\
 \text{s.t. } & \mathbf{f}_1(\mathbf{x}, \mathbf{u}) = 0 \\
 & \mathbf{f}_2(\mathbf{x}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\
 & \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u})
 \end{aligned}$$

donde la matriz  $D$  es la matriz de variancia-covariancia de las eficiencias medidas. Los conjuntos  $(\mathbf{y}^{m,k}, \mathbf{u}^{m,k}, \boldsymbol{\eta}^{m,k})$  y  $(\mathbf{y}^k, \mathbf{u}^k, \boldsymbol{\eta}^k)$  son los valores medidos y reconciliados, respectivamente, de las variables indicadas en la **Tabla 3-6**, y las restricciones de igualdad del problema son el modelo completo del sistema de energía, tal cual se lo describe en el Apéndice A. En este caso, las matrices  $A$ ,  $B$  y  $D$  son matrices diagonales (no hay correlación entre las mediciones), y los elementos de la diagonal son las variancias indicadas en la **Tabla 3-6**.

El paso siguiente es la adaptación del modelo. Para este caso de estudio, entre las estrategias propuestas en esta tesis, se ha seleccionado la estrategia MA-E2, ya que de acuerdo a los resultados del caso de estudio 1, esta estrategia mostró comparativamente una menor variabilidad en la cercanía del óptimo de la planta con respecto a MA-E1. Además, la inclusión del modificador cuadrático de los

datos en MA-E2 proporciona una corrección que es válida en una región más amplia que el modificador del gradiente de MA-E1, lo cual es útil para los escenarios 2, 3 y 4, en los que las condiciones óptimas de operación cambian frecuentemente.

El procedimiento de adaptación en este caso se formula como:

$$\begin{aligned}
 \beta_j^k &= \eta_j^k - p_{mj}(v_j^k) \\
 (a_j^k, b_j^k, c_j^k) &= \arg \min \left( \sum_{l=k-15}^k 0.9^l \cdot (\beta_j^l - \gamma_j^l)^2 \right) \\
 \text{s.t. } \gamma_1^l &= a_1^k + b_1^k \cdot F_{cb1}^{o,l} + c_1^k \cdot (F_{cb1}^{o,l})^2 \\
 \gamma_2^l &= a_2^k + b_2^k \cdot F_{cb2}^{o,l} + c_2^k \cdot (F_{cb2}^{o,l})^2 \\
 \gamma_3^l &= a_3^k + b_3^k \cdot F_{tg1}^{o,l} + c_3^k \cdot (F_{tg1}^{o,l})^2 \\
 \gamma_4^l &= a_4^k + b_4^k \cdot F_{tg2}^{o,l} + c_4^k \cdot (F_{tg2}^{o,l})^2 \\
 \gamma_j^l &= a_j^k + b_j^k \cdot V_0^l \quad j = 5 \dots 9 \\
 a_j^L &\leq a_j^k \leq a_j^U \quad j = 1 \dots nj \\
 b_j^L &\leq b_j^k \leq b_j^U \quad j = 1 \dots nj \\
 c_j^L &\leq c_j^k \leq c_j^U \quad j = 1 \dots nj
 \end{aligned} \tag{3-73}$$

Puede observarse que las eficiencias de las calderas y las turbinas se corrigen con una función cuadrática de la variable auxiliar correspondiente a cada eficiencia,  $v_j$ , mientras que los coeficientes de evaporación se modifican con una función lineal del vapor vivo utilizado en el efecto 1 ( $V_0$ ). Los parámetros  $c_j^k$  sólo toman valores diferentes de cero para ciclos de RTO  $k \geq 4$ .

Como caso de comparación se implementó una estrategia de adaptación de restricciones (CA), que consiste en resolver la ecuación (3-73) con los coeficientes  $b_j^k$  y  $c_j^k$  iguales a 0 en todos los ciclos.

El problema de optimización en tiempo real se formula como

$$\begin{aligned}
& \min Q_1 + Q_2 - c_W^k \cdot W_{TG} \\
& \text{s.t. } \mathbf{f}_1(\mathbf{x}, \mathbf{u}) = 0 \\
& \quad \mathbf{f}_2(\mathbf{x}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\
& \quad \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}) \\
& \quad \eta_j = p_{mj}(\mathbf{u}) + \gamma_j^k(\mathbf{v}_j) \\
& \quad F_{35} = d^k \\
& \quad \|F_{cb1} - F_{cb1}^k\| \leq 5 \\
& \quad \|F_{cb2} - F_{cb2,k}\| \leq 5 \\
& \quad \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \\
& \quad \mathbf{y}^L \leq \mathbf{y} \leq \mathbf{y}^U
\end{aligned} \tag{3-74}$$

donde  $\gamma_j^k$  son los términos de adaptación para el ciclo  $k$ . Los modelos de las eficiencias se expresan como  $p_{mj}(\mathbf{u})$  indicando que el cálculo no se realiza en forma explícita en función de las variables  $\mathbf{v}_j$ . El precio de venta de la electricidad es  $c_W^k$ ; la demanda de vapor para destilación y otros usos es  $d^k$ .

### 3.3.2.3. Resultados y discusión

El modelo se ha implementado en GAMS 24.1 y los problemas de optimización se resolvieron con CONOPT 3. El modelo RTO generado tiene 254 ecuaciones y 218 variables. Los tiempos computacionales de solución de un ciclo de RTO  $k$  están siempre por debajo de 1 s.

La **Figura 3.10** muestra la evolución del costo correspondiente al escenario 1 para 30 ciclos de RTO. También se muestra la evolución del caso sin ruido para ilustrar la convergencia de las estrategias de RTO. Puede observarse que para el caso sin ruido la estrategia MA-E2 converge al óptimo real de la planta, mientras que la estrategia CA, al no corregir los gradientes, converge a un punto operativo diferente. Cuando hay ruido en las mediciones, el desempeño de las dos estrategias disminuye, pero MA-E2 sigue mostrando una mejora significativa con respecto a la adaptación de restricciones CA.

A pesar de que la corrección del gradiente ignora la dependencia de la eficiencia de la etapa 2 con la temperatura de la etapa 1 para el turbogenerador, se alcanza la convergencia a las cercanías del óptimo. Esto se debe a que el efecto de la temperatura es pequeño. De hecho, el impacto de la temperatura sí se considera en forma implícita en el término de corrección  $a_4^k$ , sólo se desprecia el error en el gradiente. Lo mismo puede decirse de los factores distintos de  $V_0$  que impactan en los coeficientes de evaporación.

La **Figura 3.11** muestra los resultados para el escenario 2, junto al perfil de la demanda de vapor a lo largo de los ciclos de RTO. Este escenario y los siguientes incluyen ruido gaussiano en todas las mediciones. La estrategia MA-E2 muestra un desempeño superior a CA para el perfil de demanda propuesto en este escenario.

La **Figura 3.12** muestra los resultados obtenidos para el escenario 3, junto al perfil de precios de venta de electricidad. La estrategia MA-E2 alcanza un costo menor a CA en todos los ciclos, excepto en el ciclo  $k = 2$ , donde aún el problema de regresión cuenta con poca información para adaptar las ecuaciones de eficiencia. Este comportamiento también se observa en los escenarios 1, 2 y 4, ya que los precios y demandas para  $0 \leq k \leq 4$  son los mismos en todos los escenarios. También se observa que en el primer ciclo  $k$  que sigue a un cambio en el precio, la función objetivo se aleja del óptimo. Esto se debe a que las entradas  $\mathbf{u}^k$  se calcularon con las condiciones del ciclo  $k - 1$ . Para la estrategia MA-E2, después de cada cambio de precio, el costo operativo en general se acerca al óptimo en los ciclos siguientes; esto no ocurre para la estrategia CA. Incluso, para la estrategia CA, para muchos cambios de precio, las entradas óptimas calculadas con el precio anterior generan un costo operativo menor que aquellas calculadas con el precio correcto. Esto muestra una vez más que si no se corrigen los errores estructurales del modelo, no hay garantía de que la optimización en tiempo real mejore la función objetivo real.

La **Figura 3.13** muestra los resultados obtenidos para el escenario 4, que combina los cambios en la demanda de vapor y el precio de venta de electricidad de los escenarios anteriores. Nuevamente, la estrategia MA-E2 conduce a menores costos que la adaptación de restricciones CA.

La **Tabla 3-7** presenta la evolución del costo de diseño extendido ( $C_E$ ) para todos los escenarios estudiados. Este indicador del desempeño del sistema de RTO se calcula de acuerdo con (3-58), y se evalúa para el escenario completo ( $k = 0$  a  $k = 30$ ) y para el período desde  $k = 6$  hasta el final del escenario. Se considera que a partir de  $k = 6$  el sistema está inicializado correctamente (es decir, se cuenta con suficiente información para estimar correctamente los modificadores), además de haber alcanzado las cercanías del óptimo real para el escenario 1 y la estrategia MA-E2.

Una comparación de los casos con ruido gaussiano y sin ruido para el escenario 1 muestra que el desempeño de MA-E2 empeora significativamente con la incertidumbre en las mediciones. Para este escenario, el  $C_E$  para  $k \geq 6$  es casi 3

veces mayor para el caso con ruido que para el caso sin incertidumbre en las mediciones (131.4 contra 48.5).

El costo de diseño relativo para este caso de estudio se calcula tomando como referencia el costo de diseño de la estrategia CA:

$$C_E^{rel,CA} = \frac{C_E}{C_E^{CA}} \cdot 100 \quad (3-75)$$

Los valores de  $C_E^{rel,CA}$  muestran la ventaja de utilizar el método MA-E2 con respecto a CA. Para los escenarios 2, 3 y 4 el costo relativo es similar, con una mejora de alrededor del 50% si se considera el escenario completo, y del 60-70% para  $k \geq 6$ .

Los resultados obtenidos en este caso de estudio muestran que la estrategia de adaptación por modificadores utilizando ecuaciones de eficiencia puede ser implementada exitosamente en procesos con un número elevado de entradas, gracias a la simplicidad en la implementación y logrando una mejora en el desempeño con respecto a estrategias tradicionales (sin corrección del gradiente). En particular, se ha ilustrado la utilidad de esta estrategia para sistemas de vapor y potencia, que representan un ejemplo típico e un proceso que puede modelarse con ecuaciones de eficiencia.

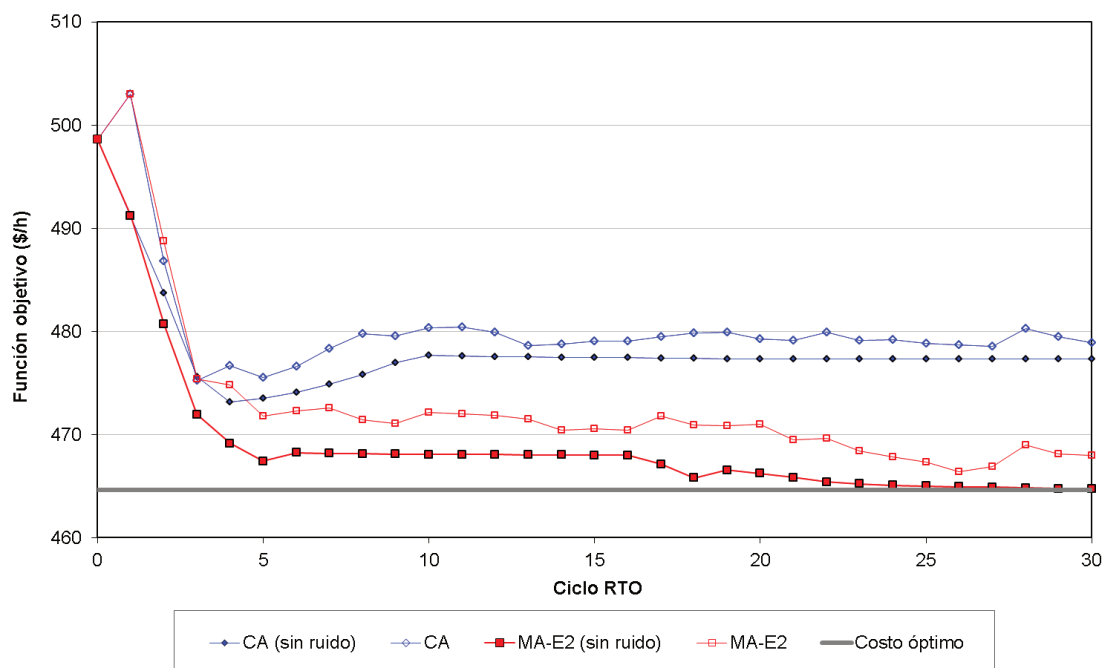


Figura 3.10. Caso de estudio 2 – Escenario 1. Evolución de la función objetivo.

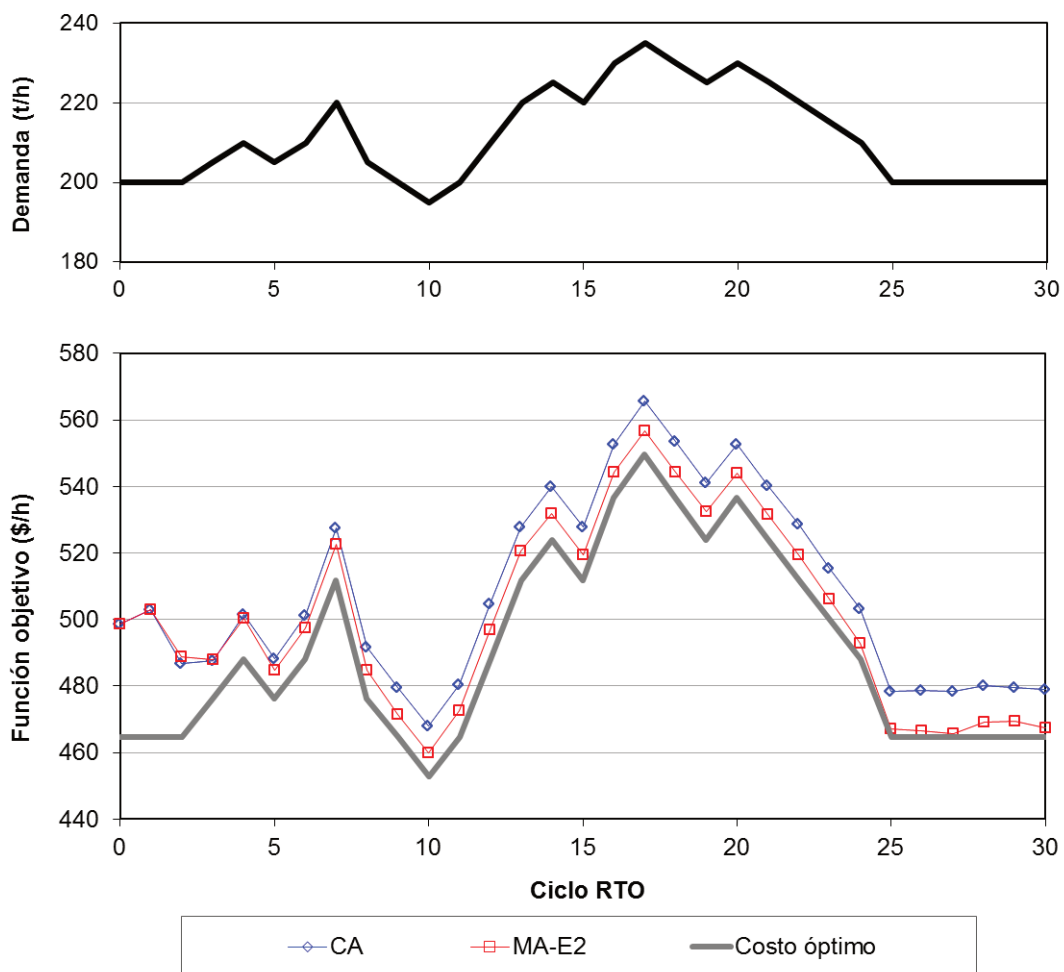


Figura 3.11. Caso de estudio 2 – Escenario 2. Evolución de la función objetivo.

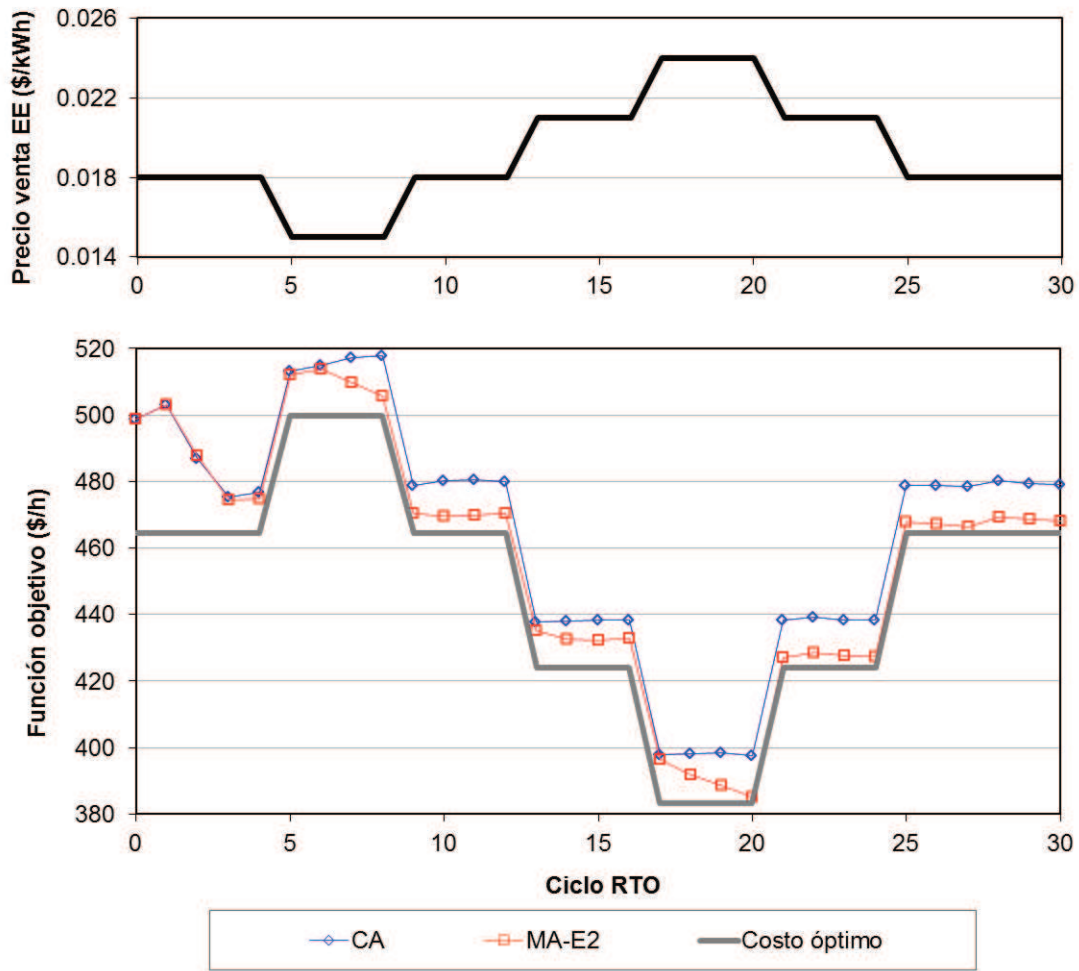


Figura 3.12. Caso de estudio 2 – Escenario 3. Evolución de la función objetivo.

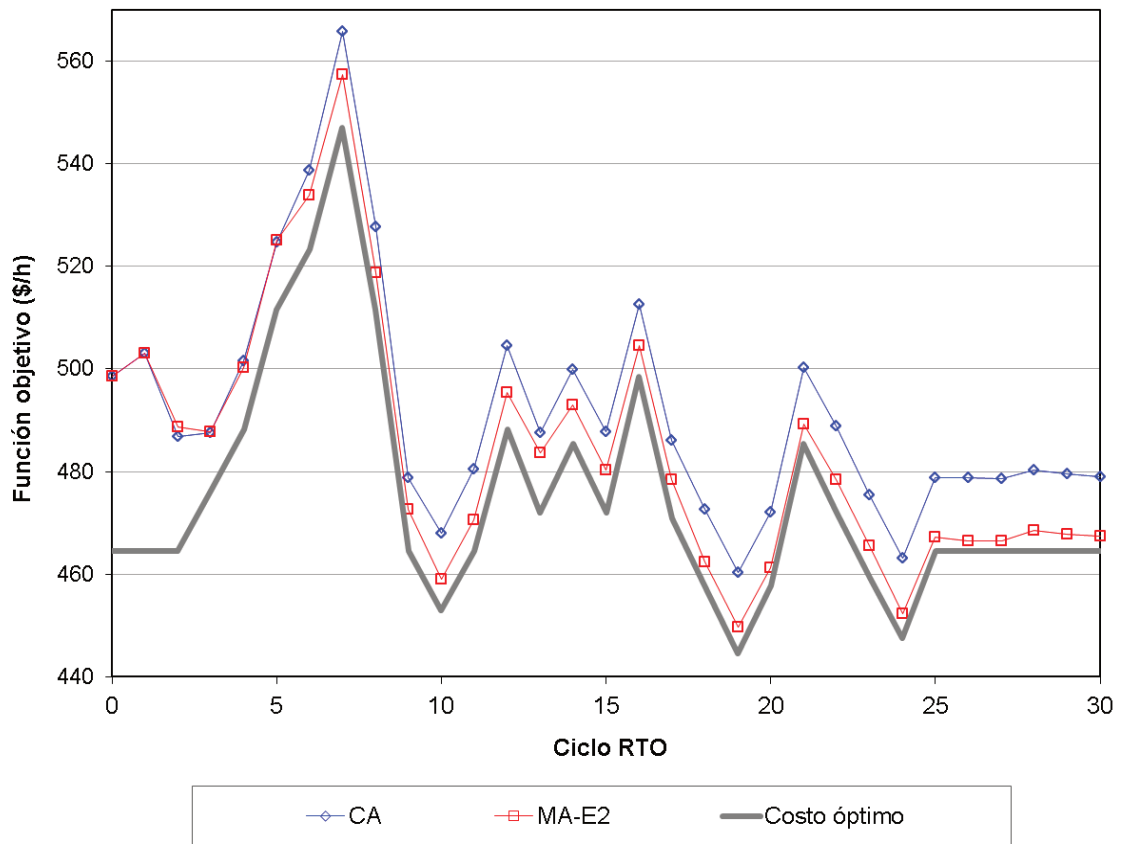


Figura 3.13. Caso de estudio 2 – Escenario 4. Evolución de la función objetivo.

**Tabla 3-7.** Caso de estudio 2. Costo de diseño extendido ( $C_E$ )

Escenarios	$k=0$ a 30		$k=6$ a 30	
	$C_E$	$C_E^{rel,CA}$	$C_E$	$C_E^{rel,CA}$
<b>Escenario 1 – Sin ruido</b>				
CA	397.1	100.0	301.1	100.0
MA-E2	124.7	31.4	48.5	16.1
<b>Escenario 1</b>				
CA	470.3	100.0	353.0	100.0
MA-E2	243.0	51.7	131.4	37.2
<b>Escenario 2</b>				
CA	494.6	100.0	362.1	100.0
MA-E2	275.4	55.7	147.5	40.7
<b>Escenario 3</b>				
CA	479.9	100.0	376.4	100.0
MA-E2	198.1	41.3	91.7	24.4
<b>Escenario 4</b>				
CA	491.2	100.0	338.7	100.0
MA-E2	262.5	53.4	111.3	32.9

## CONCLUSIONES PARCIALES

Se ha propuesto una estrategia alternativa de adaptación por modificadores para RTO. Esta estrategia aprovecha el hecho de que, para algunos tipos de procesos, los errores de modelado pueden concentrarse en un conjunto de *ecuaciones de eficiencia*, mientras los balances de masa, energía y entropía se consideran libres de error. Estas ecuaciones de eficiencia son a menudo función de un conjunto de variables que tiene menos elementos que el vector de entradas del proceso; por lo tanto, el número de ciclos de RTO necesarios para proveer información al problema de estimación del gradiente y a las restricciones de RTO

dual se reduce con respecto a las estrategias de adaptación por modificadores tradicionales.

La estrategia se desarrolló considerando las características de los sistemas de calor y potencia. No obstante, puede aplicarse a otros tipos de procesos, en particular a los que involucran redes con múltiples unidades de proceso u operaciones unitarias, tales como la optimización en tiempo real de una planta completa.

Se han propuesto dos alternativas basadas en la adaptación de ecuaciones de eficiencia: (i) estimación numérica del gradiente combinada con un filtro exponencial; y (ii) una regresión cuadrática de datos de ciclos pasados. Esta regresión permite agregar más términos de corrección (además del *bias* y el modificador del gradiente), expandiendo la región operativa donde la adaptación es válida.

Los dos casos de estudio analizados en este capítulo muestran que cuando el sistema converge, lo hace al óptimo real de la planta, tal como otras estrategias que incluyen corrección del gradiente. En los escenarios analizados, la propuesta tiene un mejor desempeño que las estrategias implementadas para comparación. El caso de estudio 2, que optimiza el sistema de calor y potencia de una planta de azúcar y etanol, muestra que la adaptación de ecuaciones de eficiencia puede aplicarse a este tipo de sistemas, y que es una alternativa viable para sistemas con un número elevado de grados de libertad.

El método no implica necesariamente el uso de modelos simplificados. Por el contrario, los modelos detallados basados en primeros principios y propiedades físicas son apropiados para implementar esta estrategia, ya que permiten detectar las fuentes de error estructural y asignarlas a una ecuación de eficiencia empírica.

La aplicación de la estrategia propuesta requiere un trabajo previo de modelado e identificación, para construir un modelo que separe las ecuaciones de eficiencia del resto de los balances rigurosos y detectar qué variables impactan en forma significativa en cada ecuación de eficiencia.

La adaptación de ecuaciones de eficiencia no excluye a otros métodos como ISOPE o la adaptación por modificadores tradicional. Por el contrario, puede combinarse con estas estrategias o utilizarse sólo como estrategia de estimación del gradiente, mientras se recurre a los procedimientos y los criterios de RTO dual de los métodos originales. De este modo, el conocimiento de la estructura del proceso y el uso inteligente de la información reduciría el problema de estimación del gradiente con respecto a la estimación *sin modelo*, que evalúa la influencia de todas las entradas del proceso.



## 4. RTO CON PROGRAMACIÓN DISYUNTIVA

### RESUMEN

El capítulo 4 presenta la optimización en tiempo real con disyunciones. La sección 4.1 justifica la inclusión de disyunciones en la RTO, tradicionalmente basada en problemas continuos. La sección 4.2 presenta brevemente la programación disyuntiva generalizada (GDP). La sección 4.3 describe las características de un proceso, un sistema de control o una implementación de RTO que originan decisiones discretas. Las estrategias de RTO continua se reformulan para el caso con disyunciones (4.4). Algunas prácticas comunes en RTO continua pueden causar suboptimalidades o infactibilidades cuando se incluyen disyunciones, por lo que es necesario redefinirlas (4.5). Dos casos de estudio (un sistema genérico sencillo y un sistema de calor y potencia) muestran el potencial y las limitaciones de las estrategias propuestas (4.6).

### 4.1. NECESIDAD DEL MODELADO CON DECISIONES DISCRETAS

La programación no lineal mixta entera (MINLP) y la programación disyuntiva generalizada (GDP) no se han utilizado generalmente en RTO por dos razones principales. La primera es que la optimización en tiempo real requiere una solución rápida (en el orden de los segundos o pocos minutos), y los problemas que incluyen decisiones discretas son más complejos computacionalmente que los problemas de programación no lineal (NLP). La segunda es que muchas decisiones discretas, como la puesta en marcha o parada de equipos, no pueden abordarse correctamente con una formulación de un solo período, y por lo tanto se asignan a la etapa de optimización del *schedule*, que resuelve un problema multiperíodo y se ejecuta con una menor frecuencia que la RTO.

Sin embargo, la evolución de la capacidad de procesamiento, así como el desarrollo de algoritmos MINLP y basados en lógica, han conseguido que problemas de un tamaño típico para RTO (cientos de variables y restricciones, y unas pocas decisiones discretas) sean resueltos en el tiempo requerido para esta aplicación. Trabajos recientes muestran aplicaciones industriales de optimización MINLP actualizando el modelo y los precios con datos en tiempo real (Mariani et al., 2009; Ruiz, 2009; Ruiz et al., 2009; Puranik et al., 2013), incluyendo puesta en marcha y parada de equipos. Sin embargo, no se encuentran publicaciones científicas que analicen el impacto de los errores estructurales y paramétricos de

modelado y las estrategias de adaptación para problemas que involucren decisiones discretas.

Por otra parte, como se muestra en la sección 4.3, en una planta industrial pueden presentarse decisiones discretas que no involucran paradas y puestas en marcha y que pueden abordarse a través de la optimización en tiempo real. Por ejemplo, contratos complejos de compra de gas o electricidad, regiones operativas desconectadas, modelos con funciones definidas a trozos, o las ya mencionadas restricciones de *RTO dual* (Marchetti et al., 2010) pueden requerir una formulación con disyunciones lógicas o variables discretas.

Por este motivo, las secciones siguientes de este capítulo proponen la implementación de RTO con programación disyuntiva, reformulando las estrategias de adaptación y las técnicas normalmente utilizadas para problemas continuos.

## 4.2. PROGRAMACIÓN DISYUNTIVA GENERALIZADA (GDP)

Cuando el modelo incluye variables continuas y discretas, un problema de optimización a menudo puede formularse usando *programación disyuntiva generalizada (GDP)* (Raman y Grossmann, 1994). Esta formulación consiste en el uso de restricciones algebraicas, disyunciones lógicas y proposiciones lógicas. La estructura general puede expresarse como:

$$\begin{aligned}
 & \min f(\mathbf{x}) + \sum_d c_d \\
 & \text{s.t. } \mathbf{g}_G(\mathbf{x}) \leq \mathbf{0} \\
 & \bigvee_{i \in N_d} \begin{bmatrix} z_{id} \\ \mathbf{g}_{id}(\mathbf{x}) \leq \mathbf{0} \\ c_d = \gamma_{id} \end{bmatrix}, d \in D \\
 & \mathbf{\Omega}(\mathbf{z}) = \text{verdadero} \\
 & \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U \\
 & \mathbf{x} \in R^n, c_d \in R^1, z_{id} \in \{\text{verdadero}, \text{falso}\}
 \end{aligned} \tag{4-1}$$

donde  $f$  es una función continua de las variables continuas  $\mathbf{x}$ ; las restricciones de desigualdad globales  $\mathbf{g}_G(\mathbf{x}) \leq \mathbf{0}$  deben cumplirse en todos los casos. Cada *disyunción*  $d \in D$  está compuesta de  $N_d$  términos, relacionados a través del operador lógico OR. Cada término  $i \in N_d$  se caracteriza por la variable Booleana  $z_{id}$ ; si  $z_{id} = \text{verdadero}$ , las restricciones  $\mathbf{g}_{id}(\mathbf{x}) \leq \mathbf{0}$  deben cumplirse, y el costo fijo  $c_d$  es igual a  $\gamma_{id}$ . Las proposiciones lógicas  $\mathbf{\Omega}(\mathbf{z}) = \text{verdadero}$  son restricciones en

función de las variables Booleanas  $\mathbf{z}$  (con  $\mathbf{z}$  el vector de todas las variables  $z_{id}$ ) (Grossmann y Ruiz, 2012).

Una vez que el modelo se ha formulado como GDP, puede convertirse en un problema MINLP, a través de reformulaciones *big-M* o de *envoltura convexa* (*convex hull*) (Lee y Grossmann, 2000). De esta manera, la solución del problema puede obtenerse con alguno de los resolvidores MINLP disponibles en la actualidad (GAMS Development Corporation, 2008). También se han desarrollado algoritmos basados en lógica, diseñados para aprovechar la estructura particular de los problemas GDP (Türkay y Grossmann, 1996; Vecchietti y Grossmann, 1999).

La programación disyuntiva generalizada se ha aplicado con éxito en la síntesis de proceso y la planificación/scheduling óptima. En las siguientes secciones se describe formalmente la aplicación de la GDP para optimización en tiempo real y se ilustra con ejemplos.

### 4.3. DISYUNCIONES TÍPICAS EN RTO

Como se mencionó en la sección 4.1, la optimización en tiempo real no suele incluir puesta en marcha y parada de equipos, ya que estas decisiones se pueden abordar de una manera más apropiada en una estrategia multiperíodo, que tiene en cuenta *restricciones de transición* (períodos de puesta en marcha, mínimo tiempo entre puesta en marcha y parada, penalización del número de cambios) (Mitra et al., 2013). Además, si la RTO actúa directamente sobre el sistema de control, podría no ser posible implementar estos cambios en forma automática; si la RTO está en *lazo abierto* (los resultados son aplicados por el operador), la puesta en marcha/parada puede realizarse en forma manual.

Más allá de esto, hay otros aspectos de los procesos industriales, la estrategia de control, el modelado y la implementación de RTO que pueden requerir un modelado con decisiones discretas. Algunos ejemplos se describen en las subsecciones siguientes.

#### 4.3.1. Funciones definidas a trozos

Las funciones definidas a trozos pueden incluirse en un problema de optimización con el objetivo de modelar el comportamiento del proceso (o la estrategia de control) con diferentes funciones en distintas regiones de operación:

$$\left[ \begin{array}{c} z_{d,1} \\ g = g_1(u) \leq 0 \\ u_1^L \leq u \leq u_2^L \end{array} \right] \vee \dots \vee \left[ \begin{array}{c} z_{d,p} \\ g = g_p(u) \leq 0 \\ u_p^L \leq u \leq u_{p+1}^L \end{array} \right] \vee \dots \vee \left[ \begin{array}{c} z_{d,np} \\ g = g_{np}(u) \leq 0 \\ u_{np}^L \leq u \leq u_{np+1}^L \end{array} \right] \quad (4-2)$$

donde  $p$  indica el trozo activo (1 a  $np$ ),  $g_p$  es la restricción que debe satisfacerse en el trozo  $p$ , y  $u_p^L$  es el límite inferior para cada trozo  $p$ .

#### 4.3.2. Contratos de compra y venta – Penalidades en la función objetivo

Si la función objetivo es el costo o el beneficio, puede incluir costos fijos o penalidades que aplican bajo ciertas condiciones. Puede incluirse una penalidad fija, por ejemplo, en un contrato de compra de electricidad si la potencia excede un valor máximo; las emisiones pueden penalizarse con una multa fija si se supera el límite permitido. Frecuentemente, estos costos fijos necesitan ser modelados con variables discretas o disyunciones.

Las funciones definidas a trozos de la sección 4.3.1 también pueden utilizarse para modelar contratos complejos incluidos en la función objetivo.

#### 4.3.3. Regiones operativas desconectadas

La estructura del sistema de control de un proceso puede generar regiones operativas desconectadas en el sistema a optimizar. Por ejemplo, una válvula que puede operar con una apertura del 5% al 100% o bien estar completamente cerrada, puede modelarse con una disyunción. La política de operación del proceso también puede incluir reglas que excluyan zonas de la región factible por razones de seguridad o confiabilidad.

#### 4.3.4. RTO dual

Las técnicas de RTO dual (descritas en la sección 2.4.5) requieren de la estimación experimental de los gradientes del costo y las restricciones, utilizando datos de los ciclos de RTO pasados. Para obtener información útil para la estimación, se agregan restricciones adicionales al problema de optimización. Estas restricciones limitan el error generado por el ruido y por truncamiento (Marchetti, 2013) y permiten evaluar el gradiente en todas las direcciones del espacio de las entradas (Rodger y Chachuat, 2011).

Las restricciones adicionales de RTO dual generan dos regiones factibles desconectadas a ambos lados del hiperplano definido por los últimos  $nu$  puntos en el espacio de las entradas (donde  $nu$  es el número de entradas del sistema). Esto genera una disyunción en cada ciclo de RTO  $k$ , que puede modelarse como:

$$\left[ e^+ \left( \mathbf{u}, \mathbf{u}^k, \dots, \mathbf{u}^{k-nu+1} \right) \leq 0 \right] \vee \left[ e^- \left( \mathbf{u}, \mathbf{u}^k, \dots, \mathbf{u}^{k-nu+1} \right) \leq 0 \right] \quad (4-3)$$

donde la variable booleana  $z_{dual}$  se utiliza para seleccionar una de las regiones factibles restringidas por las funciones  $e^+$  y  $e^-$ . Las únicas variables continuas en esta disyunción son las entradas futuras  $\mathbf{u}$ , mientras que las entradas actuales  $\mathbf{u}^k$  y las pasadas  $\mathbf{u}^{k-1}, \dots, \mathbf{u}^{k-nu+1}$  son parámetros que definen las regiones factibles.

La práctica usual para modelos continuos no lineales es resolver el NLP para cada región factible, y luego aplicar la solución (factible) con mejor valor de la función objetivo. Si ya se utiliza un modelo de programación disyuntiva para RTO, las restricciones de RTO dual pueden incluirse como una nueva disyunción, para resolver el problema completo como GDP o MINLP.

#### 4.3.5. Cambios mínimos

Una propiedad deseable para una implementación de RTO es que la variabilidad en las entradas no sea excesiva (sobre todo si se opera en cercanías del óptimo, donde pequeños cambios en las variables de entrada pueden perturbar la operación normal del proceso pero causar un impacto casi nulo en la función objetivo). Por otra parte, en los casos de RTO a *lazo abierto*, las entradas óptimas son aplicadas manualmente por los operadores, que por razones prácticas preferirán evitar la aplicación de cambios pequeños en las entradas.

Para estas dos situaciones, una alternativa posible sería validar el resultado óptimo, y no aplicarlo si el cambio propuesto en las entradas es pequeño. Sin embargo, una alternativa posible cuando el modelo contiene disyunciones es incluir una restricción en el cambio mínimo permitido para las entradas  $\mathbf{u}$ , que pueden alternativamente mantenerse en su valor actual:

$$\left[ \begin{array}{c} z_i^- \\ u_i \leq u_i^k - \Delta u_{i,\min} \end{array} \right] \vee \left[ \begin{array}{c} z_i^+ \\ u_i \geq u_i^k + \Delta u_{i,\min} \end{array} \right] \vee \left[ \begin{array}{c} z_i^- \\ u_i = u_i^k \end{array} \right], \quad i = 1..nu; \Delta u_{i,\min} \geq 0 \quad (4-4)$$

Cuando se utilizan las restricciones de RTO dual de la sección 4.3.4, el punto actual  $\mathbf{u}^k$  no es factible, al igual que todo el hiperplano formado por los últimos  $nu$  puntos. Si bien esto no es incompatible con las disyunciones de (4-4), cada una de las entradas no podrá permanecer en el mismo punto por más de  $nu$  ciclos de RTO (es decir, una variable  $z_i^-$  no puede ser *verdadera* por más de  $nu$  ciclos seguidos).

Para demostrar este hecho, consideremos el hiperplano que contiene los últimos  $nu$  puntos, definido por la ecuación  $\mathbf{a}^T \cdot \mathbf{u} = b$ , donde  $\mathbf{a}$  y  $b$  son coeficientes constantes. El vector de coeficientes,  $(\mathbf{a}^T \ b)^T = (a_1 \ \dots \ a_{nu} \ b)^T$  puede obtenerse como uno de los elementos del espacio nulo de la matriz  $A$  de  $nu \times (nu+1)$

$$A = \begin{bmatrix} \mathbf{u}^k & \mathbf{u}^{k-1} & \dots & \mathbf{u}^{k-nu-1} \\ -1 & -1 & \dots & -1 \end{bmatrix}^T = \begin{bmatrix} u_1^k & \dots & u_{nu}^k & -1 \\ \vdots & \ddots & \vdots & \vdots \\ u_1^{k-nu-1} & \dots & u_{nu}^{k-nu-1} & -1 \end{bmatrix} \quad (4-5)$$

Como las restricciones de (4-3) se aplicaron en los ciclos de RTO anteriores,  $\mathbf{u}^k$  no pertenece al hiperplano definido por los vectores  $(\mathbf{u}^{k-1}, \dots, \mathbf{u}^{k-nu})$ , y por lo tanto, tampoco pertenece al subespacio definido por los puntos  $(\mathbf{u}^{k-1}, \dots, \mathbf{u}^{k-nu-1})$ . De la misma manera,  $\mathbf{u}^{k-1}$  no pertenece al subespacio definido por  $(\mathbf{u}^{k-2}, \dots, \mathbf{u}^{k-nu-1})$ , y la misma conclusión puede obtenerse para el resto de los vectores que forman la matriz  $A$ . Por lo tanto, ninguna de las entradas puede obtenerse como combinación lineal de las entradas anteriores. Esto implica que  $A$  es de rango  $nu$  y el espacio nulo de  $A$  genera un hiperplano.

Suponiendo que el elemento  $u_1$  tiene el mismo valor en los últimos  $nu$  ciclos, es decir  $u_1^k = u_1^{k-1} = \dots = u_1^{k-nu-1}$ . Entonces, la primera fila de la matriz  $A$  puede restarse a las otras filas, y se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned} a_2 \cdot (u_2^{k-1} - u_2^k) + \dots + a_{nu} \cdot (u_{nu}^{k-1} - u_{nu}^k) &= 0 \\ \vdots & \\ a_2 \cdot (u_2^{k-nu-1} - u_2^k) + \dots + a_{nu} \cdot (u_{nu}^{k-nu-1} - u_{nu}^k) &= 0 \end{aligned} \quad (4-6)$$

Como el sistema se obtuvo como combinación lineal de las filas de (4-5), que era de rango  $nu$ , la ecuación (4-6) (de  $nu-1$  ecuaciones y  $nu-1$  incógnitas) tiene solución única:  $a_2 = a_3 = \dots = a_{nu} = 0$ . Entonces, el hiperplano definido por los últimos  $nu$  puntos es:

$$u_1 = u_1^k \quad (4-7)$$

que es el último término de la disyunción (4-4). Como las restricciones de RTO dual obligan a  $\mathbf{u}$  a estar fuera del hiperplano (4-7), esta opción no es factible y el sistema estará obligado a seleccionar una de las otras dos opciones de (4-4). Reordenando los elementos de  $\mathbf{u}$ , las ecuaciones (4-5)-(4-7) pueden aplicarse a cualquiera de los elementos  $u_i$  del vector de entradas.

