

CONTROL PREDICTIVO: PASADO, PRESENTE Y FUTURO

Eduardo F. Camacho y Carlos Bordons

*Escuela Superior de Ingenieros. Universidad de Sevilla,
Camino de los descubrimientos s/n
41092 Sevilla, España
e-mail: {eduardo, bordons}@esi.us.es*

Resumen: El control predictivo basado en modelo (Model Predictive Control, MPC) se ha desarrollado considerablemente en las últimas décadas tanto en la industria como en la comunidad de investigación. Este éxito se debe a que el Control Predictivo basado en Modelo es quizás la forma más general de formular el problema de control en el dominio del tiempo. El control predictivo integra control óptimo, control de procesos con tiempos muertos, procesos multivariados y utiliza las referencias futuras cuando están disponibles. Al utilizar una estrategia con horizonte de control finito permite la consideración de restricciones y procesos no lineales. Este artículo describe el control predictivo y sus posibilidades y pretende actuar como motivación para su utilización. *Copyright ©2004 CEA-IFAC.*

Palabras claves: Control Predictivo, Restricciones, Control Predictivo Robusto, Control de Sistemas Híbridos.

1. INTRODUCCIÓN

El Control Predictivo (Model Predictive Control, MPC) se desarrolló al finales de los setenta y ha tenido un desarrollo considerable desde entonces. El término Control Predictivo no designa a una estrategia de control particular sino a un conjunto de métodos de control que hacen uso explícito de un modelo del proceso para obtener la señal de control minimizando una función objetivo. Estos métodos de control llevan a controladores que tienen básicamente la misma estructura y los mismos elementos:

- uso explícito de un modelo para predecir la evolución del proceso en los instantes futuros,
- minimización de una función objetivo y
- utilización de un horizonte de control finito y deslizante que implica el cálculo de la secuencia de control para todo el horizonte pero con la aplicación de la primera señal de

la secuencia y la repetición de todo el proceso en el siguiente instante de muestreo.

Los distintos algoritmos de control predictivo difieren en el tipo de modelo utilizado para representar al proceso y a las perturbaciones y la función objetivo considerada. Existen aplicaciones de control predictivo a diversos procesos que van desde procesos tan diversos como robots (Gómez Ortega y Camacho, 1996) a la anestesia clínica (Linkers y Mahfonf, 1994). Aplicaciones en la industria de cemento, desecadoras, brazos robóticos se pueden encontrar descritas en (Clarke, 1988), mientras que desarrollos para columnas de destilación, plantas de PVC, generadores de vapor y servos se presentan en (Richalet, 1993) y (Richalet *et al.*, 1978). El control predictivo presenta una serie de ventajas sobre otros métodos, entre las que se pueden citar las siguientes:

- es una técnica particularmente atractiva para los operadores que requiere pocos conocimientos de control porque los conceptos son muy intuitivos y la sintonización relativamente simple,
- se puede utilizar para controlar una gran variedad de procesos, desde procesos muy simples hasta procesos con dinámicas complejas como procesos con grandes tiempos muertos, procesos de fase no mínima, procesos inestables o procesos multivariados,
- su carácter predictivo lo hace compensar intrínsecamente los tiempos muertos,
- introduce un control anticipativo (*feed forward*) y de forma natural se compensan las perturbaciones medibles,
- la ley de control resultante es fácilmente implementable,
- es muy útil cuando se conocen las referencias futuras, como ocurre en el caso de robótica o procesos por lotes y
- permite tratar las restricciones de una forma sistemática y conceptualmente muy simple durante la fase de diseño.

Como es lógico, tiene también sus inconvenientes. El principal es que, aunque su implementación no es compleja, resulta más difícil que la de los clásicos controladores PID. Si la dinámica del proceso no cambia y no existen restricciones, la mayor parte de los cálculos se puede realizar fuera de línea y el controlador resultante es simple, pudiéndose aplicar a procesos de dinámicas rápidas; en caso contrario, los requisitos de cálculo son mucho mayores. Aunque hay que decir que debido a la potencia de los computadores actuales esto no es realmente una dificultad insalvable. La mayor dificultad que presenta para su aplicación es la necesidad de un modelo apropiado del proceso cuya obtención requiere unos conocimientos mínimos de control. El control predictivo ha demostrado ser en la práctica una estrategia razonable de control y ha sido aplicado con éxito a numerosos procesos industriales.

1.1 Estrategia del MPC

La metodología de todos los controladores pertenecientes a la familia MPC se caracteriza por la siguiente estrategia, representada en la figura 1:

- (1) Las salidas futuras para un horizonte determinado N , llamado horizonte de predicción, se predicen cada instante t utilizando el modelo del proceso. Estas predicciones de la salida $y(t+k | t)$ ¹ para $k = 1 \dots N$ dependen de

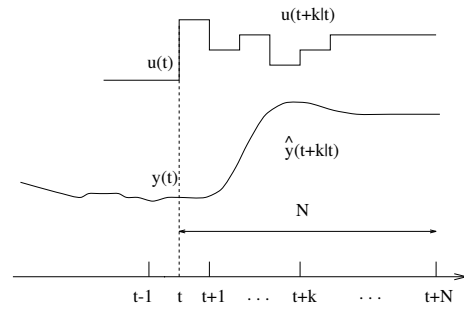


Fig. 1. Estrategia MPC

los valores conocidos hasta el instante t (entradas y salidas conocidas) y de las señales de control $u(t+k | t)$, $k = 0 \dots N-1$, que han de ser calculadas y enviadas al sistema.

- (2) La secuencia de señales de control futuras se calcula minimizando un criterio para mantener al proceso lo más cerca posible de la trayectoria de referencia $w(t+k)$. Este criterio toma normalmente la forma de una función cuadrática del error entre la salida predicha y la trayectoria de referencias futuras. En la mayor parte de los casos se incluye también el esfuerzo de control dentro de la función objetivo. La solución explícita se puede obtener cuando el criterio es cuadrático y el modelo lineal; en caso contrario se ha de utilizar un método numérico para buscar la solución.
- (3) La señal de control $u(t | t)$ se envía al proceso mientras que el resto de las señales calculadas no se consideran, ya que en el instante siguiente de muestreo $y(t+1)$ es ya conocida y los pasos anteriores se repiten con este nuevo valor. Por lo que $u(t+1 | t+1)$ se calcula con información diferente y en principio será también diferente de $u(t+1 | t)$.

La figura 2 muestra la estructura básica necesaria para implementar el control predictivo. Se usa un modelo para predecir la evolución de la salida o estado del proceso a partir de las señales de entrada y salidas conocidas. Las acciones de control futuras se calculan con el optimizador, que considera la función del coste y las posibles restricciones.

El modelo de proceso juega, en consecuencia, un papel decisivo en el controlador. El modelo elegido debe ser capaz de capturar la dinámica del proceso para predecir de forma precisa la evolución del sistema. Al mismo tiempo, debe ser suficientemente simple de implementar y entender. Las distintas metodologías del control predictivo difieren fundamentalmente en el tipo de modelo utilizado.

El optimizador es otra parte fundamental de la estructura ya que permite obtener las acciones de control a aplicar. Si la función de coste es

¹ La notación indica el valor predicho de la variable en el instante $t+k$ calculada en el instante t .

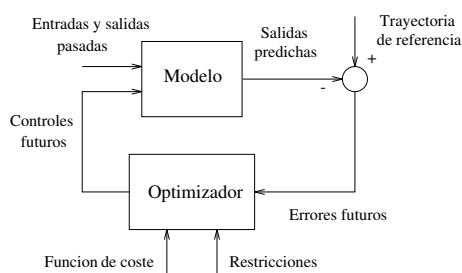


Fig. 2. Estructura básica del MPC

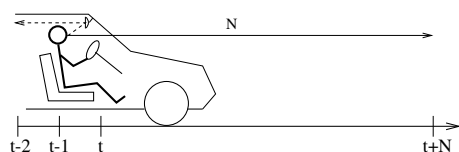


Fig. 3. Analogía MPC

cuadrática, el modelo lineal y no existen restricciones, se puede obtener una solución explícita. Si éste no es el caso se ha de acudir a un algoritmo numérico de optimización que requiere mayor capacidad de cálculo. El tamaño del problema resultante depende del número de variables, de los horizontes de control y predicción y del número de restricciones, aunque se puede decir que en general problemas de optimización resultantes en este contexto son problemas más bien modestos.

Nótese que la estrategia de control predictivo es muy similar a la estrategia que se utiliza cuando se conduce un automóvil. El conductor conoce la trayectoria de referencia deseada para un horizonte de control finito. Tomando en consideración las características del automóvil (modelo mental del automóvil) decide qué acción de control tomar (acelerador, frenos, volante, marchas) para seguir la trayectoria deseada. Sólo la primera acción de control de la secuencia calculada mentalmente es aplicada por el conductor en cada instante y el procedimiento se repite en los sucesivos instantes utilizando el concepto de horizonte deslizante. Nótese que cuando se utiliza un esquema de control clásico como PID se utilizan sólo las señales pasadas. Esta forma de conducir el automóvil sería como conducir utilizando el espejo retrovisor tal como se muestra en la figura 3. Esta analogía no es totalmente justa con los PIDs, porque el control predictivo utiliza más información (trayectoria de referencia). Nótese que si se le proporciona al PID como referencia un punto en la trayectoria futura la diferencia entre ambas estrategias de control no parecería tan abismal.