## 4.4. OPTIMIZACIÓN EN TIEMPO REAL INCLUYENDO DISYUNCIONES

### 4.4.1. Definición del problema

Un problema de optimización en tiempo real incluyendo disyunciones puede definirse como

$$\begin{aligned}
 & \min_{\mathbf{u}} Q(\mathbf{y}, \mathbf{u}) + \sum_{d \in D} c_d \\
 & \text{s.t. } \mathbf{h}_G(\mathbf{y}, \mathbf{u}) = 0 \\
 & \quad \mathbf{g}_G(\mathbf{y}, \mathbf{u}) \leq 0 \\
 & \quad \bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id} \\ \mathbf{h}_{id}(\mathbf{y}, \mathbf{u}) = 0 \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) \leq 0 \\ c_d = \gamma_{id} \end{array} \right], d \in D \\
 & \quad \mathbf{\Omega}(\mathbf{z}) = \text{verdadero} \\
 & \quad \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \\
 & \quad z_{id} \in \{\text{verdadero}, \text{falso}\} \forall i \in N_d, d \in D
 \end{aligned} \tag{4-8}$$

donde  $\mathbf{u}$  son las entradas del proceso,  $\mathbf{y}$  las salidas y  $\mathbf{z}$  las variables lógicas correspondientes a las disyunciones  $d \in D$ . Las restricciones *globales* de igualdad y desigualdad,  $\mathbf{h}_G$  y  $\mathbf{g}_G$  respectivamente, son válidas sin importar qué términos de las disyunciones  $d$  están activos (es decir, qué variables  $z_{id}$  tienen el valor *verdadero*); por el contrario,  $\mathbf{h}_{id}$  y  $\mathbf{g}_{id}$  deben cumplirse solo si  $z_{id} = \text{verdadero}$ . Las proposiciones lógicas  $\mathbf{\Omega}$  imponen restricciones adicionales en los valores de las variables  $\mathbf{z}$ .

La funcionalidad real de las salidas del proceso con respecto a las entradas no puede conocerse en la práctica. En su lugar, se dispone de un modelo del proceso que es capaz de predecir en forma aproximada las funciones  $\mathbf{h}_G$  y  $\mathbf{h}_{id}$ :

$$\begin{aligned}
 & \mathbf{h}_G(\mathbf{y}, \mathbf{u}) \approx \mathbf{f}_G(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_G) = 0 \\
 & \quad \bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id} \\ \mathbf{h}_{id}(\mathbf{y}, \mathbf{u}) \approx \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{id}) = 0 \end{array} \right], d \in D
 \end{aligned} \tag{4-9}$$

donde  $\boldsymbol{\theta}_G$  y  $\boldsymbol{\theta}_{id}$  son parámetros del modelo. Los estados del proceso (llamados frecuentemente  $\mathbf{x}$ ) aparecen implícitamente en (4-9) y no se mencionarán explícitamente en este capítulo para mantener la simplicidad en la nomenclatura.

#### 4.4.2. Estrategias de adaptación

##### 4.4.2.1. Estrategia de dos pasos

Los parámetros ajustables pueden estimarse al comienzo de cada ciclo de RTO  $k$  usando las salidas medidas  $\mathbf{y}^k$  resolviendo el siguiente problema de optimización:

$$\begin{aligned}
 \boldsymbol{\theta}_G^k, \boldsymbol{\theta}_{id}^k &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} (\mathbf{y}^k - \mathbf{y})^T \cdot (\mathbf{y}^k - \mathbf{y}) \\
 \text{s.t. } \mathbf{f}_G(\mathbf{y}, \mathbf{u}^k, \boldsymbol{\theta}_G) &= 0 \\
 \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}^k, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{id}) &= 0, \quad \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \\
 \boldsymbol{\theta}_{id} &= \boldsymbol{\theta}_{id}^{k-1}, \quad \forall i \in N_d, d \in D / z_{id}^k = \text{falso}
 \end{aligned} \tag{4-10}$$

Una práctica usual para reducir el efecto del ruido en la variabilidad de los parámetros estimados es filtrar los resultados usando valores de ciclos anteriores. Una implementación de filtrado exponencial para RTO con disyunciones para actualizar los parámetros a partir de (4-10) puede implementarse de la forma siguiente:

$$\begin{aligned}
 \boldsymbol{\theta}_G^k &:= K_G \cdot \boldsymbol{\theta}_G^k + (I_G - K_G) \cdot \boldsymbol{\theta}_G^{k-1} \\
 \boldsymbol{\theta}_{id}^k &:= K_{id} \cdot \boldsymbol{\theta}_{id}^k + (I_{id} - K_{id}) \cdot \boldsymbol{\theta}_{id}^{k-1}, \quad \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero}
 \end{aligned} \tag{4-11}$$

donde  $K_G$  y  $K_{id}$  son las matrices de filtrado, e  $I_G$  e  $I_{id}$  son matrices identidad de la dimensión apropiada.

El problema de optimización a resolver se formula entonces como:

$$\begin{aligned}
 &\min_{\mathbf{u}} Q(\mathbf{y}, \mathbf{u}) + \sum_{d \in D} c_d \\
 &\text{s.t. } \mathbf{f}_G(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_G^k) = 0 \\
 &\quad \mathbf{g}_G(\mathbf{y}, \mathbf{u}) \leq 0 \\
 &\quad \bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id} \\ \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_G^k, \boldsymbol{\theta}_{id}^k) = 0 \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) \leq 0 \\ c_d = \gamma_{id} \end{array} \right], d \in D \\
 &\quad \boldsymbol{\Omega}(\mathbf{z}) = \text{verdadero} \\
 &\quad \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \\
 &\quad z_{id} \in \{\text{verdadero}, \text{falso}\} \quad \forall i \in N_d, d \in D
 \end{aligned} \tag{4-12}$$

##### 4.4.2.2. Adaptación por modificadores

Como se ha mencionado en la sección 2.3.2, la estrategia de adaptación por modificadores no resuelve un problema de optimización para actualizar

parámetros, sino que mantiene los parámetros constantes y adapta las restricciones del modelo con un *bias* (*modificador de la restricción*) y corrige los gradientes del *costo* (función objetivo) y las restricciones con un término de corrección (*modificador del gradiente*). La extensión a RTO con disyunciones puede formularse como:

$$\begin{aligned}
 \beta_G^k &= \mathbf{g}_G(\mathbf{y}^k, \mathbf{u}^k) - \mathbf{g}_G(\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta}), \mathbf{u}^k) \\
 \lambda_G^k &= \nabla_{\mathbf{u}} \mathbf{g}_G^{real}(\mathbf{y}^k, \mathbf{u}^k) - \nabla_{\mathbf{u}} \mathbf{g}_G(\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta}), \mathbf{u}^k) \\
 \lambda_Q^k &= \nabla_{\mathbf{u}} Q^{real}(\mathbf{y}^k, \mathbf{u}^k) - \nabla_{\mathbf{u}} Q(\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta}), \mathbf{u}^k) \\
 \beta_{id}^k &= \mathbf{g}_{id}(\mathbf{y}^k, \mathbf{u}^k) - \mathbf{g}_{id}(\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta}), \mathbf{u}^k) \\
 \lambda_{id}^k &= \nabla_{\mathbf{u}} \mathbf{g}_{id}^{real}(\mathbf{y}^k, \mathbf{u}^k) - \nabla_{\mathbf{u}} \mathbf{g}_{id}(\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta}), \mathbf{u}^k) \\
 \mathbf{u}_{id}^{ref} &= \mathbf{u}^k
 \end{aligned}
 \left. \vphantom{\begin{aligned} \beta_G^k \\ \lambda_G^k \\ \lambda_Q^k \\ \beta_{id}^k \\ \lambda_{id}^k \\ \mathbf{u}_{id}^{ref} \end{aligned}} \right\} \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \quad (4-13)$$

$$\left. \vphantom{\begin{aligned} \beta_{id}^k \\ \lambda_{id}^k \end{aligned}} \right\} \forall i \in N_d, d \in D / z_{id}^k = \text{falso}$$

donde  $\beta_G^k$  son los modificadores de las restricciones globales,  $\lambda_G^k$  los modificadores del gradiente de las restricciones globales, y  $\lambda_Q^k$  el modificador del gradiente del costo.  $\beta_{id}^k$  y  $\lambda_{id}^k$  son los modificadores de las restricciones y de los gradientes correspondientes al término  $i \in N_d, d \in D$ . El superíndice *real* se usa para el gradiente obtenido experimentalmente. Las salidas estimadas a partir del modelo del proceso se expresan como  $\mathbf{y}(\mathbf{u}^k, \boldsymbol{\theta})$ . Los modificadores del gradiente de las restricciones globales y del costo se utilizan para construir una corrección lineal en torno a las entradas actuales  $\mathbf{u}^k$ . En cambio, para los modificadores del gradiente en las disyunciones,  $\lambda_{id}^k$ , la corrección lineal se construye en torno a las entradas  $\mathbf{u}_{id}^{ref}$ , que sólo se actualizan con el valor actual de las entradas cuando la variable  $z_{id}^k$  tiene el valor *verdadero*, ya que cuando  $z_{id}^k = \text{falso}$  no se tiene una nueva estimación del gradiente, y por lo tanto el último gradiente experimental disponible corresponde a las entradas  $\mathbf{u}_{id}^{ref}$ .

La variabilidad de los modificadores puede suavizarse con un filtro exponencial. El filtro puede construirse en forma análoga a la sugerida en (4-11), por lo que no se incluye aquí su formulación.

Una vez que los modificadores se han calculado con (4-13), la función objetivo y las restricciones corregidas se obtienen como:

$$\begin{aligned}
& \min_{\mathbf{u}, \mathbf{y}} Q(\mathbf{y}, \mathbf{u}) + \sum_{d \in D} c_d + (\boldsymbol{\lambda}_Q^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\
& \mathbf{g}_G(\mathbf{y}, \mathbf{u}) + \boldsymbol{\beta}_G^k + (\boldsymbol{\lambda}_G^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \leq 0 \\
& \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) + \boldsymbol{\beta}_{id}^k + (\boldsymbol{\lambda}_{id}^k)^T \cdot (\mathbf{u} - \mathbf{u}_{id}^{ref}) \leq 0, \quad i \in N_d, d \in D
\end{aligned} \tag{4-14}$$

El problema de RTO adaptado se formula entonces como:

$$\begin{aligned}
& \min_{\mathbf{u}, \mathbf{y}} Q(\mathbf{y}, \mathbf{u}) + \sum_{d \in D} c_d + (\boldsymbol{\lambda}_Q^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\
& \text{s.t.} \quad \mathbf{f}_G(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_G^k) = 0 \\
& \quad \mathbf{g}_G(\mathbf{y}, \mathbf{u}) + \boldsymbol{\beta}_G^k + (\boldsymbol{\lambda}_G^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \leq 0 \\
& \quad \left[ \begin{array}{c} z_{id} \\ \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}^k, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{id}) = 0 \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) + \boldsymbol{\beta}_{id}^k + (\boldsymbol{\lambda}_{id}^k)^T \cdot (\mathbf{u} - \mathbf{u}_{id}^{ref}) \leq 0 \\ c_d = \gamma_{id} \end{array} \right]_{i \in N_d}, d \in D \\
& \quad \left[ \begin{array}{c} z_{dual} \\ \mathbf{e}^+(\mathbf{u}) \leq 0 \end{array} \right] \vee \left[ \begin{array}{c} \neg z_{dual} \\ \mathbf{e}^-(\mathbf{u}) \leq 0 \end{array} \right] \\
& \quad \boldsymbol{\Omega}(\mathbf{z}, z_{dual}) = \text{verdadero} \\
& \quad \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \\
& \quad z_{id} \in \{\text{verdadero}, \text{falso}\} \quad \forall i \in N_d, d \in D \\
& \quad z_{dual} \in \{\text{verdadero}, \text{falso}\}
\end{aligned} \tag{4-15}$$

donde la disyunción adicional caracterizada por  $z_{dual}$  genera dos regiones factibles desconectadas a través de las restricciones  $\mathbf{e}^+(\mathbf{u})$  y  $\mathbf{e}^-(\mathbf{u})$ , descritas en la sección 4.3.4.

En realidad, la actualización del modificador del gradiente en las disyunciones podría llevar a más restricciones de RTO dual, cada una basada en los valores actuales y pasados de las entradas  $\mathbf{u}_{id}^{ref}$ . Si esto ocurre, cada disyunción puede reescribirse incluyendo una restricción de RTO dual en cada término, caracterizada por la variable lógica  $z_{id,dual}$ :

$$\bigvee_{i \in N_d} \left[ \begin{array}{c} z_{id} \\ \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}^k, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{id}) = 0 \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) + \boldsymbol{\beta}_{id}^k + (\boldsymbol{\lambda}_{id}^k)^T \cdot (\mathbf{u} - \mathbf{u}_{id}^{ref}) \leq 0 \\ c_d = \gamma_{id} \\ \left[ \begin{array}{c} z_{id,dual} \\ \mathbf{e}_{id}^+(\mathbf{u}) \leq 0 \end{array} \right] \bigvee \left[ \begin{array}{c} \neg z_{id,dual} \\ \mathbf{e}_{id}^-(\mathbf{u}) \leq 0 \end{array} \right] \end{array} \right], d \in D \quad (4-16)$$

La ecuación (4-16) muestra una posible estructura de las disyunciones en una estrategia dual de adaptación por modificadores, que no se abordará en detalle en esta tesis.

Como se explica en la sección 2.4.5, las restricciones de RTO dual sólo aparecen cuando la estimación del gradiente utiliza datos de los estados estacionarios alcanzados en cada ciclo de RTO. Si se utilizan otros métodos de cálculo experimental del gradiente (François et al., 2012; Mansour y Ellis, 2003; Srinivasan et al., 2011) estas restricciones no son necesarias.

#### 4.4.2.3. Adaptación por modificadores con ecuaciones de eficiencia

Las técnicas propuestas en el capítulo 3 pueden extenderse al caso de RTO con disyunciones. Esta alternativa a la adaptación por modificadores original puede utilizarse en sistemas de procesos formados por operaciones unitarias interconectadas. En estos sistemas, una parte del modelo puede asumirse libre de error estructural (balances de masa, energía y entropía). A menudo, el subconjunto de ecuaciones del modelo que contiene el error estructural está formado por expresiones que caracterizan la eficiencia de un equipo o de una operación.

El problema a resolver, sin adaptación, es:

$$\begin{array}{l} \min_{\mathbf{u}} Q(\mathbf{y}, \mathbf{u}) + \sum_d c_d \\ \text{s.t. } \mathbf{f}_G(\mathbf{y}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\ \boldsymbol{\eta}_G = \mathbf{p}_G(\mathbf{u}) \\ \mathbf{g}_G(\mathbf{y}, \mathbf{u}) \leq 0 \\ \bigvee_{i \in N_d} \left[ \begin{array}{c} z_{id} \\ \mathbf{f}_{id}(\mathbf{y}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\ \eta_{id} = p_{id}(\mathbf{u}) \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) \leq 0 \\ c_d = \gamma_{id} \end{array} \right], d \in D \\ \boldsymbol{\Omega}(\mathbf{z}) = \text{verdadero} \\ \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U, z_{id} \in \{\text{verdadero}, \text{falso}\} \forall i \in N_d, d \in D \end{array} \quad (4-17)$$

Los factores de eficiencia  $\boldsymbol{\eta} = \{\boldsymbol{\eta}_G, \eta_{id}, \forall i \in N_d, d \in D\}$  se estiman en forma aproximada a través de las funciones  $\mathbf{p}_G(\mathbf{u})$  y  $p_{id}(\mathbf{u})$  ( $\forall i \in N_d, d \in D$ ), que son las que contienen todo el error estructural. Como se explica en la sección 3.1, estas funciones pueden escribirse con frecuencia en términos de variables auxiliares  $\mathbf{v}$ , de dimensión menor o igual a la del vector  $\mathbf{u}$ :

$$\begin{aligned} \eta_{G,j} &= p_{G,j}(\mathbf{u}) = p_{G,j}(\mathbf{v}_{G,j}(\mathbf{u})) \quad j=1..ng_G \\ \eta_{id} &= p_{id}(\mathbf{u}) = p_{id}(\mathbf{v}_{id}(\mathbf{u})) \quad \forall i \in N_d, d \in D \end{aligned} \quad (4-18)$$

donde  $ng_G$  es la dimensión del vector  $\boldsymbol{\eta}_G$  (es decir, el número de factores de eficiencia globales).

En un ciclo de RTO  $k$ , los factores de eficiencia actuales  $\boldsymbol{\eta}^k$  se obtienen del modelo y los valores actuales  $(\mathbf{y}^k, \mathbf{u}^k)$  (medidos o conocidos) resolviendo el sistema de ecuaciones:

$$\begin{aligned} \mathbf{f}_G(\mathbf{y}^k, \mathbf{u}^k, \boldsymbol{\eta}^k) &= 0 \\ \mathbf{f}_{id}(\mathbf{y}^k, \mathbf{u}^k, \boldsymbol{\eta}^k) &= 0 \quad \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \end{aligned} \quad (4-19)$$

que incluye las restricciones de igualdad globales y los términos de las disyunciones activos al iniciar el ciclo (aquellos en las que  $z_{id}^k = \text{verdadero}$ ).

A partir de esta información, se obtienen los modificadores de las restricciones:

$$\begin{aligned} \beta_{G,j}^k &= \eta_{G,j}^{m,k} - p_{G,j}(\mathbf{v}_{G,j}^k), \quad j=1, \dots, ng_G \\ \beta_{id}^k &= \eta_{id}^{m,k} - p_{id}(\mathbf{v}_{id}^k), \quad \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \\ \beta_{id}^k &= \beta_{id}^{k-1}, \quad \forall i \in N_d, d \in D / z_{id}^k = \text{falso} \end{aligned} \quad (4-20)$$

Extendiendo la estrategia MA-E1 (sección 3.2.1) los modificadores del gradiente pueden calcularse como:

$$\begin{aligned} \boldsymbol{\lambda}_{G,j}^k &= \nabla_{\mathbf{v}_{G,j}} \eta_{G,j}^{real} - \nabla_{\mathbf{v}_{G,j}} p_{G,j}(\mathbf{v}_{G,j}^k), \quad j=1, \dots, ng_G \\ \boldsymbol{\lambda}_{id}^k &= \nabla_{\mathbf{v}_{id}} \eta_{id}^{real} - \nabla_{\mathbf{v}_{id}} p_{id}(\mathbf{v}_{id}^k), \quad \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \\ \boldsymbol{\lambda}_{id}^k &= \boldsymbol{\lambda}_{id}^{k-1}, \quad \forall i \in N_d, d \in D / z_{id}^k = \text{falso} \end{aligned} \quad (4-21)$$

Como alternativa, es posible efectuar una regresión lineal para obtener los modificadores, extendiendo la estrategia MA-E2 presentada en la sección 3.2.2:

$$\left. \begin{aligned}
 a_{G,j}^k, \mathbf{b}_{G,j}^k, C_{G,j}^k = \operatorname{argmin} & \sum_{m=k-M_{G,j}-1}^k w_{G,j}^m \cdot \left( \beta_{G,j}^m - a_{G,j} - \mathbf{b}_{G,j}^T \cdot \mathbf{v}_{G,j}^m \right. \\
 & \left. - \left( \mathbf{v}_{G,j}^m \right)^T \cdot C_{G,j} \cdot \mathbf{v}_{G,j}^m \right)^2 \\
 \text{s.t. } a_{G,j}^L & \leq a_{G,j} \leq a_{G,j}^U \\
 \mathbf{b}_{G,j}^L & \leq \mathbf{b}_{G,j} \leq \mathbf{b}_{G,j}^U \\
 C_{G,j}^L & \leq C_{G,j} \leq C_{G,j}^U
 \end{aligned} \right\}, j = 1..ng_G \quad (4-22)$$

$$\left. \begin{aligned}
 a_{id}^k, \mathbf{b}_{id}^k, C_{id}^k = \operatorname{argmin} & \sum_{m=k-M_{id}-1}^k w_{id}^m \cdot \left( \beta_{id}^m - a_{id} - \mathbf{b}_{id}^T \cdot \mathbf{v}_{id}^m \right. \\
 & \left. - \left( \mathbf{v}_{id}^m \right)^T \cdot C_{id} \cdot \mathbf{v}_{id}^m \right)^2 \\
 \text{s.t. } a_{id}^L & \leq a_{id} \leq a_{id}^U \\
 \mathbf{b}_{id}^L & \leq \mathbf{b}_{id} \leq \mathbf{b}_{id}^U \\
 C_{id}^L & \leq C_{id} \leq C_{id}^U
 \end{aligned} \right\} \begin{array}{l} \forall i \in N_d, d \in D \\ / z_{id}^k = \text{verdadero} \end{array} \quad (4-23)$$

donde  $a$ ,  $\mathbf{b}$  y  $C$  son un escalar, un vector y una matriz cuadrada, respectivamente. Los modificadores  $a$  y  $\mathbf{b}$  son equivalentes a  $\beta$  y  $\lambda$  de la adaptación por modificadores original, mientras  $C$  provee grados de libertad adicionales para la corrección del modelo.  $M_{G,j}$  es la cantidad de datos tenidos en cuenta para la regresión en una ecuación global  $j$ ;  $M_{id}$  es la cantidad de datos para la regresión en la ecuación del término  $i$  en la disyunción  $d$ .

Los pesos  $w_{G,j}^m$  y  $w_{id}^m$  ponderan el impacto en la regresión de los datos de ciclos de RTO pasados. Si el término  $i$  de la disyunción  $d$  no está activo en el ciclo  $m$  ( $z_{id}^m = \text{falso}$ ), el peso  $w_{id}^m$  será igual a 0.

Una posible estrategia para la asignación de los pesos consiste en penalizar la antigüedad de los datos (ya que cuanto más antiguos son los datos, mayor es la probabilidad de que haya disturbancias de baja frecuencia que tengan un valor diferente al valor actual). En el ciclo  $k$ , los pesos se actualizan como:

$$\left. \begin{array}{l} w_{G,j}^k = 1 \\ w_{G,j}^m := \alpha_{G,j} w_{G,i}^m \\ 0 \leq \alpha_{G,j} \leq 1 \end{array} \right\}, \quad \begin{array}{l} j = 1, \dots, ng_G \\ m = k - M_{G,j} - 1, \dots, k - 1 \end{array} \quad (4-24)$$

$$\begin{array}{ll} w_{id}^k = 1 & \forall i \in N_d, d \in D / z_{id}^k = \text{verdadero} \\ w_{id}^k = 0 & \forall i \in N_d, d \in D / z_{id}^k = \text{falso} \\ w_{id}^m := \alpha_{id} w_{id}^m & \forall i \in N_d, d \in D, m = k - M_{id} - 1, \dots, k - 1 \\ 0 \leq \alpha_{id} \leq 1 & \forall i \in N_d, d \in D \end{array}$$

Los factores  $\alpha_{G,j}$  y  $\alpha_{id}$  determinan una caída exponencial del peso de un conjunto de datos a medida que transcurre el tiempo.

Una vez que se han obtenido los modificadores, el problema de optimización con el modelo adaptado se formula como:

$$\begin{array}{l} \min Q(\mathbf{y}, \mathbf{u}) + \sum_{d \in D} c_d \\ \text{s.t.} \quad \mathbf{h}_G(\mathbf{y}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\ \left. \begin{array}{l} \eta_{G,j} = p_{G,j}(\mathbf{u}) + a_{G,j}^k + (\mathbf{b}_{G,j}^k)^T \cdot \mathbf{v}_{G,j} + (\mathbf{v}_{G,j})^T C_{G,j}^k \cdot \mathbf{v}_{G,j} \\ \mathbf{v}_{G,j} = \mathbf{v}_{G,j}(\mathbf{u}) \\ \mathbf{g}_G(\mathbf{y}, \mathbf{u}) \leq 0 \end{array} \right\}, j = 1 \dots ng_G \\ \bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id} \\ \mathbf{h}_{id}(\mathbf{y}, \mathbf{u}, \boldsymbol{\eta}) = 0 \\ \eta_{id} = p_{id}(\mathbf{u}) + a_{id}^k + (\mathbf{b}_{id}^k)^T \cdot \mathbf{v}_{id} + (\mathbf{v}_{id})^T C_{id}^k \cdot \mathbf{v}_{id} \\ \mathbf{v}_{id} = \mathbf{v}_{id}(\mathbf{u}) \\ \mathbf{g}_{id}(\mathbf{y}, \mathbf{u}) \leq 0 \\ c_d = \gamma_{id} \\ \boldsymbol{\Omega}(\mathbf{z}) = \text{verdadero} \\ \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \\ z_{id} \in \{\text{verdadero}, \text{falso}\} \forall i \in N_d, d \in D \end{array} \right], d \in D \end{array} \quad (4-25)$$

#### 4.4.3. Adaptación por modificadores de funciones definidas a trozos

Si la función definida a trozos de la ecuación (4-2) tiene continuidad  $C^0$ , es deseable que la estrategia de adaptación mantenga esta propiedad. En el caso más simple, una estrategia de adaptación de restricciones (sin corrección del gradiente) debería agregar el mismo *bias* a todos los trozos de la función. Si se incluye un modificador del gradiente en el trozo actual (es decir, el trozo correspondiente al punto operativo actual), se presentan dos opciones: la primera es usar el mismo valor del modificador del gradiente y del modificador de la restricción en todos los

trozos; la segunda es mantener el modificador del gradiente de cada trozo y corregir el modificador de la restricción para mantener la continuidad  $C^0$ . Esta última opción se desarrolla a continuación.

Suponiendo que la función está definida para una sola entrada  $u$ , la condición para continuidad  $C^0$  en el límite entre dos trozos es:

$$g_p(u_{p+1}^L) + \beta_p^k + \lambda_p^k \cdot (u_{p+1}^L - u_p^{ref}) = g_{p+1}(u_{p+1}^L) + \beta_{p+1}^k + \lambda_{p+1}^k \cdot (u_{p+1}^L - u_{p+1}^{ref}) \quad (4-26)$$

Si el modelo original ya es continuo  $C^0$ :

$$g_p(u_{p+1}^L) = g_{p+1}(u_{p+1}^L) \quad (4-27)$$

Por lo tanto, (4-26) implica:

$$\beta_p^k + \lambda_p^k \cdot (u_{p+1}^L - u_p^{ref}) = \beta_{p+1}^k + \lambda_{p+1}^k \cdot (u_{p+1}^L - u_{p+1}^{ref}) \quad (4-28)$$

Una vez que se han calculado los modificadores  $\beta_{p_{act}^k}^k$  y  $\lambda_{p_{act}^k}^k$  para el trozo activo  $p_{act}^k$  a través de (4-13), los modificadores para los otros trozos pueden actualizarse a través del siguiente procedimiento:

$$\begin{aligned} & \text{FOR } p := (p_{act}^k - 1) \text{ to } 1 \\ & \quad \lambda_p^k = \lambda_p^{k-1} \\ & \quad \beta_p^k = \beta_{p+1}^k + \lambda_{p+1}^k \cdot (u_{p+1}^L - u_{p+1}^{ref}) - \lambda_p^k \cdot (u_{p+1}^L - u_p^{ref}) \\ & \text{LOOP} \\ & \text{FOR } p := (p_{act}^k + 1) \text{ to } np \\ & \quad \lambda_p^k = \lambda_p^{k-1} \\ & \quad \beta_p^k = \beta_{p-1}^k + \lambda_{p-1}^k \cdot (u_p^L - u_{p-1}^{ref}) - \lambda_p^k \cdot (u_{p+1}^L - u_p^{ref}) \\ & \text{LOOP} \end{aligned} \quad (4-29)$$

Si la adaptación utiliza las variables auxiliares  $\mathbf{v}$ , es posible definir la función a trozos como:

$$\left[ \begin{array}{l} z_{d,1} \\ g = g_1(\mathbf{v}(\mathbf{u})) \leq 0 \\ v_1^L \leq v \leq v_2^L \end{array} \right] \vee \dots \vee \left[ \begin{array}{l} z_{d,p} \\ g = g_p(\mathbf{v}(\mathbf{u})) \leq 0 \\ v_p^L \leq v \leq v_{p+1}^L \end{array} \right] \vee \dots \vee \left[ \begin{array}{l} z_{d,np} \\ g = g_{np}(\mathbf{v}(\mathbf{u})) \leq 0 \\ v_{np}^L \leq v \leq v_{np+1}^L \end{array} \right] \quad (4-30)$$

En este caso, si se puede asegurar que la estructura de la función modelada a trozos es función de la variable unidimensional  $v$ , es posible adaptar esta disyunción con modificadores basados en esta variable. Si, en cambio, fuera

necesario corregir los gradientes en todas las direcciones de  $\mathbf{u}$ , se obtendría una formulación más compleja de la función definida a trozos, que dificultaría mantener la continuidad en los puntos que conectan cada tramo de la función.

Si se utiliza una regresión lineal como (4-23), sin coeficientes cuadráticos, la condición para la continuidad  $C^0$  en la unión de dos tramos es:

$$g_p(v_{p+1}^L) + a_p^k + b_p^k \cdot v_{p+1}^L = g_{p+1}(v_{p+1}^L) + a_{p+1}^k + b_{p+1}^k \cdot v_{p+1}^L \quad (4-31)$$

Usando (4-27) se obtiene:

$$a_p^k + b_p^k \cdot v_{p+1}^L = a_{p+1}^k + b_{p+1}^k \cdot v_{p+1}^L \quad (4-32)$$

y la estrategia de (4-29) puede reescribirse como:

$$\begin{aligned} & \text{FOR } p := (p_{act}^k - 1) \text{ to } 1 \\ & \quad b_p^k = b_p^{k-1} \\ & \quad a_p^k = a_{p+1}^k + (b_{p+1}^k - b_p^k) \cdot v_{p+1}^L \\ & \text{LOOP} \\ & \text{FOR } p := (p_{act}^k + 1) \text{ to } np \\ & \quad b_p^k = b_p^{k-1} \\ & \quad a_p^k = a_{p-1}^k + (b_{p-1}^k - b_p^k) \cdot v_p^L \\ & \text{LOOP} \end{aligned} \quad (4-33)$$

#### 4.5. REVISIÓN Y REFORMULACIÓN DE PRÁCTICAS DE RTO CONTINUA

Las estrategias de adaptación presentadas en la sección 4.4 generan un modelo que sólo es confiable en las cercanías del punto operativo actual (definido por el vector de entradas  $\mathbf{u}^k$ ). Si el nuevo punto  $\mathbf{u}^{k+1}$  se encuentra en una región en la que el modelo adaptado no es capaz de predecir satisfactoriamente las salidas de la planta, la aplicación de las entradas óptimas en la planta puede resultar en violación de restricciones o en una desmejora de la función objetivo.

En la práctica, muchas implementaciones de RTO aplican una estrategia para mantener el siguiente punto operativo  $\mathbf{u}^{k+1}$  en las cercanías de  $\mathbf{u}^k$ . En algunos casos, se modifica el problema de optimización incluyendo restricciones que limitan el cambio máximo de las variables de entrada (sección 2.9.1) o penalizan este cambio en la función objetivo (sección 2.3.2). Otro enfoque resuelve un problema de optimización, calculando las entradas óptimas  $\mathbf{u}^*$  y luego efectúa un *filtrado* de estas entradas: el nuevo vector  $\mathbf{u}^{k+1}$  se obtiene a partir del

punto anterior, efectuando un paso en la dirección del vector  $(\mathbf{u}^* - \mathbf{u}^k)$  (sección 2.7.2).

Estas estrategias no pueden aplicarse en su forma original a un sistema de RTO con disyunciones, porque pueden conducir a puntos infactibles o subóptimos. Las siguientes secciones ilustran los problemas mencionados, y proponen una extensión de las estrategias indicadas en el párrafo anterior que permite su implementación en sistemas con decisiones discretas.

#### 4.5.1. Restricción de máximo cambio o penalización del cambio en las entradas

Una práctica común en RTO es limitar el cambio en las entradas agregando una restricción de máximo cambio en el problema de optimización (Darby et al., 2011). La forma usual de esta restricción se expresa en la ecuación (2-77)

Otras restricciones que limitan el cambio en las entradas son las propuestas en algunas implementaciones de control dual (Marchetti et al., 2010; Rodger, 2010) para acotar el error por ruido y truncamiento en la estimación del gradiente.

Brdýs y Tatjewski (2005) proponen la inclusión de un *término de convexificación* en la función objetivo, que penaliza los cambios con respecto al punto actual (y, adicionalmente, consigue que la función objetivo sea estrictamente convexa si el resto de los términos son lineales). El término de convexificación  $C_T$  se define como:

$$C_T = \rho \cdot \|\mathbf{u} - \mathbf{u}^k\|^2 \quad (4-34)$$

donde  $\rho$  es un parámetro que se selecciona para cada implementación. Valores más pequeños de  $\rho$  permiten cambios más grandes en las entradas, mientras que valores más altos de  $\rho$  generan resultados más robustos pero una convergencia más lenta del sistema.

Si el modelo de RTO incluye disyunciones que involucran costos fijos o regiones factibles desconectadas, estas estrategias pueden afectar la convergencia al óptimo real de la planta (e incluso al óptimo del modelo). Este comportamiento puede ilustrarse con un ejemplo sencillo, que consiste en resolver el siguiente problema de optimización:

$$\begin{aligned}
 \min f &= 10 - 0.4x + \gamma \\
 & \text{s.t.} \\
 & \left[ \begin{array}{c} z \\ x \geq 5 \\ \gamma = 3.5 + 0.05x \end{array} \right] \vee \left[ \begin{array}{c} -z \\ x \leq 5 \\ \gamma = 0 \end{array} \right] \\
 & 2 \leq x \leq 10
 \end{aligned} \tag{4-35}$$

Se asume que (4-35) es un modelo perfecto, es decir que describe el sistema a optimizar sin error estructural ni paramétrico. De esta manera, no hay necesidad de adaptación. El punto inicial es  $x^0 = 8$ .

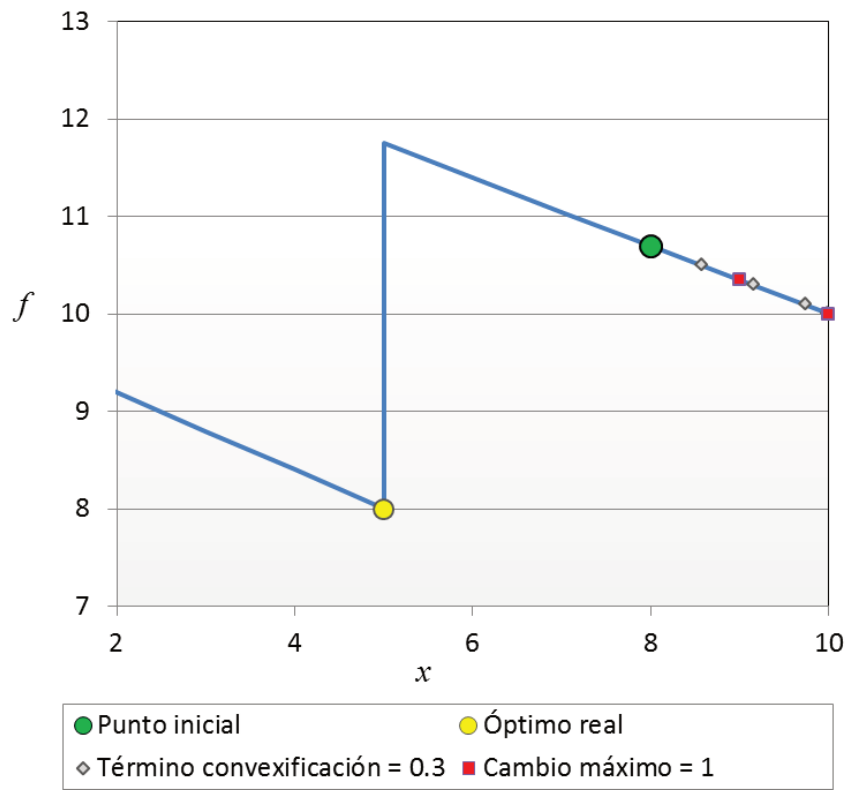
Como se propone en (2-77), se incluye una restricción al máximo cambio permitido:

$$-1 \leq x - x^k \leq 1 \tag{4-36}$$

En forma alternativa, se agrega un término de convexificación a la función objetivo, según (4-34):

$$C_T = 0.3(x - x^k)^2 \tag{4-37}$$

La solución óptima de (4-35) es  $(x=5, f=8)$ . Puede decirse que un "sistema de RTO" con el modelo de (4-35) converge al óptimo en un ciclo (ya que no hay error en el modelo, y los resultados no se filtran). Si se agrega (4-36) al problema, el sistema converge en dos ciclos de RTO a un óptimo local  $(x=10, f=10)$ . Si, en cambio, se incluye (4-37) en la función objetivo, el sistema converge en 4 ciclos al mismo óptimo local. Las evoluciones se muestran en la **Figura 4-1**.



**Figura 4-1.** Convergencia utilizando restricciones de cambio máximo y términos de convexificación.

Para evitar estas soluciones subóptimas, la restricción de máximo cambio o la penalidad en la función objetivo pueden incluirse sólo si no cambia el término activo de la disyunción (es decir, si  $z = z^k$ ). Por su parte, el cambio del término activo puede penalizarse con un costo fijo  $\pi$ , de manera que el cambio sólo ocurra cuando el modelo predice un beneficio significativo. Por ejemplo, para la siguiente disyunción puede escribirse para el término de convexificación:

$$\left[ \begin{array}{l} z \\ \mathbf{g}_1(\mathbf{u}) \leq 0 \\ \text{IF } z^k = \text{verdadero THEN} \\ C_T = \rho \|\mathbf{u} - \mathbf{u}^k\|^2 \\ \text{ELSE} \\ C_T = \pi_1 \\ \text{END IF} \end{array} \right] \vee \left[ \begin{array}{l} -z \\ \mathbf{g}_2(\mathbf{u}) \leq 0 \\ \text{IF } z^k = \text{verdadero THEN} \\ C_T = \pi_2 \\ \text{ELSE} \\ C_T = \rho \|\mathbf{u} - \mathbf{u}^k\|^2 \\ \text{END IF} \end{array} \right] \quad (4-38)$$

donde  $\mathbf{g}_1$  y  $\mathbf{g}_2$  son las restricciones válidas cuando  $z^k = \text{verdadero}$  y cuando  $z^k = \text{falso}$ , respectivamente.  $\pi_1$  es la penalidad por cambiar de *verdadero* a *falso*,

y  $\pi_2$  la penalidad por el cambio inverso. Puede observarse que para cada término de la disyunción, el uso del término de convexificación o de la penalidad no se determina durante la resolución del problema de optimización, sino que es una decisión previa basada en el valor actual de  $z^k$ ; por este motivo, no requiere de variables discretas adicionales.

Si al ejemplo de la ecuación (4-35) se agrega el término de convexificación de (4-37) a través de la disyunción (4-38), el sistema de RTO converge al óptimo en una iteración si las penalidades  $\pi_1$  y  $\pi_2$  son menores a 2; para penalidades mayores, converge al óptimo local.

La formulación equivalente a (4-38) para la restricción de cambio máximo se obtiene incluyendo la restricción en la siguiente disyunción:

$$\left[ \begin{array}{c} z \\ \mathbf{g}_1(\mathbf{u}) \leq 0 \\ \text{IF } z^k = \text{verdadero THEN} \\ \mathbf{u}^k - \Delta \mathbf{u}_{\max} \leq \mathbf{u} \leq \mathbf{u}^k + \Delta \mathbf{u}_{\max} \\ \text{ELSE} \\ C_T = \pi_1 \\ \text{END IF} \end{array} \right] \vee \left[ \begin{array}{c} -z \\ \mathbf{g}_2(\mathbf{u}) \leq 0 \\ \text{IF } z^k = \text{verdadero THEN} \\ C_T = \pi_2 \\ \text{ELSE} \\ \mathbf{u}^k - \Delta \mathbf{u}_{\max} \leq \mathbf{u} \leq \mathbf{u}^k + \Delta \mathbf{u}_{\max} \\ \text{END IF} \end{array} \right] \quad (4-39)$$

#### 4.5.2. Filtrado de las entradas óptimas

Como se ha mencionado, otra práctica común en RTO es el filtrado de las entradas. Los resultados óptimos  $\mathbf{u}^*$  obtenidos de la solución del problema de optimización en el ciclo  $k$  se filtran con el propósito de generar una evolución más robusta y facilitar la convergencia al óptimo real de la planta:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + K \cdot (\mathbf{u}^* - \mathbf{u}^k) \quad 0 \leq K \leq 1 \quad (4-40)$$

Las estrategias para generar una ganancia de filtrado  $K$  que garantice la factibilidad y la mejora en la función objetivo se han analizado en la sección 2.7.2.

En el caso de RTO con disyunciones, el filtrado de las entradas puede causar soluciones subóptimas o infactibles. Este comportamiento indeseado puede aparecer cuando las disyunciones incluyen costos fijos o generan regiones factibles desconectadas, como se analiza en los siguientes párrafos.

En problemas con regiones factibles desconectadas, el filtrado de las entradas puede llevar a puntos fuera de la región factible. Para ilustrar este comportamiento, se considera el siguiente problema:

$$\begin{aligned}
 \min f &= (x-4)^2 \\
 \text{s.t.} & \\
 & \left[ \begin{array}{c} z \\ x \leq 4.5 \end{array} \right] \vee \left[ \begin{array}{c} -z \\ x \geq 7 \end{array} \right] \\
 & 3 \leq x \leq 8
 \end{aligned} \tag{4-41}$$

Se selecciona un punto inicial  $x^0 = 8$  y una ganancia para el filtrado de las entradas  $K = 0.4$ . El óptimo del problema es el mínimo de  $f$  en el problema no restringido,  $x = 4$ . En la evolución hacia este punto óptimo, el filtrado de las entradas ocasiona cuatro puntos intermedios que están en la zona no factible ( $4.5 < x < 7$ ), como puede observarse en la **Figura 4-2**.

Si las restricciones son *duras* (no se toleran violaciones temporarias), el filtrado de las entradas no debe permitir estos puntos infactibles. Considerando una disyunción que genera regiones factibles desconectadas:

$$\left[ \begin{array}{c} z_d \\ \mathbf{g}_{1d}(\mathbf{u}) \leq 0 \\ c_d = \gamma_d \end{array} \right] \vee \left[ \begin{array}{c} -z_d \\ \mathbf{g}_{2d}(\mathbf{u}) \leq 0 \\ c_d = 0 \end{array} \right] \tag{4-42}$$

se pueden obtener límites para el filtro de las entradas en el caso de restricciones convexas:

$$\left. \begin{aligned}
 K_{U,d}^{\max} &= \operatorname{argmax}_K K \\
 \text{s.t.} \quad & \mathbf{u}^{k+1} = \mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k) \\
 & \mathbf{g}_{1d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{si } (z_d^k = \text{verdadero}) \wedge (z_d^* = \text{falso}) \\
 & \mathbf{g}_{2d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{si } (z_d^k = \text{falso}) \wedge (z_d^* = \text{verdadero})
 \end{aligned} \right\} \forall d \in D_U \tag{4-43}$$

$$\left. \begin{aligned}
 K_{U,d}^{\min} &= \operatorname{argmin}_K K \\
 \text{s.t.} \quad & \mathbf{u}^{k+1} = \mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k) \\
 & \mathbf{g}_{1d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{si } (z_d^k = \text{falso}) \wedge (z_d^* = \text{verdadero}) \\
 & \mathbf{g}_{2d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{si } (z_d^k = \text{verdadero}) \wedge (z_d^* = \text{falso})
 \end{aligned} \right\} \forall d \in D_U \tag{4-44}$$

donde  $D_U$  es el subconjunto de las disyunciones que generan regiones factibles desconectadas. Puede verse que las restricciones de (4-43) implican que  $z_d = z_d^k$ , mientras que las de (4-44) implican  $z_d = -z_d^k$ .

Para  $z_d^k = \text{verdadero}$  y  $z_d^* = \text{falso}$ :

$$\begin{aligned} \mathbf{g}_{1d}(\mathbf{u}^k) &\leq 0 \\ \mathbf{g}_{1d}\left(\mathbf{u}^k + K_{U,d}^{\max}(\mathbf{u}^* - \mathbf{u}^k)\right) &\leq 0 \end{aligned} \quad (4-45)$$

Como  $\mathbf{g}_{1d}(\mathbf{u})$  es convexa, (4-45) implica que

$$\mathbf{g}_{1d}\left(\mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k)\right) \leq 0 \quad \forall 0 \leq K \leq K_{U,d}^{\max} \quad (4-46)$$

Al mismo tiempo,

$$\begin{aligned} \mathbf{g}_{2d}(\mathbf{u}^*) &\leq 0 \\ \mathbf{g}_{2d}\left(\mathbf{u}^k + K_{U,d}^{\min}(\mathbf{u}^* - \mathbf{u}^k)\right) &\leq 0 \end{aligned} \quad (4-47)$$

que, como  $\mathbf{g}_{2d}(\mathbf{u})$  es convexa, implica que:

$$\mathbf{g}_{2d}\left(\mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k)\right) \leq 0 \quad \forall K_{U,d}^{\min} \leq K \leq 1 \quad (4-48)$$

Una prueba similar puede obtenerse para  $z_d^k = falso$  y  $z_d^* = verdadero$  para mostrar que todos los valores de  $K$  mayores que  $K_{U,d}^{\min}$  o menores que  $K_{U,d}^{\max}$  generan puntos factibles para las restricciones contenidas en la disyunción  $d$  de (4-42).

Como resultado de (4-43) y (4-44), puede obtenerse la siguiente condición para  $K$ , para cada disyunción  $d$ :

$$\left(0 \leq K \leq K_{U,d}^{\max}\right) \vee \left(K_{U,d}^{\min} \leq K \leq 1\right) \quad (4-49)$$

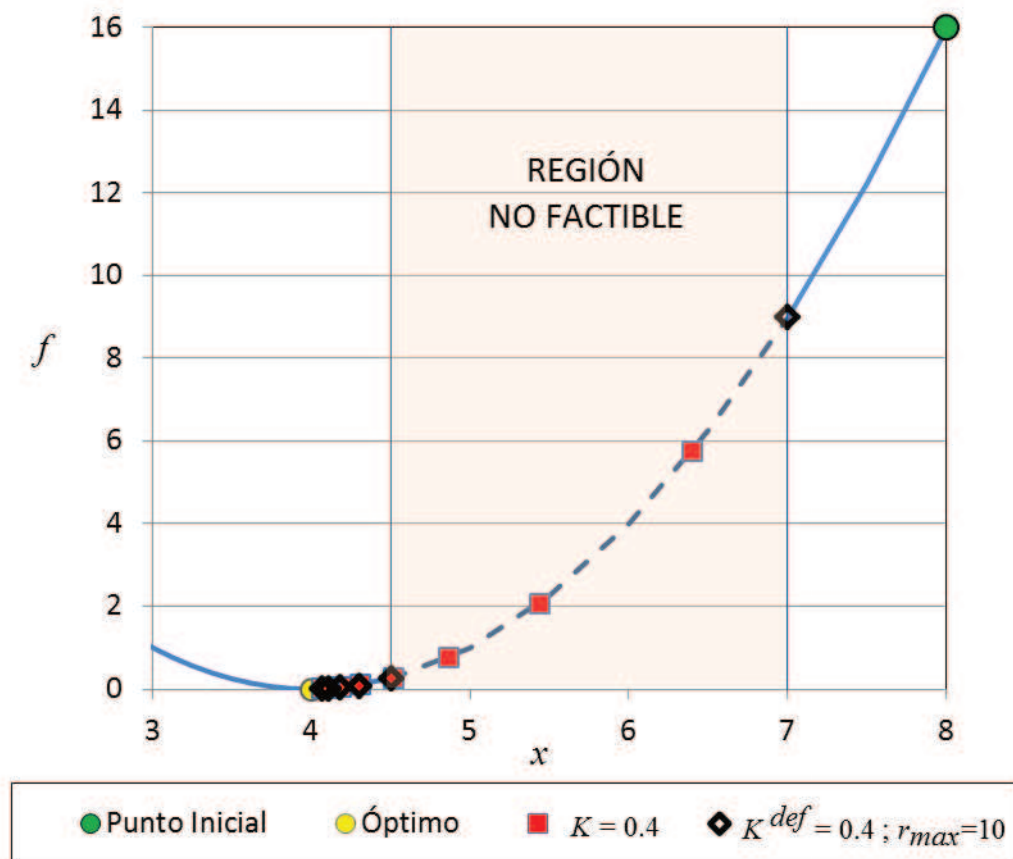
La condición (4-49) genera dos rangos posibles para  $K$ . Si se selecciona siempre la opción  $\left(K \leq K_{U,d}^{\max}\right)$ , el sistema podría estacionarse en el mismo término de una disyunción, mientras los resultados de la optimización indican que el óptimo podría alcanzarse en otra región operativa. Por el contrario, si se selecciona siempre  $\left(K \geq K_{U,d}^{\min}\right)$ , el resultado podría ser un sistema menos robusto y con excesivos cambios en la región de operación. Un criterio práctico para seleccionar una de las dos opciones es definir un valor máximo  $r_{\max}$  para el cociente  $\left(K_{U,d}^{\min} / K_{U,d}^{\max}\right)$ . Para una sola disyunción  $d$ , el criterio se aplica de la siguiente manera:

$$\begin{aligned} &\text{IF } \left( K_{U,d}^{\min} / K_{U,d}^{\max} \leq r_{\max} \right) \text{ THEN } K \leq K_{U,d}^{\max} \\ &\text{ELSE } K \geq K_{U,d}^{\min} \end{aligned} \tag{4-50}$$

La fórmula (4-50) sólo establece una desigualdad, es decir un rango de valores posibles de  $K$ . Si se propone un filtro a utilizar por defecto,  $K^{def}$ , (4-50) puede transformarse en:

$$\begin{aligned} &\text{IF } \left( K_{U,d}^{\min} / K_{U,d}^{\max} \leq r_{\max} \right) \text{ THEN } K = \min(K^{def}, K_{U,d}^{\max}) \\ &\text{ELSE } K = \max(K^{def}, K_{U,d}^{\min}) \end{aligned} \tag{4-51}$$

La evolución para  $r_{\max} = 10$  se muestra en la **Figura 4-2**. Puede verse que el sistema de RTO nunca selecciona un valor de  $x$  en la región no factible.



**Figura 4-2**. Filtrado de las entradas óptimas y regiones factibles desconectadas.

Cuando las disyunciones incluyen costos fijos, un sistema de RTO con filtrado de las entradas puede converger a un punto operativo subóptimo. Un ejemplo de este comportamiento puede observarse al resolver el sistema (4-35) y aplicar los resultados con una ganancia de filtrado  $K < 1$ . La **Figura 4-3** ilustra la evolución obtenida para  $K = 0.6$ . Las entradas óptimas  $x^*$  después de cada ciclo

$k$  son siempre el óptimo real ( $x = 5$ ), pero la estrategia de filtrado no permite que el sistema alcance este punto.

El ejemplo de la **Figura 4-3** converge en forma no deseada al máximo valor posible de la función objetivo ( $f = 11.75$ ). En la práctica, después de un cierto número de iteraciones,  $x^k$  estará suficientemente cerca de  $x = 5$  (por ejemplo,  $x^8 = 5.001$ ) como para asumirse que se ha alcanzado el punto óptimo ( $x = 5$ ,  $f = 8$ ) (aunque rigurosamente función objetivo toma el valor 11.7497). Incluso en este caso se observa otro comportamiento indeseado: la evolución seguida hasta alcanzar el óptimo causa una desmejora continua en la función objetivo después de cada ciclo de RTO, por lo que el sistema podría operar a valores de la función objetivo mayores que el del punto inicial durante varios ciclos (lo que, en el contexto de la optimización en tiempo real, podría implicar varias horas o más de un día).

Una posible solución para este problema es definir un mínimo cambio aceptable,  $\delta_d$ , para las entradas. Este cambio mínimo se aplicará cuando el valor óptimo de la variable booleana asociada a una disyunción  $d$ ,  $z_d^*$ , sea diferente al valor actual de esta variable,  $z_d^k$ . Si el filtrado de las entradas genera un paso  $\|K(\mathbf{u}^* - \mathbf{u}^k)\|$  que es menor que el mínimo cambio aceptado,  $\delta_d$ , entonces se obligará a la variable  $z_d$  a cambiar de valor.

En primer lugar, para cada disyunción escrita como (4-42) (pero que no necesariamente genera regiones factibles desconectadas) el valor máximo de la constante de filtrado  $K$  que hace que  $z_d^{k+1} = z_d^k$  se calcula como:

$$\begin{aligned}
 & K_d^{\max} = \arg \max_K \\
 & \text{s.t. } \mathbf{u}^{k+1} = \mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k) \\
 & \left. \begin{aligned} & \mathbf{g}_{1d}(\mathbf{u}^{k+1}) \leq 0 \\ & \mathbf{g}_{2d}(\mathbf{u}^{k+1}) \geq 0 \end{aligned} \right\} \text{ si } (z_d^k = \text{verdadero}) \wedge (z_d^* = \text{falso}) \\
 & \left. \begin{aligned} & \mathbf{g}_{2d}(\mathbf{u}^{k+1}) \leq 0 \\ & \mathbf{g}_{1d}(\mathbf{u}^{k+1}) \geq 0 \end{aligned} \right\} \text{ if } (z_d^k = \text{falso}) \wedge (z_d^* = \text{verdadero})
 \end{aligned} \tag{4-52}$$

Si el cambio en  $\mathbf{u}$  para  $K_d^{\max}$  es menor que el mínimo cambio aceptado  $\delta_d$  (elegido como un parámetro de diseño del sistema de RTO), entonces  $z_d$  debe ser igual al óptimo  $z_d^*$ :

$$\left( (z_d^k = -z_d^*) \wedge \left( \|K_d^{\max}(\mathbf{u}^* - \mathbf{u}^k)\| \leq \delta_d \right) \right) \Rightarrow z_d^{k+1} = z_d^* \quad (4-53)$$

Utilizando (4-52) y (4-53), puede calcularse un valor mínimo para  $K$  para cada disyunción  $d$ :

$$\left. \begin{array}{l} K_d^{\min} = \arg \min_K \\ s.t. \quad \mathbf{u}^{k+1} = \mathbf{u}^k + K(\mathbf{u}^* - \mathbf{u}^k) \\ \mathbf{g}_{1d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{if } \left( \|K_d^{\max}(\mathbf{u}^* - \mathbf{u}^k)\| \leq \delta_d \right) \wedge (\neg z_d^k) \wedge (z_d^*) \\ \mathbf{g}_{2d}(\mathbf{u}^{k+1}) \leq 0 \quad \text{if } \left( \|K_d^{\max}(\mathbf{u}^* - \mathbf{u}^k)\| \leq \delta_d \right) \wedge (z_d^k) \wedge (\neg z_d^*) \end{array} \right\} \forall d \in D \quad (4-54)$$

y, a partir de (4-54), se puede calcular un límite inferior para la constante de filtrado  $K$ :

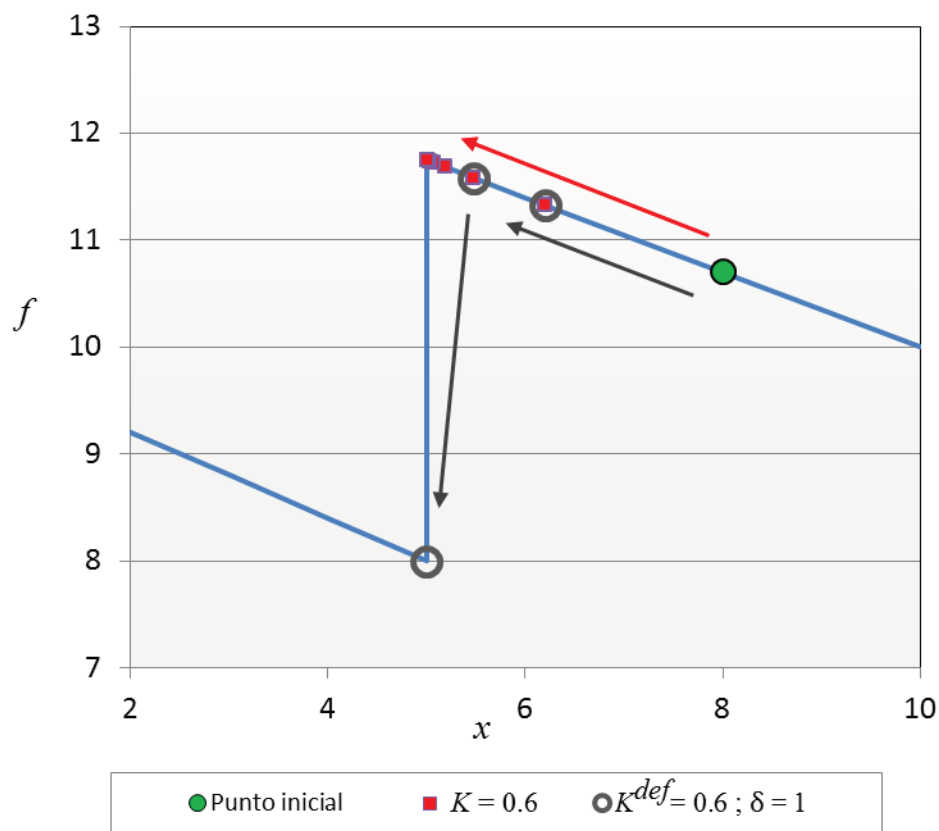
$$K \geq \max_d(K_d^{\min}) \quad (4-55)$$

En (4-54), puede destacarse que si la optimización no propone un cambio en la variable booleana  $z_d$  o si no se cumple el criterio  $\delta_d$ , entonces el valor  $K_d^{\min}$  es 0 para la correspondiente disyunción  $d$ .

Al igual que en (4-51), si se utiliza una ganancia de filtrado por defecto  $K^{def}$ , (4-55) puede transformarse en:

$$K = \max\left(K^{def}, \max_d(K_d^{\min})\right) \quad (4-56)$$

Aplicando el criterio  $\delta_d$  al ejemplo (4-35), el sistema converge al óptimo en 3 ciclos de RTO, para un mínimo cambio  $\delta = 1$ . La evolución se muestra en la **Figura 4-3**.



**Figura 4-3.** Problemas de convergencia y estrategia corregida para el filtrado de las entradas incluyendo costos fijos.

Si se combinan las ecuaciones (4-49) y (4-55), es posible obtener el conjunto  $V_K$  de posibles valores de  $K$  :

$$V_K = \left\{ K / \bigwedge_{d \in D} (K \geq K_d^{\min}) \bigwedge_{d \in D_U} ((K \leq K_{U,d}^{\max}) \vee (K \geq K_{U,d}^{\min})) \right\} \quad (4-57)$$

El criterio presentado en (4-50), formulado para una sola disyunción, no puede aplicarse en forma independiente si hay múltiples disyunciones presentes que generan regiones factibles desconectadas. Para el conjunto  $V_K$ , se propone el siguiente algoritmo para encontrar el límite inferior  $L$  y el límite superior  $U$  para la constante de filtrado:

```

U := 1
L := max_d(K_d^min)

FOR EACH d ∈ D_U
  IF K_{U,d}^min / K_{U,d}^max > r_max THEN L := max(L, K_{U,d}^min)
ENDFOR

REPEAT
  count := 0
  FOR EACH d ∈ D_U / (K_{U,d}^min / K_{U,d}^max ≤ r_max)
    IF K_{U,d}^max ≤ L THEN
      L := max(L, K_{U,d}^min)
      count := count + 1
    ENDIF
  ENDFOR
UNTIL count = 0

FOR EACH d ∈ D_U / ((K_{U,d}^min / K_{U,d}^max ≤ r_max) ∧ (K_{U,d}^max ≥ L))
  U := min(U, K_{U,d}^max)
ENDFOR

```

(4-58)

que resulta en las cotas superior e inferior para  $K$  :

$$L \leq K \leq U \quad (4-59)$$

Los límites de (4-59) podrían estar en conflicto con otros límites para el filtrado de las entradas óptimas que se han formulado en publicaciones recientes para RTO continua (Bunin et al., 2011, 2013a). Además, la presencia de restricciones lógicas entre las variables booleanas (ecuación (4-1)) puede reducir aún más el tamaño del conjunto  $V_K$ . Por estos motivos, el uso del filtrado de las entradas óptimas para RTO con disyunciones presenta aspectos no resueltos que deben investigarse en detalle.

## 4.6. CASOS DE ESTUDIO

La aplicación de la optimización en tiempo real a sistemas incluyendo decisiones discretas se ilustra a través de dos casos de estudio. El primero consiste en la optimización de un sistema genérico que incluye tres procesos modelados

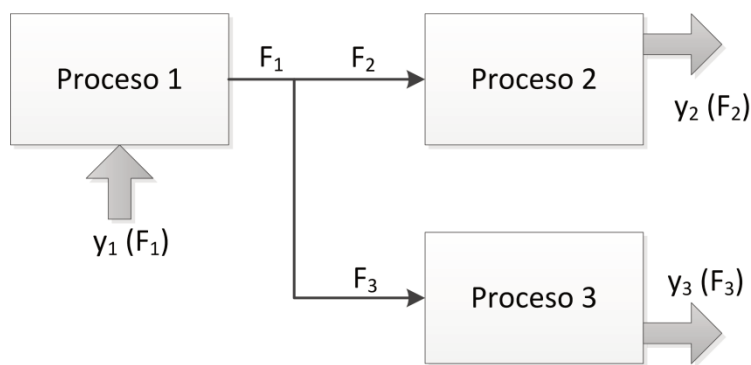
por ecuaciones simples. El segundo implementa la estrategia en un sistema de calor y potencia.

Como es usual en la literatura (Bunin et al., 2011; Yip y Marlin, 2004) y como ya se ha aplicado en el capítulo 3, para analizar el comportamiento de las técnicas de RTO se utilizan dos modelos: uno llamado *planta real*, que se asume como un modelo perfecto del proceso, y el *modelo RTO*, que difiere estructural y paramétricamente de la *planta real*. El uso de cada uno de los modelos se explica en detalle en la sección 3.3.

#### 4.6.1. Caso de estudio 1: Sistema genérico

##### 4.6.1.1. Descripción y modelado del sistema

El sistema a optimizar en este caso de estudio es un proceso genérico formado por tres subprocesos interconectados, con una estructura similar al caso de estudio de la sección 3.3.1. El diagrama del proceso se muestra en la **Figura 4-4**. El vector de entradas  $\mathbf{u}$  está formado por el par de corrientes  $(F_1, F_2)$ , y las salidas medidas son  $\mathbf{y} = (y_1, y_2, y_3)$ .



**Figura 4-4.** Caso de estudio 1. Diagrama del sistema estudiado.

La funcionalidad de las salidas del proceso sólo se conoce en forma aproximada a través del modelo RTO.  $y_2$  se modela como una función lineal definida a trozos.  $F_3$  puede calcularse por balance como la diferencia entre  $F_1$  y  $F_2$ . La funcionalidad real y del modelo RTO para todas las salidas  $\mathbf{y}$  se muestra en la **Tabla 4-1**.

**Tabla 4-1.** Caso de estudio 1. Funcionalidad de las salidas  $y$  para la planta real y el modelo RTO.

$y$	Planta real	Modelo RTO
$y_1$	$5 \cdot \exp(-(F_1 - 80)/18) + 60$	$-18.5 + 0.98 \cdot F_1$
$y_2$	$0.03 \cdot F_2^2 + 0.04 \cdot F_2 - 60$	$-177 + 3.79 \cdot F_2 \quad F_2 \leq 70$ $-250 + 4.84 \cdot F_2 \quad F_2 \geq 70$
$y_3$	$30 \cdot (F_3 - 18)^{0.5} - 20$	$-46.38 + 4.08 \cdot F_3 - 0.001 \times ((F_1 - 110)^2 + (F_2 - 80)^2)$

Si  $F_1$  es mayor que 100, se adiciona un costo  $C$  a la función objetivo:

$$\begin{bmatrix} z_1 \\ F_1 \geq 100 \\ C = 100 + 0.1 \cdot (F_1 - 100) \end{bmatrix} \vee \begin{bmatrix} -z_1 \\ F_1 \leq 100 \\ C = 0 \end{bmatrix} \quad (4-60)$$

La función definida a trozos para  $y_2$  (**Tabla 4-1**) se expresa como otra disyunción:

$$\begin{bmatrix} z_2 \\ F_2 \geq 70 \\ y_2 \leq -250 + 4.84 \cdot F_2 \end{bmatrix} \vee \begin{bmatrix} -z_2 \\ F_2 \leq 70 \\ y_2 \leq -177 + 3.79 \cdot F_2 \end{bmatrix} \quad (4-61)$$

El proceso sólo puede operar para valores de  $F_3$  menores que 33 o mayores a 37:

$$\begin{bmatrix} z_3 \\ F_3 \geq 37 \end{bmatrix} \vee \begin{bmatrix} -z_3 \\ F_3 \leq 33 \end{bmatrix} \quad (4-62)$$

En cada ciclo de RTO  $k$  se incluye una restricción con el fin de generar información para la estimación del gradiente. Definiendo el vector unitario  $\mathbf{d}$ , normal a la recta generada por las dos últimas entradas:

$$\mathbf{d}^k = \frac{\begin{bmatrix} F_2^k - F_2^{k-1} & F_1^{k-1} - F_1^k \end{bmatrix}^T}{\left\| \begin{bmatrix} F_2^k - F_2^{k-1} & F_1^{k-1} - F_1^k \end{bmatrix} \right\|} \quad (4-63)$$

se incluye la siguiente disyunción en el modelo de optimización:

$$\begin{bmatrix} z_4 \\ [\mathbf{u} - \mathbf{u}^k]^T \cdot \mathbf{d}^k \geq 0.2 \end{bmatrix} \vee \begin{bmatrix} -z_4 \\ [\mathbf{u} - \mathbf{u}^k]^T \cdot \mathbf{d}^k \leq -0.2 \end{bmatrix} \quad (4-64)$$

Con el fin de limitar los cambios en las entradas, se incluye un término de convexificación en la función objetivo, como el descrito en 4.5.1, con una constante de penalidad  $\rho = 0.05$ . Este costo de penalización sólo está activo si los valores óptimos de las variables booleanas en las disyunciones 1 y 3 ( $z_1$  y  $z_3$ ) no cambian con respecto al valor actual. En caso contrario, los cambios no se penalizan:

$$\left[ \begin{array}{c} z_5 \\ C_T = 0.05 \cdot \left( (F_1 - F_1^k)^2 + (F_2 - F_2^k)^2 \right) \end{array} \right] \vee \left[ \begin{array}{c} \neg z_5 \\ C_T = 0 \end{array} \right] \quad (4-65)$$

$$\left( (z_1 = z_1^k) \wedge (z_3 = z_3^k) \right) \Rightarrow z_5$$

Se implementaron dos estrategias de adaptación por modificadores aplicadas a este sistema. La primera es la simple pero efectiva adaptación de las restricciones, que no incluye una corrección del gradiente sino únicamente un término constante para corregir el error entre las salidas reales y la predicción del modelo. La segunda implementa la estrategia de adaptación por modificadores completa, con corrección del gradiente, presentada en la sección 2.3.2

La adaptación por modificadores se implementa siguiendo el procedimiento propuesto por Rodger y Chachuat (2011). Las ecuaciones que se adaptan son aquellas que predicen las variables de salida (es decir, las de la **Tabla 4-1**). En primer lugar, se calcula el error entre el modelo RTO y las salidas de la planta real:

$$\begin{aligned} \beta_1^k &= y_{1,real}^k - y_1(\mathbf{u}^k) \\ \beta_2^k &= y_{2,real}^k - y_2(\mathbf{u}^k) \\ \beta_3^k &= y_{3,real}^k - y_3(\mathbf{u}^k) \end{aligned} \quad (4-66)$$

A continuación, para el caso con corrección del gradiente, se estiman los gradientes utilizando el algoritmo recursivo de Broyden:

$$\begin{aligned}
 \mathbf{BR}_i^k &= \mathbf{BR}_i^{k-1} + \frac{y_i^k - y_i^{k-1} - (\mathbf{BR}_i^{k-1})^T (\mathbf{u}^k - \mathbf{u}^{k-1})}{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|^2} (\mathbf{u}^k - \mathbf{u}^{k-1}) \quad i = 1,3 \\
 \left. \begin{aligned}
 \mathbf{BR}_{2,1}^k &= \mathbf{BR}_{2,1}^{k-1} + \frac{y_2^k - y_{2,1}^{ref} - (\mathbf{BR}_{2,1}^{k-1})^T (\mathbf{u}^k - \mathbf{u}_{2,1}^{ref})}{\|\mathbf{u}^k - \mathbf{u}_{2,1}^{ref}\|^2} (\mathbf{u}^k - \mathbf{u}_{2,1}^{ref}) \\
 \mathbf{BR}_{2,2}^k &= \mathbf{BR}_{2,2}^{k-1}
 \end{aligned} \right\} \text{si } z_2 = \text{verdadero} \\
 \left. \begin{aligned}
 \mathbf{BR}_{2,2}^k &= \mathbf{BR}_{2,2}^{k-1} + \frac{y_2^k - y_{2,2}^{ref} - (\mathbf{BR}_{2,2}^{k-1})^T (\mathbf{u}^k - \mathbf{u}_{2,2}^{ref})}{\|\mathbf{u}^k - \mathbf{u}_{2,2}^{ref}\|^2} (\mathbf{u}^k - \mathbf{u}_{2,2}^{ref}) \\
 \mathbf{BR}_{2,1}^k &= \mathbf{BR}_{2,1}^{k-1}
 \end{aligned} \right\} \text{si } z_2 = \text{falso}
 \end{aligned} \tag{4-67}$$

donde  $\mathbf{BR}$  es el estimador del gradiente de Broyden (Mansour y Ellis, 2003; Rodger, 2010). Para la salida  $y_2$ , el gradiente se actualiza sólo para el tramo activo de la función definida a trozos.  $y_{2,1}^{ref}$  y  $\mathbf{u}_{2,1}^{ref}$  son los valores de  $y_2$  y  $\mathbf{u}$  del último ciclo en el que el trozo 1 de la función estuvo activo (es decir, que  $z_2$  fue *verdadero*).  $y_{2,2}^{ref}$  y  $\mathbf{u}_{2,2}^{ref}$  se definen igualmente para la última vez que el trozo 2 estuvo activo. Después de la actualización del gradiente, se actualizan los valores de  $y_2^{ref}$  y  $\mathbf{u}_2^{ref}$  para el tramo actual:

$$\left. \begin{aligned}
 \mathbf{u}_{2,1}^{ref} &= \mathbf{u}^k \\
 y_{2,1}^{ref} &= y_2^k
 \end{aligned} \right\} \text{si } z_2 = \text{verdadero} \\
 \left. \begin{aligned}
 \mathbf{u}_{2,2}^{ref} &= \mathbf{u}^k \\
 y_{2,2}^{ref} &= y_2^k
 \end{aligned} \right\} \text{si } z_2 = \text{falso}
 \end{aligned} \tag{4-68}$$

Utilizando el conocimiento del sistema, algunos elementos de los gradientes pueden acotarse o despreciarse (Bunin et al., 2013b). En este caso de estudio, se sabe al construir el modelo que  $y_1$  es sólo función de  $F_1$ , y por lo tanto la derivada parcial con respecto a  $F_2$  se considera siempre nula. De la misma manera, se sabe que  $y_2$  es sólo función de  $F_2$ , por lo que el gradiente estimado con respecto a  $F_1$  se desprecia en los dos tramos de la función definida a trozos. Además, se supone que existe un conocimiento aproximado de la magnitud de los restantes elementos del gradiente, que permite acotar las derivadas parciales estimadas como:

$$\frac{2}{3} \min_{\mathbf{u}} \left\| \frac{\partial y_p^{real}}{\partial u_i} \right\| \leq \left\| \frac{\partial y_p}{\partial u_i} \right\| \leq \frac{3}{2} \max_{\mathbf{u}} \left\| \frac{\partial y_p^{real}}{\partial u_i} \right\| \quad i = 1, 2; p = 1, 2, 3 \quad (4-69)$$

donde los valores máximos y mínimos se buscan en toda la región factible del problema de RTO.

El paso siguiente es obtener los modificadores del gradiente, de acuerdo con (4-13):

$$\left. \begin{aligned} \lambda_i^k &= \mathbf{BR}_i^k - \nabla_{\mathbf{u}} y_i(\mathbf{u}^k), \quad i = 1, 3 \\ \lambda_{2,1}^k &= \mathbf{BR}_{2,1}^k - \nabla_{\mathbf{u}} y_2(\mathbf{u}^k) \\ \lambda_{2,2}^k &= \lambda_{2,2}^{k-1} \end{aligned} \right\} \text{si } z_2^k = \text{verdadero} \quad (4-70)$$

$$\left. \begin{aligned} \lambda_{2,2}^k &= \mathbf{BR}_{2,2}^k - \nabla_{\mathbf{u}} y_2(\mathbf{u}^k) \\ \lambda_{2,1}^k &= \lambda_{2,1}^{k-1} \end{aligned} \right\} \text{si } z_2^k = \text{falso}$$

A continuación, se aplica un filtro exponencial a los modificadores, de acuerdo con la ecuación (4-11). Las constantes de filtrado (los elementos de la diagonal de la matriz de filtrado) para los modificadores del gradiente  $\lambda$  son todas iguales a 0.7, mientras que para los modificadores de las restricciones son iguales a 1 (no se aplica el filtro).

Para mantener la continuidad  $C^0$  en la función definida a trozos que predice  $y_2$ , se aplica la estrategia propuesta en (4-29).

Una vez que se ha adaptado el modelo, se resuelve el siguiente problema de optimización para encontrar el conjunto de entradas óptimas:

$$\begin{aligned}
 & \min y_1 - y_2 - y_3 + C + C_T \\
 & \text{s.t.} \\
 & \quad F_1 - F_2 - F_3 = 0 \\
 & \quad y_1 = -18.5 + 0.98 \cdot F_1 + \beta_1^k + (\lambda_1^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\
 & \quad y_3 = -46.38 + 4.083 \cdot F_3 + \beta_3^k + (\lambda_3^k)^T \cdot (\mathbf{u} - \mathbf{u}^k) \\
 & \quad \left[ \begin{array}{c} z_1 \\ F_1 \geq 100 \\ C = 100 + 0.1 \cdot (F_1 - 100) \end{array} \right] \vee \left[ \begin{array}{c} \neg z_1 \\ F_1 \leq 100 \\ C = 0 \end{array} \right] \\
 & \quad \left[ \begin{array}{c} z_{2,1} \\ F_2 \geq 70 \\ y_2 \leq -250 + 4.84 \cdot F_2 + \beta_{2,1}^k + (\lambda_{2,1}^k)^T \cdot (\mathbf{u} - \mathbf{u}_{2,1}^{ref}) \end{array} \right] \vee \\
 & \quad \left[ \begin{array}{c} z_{2,2} \\ F_2 \leq 70 \\ y_2 \leq -177 + 3.79 \cdot F_2 + \beta_{2,2}^k + (\lambda_{2,2}^k)^T \cdot (\mathbf{u} - \mathbf{u}_{2,1}^{ref}) \end{array} \right] \\
 & \quad \left[ \begin{array}{c} z_3 \\ F_3 \geq 37 \end{array} \right] \vee \left[ \begin{array}{c} \neg z_3 \\ F_3 \leq 33 \end{array} \right] \\
 & \quad \left[ \begin{array}{c} z_4 \\ \left[ \begin{array}{c} F_1 - F_1^k \\ F_2 - F_2^k \end{array} \right]^T \cdot \mathbf{d}^k \geq 0.2 \end{array} \right] \vee \left[ \begin{array}{c} \neg z_4 \\ \left[ \begin{array}{c} F_1 - F_1^k \\ F_2 - F_2^k \end{array} \right]^T \cdot \mathbf{d}^k \leq -0.2 \end{array} \right] \\
 & \quad \left[ \begin{array}{c} z_5 \\ C_T = 0.3 \cdot \left( (F_1 - F_1^k)^2 + (F_2 - F_2^k)^2 \right) \end{array} \right] \vee \left[ \begin{array}{c} \neg z_5 \\ C_T = 0 \end{array} \right] \\
 & \quad \left( (z_1 = z_1^k) \wedge (z_3 = z_3^k) \right) \Rightarrow z_5 \\
 & \quad \left[ \begin{array}{c} 80 \\ 55 \leq F_1 \leq 120 \\ 20 \end{array} \right] \leq \left[ \begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array} \right] \leq \left[ \begin{array}{c} 120 \\ 90 \\ 50 \end{array} \right] \\
 & \quad \mathbf{z} \in \{\text{verdadero, falso}\}^5
 \end{aligned} \tag{4-71}$$

Para el caso de adaptación de las restricciones (CA), el problema a resolver se obtiene de (4-71) haciendo 0 los modificadores del gradiente  $\lambda_1^k$ ,  $\lambda_{2,1}^k$ ,  $\lambda_{2,2}^k$  y  $\lambda_3^k$ .

Las entradas óptimas  $(F_1^*, F_2^*)^T$  obtenidas de (4-71) se filtran, con una constante de filtrado por defecto  $K^{def} = 0.6$ :

$$(F_1^{k+1}, F_2^{k+1})^T = (F_1^k, F_2^k)^T + 0.6 \cdot (F_1^* - F_1^k, F_2^* - F_2^k)^T \quad (4-72)$$

La disyunción (4-60) incluye un costo fijo; por lo tanto, un filtro de las entradas fijo podría causar los problemas de convergencia discutidos en la sección 4.5.2 e ilustrados en la **Figura 4-3**. Por este motivo, se aplica la estrategia presentada en las ecuaciones (4-52)-(4-55), fijando el parámetro  $\delta_1 = 4$ .

La disyunción (4-62) genera dos regiones factibles desconectadas, lo que podría llevar a los problemas presentados en la **Figura 4-2**. Para evitar puntos no factibles intermedios, se utiliza la estrategia formulada en las ecuaciones (4-43)-(4-50) y (4-57)-(4-59), seleccionando el parámetro  $r_{max} = 10$ .

El punto  $(F_1^0 = 110, F_2^0 = 80)$  se seleccionó como punto inicial. En este punto, el valor de la función objetivo calculado para la planta real (es decir, con la funcionalidad real de las variables de salida en la **Tabla 4-1**) es -31.65. El óptimo real del sistema es  $(F_1^* = 100, F_2^* = 69.2)$ , y el valor óptimo de la función objetivo es -98.57. Con el fin de proveer información para la estimación inicial del gradiente, el punto inicial se perturba con pequeños cambios en  $F_1$  y  $F_2$ , lo que fija la operación en  $k = 1$  y  $k = 2$  como  $(F_1^1 = 110.5, F_2^1 = 80)$  y  $(F_1^2 = 110, F_2^2 = 80.5)$ , respectivamente. Los estimadores de Broyden se inicializan para  $k = 0$  con los gradientes del modelo RTO.

Se analizan dos escenarios para cada estrategia. En el primero, las salidas se miden *perfectamente*: es decir, sin errores o ruido. En el segundo, se adiciona ruido gaussiano de media 0 y variancia 1.0 a las variables de salida, antes del paso de adaptación del modelo.

#### 4.6.1.2. Resultados y discusión

El modelo RTO y el modelo de la planta se implementaron en GAMS (GAMS Development Corporation, 1997) como problemas MINLP. Las disyunciones se tradujeron al problema mixto entero a través de una formulación tipo big-M. El problema se resolvió utilizando DICOPT, con XPRESS como resolvidor lineal y CONOPT 3 como resolvidor NLP (GAMS Development Corporation, 2008). La disyunción 5 no requirió de una variable binaria para ser modelada ya que la variable  $z_5$  presente en (4-65) y (4-71) se expresó en función de las variables  $z_1$  y  $z_3$ ; cada una de las disyunciones restantes utiliza una variable

binaria. El paso de optimización para la estrategia de adaptación por modificadores (MA-Broyden) involucra 16 ecuaciones y 14 variables (4 binarias). Para la adaptación de restricciones (CA), el modelo MINLP contiene 14 ecuaciones y 13 variables (3 binarias), ya que la disyunción 4 no es necesaria.

El sistema de RTO se resolvió para 40 ciclos. Cada ciclo incluye la adaptación del modelo utilizando *mediciones* de la planta real, optimización y filtrado de las entradas óptimas. Cada ciclo requirió menos de 1 segundo de tiempo de CPU en una computadora de 64 bits, con un procesador de 4 núcleos y 2.20 GHz, y con 8 GB de memoria RAM.

La **Figura 4-5** y la **Figura 4-6** muestran la evolución de las entradas para las estrategias CA y MA-Broyden, para los escenarios con y sin ruido en las mediciones. La **Figura 4-5** muestra la evolución en el espacio de las entradas; se incluyen las curvas de nivel de la función objetivo, y se resaltan en verde las regiones infactibles. La **Figura 4-6**, en cambio, muestra los valores de las entradas vs. el ciclo de RTO  $k$ .

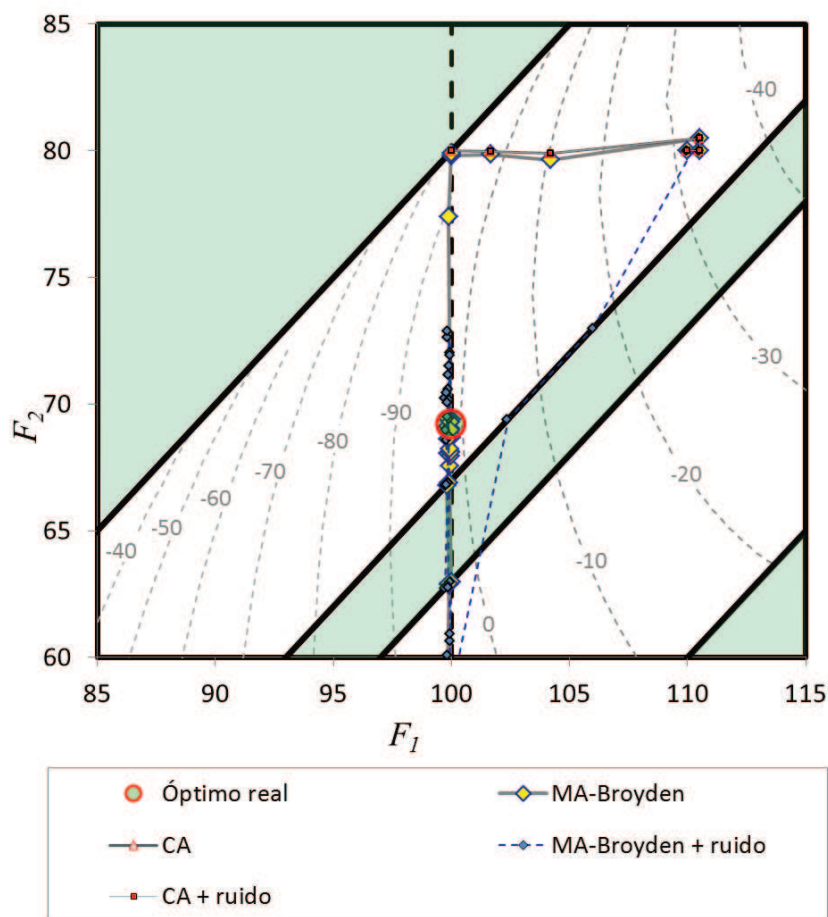
En las dos figuras puede observarse que la adaptación de las restricciones (CA) muestra menos variabilidad en los resultados y converge a un punto que mejora en forma significativa el valor de la función objetivo con respecto al punto inicial. Sin embargo, como no se corrigen los gradientes, no consigue converger a un punto KKT de la planta real. La adaptación por modificadores *completa* (MA-Broyden), en cambio, converge a las proximidades del óptimo; al mismo tiempo, presenta una mayor variabilidad en los resultados debida a los errores en la estimación del gradiente.

Un resultado secundario pero interesante es que la estrategia CA genera la misma evolución en los dos escenarios (con y sin ruido). En realidad, los resultados de la optimización siempre seleccionan el mismo punto ( $F_1 = 100, F_2 = 80$ ), que no se alcanza en un solo ciclo debido al filtrado de las entradas óptimas. Incluso el modelo original sin adaptación (que no se muestra en las figuras) también tiene la misma solución óptima, mostrando que en algunos casos un modelo sin corrección puede identificar las mismas restricciones activas que un modelo adaptado para la optimización en tiempo real.

En el escenario sin ruido, la estrategia MA-Broyden comienza siguiendo una trayectoria similar a la adaptación de restricciones. Después de 6 ciclos de RTO, comienza a evolucionar hacia el óptimo real de la planta, y después de 20 ciclos de RTO, los valores de  $F_1$  y  $F_2$  convergen a las cercanías del óptimo real. En el escenario con ruido gaussiano, esta estrategia evoluciona inicialmente más

rápido hacia el óptimo, pero la entrada  $F_2$  presenta una variabilidad elevada durante todo el período estudiado.