1.2 Perspectiva Histórica

Desde el final de la década de los 70 aparecieron varios artículos mostrando un interés incipiente en el control predictivo en la industria, principalmente en (Richalet *et al.*, 1976)(Richalet *et al.*, 1978) presentando el control predictivo heurístico basado en modelo ("Model Predictive Heuristic Control", MPH) (más tarde conocido como control algorítmico basado en modelo ("Model Algorithmic Control", MAC)) y (Cutler y Ramaker, 1980) sobre control con matriz dinámica ("Dynamic Matrix Control", DMC). Ambos algoritmos utilizan explícitamente un modelo dinámico del proceso (la respuesta impulso en el primer caso y la respuesta escalón en el segundo) para predecir el efecto de las futuras señales de control en las variables a controlar. Estas formulaciones eran heurísticas e hicieron uso del potencial cada vez mayor de los computadores digitales en aquellos tiempos.

Estos controladores estaban íntimamente ligados al problema de control óptimo en tiempo mínimo y a la programación lineal (Zadeh y Whalen, 1962). El concepto de horizonte deslizante, una de las ideas centrales del control predictivo, fue propuesto por Propoi ya en el 1963 (Propoi, 1963), en el marco de realimentación óptima en bucle abierto (*open-loop optimal feedback*) que fue utilizada extensamente en los años 70.

El control predictivo llegó a ser popular, y particularmente en la industria de procesos químicos, debido a la simplicidad del algoritmo y a la utilización del modelo de respuesta ante impulso que, aunque requiriendo muchos más parámetros que las formulaciones en el espacio de estado o en el dominio de entrada y salida, resulta más intuitivo y requiere mucha menos información a priori para la identificación. Un informe bastante completo sobre las aplicaciones del control predictivo en el sector petroquímico durante los años 80 se puede encontrar en (García *et al.*, 1989).

La mayoría de estas aplicaciones se llevaron a cabo en sistemas multivariables que incluían restricciones. A pesar de este éxito, estas formulaciones carecían de una teoría formal para proveer resultados sobre la estabilidad y robustez. De hecho, el caso de horizonte finito parecía demasiado difícil de analizar excepto en casos muy específicos.

Otra línea de trabajo se desarrolló independientemente en torno a las ideas de control adaptativo, desarrollándose estrategias de control predictivo para sistemas monovariantes y formulada sobre modelos de entrada y salida. El control autosintonizado basado en predictores ("Predictor-Based Self-Tuning Control") (Peterka, 1984) y el control adaptativo de horizonte extendido ("Extended Horizon Adaptive Control" EHAC) (Ydstie, 1984),

el controlador autosintonizado ("Extended Prediction Self Adaptive Control" EPSAC) (Keyser y Cuawenberghe, 1985), y el control predictivo generalizado ("Generalized Predictive Control GPC) desarrollado por Clarke *et al.* en 1987 (Clarke *et al.*, 1987) pueden ser mencionado en este contexto. El GPC utiliza ideas de los controladores de mínima varianza generalizada ("Generalized Minimum Variance GMV) (Clarke y Gawthrop, 1979) y es en la actualidad uno de los métodos más utilizados a nivel académico.

Existen numerosas formulaciones de control predictivo basadas en las mismas ideas comunes, entre las que se puede incluir control adaptativo multivariable multipaso ("Multistep Multivariable Adaptive Control" MUSMAR) (Greco *et al.*, 1984), control predictivo funcional ("Predictive Functional Control" PFC) (Richalet *et al.*, 1987). El MPC ha sido formulado en el espacio de estados (Morari, 1994), lo que permite una utilización de resultados bien conocidos sobre estabilidad y también la generalización a casos más complejos como procesos multivariables, procesos no lineales y sistemas con perturbaciones estocásticas. Aunque los primeros trabajos sobre GPC contienen algunos resultados de estabilidad para el caso nominal, la falta de resultados generales sobre la estabilidad de los controladores de horizontes finito y deslizante constituyó un inconveniente para su utilización al principio. Para hacer frente a esto, apareció en los 90s una nueva línea de trabajo sobre controladores predictivos con estabilidad garantizada. Se pueden citar dos métodos el CRHPC (Clarke y Scattolini, 1991)) y el SIORHC (Mosca *et al.*, 1990)) que fueron desarrollados independientemente y garantizaban estabilidad, para el caso nominal, imponiendo que la señal de salida alcanzara a la referencia al final del horizonte.

Para el caso de restricciones, el análisis de estabilidad parecía ser un problema demasiado complicado de resolver. Aún en el caso de que el optimizador fuera capaz de encontrar una solución, no estaba garantizada la estabilidad del bucle cerrado. La utilización de penalizaciones terminales y/o restricciones, funciones de Lyapunov, o conjuntos invariantes han dado lugar a una familia de técnicas que garantizan la estabilidad del sistema. Este problema ha sido atacado de distintos puntos de vista y han aparecido numerosas contribuciones en años recientes, casi siempre analizando el problema del regulador (llevar el estado al reposo) y normalmente en el espacio de estados. Las principales formulaciones propuestas que garantizan estabilidad están resumidas en (Mayne *et al.*, 2000) donde se dan condiciones suficientes para diseñar un regulador predictivo con garantías de estabilidad.

Otra de las líneas de investigación abiertas en los últimos años es el control predictivo robusto. La idea básica es tener en cuenta las incertidumbres sobre el proceso de una manera explícita y diseñar el controlador predictivo para optimizar la función objetivo ante la peor situación posible de las incertidumbres.

Estos resultados prometedores permiten pensar que el control predictivo experimentará una mayor diseminación tanto en el mundo académico como en la comunidad industria. En este contexto uno de los mayores fabricantes de sistema de control distribuido, Honeywell, incorpora control predictivo robusto en uno de sus productos ("Robust Multivariable Predictive Control RMPCTM) para el TDC 3000.

El control predictivo se puede considerar una técnica madura para sistemas lineales y no muy rápidos como los encontrados normalmente en la industria de procesos. Sistemas más complejos, tales como sistemas no lineales, híbridos, y o sistemas muy rápidos, eran considerados como fuera del alcance de los controladores predictivos. Durante los últimos años se han producido resultados espectaculares en estos campos. Se ha demostrado (Bemporad *et al.*, 2002) que un controlador predictivo con restricciones lineales resulta ser un controlador afín a trozos que puede ser implementado con poca carga de computación. Recientemente han aparecido en la literatura aplicaciones de los controladores predictivos a sistemas no lineales o híbridos.

1.3 Tecnología Industrial

Aunque hay compañías que hacen uso de controladores predictivos que han desarrollado la tecnología dentro de sus empresas y no la ofrecen externamente, existe un número de ellas que suministran controladores predictivos, entre las que podemos destacar las siguientes con el acrónimo del producto que suministran :

- AspenTech: Dynamic Matrix Control (DMC)
- Adersa: Identification and Command (IDCOM), Hierarchical Constraint Control (HIECON) and Predictive Functional Control (PFC)
- Honeywell Profimatics: Robust Model Predictive Control Technology (RMPCT) and Predictive Control Technology (PCT)
- Setpoint Inc.: Setpoint Multivariable Control Architecture (SMCA) and IDCOM-M (multivariable)
- Treiber Controls: Optimum Predictive Control (OPC)
- ABB: 3dMPC
- Pavillion Technologies Inc.: Process Perfecter
- Simulation Sciences: Connoisseur

Nótese que cada producto no es solamente un algoritmo, sino que viene acompañado por paquetes adicionales para la identificación de la planta o realización de pruebas.

Existen miles de aplicaciones del control predictivo en la industria. La mayor parte de las aplicaciones están en el sector petroquímico (Qin y Badgwell, 1997) (Qin y Badgwell, 1998) en el área de refino pero también existen numerosas aplicaciones en los sectores de pulpa y papel, procesamiento de alimentos, gas, minería, hornos, metalurgia, industria aeroespacial e industria del automóvil. Una excelente revisión sobre MPC dirigida principalmente a personal de la industria con experiencia en control se puede encontrar en (Rawlings, 2000).

2. FORMULACIÓN DEL PROBLEMA

En la sección anterior se ha mostrado la forma genérica de resolver el problema MPC, que consiste básicamente en minimizar la función de coste haciendo uso del modelo del sistema para calcular las predicciones. La solución del problema depende principalmente del tipo de modelo que se use para capturar la dinámica del proceso. Aunque el procedimiento general es prácticamente el mismo, los pasos necesarios para formular los algoritmos de control son ligeramente distintos.

Se analizan aquí tres tipos de modelo: función de transferencia, modelos de convolución y modelo de espacio de estados.

2.1 Modelo de función de transferencia

La formulación más conocida que usa este tipo de modelo es sin duda el Control Predictivo Generalizado (Generalized Predictive Control, GPC) (Clarke *et al.*, 1987), aunque existen otras formulaciones que también usan modelos de función de transferencia (EPSAC y EHAC, por ejemplo). El GPC se ha convertido en uno de los métodos más populares tanto en el mundo industrial como en el académico y ha funcionado con éxito en muchas aplicaciones industriales (Clarke, 1988), pudiendo tratar plantas inestables y de fase no mínima a la vez que incorpora la idea de horizonte de control y la consideración de pesos en los incrementos de la señal de control.