La **Figura 4-7** muestra la evolución de la función objetivo para los 40 ciclos de RTO. Puede observarse que la adaptación de las restricciones converge más rápido, pero el valor final de la función objetivo está distante del óptimo. La adaptación por modificadores requiere más ciclos para converger, pero el valor alcanzado es el óptimo real. Una vez que se alcanza la convergencia, el valor de la función objetivo se mantiene prácticamente constante para el caso sin ruido. En el caso con ruido, a pesar de la alta variabilidad de las entradas mostrada en la **Figura 4-6**, la función objetivo se mantiene en las cercanías del óptimo real a partir del ciclo  $k = 5$ .



**Figura 4-5.** Caso de estudio 1. Evolución de las entradas para adaptación de restricciones (CA) y adaptación por modificadores (MA-Broyden).

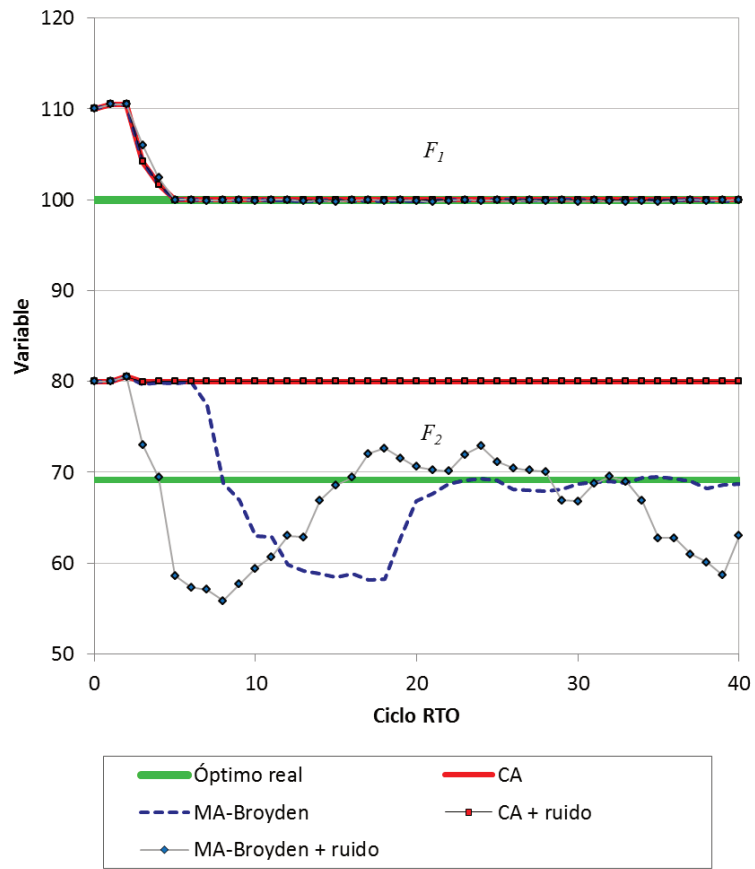


Figura 4-6. Caso de estudio 1. Evolución de las entradas vs. ciclo RTO.

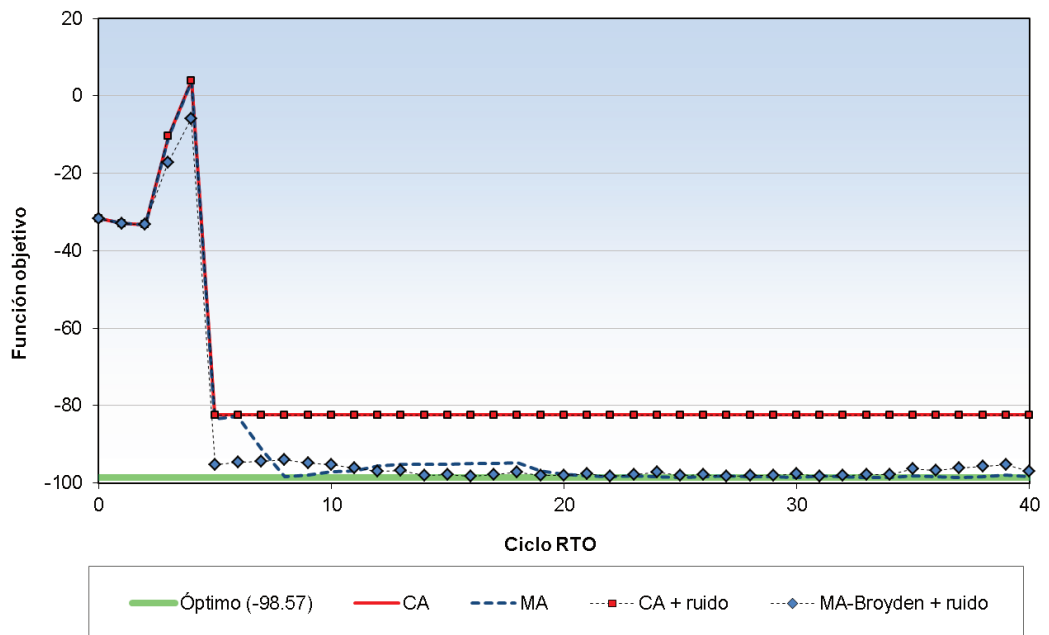


Figura 4-7. Caso de estudio 1. Evolución de la función objetivo vs. ciclo RTO.

La **Tabla 4-2** presenta además el Costo de Diseño relativo  $C_E^{rel}$ , calculado según (3-59).

Puede observarse que el desempeño de la estrategia de adaptación por modificadores (MA-Broyden) es mucho mejor que el de la adaptación de las restricciones (CA), debido a la corrección del error estructural a través de los modificadores del gradiente. De cualquier manera, las dos estrategias consiguen una reducción significativa en el valor de la función objetivo, y por lo tanto ambas alternativas son potencialmente útiles en una implementación de RTO con disyunciones.

Como se mencionó anteriormente, los valores del  $C_E$  para la adaptación de las restricciones son los mismos para los escenarios con y sin ruido, ya que las evoluciones son idénticas.

En el caso de la adaptación por modificadores, el  $C_E$  para  $k = 0$  a 40 para el escenario con ruido gaussiano (394.3) es algo menor que para el escenario sin ruido (427.6). Esto se debe a la evolución inicial más rápida en este caso particular para el escenario con ruido. De cualquier manera, como es esperado, el  $C_E$  para  $k \geq 20$  es mucho mayor (casi tres veces) en el escenario con ruido, debido a los errores en la estimación de los modificadores.

**Tabla 4-2.** Caso de estudio 1. Costo de diseño extendido ( $C_E$ )

Escenarios	k=0 a 40		k=20 a 40	
	$C_E$	$C_E^{rel}$	$C_E$	$C_E^{rel}$
Sin RTO	2676.8	100.0	1338.4	100.0
<b>Sin ruido</b>				
CA	927.7	34.7	338.7	25.3
MA-Broyden	427.6	16.0	6.3	0.5
<b>Ruido gaussiano</b>				
CA	927.7	34.7	338.7	25.3
MA-Broyden	394.3	14.7	23.4	1.7

## 4.6.2. Caso de estudio 2: Sistema de calor y potencia

### 4.6.2.1. Descripción y modelado del sistema

Este caso de estudio considera un sistema combinado de calor y potencia (CHP). El sistema incluye:

- 2 calderas de vapor utilizando un solo combustible
- 1 turbina de gas (GT) con caldera de recuperación de calor (HRSG). La HRSG puede generar vapor adicional quemando combustible en el conducto de humos. Este proceso, conocido como *postcombustión*, tiene una eficiencia alta (cercana al 100%) porque la turbina de gas opera con alto exceso de aire, y este aire está disponible a alta temperatura para la combustión a la entrada de la HRSG.
- 3 turbinas de vapor de contrapresión, utilizadas para mover bombas y compresores
- 1 turbina de extracción-condensación (STG) que genera potencia eléctrica.
- demanda de vapor de baja presión para intercambio de calor.
- 1 desaireador para el tratamiento del agua de calderas.
- válvulas reductoras y de venteo a la atmósfera para control de la presión del vapor.
- demanda de potencia eléctrica.

La **Figura 4-8** muestra un diagrama del sistema. Las ecuaciones detalladas del modelo pueden consultarse en el Apéndice B de esta tesis.

Si se mantienen constantes las temperaturas del vapor en las calderas y las presiones en los colectores de vapor, el sistema tiene 5 grados de libertad. Por lo tanto, se han seleccionado 5 variables independientes (entradas): producción de vapor en la Caldera 2, producción de electricidad en la GT, uso de combustible en la postcombustión de la HRSG, y admisión y extracción de vapor en la STG. En realidad, el sistema de RTO podría usar más de 5 entradas, pero luego la adaptación debe tener en cuenta los grados de libertad reales del sistema, como lo han demostrado Costello et al. (2013).

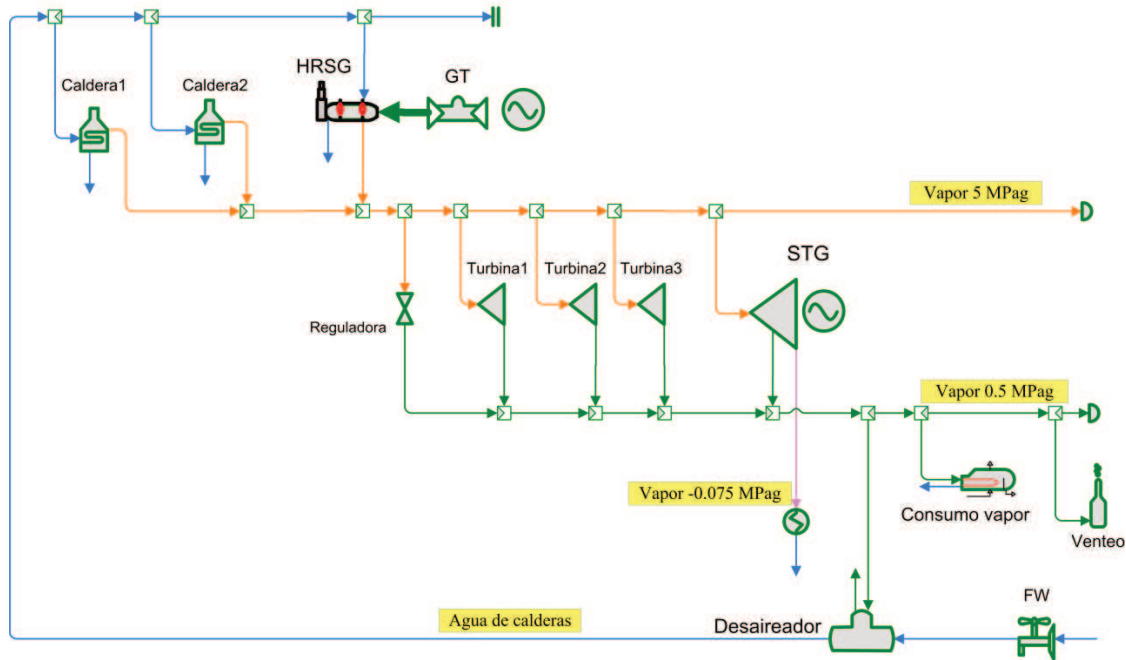


Figura 4-8 . Caso de estudio 2. Diagrama del sistema de calor y potencia.

El sistema presenta tres características que pueden modelarse con disyunciones o variables discretas:

- El uso de combustible en la postcombustión ( $Q_{pc}$ ) puede variar entre 30 y 100 MW, o bien ser igual a 0 MW.

$$\left[ \begin{array}{c} z_1 \\ 30 \leq Q_{pc} \leq 100 \end{array} \right] \vee \left[ \begin{array}{c} \neg z_1 \\ Q_{pc} = 0 \end{array} \right] \quad (4-73)$$

- La compra de electricidad de la red está penalizada si la potencia importada es mayor a 11 MW. La penalidad incluye un término fijo y uno variable:

$$\left[ \begin{array}{c} z_2 \\ D_{ee} - W_{GT} - W_{STG} \geq 11 \\ pen = 2C_{ee} \cdot (1 + (D_{ee} - W_{GT} - W_{STG} - 11)) \end{array} \right] \vee \left[ \begin{array}{c} \neg z_2 \\ D_{ee} - W_{GT} - W_{STG} \leq 11 \\ pen = 0 \end{array} \right] \quad (4-74)$$

donde  $D_{ee}$  es la demanda de electricidad de la planta,  $W_{GT}$  es la potencia generada en la turbina de gas,  $W_{STG}$  es la potencia generada en la turbina de vapor STG,  $C_{ee}$  el precio de compra de la electricidad para valores menores a 11 MW, y  $pen$  la penalidad a pagar.

- El caudal de la válvula reductora de vapor de alta presión a baja presión ( $F_{vl}$ ) puede variar entre 4 y 80 t/h, o bien la válvula está cerrada:

$$\left[ \begin{array}{c} z_3 \\ 4 \leq F_{vl} \leq 80 \end{array} \right] \vee \left[ \begin{array}{c} -z_3 \\ F_{vl} = 0 \end{array} \right] \quad (4-75)$$

La función objetivo a minimizar a través de la optimización en tiempo real es el costo operativo ( $CO$ ), que se formula como:

$$CO = C_F(Q_{b1} + Q_{b2} + Q_{GT} + Q_{db}) + C_W F_{FW} + C_{pow}(D_P - W_{GT} - W_{STG}) + pen \quad (4-76)$$

donde  $C_C$  y  $C_{AD}$  son los costos unitarios del combustible y el agua demineralizada, respectivamente, y  $Q_{c1}$ ,  $Q_{c2}$  y  $Q_{GT}$  son los consumos de combustible en las calderas y la turbina de gas. En el caso de estudio, el costo de combustible está fijo en 7.2 \$/GJ y el del agua en 0.5 \$/t.

La **Tabla 4-3** presenta las diferencias entre la planta real y el modelo RTO. Las diferencias corresponden a *ecuaciones de eficiencia* como las presentadas en la sección 3.1, que no necesariamente corresponden a la eficiencia de un proceso, sino a ecuaciones empíricas que describen el comportamiento de un equipo o una unidad.

**Tabla 4-3.** Caso de estudio 2. Ecuaciones de eficiencia para la planta real y el modelo RTO.

Ecuaciones empíricas	Modelo RTO	Planta real
	$\eta_j = p_m(v_j)$	$\eta_j = p(v_j)$
Eficiencia Caldera 1 (%)	92	$55 + 0.51F_1 - 0.0010(F_1)^2$
Eficiencia Caldera 2 (%)	92	$55 + 1.17F_2 - 0.0010(F_2)^2$
Eficiencia 1° etapa STG (%)	$70 - 0.005(130 - F_3)^2$	$70 - 140/(1 + F_3)$
Eficiencia 2° etapa STG (%)	$70 - 0.004(150 - F_4)^2$	$70 - 100/(1 + F_4)$
Relación caudal/potencia - Turbinas 1-3 (kg/kWh)	12	Eficiencia = 50%
Potencia carga base GT (MW)	$1.7E-6T^3 - 2.72E-4T^2 - 0.24T + 43.68$	$39 + 1.65E-6(T-15)^5 - 2.E-4(T-15)^2 - 2.43(T-15)$
Heat rate a carga base GT (GJ/MWh)	$1.5.10^{-7}T^3 + 9.8.10^{-5}T^2 + 0.012T + 10.59$	$11 + 1.49E-7(T-15)^5 + 1E-4(T-15)^2 + 0.016(T-15)$

$F_1$  y  $F_2$  son las producciones de vapor de las calderas 1 y 2, respectivamente, expresadas en t/h.  $F_3$  es el caudal de admisión de la turbina STG en t/h, y  $F_4$  el caudal de salida al condensador.  $T$  es la temperatura de entrada del aire a la turbina de gas ( $^{\circ}\text{C}$ ), considerada como la temperatura ambiente para este estudio.

El *heat rate* es un indicador de la eficiencia de una turbina de gas, que expresa la relación entre el combustible consumido y la potencia generada. La operación a *carga base* implica generar la mayor potencia posible, controlando la temperatura de los gases a la salida de la turbina (Brooks, 2000).

La adaptación del modelo se implementa de acuerdo con la estrategia descrita en la sección 4.4.2.3, corrigiendo las eficiencias de las calderas y de las dos etapas de la turbina STG. La estrategia utilizada es la adaptación por modificadores con ecuaciones de eficiencia y estimación de los modificadores por regresión (MA-E2). Como caso de comparación, se incluye sólo un término de corrección constante para corregir las ecuaciones de eficiencia (adaptación de las restricciones, CA). Las variables  $\nu$ , de las ecuaciones (3-2)-(3-6) y (4-18), utilizadas para corregir los gradientes se obtienen de la funcionalidad de la planta real especificada en la **Tabla 4-3** ( $F_1$  para la eficiencia de la Caldera 1,  $F_2$  para la Caldera 2,  $F_3$  para la primera etapa de la turbina STG y  $F_4$  para la segunda etapa).

Las eficiencias de las turbinas de contrapresión y la predicción del *heat rate* y la potencia máxima para la turbina de gas se corrigen sólo con un término constante.

Las propiedades del vapor (temperatura y presión de saturación, entalpías y entropías del vapor saturado y sobrecalentado) y del agua se calculan según Irvine y Liley (1984).

En este caso de estudio, las *mediciones* obtenidas del modelo *planta real* son los factores de eficiencia listados en la **Tabla 4-3**. En un sistema real, estos factores podrían obtenerse en base a otras variables medidas. Por ejemplo, la eficiencia de una caldera puede obtenerse a través de la medición de  $\text{O}_2\%$ ,  $\text{CO}_2\%$  y temperatura en los humos, temperatura de entrada del aire y del combustible y poder calorífico del combustible; la eficiencia de una turbina puede calcularse a través de las temperaturas y presiones a la entrada y la salida del equipo.

Para evitar que errores de modelado o perturbaciones en el proceso determinen el pago de la penalidad dada por (4-74), el límite real de 11 MW se reemplaza en el modelo RTO por 10.8 MW. De este modo, el sistema puede tolerar una perturbación en el proceso o un error en las predicciones del modelo que

causen una demanda eléctrica extra de 200 kW, aunque probablemente a expensas de una pérdida de beneficio. Esta estrategia es similar al margen de seguridad en las restricciones presentado en la sección 2.9 (Chachuat et al., 2009), utilizado con frecuencia en RTO para hacer frente a la incertidumbre.

Los valores iniciales de las variables de entrada se resumen en la **Tabla 4-4**.

**Tabla 4-4.** Caso de estudio 2. Variables de entrada iniciales.

Variable de entrada	Valor
Caudal caldera 2 (t/h)	40
Potencia generada GT (MW)	38.9
Caudal de admisión STG (t/h)	132.4
Caudal de extracción STG (t/h)	60
Combustible a postcombustión (MW)	0

Se plantean dos escenarios para este caso de estudio. En el escenario 1, las mediciones están libres de ruido. El precio de la electricidad, la temperatura ambiente y las demandas de potencia y vapor se mantienen constantes; sus valores se listan en la **Tabla 4-5**. El escenario 2 tiene un perfil variable para el precio de la electricidad, la temperatura ambiente y las demandas de electricidad y vapor, que se muestran en la **Figura 4-10a**. Las mediciones se contaminan con ruido gaussiano de media 0 y desviación estándar  $\sigma$ . Los valores de  $\sigma$  se muestran en la **Tabla 4-6**.

**Tabla 4-5.** Caso de estudio 2. Especificaciones para el escenario 1.

Parámetro	Valor
Precio compra electricidad, $C_{ee}$ (\$/MWh)	90
Temperatura ambiente, $T$ (°C)	15
Demanda de vapor, $D_v$ (t/h)	75
Demanda de potencia, $D_{ee}$ (MW)	63

**Tabla 4-6.** Caso de estudio 2. Desviación estándar de las mediciones en el escenario 2.

Factor de eficiencia	$\sigma$
Eficiencia calderas (%)	0.1
Eficiencia turbinas (%)	0.2
Potencia turbina de gas (MW)	0.05
Heat rate turbina de gas (MW/MW)	0.05

#### 4.6.2.2. Resultados y discusión

El modelo RTO y el de la planta real se implementaron en GAMS como problemas MINLP usando una formulación de tipo *big-M*. Para cada escenario, se resolvieron 30 ciclos de RTO. Las regresiones lineales para obtener los modificadores de la estrategia MA-E2 se resolvieron con CPLEX. El problema MINLP de minimización del costo se resolvió con DICOPT, usando CPLEX para el problema MILP *master* y CONOPT 3 para los *subproblemas* NLP (Kocis y Grossmann, 1989). El modelo MINLP tiene 76 ecuaciones y 86 variables (3 discretas). Cada ciclo (obtención de los modificadores, minimización del costo y pasos intermedios) requirió menos de 0.7 s de tiempo de CPU para ser resuelto en una computadora de 64 bits, con un procesador de 4 núcleos y 2.20 GHz, y 8 GB de memoria RAM.

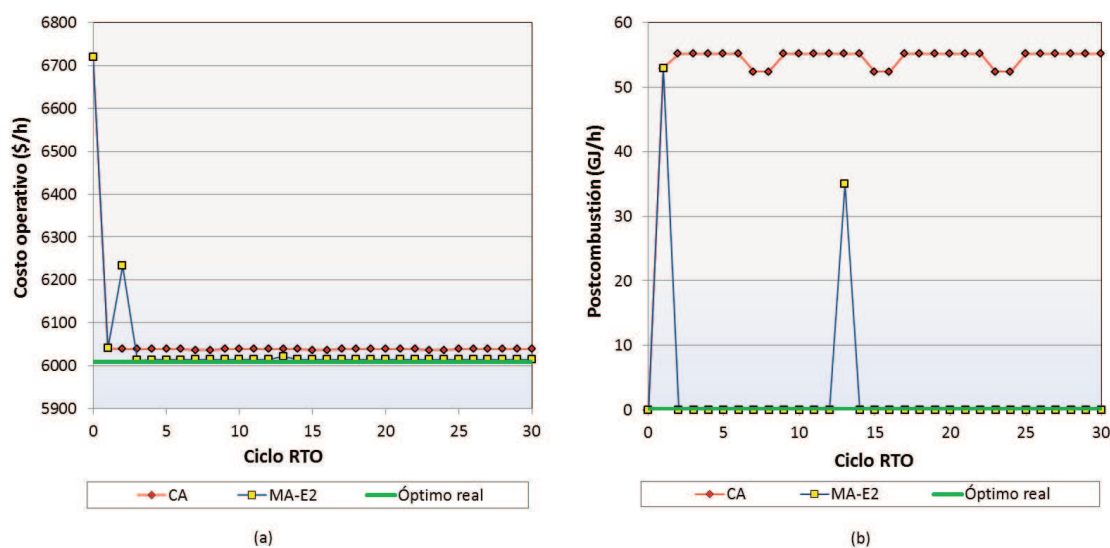
Si bien el análisis de los tiempos de procesamiento está fuera de los alcances de este trabajo, puede observarse que los tiempos de CPU requeridos para este estudio son satisfactorios para una implementación de RTO. Si bien el modelo es pequeño, existen sistemas de calor y potencia (y procesos industriales en general) que pueden representarse con modelos de un tamaño similar al de este caso de estudio. El principal requerimiento de tiempo corresponde a la solución

del problema MINLP de minimización del costo. El estudio de la influencia del tamaño del modelo en los tiempos de CPU requeriría un análisis más detallado, y dependerá fuertemente del algoritmo de solución, de la capacidad de procesamiento de la computadora y de las restricciones del modelo. De cualquier manera, un tiempo de cálculo de 10 minutos o más podría ser aceptable para una implementación de RTO (dependiendo de la frecuencia de los ciclos) por lo que es de esperar que problemas más complejos de RTO con disyunciones puedan ser implementados en la escala industrial. De hecho, como se ha mencionado en la sección 2.12, en la práctica existen implementaciones de RTO que resuelven problemas MINLP (Puranik et al., 2013; Ruiz et al., 2005).

La **Figura 4-9a** muestra la evolución del costo operativo para el escenario 1. Puede observarse que la estrategia MA-E2 converge a un costo prácticamente idéntico al óptimo, mientras que la adaptación de las restricciones (CA), que no corrige los gradientes, converge a un costo mayor. Este comportamiento es el que se observa generalmente en las implementaciones de RTO con modelos NLP, como puede verificarse en los casos de estudio de la sección 3.3.

La **Figura 4-9b** presenta la evolución del uso de combustible en la postcombustión (*PC*). Las dos seleccionan la *PC* en el primer ciclo de RTO ( $k=1$ ). La estrategia CA converge a un punto en el que se quema este combustible adicional, mientras que MA-E2 converge a un punto sin *PC*, en coincidencia con el óptimo real de la planta. En un ciclo intermedio ( $k=13$ ), MA-E2 utiliza la *PC* para detenerla en el ciclo siguiente. Este puede ser un comportamiento indeseado, ya que las puestas en marcha y paradas frecuentes pueden incrementar los costos de mantenimiento del equipo. Para evitarlo, es posible incluir *restricciones de transición* (Mitra et al., 2013) que fijen un número mínimo de ciclos en los que la *PC* debe permanecer en el mismo estado (en operación o detenida). Si estas restricciones no son suficientes para producir resultados satisfactorios en una formulación de un solo período (como la presentada en este capítulo), podría ser necesario recurrir a una formulación multiperíodo.

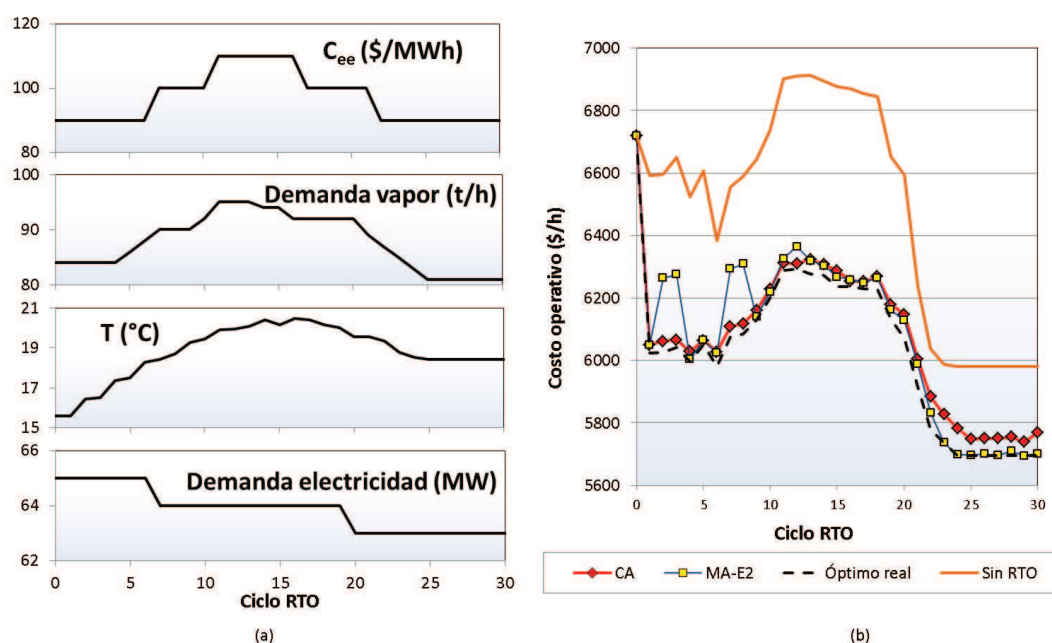
Para las dos estrategias (CA y MA-E2) no se paga penalidad por exceso de consumo de electricidad en ninguno de los ciclos, y la válvula reguladora de (4-75) permanece abierta en todo momento (estos resultados no se ilustran en ninguna figura).



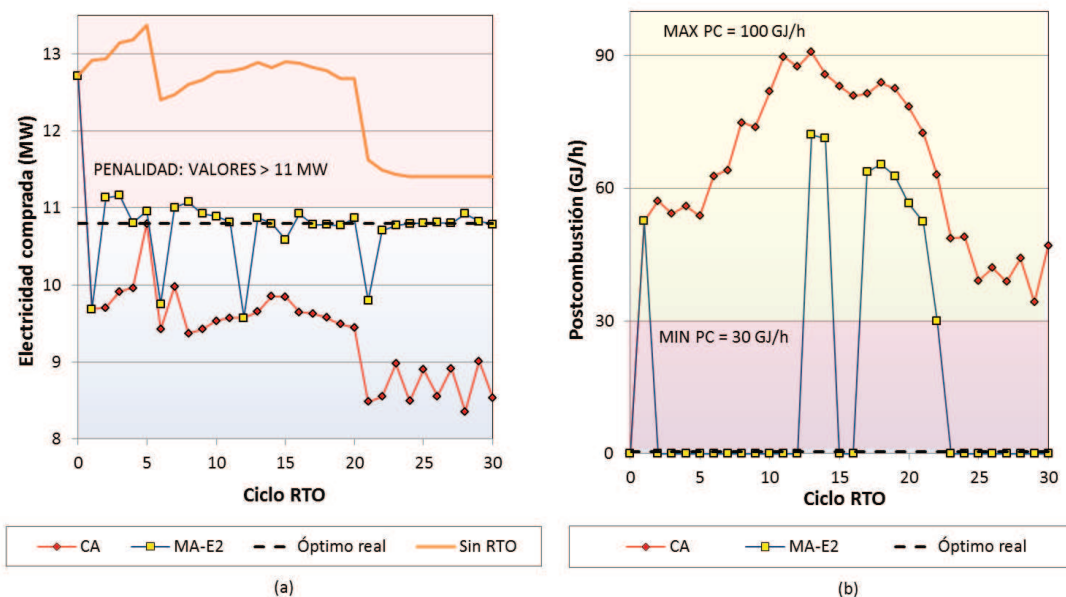
**Figura 4-9.** Caso de estudio 2, escenario 1. (a) Evolución del costo operativo. (b) Evolución del uso de la postcombustión.

La **Figura 4-10b** muestra la evolución del costo operativo para el escenario 2. Puede observarse que las dos estrategias de RTO operan en cercanías del costo óptimo real, en comparación con la evolución sin optimización. A pesar de que en la mayoría de los ciclos la estrategia MA-E2 alcanza un costo menor que CA, en cuatro ciclos ( $k = 2, 3, 7$  y  $8$ ) se observa un costo elevado que afecta el desempeño de la estrategia. La **Figura 4-11a** muestra la razón de estos picos: a pesar de que la adaptación por modificadores consigue operar en cercanías de la compra óptima de electricidad (10.8 MW), los cambios en precios, temperatura ambiente y demandas, sumados a la incertidumbre del modelo, fuerzan al sistema a comprar más potencia en algunos casos. Esto conduce a algunas penalizaciones por compras mayores a 11 MW. La estrategia CA opera con valores más bajos de compra de electricidad y por eso nunca es penalizada. A partir del ciclo 23, cuando las condiciones son más estables, se observan las ventajas de utilizar una estrategia con corrección del gradiente como la MA-E2.

La **Figura 4-11b** muestra la evolución del uso de la postcombustión. MA-E2 decide correctamente operar sin postcombustión en 21 de los 30 ciclos, mientras que CA opera con postcombustión en todo el escenario.



**Figura 4-10.** Caso de estudio 2, escenario 2. (a) Evolución del costo operativo. (b) Perfiles de precio eléctrico, demanda de vapor, temperatura ambiente y demanda de potencia.



**Figura 4-11.** Caso de estudio 2, escenario 2. (a) Evolución de la compra de electricidad. (b) Uso de la postcombustión.

La **Tabla 4-7** muestra el costo de diseño extendido ( $C_E$ ) para las dos estrategias y los dos escenarios, y los compara con los resultados obtenidos cuando no se implementa ninguna estrategia de optimización.

En el escenario 1, ambas estrategias *capturan* más del 95% de los beneficios posibles. El  $C_E$  para  $k \geq 5$  muestra cómo la estrategia MA-E2, gracias a la

corrección de los gradientes, tiene un desempeño mucho mejor que la CA. Sin embargo, gracias a su simplicidad y a los buenos resultados obtenidos, CA también se constituye en una estrategia adecuada para implementar en este sistema.

En el escenario 2, los cambios frecuentes en precios, temperatura ambiente y demandas, junto con el ruido en las mediciones, impactan negativamente en el desempeño de las estrategias de RTO. La adaptación por modificadores MA-E2, a pesar de los buenos resultados obtenidos en situaciones estables (por ejemplo, para ciclos  $k \geq 23$ ) consigue un  $C_E$  peor que la estrategia CA (1578.2 vs. 1362.9), sobre todo por las penalidades en la compra de electricidad. A pesar de que las dos estrategias consiguen una buena reducción en el costo, para este escenario con frecuentes cambios en las condiciones operativas la estrategia de CA se muestra como la mejor opción.

**Tabla 4-7.** Caso de estudio 2. Costo de diseño extendido ( $C_E$ ) y Costo de diseño extendido relativo ( $C_E^{rel}$ )

Escenarios	$k=0$ a 30		$k=5$ a 30	
	$C_E$	$C_E^{rel}$	$C_E$	$C_E^{rel}$
<b>Escenario 1</b>				
Sin RTO	21321.9	100	17768.3	100.0
CA	1212.2	5.7	722.4	4.1
MA-E2	746.4	3.5	127.5	0.7
<b>Escenario 2</b>				
Sin RTO	14004.0	100		
CA	1362.9	9.7		
MA-E2	1578.2	11.3		

## CONCLUSIONES PARCIALES

Este capítulo presentó y discutió una formulación de RTO con disyunciones. Se mostró que en la optimización de procesos industriales pueden existir decisiones discretas de distinta naturaleza y que pueden ser abordadas por la optimización en tiempo real, incluso cuando no se considera la puesta en marcha y parada de equipos.

La estrategia de dos pasos y dos versiones de la estrategia de adaptación por modificadores se reformularon para su aplicación a problemas que involucren variables de decisión discretas. La reformulación incluye la selección de los parámetros/modificadores a actualizar en cada ciclo, el filtrado exponencial y los ciclos pasados que se utilizan para la corrección del gradiente.

Las restricciones que fijan cambios máximos en las variables de entrada y los términos de convexificación pueden causar una convergencia a puntos subóptimos si el problema incluye decisiones discretas, aun en ausencia de errores estructurales. Se presentó una estrategia para evitar este comportamiento indeseado y se la ilustró con ejemplos.

Cuando se utiliza el filtrado de las entradas óptimas y el problema tiene disyunciones, puede originarse una convergencia lenta al óptimo, empeorando la función objetivo en puntos intermedios. Además, si hay regiones factibles desconectadas, el filtro puede llevar a puntos intermedios no factibles. Se propuso una formulación para determinar los límites superior e inferior de las constantes de filtrado que evitan estos inconvenientes. El problema requiere ser investigado con mayor detalle, ya que los límites obtenidos pueden estar en conflicto con otros límites propuestos en la literatura.

Se presentaron dos casos de estudio, que muestran que una implementación práctica de RTO con disyunciones es factible. El primer caso estudia un sistema simple. Se analizaron dos estrategias (adaptación de restricciones y adaptación por modificadores); ambas mostraron un comportamiento similar al observado en RTO con modelos continuos: convergencia al óptimo para los casos con corrección del gradiente, y menor variabilidad en los resultados para la corrección de las restricciones. El segundo caso de estudio analizó la optimización de un sistema de calor y potencia con decisiones discretas; las estrategias analizadas nuevamente mostraron resultados satisfactorios.

Los tiempos de CPU requeridos para resolver un ciclo de RTO (adaptar el modelo y minimizar la función objetivo) fueron muy bajos. Esta es una característica deseada en la RTO; sin embargo, es necesario evaluar el

comportamiento en problemas con mayor número de variables y restricciones. De cualquier manera, el tamaño del problema del caso de estudio 2 puede ser equivalente al de sistemas industriales reales con un número pequeño de equipos.

Si las decisiones discretas abordadas por la RTO incluyen la puesta en marcha y parada de equipos o restricciones que involucran el tiempo, podría ser necesaria una formulación de RTO multiperíodo. La integración de RTO con *scheduling* óptimo mediante una formulación multiperíodo se aborda en el próximo capítulo de esta tesis.

## 5. INTEGRACIÓN DE RTO Y SCHEDULING ÓPTIMO EN SISTEMAS DE ENERGÍA

### RESUMEN

Este capítulo presenta una propuesta para integrar el *scheduling* óptimo con la optimización en tiempo real (*SO+RTO*) para sistemas de energía. En la sección 5.1 se presenta la problemática y se resumen antecedentes. El problema a resolver se formula en la sección 5.2. La sección 5.3 describe algunos fenómenos que requieren una formulación multiperíodo para la RTO: las *restricciones de transición* y las *variables de acumulación*. En la sección 5.4 se describe la metodología completa de integración *SO+RTO*, incluyendo una estrategia para reaccionar a soluciones no factibles del problema de optimización. El método presentado se ilustra a través de un caso de estudio de un sistema de calor y potencia (sección 5.5), que incluye restricciones en las emisiones de NO<sub>x</sub> y analiza un horizonte de 7 días.

### 5.1. INTRODUCCIÓN

La optimización de los sistemas de calor y potencia suele conducir a ahorros significativos en la operación de los procesos industriales. Estos sistemas están sujetos a cambios frecuentes en la demanda de vapor y potencia a lo largo del día, así como a variaciones en los precios de la electricidad (que pueden cambiar incluso en *tiempo real* de acuerdo con la oferta y demanda en mercados desregulados). Las variables de decisión para la optimización pueden ser continuas (producción de vapor en calderas, producción de electricidad en turbinas de gas) y discretas (puesta en marcha y parada de equipos). Las restricciones involucradas pueden incluir límites operativos de los equipos, decisiones operativas (por ejemplo, estrategias de control) y regulaciones ambientales.

Un enfoque posible para la operación óptima de estos sistemas es la *optimización del scheduling*. Mitra et al. (2013) sugieren una formulación MILP multiperíodo que tiene en cuenta diferentes *modos de operación* y las *restricciones de transición* que limitan los cambios de modo de operación. Además, el trabajo de Mitra et al. (2013) incluye una completa revisión de otros enfoques en la optimización de sistemas de calor y potencia. Otros autores presentan la aplicación industrial de una formulación MILP multiperíodo que optimiza el sistema de servicios auxiliares de una refinería, aunque está orientado a la planificación en el mediano plazo (Micheletto et al., 2008). Velasco-García et al. (2011) desarrollaron

una estrategia para determinar la forma óptima de operación de un sistema de calor y potencia para un solo período, que tiene en cuenta costos fijos de operación y costos de transición para las puestas en marcha y paradas, así como la duración del período; para resolver las no linealidades de los balances de energía, este trabajo propone una técnica llamada *successive-MILP*. Todas estas estrategias necesitan un pronóstico de los precios y la demanda, y resuelven el *schedule* óptimo asumiendo que no hay incertidumbre en los pronósticos ni errores de modelado. Por lo tanto, la implementación de los resultados en la planta podría llevar a una operación subóptima o causar violaciones en las restricciones del proceso.

La optimización en tiempo real (RTO), en cambio, hace uso de las mediciones del proceso para corregir los errores de modelado, y actualiza periódicamente los precios y las demandas del proceso (Chachuat et al., 2009). En el capítulo 2 se han mencionado algunas aplicaciones industriales de RTO en la operación de los servicios auxiliares de refinerías (Ruiz et al., 2005; Tani y Matsuo, 2009). En el capítulo 3 se ha presentado una estrategia que aprovecha la estructura particular de los sistemas de calor y potencia, incluyendo un ejemplo del sistema de energía de una planta de azúcar y etanol. Uno de los problemas principales de la RTO para sistemas de calor y potencia es que en general se optimiza un solo estado estacionario, lo que impide incluir en forma apropiada las restricciones de transición (como el tiempos mínimos de operación de un equipo antes de poder pararlo) y las variables de acumulación (tales como inventarios de tanques o emisiones diarias). Además, la gran mayoría de las formulaciones de RTO son de tipo NLP, y en este caso no es posible modelar la puesta en marcha y parada de equipos. En el capítulo 4 se presentó una formulación basada en disyunciones para la RTO en un solo período; en el capítulo 2 se mencionan aplicaciones industriales de RTO con formulaciones MINLP.

En otras áreas de investigación, algunas propiedades útiles de la RTO se han propuesto para el *scheduling* óptimo y viceversa. En el *scheduling* de procesos batch, las incertidumbres en los pronósticos y en el modelo pueden resolverse parcialmente a través de una resolución iterativa del problema de *scheduling* en un horizonte deslizante (Engell y Harjunkoski, 2012). Para el *blending* óptimo de gasolinas, Singh et al. (2000) proponen una formulación NLP multiperíodo para RTO con un horizonte fijo. Otros ejemplos pueden encontrarse en la sección 2.12. Por otra parte, existen aplicaciones industriales que utilizan el modelo adaptado por la RTO para generar el modelo multiperíodo y generar un *schedule* óptimo para un sistema de energías (García y Ruiz, 2013).

En este capítulo se propone un marco teórico para integrar la optimización en tiempo real (RTO) y el *scheduling* óptimo (SO), aplicado a sistemas de energía pero directamente extensible a otros sistemas. El enfoque está basado en una formulación GDP y se resuelve en dos niveles: En el primero, se resuelve el *schedule* óptimo para la totalidad del horizonte a analizar (por ejemplo, 7 días). Esta solución define las puestas en marcha y parada de los equipos con tiempos de transición largos (turbina de gas, calderas). En el segundo nivel, se resuelve la operación para cada período a través de optimización en tiempo real (adaptando el modelo y los pronósticos), con una formulación multiperíodo y un horizonte fijo (por ejemplo, el final del día).

## 5.2. PLANTEO DEL PROBLEMA

El problema a resolver puede plantearse en forma simplificada como:

$$\begin{aligned}
 & \arg \min \sum_{\theta=\theta_0}^{\theta_f} (Q^\theta(\mathbf{y}^\theta, \mathbf{u}^\theta) + \sum_d c_d^\theta + \sum_{d,i,i'} t_d^\theta) \\
 & \text{s.t.} \\
 & \quad \mathbf{h}_G(\mathbf{y}^\theta, \mathbf{u}^\theta, \boldsymbol{\eta}^\theta) = \mathbf{0} \\
 & \quad \boldsymbol{\eta}_G = \mathbf{p}_G(\mathbf{u}^\theta) \\
 & \quad \mathbf{g}_G(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq \mathbf{0} \\
 & \quad \bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id}^\theta \\ \mathbf{h}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta, \boldsymbol{\eta}^\theta) = \mathbf{0} \\ \eta_{id}^\theta = p_{id}(\mathbf{u}^\theta) \\ \mathbf{g}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq \mathbf{0} \\ c_d^\theta = \gamma_{id} \end{array} \right], \quad d \in D \\
 & \quad \mathbf{u}^L \leq \mathbf{u}^\theta \leq \mathbf{u}^L, z_{id}^\theta \in \{\text{verdadero, falso}\}, i \in N_d, d \in D \\
 & \quad \boldsymbol{\Omega}(\mathbf{z}^{\theta_0}, \dots, \mathbf{z}^{\theta_f}) \leq \mathbf{0} \\
 & \quad \Psi(\mathbf{u}^{\theta_0}, \dots, \mathbf{u}^{\theta_f}, \mathbf{y}^{\theta_0}, \dots, \mathbf{y}^{\theta_f}) \leq \mathbf{0}
 \end{aligned} \quad \left. \vphantom{\begin{array}{l} \arg \min \\ \text{s.t.} \\ \mathbf{h}_G \\ \boldsymbol{\eta}_G \\ \mathbf{g}_G \\ \bigvee_{i \in N_d} \\ \mathbf{u}^L \\ \boldsymbol{\Omega} \\ \Psi \end{array}} \right\} \theta = \theta_0, \dots, \theta_f \quad (5-1)$$

donde  $\theta$  (de  $\theta_0$  a  $\theta_f$ ) es el índice correspondiente al período,  $\mathbf{u}^\theta$  es el vector de entradas del proceso,  $\mathbf{y}^\theta$  son las salidas medidas y  $\mathbf{z}^\theta$  las variables booleanas correspondientes a las disyunciones  $d$ . Las variables  $\boldsymbol{\eta}^\theta = (\boldsymbol{\eta}_G^\theta, \eta_{id}^\theta, i \in N_d, d \in D)$  son factores de eficiencia (presentados en el capítulo 3), cuya funcionalidad real con respecto a  $\mathbf{u}$  sólo se conoce en forma aproximada a través de las funciones  $(\mathbf{p}_G, p_{id}, i \in N_d, d \in D)$ . Estas funciones contienen todo el error del modelo, mientras que el resto del modelo del proceso  $(\mathbf{h}_G, \mathbf{h}_{id}, i \in N_d, d \in D)$  comprende los balances de masa, energía y entropía, incluyendo los pronósticos. Las ecuaciones

de desigualdad se expresan a través de las funciones  $(\mathbf{g}_G, \mathbf{g}_{id}, i \in N_d, d \in D)$ . Las funciones  $\Omega$  incluyen proposiciones lógicas que relacionan las variables booleanas. Las ecuaciones  $\Psi$  contienen las relaciones entre períodos.  $Q^\theta$  es la función costo, incluyendo los precios pronosticados, y  $c_d^\theta$  son costos fijos. Por simplicidad, las restricciones de transición, las variables discretas asociadas a las transiciones y los costos de transición  $t_d^\theta$  (relacionados con el cambio en las variables  $z_{id}^\theta$ ) se incluyen implícitamente dentro de  $\Psi$ . La sección siguiente de este capítulo presenta una breve descripción de los posibles costos, restricciones y variables de transición.

### 5.3. FENÓMENOS QUE REQUIEREN UNA FORMULACIÓN MULTIPERÍODO EN RTO

Como hemos visto en los capítulos anteriores, la optimización en tiempo real utiliza generalmente un modelo de estado estacionario para un solo período. Sin embargo, hay decisiones continuas y discretas que pueden involucrar el tiempo y que pueden abordarse de una manera más adecuada a través de una formulación multiperíodo. A continuación, se presentan algunos aspectos de la operación de procesos industriales que, aun manteniendo la hipótesis de estado estacionario, no pueden modelarse considerando un solo período de operación.

#### 5.3.1. Restricciones de transición

Con frecuencia, los equipos industriales pueden funcionar en distintos *modos de operación*. Por ejemplo, una caldera podría tener tres modos de operación: *parada*, *reserva en caliente (stand-by)* o *en marcha*. Las decisiones discretas que involucran cambios en el modo de operación de equipos suelen presentar restricciones relacionadas con el tiempo. Para modelarlas, Mitra et al. (2012) proponen la inclusión de *variables de transición* que toman el valor *verdadero* cuando se produce un cambio.

Para una disyunción representando diferentes modos de operación  $i$  en un equipo:

$$\bigvee_{i \in N_d} \left[ \begin{array}{l} z_{id}^\theta \\ \mathbf{h}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta, \boldsymbol{\eta}^\theta) = 0 \\ \eta_{id}^\theta = p_{id}(\mathbf{u}^\theta) \\ \mathbf{g}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq 0 \\ c_d^\theta = \gamma_{id} \end{array} \right], \quad \theta = \theta_0, \dots, \theta_f \quad (5-2)$$

pueden definirse las *variables de transición*  $z_{dd}^{i,i',\theta}$  a través de la siguiente proposición lógica:

$$\left(z_{id}^{\theta} \wedge z_{i'd}^{\theta-1}\right) \Leftrightarrow z_{dd}^{i,i',\theta} \quad , i \in N_d, i' \in N_d, \theta = \theta_0, \dots, \theta_f \quad (5-3)$$

donde  $i$  e  $i'$  son modos de operación correspondientes a la disyunción  $d$ . Es decir que la variable de transición asociada a los modos de operación  $i$  e  $i'$  (en ese orden) tomará el valor verdadero en el período  $\theta$  si y solo si el equipo cambia el modo  $i'$  al modo  $i$  al comienzo de dicho período.

A partir de estas variables, pueden implementarse diferentes *restricciones de transición* (Mitra et al., 2013, 2012). A continuación, se enumeran algunos posibles ejemplos:

- *Mínimo tiempo de permanencia*: en algunos casos, un equipo debe permanecer en un modo de operación por un tiempo mayor a un límite mínimo. Por ejemplo, al parar una caldera podría ser necesario que el equipo no se ponga en marcha nuevamente por un tiempo, para evitar daños en los materiales por los cambios frecuentes de temperatura. La restricción se formula como:

$$z_{dd}^{i,i',\theta} \Rightarrow z_{i'd}^{\theta+\zeta} \quad \zeta = 0, \dots, T_d^{i,i',\min} \quad , \theta = \theta_0, \dots, \theta_f \quad (5-4)$$

donde  $T_d^{i,i',\min}$  es el número mínimo de períodos en que el equipo debe operar en el modo de operación  $i'$  luego de una transición desde el modo  $i$ .

- *Transiciones prohibidas*: Algunos cambios entre modos de operación pueden estar prohibidos. Por ejemplo, es posible que la caldera no pueda pasar del estado *parada* al estado *en marcha* directamente, sino que deba pasar antes por el estado *reserva en caliente*. La formulación en este caso es:

$$z_{dd}^{i,i',\theta} = \text{falso} \quad , \quad \theta = \theta_0, \dots, \theta_f \quad (5-5)$$

- *Transiciones dependientes de los estados pasados*: la posibilidad de algunas transiciones puede estar relacionada con las transiciones ocurridas en el pasado. Siguiendo con el ejemplo de la caldera, otro caso posible es que el cambio de *parada* a *en marcha* sin pasar por un modo intermedio sólo sea posible cuando se realizó el cambio inverso (*en marcha* a *parada*) en las últimas horas. La restricción puede formularse en este caso como:

$$z_{\overline{ad}}^{i,i',\theta} \Rightarrow \bigvee_{\zeta=T_d^{i,i',\min}, \dots, T_d^{i,i',\max}} z_{\overline{ad}}^{i,i',\theta-\zeta}, \theta = \theta_0, \dots, \theta_f \quad (5-6)$$

es decir, el cambio de  $i$  a  $i'$  sólo es posible si el cambio inverso ocurrió entre los períodos  $(\theta - T_d^{i,i',\max})$  y  $(\theta - T_d^{i,i',\min})$ .

Las transiciones también pueden involucrar un costo (expresado como  $t_d^\theta$ ) en (5-1), así como también generar otras restricciones en las variables continuas de operación. Esto puede expresarse a través de la disyunción:

$$\left[ \begin{array}{c} z_{\overline{ad}}^{i,i',\theta} \\ \mathbf{g}_{\overline{ad}}^{i,i'}(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq 0 \\ t_d^\theta = \gamma_{\overline{ad}}^{i,i'} \end{array} \right] \vee \left[ \begin{array}{c} \neg z_{\overline{ad}}^{i,i',\theta} \\ t_d^\theta = 0 \end{array} \right], \quad i, i' \in N_d, d \in D, \theta = \theta_0 \dots \theta_f \quad (5-7)$$

donde  $\mathbf{g}_{\overline{ad}}^{i,i'}$  son las restricciones adicionales generadas por la transición del estado  $i'$  al  $i$ , y  $\gamma_{\overline{ad}}^{i,i'}$  el costo fijo ocasionado por la transición.

### 5.3.2. Variables de acumulación

Si bien la RTO suele utilizarse en procesos continuos donde puede asumirse el estado estacionario, en algunos casos es inevitable la presencia de variables que representan propiedades o cantidades que evolucionan con el tiempo y se acumulan. En sistemas de calor y potencia, es posible enumerar algunos ejemplos:

- Contratos de consumo de gas natural con un valor máximo diario a partir del cual se paga una penalidad.
- Disponibilidad de combustibles líquidos o sólidos (por ejemplo, inventario de bagazo en una planta de azúcar y etanol, nivel de tanques de combustibles líquidos residuales en la industria petroquímica).
- Legislación ambiental que pone límites a las emisiones acumuladas o promediadas durante un cierto período.

En una formulación multiperíodo, una variable acumulada  $x_{Ac}$  se estima para cada período  $\theta$  en base a las entradas y salidas (*flujos*)  $x_e$  y  $x_s$ , y a las velocidades de reacción  $R_i$  como:

$$x_{Ac}^\theta = x_{Ac}^{\theta_0} + \sum_{l=\theta_0+1}^k \left( \sum_{e=1}^{N_e} x_e^l - \sum_{s=1}^{N_s} x_s^l + \sum_{i=1}^{N_R} \nu_{xi} R_i \right) \cdot \Delta t, \theta = \theta_0 + 1, \dots, \theta_f \quad (5-8)$$

donde  $N_e$  es el número de *flujos* de entrada,  $N_s$  el número de *flujos* de salida,  $N_R$  es el número de reacciones químicas,  $\nu_{xi}$  el coeficiente estequiométrico de  $x$  en la reacción  $i$ , y  $\Delta t$  la duración de cada período.

## 5.4. INTEGRACIÓN DE *SCHEDULING* ÓPTIMO Y OPTIMIZACIÓN EN TIEMPO REAL (SO+RTO)

### 5.4.1. Descripción de la metodología propuesta

En una aplicación industrial, el tamaño de un problema del tipo de (5-1) puede determinar un costo computacional demasiado elevado como para resolverlo con una frecuencia alta y aplicar los resultados en tiempo real. Por otra parte, si sólo se lo resuelve fuera de línea y con una frecuencia baja, los errores de modelado y las desviaciones entre los precios y demandas reales y los pronosticados pueden llevar a una operación subóptima o a la violación de restricciones.

Por este motivo, se propone una resolución del problema (5-1) en dos niveles, 1 y 2. Para esto, se divide el conjunto de decisiones discretas  $D$  en dos subconjuntos,  $D_1$  y  $D_2$ .  $D_1$  incluye las decisiones con tiempos de transición relativamente largos y que no pueden cambiarse frecuentemente. El resto de las decisiones discretas pertenecen a  $D_2$ .

El nivel 1 de la estrategia resuelve el *scheduling* óptimo (SO) para la totalidad del horizonte e incluyendo todas las disyunciones  $D$ . Luego se fijan los valores obtenidos para las variables de decisión discretas correspondientes al subconjunto  $D_1$  y se aplican en la planta.

El nivel 2 de resolución utiliza las propiedades de *feedback* de la optimización en tiempo real (RTO), y se resuelve con mayor frecuencia que el nivel 1. Esta etapa tiene dos objetivos principales: el primero es actualizar los pronósticos de precios y demandas; el segundo es adaptar el modelo a través de las mediciones disponibles para atenuar el efecto de los errores de modelado. Una vez que se han actualizado el modelo y la información, se recalcula la operación óptima. Los grados de libertad disponibles son las decisiones correspondientes al subconjunto  $D_2$  y las variables de entrada continuas  $\mathbf{u}$ . Además, para reducir el tamaño del problema, no se tiene en cuenta el horizonte completo del nivel 1, sino que el tiempo total analizado es más corto. Dependiendo del problema, puede utilizarse un horizonte fijo o móvil. En esta sección se propone una estrategia de horizonte fijo.

## 5.4.2. Formulación de la metodología

### 5.4.2.1. Nivel 1: Optimización del schedule completo (SO)

El problema (5-1) se resuelve desde el período inicial hasta el horizonte final ( $\theta_0 = 1, \theta_f = N$ ), con un pronóstico inicial para las demandas y los precios. El valor óptimo de las variables discretas pertenecientes a  $D_1$ ,  $\{z_{id}^{\theta, opt}, i \in N_d, d \in D_1, \theta = 1, \dots, N\}$ , se almacenan y no serán modificados en el nivel 2. Esto se expresa a través de la siguiente restricción:

$$z_{id}^{\theta} = z_{id}^{\theta, opt}, \quad i \in N_d, d \in D_1, \theta = 1 \dots N \quad (5-9)$$

El nivel 1 puede plantearse con un horizonte fijo, y repetirse cuando se alcanza el tiempo final  $N$ ; también puede resolverse con una estrategia de horizonte móvil (Cui y Engell, 2009; Sand y Engell, 2004), o actualizarse cuando se detecta un cambio significativo en los pronósticos o en las condiciones de operación. El problema de scheduling óptimo puede resolverse usando un modelo determinístico o recurriendo a alguna de las formulaciones con incertidumbre de la sección 2.12 (scheduling en dos etapas, multi-etapa, formulaciones robustas)

### 5.4.2.2. Nivel 2: RTO multiperíodo (RTO)

1. **Estrategia de adaptación:** En cada período  $\theta$ , las mediciones del proceso se utilizan para adaptar las ecuaciones de eficiencia de, siguiendo la estrategia presentada en el capítulo 3:

$$\begin{aligned} \eta_{G,j} &= p_{G,j}(\mathbf{v}_{G,j}) + a_{G,j}^k + \left(\mathbf{b}_{G,j}^k\right)^T \cdot \mathbf{v}_{G,j} \quad j = 1 \dots n_j \\ \eta_{id} &= p_{id}(\mathbf{v}_{id}) + a_{id}^k + \left(\mathbf{b}_{id}^k\right)^T \cdot \mathbf{v}_{id} \quad i \in N_d, d \in D \end{aligned} \quad (5-10)$$

donde  $a$  y  $\mathbf{b}$  son modificadores de las restricciones y del gradiente de las restricciones, respectivamente, y  $\mathbf{v}_{G,j}$  y  $\mathbf{v}_{id}$  son las variables auxiliares definidas en el capítulo 3.  $n_j$  es el número de factores de eficiencia globales (es decir, la dimensión del vector  $\boldsymbol{\eta}_G$ ). Los modificadores se calculan para las ecuaciones de eficiencia globales y para los términos de las disyunciones *activos* al comienzo del período  $k$ , siguiendo la estrategia presentada en la sección 4.4.2.

2. **Actualización de los pronósticos:** Además del modelo, es necesario corregir los pronósticos para los precios, las demandas y las condiciones ambientales. Si se dispone de un nuevo pronóstico actualizado, puede utilizarse directamente. En caso contrario, los

pronósticos originales pueden corregirse teniendo en cuenta el error entre el valor real y el valor pronosticado para el período actual:

$$\xi_{\pi}^k = \hat{\pi}^k - \pi_p^k \quad (5-11)$$

donde  $\hat{\pi}^k$  es el valor observado para el parámetro  $\pi$  en el período actual  $k$ ,  $\pi_p^k$  es el pronóstico original y  $\xi_{\pi}^k$  es el error del pronóstico. Para el problema de optimización, los pronósticos corregidos a utilizar son:

$$\pi^{\theta} = \pi_p^{\theta} + \xi_{\pi}^k, \quad \theta = k, \dots, \theta_f \quad (5-12)$$

3. Optimización: Después de actualizar el modelo y los pronósticos, se resuelve la optimización del sistema al comienzo del período actual  $k$ . Para reducir los tiempos de cálculo, el horizonte completo de scheduling se divide en subconjuntos  $l$ :

$$\{1 \dots N\} = \left\{ \left\{ 1 \dots \theta_{f1} \right\} \cup \left\{ \theta_{f1} + 1 \dots \theta_{f2} \right\} \cup \dots \cup \left\{ \theta_{f, T-1} + 1 \dots N \right\} \right\} \quad (5-13)$$

donde  $\theta_{fl}$  ( $l = 1, \dots, T$ ) es el período final de cada subconjunto  $l$ . En la práctica, un posible tamaño para cada subconjunto  $l$  es 1 día (con  $\theta_{fl}$  el último período del día). Esto permite trabajar con restricciones y esquemas de costos presentes con frecuencia en sistemas de calor y potencia (máximo consumo diario de electricidad, uso máximo de gas natural, emisiones diarias, entre otros).

En este caso, el horizonte está fijo, y el problema se resuelve desde el período actual  $k$  hasta el período final  $\theta_{f\alpha}$  del subconjunto  $\alpha = \{l / k \in l\}$  (es decir, para todo el tiempo restante hasta el final del horizonte fijo correspondiente al instante actual; por ejemplo, si la duración de  $l$  es un día, el problema se resuelve teniendo en cuenta todos los períodos restantes hasta el final del día).

El problema de optimización a resolver es (5-1), restringido a través de (5-9), con los modelos actualizados según (5-10) y los pronósticos de (5-12):

$$\begin{aligned}
& \underset{\mathbf{y}, \mathbf{u}, \boldsymbol{\eta}}{\operatorname{argmin}} \sum_{\theta=k}^{\theta_{f\alpha}} (Q^\theta(\mathbf{y}^\theta, \mathbf{u}^\theta) + \sum_d c_d^\theta + \sum_{d,i,i'} t_d^{i,i',\theta}) \\
& \quad \text{s.t.} \\
& \quad \mathbf{h}_G(\mathbf{y}^\theta, \mathbf{u}^\theta, \boldsymbol{\eta}^\theta) = 0 \\
& \quad \eta_G^k = p_{G,j}(\mathbf{v}_{G,j}^\theta) + a_{G,j}^k + (\mathbf{b}_{G,j}^k)^T \cdot \mathbf{v}_{G,j}^\theta, \quad j=1, \dots, ng_G \\
& \quad \mathbf{g}_G(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq 0 \\
& \quad \left[ \begin{array}{c} z_{id}^\theta \\ \mathbf{h}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta, \boldsymbol{\eta}^\theta) = 0 \\ \eta_{id}^\theta = p_{id}(\mathbf{v}_{id}^\theta) + a_{id}^k + (\mathbf{b}_{id}^k)^T \cdot \mathbf{v}_{id}^\theta, \quad d \in D \\ \mathbf{g}_{id}(\mathbf{y}^\theta, \mathbf{u}^\theta) \leq 0 \\ c_d^\theta = \gamma_{id} \end{array} \right]_{i \in N_d}, \quad \theta = k, \dots, \theta_{f\alpha} \\
& \quad \mathbf{u}^L \leq \mathbf{u}^\theta \leq \mathbf{u}^U, \quad z_{id}^\theta \in \{\text{verdadero}, \text{falso}\}, i \in N_D, d \in D \\
& \quad \boldsymbol{\Omega}(\mathbf{z}^{\theta_0}, \dots, \mathbf{z}^{\theta_{f\alpha}}) \leq \mathbf{0} \\
& \quad \Psi(\mathbf{u}^{\theta_0}, \dots, \mathbf{u}^{\theta_{f\alpha}}, \mathbf{y}^{\theta_0}, \dots, \mathbf{y}^{\theta_{f\alpha}}) \leq \mathbf{0}
\end{aligned} \tag{5-14}$$

Si  $k = \theta_{f\alpha} - 1$ , se resuelve una optimización de un solo período, y en la siguiente instancia del sistema se definirá un nuevo horizonte  $\theta_{f,\alpha+1}$ .

Las restricciones  $\boldsymbol{\Omega}$  y  $\Psi$  siguen manteniendo la información de ciclos anteriores, ya que las restricciones de transición y las variables acumuladas pueden requerir valores pasados. Cuando estos valores anteriores están disponibles como mediciones, no se utiliza el valor estimado en las ejecuciones de RTO anteriores, sino el valor real medido.

4. Aplicación de la solución óptima: Los resultados óptimos pueden validarse o filtrarse antes de implementarse en la planta (Bunin et al., 2011; Miletic y Marlin, 1996). Sólo se aplican las entradas óptimas  $\mathbf{u}^{k+1,*}$  y las decisiones discretas óptimas  $z_{id}^{k+1,*}$ , ya que las decisiones para los períodos siguientes ( $k+2$ ,  $k+3$ , etc.) estarán sujetas a un nuevo ciclo de optimización en tiempo real, una vez transcurrido el período  $k+1$ .

#### 5.4.3. Reacción a las soluciones no factibles

Los algoritmos de resolución de problemas MINLP y GDP pueden presentar dificultades para la convergencia si el problema planteado es complejo,

los valores iniciales de las variables están muy alejados de la solución óptima o los parámetros de configuración del algoritmo no son los adecuados. Por este motivo, es probable que en algunos ciclos de optimización el problema (5-14) conduzca a una solución no factible, o que no consiga converger a una solución factible en el tiempo disponible antes de aplicar los resultados de la RTO.

Para hacer frente a este inconveniente, el sistema *SO+RTO* debe incluir alguna estrategia que, ante la presencia de soluciones no factibles, proponga un conjunto de entradas para aplicar en la planta en el período actual. En la optimización en tiempo real tradicional, como se supone un estado estacionario, la reacción a una solución no factible es mantener, para el ciclo  $k+1$ , las entradas aplicadas en el ciclo anterior,  $\mathbf{u}^k$ , y esperar al próximo ciclo para obtener una nueva solución.

En el caso del sistema *SO+RTO*, existen alternativas intermedias a esperar al próximo ciclo. Si se eliminan las decisiones discretas y se mantienen sólo los términos activos de las disyunciones en el ciclo actual  $k$ :

$$z_{id}^{\theta} = z_{id}^k, \quad i \in N_d, d \in D, \theta = k+1, \dots, \theta_{fl} \quad (5-15)$$

se obtendrá un problema NLP de una complejidad mucho menor al problema original.

Si se desea una complejidad aún menor, es posible resolver el problema de RTO para el período actual, y sólo con variables continuas. En este caso, se eliminarían las restricciones que conectan períodos y se reduciría considerablemente el tamaño y la dificultad para obtener el óptimo.

De esta manera, se sugiere la siguiente estrategia de solución del nivel 2 (RTO) de la metodología *SO+RTO*, que contempla la ocurrencia de soluciones no factibles. Para determinar las entradas óptimas  $(\mathbf{u}^{k+1,*}, \mathbf{z}^{k+1,*})$  en el período  $k+1$ , se propone:

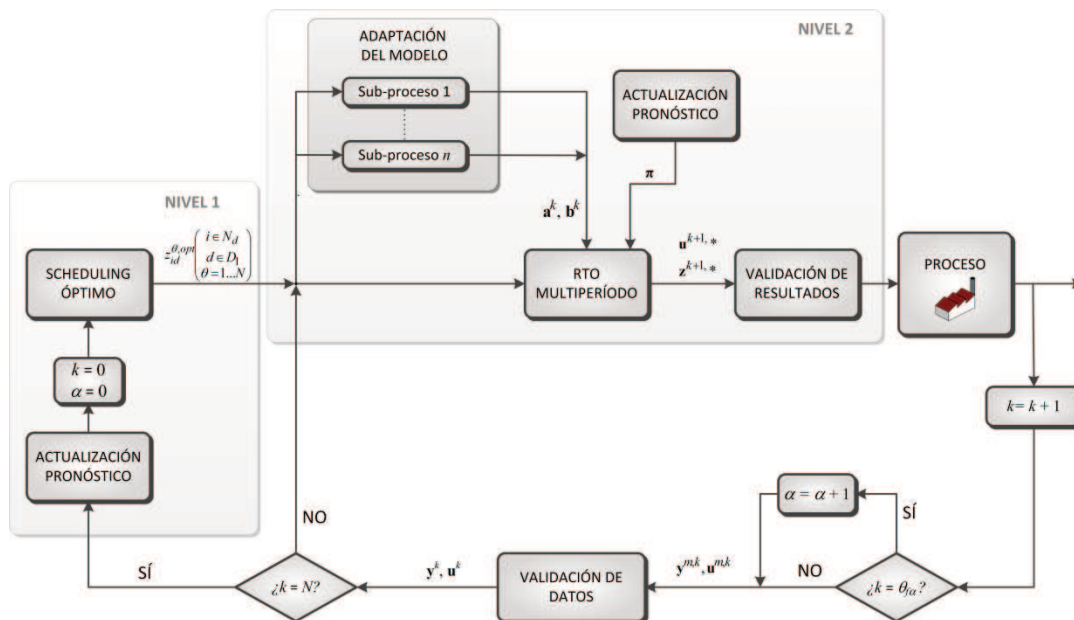
1. Resolver el problema (5-14). Si se alcanza una solución, ir al paso 5.
2. Agregar la restricción (5-15) y resolver el NLP correspondiente. Si se alcanza una solución, ir al paso 5.
3. Fijar el horizonte final en  $k+1$ . Resolver el NLP resultante (de un solo período). Si se alcanza una solución, ir al paso 5.
4. Fijar  $\mathbf{u}^{k+1,*} = \mathbf{u}^k, z_{id}^{k+1,*} = z_{id}^k, i \in N_d, d \in D$ .

- Validar los resultados óptimos alcanzados y aplicar en la planta la solución obtenida para el período  $k + 1$ .

El problema de las soluciones no factibles y los resultados subóptimos en la RTO se analiza con mayor detalle en el Capítulo 6

#### 5.4.4. Estructura completa propuesta para el sistema integrado SO+RTO

La **Figura 5-1** muestra un diagrama del sistema integrado SO+RTO. Además de los niveles 1 y 2 de resolución, el sistema incluye un paso de validación de datos para corregir las entradas y salidas medidas  $\mathbf{u}^{m,k}$  e  $\mathbf{y}^{m,k}$  (Darby et al., 2011; Mansour y Ellis, 2008). El bloque *RTO multiperíodo* incluye la estrategia de reacción a las infactibilidades presentada en la sección 5.4.3.



**Figura 5-1.** Estructura completa propuesta para el sistema integrado SO+RTO.

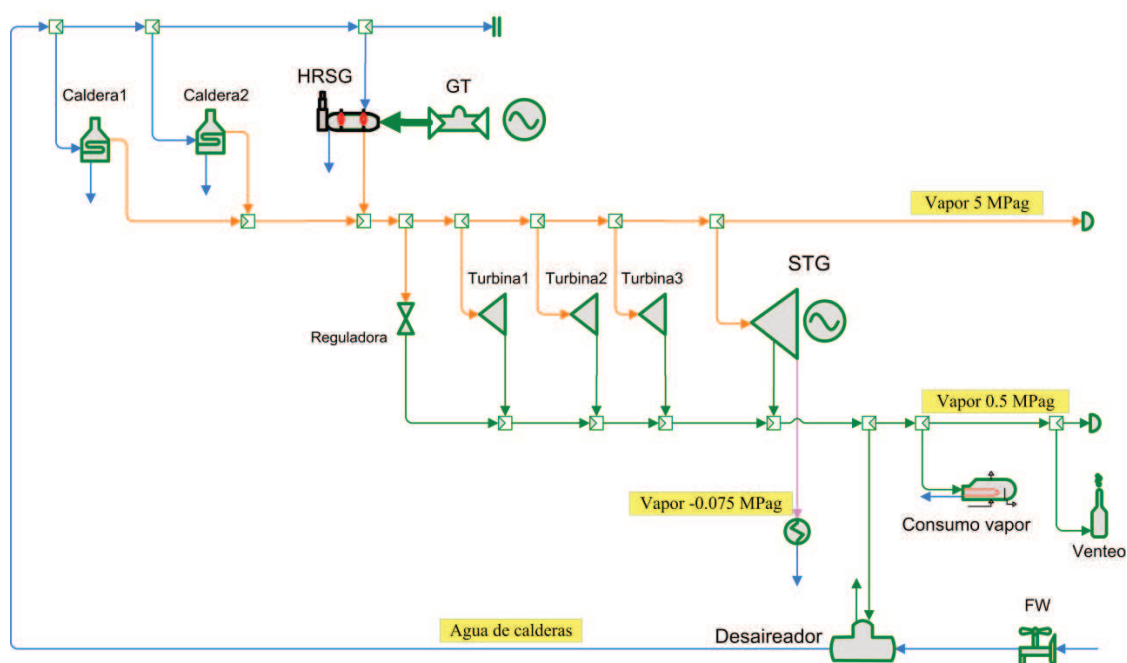
## 5.5. CASO DE ESTUDIO: SISTEMA DE CALOR Y POTENCIA

### 5.5.1. Descripción del caso

La metodología desarrollada se aplica a un sistema de calor y potencia, que consiste en 2 calderas, 1 turbina de gas con caldera de recuperación y quemado de combustible suplementario (postcombustión), 1 turbina de extracción-condensación con generación de energía eléctrica, 3 turbinas de vapor de contrapresión para impulsar bombas (que pueden reemplazarse por motores eléctricos), una demanda de vapor de un proceso industrial, válvulas reguladoras de presión y venteo de vapor a la atmósfera. Las emisiones diarias de NO<sub>x</sub> están limitadas a 1000 kg/día. El sistema se estudia para un horizonte de 7 días. La

puesta en marcha y parada de calderas se selecciona en el nivel 1 de la estrategia SO+RTO (optimización del *schedule* inicial); mientras que las decisiones de puesta en marcha y parada para las turbinas de vapor y para la postcombustión se definen en el nivel 2, a través de optimización en tiempo real, multiperíodo (con un horizonte fijo al final de cada día), resuelta cada 6 horas.

El diagrama del sistema modelado se muestra en la **Figura 5-2**. El modelo se basa en el Caso de Estudio 2 de la sección 4.6.2; que se desarrolla en detalle en el Apéndice B. Las ecuaciones particulares de este caso de estudio (que involucran puesta en marcha y parada de equipos, transiciones y variables de acumulación) se presentan en el Apéndice C.



**Figura 5-2.** Diagrama del sistema de calor y potencia estudiado.

Todos los balances de masa, energía y entropía se modelan rigurosamente, mientras que para los factores de eficiencia (eficiencias de calderas, *heat rate* y máxima potencia de la turbina de gas, eficiencias de las turbina de vapor) se ha construido un modelo aproximado. Las emisiones de NOx se modelaron a través de factores de emisión (Office of Air Quality Planning and Standards y Office of Air and Radiation, 1995).

Con el propósito de evaluar la estrategia propuesta, se ha aplicado el método ya utilizado en capítulos anteriores, empleando dos modelos: el modelo de la *planta real* (que se supone representa exactamente el proceso) y el *modelo RTO* que difiere estructuralmente de la planta real. Las principales diferencias entre el modelo RTO y la planta real se resumen en la **Tabla 5-1**. El resto de las

diferencias (comportamiento de las turbinas de vapor y la turbina de gas) pueden obtenerse de la **Tabla 4-2**(en el capítulo 4).

Para resolver el problema conjunto de *scheduling* óptimo y RTO (*SO+RTO*) se dispone de un pronóstico para el precio de la electricidad, la demanda de potencia eléctrica y vapor y la temperatura ambiente. Para evaluar el desempeño de la técnica, un conjunto diferente de demandas, precio y temperatura representa los valores *reales* de estos parámetros a lo largo del horizonte de *scheduling* (estos valores no pueden ser conocidos de antemano). Los valores *reales* se utilizan para simulación (junto con el modelo de la *planta real*) y para actualizar los pronósticos del modelo RTO a través de (5-11)-(5-12). Los valores pronosticados y reales se muestran en la **Figura 5-3**.

**Tabla 5-1.** Diferencias entre el modelo aproximado (RTO) y la *planta real*.  $F$ : caudal de vapor generado en las calderas (t/h);  $Q$ : uso de combustible en calderas, turbina de gas y postcombustión (Gcal/h);  $\theta$ : período del horizonte de optimización.

Índice de eficiencia	Modelo RTO	Planta real
Eficiencia calderas (%)	92	$75 + 0.48 \cdot F - 0.0035 \cdot F^2$
NOx Caldera 1 (kg/GJ)	0.08	$0.07 + 2 \cdot Q^{-1}$
NOx Caldera 2 (kg/GJ)	0.08	$0.08 + 3 \cdot Q^{-1}$
NOx Turbina de gas (kg/GJ)	0.055	$0.03 + 10 \cdot Q^{-1}$
NOx Postcombustión (kg/GJ)	0.07	$0.05 + 5 \cdot Q^{-1} + 0.001 \cdot \theta$

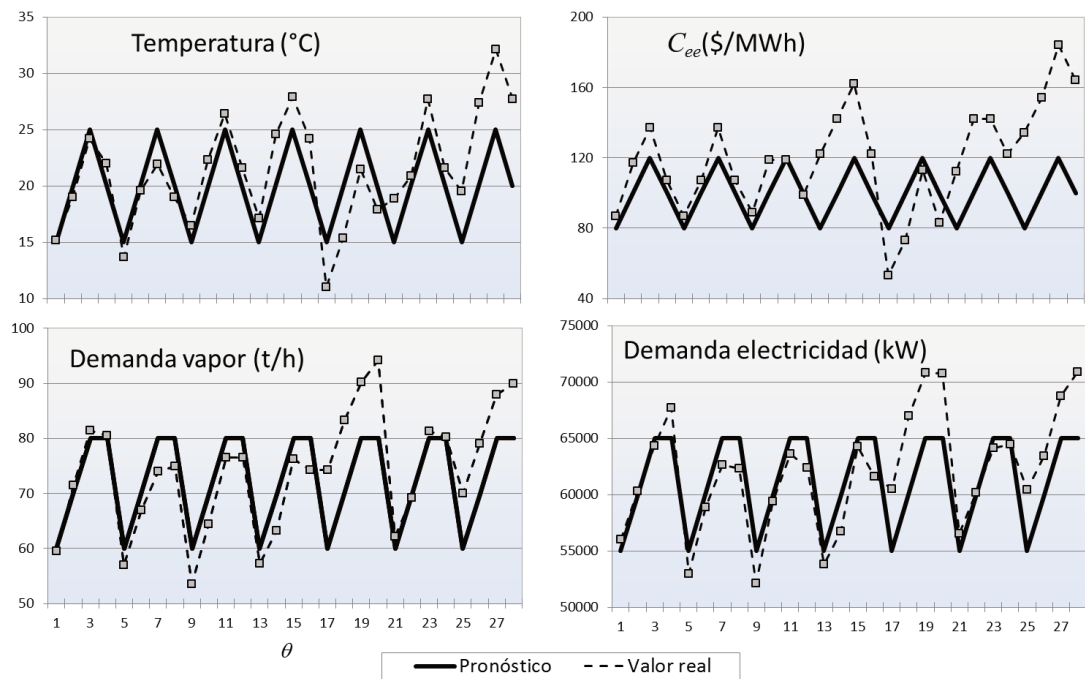
La función objetivo a minimizar es el costo operativo  $CO$  del sistema:

$$CO = \sum_{\theta=\theta_0}^{\theta_f} \left( \left( 7.2 \cdot \left( Q_{TG}^{\theta} + Q_{PC}^{\theta} + \sum_{cb=1}^2 Q_{cb}^{\theta} \right) + 0.5 \cdot FW^{\theta} + C_{ee}^{\theta} \cdot P^{\theta} \right) \cdot \Delta \right. \\ \left. + 120 \cdot \sum_{t=1}^3 \left( z_t^{on-off,\theta} + z_t^{off-on,\theta} \right) \right) + 300 \cdot \sum_j EX_{NO_x}^j \quad (5-16)$$

donde  $Q_{TG}^{\theta}$ ,  $Q_{PC}^{\theta}$  y  $Q_{cb}^{\theta}$  son los consumos de combustible en la turbina de gas, la postcombustión y cada caldera  $bl$ , respectivamente (GJ/h);  $FW^{\theta}$  es el consumo de agua tratada para las calderas (t/h);  $C_{ee}^{\theta}$  es el precio de compra de la electricidad (\$/MWh) y  $P^{\theta}$  es la importación neta de potencia eléctrica (MW);  $z_t^{on-off,\theta}$  y

$z_t^{off-on,\theta}$  son variables binarias de transición, que indican que una turbina  $t$  fue parada o puesta en marcha, respectivamente, en el período  $\theta$ . La duración de cada período,  $\Delta$ , es de 6 horas. Al final de cada día  $j$ , el exceso de NOx generado,  $EX_{NO_x}^j$  (es decir la producción por encima de 1000 kg/d) se penaliza con un costo de 300 \$/kg.

Otras restricciones incluyen la necesidad de mantener al menos una caldera en operación. Una caldera puede cambiar de estado (pasar de *parada* a *en marcha*) como máximo una vez cada 48 horas (8 períodos), mientras que cada turbina de contrapresión puede ser puesta en marcha o parada sólo una vez cada 24 horas. La turbina de dos etapas debe estar siempre en marcha.



**Figura 5-3.** Pronósticos y valores reales para precio de la electricidad, demandas y temperatura ambiente.

Para que la solución se acepte como factible, se impone la restricción de que se alcance el resultado en un tiempo menor a 1.5 s. En caso de que el problema de RTO multiperíodo no devuelva una solución factible, se aplica la estrategia presentada en la sección 5.4.3. En una implementación real, el tiempo máximo aceptable para alcanzar la solución de RTO puede ser de varios minutos, dependiendo de la duración de los períodos; en este caso de estudio, se fijó arbitrariamente un tiempo bajo para forzar la aplicación de la estrategia, ya que en los escenarios planteados el modelo se resuelve rápidamente y converge en todos los casos a soluciones factibles del problema MINLP.

### 5.5.2. Resultados y discusión

El caso de estudio se implementó en GAMS con una formulación MINLP *big-M* y se resolvió utilizando DICOPT en un ordenador con un procesador Intel Core i7-2670QM (2.20 GHz) y una memoria RAM de 8 GB. El problema del nivel 1 (*schedule* del horizonte completo) tiene 2541 ecuaciones y 2745 variables (196 discretas), y requirió un tiempo de solución de 26.61s. Los problemas del nivel 2 (RTO multiperíodo) tienen un tamaño máximo de 362 ecuaciones y 418 variables (16 discretas), que corresponde a 4 períodos. 23 de los ciclos del nivel 2 (RTO multiperíodo) se resolvieron en un tiempo menor a 1.5 s. En los ciclos restantes (13, 14, 17, 21 y 26) se aplicó la solución multiperíodo fijando las decisiones discretas.

Con el fin de comparar los resultados alcanzados por la estrategia SO+RTO, se calcularon y presentan los resultados obtenidos al aplicar en la *planta real* sólo el nivel 1 (*scheduling* fuera de línea). Los costos y las restricciones reales alcanzados se calculan a través del modelo *planta real*.

El costo operativo total para el caso de comparación es de \$1229964, mientras que la estrategia SO+RTO alcanza un costo de \$976842. La reducción relativa en el costo es del 20.6%.

La **Figura 5-4** presenta la evolución del costo para la estrategia propuesta (SO+RTO) y el caso de comparación (*Scheduling fuera de línea*). Puede observarse que la actualización del modelo y los pronósticos provista por la RTO impacta en una reducción del costo total en todos los períodos analizados. La penalidad por exceso en las emisiones de NOx no puede aplicarse a cada período sino a cada día completo; por este motivo, para ilustrar los resultados se distribuye este costo en forma equitativa entre todos los períodos del día correspondiente (25% de la penalidad diaria se asigna al costo de cada período).

La **Figura 5-5** muestra la evolución de la emisión total de NOx. Una vez más, la realimentación del modelo con datos de la planta en la RTO reduce el número y el tamaño de las violaciones en la restricción de emisión diaria de NOx. Los errores de modelado y la incertidumbre en las demandas pronosticadas son responsables de una diferencia en el uso de combustible y la producción de NOx en las calderas y la turbina de gas, que se corrige (al menos parcialmente) a través de la RTO. En este ejemplo, el modelo no predice ninguna violación de las restricciones en la emisión de NOx en ninguno de los días; no obstante, la aplicación de los resultados del caso de comparación a la *planta real* genera una cantidad mayor a la permitida. En este caso de estudio, el límite se modeló como una restricción *blanda*, es decir que puede ser ligeramente violada pero es fuertemente penalizada en la función objetivo (5-16). En otros casos, el límite

podría ser una restricción *estricta*, y por lo tanto el *schedule* propuesto no sería factible.

La **Figura 5-6** muestra que los venteos de vapor para cada período se reducen significativamente cuando se aplica la metodología *SO+RTO*. Este comportamiento se debe a la gestión óptima en tiempo real de las turbinas de contrapresión, mostrado en la **Figura 5-7**. Esta reducción de los venteos tiene un impacto directo en el costo, ya que corresponde a una disminución en la producción de vapor, que conlleva un menor consumo de combustible.

La **Figura 5-8** muestra el uso de cada turbina de contrapresión en la estrategia *SO+RTO*. Puede observarse que el número de turbinas en marcha varía para operar en forma óptima, pero a la vez cada turbina respeta la restricción de no cambiar de estado (en marcha – parada) más de una vez cada 4 períodos (es decir, más de una vez por día). Lo mismo puede observarse para las calderas (**Figura 5-9**), que se fijan en el nivel 1 (*schedule* inicial) y por lo tanto se paran o ponen en marcha de la misma manera en las dos estrategias (*SO+RTO* y *schedule fuera de línea*).

La **Figura 5-10** muestra el uso de la turbina de gas. Mientras el *schedule* original propone operar en todo momento la turbina en su máxima potencia (coincidente con su máxima eficiencia), la metodología *SO+RTO* opera por debajo de la capacidad máxima, para evitar las emisiones por encima del límite y la generación de vapor en exceso en la caldera de recuperación, que debería ventearse a la atmósfera.

Incluso, como puede verse en la **Figura 5-11**, el *scheduling* inicial propone quemar combustible en la caldera de recuperación (*postcombustión*) para generar vapor adicional. Por el contrario, la estrategia *SO+RTO* no propone en ningún período utilizar la *postcombustión*.