En el GPC el modelo de la planta viene dado por una función de transferencia discreta en la forma de un modelo CARIMA (Controller Auto-Regressive Integrated Moving-Average):

$$A(z^{-1})y(t) = B(z^{-1})z^{-d} u(t-1) + C(z^{-1})\frac{e(t)}{\Delta} \quad (1)$$

donde $\Delta = 1 - z^{-1}$, $u(t)$ e $y(t)$ son las variables de entrada y salida respectivamente y $e(t)$ es un ruido blanco de media cero. A , B y C son polinomios en el operador retardo z^{-1} y d es el tiempo muerto del sistema. Este modelo es muy apropiado para muchas aplicaciones industriales en las que las perturbaciones no son estacionarias, según se justifica en (Clarke *et al.*, 1987). A partir de ahora el polinomio C se toma igual a 1. Nótese que si se puede truncar entonces puede ser incluido en A y B . En (Camacho y Bordons, 2004) se trata el caso general de ruido coloreado. En general es difícil determinar el valor real de este polinomio, por lo que se puede usar para rechazo óptimo de perturbaciones, aunque es más adecuado su papel en la mejora de la robustez.

El algoritmo GPC consiste en aplicar la secuencia de control que minimiza una función de coste de la forma:

$$J = \sum_{j=N_1}^{N_2} \delta(j)[\hat{y}(t+j|t) - w(t+j)]^2 + \sum_{j=1}^{N_u} \lambda(j)[\Delta u(t+j-1)]^2 \quad (2)$$

donde $\hat{y}(t+j|t)$ es la predicción óptima de la salida j pasos hacia delante calculada con datos conocidos en el instante t , N_1 y N_2 son los horizontes mínimo y máximo de predicción y N_u es el horizonte de control, $\delta(j)$ y $\lambda(j)$ son secuencias de ponderación (normalmente constantes) y $w(t+j)$ es la futura trayectoria de referencia.

Con objeto de minimizar la función de coste, hay que calcular la predicción óptima $y(t+j)$ para $j \geq N_1$ y $j \leq N_2$. Esto se lleva a cabo resolviendo una ecuación diofántica cuya solución se puede obtener mediante un algoritmo recursivo. Si se hace esto (Clarke *et al.*, 1987), se pueden escribir los valores futuros de la salida como:

$$y(t+j) = F_j y(t) + E_j B(z^{-1}) \Delta u(t+j-d-1) + E_j e(t+j) \quad (3)$$

Los polinomios E_j y F_j se derivan de la ecuación diofántica y vienen unívocamente definidos con grados $j-1$ y na respectivamente. Se pueden obtener dividiendo 1 entre $\tilde{A}(z^{-1})$ hasta que el resto se pueda factorizar como $z^{-j}F_j(z^{-1})$. El cociente de la división es el polinomio $E_j(z^{-1})$. Como el grado del polinomio $E_j(z^{-1}) = j-1$, los términos de ruido de la ecuación (3) se encuentran en el futuro y por tanto la mejor predicción viene dada por:

$$\hat{y}(t+j|t) = G_j(z^{-1})\Delta u(t+j-d-1) + F_j(z^{-1})y(t)$$

donde $G_j(z^{-1}) = E_j(z^{-1})B(z^{-1})$.

Para resolver el problema del GPC, es necesario obtener la secuencia de señales de control $u(t)$, $u(t+1)$, ..., $u(t+N)$ que optimice la expresión (2). Como el proceso tiene un tiempo muerto de d períodos de muestreo, la salida sólo se verá influenciada por la señal de control $u(t)$ tras $d+1$ períodos. Los horizontes se pueden definir por tanto como $N_1 = d+1$, $N_2 = d+N$ y $N_u = N$. Obsérvese que no tiene sentido hacer $N_1 < d+1$ ya que los términos que se añadirían a (2) sólo dependerían de acciones de control pasadas. Por otro lado, si $N_1 > d+1$ los primeros puntos de la predicción, que son los mejor estimados, no se tendrían en cuenta.

Considérese ahora el siguiente conjunto de predicciones de j pasos:

$$\begin{aligned}\hat{y}(t+d+1) &= G_{d+1} \Delta u(t) + F_{d+1}y(t) \\ \hat{y}(t+d+2) &= G_{d+2} \Delta u(t+1) + F_{d+2}y(t) \\ &\vdots \\ \hat{y}(t+d+N) &= G_{d+N} \Delta u(t+N-1) + F_{d+N}y(t)\end{aligned}$$

que puede escribirse como:

$$\mathbf{y} = \mathbf{G}\mathbf{u} + \mathbf{F}(z^{-1})y(t) + \mathbf{G}'(z^{-1}) \Delta u(t-1) \quad (4)$$

con

$$\begin{aligned}\mathbf{y} &= [\hat{y}(t+d+1|t) \dots \hat{y}(t+d+N|t)]^T \\ \mathbf{u} &= [\Delta u(t) \Delta u(t+1) \dots \Delta u(t+N-1)]^T\end{aligned}$$

$$\mathbf{G} = \begin{bmatrix} g_0 & 0 & \dots & 0 \\ g_1 & g_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{N-1} & g_{N-2} & \dots & g_0 \end{bmatrix}$$

$$\mathbf{G}'(z^{-1}) = \begin{bmatrix} (G_{d+1}(z^{-1}) - g_0)z \\ (G_{d+2}(z^{-1}) - g_0 - g_1z^{-1})z^2 \\ \vdots \\ (G_{d+N}(z^{-1}) - \dots - g_{N-1}z^{-(N-1)})z^N \end{bmatrix}$$

$$\mathbf{F}(z^{-1}) = \begin{bmatrix} F_{d+1}(z^{-1}) \\ F_{d+2}(z^{-1}) \\ \vdots \\ F_{d+N}(z^{-1}) \end{bmatrix}$$

Nótese que los dos últimos términos en la ecuación (4) dependen sólo del pasado y se pueden agrupar por tanto dentro del vector \mathbf{f} (respuesta libre) dando lugar a:

$$\mathbf{y} = \mathbf{G}\mathbf{u} + \mathbf{f}$$

Si las condiciones iniciales son nulas, la respuesta libre también lo es. Si se aplica un escalón unitario a la entrada en el instante t , es decir,

$$\Delta u(t) = 1, \Delta u(t+1) = 0, \dots, \Delta u(t+N-1) = 0$$

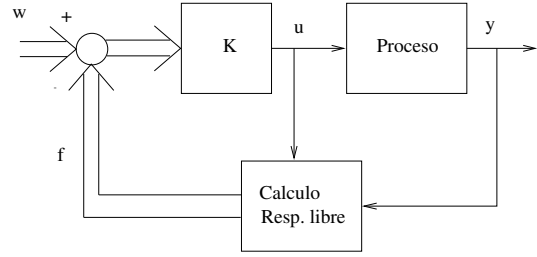


Fig. 4. Ley de control del GPC

la secuencia de salida $t [\hat{y}(t+1), \hat{y}(t+2), \dots, \hat{y}(t+N)]^T$ es igual a la primera columna de la matriz \mathbf{G} . Es decir, la primera columna de la matriz \mathbf{G} se puede calcular como la respuesta de la planta cuando se aplica un escalón unitario en la variable manipulable. La respuesta libre se puede calcular de forma recursiva como

$$\mathbf{f}_{j+1} = z(1 - \tilde{A}(z^{-1}))\mathbf{f}_j + B(z^{-1}) \Delta u(t-d+j)$$

con $\mathbf{f}_0 = y(t)$ y $\Delta u(t+j) = 0$ para $j \geq 0$.

Por otra parte, la función de coste se puede escribir como

$$J = (\mathbf{G}\mathbf{u} + \mathbf{f} - \mathbf{w})^T (\mathbf{G}\mathbf{u} + \mathbf{f} - \mathbf{w}) + \lambda \mathbf{u}^T \mathbf{u} \quad (5)$$

donde

$$\mathbf{w} = [w(t+d+1) \dots w(t+d+N)]^T$$

El mínimo de J , considerando que no hay restricciones se puede encontrar igualando a cero el gradiente de J , lo que conduce a:

$$\mathbf{u} = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T (\mathbf{w} - \mathbf{f}) \quad (6)$$

En realidad, al usar una estrategia de horizonte deslizante, la señal de control que verdaderamente se envía al proceso es el primer elemento del vector \mathbf{u} , dado por:

$$\Delta u(t) = \mathbf{K}(\mathbf{w} - \mathbf{f}) \quad (7)$$

donde \mathbf{K} es la primera fila de la matriz $(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T$. Este hecho tiene un claro significado, que se puede derivar fácilmente de la figura 4: si no hay errores futuros (predichos), es decir, si $\mathbf{w} - \mathbf{f} = \mathbf{0}$, entonces no hay movimientos en la señal de control, ya que el objetivo se satisfará con la evolución libre del proceso. Sin embargo, en cualquier otro caso existirá un incremento de la acción de control proporcional (con factor \mathbf{K}) al error futuro. Obsérvese que la acción se toma respecto a errores *futuros*, no a errores *pasados* como en el caso de los controladores realimentados convencionales

La solución del GPC precisa de la inversión (o al menos la triangularización) de una matriz de dimensión $N \times N$ que requiere una carga de cómputo considerable. En (Clarke *et al.*, 1987) se introduce el concepto de horizonte de control

para reducir la carga de cálculo, considerando que las señales de control futuras se mantendrán constantes tras $N_u < N$. Esto da lugar a la inversión de una matriz de dimensión $N_u \times N_u$ que reduce la carga de cálculo (en particular, si $N_u = 1$ se reduce a un escalar, aunque se restringe la optimalidad del GPC).