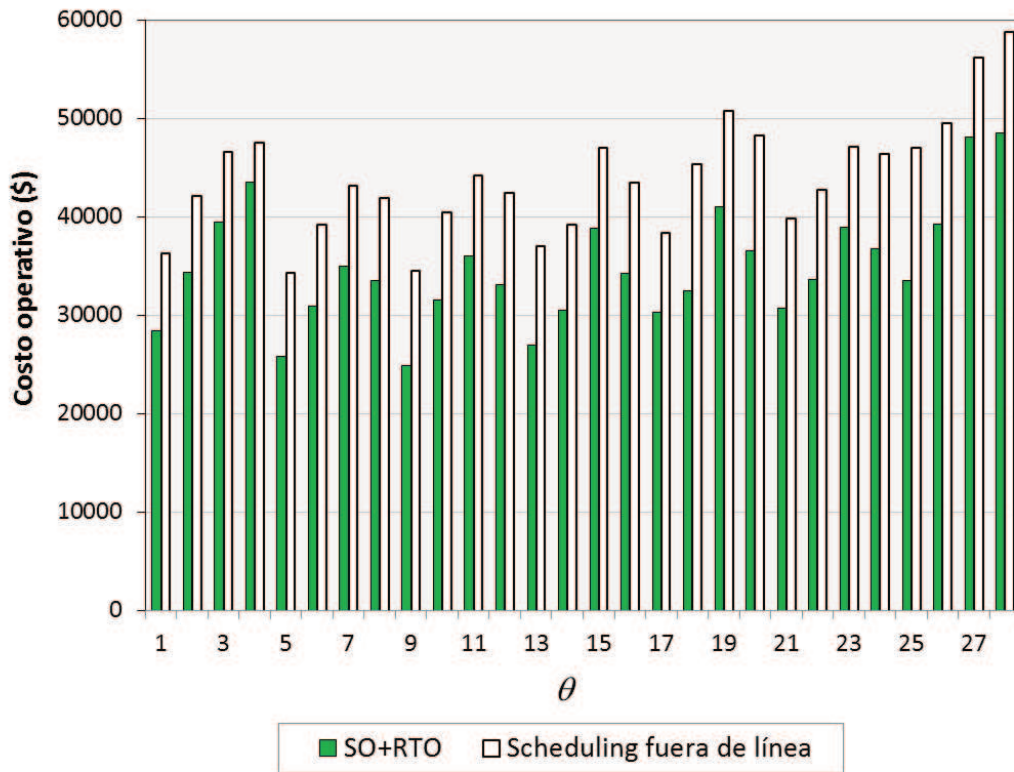


Figura 5-4. Costo operativo por período.

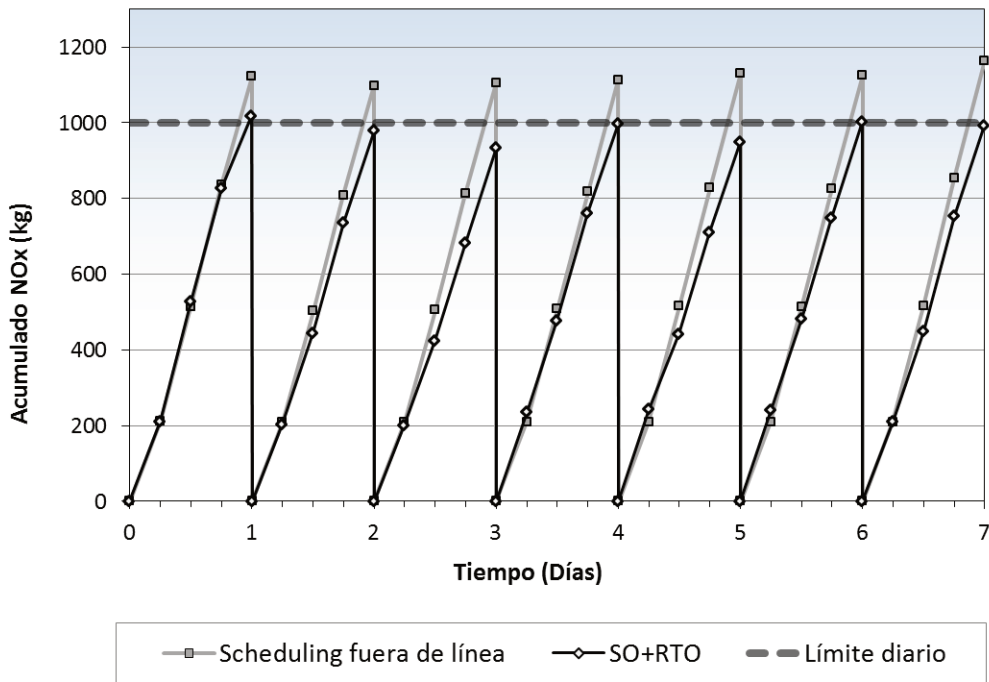


Figura 5-5. Producción acumulada de NOx y límite diario.

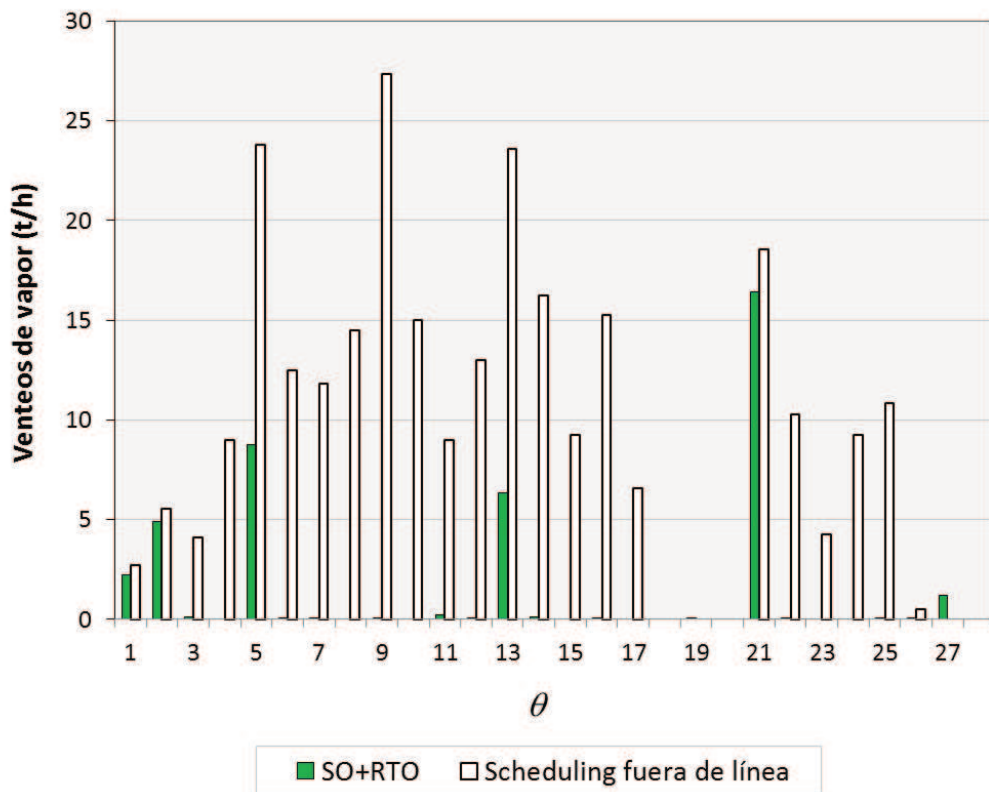


Figura 5-6. Exceso de vapor venteado a la atmósfera.

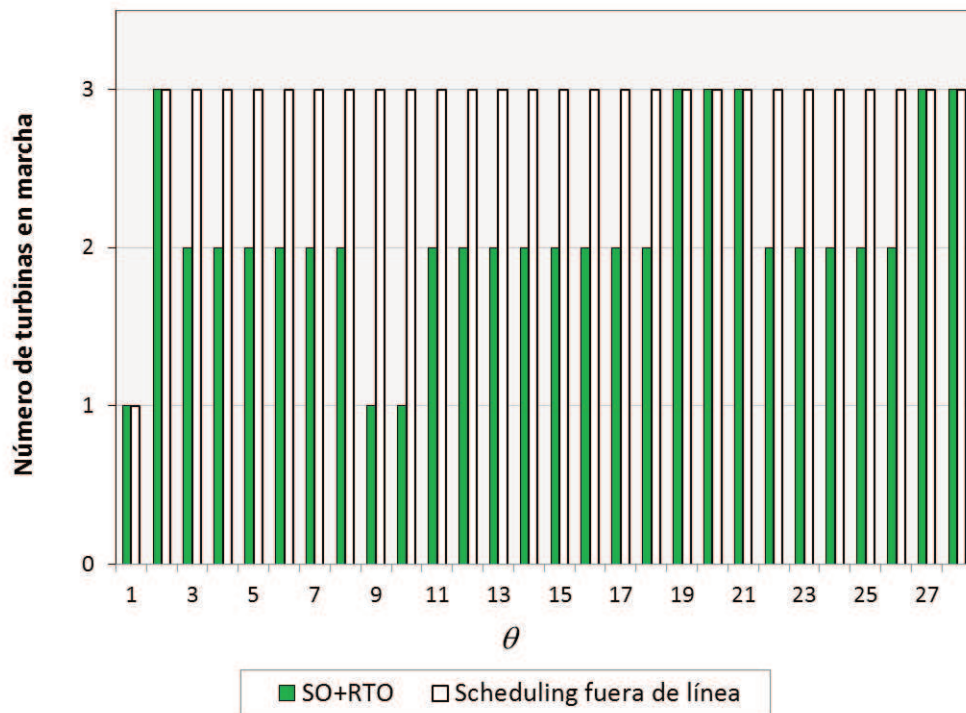
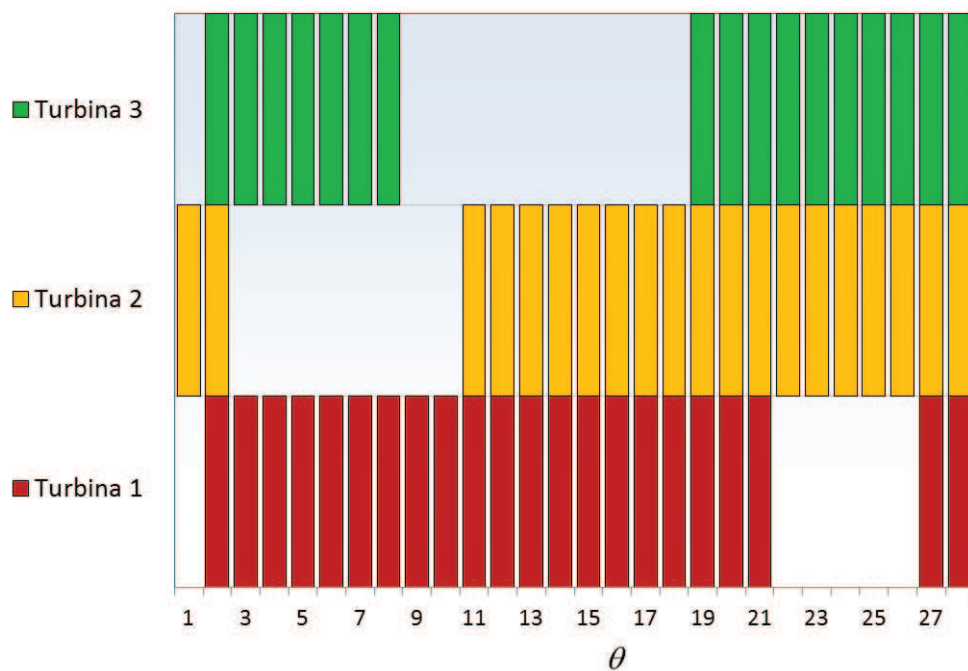
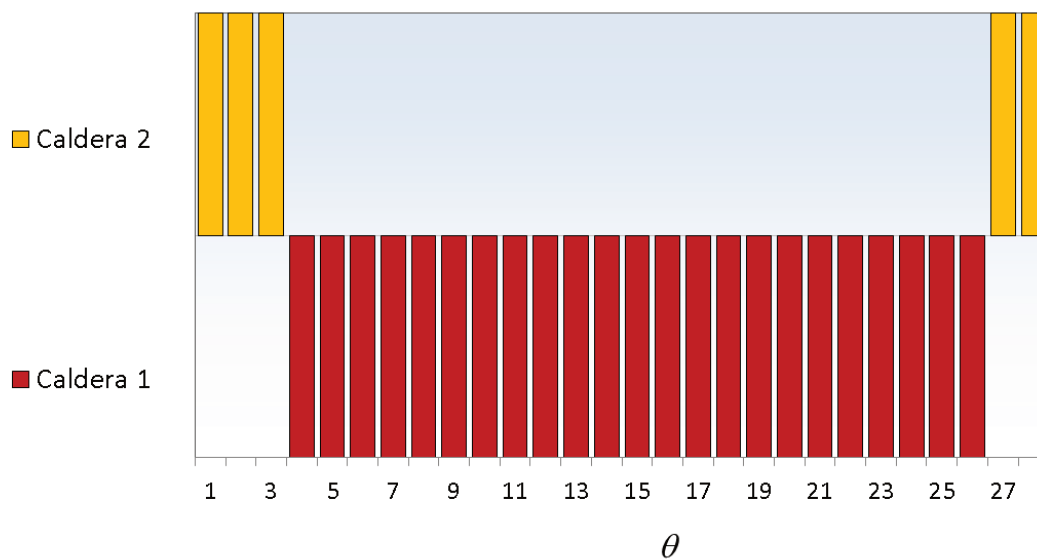


Figura 5-7. Número de turbinas de contrapresión en marcha



**Figura 5-8.** Turbinas de contrapresión en operación por período en la implementación *SO-RTO*.



**Figura 5-9.** Calderas en operación por período para las dos estrategias.

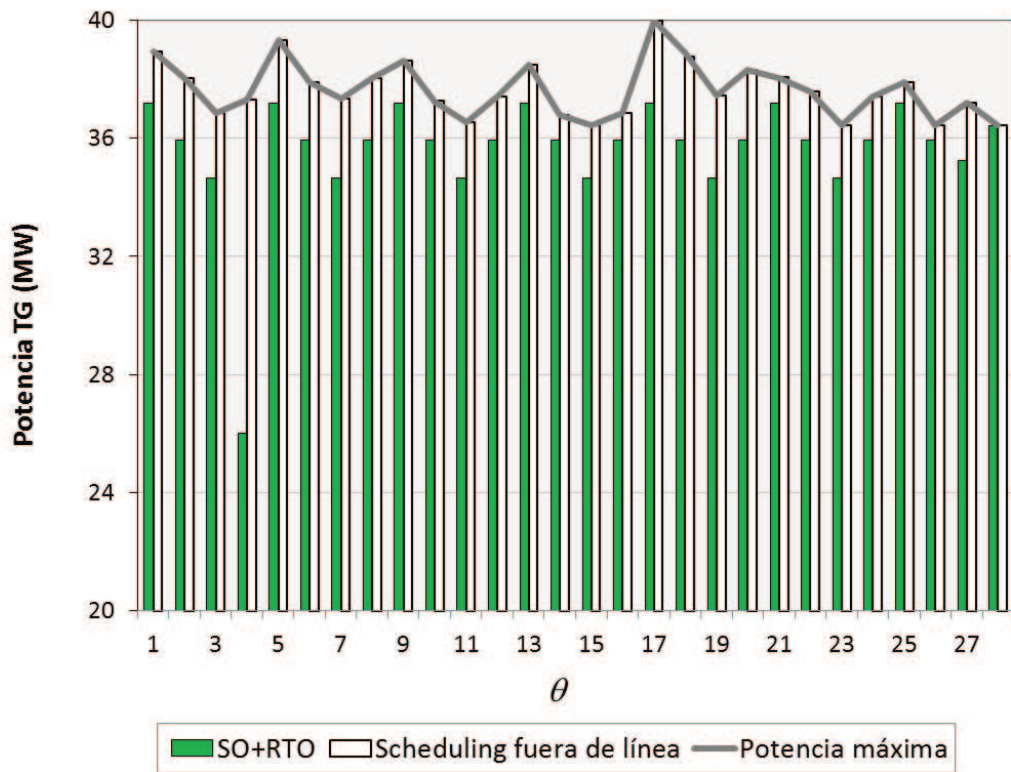


Figura 5-10. Producción de electricidad en la turbina de gas y potencia máxima.

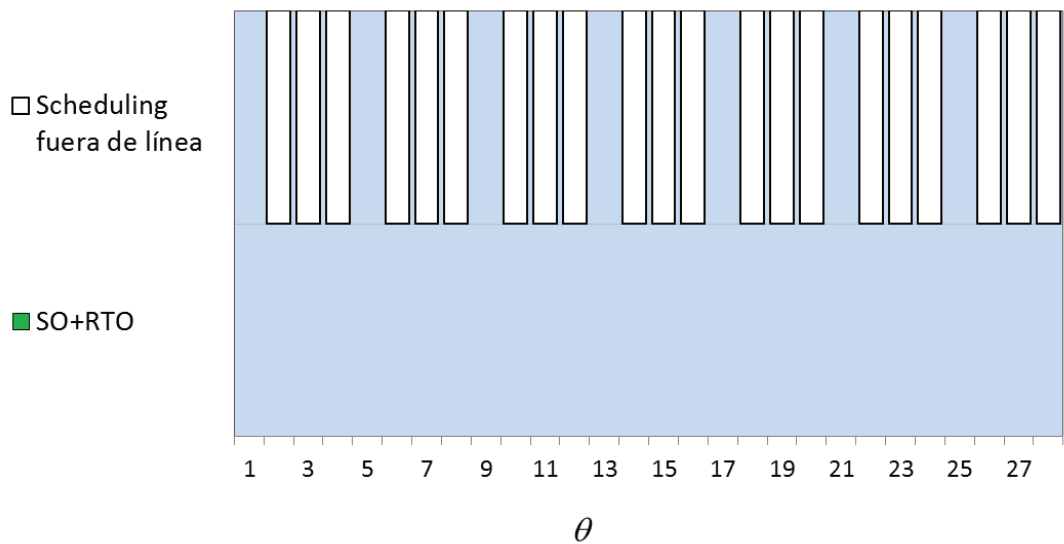


Figura 5-11. Uso de la postcombustión por período para cada estrategia.

## CONCLUSIONES PARCIALES

En este capítulo se ha presentado un marco metodológico para la integración del *scheduling* óptimo y la optimización en tiempo real de procesos continuos (*SO+RTO*). Es particularmente aplicable y útil para la mejora continua de la operación de sistemas de calor y potencia, aunque puede extenderse a otros procesos o sistemas.

Las metodologías propuestas aprovechan el mayor número de grados de libertad de una formulación de *scheduling* óptimo, haciendo uso al mismo tiempo de las propiedades de *feedback* y el menor costo computacional de la RTO. Las variables de decisión con tiempos de transición más largos se fijan en la etapa de *scheduling*, mientras que el resto de los grados de libertad se fijan con una frecuencia mayor a través de un sistema de RTO multiperíodo. Excepto por la adaptación del modelo a través de las mediciones disponibles en RTO, los modelos MINLP utilizados en los dos niveles de la estrategia de solución son los mismos.

El caso de estudio presentado optimiza la operación de 7 días de un sistema de generación de calor y potencia. La estrategia *SO+RTO* implementada utiliza un horizonte fijo para los dos niveles; para la RTO, el horizonte está al final de cada día, y se renueva al comenzar el día siguiente. Los resultados muestran una mejora significativa en el costo operativo y el cumplimiento de las restricciones, en comparación con la aplicación directa del *schedule* óptimo completo, obtenido fuera de línea (sin corrección por las mediciones de la planta).

Como los problemas a resolver en el nivel de RTO pueden presentar una complejidad importante, es necesario generar estrategias para reaccionar ante la ocurrencia de una solución no factible o a la ausencia de solución en los plazos requeridos para la implementación en tiempo real. Se ha propuesto una metodología para obtener una solución cuando el problema no converge o es infactible. La problemática del aprovechamiento del tiempo y la reacción a las soluciones no factibles se aborda con mayor detalle en el capítulo 6 de esta tesis.

## 6. USO EFICIENTE DEL TIEMPO EN RTO

### RESUMEN

El capítulo 6 presenta estrategias para el aprovechamiento del tiempo y los recursos computacionales disponibles en un sistema de RTO. La sección 6.1 introduce la problemática. Las limitaciones presentes al resolver en tiempo real los problemas de optimización se enumeran en la sección 6.2. La sección 6.3.1 presenta estrategias para utilizar el tiempo disponible para generar una solución. En la sección 6.3.2 se propone el uso del tiempo *ocioso* que transcurre entre el fin de la optimización y el comienzo de un nuevo ciclo.

### 6.1. INTRODUCCIÓN

La optimización en tiempo real requiere que las entradas óptimas a aplicar en la planta se obtengan en un plazo relativamente rápido, para que el modelo adaptado y las restricciones incorporadas al modelo sean válidos, y el sistema pueda responder en forma eficaz a las *disturbancias* (en precios, condiciones ambientales y variables operativas no medidas) que afectan al proceso.

Las limitaciones en el tiempo de solución dependen del tamaño del modelo y de la frecuencia de ejecución de los ciclos de RTO. Darby et al. (2011) mencionan tiempos aceptables de solución de hasta 40 minutos para un problema de  $3 \times 10^5$  variables y 40 grados de libertad, incluyendo la actualización de los parámetros del modelo y la resolución del modelo; en muchos casos el tiempo debe ser menor.

En un ciclo de RTO, el paso que consume más tiempo de cálculo es la optimización del costo con el modelo adaptado. La inclusión de variables discretas y la formulación multiperíodo propuestas en los capítulos 4 y 5 aumentan la complejidad de los problemas de optimización con respecto a la formulación usual, (continua y de un solo período). En estos casos, dependiendo del tiempo aceptable para alcanzar una solución, podría ser necesario reducir el grado de detalle en el modelo para permitir la inclusión de variables discretas y múltiples períodos. El número de períodos a incluir en la RTO también estará influenciada por el tiempo máximo aceptable.

Sin embargo, el tiempo, que generalmente es visto como una limitación en el contexto de la RTO, también puede representar una oportunidad. El *tiempo máximo admisible* para obtener una solución también puede ser visto como *tiempo disponible* para obtener uno o más resultados (Ruiz et al., 2005) o para generar nueva información. Los algoritmos de procesamiento en paralelo (que han

evolucionado notablemente en los últimos años) pueden ser de utilidad para obtener distintas soluciones factibles. Algunas estrategias para este aprovechamiento del tiempo disponible se proponen en la sección 6.3.1.

Por otra parte, el tiempo para alcanzar la solución es generalmente una pequeña parte de la duración de un ciclo de RTO. El resto del tiempo consiste en la implementación de la solución a través del sistema de control de la planta, y en la espera del comienzo de un nuevo ciclo. Si los recursos computacionales asignados a la RTO permanecen ociosos en todo este tiempo entre ejecuciones, pueden utilizarse para obtener nuevas soluciones de optimización, o información para adaptar el modelo. La sección 6.3.2 incluye algunas propuestas para el uso del *tiempo ocioso* entre ejecuciones de RTO.

La siguiente sección (6.2) enumera algunos problemas frecuentes en la optimización en tiempo real, que pueden abordarse total o parcialmente mediante las estrategias de la sección 6.3.

## 6.2. PROBLEMAS FRECUENTES EN LA OPTIMIZACIÓN EN SISTEMAS DE RTO

### 6.2.1. Soluciones no factibles

La resolución de un problema de optimización puede finalizar con una *solución no factible*, por dos motivos: o bien el problema es efectivamente no factible, o bien el algoritmo utilizado no ha conseguido encontrar una solución que cumpla con todas las restricciones impuestas por el modelo. En ambos casos, la ausencia de solución supone un problema para un sistema de RTO. La reacción más sencilla a esta inconveniente es mantener las entradas óptimas del ciclo anterior; para la estrategia *SO+RTO*, en la sección 5.4.3 de esta tesis se ha propuesto una metodología para obtener una solución a través de una reducción del tamaño del problema.

En muchos casos, la convergencia de los algoritmos de optimización a una solución factible depende de los valores iniciales de las variables. Al resolver un problema de optimización *fuera de línea*, es posible calcular valores iniciales para el modelo a optimizar, y resolver el problema con distintos conjuntos de valores iniciales si la solución no es factible. En la optimización *en línea* (RTO), los valores iniciales apropiados en un momento pueden no serlo en otro, ya que cada ciclo de RTO corresponde a un problema diferente (con un modelo distinto y condiciones de operación que cambian). Una posible inicialización podría utilizar el punto de operación actual; dependiendo de las restricciones del modelo, incluso este punto podría ser infactible y hacer que la convergencia sea complicada.

En otros casos puede ocurrir que no exista una solución factible para las restricciones del modelo. Las estrategias de adaptación del modelo modifican la región factible del problema de optimización. Las restricciones adicionales impuestas por las estrategias de RTO dual (Marchetti et al., 2010) pueden excluir a la región factible original del problema; cuando esto ocurre, es posible reformular estas restricciones aceptando un error mayor en la estimación del gradiente, pero esto implica la resolución de un nuevo problema de optimización, lo que podría ser impracticable en el tiempo disponible.

### 6.2.2. Límite de tiempo

Es posible que, en algunos ciclos de RTO, el algoritmo de optimización no finalice en el *tiempo máximo aceptable* (aunque probablemente llegaría al óptimo en un tiempo mayor). Para lograr menores tiempos de convergencia, también pueden ser importantes los valores iniciales de las variables.

Si se alcanza el tiempo máximo aceptable (o el número máximo de iteraciones del algoritmo), el algoritmo puede presentar la mejor solución factible encontrada durante el cálculo. Sin embargo, en algunos casos podría no haber una solución factible intermedia. Esto puede deberse a que el problema se inició en un punto infactible y no se consiguió resolver la violación de las restricciones. En los casos con variables discretas, algunos resolvers como SBB (GAMS Development Corporation, 2008) se inician con una relajación de las variables enteras, y gran parte del tiempo de solución podría transcurrir sin que se encuentre una solución factible (es decir, una solución en la que las variables discretas tomen valores enteros y se cumplan las restricciones del modelo).

De cualquier manera, las soluciones factibles intermedias encontradas serán muy probablemente subóptimas, por lo que no se consigue todo el beneficio posible de la RTO. En algunos casos, es posible comparar el valor de la función objetivo en la solución intermedia propuesta con el valor actual (evaluado en las condiciones de la planta al iniciar la optimización) (Ruiz et al., 2005). Si la solución intermedia mejora la situación actual, puede aplicarse en la planta.

### 6.2.3. Óptimos locales

Con frecuencia, los modelos rigurosos de los procesos industriales abordados por la RTO son no convexos (debido, por ejemplo, a los balances de energía y de materia). Aun cuando el modelo nominal (original) sea convexo, si se utiliza la estrategia de adaptación de dos pasos, los parámetros ajustados del modelo podrían generar no convexidades (esto no sucede en la adaptación por modificadores tradicional, ya que sólo agrega términos constantes y lineales).

Si bien el objetivo de la RTO es mejorar la operación de la planta, y en este sentido la convergencia a un óptimo local puede ser suficiente, es siempre deseable encontrar el óptimo global, que corresponde a la mejor forma posible de operar. Aunque la optimización global es un área que aún está en desarrollo, actualmente existen algoritmos capaces de resolver problemas no convexos a optimalidad global (Floudas et al., 2005). Sin embargo, los tiempos de solución requeridos pueden ser prohibitivos para una implementación de RTO. De todos modos, existen aplicaciones industriales de optimización en tiempo real que resuelven modelos MINLP no convexos con un resolvidor global (Puranik et al., 2013).

#### 6.2.4. Algoritmos de optimización

En los capítulos anteriores de esta tesis se han utilizado en todo momento resolvidores que aplican métodos de programación matemática (LP, NLP, MINLP y GDP). Para cada tipo de problemas, existe una gran variedad de *solvers* disponibles (GAMS Development Corporation, 2008).

Para un problema particular, cada *solver* puede comportarse de manera diferente en cuanto al tiempo de convergencia, el punto *óptimo* alcanzado y la factibilidad o no de la solución. Además, cada algoritmo contiene parámetros ajustables para el escalado de las variables, los criterios de convergencia, las estrategias de exploración de nodos, y otras variantes; los valores óptimos de estos parámetros deben encontrarse por prueba y error. En una aplicación *fuera de línea*, el problema puede resolverse con diferentes conjuntos de parámetros y con distintos *solvers* hasta encontrar la mejor solución; en la optimización en tiempo real, por el contrario, la selección del *solver* suele realizarse al diseñar el sistema, y los parámetros del algoritmo se ajustan rara vez, si se observa que aumenta la ocurrencia de soluciones no factibles.

Por otra parte, además de las técnicas de programación matemática, hoy en día pueden utilizarse otros algoritmos de optimización *metaheurísticos* tales como algoritmos genéticos, *simulated annealing*, optimización de enjambre de partículas (*particle swarm optimization*) y búsqueda tabú (Floudas y Pardalos, 2008). Si bien estas técnicas pueden no ser rigurosas y no garantizar la optimalidad (local o global), para algunos problemas pueden alcanzar un mejor resultado que las técnicas de programación matemática.

Además, en la planta industrial donde se aplica la RTO puede haber reglas heurísticas de operación, que permiten estimar rápidamente una solución que frecuentemente estará cerca de la óptima. Estas reglas pueden obtenerse de la experiencia de los operadores o desarrollarse algorítmicamente a través de

aprendizaje por refuerzos (Palombarini y Martínez, 2012). Si bien estas reglas no garantizan la operación óptima, pueden utilizarse para inicializar el problema de optimización, o para contrastarlas con el *óptimo* alcanzado al resolver el problema (evaluando el valor de la función objetivo alcanzado al aplicar las reglas, utilizando el modelo de RTO adaptado).

### 6.3. ESTRATEGIAS PARA EL APROVECHAMIENTO DEL TIEMPO Y LOS RECURSOS DISPONIBLES

#### 6.3.1. Uso del tiempo máximo disponible para obtener la solución

##### 6.3.1.1. Generación de soluciones candidatas

Si el *tiempo máximo disponible* para generar la solución de RTO se utiliza en su totalidad, es posible resolver más de un problema de optimización, de manera de generar un conjunto de soluciones factibles *candidatas* a ser aplicadas en la planta. Estas soluciones pueden obtenerse a través de:

- *Inicialización desde diferentes puntos operativos*: al usar diferentes valores iniciales para la optimización, aumenta la probabilidad de encontrar una solución factible y que encuentre el óptimo en el tiempo dado. Además, si el problema presenta óptimos locales, las soluciones obtenidas pueden ser diferentes y se selecciona la mejor.
- *Contracciones de la región factible*: el problema original puede modificarse fijando los valores de algunas variables o agregando restricciones (*contracción*). Un ejemplo es la primera reacción a las infactibilidades propuesta en la sección 5.4.3: ante una solución no factible, se fijan los valores de las variables discretas, transformando un problema MINLP en NLP. Si bien se espera que el costo alcanzado a través de una contracción sea mayor o igual al obtenido con el problema original, si hay óptimos locales puede ocurrir que el algoritmo converja a un mejor valor de la función objetivo. Como la región factible contraída está incluida en la región factible original, el resultado alcanzado puede aplicarse en la planta si es conveniente.
- *Relajaciones de la región factible*: si se tolera, al menos temporalmente, la violación de ciertas restricciones, éstas pueden eliminarse o flexibilizarse temporalmente (*relajación*). También pueden transformarse en un término de penalidad en la función objetivo. La segunda reacción a las infactibilidades discutida en la sección 5.4.3 es un ejemplo, aunque no directo, de una relajación: se transforma el

problema multiperíodo en uno de período único, eliminando las restricciones relativas a los períodos futuros y las que relacionan el período actual con los futuros. Como la relajación puede generar puntos fuera de la región factible inicial, es de esperar que el costo alcanzado sea menor o igual al del problema original, pero probablemente la solución sea no factible para las restricciones originales. Los resultados del problema relajado pueden implementarse si el óptimo está dentro de la región factible inicial, o si hay una mejora significativa en la función objetivo y la violación a las restricciones del problema original puede considerarse aceptable para la operación del proceso. En caso contrario, el resultado no puede considerarse una solución candidata.

- *Cambios en los parámetros del algoritmo de optimización:* el mismo problema puede resolverse con distintos conjuntos de parámetros del algoritmo de optimización (criterios de convergencia, estrategias de exploración de nodos, escalado, entre otros). Las soluciones factibles alcanzadas son candidatas a implementarse en la planta.
- *Uso de varios algoritmos de resolución:* diferentes algoritmos (de programación matemática o metaheurísticos) pueden utilizarse para resolver el problema de RTO y generar soluciones candidatas.
- *Empleo de reglas de operación:* las reglas de operación de la planta, obtenidas de la experiencia o con técnicas de aprendizaje de inteligencia artificial, también pueden ser fuente de soluciones candidatas a ser aplicadas en la planta. El valor de la función objetivo y la factibilidad de las soluciones candidatas se evalúan a través del modelo de RTO adaptado.
- *Uso de diferentes modelos:* además del modelo original (adaptado con las mediciones de la planta), es posible resolver modelos simplificados para buscar el óptimo y generar soluciones candidatas. Una vez más, el modelo de RTO adaptado se utiliza para evaluar la función objetivo y la factibilidad de las soluciones propuestas por el modelo simplificado.

En el ciclo de RTO  $k$ , las entradas *óptimas* calculadas  $\mathbf{u}^{*,k+1}$  se obtienen a partir del conjunto de soluciones candidatas  $S^k = \{\mathbf{u}^k, \mathbf{u}^{*,k+1,1}, \dots, \mathbf{u}^{*,k+1,ns}\}$  (donde  $ns$  es el número de soluciones candidatas) como:

$$\begin{aligned} \mathbf{u}^{*,k+1} &= \underset{\mathbf{u} \in S^k}{\operatorname{argmin}} Q(\mathbf{u}) \\ \text{s.a. } &\mathbf{g}(\mathbf{u}) \leq \mathbf{0} \end{aligned} \quad (6-1)$$

donde  $Q$  y  $\mathbf{g}$  son el costo y las restricciones del modelo original de RTO (adaptados con las mediciones del ciclo  $k$ ). En el conjunto  $S^k$  se incluye también el valor actual de las entradas  $\mathbf{u}^k$ .

Alternativamente, si se pueden tolerar violaciones a las restricciones (por ejemplo, para relajaciones de la región factible, las restricciones *blandas* pueden considerarse como una penalización en el costo de la siguiente manera:

$$\begin{aligned} \mathbf{u}^{*,k+1} &= \underset{\mathbf{u} \in S^k}{\operatorname{argmin}} Q(\mathbf{u}) + \sum_{i \in G_B} \mu_i \cdot \max(0, (g_i(\mathbf{u}))^2) \\ \text{s.a. } &g_j(\mathbf{u}) \leq 0, j \in G_D \\ &g_i(\mathbf{u}) \leq \xi_i, i \in G_B \end{aligned} \quad (6-2)$$

donde  $\mu_i$  es la penalidad asignada a la violación de la restricción  $g_i(\mathbf{u})$ .  $G_D$  es el conjunto de restricciones *duras*, que no toleran violaciones, y  $G_B$  el de restricciones *blandas*. Cuanto mayor es la penalidad  $\mu_i$ , menor será la violación tolerada. Además, se puede fijar un valor máximo a la violación a través de la constante (positiva)  $\xi_i$ , de manera tal de no aceptar soluciones candidatas que no cumplan este valor máximo. La ecuación (6-2) muestra una penalidad cuadrática; en realidad, puede seleccionarse otra forma (por ejemplo, lineal), que puede ser diferente para cada restricción  $g_i$ ,  $i \in G_B$ .

La solución seleccionada  $\mathbf{u}^{*,k+1}$  puede estar sujeta a la validación de resultados o al filtrado de las entradas óptimas (presentados en capítulos anteriores de esta tesis) antes de ser implementada en la planta.

### 6.3.1.2. Resolución en serie

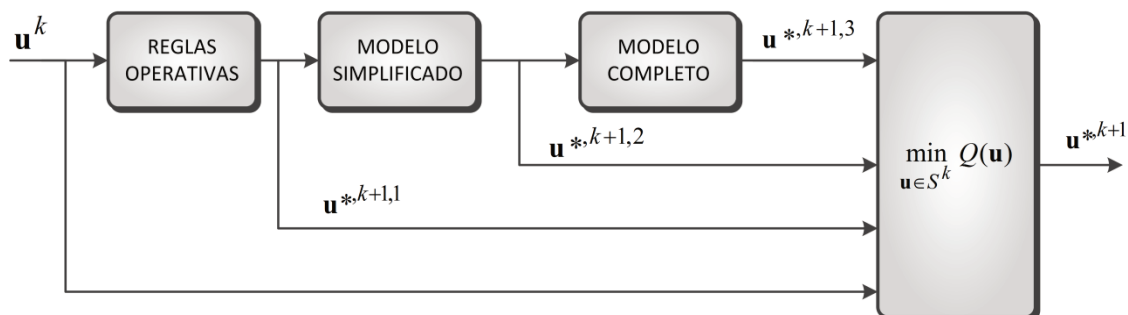
Las soluciones candidatas pueden generarse en serie: una vez que finaliza un algoritmo, se da comienzo a la resolución del siguiente. De esta manera, si la computadora utilizada y los algoritmos tienen posibilidades de procesamiento en paralelo, esta capacidad se aprovecha en su totalidad. Al alcanzarse el tiempo máximo disponible, si alguno de los algoritmos está en ejecución, se interrumpe, y devolverá como resultado la mejor solución factible que tenga disponible.

La resolución en serie puede utilizarse para inicializar el problema de RTO, y a su vez generar soluciones intermedias. Por ejemplo, pueden utilizarse las entradas actuales para inicializar las reglas de operación; estas reglas pueden inicializar un modelo simplificado, y el modelo simplificado inicializar el modelo

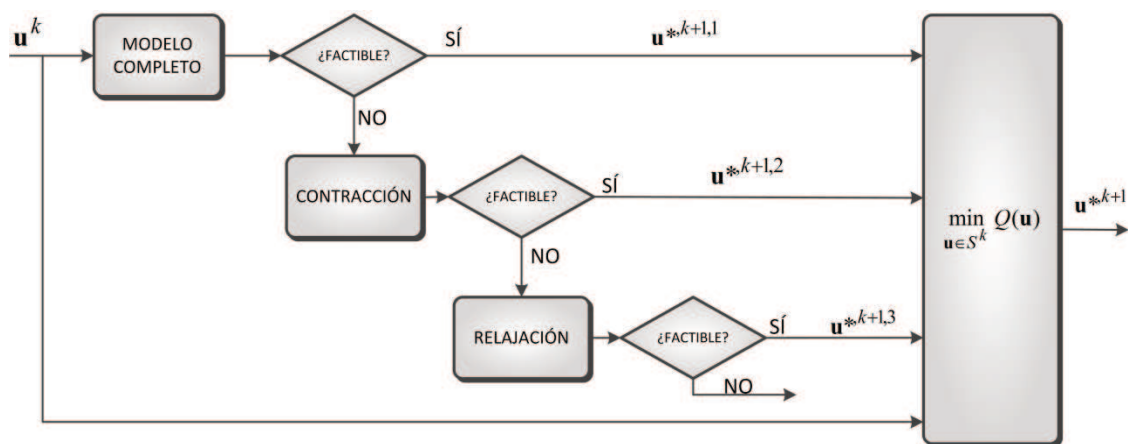
completo (riguroso). Todas las soluciones intermedias son candidatas. Este ejemplo se muestra en la **Figura 6-1**.

También puede utilizarse como una reacción a las soluciones no factibles, como se ha presentado en la sección 5.4.4: se resuelve el modelo original, y si esta solución no es factible, se incluyen contracciones o relajaciones de la región factible, o simplificaciones del modelo. En este caso, el tiempo máximo de solución para el modelo original debe ser menor al tiempo máximo disponible, con el fin de permitir la resolución de los siguientes problemas (que, al ser menos complejos, es de esperar que converjan en un tiempo menor). La estructura propuesta se ilustra en la **Figura 6-2**.

Si hay tiempo disponible, el esquema anterior puede implementarse incluso cuando la solución del problema original es factible, sobre todo si el algoritmo no garantiza convergencia al óptimo global. Las soluciones candidatas generadas por los problemas menos complejos podrían alcanzar un mejor valor de la función objetivo que la solución candidata original.



**Figura 6-1.** Resolución en serie – Inicialización.



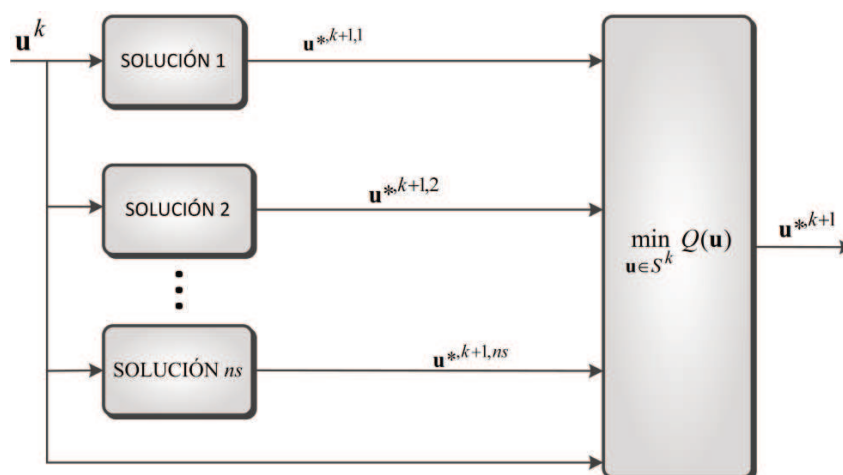
**Figura 6-2.** Resolución en serie – Reacción a las infactibilidades (ejemplo).

Finalmente, si los modelos de inicialización consumen poco tiempo de CPU, es posible implementar en serie las dos estructuras anteriores, comenzando

por la inicialización, siguiendo por el modelo completo y terminando con la reacción a las infactibilidades.

### 6.3.1.3. Resolución en paralelo

Si el procesador utilizado admite el procesamiento en paralelo, es posible generar soluciones candidatas resolviendo simultáneamente el mismo problema con diferentes algoritmos de optimización (de programación matemática o metaheurísticos); o bien, resolviendo diferentes problemas (contracciones, relajaciones o simplificaciones del problema original). Una posible estructura se muestra en la **Figura 6-3**.