2.2 Modelo de convolución

Los modelos de convolución engloban los modelos de respuesta impulsional y respuesta ante escalón, de gran éxito en la industria por ser muy intuitivos y permitir un procedimiento de identificación relativamente sencillo. Estos tipos de modelo han dado lugar a dos de los controladores más extendidos en la práctica: Dynamic Matrix Control, (DMC) y Model Algorithmic Control (MAC). Se muestra aquí la formulación del primero de ellos, siendo la del segundo muy similar.

El controlador DMC se desarrolló a finales de los años setenta del siglo pasado por Cutler y Ramaker (Cutler y Ramaker, 1980) de la Shell Oil Co. y ha sido ampliamente aceptado en el entorno industrial, principalmente por la industria petroquímica (Qin y Badgwell, 1997). Su gran éxito proviene de su capacidad para tratar procesos multivariables, aunque en esta sección sólo se aborda el caso monovariante ya que permite comprender mejor los fundamentos sin perderse en notación matemática.

Como se ha indicado previamente, se usa un modelo de respuesta ante escalón para capturar la dinámica del proceso, mientras que la perturbación se considera constante a lo largo del horizonte. El procedimiento para obtener las predicciones se muestra a continuación.

Al usar el siguiente modelo de respuesta ante escalón

$$y(t) = \sum_{i=1}^{\infty} g_i \Delta u(t-i)$$

los valores predichos de la salida a lo largo del horizonte serán:

$$\begin{aligned} \hat{y}(t+k|t) &= \sum_{i=1}^{\infty} g_i \Delta u(t+k-i) + \hat{n}(t+k|t) \\ &= \sum_{i=1}^k g_i \Delta u(t+k-i) + \sum_{i=k+1}^{\infty} g_i \Delta u(t+k-i) \\ &\quad + \hat{n}(t+k|t) \end{aligned}$$

Como se considera que las perturbaciones son constantes en el futuro, $\hat{n}(t+k|t) = \hat{n}(t|t) = y_m(t) - \hat{y}(t|t)$, se puede escribir:

$$\begin{aligned} \hat{y}(t+k|t) &= \sum_{i=1}^k g_i \Delta u(t+k-i) + y_m(t) \\ &+ \sum_{i=k+1}^{\infty} g_i \Delta u(t+k-i) - \sum_{i=1}^{\infty} g_i \Delta u(t-i) \end{aligned}$$

Es decir, la predicción viene dada por:

$$\hat{y}(t+k|t) = \sum_{i=1}^k g_i \Delta u(t+k-i) + f(t+k)$$

donde $f(t+k)$ es la respuesta libre del sistema, que como se sabe no depende de las acciones de control futuras, y viene dada por:

$$f(t+k) = y_m(t) + \sum_{i=1}^{\infty} (g_{k+i} - g_i) \Delta u(t-i) \quad (8)$$

Si el proceso es asintóticamente estable los coeficientes g_i de la respuesta ante escalón tienden a valores constantes tras un cierto número N de periodos de muestreo, por lo que se puede considerar que

$$g_{k+i} - g_i \approx 0, \quad i > N$$

y por lo tanto se puede calcular la respuesta libre como

$$f(t+k) = y_m(t) + \sum_{i=1}^N (g_{k+i} - g_i) \Delta u(t-i)$$

Nótese que si el sistema no es asintóticamente estable, no existe tal valor N y $f(t+k)$ no se puede calcular (aunque existe una generalización en el caso de que la inestabilidad sea debida a integradores puros).

Ahora se pueden calcular las predicciones a lo largo del horizonte de predicción ($k = 1, \dots, p$), considerando m acciones de control:

$$\begin{aligned} \hat{y}(t+1|t) &= g_1 \Delta u(t) + f(t+1) \\ \hat{y}(t+2|t) &= g_2 \Delta u(t) + g_1 \Delta u(t+1) + f(t+2) \\ &\vdots \\ \hat{y}(t+p|t) &= \sum_{i=p-m+1}^p g_i \Delta u(t+p-i) + f(t+p) \end{aligned}$$

Si se define ahora la *matriz dinámica* del sistema \mathbf{G} (que da nombre al algoritmo) como

$$\mathbf{G} = \begin{bmatrix} g_1 & 0 & \cdots & 0 \\ g_2 & g_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_m & g_{m-1} & \cdots & g_1 \\ \vdots & \vdots & \ddots & \vdots \\ g_p & g_{p-1} & \cdots & g_{p-m+1} \end{bmatrix}$$

se puede escribir la predicción como

$$\hat{\mathbf{y}} = \mathbf{G}\mathbf{u} + \mathbf{f} \quad (9)$$

que tiene la misma forma que las predicciones calculadas por el GPC. Nótese que \mathbf{G} está formada por m (horizonte de control) columnas de la respuesta ante escalón del sistema deslizadas hacia abajo de forma apropiada. $\hat{\mathbf{y}}$ es un vector de dimensión p que contiene las predicciones, \mathbf{u} tiene dimensión m y contiene los incrementos en las acciones de control y \mathbf{f} es el vector de la respuesta libre.

Nótese que \mathbf{f} depende del vector de estados $x(t)$, que en este caso viene dado por $x(t)^T = [y_m(t) \ u(t-1) \ u(t-2) \ \dots \ u(t-N-1)]$, y se puede expresar como $\mathbf{f} = \mathbf{F}x(t)$, con lo que se puede escribir la predicción como:

$$\hat{\mathbf{y}} = \mathbf{G}\mathbf{u} + \mathbf{F}x(t)$$

Las variables manipuladas se calculan, como en cualquier controlador predictivo, de forma que minimicen la siguiente función objetivo cuadrática, que incluye el esfuerzo de control:

$$J = \sum_{j=1}^p [\hat{y}(t+j|t) - w(t+j)]^2 + \sum_{j=1}^m \lambda [\Delta u(t+j-1)]^2$$

Si no hay restricciones, la solución a la minimización de la función de coste $J = \mathbf{e}\mathbf{e}^T + \lambda\mathbf{u}\mathbf{u}^T$, donde \mathbf{e} es el vector de errores futuros a lo largo del horizonte, se puede obtener de forma analítica igualando a cero la derivada de J (igual que en el GPC), proporcionando:

$$\mathbf{u} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^T(\mathbf{w} - \mathbf{f})$$

Conviene recordar que, como ocurre en todas las estrategias de control predictivo, sólo el primer elemento del vector \mathbf{u} ($\Delta u(t)$) se envía realmente a la planta. No es aconsejable implementar la secuencia de control calculada completa (m valores), ya que es imposible estimar perfectamente las perturbaciones en el horizonte y por tanto es imposible anticiparse de forma precisa a las inevitables perturbaciones que hacen que la salida real difiera de las predicciones que se han usado en el cálculo. Además, la referencia podría cambiar en dicho intervalo.

Como resumen, obsérvese que ambos controladores descritos hasta ahora tienen un procedimiento similar, dando lugar a leyes de control

donde la única diferencia es el cálculo de la respuesta libre, debido al distinto tipo de modelo usado. En ambos casos si existen restricciones, se pasa de una solución analítica a la resolución de un problema de programación cuadrática, como se verá más adelante.

2.3 Modelo en el espacio de estados

Los modelos de espacio de estados o descripción interna se pueden usar también para formular el problema del control predictivo. Además, los principales resultados teóricos relacionados con la estabilidad provienen de este tipo de formulación, que puede ser usada tanto para problemas monovariabes como multivariabes y se puede extender fácilmente al caso no lineal.

Se usan las siguientes ecuaciones para capturar la dinámica del proceso:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \quad (10)$$

En el caso monovariable $y(t)$ y $u(t)$ son escalares y $x(t)$ es el vector de estados. Un proceso multivariable tiene la misma descripción pero con el vector u de dimensión m y el vector y de dimensión n . En esta sección, por simplicidad en la notación, sólo se considera el primer caso. La extensión al caso multivariable es directa.

También se puede usar un modelo incremental si se considera como variable de entrada el incremento de la señal de control $\Delta u(t)$ en lugar de $u(t)$. Este modelo se puede escribir en la forma genérica de espacio de estados simplemente teniendo en cuenta que $\Delta u(t) = u(t) - u(t-1)$. Combinando esta expresión con (10) se obtiene la siguiente representación:

$$\begin{aligned} \begin{bmatrix} x(t+1) \\ u(t) \end{bmatrix} &= \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \begin{bmatrix} x(t) \\ u(t-1) \end{bmatrix} + \begin{bmatrix} B \\ I \end{bmatrix} \Delta u(t) \\ y(t) &= [C \ 0] \begin{bmatrix} x(t) \\ u(t-1) \end{bmatrix} \end{aligned}$$

Al definir el nuevo vector de estados como $\bar{x}(t) = [x(t) \ u(t-1)]^T$, el modelo incremental toma la forma general:

$$\begin{aligned} \bar{x}(t+1) &= M\bar{x}(t) + N \Delta u(t) \\ y(t) &= Q\bar{x}(t) \end{aligned} \quad (11)$$

donde la relación entre (M, N, Q) y las matrices de la forma no-incremental (A, B, C) se pueden derivar fácilmente combinando (10) y (11).

Con la idea de minimizar la función objetivo, hay que calcular las predicciones a lo largo del

horizonte. En el caso del modelo incremental, éstas se obtienen usando (11) de manera recursiva, dando lugar a:

$$\hat{y}(t+j) = QM^j \hat{x}(t) + \sum_{i=0}^{j-1} QM^{j-i-1} N \Delta u(t+i)$$

Nótese que las predicciones necesitan una estimación insesgada del vector de estados $x(t)$. Si éste no es accesible será necesario incluir un observador, que calcula la estimación por medio de

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(y_m(t) - y(t|t-1))$$

donde $y_m(t)$ es la salida medida. Si la planta está sujeta a perturbaciones en forma de ruido blanco que afectan a la salida y al proceso con matrices de covarianza conocidas, el observador se convierte en un filtro de Kalman (Aström y Wittenmark, 1984) y la ganancia K se calcula resolviendo una ecuación de Riccati.

Ahora, las predicciones a lo largo del horizonte vienen dadas por:

$$\mathbf{y} = \begin{bmatrix} \hat{y}(t+1|t) \\ \hat{y}(t+2|t) \\ \vdots \\ \hat{y}(t+N_2|t) \end{bmatrix} = \begin{bmatrix} QM\hat{x}(t) + QN \Delta u(t) \\ QM^2\hat{x}(t) + \sum_{i=0}^1 QM^{1-i} N \Delta u(t+i) \\ \vdots \\ QM^{N_2}\hat{x}(t) + \sum_{i=0}^{N_2-1} QM^{N_2-1-i} N \Delta u(t+i) \end{bmatrix}$$

que se puede expresar de forma vectorial como

$$\mathbf{y} = \mathbf{F}\hat{x}(t) + \mathbf{H}\mathbf{u} \quad (12)$$

donde $\mathbf{u} = [\Delta u(t) \ \Delta u(t+1) \ \dots \ \Delta u(t+N_u-1)]^T$ es el vector de incrementos de control futuros, \mathbf{H} es una matriz triangular inferior por bloques cuyos elementos no nulos vienen definidos por $\mathbf{H}_{ij} = QM^{i-j}N$ y la matriz \mathbf{F} viene dada por

$$\mathbf{F} = \begin{bmatrix} QM \\ QM^2 \\ \vdots \\ QM^{N_2} \end{bmatrix}$$

Nótese que la expresión (12) está compuesta por dos términos: el primero de ellos depende del estado actual y por lo tanto se conoce en el instante t , mientras que el segundo depende las acciones de control futuras, que son las variables de decisión que deben ser calculadas. La secuencia de control \mathbf{u} se calcula minimizando la función

objetivo, que (en el caso de $\delta(j) = 1$ y $\lambda(j) = \lambda$) se puede escribir como:

$$J = (\mathbf{H}\mathbf{u} + \mathbf{F}\hat{x}(t) - \mathbf{w})^T (\mathbf{H}\mathbf{u} + \mathbf{F}\hat{x}(t) - \mathbf{w}) + \lambda \mathbf{u}^T \mathbf{u}$$

Si no hay restricciones, la solución analítica que proporciona el óptimo es:

$$\mathbf{u} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T (\mathbf{w} - \mathbf{F}\hat{x}(t))$$

Como se usa una estrategia de control deslizante, sólo se envía el primer elemento de la secuencia, $\Delta u(t)$, y se vuelven a repetir los cálculos en el siguiente periodo de muestreo. Nótese que se necesita un observador, ya que la ley de control depende de $\hat{x}(t)$.

Se debe notar que cuando los horizontes de control y de predicción tienden a infinito y no existen restricciones, este controlador se convierte en el conocido problema de regulador lineal óptimo cuadrático (LQR). La secuencia de control óptima se genera entonces por una realimentación lineal del vector de estado donde la matriz de ganancia se calcula mediante la resolución de una ecuación de Riccati. Esta equivalencia permite un estudio teórico de los problemas de MPC basado en los resultados del campo del Control Óptimo, como es el caso de la estabilidad en bucle cerrado.

Si se emplea el modelo de espacio de estados de la ecuación (10), las predicciones se calculan de forma ligeramente diferentes, como se muestra en (Maciejowski, 2001). Sea cual sea el modelo usado la ley de control es una realimentación del vector de estado que necesita un observador.

2.4 Perturbaciones medibles

Muchos procesos se encuentran sometidos a perturbaciones externas provocadas por cambios en las variables que se pueden medir. Ésta es una situación típica en procesos cuyas salidas se ven afectadas por variaciones en el régimen de carga. Considérese, por ejemplo, un reactor encamisado donde la temperatura se controla manipulando el caudal de agua que entra en la camisa refrigerante. Cualquier variación del caudal de reactivo influirá en la temperatura del reactor. Este tipo de perturbaciones, también conocidas como perturbaciones en la carga, pueden ser abordadas mediante el uso de acción *feedforward*. Como se verá a continuación, las perturbaciones conocidas pueden ser consideradas de forma explícita en MPC.

La idea es que si se conoce el efecto de estas perturbaciones en la salida del sistema entonces pueden ser incluidas en las predicciones. Esto permite que los controladores predictivos incluyan acción *feedforward*, usada con gran éxito en conjunción con los controladores PID. Las perturbaciones medibles se pueden incluir fácilmente en

las ecuaciones de predicción, ya que pueden ser tratadas como entradas al sistema.

Para sistemas lineales, el principio de superposición permite calcular las predicciones de la salida como la suma del efecto de ambas entradas: la variable manipulable y la perturbación medible. Por tanto, cualquiera que sea el controlador usado (GPC, DMC, etc.), la predicción viene dada por:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}_u + \hat{\mathbf{y}}_d = \mathbf{G} \mathbf{u} + \mathbf{f} + \mathbf{D} \mathbf{d} + \mathbf{f}_d$$

donde $\hat{\mathbf{y}}_d$ es la contribución de la perturbación medible a la salida del sistema, \mathbf{D} es una matriz similar a \mathbf{G} que contiene los coeficientes de la respuesta del sistema a un escalón en la perturbación, \mathbf{d} es el vector de incrementos en la perturbación y \mathbf{f}_d es la parte de la respuesta que no depende de la perturbación.

El término $\mathbf{D} \mathbf{d}$ depende de las perturbaciones futuras. En algunos casos, cuando éstas están relacionadas con la carga, las perturbaciones futuras son conocidas o se pueden predecir usando tendencias u otros medios (ver por ejemplo (Bordons y Cueli, 2004)). Si éste es el caso, el término correspondiente a las perturbaciones futuras puede ser calculado. En la mayoría de los casos se considera que las futuras perturbaciones de carga van a ser constantes durante el horizonte e iguales al último valor medido (es decir, $d(t+j) = d(t)$) y entonces $\Delta d(t+j) = 0$ y el segundo término de esta ecuación desaparece.

Entonces, la predicción viene dada en la forma genérica de respuesta libre y forzada:

$$\hat{\mathbf{y}} = \mathbf{G} \mathbf{u} + \mathbf{f}'$$

donde \mathbf{f}' incluye el efecto de la respuesta libre debida a la variable manipulada y el efecto de la perturbación, que es parte de la nueva respuesta libre, ya que no depende de la secuencia de acciones de control y por tanto es una constante que no influye en la minimización. Esta ecuación es el único cambio que se debe hacer en la formulación del controlador para incluir el efecto de las perturbaciones medibles.

3. RESTRICCIONES EN CONTROL PREDICTIVO

El problema de control se ha formulado en las secciones anteriores considerando que todas las señales poseían un rango ilimitado. Esta suposición no es demasiado realista, ya que en la práctica todos los procesos están sometidos a restricciones. Los actuadores tienen un rango limitado de actuación y una velocidad de movimiento limitadas, como ocurre con las válvulas de control

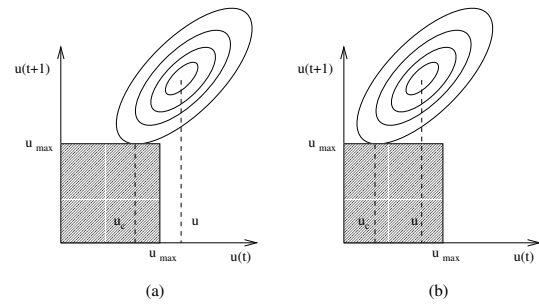


Fig. 5. Restricciones en la señal de control

que están limitadas por la posición correspondiente a su cierre total o a su apertura total. Por razones constructivas o de seguridad las variables de proceso están también limitadas, tal es el caso, por ejemplo, de los niveles en tanques o caudales en tuberías o la presión en depósitos. Mas aún, en muchos casos, el punto de operación de la planta se elige para satisfacer objetivos económicos que suelen estar situados en los límites de operación del proceso, por lo que es previsible que el proceso operando en esos entornos y debido a las perturbaciones se salga de límites.