**Figura 6-3.** Resolución en paralelo.

Además, es posible ejecutar un algoritmo de optimización global entre los procesos en paralelo. Es posible que no alcance a converger en el tiempo disponible; sin embargo, si el algoritmo obtiene soluciones factibles intermedias, serán soluciones candidatas que podrían alcanzar un menor costo que aquellas de un optimizador local o un metaheurístico.

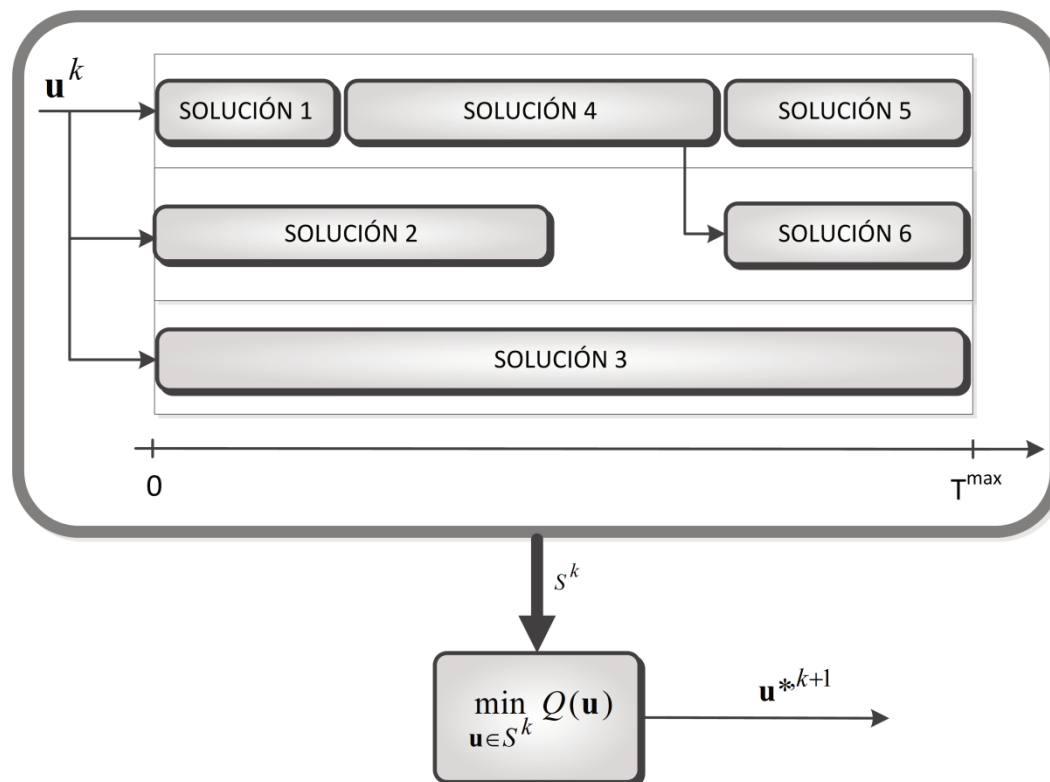
Esta estrategia no excluye el uso de algoritmos con capacidad de procesamiento en paralelo; si se cuenta con un servidor de gran capacidad, con una cantidad suficiente de núcleos de procesamiento, es posible resolver al mismo tiempo un cierto número de problemas, con cada algoritmo ejecutando a su vez procesos en paralelo.

### 6.3.1.4. Combinación serie-paralelo

Las estrategias de las secciones 6.3.1.2 y 6.3.1.3 pueden combinarse, resolviendo problemas en serie y en paralelo, y estableciendo un orden de prioridades en la ejecución de cada algoritmo. Al finalizar el tiempo máximo disponible, se contará con todas las soluciones candidatas que se hayan generado.

Si un algoritmo o un modelo no se ha llegado a ejecutar, simplemente no aportará una solución candidata al conjunto  $S^k$ .

La **Figura 6-4** ilustra la estructura serie-paralelo. En la figura, las soluciones 5 y 6 dependen de la solución 4, por lo que comienzan una vez que la solución 4 ha convergido. Los procesos encargados de proporcionar las soluciones 3, 5 y 6 están en ejecución cuando se alcanza el tiempo máximo  $T^{\max}$ , por lo que aportarán la mejor solución factible que se haya encontrado durante la ejecución del algoritmo.



**Figura 6-4.** Combinación serie-paralelo.

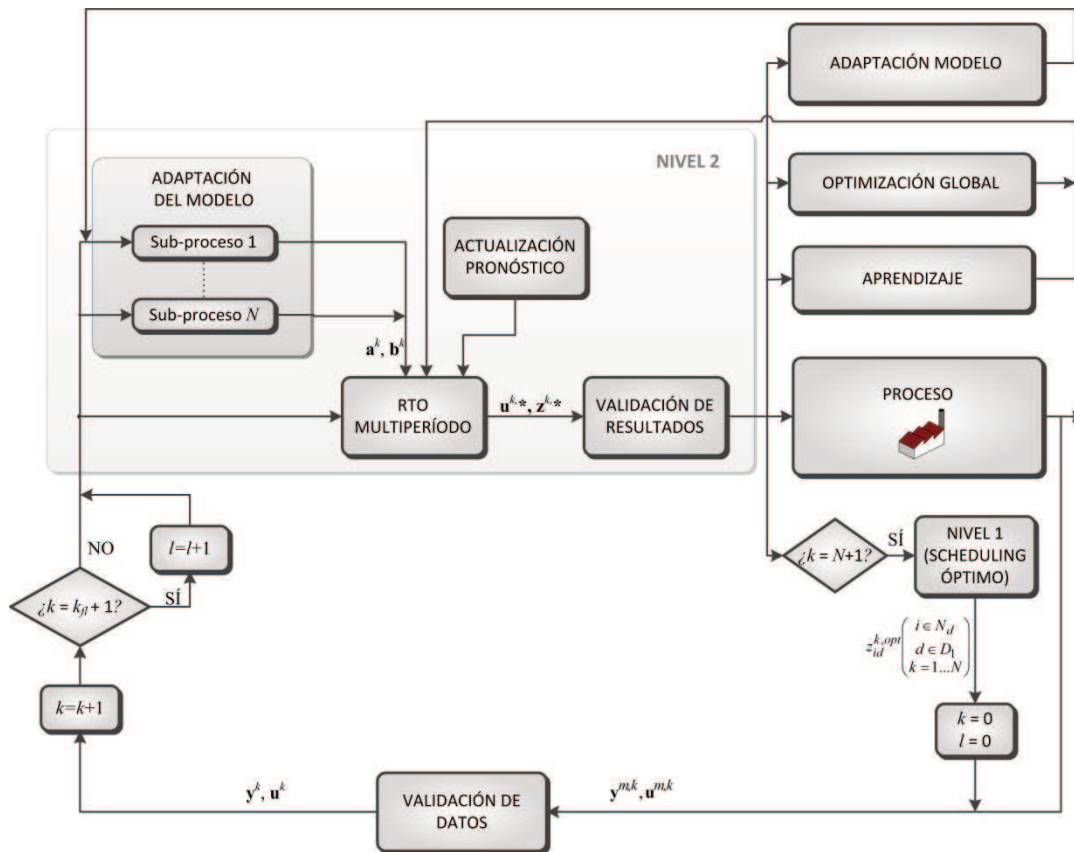
### 6.3.2. Uso del tiempo entre ejecuciones

Si luego de exportar las nuevas entradas  $\mathbf{u}^{k+1}$ , y hasta el inicio de un nuevo proceso de validación-adaptación-optimización, existe capacidad de procesamiento ociosa, es posible aprovecharla para ejecutar tareas útiles para la optimización en tiempo real. Estas tareas pueden consistir en:

- *Optimización global:* el problema de optimización puede resolverse con un *solver* global (que podría requerir un tiempo de CPU mucho mayor al tiempo máximo disponible para RTO). Los resultados alcanzados pueden utilizarse como inicialización de la optimización en el ciclo siguiente, o simplemente como soluciones candidatas si las condiciones de la planta no han cambiado significativamente.

- *Aprendizaje – generación de reglas*: los resultados de la optimización del ciclo anterior, combinados con resultados pasados, pueden usarse para actualizar las reglas de operación (o generar nuevas reglas), utilizando técnicas de inteligencia artificial como el aprendizaje por refuerzos (Palombarini y Martínez, 2012). Estas reglas, como se mencionó en secciones anteriores, pueden utilizarse para generar soluciones candidatas o para inicializar los algoritmos de optimización.
- *Finalización de soluciones inconclusas*: los problemas de optimización que no pudieron finalizarse en el tiempo máximo disponible, pueden continuarse durante el tiempo entre ciclos. También puede modificarse el criterio de finalización para el algoritmo (por ejemplo, criterio de óptimo de un algoritmo MINLP, número máximo de iteraciones) para verificar si se genera una nueva solución que puede utilizarse como inicialización del nuevo ciclo.
- *Adaptación del modelo*: las mediciones obtenidas durante el tiempo entre ciclos pueden utilizarse para actualizar los parámetros o los modificadores del modelo. Puede ser necesario recurrir a modelos dinámicos para estimar los parámetros del modelo estacionario (Mansour y Ellis, 2003; Srinivasan et al., 2011).
- *Scheduling óptimo*: el nivel 1 de la estrategia *SO+RTO* presentada en el capítulo 5 requiere la solución de un problema de *scheduling* óptimo que puede ser complejo. Este problema puede resolverse luego de la última ejecución de RTO en el período final  $N$ , para fijar los valores de las variables discretas en el nuevo horizonte,  $z_{id}^{\theta}, i \in N_d, d \in D_1, \theta = 1, \dots, N$ .

La **Figura 6-5** muestra un diagrama de la estrategia *SO+RTO* (basado en la **Figura 5-1**) que incluye el aprovechamiento del tiempo ocioso entre ciclos. Las tareas adicionales (adaptación, optimización global, aprendizaje y *scheduling* óptimo) se realizan mientras se implementan en la planta las entradas óptimas calculadas por la RTO multiperíodo. La optimización global y el desarrollo de reglas (*aprendizaje*) se utilizan para inicializar el o los problemas de optimización del siguiente ciclo. El paso de RTO multiperíodo, que se muestra como un solo bloque, incluye las estrategias de uso del tiempo máximo disponible para optimización presentadas en la sección 6.3.1.



**Figura 6-5.** Estructura SO+RTO con aprovechamiento del tiempo ocioso entre ejecuciones de RTO.

## CONCLUSIONES PARCIALES

El capítulo 6 presenta estrategias para aprovechar el tiempo y la capacidad de procesamiento disponibles en un sistema de RTO. El *tiempo máximo admisible* para obtener una solución se toma como *tiempo disponible* para generar una serie de *soluciones candidatas*, a través del uso de diferentes modelos, algoritmos de resolución y puntos iniciales. De estas soluciones, se escoge aquella con mejor valor de la función objetivo y se aplica en la planta.

Además, el tiempo entre ejecuciones de RTO puede utilizarse para resolver algoritmos de optimización global, generar reglas heurísticas de operación o resolver problemas de *scheduling* óptimo, tareas que por su complejidad no pueden abordarse en el tiempo disponible para la optimización en tiempo real. Además, pueden utilizarse las mediciones intermedias para adaptar el modelo utilizando modelos dinámicos.

Estas estrategias se han presentado desde un enfoque conceptual y metodológico, y de manera simplificada. Para avanzar en el aprovechamiento del tiempo y los recursos de CPU, es necesario implementar estas estrategias en forma

experimental, de manera de evaluar su factibilidad práctica en la industria y sus potenciales beneficios.



## 7. CONCLUSIONES Y LÍNEAS DE TRABAJO A FUTURO

### 7.1. CONCLUSIONES

Este trabajo de tesis ha explorado los potenciales usos de la optimización en tiempo real basada en modelos de estado estacionario, con énfasis en la estrategia de adaptación por modificadores. Si bien las estrategias y las técnicas propuestas están orientadas a la optimización de sistemas de energía, la formulación y las metodología son lo suficientemente generales para su aplicación a otros procesos industriales.

A lo largo de la tesis se ha propuesto una evolución partiendo de la formulación tradicional con un modelo no lineal, continuo y de un solo período, a través de la incorporación progresiva de decisiones discretas (con modelos de programación disyuntiva) y de formulaciones multiperíodo que facilitan la integración de las técnicas de RTO con el scheduling óptimo.

Cada una de las formulaciones se ha aplicado a modelos de sistemas de generación de calor y potencia, con el fin de mostrar la utilidad de la RTO en estos sistemas de energía. Estos sistemas son apropiados para la optimización en tiempo real con modelos estacionarios por su (frecuente) disponibilidad grados de libertad, su dinámica rápida, y la necesidad de responder a cambios en la demanda de potencia y calor, precios y condiciones ambientales, teniendo en cuenta restricciones de capacidad, operativas y ambientales.

La estructura particular de ciertos procesos (como muchos sistemas de energía o, en general, sistemas con redes de equipos o unidades interconectados) permitió generar formulaciones alternativas para la estrategia de adaptación por modificadores, a las que se dio el nombre de adaptación por modificadores basada en ecuaciones de eficiencia. Estas formulaciones, en las que la corrección del gradiente se realiza en función de variables auxiliares (en lugar de las variables de entrada) permiten reducir la cantidad de datos necesarios para la estimación del gradiente. Además, al utilizar el conocimiento del sistema para determinar las direcciones del espacio de las entradas en las que el gradiente es nulo, permiten eliminar algunos de los errores ocasionados por el ruido en las mediciones. Por este motivo, las estrategias propuestas aumentan la posibilidad de implementar la adaptación por modificadores (con corrección del gradiente) en sistemas con un gran número de variables de entrada.

Se presentaron dos variantes para la implementación de la adaptación basada en ecuaciones de eficiencia: en la primera, la estimación numérica del

gradiente se combina con un filtro exponencial (como las estrategias de RTO dual con adaptación por modificadores tradicionales). En la segunda, los modificadores del gradiente se estiman por regresión lineal, y se incluye una corrección cuadrática.

Las dos variantes se ilustraron a través de casos de estudio, que muestran la convergencia al óptimo real de la planta en ausencia de error en las mediciones, y el buen desempeño de las alternativas propuestas en comparación con otras estrategias cuando las mediciones están afectadas por ruido. De cualquier manera, la presencia de ruido y perturbaciones no medidas en el proceso continúa afectando significativamente el desempeño de los sistemas de RTO.

Si bien existen implementaciones industriales de RTO que involucran variables binarias, en este trabajo se ha abordado formalmente la inclusión de decisiones discretas en la optimización en tiempo real, a través de una formulación basada en programación disyuntiva. Este enfoque incluyó la reformulación de las estrategias de adaptación (incluyendo la adaptación por modificadores basada en ecuaciones de eficiencia) para obtener una técnica apropiada para la RTO con disyunciones.

Por otra parte, se ha mostrado que, además de las decisiones de puesta en marcha/parada de equipos, existen otros aspectos o fenómenos relacionados con los procesos, con las técnicas de modelado y con los sistemas de control que requieren una formulación con disyunciones. Estos fenómenos pueden incluirse en una implementación de RTO en *lazo cerrado* (implementada directamente en el proceso a través del sistema de control), mientras que las puestas en marcha y paradas suelen requerir la acción manual de los operadores de planta. Además, los fenómenos discretos presentados no presentan inconvenientes a la optimización con una formulación de un solo período; la decisión de parar o arrancar un equipo, en cambio, está sujeta generalmente a restricciones temporales, por lo que se aborda de una manera más apropiada a través de una formulación multiperíodo.

Algunas técnicas usuales en la RTO con modelos continuos (tales como el uso de términos de *convexificación*, el filtrado de las entradas óptimas o la imposición de cambios máximos permitidos en las variables) pueden presentar problemas si se aplican directamente a un sistema de RTO con disyunciones. En esta tesis, estas técnicas se han reformulado para hacer factible su uso cuando hay decisiones discretas involucradas. Sin embargo, es necesario avanzar en el desarrollo de estas reformulaciones: por ejemplo, el filtrado de las entradas óptimas no es de utilidad para equipos que presentan sólo dos puntos de operación posibles (como una bomba que puede estar parada u operar a una

potencia determinada), ya que no existen puntos factibles intermedios para aplicar el filtro cuando el sistema propone cambiar de punto de operación.

A través de casos de estudio, se ha mostrado el comportamiento de sistemas de RTO con disyunciones, y se ha comprobado la utilidad y la factibilidad de los métodos con corrección del gradiente.

La integración de la RTO con los demás sistemas de toma de decisiones (*scheduling*, control) es clave para la operación eficiente de los procesos. Otros autores ya han desarrollado variadas estrategias de integración de RTO con control, por lo que este trabajo ha abordado la integración con el *scheduling* óptimo, a través de una formulación de RTO multiperíodo, y una resolución en dos niveles (SO+RTO) que establece una jerarquía de variables discretas: algunas se fijan en un problema de *scheduling* óptimo inicial (resuelto para un horizonte dado), y el resto, junto con las variables continuas, en subproblemas de RTO (resueltos en tiempo real).

Para fijar el número de períodos considerados por cada subproblema de RTO, se debe buscar un compromiso entre la complejidad del problema y la necesidad de resolverlo en un tiempo rápido. Al no optimizar la totalidad del horizonte, la descomposición en subproblemas podría no ser óptima; sin embargo, permite responder a las perturbaciones de la planta y utilizar las estrategias de adaptación propias de la RTO para corregir el modelo, sin exceder el tiempo requerido para la operación en tiempo real. Además, al trabajar con horizontes más cortos, la incertidumbre en las demandas, precios y condiciones ambientales es menor que en el *scheduling* del horizonte completo.

Dependiendo del proceso, los subproblemas de RTO pueden plantearse con un horizonte fijo o móvil. Probablemente una formulación con horizonte móvil se adapte a mayor cantidad de casos; en esta tesis se ha planteado una estrategia con horizonte fijo, apropiada para sistemas de energía.

La integración de RTO con *scheduling* (SO+RTO) presenta desafíos no resueltos, y para evolucionar requiere de una investigación más profunda y de la aplicación a una mayor variedad de procesos. La mejor forma de integración dependerá del proceso a optimizar. El tamaño de los problemas a abordar en la etapa de tiempo real (RTO) estará ligada a la evolución en la capacidad de procesamiento y los algoritmos de optimización: en algún momento y para ciertos procesos será factible resolver, a la frecuencia de la RTO, el problema de *scheduling* completo, con un horizonte móvil (SO) y adaptando el modelo con las mediciones del proceso.

En su capítulo final esta tesis ha abordado la problemática del tiempo en la RTO. En la optimización fuera de línea, para obtener una solución es posible inicializar un problema con distintos conjuntos de valores, recurrir a distintos algoritmos de optimización y realizar pruebas agregando restricciones o relajando las restricciones existentes; además, no existen en principio limitaciones al tiempo máximo de cálculo. Por el contrario, en la RTO el resultado debe obtenerse en segundos o pocos minutos, dependiendo de la duración de los ciclos.

Se ha propuesto una metodología para aprovechar el tiempo máximo disponible para obtener una solución. Durante este tiempo, es posible resolver el problema de optimización con distintos algoritmos: programación matemática, metaheurísticos (por ejemplo, algoritmos genéticos) y reglas heurísticas. Además, se puede recurrir a simplificaciones del modelo, y relajaciones y contracciones de la región factible. De esta manera, se genera un conjunto de *soluciones candidatas* a ser implementadas en el siguiente ciclo de RTO.

Los problemas pueden resolverse en serie, con un objetivo de inicialización (resolviendo problemas de complejidad creciente hasta llegar al modelo completo) o de reacción a las soluciones no factibles (partiendo del problema completo y recurriendo a relajaciones, contracciones y simplificaciones hasta obtener una solución factible). También pueden resolverse en paralelo, o en combinaciones serie-paralelo, estableciendo prioridades y dependencias para la ejecución de los algoritmos, hasta que finaliza el tiempo disponible. La cantidad de problemas a resolver y de procesos en paralelo a ejecutar dependerá de la capacidad de cálculo del sistema en el que está implementada la RTO.

Además, se propusieron alternativas para aprovechar el tiempo y los recursos computacionales en el lapso en que se aplican las soluciones óptimas en la planta y se espera al próximo ciclo de optimización (lapso que puede ser mucho mayor al tiempo efectivamente usado para la adaptación del modelo y la optimización). Durante este tiempo es posible resolver el problema de optimización con algoritmos de optimización global; procesar los resultados de ciclos anteriores para generar reglas heurísticas de operación; resolver el schedule completo de la estrategia *SO+RTO*; y utilizar las mediciones para continuar la adaptación del modelo.

Las estrategias de aprovechamiento del tiempo se han planteado en forma conceptual y para generar líneas de trabajo futuras. Para desarrollarlas completamente es necesario implementarlas en sistemas de RTO reales o simulados, de modo de verificar las ventajas descriptas y la factibilidad de la propuesta.

Los avances tecnológicos en el control de procesos y el procesamiento de la información, acompañados del desarrollo de estrategias de optimización, tratamiento de datos y desarrollo de modelos, permitirán una operación más eficiente de los procesos industriales. Para alcanzar este objetivo, es necesario avanzar en la integración de los distintos niveles de toma de decisiones (*scheduling*, RTO, control), con el fin de hacer uso de todo el potencial de las herramientas de optimización y de aprovechar al máximo la tecnología y la información disponible.

## 7.2. LÍNEAS DE TRABAJO FUTURAS

### 7.2.1. Estrategias de adaptación

#### 7.2.1.1. Análisis de convergencia

En esta tesis, las estrategias de adaptación se han formulado para que, *si el sistema converge*, lo haga a un punto que cumple las condiciones de óptimo de la planta. Para completar el planteo de las estrategias, sería deseable establecer las condiciones que *garantizan* que la estrategia converja (Marchetti et al., 2009).

#### 7.2.1.2. Formulación de restricciones de RTO dual

Para la adaptación por modificadores basada en ecuaciones de eficiencia, las restricciones que acotan el error del gradiente en la adaptación por modificadores tradicional podrían no ser útiles. Por este motivo, es necesario desarrollar nuevas restricciones para limitar el error de estimación.

#### 7.2.1.3. Adaptación en base a variables discretas

Al presentar las estrategias de adaptación para RTO con disyunciones, los modificadores del gradiente se incluyeron solo en función de las variables de entrada continuas. Es necesario explorar la necesidad de incluir un modificador similar para las variables discretas (que, si bien podría no estar relacionado con las condiciones de óptimo local, sí podría ser útil para la resolución de relajaciones del problema de optimización y para que el modelo represente con mayor fidelidad la planta real).

#### 7.2.1.4. Adaptación en base a perturbaciones medidas u observables

Además de la adaptación del modelo con respecto a las variables de entrada, podría ser útil actualizar el modelo teniendo en cuenta la funcionalidad de las variables de salida con respecto a las perturbaciones directamente medidas u observables en el proceso. De esta manera, puede mejorarse la respuesta del

sistema de RTO a las perturbaciones y mantener la validez (para el cálculo de los modificadores) de los datos medidos antes de la ocurrencia de la perturbación.

### **7.2.2. Filtrado de las entradas óptimas para RTO con disyunciones**

Las técnicas de filtrado propuestas en la sección 4.5.2 no son de utilidad para equipos con dos puntos de operación factibles (por ejemplo, parado y en marcha a potencia constante), ya que no existen puntos intermedios factibles. Es necesario avanzar en estas técnicas, evaluando la posibilidad de incluir proyecciones (Bunin et al., 2013a).

### **7.2.3. Simulaciones y trabajo experimental**

#### *7.2.3.1. Procesos de mayor escala*

Si bien se ha trabajado con modelos relativamente complejos (cuyo tamaño podría ser similar al de modelos aplicables a procesos industriales) es necesario implementar las estrategias de RTO en modelos más complejos y con mayor cantidad de variables de decisión. Esto permitirá verificar el desempeño de las estrategias de adaptación al aumentar la cantidad de variables de entrada, y analizar el tiempo de cálculo requerido por los modelos de mayor tamaño, sobre todo en RTO con disyunciones y multiperíodo.

#### *7.2.3.2. Modelado dinámico del proceso*

Aun cuando el *modelo RTO* sea de estado estacionario, resulta de interés desarrollar modelos de la *planta real* dinámicos y que incluyan el comportamiento del sistema de control. De esta manera, es posible tener en cuenta los transientes, implementar estrategias de detección de estado estacionario y verificar el funcionamiento de los sistemas en condiciones más cercanas a la realidad.

#### *7.2.3.3. Trabajo experimental*

El desempeño de los sistemas analizados en esta tesis está basado en la comparación con modelos que simulan la operación real de la planta. La implementación de los sistemas en un laboratorio o en un proceso industrial real permitirá validar los resultados obtenidos y detectar potenciales puntos de mejora de las propuestas.

## 7.2.4. Integración de sistemas

### 7.2.4.1. Integración entre métodos de RTO

Los métodos de adaptación por modificadores propuestos pueden integrarse con estrategias de adaptación directa de las entradas ejecutadas a una frecuencia mayor, como han sugerido Chachuat et al. (2009). De esta manera, la adaptación por modificadores permite identificar las restricciones activas, y la adaptación directa de las entradas mantiene el sistema en una operación *óptima* asumiendo constante el conjunto de restricciones activas. También puede implementarse para RTO con disyunciones: las disyunciones se abordan en la RTO con adaptación por modificadores, mientras que la adaptación directa de las entradas opera a valores constantes de las variables discretas.

### 7.2.4.2. Integración SO+RTO+control

Como se ha mencionado al finalizar la sección 7.1, es necesario avanzar en la integración y la coordinación de los diferentes niveles de decisión, mejorando los detalles de la estrategia SO+RTO e incorporando el modelado y la interacción con el sistema de control.

## 7.2.5. Uso del tiempo disponible

Las metodologías del capítulo 6 se han presentado en forma conceptual y metodológica. Es necesario implementarlas en un sistema de RTO real o simulado, para determinar la estructura más apropiada de aprovechamiento del tiempo máximo disponible y del tiempo ocioso entre ciclos.

## 7.2.6. Extensión a otros procesos

### 7.2.6.1. Procesos continuos – modelos de estado estacionario

Además de los sistemas de calor y potencia, es posible explorar la utilidad de las técnicas propuestas en esta tesis para otros sistemas industriales, incluyendo procesos de separación y procesos con reacción química.

### 7.2.6.2. Optimización dinámica en tiempo real – DRTO

Las técnicas de optimización dinámica en tiempo real (DRTO) evolucionan rápidamente y pueden estar sujetas a estrategias de adaptación similares a las de la RTO con modelos de estado estacionario (Bonvin y Srinivasan, 2013). Por este motivo, una forma natural de extender el trabajo de esta tesis es aplicarlo a modelos dinámicos. Los procesos biológicos, por el alto grado de incertidumbre en el modelado, pueden ser objeto de la DRTO.

### 7.2.6.3. *Run-to-run optimization*

El perfil dinámico de un proceso discontinuo puede optimizarse utilizando los datos de corridas anteriores (François et al., 2004; Srinivasan y Bonvin, 2002). Las estrategias utilizadas para la adaptación en estas técnicas son muy similares a las de RTO de procesos continuos; por lo tanto, son un campo hacia el cual es posible extender las estrategias de esta tesis. En este caso, los procesos biológicos también son candidatos ideales para la aplicación de la optimización en tiempo real.

## BIBLIOGRAFÍA

- Adetola, V., Guay, M., 2007. Parameter convergence in adaptive extremum-seeking control. *Automatica* 43, 105 - 110.
- Adetola, V., Guay, M., 2010. Integration of real-time optimization and model predictive control. *Journal of Process Control* 20, 125 - 133.
- Alamo, T., Ferramosca, A., González, A.H., Limon, D., Odloak, D., 2014. A gradient-based strategy for the one-layer RTO + MPC controller. *Journal of Process Control* 24, 435-447.
- Alamo, T., Ferramosca, A., Gonzalez, A., Limon, D., Odloak, D., 2012. A gradient-based strategy for integrating real time optimizer (RTO) with model predictive control (MPC), en: 4th IFAC Nonlinear Model Predictive Control Conference. Leeuwenhorst, Netherlands, pp. 33–38.
- Ariyur, K.B., Krstic, M., 2003. *Real-Time Optimization by Extremum-Seeking Control*. John Wiley & Sons.
- Arora, N., Biegler, L.T., 2001. Redescending estimators for data reconciliation and parameter estimation. *Computers & Chemical Engineering* 25, 1585-1599.
- Avriel, M., 2012. *Nonlinear Programming: Analysis and Methods*. Courier Dover Publications.
- Aytug, H., Lawley, M.A., McKay, K., Mohan, S., Uzsoy, R., 2005. Executing production schedules in the face of uncertainties: A review and some future directions. *European Journal of Operational Research, IEPM: Focus on Scheduling* 161, 86-110.
- Bagajewicz, M.J., Jiang, Q., 2000. Comparison of steady state and integral dynamic data reconciliation. *Computers & Chemical Engineering* 24, 2367-2383.
- Benqilou, C., 2004. *Data Reconciliation as a Framework for Chemical Processes Optimization and Control (Tesis de doctorado)*. Universitat Politècnica de Catalunya.
- Bhat, S.A., Saraf, D.N., 2004. Steady-State Identification, Gross Error Detection, and Data Reconciliation for Industrial Process Units. *Industrial & Engineering Chemistry Research* 43, 4323-4336.
- Biegler, L.T., 2009. Technology Advances for Dynamic Real-Time Optimization, en: 10th International Symposium on Process Systems Engineering: Part A, Computer Aided Chemical Engineering. Elsevier, pp. 1 - 6.
- Bonvin, D., François, G., 2013. Model Control and Optimisation of Process Systems, en: *Advances in Chemical Engineering*. Academic Press.
- Bonvin, D., Srinivasan, B., 2013. On the role of the necessary conditions of optimality in structuring dynamic real-time optimization schemes. *Computers & Chemical Engineering* 51, 172-180.
- Brdys, M.A., Tatjewski, P., 2005. *Iterative Algorithms for Multilayer Optimization Control*, 1st ed. Imperial College Press.
- Brdys, M., Ellis, J.E., Roberts, P.D., 1987. Augmented integrated system optimisation and parameter estimation technique: derivation, optimality and convergence. *Control Theory and Applications, IEE Proceedings D* 134, 201-209.

- Brooks, F.J., 2000. GE gas turbine performance characteristics. GE Power Systems, Schenectady, NY.
- Bunin, G.A., 2013. On the equivalence between the modifier-adaptation and trust-region frameworks. arXiv:1312.6464 [math].
- Bunin, G.A., François, G., 2012. Exploiting Local Quasiconvexity for Gradient Estimation in Modifier-Adaptation Schemes, en: 2012 American Control Conference. Montréal, Canadá.
- Bunin, G.A., François, G., Bonvin, D., 2013a. Sufficient Conditions for Feasibility and Optimality of Real-Time Optimization Schemes - I. Theoretical Foundations. arXiv:1308.2620 [math.OA].
- Bunin, G.A., François, G., Bonvin, D., 2013b. From Discrete Measurements to Bounded Gradient Estimates: A Look at Some Regularizing Structures. *Industrial & Engineering Chemistry Research* 52, 12500-12513.
- Bunin, G.A., François, G., Bonvin, D., 2013c. Sufficient Conditions for Feasibility and Optimality of Real-Time Optimization Schemes - II. Implementation Issues. arXiv:1308.2625 [math.OA].
- Bunin, G.A., Tirado, F.F., François, G., Bonvin, D., 2012a. Run-to-Run MPC Tuning via Gradient Descent, en: Computer Aided Chemical Engineering, 22nd European Symposium on Computer Aided Process Engineering. Elsevier, pp. 927-931.
- Bunin, G.A., Wuillemin, Z., François, G., Nakajo, A., Tsikonis, L., Bonvin, D., 2012b. Experimental real-time optimization of a solid oxide fuel cell stack via constraint adaptation. *Energy* 39, 54 - 62.
- Bunin, G., François, G., Srinivasan, B., Bonvin, D., 2011. Input Filter Design for Feasibility in Constraint-Adaptation Schemes, en: 18th IFAC World Congress. Milán, Italia, pp. 5585-5590.
- Cao, S., Rhinehart, R.R., 1995. An efficient method for on-line identification of steady state. *Journal of Process Control* 5, 363 - 374.
- Castro, P.M., Harjunkoski, I., Grossmann, I.E., 2011. Optimal scheduling of continuous plants with energy constraints. *Computers & Chemical Engineering* 35, 372 - 387.
- Chachuat, B., Marchetti, A., Bonvin, D., 2008. Process optimization via constraints adaptation. *Journal of Process Control* 18, 244 - 257.
- Chachuat, B., Srinivasan, B., Bonvin, D., 2008. Model parameterization tailored to real-time optimization, en: Braunschweig, B., Joulia, X. (Eds.), 18th European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering. Elsevier, pp. 1 - 13.
- Chachuat, B., Srinivasan, B., Bonvin, D., 2009. Adaptation strategies for real-time optimization. *Computers & Chemical Engineering* 33, 1557 - 1567.
- Chen, C.Y., Joseph, B., 1987. On-line optimization using a two-phase approach: an application study. *Industrial & Engineering Chemistry Research* 26, 1924-1930.
- Chen, J., Bandoni, A., Romagnoli, J.A., 1997. Robust estimation of measurement error variance/covariance from process sampling data. *Computers & Chemical Engineering* 21, 593 - 600.

- Chen, J., Romagnoli, J.A., 1998. A strategy for simultaneous dynamic data reconciliation and outlier detection. *Computers & Chemical Engineering* 22, 559 - 562.
- Costello, S., François, G., Bonvin, D., 2013. Real-Time Optimization when the Plant and the Model have Different Inputs, en: *Dynamics and Control of Process Systems*. Mumbai.
- Cui, J., Engell, S., 2009. Scheduling of a Multiproduct Batch Plant under Multiperiod Demand Uncertainties by Means of a Rolling Horizon Strategy, en: *Computer Aided Chemical Engineering, 19th European Symposium on Computer Aided Process Engineering*. Elsevier, pp. 423-428.
- Cui, J., Engell, S., 2010. Medium-term planning of a multiproduct batch plant under evolving multi-period multi-uncertainty by means of a moving horizon strategy. *Computers & Chemical Engineering* 34, 598-619.
- Darby, M.L., Nikolaou, M., Jones, J., Nicholson, D., 2011. RTO: An overview and assessment of current practice. *Journal of Process Control* 21, 874 - 884.
- Drud, A., 2012. CONOPT 3 solver manual. ARKI Consulting and Development A/S, Bagsvaerd, Denmark.
- Engell, S., Harjunkoski, I., 2012. Optimal operation: Scheduling, advanced control and their integration. *Computers & Chemical Engineering* 47, 121 - 133.
- Faulwasser, T., Bonvin, D., 2014. On the Use of Second-Order Modifiers in Real-Time Optimization.
- Floudas, C.A., Akrotirianakis, I.G., Caratzoulas, S., Meyer, C.A., Kallrath, J., 2005. Global optimization in the 21st century: Advances and challenges. *Computers & Chemical Engineering* 29, 1185 - 1202.
- Floudas, C.A., Pardalos, P.M., 2008. *Encyclopedia of Optimization*. Springer Science & Business Media.
- Forbes, J.F., 1994. Model structure and adjustable parameter selection for operations optimization. ETD Collection for McMaster University 1-260.
- Forbes, J.F., Marlin, T.E., 1994. Model Accuracy for Economic Optimizing Controllers: The Bias Update Case. *Industrial & Engineering Chemistry Research* 33, 1919-1929.
- Forbes, J.F., Marlin, T.E., 1996. Design cost: a systematic approach to technology selection for model-based real-time optimization systems. *Computers & Chemical Engineering* 20, 717 - 734.
- François, G., Bonvin, D., 2013. Use of Convex Model Approximations for Real-Time Optimization via Modifier Adaptation. *Industrial & Engineering Chemistry Research* 52, 11614-11625.
- François, G., Bonvin, D., 2014. Use of Transient Measurements for the Optimization of Steady-State Performance via Modifier Adaptation. *Industrial & Engineering Chemistry Research* 53, 5148-5159.
- François, G., Srinivasan, B., Bonvin, D., 2005. Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *Journal of Process Control* 15, 701-712.
- François, G., Srinivasan, B., Bonvin, D., 2012. Comparison of six implicit real-time optimization schemes. *Journal Européen des Systèmes Automatisés* 45, 291-305.

- François, G., Srinivasan, B., Bonvin, D., 2014. Equivalence between Neighboring-Extremal Control and Self-Optimizing Control for the Steady-State Optimization of Dynamical Systems. *Ind. Eng. Chem. Res.* 53, 7470-7478.
- François, G., Srinivasan, B., Bonvin, D., Hernandez Barajas, J., Hunkeler, D., 2004. Run-to-Run Adaptation of a Semiadiabatic Policy for the Optimization of an Industrial Batch Polymerization Process. *Ind. Eng. Chem. Res.* 43, 7238-7242.
- GAMS Development Corporation, 1997. GAMS Language Guide, RELEASE 2.25, Version 92. Washington D.C.
- GAMS Development Corporation, 2008. GAMS - The Solver Manuals. Washington D.C.
- Gao, W., Engell, S., 2005. Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering* 29, 1401-1409.
- García, J., Ruiz, J.P., 2013. Multi-period Optimization of District Energy Supply. Development of Day-Ahead Scheduling to Supplement Real-Time Optimization at TECO. Presentado en International District Energy Association (IDEA) Annual Conference, Miami, Estados Unidos.
- Golato, M.A., Franck Colombres, F.J., Aso, G., Correa, C.A., Paz, D., 2008. Metodología de cálculo de la eficiencia térmica de generadores de vapor. *Rev. Ind. Agríc. de Tucumán* 85.
- Grossmann, I., Ruiz, J., 2012. Generalized Disjunctive Programming: A Framework for Formulation and Alternative Algorithms for MINLP Optimization, en: Lee, J., Leyffer, S. (Eds.), *Mixed Integer Nonlinear Programming, The IMA Volumes in Mathematics and Its Applications*. Springer New York, pp. 93-115.
- Gros, S., Srinivasan, B., Bonvin, D., 2009. Optimizing control based on output feedback. *Computers & Chemical Engineering* 33, 191-198.
- Guay, M., Zhang, T., 2003. Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties. *Automatica* 39, 1283 - 1293.
- Harjunkoski, I., Maravelias, C.T., Bongers, P., Castro, P.M., Engell, S., Grossmann, I.E., Hooker, J., Méndez, C., Sand, G., Wassick, J., 2014. Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering* 62, 161 - 193.
- Henning, G.P., Cerdá, J., 2000. Knowledge-based predictive and reactive scheduling in industrial environments. *Computers & Chemical Engineering* 24, 2315-2338.
- Higa, M., 2003. Cogeração e integração termica em usinas de açúcar e alcool (Tesis de doctorado). Universidad de Campinas.
- Hugot, E., 1986. *Handbook of Cane Sugar Engineering*, 3rd ed. Elsevier Science Publishers B.V.
- Irvine Jr., T.F., Liley, P.E., 1984. *Steam and Gas Tables with Computer Equations*. Academic Press.
- Jäschke, J., Skogestad, S., 2011. NCO tracking and self-optimizing control in the context of real-time optimization. *Journal of Process Control* 21, 1407 - 1416.
- Kocis, G.R., Grossmann, I.E., 1989. Computational experience with dicopt solving MINLP problems in process systems engineering. *Computers & Chemical Engineering* 13, 307-315.

- Kouvelis, P., Daniels, R.L., Vairaktarakis, G., 2000. Robust scheduling of a two-machine flow shop with uncertain processing times. *Iie Transactions* 32, 421–432.
- Lee, S., Grossmann, I.E., 2000. New algorithms for nonlinear generalized disjunctive programming. *Computers & Chemical Engineering* 24, 2125 - 2141.
- Liporace, F.S., Gomes, M.V.C., Katata, A.C., Zanin, A.C., Moro, L.F.L., Porfírio, C.R., 2009. PETROBRAS Experience Implementing Real Time Optimization, en: *Computer Aided Chemical Engineering, 10th International Symposium on Process Systems Engineering: Part A*. Elsevier, pp. 1245-1250.
- Lorenz, M.G., Roberto, F., 2009. *Steam and Power Optimizer*, ExxonMobil Baytown, Texas.
- Majchrowicz, J., Herra, M., Serralunga, F., Ruiz, D., 2011. Energy management: A polish case study. *Hydrocarbon Engineering* 16, 40–46.
- Mansour, M., Ellis, J.E., 2003. Comparison of methods for estimating real process derivatives in on-line optimization. *Applied Mathematical Modelling* 27, 275 - 291.
- Mansour, M., Ellis, J.E., 2008. Methodology of on-line optimisation applied to a chemical reactor. *Applied Mathematical Modelling* 32, 170 - 184.
- Marchetti, A., Chachuat, B., Bonvin, D., 2008. Real-time optimization via adaptation and control of the constraints, en: *18th European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering*. Elsevier, pp. 393 - 398.
- Marchetti, A., Chachuat, B., Bonvin, D., 2009. Modifier-Adaptation Methodology for Real-Time Optimization. *Industrial & Engineering Chemistry Research* 48, 6022–6033.
- Marchetti, A., Chachuat, B., Bonvin, D., 2010. A dual modifier-adaptation approach for real-time optimization. *Journal of Process Control* 20, 1027 - 1037.
- Marchetti, A.G., 2013. A new dual modifier-adaptation approach for iterative process optimization with inaccurate models. *Computers & Chemical Engineering* 59, 89 - 100.
- Marchetti, A.G., Basualdo, M., 2012. A New Dual Modifier-Adaptation Approach for Real-Time Optimization with Inaccurate Models, en: *22nd European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering*. Elsevier, pp. 942 - 946.
- Marchetti, A.G., Ferramosca, A., González, A.H., 2014. Steady-state target optimization designs for integrating real-time optimization and model predictive control. *Journal of Process Control* 24, 129-145.
- Marchetti, A.G., Zumoffen, D., 2013. On the links between real-time optimization, neighboring-extremal control, and self-optimizing control, en: *2013 European Control Conference, ECC 2013*. pp. 4466-4471.
- Mariani, D.C., Kihn, M.A., Ruiz, C.A., 2009. Industrial Experience on the Implementation of Real Time On Line Energy Management Systems in Sugar and Alcohol Industry, en: *10th International Symposium on Process Systems Engineering: Part A, Computer Aided Chemical Engineering*. Elsevier, pp. 459 - 464.