La forma usual de tratar ese problema es calcular la señal de control $u(t)$ y aplicarla al proceso. Si $u(t)$ viola las restricciones, se satura a su límite bien por el programa de control que vigila este aspecto o por el propio actuador que no puede dar mayor actuación que la máxima. El caso en que la violación se vaya a producir en los instantes de tiempo futuro $u(t+1), \dots, u(t+N)$ ni siquiera se considera. Esta forma de operar no garantiza la optimalidad del problema cuando las restricciones son violadas, por lo que el principal objetivo del control predictivo que es calcular la *mejor* señal de control posible optimizando un índice de funcionamiento no se cumple. Para ilustrar ese punto considérense los casos de violación de restricciones de la figura 5 de un controlador predictivo con horizonte dos. La figura 5(a) muestra el caso donde $u(t) > u_{\max}$. En este caso lo normal sería aplicar u_{\max} al proceso en lugar de u_c que es el verdadero mínimo cuando se consideran restricciones. En el caso que se muestra en la figura 5(b), $u(t)$ no viola las restricciones y sería aplicada al sistema en lugar de la señal u_c que es la que se debía haber aplicado.

El no considerar plenamente las restricciones en las variables manipuladas puede resultar en mayores valores de la función objetivo y por lo tanto peores resultados. En cualquier caso las variables manipuladas pueden ser siempre acotadas por el programa de control o por los propios actuadores que no van más allá de sus límites. El principal motivo para considerar restricciones consiste en que violar las variables controladas puede ser mucho más costoso y peligroso ya que esto puede originar daños en los equipos y pérdidas en la

producción. Por ejemplo, en la mayoría de los reactores por lotes, la calidad de la producción requiere que las variables del proceso se mantengan dentro de límites durante la reacción; violar estos límites puede llevar a una producción de calidad inferior a la requerida. Cuando los límites han sido impuestos debido a razones de seguridad, la violación de estos límites puede producir daños en los equipos, derrames de productos nocivos, o en la mayoría de los casos que se activen los enclavamientos de seguridad produciendo paradas de emergencia con las pérdidas aparejadas de producción y costosos procesos de puesta en marcha.

Al contrario de lo que ocurre con las variables manipuladas, que en último extremo podemos saturar para mantenerlas dentro de límites, las variables de proceso son fijadas por este y no podemos manipularlas a voluntad. Más aún, la violación de las restricciones en una variable de proceso puede ser debida a acciones previamente tomadas. Una de las ventajas del control predictivo es su capacidad de predicción y anticiparse a la violación de las restricciones.

Las restricciones que normalmente actúan en un proceso sobre las variables manipuladas, la velocidad de cambio de las variables manipuladas y las variables de proceso se pueden expresar como:

$$\begin{aligned} \underline{U} &\leq u(t) \leq \bar{U} \quad \forall t \\ \underline{u} &\leq u(t) - u(t-1) \leq \bar{u} \quad \forall t \\ \underline{y}(t) &\leq y(t) \leq \bar{y}(t) \quad \forall t \end{aligned}$$

Nótese que el último conjunto de restricciones fuerza a las variables a seguir trayectorias dentro de bandas (normalmente formadas por una trayectoria nominal a seguir más una banda de tolerancia).

En algunos casos, el estado final del proceso (estado que se debe alcanzar al final del horizonte de predicción) se fuerza a alcanzar un conjunto terminal con el fin de satisfacer condiciones de operación o de estabilidad. Es fácil ver que el conjunto terminal induce un conjunto de restricciones sobre el vector de acciones de control (ver (Camacho y Bordons, 2004)).

3.1 Forma general de las restricciones

La mayor parte de las restricciones que pueden actuar sobre un proceso pueden ser descritas por

$$\mathbf{R}\mathbf{u} \leq \mathbf{r} + \mathbf{V}\mathbf{x}(t) \quad (13)$$

Las matrices \mathbf{R} , \mathbf{r} , y \mathbf{V} dependen de los parámetros del proceso y de los valores límites de las variables del proceso, por lo que sólo tienen que ser calculadas cada vez que los parámetros o los límites cambien (lo que suele ocurrir con poca frecuencia).

El último sumando de la expresión (13) depende del estado del proceso y ha de ser recalculado cada instante de muestreo.

El problema completo puede expresarse como

$$\begin{aligned} \min_{\mathbf{u}} \quad & \frac{1}{2} \mathbf{u}^T \mathbf{H}\mathbf{u} + \mathbf{b}\mathbf{u} + \mathbf{f}_0 \quad (14) \\ \text{s.a:} \quad & \mathbf{R}\mathbf{u} \leq \mathbf{r} + \mathbf{V}\mathbf{x}(t) \quad (15) \end{aligned}$$

Este problema recibe el nombre de problema de programación cuadrática (QP). Desafortunadamente no existe una solución explícita del problema anterior como en el caso de MPC sin restricciones. El problema QP se tiene que resolver aplicando algún algoritmo numérico. Aunque los algoritmos QP son muy eficientes y algunos de ellos tienen un tiempo de ejecución garantizado, la solución del problema con restricciones es mucho más compleja que el caso sin restricciones

3.2 Optimización y restricciones

La implementación del control predictivo es más difícil que la de los esquemas de control clásicos basados en controladores PID, requiriendo más tiempo y personal con más conocimiento de control. Para empezar hay que encontrar un modelo del sistema, lo que requiere un conjunto de pruebas sobre el proceso que en muchos casos implican perturbar las condiciones normales de operación del proceso. Los equipos de control que se requieren son más complicados, con computadores más potentes, y software de control más caros.

A pesar de todas estas dificultades, el control predictivo ha demostrado que es económicamente rentable, reduciendo los costes de operación y aumentando la producción y la calidad. La razón de este éxito radica en que el control predictivo opera optimizando funciones de costo y teniendo en cuenta las restricciones explícitamente, lo que permite:

- optimizar el punto de operación,
- optimizar las transiciones del proceso de un punto otro
- minimizar la varianza de operación del proceso. Una varianza más pequeña permite aumentar la calidad del producto y también operar más cerca de las restricciones. Como la mayoría de los procesos están sometidos a restricciones el punto óptimo suele estar en la intersección del conjunto de restricciones, el operar más cerca de estos puntos implica operar más cerca del óptimo. Considérese, por ejemplo, un proceso donde la producción (caudal, por ejemplo) depende de la presión de operación como se ilustra en la figura 6. Al estar la presión de operación

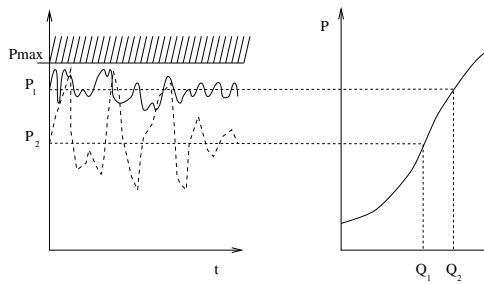


Fig. 6. Puntos de operación óptimos y restricciones

limitada a p_{\max} es evidente que para maximizar la producción se tendría que operar en ese punto. Cuando el proceso está sometido a perturbaciones, no podemos operar en ese punto porque cualquier perturbación positiva haría al proceso salir de la zona permitida y posiblemente disparar los enclavamientos que harían parar al proceso en el mejor de los casos. Si el controlador es capaz de mantener la varianza pequeña, el punto de consigna se puede situar mucho más cercano tal como se ilustra en la figura 6.

- La consideración explícita de las restricciones permite eliminar o al menos reducir el número de violaciones de restricciones y de paradas de emergencia por lo que el proceso puede operar más cerca del óptimo.

3.3 Gestión de restricciones

En algunos casos, a la hora de optimizar, la región definida por las restricciones está vacía, es decir, el problema de optimización no es factible y no se puede encontrar una solución. En algunos casos la falta de factibilidad está provocada por unos objetivos de control (en forma de restricciones) demasiado estrictos en otras porque una perturbación aleja al proceso de la región permitida y no existe ninguna secuencia de control capaz de meter a las variables del proceso en la región permitida en los próximos instantes de muestreo.

La no-factibilidad puede aparecer por las especificaciones del régimen estacionario cuando los valores de consigna que se fijan no son alcanzables debido a que las variables manipuladas están acotadas. Cuando las variables manipulables están acotadas por un poliedro (en general un hiper-cubo), la región alcanzable en régimen permanente es el poliedro cuyos vértices se obtienen a partir de los vértices del poliedro delimitando las variables manipulables por la matriz de ganancia estáticas del proceso. Para evitar problemas de factibilidad debido a puntos de consigna inalcanzables sólo hay que verificar que el vector de consignas está incluido dentro del poliedro de valores alcanzables por las variables a controlar en régimen permanente.

La no factibilidad puede aparecer en el régimen transitorio; en algunos casos cuando una perturbación aparta al proceso de la zona admisible, tal que es imposible introducirla de nuevo en la zona permitida en los instantes de muestreo próximos con una señal manipulable de energía limitada. En otros casos, el operador puede variar los límites operacionales de las variables, haciendo también el problema no factible como se menciona en (Alvarez y Prada, 1997). La factibilidad es crucial para el control predictivo ya que el algoritmo numérico no produce una solución si no es factible.

Dado que las circunstancias descritas pueden ocurrir y que los algoritmos de optimización utilizados por los controladores predictivos necesitan que existan soluciones factibles, es necesario tratar los problemas de factibilidad. En (Camacho y Bordons, 2004) se dan algunas ideas para tratar los problemas de no factibilidad que se puede resumir en lo que sigue:

- (1) Desconexión del controlador predictivo y utilización de un controlador *backup*.
- (2) Eliminación de restricciones hasta que el problema sea factible. La eliminación es temporal y se suele realizar de forma jerarquizada.
- (3) Relajación de restricciones: algunas de las restricciones *duras* ($\mathbf{Ru} < \mathbf{a}$) se convierten en restricciones *blandas* ($\mathbf{Ru} < \mathbf{a} + \epsilon$) introduciendo variables de holgura (ϵ) que se incluye en la función objetivo añadiendo $\epsilon^T \mathbf{T} \epsilon$ con una fuerte penalización.
- (4) Modificación de los horizontes de control: se utiliza cuando la no-factibilidad ocurre en los primeros instantes debido a que una perturbación ha sacado al proceso de la zona factible. El horizonte de control (a efectos de cumplir restricciones) se desliza; eliminándose de esta forma las restricciones imposibles de cumplir.

3.4 MPC Multiobjetivo

Todas las estrategias MPC analizadas hasta ahora se basan en optimizar una única función de coste, usualmente cuadrática, con objeto de determinar la secuencia de control futura que consigue el mejor comportamiento de la planta. Sin embargo, en muchos problemas de control el comportamiento del proceso no se puede representar por una simple función objetivo, sino que, la mayoría de las veces, existen objetivos de control diferentes y a veces conflictivos. Las razones para objetivos múltiples son variadas. La operación del proceso puede ser diferente en distintos puntos de trabajo (por ejemplo, tiempo mínimo en el arranque y varianza mínima en régimen nominal) e incluso en un mismo punto de trabajo, el objetivo de control

puede depender del valor de las variables (en el caso de que una variable controlada se salga de límites, puede ser prioritario minimizar su error en detrimento de otras variable).

Además, en muchos casos, el objetivo de control no es optimizar la suma de los errores al cuadrado, sino mantener ciertas variables dentro de los límites marcados. Este tipo de objetivos se puede expresar como una penalización de la cantidad que la variable se excede de su límite.

Algunas veces todos los objetivos de control se pueden representar con una sola función objetivo. Considérese, por ejemplo, un proceso con una serie de objetivos de control J_1, J_2, \dots, J_m , cada uno de ellos con restricciones lineales. Alguno de los objetivos puede ser mantener algunas variables lo más cerca posible de sus referencias mientras que otros pueden estar relacionados con mantener ciertas variables dentro de ciertas regiones. La secuencia de control se obtendrá resolviendo el siguiente problema de optimización:

$$\begin{aligned} \min J &= \sum_{i=1}^m \beta_i J_i \\ \text{s. a. } \mathbf{R}_i \mathbf{u} &\leq \mathbf{a}_i \text{ for } i = 1, \dots, m \end{aligned}$$

La importancia de cada objetivo se puede ponderar con la elección de los β_i . Desde luego esto no es trivial, ya que es muy difícil determinar el conjunto de pesos que representarán la importancia relativa de los objetivos de control. Además, los objetivos de control en la práctica son a veces cualitativos, haciendo muy difícil la tarea de determinar los pesos.

En algunos casos, la importancia relativa de los objetivos de control se puede establecer mediante priorización. Es decir, los objetivos de mayor prioridad (por ejemplo los relacionados con la seguridad) deben ser satisfechos antes de considerar otros de prioridad menor. Nótese que aunque los objetivos se pueden priorizar dándole mayor valor a los pesos, esto resulta una tarea complicada.

Considérese un proceso con una serie de m objetivos de control priorizados, O_i , de tal manera que el objetivo O_i tiene mayor prioridad que el O_{i+1} y que los objetivos se pueden expresar como:

$$\mathbf{R}_i \mathbf{u} \leq \mathbf{a}_i$$

El procedimiento para tener en cuenta dicha prioridad consiste en introducir variables enteras L_i que toman valor 1 cuando el objetivo se cumple y 0 en caso contrario. Los objetivos se expresan como:

$$\mathbf{R}_i \mathbf{u} \leq \mathbf{a}_i + K_i(1 - L_i) \quad (16)$$

donde K_i es un límite superior conservador de $\mathbf{R}_i \mathbf{u} - \mathbf{a}_i$. Si se cumple el objetivo O_i , $L_i = 1$

y el objetivo reformulado coincide con el objetivo de control original. Al introducir K_i , el objetivo reformulado (restricción) se satisface siempre incluso cuando no lo hace el correspondiente objetivo de control O_i ($L_i = 0$).

La priorización de los objetivos se establece imponiendo las siguientes restricciones:

$$L_i - L_{i+1} \geq 0 \text{ for } i = 1, \dots, m - 1$$

El problema es maximizar el número de objetivos de control satisfechos:

$$\sum_{i=1}^m L_i$$

Si el modelo del proceso es lineal, el problema se puede resolver mediante programación lineal mixta (Mixed Integer Linear Programming, MILP). El número de variables enteras se puede reducir tal como se indica en (Tyler y Morari, 1996). La idea es usar la misma variable L_i para restricciones que no pueden ser violadas al mismo tiempo, como es el caso de los límites inferior y superior de la misma variable. Aunque hay algoritmos eficientes para resolver este tipo de problemas, la carga de cálculo es mucho mayor que la de problemas del tipo LP o QP.

4. EJEMPLO

Esta sección muestra un ejemplo realizado con una herramienta que puede ser empleada para simular la mayoría de los controladores predictivos existentes. Este paquete de programas permite al usuario centrarse sólo en la definición del problema y el ajuste del controlador, olvidándose de la carga de programación del algoritmo. La herramienta se llama IMPACT, y se trata de un programa que se ejecuta en el entorno Matlab con una cómoda interface de usuario y que se puede descargar de forma gratuita de www.esi.us.es/MPCBOOK. Este paquete de software permite la simulación de la mayoría de los algoritmos existentes (GPC, DMC, PFC, state-space MPC, etc.) para sistemas tanto monovariabes como multivariabes con la inclusión de restricciones, incertidumbres, ruido y cambios en la referencia.

El proceso elegido se trata de una columna de destilación que es usada como banco de pruebas para estrategias de control de destilación y se conoce en la literatura como *Shell Oil's heavy oil fractionator* (Arruda *et al.*, 1994), (Chia y Brosilow, 1991).

El proceso y el experimento están descritos con detalle en (Camacho y Bordons, 2004). Hay tres

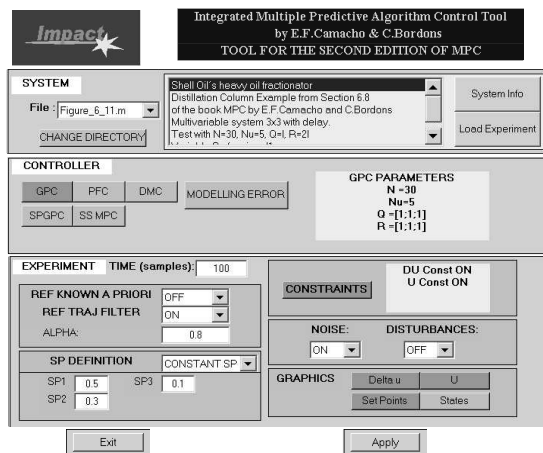


Fig. 7. Interface de usuario

variables que tienen que ser controladas: las composiciones de cabeza (y_1) y lateral (y_2) (medidas mediante analizadores) y la temperatura de cola (y_3). Las variables manipuladas son las extracciones superior (u_1) y media (u_2) y el caudal de reflujo (u_3).

La composición de cabeza se debe mantener restringida entre unos valores mínimo y máximo de -0.5 y 0.5 . Las variables manipuladas están también restringidas de forma que todas las extracciones estén entre -0.5 y 0.5 , al igual que el caudal de reflujo. A su vez, tienen unas variaciones máximas de 0.05 por minuto.

IMPACT resulta especialmente útil en este caso, ya que el esfuerzo de programar un controlador multivariable es muy elevado. Para llevar a cabo una simulación, el usuario debe definir el sistema, el controlador y el experimento, es decir, elegir qué controlador quiere usar, introducir los coeficientes del modelo y definir las restricciones y los ruidos e incertidumbres en su caso.

La herramienta lleva un conjunto de ejemplos ya incorporados, no siendo necesario introducir los coeficientes del modelo. Este ejemplo se encuentra en la biblioteca de experimentos y por tanto basta con cargar el fichero. La pantalla que se le presenta al usuario se muestra en la figura 7.

Una vez que el modelo se ha cargado y se ha dado valor a los parámetros de sintonía puede empezar la simulación. El usuario puede probar con diferentes parámetros y añadir incertidumbres y ruido para obtener el controlador que mejor se ajuste a su planta real en distintas situaciones. La potencia de esta herramienta se aprecia en la fase de diseño del controlador, ya que se puede llevar a cabo una gran cantidad de experimentos en la planta simulada antes de aplicarlo al proceso real.