- Méndez, C.A., Cerdá, J., 2003. Dynamic scheduling in multiproduct batch plants. *Computers & Chemical Engineering*, 2nd Pan American Workshop in Process Systems Engineering 27, 1247-1259.
- Micheletto, S.R., Carvalho, M.C.A., Pinto, J.M., 2008. Operational optimization of the utility system of an oil refinery. *Computers & Chemical Engineering* 32, 170 - 185.
- Miletic, I., Marlin, T., 1996. Results analysis for real-time optimization (RTO): Deciding when to change the plant operation. *Computers & Chemical Engineering*, European Symposium on Computer Aided Process Engineering-6 20, Supplement 2, S1077-S1082.
- Miletic, I.P., Marlin, T.E., 1998a. Results diagnosis for real-time process operations optimization. *Computers & Chemical Engineering* 22, S475 - S482.
- Miletic, I.P., Marlin, T.E., 1998b. On-line Statistical Results Analysis in Real-Time Operations Optimization. *Ind. Eng. Chem. Res.* 37, 3670-3684.
- Mitra, S., Grossmann, I.E., Pinto, J.M., Arora, N., 2012. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering* 38, 171-184.
- Mitra, S., Sun, L., Grossmann, I.E., 2013. Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices. *Energy* 54, 194 - 211.
- Narasimhan, S., Jordache, C., 1999. *Data Reconciliation & Gross Error Detection: An Intelligent Use of Process Data*. Gulf Professional Publishing.
- Navia, D., Martí, R., Sarabia, D., Gutiérrez, D., de Prada, C., 2012. Handling Infeasibilities in Dual Modifier-Adaptation Methodology for Real-Time Optimization, en: 8th IFAC Symposium on Advanced Control of Chemical Processes. Singapore.
- Novas, J.M., Henning, G.P., 2010. Reactive scheduling framework based on domain knowledge and constraint programming. *Computers & Chemical Engineering*, 10th International Symposium on Process Systems Engineering, Salvador, Bahia, Brasil, 16-20 August 2009 34, 2129-2148.
- Ochoa, S., Repke, J.-U., Wozny, G., 2010. Integrating real-time optimization and control for optimal operation: Application to the bio-ethanol process. *Biochemical Engineering Journal* 53, 18 - 25.
- Office of Air Quality Planning and Standards, Office of Air and Radiation, 1995. *AP-42 Compilation of Air Pollutant Emission Factors*. Volume I: Stationary Point and Area Sources, 5th ed. US Environmental Protection Agency (EPA).
- Özyurt, D.B., Pike, R.W., 2004. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers & Chemical Engineering* 28, 381-402.
- Palombarini, J., Martínez, E., 2012. SmartGantt – An interactive system for generating and updating rescheduling knowledge using relational abstractions. *Computers & Chemical Engineering*, FOCAPO 2012 47, 202-216.
- Peters, N., Guay, M., DeHaan, D., 2007. Real-time dynamic optimization of batch systems. *Journal of Process Control* 17, 261-271.

- Poulin, É., Hodouin, D., Lachance, L., 2010. Impact of plant dynamics on the performance of steady-state data reconciliation. *Computers & Chemical Engineering* 34, 354 - 360.
- Puranik, Y., Sahinidis, N., Li, T., Feather, D., Bensacon, B., 2013. Real-time optimization of a Complex Industrial Gas Network. 2013 AIChE Annual Meeting, San Francisco, Estados Unidos.
- Raman, R., Grossmann, I.E., 1994. Modelling and computational techniques for logic based integer programming. *Computers & Chemical Engineering* 18, 563 - 578.
- Rejowski, R., Shah, V., Fontenot, C.E., de Tarso, P., Santos, V.E.N., 2009. Sustain Activities for Real-Time Optimization Models of Ethylene Plants, en: *Computer Aided Chemical Engineering, 10th International Symposium on Process Systems Engineering: Part A*. Elsevier, pp. 351-356.
- Roberts, P.D., 1979. An algorithm for steady-state system optimization and parameter estimation. *International Journal of System Sciences* 10, 719-734.
- Roberts, P.D., 1995. Coping with model-reality differences in industrial process optimisation – A review of integrated system optimisation and parameter estimation (ISOPE). *Computers in Industry* 26, 281 - 290.
- Rodger, E., 2010. *Dual Modifier Adaptation Methodology For the On-line Optimization of Uncertain Processes (Tesis de maestría)*. Mc Master University.
- Rodger, E., Chachuat, B., 2011. Design Methodology of Modifier Adaptation for On-Line Optimization of Uncertain Processes, en: 18th IFAC World Congress. Milán, Italia.
- Romagnoli, J., Sanchez, M.C., 1999. *Data Processing and Reconciliation for Chemical Process Operations, Process Systems Engineering*. Academic Press.
- Roux, G.A.C.L., Santoro, B.F., Sotelo, F.F., Teissier, M., Joulia, X., 2008. Improving steady-state identification, en: 18th European Symposium on Computer Aided Process Engineering, *Computer Aided Chemical Engineering*. Elsevier, pp. 459 - 464.
- Ruiz, C.A., 2009. Real Time Industrial Process Systems: Experiences from the Field, en: 10th International Symposium on Process Systems Engineering: Part A, *Computer Aided Chemical Engineering*. Elsevier, pp. 133 - 138.
- Ruiz, D., Mamprin, J., Ruiz, C., Nelson, D., Roseme, G., 2005. Utilities systems on-line optimization and monitoring: Experiences from the real world, en: *European Symposium on Computer-Aided Process Engineering-15, Computer Aided Chemical Engineering*. Elsevier, pp. 1159 - 1164.
- Ruiz, D., Ruiz, C., Serralunga, F., 2009. Real-time refinery energy management. *Petroleum technology quarterly* 14, 115-118.
- Sánchez, M., Romagnoli, J., 1996. Use of orthogonal transformations in data classification-reconciliation. *Computers & Chemical Engineering* 20, 483-493.
- Sand, G., Engell, S., 2004. Modeling and solving real-time scheduling problems by stochastic integer programming. *Computers & Chemical Engineering* 28, 1087-1103.

- Seborg, D.E., Mellichamp, D.A., Edgar, T.F., III, F.J.D., 2010. *Process Dynamics and Control*. John Wiley & Sons.
- Sequeira, S.E., Herrera, M., Graells, M., Puigjaner, L., 2004. On-line process optimization: parameter tuning for the real time evolution (RTE) approach. *Computers & Chemical Engineering* 28, 661 - 672.
- Serralunga, F.J., Aguirre, P.A., Mussati, M.C., 2014. Including disjunctions in real-time optimization. *Industrial & Engineering Chemistry Research*.
- Serralunga, F.J., Mussati, M.C., Aguirre, P.A., 2013. Model Adaptation for Real-Time Optimization in Energy Systems. *Industrial & Engineering Chemistry Research* 52, 16795-16810.
- Serralunga, F., Mussati, M.C., Aguirre, P.A., 2012a. Real-time optimization of energy systems in sugar and ethanol facilities: a modifier adaptation approach, en: 11th International Symposium on Process Systems Engineering, Computer Aided Chemical Engineering. Elsevier, pp. 375 - 379.
- Serralunga, F., Mussati, M.C., Aguirre, P.A., 2012b. An alternative real-time optimization algorithm with modifier adaptation: Application to heat and power systems, en: 22nd European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering. Elsevier, pp. 367 - 371.
- Singh, A., Forbes, J.F., Vermeer, P.J., Woo, S.S., 2000. Model-based real-time optimization of automotive gasoline blending operations. *Journal of Process Control* 10, 43 - 58.
- Skogestad, S., 2000a. Self-optimizing control: the missing link between steady-state optimization and control. *Computers & Chemical Engineering* 24, 569-575.
- Skogestad, S., 2000b. Plantwide control: the search for the self-optimizing control structure. *Journal of Process Control* 10, 487-507.
- Souza, G.D., Odloak, D., Zanin, A.C., 2009. Real Time Optimization (RTO) with Model Predictive Control (MPC), en: 10th International Symposium on Process Systems Engineering: Part A, Computer Aided Chemical Engineering. Elsevier, pp. 1365 - 1370.
- Souza, G.D., Odloak, D., Zanin, A.C., 2010. Real time optimization (RTO) with model predictive control (MPC). *Computers & Chemical Engineering* 34, 1999 - 2006.
- Srinivasan, B., Biegler, L.T., Bonvin, D., 2008. Tracking the necessary conditions of optimality with changing set of active constraints using a barrier-penalty function. *Computers & Chemical Engineering* 32, 572 - 579.
- Srinivasan, B., Bonvin, D., 2002. Interplay between identification and optimization in run-to-run optimization schemes, en: American Control Conference. Anchorage, Estados Unidos, pp. 2174-2179 vol.3.
- Srinivasan, B., Bonvin, D., 2003. Convergence Analysis of Iterative Identification and Optimization Schemes, en: American Control Conference. Denver, Estados Unidos, pp. 1956-1961.
- Srinivasan, B., François, G., Bonvin, D., 2011. Comparison of Gradient Estimation Methods for Real-time Optimization, en: 21st European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering. Elsevier, pp. 607 - 611.

- Tani, T., Matsuo, K., 2009. Robust Closed-Loop Real-Time Optimization for refinery utility plant with Model Predictive Control for constraint handling, en: *Industrial Technology, 2009. ICIT 2009. IEEE International Conference on*. pp. 1-5.
- Tatjewski, P., 2002. Iterative Optimizing Set-Point Control - The Basic Principle Redesigned.
- Tosukhowong, T., Lee, J.M., Lee, J.H., Lu, J., 2004. An introduction to a dynamic plant-wide optimization strategy for an integrated plant. *Computers & Chemical Engineering, PSE 2003 29*, 199-208.
- Türkay, M., Grossmann, I.E., 1996. Logic-based MINLP algorithms for the optimal synthesis of process networks. *Computers & Chemical Engineering 20*, 959 - 978.
- Valleru, S., Kelly, T., Hackney, J., 2012. LyondellBasell Industries multi-unit RTO implementation using combination of rigorous and simplified methodology. Presentado en AIChE 2012 Spring National Meeting, Houston, Texas, USA.
- Vecchiotti, A., Grossmann, I.E., 1999. LOGMIP: a disjunctive 0-1 non-linear optimizer for process system models. *Computers & Chemical Engineering 23*, 555 - 565.
- Velasco-Garcia, P., Varbanov, P.S., Arellano-Garcia, H., Wozny, G., 2011. Utility systems operation: Optimisation-based decision making. *Applied Thermal Engineering 31*, 3196 - 3205.
- Woodward, L., Perrier, M., Srinivasan, B., 2009. Improved performance in the multi-unit optimization method with non-identical units. *Journal of Process Control 19*, 205-215.
- Woodward, L., Perrier, M., Srinivasan, B., 2010. Real-time optimization using a jamming-free switching logic for gradient projection on active constraints. *Computers & Chemical Engineering 34*, 1863 - 1872.
- Yip, W.S., 2002. Modeling Updating in Real-Time Optimization (Tesis de doctorado). Mc Master University.
- Yip, W.S., Marlin, T.E., 2002. Multiple data sets for model updating in real-time operations optimization. *Computers & Chemical Engineering 26*, 1345 - 1362.
- Yip, W.S., Marlin, T.E., 2003. Designing plant experiments for real-time optimization systems. *Control Engineering Practice 11*, 837 - 845.
- Yip, W.S., Marlin, T.E., 2004. The effect of model fidelity on real-time optimization performance. *Computers & Chemical Engineering 28*, 267 - 280.
- Zanin, A.C., Gouvêa, M.T. de, Odloak, D., 2002. Integrating real-time optimization into the model predictive controller of the FCC system. *Control Engineering Practice 10*, 819 - 831.
- Zhang, Y., Forbes, J.F., 2000. Extended design cost: a performance criterion for real-time optimization systems. *Computers & Chemical Engineering 24*, 1829 - 1841.
- Zhang, Y., Monder, D., Forbes, J.F., 2002. Real-time optimization under parametric uncertainty: a probability constrained approach. *Journal of Process Control 12*, 373 - 389.

Zhang, Y., Nadler, D., Forbes, J.F., 2001. Results analysis for trust constrained real-time optimization. *Journal of Process Control* 11, 329 - 341.

# APÉNDICE A: MODELO DEL SISTEMA DE ENERGÍAS DE LA PLANTA DE AZÚCAR Y ETANOL

## A.1. NOMENCLATURA

**Tabla A- 1.** Nomenclatura completa del modelo.

$A$	Área del evaporador ( $m^2$ )	<b>Superíndices</b>	
$BD$	Fracción de agua purgada en caldera	$cond$	condensado
$BPE$	Ascenso ebulloscópico (K)	$i$	corriente de entrada
$Bx$	°Brix	$o$	corriente de salida
$c$	Coefficiente de evaporación ( $kg/m^2 \cdot ^\circ C$ )	$s$	vapor
$Cp$	Calor específico (kJ/kg/K)	$sat$	saturado (agua o vapor) en el evaporador
$F$	Caudal (t/h)	$w$	agua
$Fj$	Caudal de jugo (kg/h)	<b>Subíndices</b>	
$H$	Entalpía (kJ/kg)	$cb$	caldera
$Q$	Combustible en la caldera (kJ/h)	$d$	desaireador
$P$	Presión (Pa)	$dist$	destilación y otros usos
$S$	Entropía (kJ/kg/K)	$e$	efecto del evaporador
$T$	Temperatura (K)	$i$	corriente
$V$	Vapor del evaporador utilizado en el siguiente efecto (kg/h)	$j$	nodo (mezclador o divisor)
$VB$	Extracción de vapor en evaporador (kg/h)	$tb$	turbina de contrapresión
$W$	Potencia de la turbina (kW)	$tg$	etapa de turbina de extracción-condensación
$X^{med}$	Fracción másica de sólido	$vl$	válvula

## A.2. LÍMITES OPERATIVOS

**Tabla A- 2.** Límites en las variables operativas (LI: Límite inferior; LS: Límite superior)

Variable	LI	LS
$F_{cb1}^o$ (t/h)	70	200
$F_{cb2}^o$ (t/h)	70	200
$F_{tg1}^i$ (t/h)	80	140
$F_{tg1}^o$ (t/h)	50	90
$F_{tg2}^o$ (t/h)	30	50
$(T_0^{sat} - T_1^{sat})(^{\circ}\text{C})$	5	20
$(T_1^{sat} - T_2^{sat})(^{\circ}\text{C})$	6	20
$(T_2^{sat} - T_3^{sat})(^{\circ}\text{C})$	7	20
$(T_3^{sat} - T_4^{sat})(^{\circ}\text{C})$	8	20
$(T_4^{sat} - T_5^{sat})(^{\circ}\text{C})$	9	20
$T_5^{sat}(^{\circ}\text{C})$	54	-

### A.3. MODELOS DE LOS EQUIPOS

Nota: Cuando se indican los valores calculados como funciones  $H(\cdot)$ ,  $S(\cdot)$ ,  $T(\cdot)$  y  $P(\cdot)$ , se utilizan las correlaciones empíricas de Irvine y Liley (1984).

#### A.3.1. Calderas ( $cb = \{1,2\}$ )

$$\begin{aligned}
 Q_{cb} \cdot \eta_{cb} &= F_{cb}^o \cdot H_{cb}^o + F_{cb}^i \cdot BD \cdot H_{cb}^{sat} - F_{cb}^i \cdot H_{cb}^i \\
 F_{cb}^o &= (1 - BD_{cb}) \cdot F_{cb}^i \\
 BD_{cb} &= 0.02 \\
 H_{cb}^o &= H(T_{cb}^o, P_{cb}^o) \\
 H_{cb}^i &= H(T_{cb}^i, P_{cb}^i) \\
 H_{cb}^{sat} &= H^W(T^{sat}(P_{cb}^o))
 \end{aligned} \tag{A-1}$$

#### A.3.2. Etapas del turbogenerador ( $tg = \{1,2\}$ )

$$\begin{aligned}
 W_{TG} &= F_{tg1}^o \cdot (H_{tg1}^o - H_{tg1}^i) + F_{tg2}^o \cdot (H_{tg2}^o - H_{tg2}^i) \\
 F_{tg1}^o &= F_{tg2}^i + F_{tg1}^i \\
 F_{tg2}^o &= F_{tg2}^i \\
 H_{tg1}^o &= H_{tg2}^i \\
 S_{tg1}^o &= S_{tg2}^i \\
 P_{tg1}^o &= P_{tg2}^i \\
 H_{tg1}^o &= H_{tg1}^i - \eta_{tg1} \cdot (H_{tg1}^i - H_{tg1}^{iso}) \\
 H_{tg2}^o &= H_{tg2}^i - \eta_{tg2} \cdot (H_{tg2}^i - H_{tg2}^{iso}) \\
 H_{tg1}^i &= H(T_{tg1}^i, P_{tg1}^i) \\
 S_{tg1}^{iso} &= S_{tg1}^i = S(T_{tg1}^i, P_{tg1}^i) \\
 H_{tg1}^{iso} &= H(S_{tg1}^{iso}, P_{tg1}^o) \\
 S_{tg2}^{iso} &= S_{tg2}^i = S(H_{tg2}^i, P_{tg2}^i) \\
 H_{tg2}^{iso} &= H(S_{tg2}^{iso}, P_{tg2}^o)
 \end{aligned} \tag{A-2}$$

### A.3.3. Válvulas reductoras de presión ( $vl = \{1,2\}$ )

$$\begin{aligned} F_{vl}^o &= F_{vl}^i \\ H_{vl}^o &= H_{vl}^i \end{aligned} \quad (A-3)$$

### A.3.4. Turbinas de vapor de contrapresión ( $tb = \{1..4\}$ )

$$\begin{aligned} W_{tb} &= F_{tb}^o \cdot (H_{cb}^o - H_{cb}^i) \\ F_{tb}^o &= F_{tb}^i \\ H_{tb}^o &= H_{tb}^i - \eta_{tb} \cdot (H_{tb}^i - H_{tb}^{iso}) \\ H_{tb}^i &= H(T_{tb}^i, P_{tb}^i) \\ S_{tb}^{iso} &= S_{tb}^i = S(T_{tb}^i, P_{tb}^i) \\ H_{tb}^{iso} &= H(S_{tb}^{iso}, P_{tb}^o) \end{aligned} \quad (A-4)$$

### A.3.5. Efectos del evaporador ( $e = \{0..5\}$ (el índice 0 se utiliza para el jugo clarificado a la entrada y el vapor al primer efecto))

$$\begin{aligned} X_{e+1}^{med} &= (Bx_e + Bx_{e+1}) / 200 \\ BPE_e &= X_e^{med} \cdot (0.3 + X_e^{med}) \cdot (0.22 + 0.0078 \cdot T_e^{sat}) / (0.355 \cdot (1.036 - X_e^{med})) \\ Cp_e &= 4.187 \cdot (1 - 0.006 \cdot Bx_e) \\ Fj_e &= Fj_{e+1} + V_{e+1} + VB_{e+1} \\ Fj_e \cdot Bx_e &= Fj_{e+1} \cdot Bx_{e+1} \\ Fj_e \cdot Cp_e \cdot (T_e^{sat} + BPE_e) + V_e \cdot (H_e^{sat,s} - H_e^{sat,w}) &= Fj_{e+1} \cdot Cp_{e+1} \cdot (T_{e+1}^{sat} + BPE_{e+1}) \\ &\quad + (V_{e+1} + VB_{e+1}) \cdot H_{e+1}^{sat,s} \end{aligned} \quad (A-5)$$

$$V_{e+1} + VB_{e+1} = c_{e+1} \cdot A_{e+1} \cdot (T_e^{sat} - T_{e+1}^{sat} - BPE_{e+1}), \quad e = 1,2,3,4$$

$$V_1 + VB_1 \leq c_1 \cdot A_1 \cdot (T_0 - T_1^{sat} - BPE_1)$$

$$P_e = P^{sat}(T_e^{sat})$$

$$H_e^{sat,s} = H(T_e^{sat}, P_e)$$

## A.3.6. Desaireador

$$\begin{aligned}
F_d^{i,w} + F_d^{i,s} &= F_d^{o,w} + F_d^{o,s} \\
F_d^{i,w} \cdot H_d^{i,w} + F_d^{i,s} \cdot H_d^{i,s} &= F_d^{o,w} \cdot H_d^{o,w} + F_d^{o,s} \cdot H_d^{o,s} \\
H_d^{o,w} &= H^{sat,w}(P_d) \\
H_d^{o,s} &= H^{sat,s}(P_d) \\
F_{cb1}^i + F_{cb2}^i &= F_d^{i,w} \\
H_d^{o,w} &= H_{cb}^i \\
F_{cb1}^i + F_{cb2}^i &= F_d^{i,w} \\
H_d^{i,w} &= \left( F_{tg2}^{cond} H_{tg2}^{cond} + F_{dist}^{cond} H_{dist}^{cond} \right) / \left( F_{tg2}^{cond} + F_{dist}^{cond} \right) - 100 \text{ kJ/kg}
\end{aligned} \tag{A-6}$$



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
30																1
31													-1	1		
32														-1	1	
33															-1	1
34																-1
35																-1

**A.4.2. Balances de masa, energía y entropía:**

$$\sum_{j=1}^{35} F_j a_{ij} = 0, \quad j = 1 \dots 16$$

$$\left( \sum_{k=1}^{35} F_k H_k \right)_{a_{ik}=1} / \left( \sum_{k=1}^{35} F_k \right)_{a_{ik}=1} = H_l \quad j = 1 \dots 16, \quad l = \{i / a_{ij} = -1\} \quad (\text{A-7})$$

$$\left( \sum_{k=1}^{35} F_k S_k \right)_{a_{ik}=1} / \left( \sum_{k=1}^{35} F_k \right)_{a_{ik}=1} = S_l \quad j = 1 \dots 16, \quad l = \{i / a_{ij} = -1\}$$



## APÉNDICE B: MODELO DEL SISTEMA DE ENERGÍAS (CASO DE ESTUDIO 2, CAPÍTULO 4)

### B.1. NOMENCLATURA

**Tabla B- 1.** Nomenclatura completa del modelo.

		<b>Subíndices</b>	
BD	Purga de caldera		
D	Demanda	<i>cb</i>	caldera
F	Caudal (t/h)	<i>d</i>	desaireador
FW	Agua de alimentación calderas (t/h)	<i>dem</i>	demanda de vapor
H	Entalpía (kJ/kg)	<i>ee</i>	Energía eléctrica
HR	Heat Rate (turbina de gas)	<i>GT</i>	turbina de gas
Q	Combustible en la caldera (kJ/h)	<i>i</i>	corriente
P	Presión (Pa)	<i>j</i>	equipo
S	Entropía (kJ/kg/K)	<i>max</i>	máximo
T	Temperatura (K)	<i>mín</i>	mínimo
W	Potencia (kW)	<i>pc</i>	postcombustión
<i>y</i>	Variable de decisión discreta	<i>R</i>	Caldera de recuperación
$\eta$	Eficiencia	<i>STG</i>	turbogenerador
<b>Superíndices</b>		<i>tb</i>	turbina de contrapresión
<i>B</i>	a carga base (turbina de gas)	<i>tg</i>	etapa de turbina de extracción-condensación
<i>cond</i>	condensado	<i>vl</i>	válvula
<i>i</i>	corriente de entrada	<i>vt</i>	venteo
<i>o</i>	corriente de salida		
<i>s</i>	vapor		
<i>sat</i>	saturado (agua o vapor) en el evaporador		
<i>w</i>	agua		

## B.2. LÍMITES OPERATIVOS

**Tabla B- 2.** Límites en las variables operativas (LI: Límite inferior; LS: Límite superior)

<b>Variable</b>	<b>LI</b>	<b>LS</b>
$F_{cb1}^o$ (t/h)	40	80
$F_{cb2}^o$ (t/h)	40	80
$F_{tg1}^i$ (t/h)	80	140
$F_{tg1}^o$ (t/h)	0	60
$F_{tg2}^o$ (t/h)	20	140
$W_{GT}$ (kW)	20000	
$W_{tb1}$ (kW)	1000	1000
$W_{tb2}$ (kW)	1000	1000
$W_{tb3}$ (kW)	0	0

### B.3. MODELOS DE LOS EQUIPOS

Nota: Cuando se indican los valores calculados como funciones  $H(\cdot)$ ,  $S(\cdot)$ ,  $T(\cdot)$  y  $P(\cdot)$ , se utilizan las correlaciones empíricas de Irvine y Liley (1984).

#### B.3.1. Calderas ( $cb = \{1,2\}$ )

$$\begin{aligned}
 Q_{cb} \cdot \eta_{cb} &= F_{cb}^o \cdot H_{cb}^o + F_{cb}^i \cdot BD \cdot H_{cb}^{sat} - F_{cb}^i \cdot H_{cb}^i \\
 F_{cb}^o &= (1 - BD_{cb}) \cdot F_{cb}^i \\
 BD_{cb} &= 0.02 \\
 H_{cb}^o &= H(T_{cb}^o, P_{cb}^o) \\
 H_{cb}^i &= H(T_{cb}^i, P_{cb}^i) \\
 H_{cb}^{sat} &= H^W(T^{sat}(P_{cb}^o))
 \end{aligned} \tag{B-1}$$

#### B.3.2. Etapas del turbogenerador ( $tg = \{1,2\}$ )

$$\begin{aligned}
 W_{STG} &= F_{tg1}^o \cdot (H_{tg1}^o - H_{tg1}^i) + F_{tg2}^o \cdot (H_{tg2}^o - H_{tg2}^i) \\
 F_{tg1}^o &= F_{tg2}^i + F_{tg1}^o \\
 F_{tg2}^o &= F_{tg2}^i \\
 H_{tg1}^o &= H_{tg2}^i \\
 S_{tg1}^o &= S_{tg2}^i \\
 P_{tg1}^o &= P_{tg2}^i \\
 H_{tg1}^o &= H_{tg1}^i - \eta_{tg1} \cdot (H_{tg1}^i - H_{tg1}^{iso}) \\
 H_{tg2}^o &= H_{tg2}^i - \eta_{tg2} \cdot (H_{tg2}^i - H_{tg2}^{iso}) \\
 H_{tg1}^i &= H(T_{tg1}^i, P_{tg1}^i) \\
 S_{tg1}^{iso} &= S_{tg1}^i = S(T_{tg1}^i, P_{tg1}^i) \\
 H_{tg1}^{iso} &= H(S_{tg1}^{iso}, P_{tg1}^o) \\
 S_{tg2}^{iso} &= S_{tg2}^i = S(H_{tg2}^i, P_{tg2}^i) \\
 H_{tg2}^{iso} &= H(S_{tg2}^{iso}, P_{tg2}^o)
 \end{aligned} \tag{B-2}$$

**B.3.3. Válvula reductora de presión (*vl*)**

$$\begin{aligned}
F_{vl}^o &= F_{vl}^i \\
H_{vl}^o &= H_{vl}^i \\
F_{vl}^{i,\min} \cdot y_{vl} - F_{vl}^i &\leq 0 \\
F_{vl}^i - F_{vl}^{i,\max} \cdot y_{vl} &\leq 0 \\
F_{vl}^{i,\min} &= 4 \\
F_{vl}^{i,\max} &= 40
\end{aligned}
\tag{B-3}$$

**B.3.4. Turbinas de vapor de contrapresión (*tb* = {1..3})**

$$\begin{aligned}
W_{tb} &= F_{tb}^o \cdot (H_{cb}^o - H_{cb}^i) \\
F_{tb}^o &= F_{tb}^i \\
H_{tb}^o &= H_{tb}^i - \eta_{tb} \cdot (H_{tb}^i - H_{tb}^{iso}) \\
H_{tb}^i &= H(T_{tb}^i, P_{tb}^i) \\
S_{tb}^{iso} &= S_{tb}^i = S(T_{tb}^i, P_{tb}^i) \\
H_{tb}^{iso} &= H(S_{tb}^{iso}, P_{tb}^o)
\end{aligned}
\tag{B-4}$$

### B.3.5. Turbina de gas

$$\begin{aligned}
 Q_{GT} &= HR^B \cdot W_{GT}^B \cdot \left( 0.2 + 0.8 \cdot \frac{W_{GT}}{W_{GT}^B} \right) \\
 (Q_T - W_{GT}) \cdot \eta_R + Q_{pc} \cdot \eta_{pc} &= F_R^o \cdot H_R^o + F_R^i \cdot BD \cdot H_R^{sat} - F_R^i \cdot H_R^i \\
 \eta_R &= 0.7 \\
 \eta_{pc} &= 0.98 \\
 F_{cb}^o &= (1 - BD_R) \cdot F_R^i \\
 BD_R &= 0.02 \\
 H_R^o &= H(T_R^o, P_R^o) \\
 H_R^i &= H(T_R^i, P_R^i) \\
 H_R^{sat} &= H^W(T^{sat}(P_R^o)) \\
 Q_{pc}^{\min} \cdot y_{pc} - Q_{pc} &\leq 0 \\
 Q_{pc} - Q_{pc}^{\max} \cdot y_{pc} &\leq 0 \\
 W_{GT} - W_{GT}^B &\leq 0
 \end{aligned} \tag{B-5}$$

### B.3.6. Desaireador

$$\begin{aligned}
 F_d^{i,w} + F_d^{i,s} &= F_d^{o,w} + F_d^{o,s} \\
 F_d^{i,w} \cdot H_d^{i,w} + F_d^{i,s} \cdot H_d^{i,s} &= F_d^{o,w} \cdot H_d^{o,w} + F_d^{o,s} \cdot H_d^{o,s} \\
 H_d^{o,w} &= H^{sat,w}(P_d) \\
 H_d^{o,s} &= H^{sat,s}(P_d) \\
 F_{cb1}^i + F_{cb2}^i &= F_d^{i,w} \\
 H_d^{o,w} &= H_{cb}^i \\
 F_{cb1}^i + F_{cb2}^i &= F_d^{i,w} \\
 H_d^{i,w} &= \left( F_{tg2}^{cond} H_{tg2}^{cond} + F_{dist}^{cond} H_{dist}^{cond} \right) / \left( F_{tg2}^{cond} + F_{dist}^{cond} \right) - 100 \text{ kJ/kg}
 \end{aligned} \tag{B-6}$$

### B.3.7. Penalidad de compra de electricidad

$$\begin{aligned}
 \frac{D_{ee} - W_{GT} - W_{STG} - 11000 \text{ kW}}{D_{ee}} - y_{pen} &\leq 0 \\
 \frac{D_{ee} - W_{GT} - W_{STG} - 11000 \text{ kW}}{D_{ee}} + (1 - y_{pen}) &\geq 0 \\
 pen &\geq 2 \cdot C_{ee} \cdot (1000 \cdot y_{pen} + D_{ee} - W_{GT} - W_{STG} - 11000 \text{ kW})
 \end{aligned} \tag{B-7}$$

## B.4. BALANCES

### B.4.1. Matriz de incidencias

Tabla B- 3. Matriz de incidencias (Elementos  $a_{ij}$ : fila  $i$ : colector de vapor (según Figura 4-8), columna  $j$ : equipo; 1: entra al colector, -1: sale del colector).

Colector	Índice	cb1	cb2	R	vl	tg1	tg2	tb1	tb2	tb3	dem	d	vt
5 Mpag	1	1	1	1	-1	-1		-1	-1	-1			
0.5 Mpag	2				1	1		1	1	1	-1	-1	-1
-0.075 Mpag	3						1						

### B.4.2. Balances de masa, energía y entropía:

$$\sum_j F_i a_{ij} = 0, \quad j = 1, 2$$

$$\left( \sum_k F_k H_k \right)_{a_{ik}=1} / \left( \sum_k F_k \right)_{a_{ik}=1} = H_l \quad j = 1, 2, \quad l = \{i / a_{ij} = -1\} \quad (7-1)$$

$$\left( \sum_{k=1}^{35} F_k S_k \right)_{a_{ik}=1} / \left( \sum_{k=1}^{35} F_k \right)_{a_{ik}=1} = S_l \quad j = 1, 2, \quad l = \{i / a_{ij} = -1\}$$

## B.5. PERFIL DE TEMPERATURAS, PRECIOS Y DEMANDAS

**Tabla B- 4.** Temperatura ambiente ( $T$ ), precio de compra de electricidad ( $C_{ee}$ ), demanda de vapor ( $D_v$ ) y demanda de potencia ( $D_{ee}$ ) para cada ciclo de RTO.

Ciclo RTO	$T$ (°C)	$C_{ee}$ (\$/kWh)	$D_v$ (t/h)	$D_{ee}$ (kW)
1	15.59	0.1	84	65000
2	16.43	0.1	84	65000
3	16.50	0.1	84	65000
4	17.37	0.1	84	65000
5	17.50	0.09	86	65000
6	18.28	0.09	88	65000
7	18.41	0.09	90	64000
8	18.70	0.1	90	64000
9	19.26	0.1	90	64000
10	19.47	0.1	92	64000
11	19.93	0.1	95	64000
12	19.96	0.11	95	64000
13	20.09	0.11	95	64000
14	20.42	0.11	94	64000
15	20.15	0.11	94	64000
16	20.47	0.11	92	64000
17	20.40	0.11	92	64000
18	20.15	0.11	92	64000
19	20.01	0.11	92	64000
20	19.56	0.1	92	64000
21	19.57	0.1	89	64000
22	19.33	0.1	87	63000
23	18.77	0.09	85	63000
24	18.54	0.09	83	63000
25	18.43	0.09	81	63000
26	18.43	0.09	81	63000
27	18.43	0.09	81	63000
28	18.43	0.09	81	63000
29	18.43	0.09	81	63000
30	18.43	0.09	81	63000



## APÉNDICE C: MODELO DEL SISTEMA DE ENERGÍAS (CASO DE ESTUDIO, CAPÍTULO 5)

### C.1. MODELO BASE

El modelo resuelto para cada período de este caso de estudio está basado en el modelo presentado en el Apéndice B. Por este motivo, sólo se describen las diferencias (relativas a la puesta en marcha de equipos), las ecuaciones de transición y las que involucran variables acumuladas.

### C.2. NOMENCLATURA

**Tabla C- 1.** Nomenclatura del del modelo.

a	Término emisión NO <sub>x</sub> (kg/GJ)	<b>Superíndices</b>	
Ac	Valor acumulado	<i>j</i>	Diario
b	Término emisión NO <sub>x</sub> (kg/h)	<i>off-on</i>	Parado a <i>en marcha</i>
c	Término emisión NO <sub>x</sub> (kg/h)	<i>off-on</i>	<i>En marcha</i> a parado
e	Factor de emisión NO <sub>x</sub> (kg/GJ) (modelo RTO)	<i>θ</i>	Período
<b>Subíndices</b>			
NO <sub>x</sub>	Producción de NO <sub>x</sub> (kg/h)	<i>e</i>	emisión
z	Variable de transición	<i>NO<sub>x</sub></i>	óxidos de nitrógeno
Δ	Período de tiempo (h)	<i>τ</i>	de transición

### C.3. MODELOS DE LOS EQUIPOS

Nota: Cuando se indican los valores calculados como funciones  $H(\cdot)$ ,  $S(\cdot)$ ,  $T(\cdot)$  y  $P(\cdot)$ , se utilizan las correlaciones empíricas de Irvine y Liley (1984). Se describen sólo las diferencias con respecto al Apéndice B.

#### C.3.1. Calderas ( $cb = \{1,2\}$ )

$$\begin{aligned}
 F_{cb}^{o,\theta} - F_{cb}^{o,\max} y_{cb}^\theta &\leq 0. \\
 F_{cb}^{o,\min} \cdot y_{cb}^\theta - F_{cb}^{o,\theta} &\leq 0 \\
 F_{cb}^{o,\max} &= 80 t / h \\
 F_{cb}^{o,\min} &= 40 t / h \\
 z_{\tau,cb}^{off-on,\theta} - z_{\tau,cb}^{on-off,\theta} &= y_{cb}^\theta - y_{cb}^{\theta-1} \\
 \sum_{cb} y_{cb}^\theta &\geq 1 \\
 \sum_{\theta=7}^{\theta} \left( z_{\tau,cb}^{off-on,\theta} + z_{\tau,cb}^{on-off,\theta} \right) &\leq 1
 \end{aligned} \tag{C-1}$$

#### C.3.2. Válvula reductora de presión ( $vl$ )

$$F_{vl}^{i,\min} = 1 t / h \tag{C-2}$$

#### C.3.3. Turbinas de vapor de contrapresión ( $tb = \{1..3\}$ )

$$\begin{aligned}
 W_{tb} &= 1000 kW \\
 F_{tb}^{i,\theta} &= f_{tb} \cdot y_{tb}^\theta && \text{Modelo RTO} \\
 F_{tb}^{i,\theta} \cdot (H_{tb}^i - H_{tb}^o) &= W_{tb} \cdot y_{tb}^\theta \\
 H_{tb}^{o,\theta} &= H_{tb}^{i,\theta} - \eta_{tb} \cdot (H_{tb}^{i,\theta} - H_{tb}^{iso,\theta}) \\
 \eta_{tb} &= 0.5 \\
 H_{tb}^{i,\theta} &= H(T_{tb}^{i,\theta}, P_{tb}^{i,\theta}) \\
 S_{tb}^{iso,\theta} &= S_{tb}^{i,\theta} = S(T_{tb}^{i,\theta}, P_{tb}^{i,\theta}) \\
 H_{tb}^{iso,\theta} &= H(S_{tb}^{iso,\theta}, P_{tb}^{o,\theta}) \\
 z_{\tau,tb}^{off-on,\theta} - z_{\tau,tb}^{on-off,\theta} &= y_{tb}^\theta - y_{tb}^{\theta-1} \\
 \sum_{\theta=3}^{\theta} \left( z_{\tau,tb}^{off-on,\theta} + z_{\tau,tb}^{on-off,\theta} \right) &\leq 1
 \end{aligned} \left. \vphantom{\begin{aligned} F_{tb}^{i,\theta} \\ H_{tb}^{o,\theta} \\ \eta_{tb} \\ H_{tb}^{i,\theta} \\ S_{tb}^{iso,\theta} \\ H_{tb}^{iso,\theta} \end{aligned}} \right\} \text{PlantaReal} \tag{C-3}$$

## C.3.4. Producción de NOx

$$\begin{aligned}
 & \left. \begin{aligned}
 NOx_{cb}^{\theta} &= Q_{cb}^{\theta} \cdot e_{cb} \quad , cb=1,2 \\
 NOx_{pc}^{\theta} &= Q_{pc}^{\theta} \cdot e_{pc} \\
 NOx_{GT}^{\theta} &= Q_{GT}^{\theta} \cdot e_{GT}
 \end{aligned} \right\} \text{Modelo RTO} \\
 & \left. \begin{aligned}
 NOx_{cb}^{\theta} &= a_{e,cb} \cdot Q_{cb}^{\theta} + b_{e,cb} \cdot y_{cb}^{\theta} \quad , cb=1,2 \\
 NOx_{pc}^{\theta} &= a_{e,pc} \cdot Q_{pc}^{\theta} + b_{e,pc} \cdot y_{pc}^{\theta} + c_{e,pc} \cdot \theta \\
 NOx_{GT}^{\theta} &= a_{e,GT} \cdot Q_{GT}^{\theta} + b_{e,GT} \cdot
 \end{aligned} \right\} \text{Plantareal} \\
 & Ac_{NOx}^j = \sum_{\theta \in j} \left( \sum_{k=cb,pc,TG} NOx_k^{\theta} \cdot \Delta \right) \\
 & EX_{NOx}^{\theta} \geq Ac_{NOx}^j - Ac_{NOx}^{\max} \quad \forall \theta / \theta \bmod 4 = 0 \\
 & EX_{NOx}^j \geq 0 \\
 & Ac_{NOx}^{\max} = 1000 \text{ kg} \\
 & \Delta = 6 \text{ h}
 \end{aligned} \tag{C-4}$$

#### C.4. PERFIL DE TEMPERATURAS, PRECIOS Y DEMANDAS

**Tabla C- 2.** Temperatura ambiente (°C). Pronóstico y valor real para cada período  $\theta$ .

Período	Pronóstico	Valor real
1	15	15.2
2	20	19
3	25	24.2
4	20	22
5	15	13.7
6	20	19.6
7	25	21.9
8	20	19
9	15	16.5
10	20	22.3
11	25	26.4
12	20	21.6
13	15	17.1
14	20	24.6
15	25	27.9
16	20	24.2
17	15	11
18	20	15.4
19	25	21.5
20	20	17.9
21	15	18.9
22	20	20.9
23	25	27.7
24	20	21.6
25	15	19.5
26	20	27.4
27	25	32.1
28	20	27.7

**Tabla C- 3.** Precio de compra de electricidad (\$/kW). Pronóstico y valor real para cada período  $\theta$ .

Período	Pronóstico	Valor real
1	0.08	0.087
2	0.1	0.117
3	0.12	0.137
4	0.1	0.107
5	0.08	0.087
6	0.1	0.107
7	0.12	0.137
8	0.1	0.107
9	0.08	0.089
10	0.1	0.119
11	0.12	0.119
12	0.1	0.099
13	0.08	0.122
14	0.1	0.142
15	0.12	0.162
16	0.1	0.122
17	0.08	0.053
18	0.1	0.073
19	0.12	0.113
20	0.1	0.083
21	0.08	0.112
22	0.1	0.142
23	0.12	0.142
24	0.1	0.122
25	0.08	0.134
26	0.1	0.154
27	0.12	0.184
28	0.1	0.164

**Tabla C- 4.** Demanda de vapor (t/h). Pronóstico y valor real para cada período  $\theta$ .

Período	Pronóstico	Valor real
1	60	59.5
2	70	71.5
3	80	81.5
4	80	80.5
5	60	57
6	70	67
7	80	74
8	80	75
9	60	53.5
10	70	64.5
11	80	76.5
12	80	76.5
13	60	57.25
14	70	63.25
15	80	76.25
16	80	74.25
17	60	74.25
18	70	83.25
19	80	90.25
20	80	94.25
21	60	62.25
22	70	69.25
23	80	81.25
24	80	80.25
25	60	70
26	70	79
27	80	88
28	80	90

**Tabla C- 5.** Demanda de potencia eléctrica (kW). Pronóstico y valor real para cada período  $\theta$ .

Período	Pronóstico	Valor real
1	55000	56027
2	60000	60320
3	65000	64324
4	65000	67731
5	55000	52961
6	60000	58846
7	65000	62672
8	65000	62337
9	55000	52099
10	60000	59380
11	65000	63626
12	65000	62352
13	55000	53814
14	60000	56754
15	65000	64263
16	65000	61592
17	55000	60531
18	60000	66966
19	65000	70823
20	65000	70743
21	55000	56542
22	60000	60158
23	65000	64114
24	65000	64448
25	55000	60415
26	60000	63406
27	65000	68734
28	65000	70861