Los resultados que se obtienen al aplicar un GPC multivariable al citado proceso se muestran en la figura 8 para el caso de un horizonte de predicción

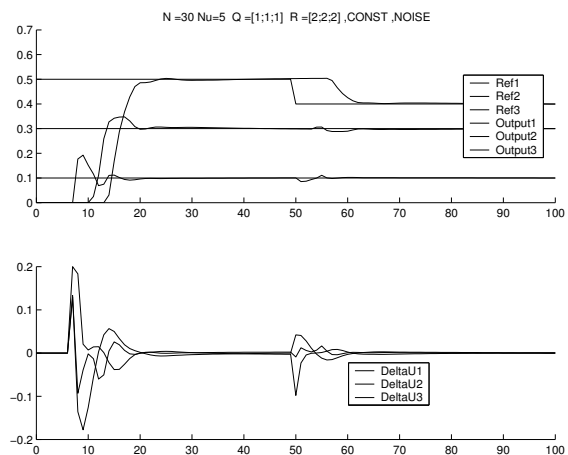


Fig. 8. Columna de destilación

común a todas las salidas de 30 y uno de control de 5 con un ruido añadido a las salidas. Las matrices de ponderación se han tomado como $Q = I$ y $R = 2I$. Hacia la mitad de la simulación se provocó un cambio en la referencia de la composición de cabeza de 0.5 a 0.4 . Las restricciones anteriormente definidas también se han incluido en el experimento y se muestran en la misma figura. Nótese que el incremento de las variables manipuladas están restringidas a un valor de 0.2 , que equivale a 0.05 por minuto, ya que el tiempo de muestreo es de 4 minutos. Durante el primer cambio de referencia, la primera señal de control trata de responder de forma rápida para alcanzar cuanto antes el valor de referencia, pero el controlador la mantiene restringida a su valor máximo durante un periodo de muestreo.

5. CONTROL PREDICTIVO ROBUSTO

Los modelos matemáticos son necesariamente simplificaciones de la realidad, especialmente en el caso de los modelos de control que han de tener estructuras simples (lineal en la mayor parte de los casos) y con un orden suficientemente pequeño.

La mayor parte de las técnicas de diseño de control necesitan un modelo de la planta, con estructuras y parámetros conocidos (*modelo nominal*), que se usa en el diseño. Si el modelo de control fuera exacto, en lugar de aproximado y no hubiera perturbaciones externas podría ser controlado por un controlador en bucle abierto. La realimentación es necesaria en control debido a las perturbaciones externas y a los errores de modelado. El objetivo de un controlador robusto es hacer frente a los errores en el modelo. Aunque la realimentación hace frente a estos problemas de forma implícita, el término Control Robusto se utiliza en la literatura para aquellos controladores que tienen en cuenta las incertidumbres de forma implícita.

Normalmente se supone que existe una familia de modelos y que la planta se puede describir exactamente por un modelo de esa familia. Las técnicas más utilizadas en el contexto de Control Robusto son incertidumbres en la respuestas frecuenciales o incertidumbres en los parámetros de la función de transferencia del proceso. Ambos modelos asumen que la planta es lineal con una respuesta en frecuencia dentro de la banda de incertidumbre en el primer caso y en el segundo caso que la planta es además del mismo orden. En el caso de control predictivo, las incertidumbres afectan a la capacidad de predicción del modelo, la predicción en lugar de ser una trayectoria es una banda alrededor del modelo nominal. Considérese un sistema discretizado en el tiempo cuya dinámica viene definida por:

$$\begin{aligned} y(t+1) &= f(y(t), \dots, y(t-n_y), u(t), \dots, \\ u(t-n_u), z(t), \dots, z(t-n_z), \psi) \end{aligned} \quad (17)$$

donde $y(t) \in \mathbf{Y}$ y $u(t) \in \mathbf{U}$ son vectores de salida y entradas de dimensiones n y m respectivamente, $\psi \in \Psi$ es un vector de parámetros y $z(t) \in \mathbf{Z}$ es un vector de variables exógenas.

Considérese una familia de modelos definida por:

$$\begin{aligned} \hat{y}(t+1) &= \hat{f}(y(t), \dots, y(t-n_{n_a}), \\ u(t), \dots, u(t-n_{n_b}), \theta) \end{aligned} \quad (18)$$

donde $\hat{y}(t+1)$ es la predicción de la salida para el instante $t+1$ generada por los modelos de la familia; \hat{f} es una función vector, usualmente una simplificación de f ; n_{n_a} y n_{n_b} son el número de salidas pasadas, entradas pasadas consideradas por el modelo; y $\theta \in \Theta$ es un vector de incertidumbres sobre el modelo. Nótese que las variables $z(t)$ que afectan a la respuesta del proceso no se consideran por el modelo.

La dinámica del proceso (17) está completamente descrita por la familia de modelos (18) si para cada posible $y(t), \dots, y(t-n_y) \in \mathbf{Y}$, $u(t), \dots, u(t-n_u) \in \mathbf{U}$, $z(t), \dots, z(t-n_z) \in \mathbf{Z}$ and $\psi \in \Psi$ existe un valor del vector de parámetros $\theta_i \in \Theta$ tal que $y(t+1) = \hat{y}(t+1)$. La forma de definir los parámetros de incertidumbres θ y su dominio Θ depende de la estructura de f y \hat{f} y el grado de certeza sobre el modelo. Usualmente se consideran que los parámetros del modelo están definidos dentro de una banda de incertidumbre alrededor de los parámetros nominales. Lo más práctico es considerar todos los errores *globalizados* (Camacho y Bordons, 2004) de forma que la planta se describe por la siguiente familia de modelos:

$$\begin{aligned} \hat{y}(t+1) &= \hat{f}(y(t), \dots, y(t-n_{n_a}), \\ u(t), \dots, u(t-n_{n_b})) + \theta(t) \end{aligned}$$

con $\dim(\theta(t))=n$. Aunque esta descripción es más conservadora, es más fácil de obtener y mucho más intuitiva y refleja lo buena que es la predicción a un paso del modelo. Otra ventaja a destacar de esta descripción es que la ecuación de predicción es afín en el vector de incertidumbre pudiéndose escribir como:

$$\mathbf{y} = \mathbf{G}_u \mathbf{u} + \mathbf{G}_\theta \theta + \mathbf{G}_x x(t)$$

Nótese que al utilizar esta descripción no se asume que el proceso sea lineal, aunque el modelo lo sea, como es el caso de las incertidumbres paramétricas o frecuenciales. Esta descripción asume únicamente que el modelo puede predecir con toda certeza cuál va a ser el valor de las variables de salida del proceso en el próximo instante de muestreo con un grado de tolerancia determinado. Se podría argumentar que estas incertidumbres son en realidad perturbaciones no medibles por la forma que afectan al modelo. Pero aquí la hipótesis que se hace es que las incertidumbres están acotadas; de hecho, $\theta(t)$ podrían ser funciones de las entradas y salidas pasadas (estado).

La forma normal de operar al considerar incertidumbres estocásticas es minimizar la función objetivo J para la situación más esperada. En el caso de control robusto, la estrategia más extendida es considerar la peor de las situaciones posibles, que en el contexto de control predictivo equivale a minimizar la función objetivo en el peor de los casos. Es decir,

$$\min_{\mathbf{u} \in \mathbf{U}} \max_{\theta \in \Theta} J(\mathbf{u}, \theta)$$

Si se utilizan normas ∞ o normas 1, el problema puede expresarse como (ver (Camacho y Bordons, 2004), capítulo 8) como un problema LP. En cualquier caso el problema LP resultante es mucho más complejo que el caso nominal porque el número de restricciones aumenta considerablemente.

Cuando se utiliza una función cuadrática, la función objetivo es convexa en θ , lo que implica que el máximo de J se alcanza en uno de los vértices del Θ (Bazarra y Shetty, 1979). La función es también convexa en \mathbf{u} lo que implica que si se alcanza un óptimo, éste es global.

5.1 Robustez y restricciones

Una forma de hacer un controlador predictivo robusto es imponer que las condiciones de estabilidad se verifiquen para cualquier realización de las restricciones. Las condiciones de estabilidad (Mayne *et al.*, 2000) en control predictivo requieren en general que la función objetivo contenga un coste terminal y que el estado al final del

horizonte de control este incluido en una región terminal.

El control predictivo robusto consiste en encontrar los movimientos de control futuro minimizando la función objetivo y forzando al estado terminal alcanzar la región terminal para todos los valores posibles de las incertidumbres; esto es,

$$\begin{aligned} & \min_{\mathbf{u}} J(x(t), \mathbf{u}) \\ \text{s. a. } & \left. \begin{aligned} \mathbf{R}\mathbf{u} \leq \mathbf{r} + \mathbf{V}x(t) \\ x(t+N) \in \Omega_T \end{aligned} \right\} \forall \theta \in \Theta \end{aligned} \quad (19)$$

donde el conjunto terminal Ω_T se define usualmente como un poliedro $\Omega_T = \{x : \mathbf{R}_T x \leq \mathbf{r}_T\}$. Las restricciones $\mathbf{R}\mathbf{u} \leq \mathbf{r} + \mathbf{V}x(t)$ corresponden a las restricciones de operación. El vector \mathbf{r} depende de forma afín de las incertidumbres θ ; i.e., $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}_\theta \theta$. El vector de predicciones se puede escribir como:

$$\mathbf{x} = \mathbf{G}_u \mathbf{u} + \mathbf{G}_\theta \theta + \mathbf{f}_x x(t) \quad (20)$$

Tomando las filas correspondientes a $x(t+N)$ y sustituyendo en las restricciones de región terminal,

$$\mathbf{R}_T(\mathbf{g}_{u_N} \mathbf{u} + \mathbf{g}_{\theta_N} \theta + \mathbf{f}_{x_N} x(t)) \leq \mathbf{r}_T \quad (21)$$

donde \mathbf{g}_{u_N} , \mathbf{g}_{θ_N} , and \mathbf{f}_{x_N} son las últimas n filas de \mathbf{G}_u , \mathbf{G}_θ , and \mathbf{F}_x respectivamente, con $n = \dim(x)$. Las parte izquierda de las desigualdades (21) dependen de forma afín del vector de incertidumbre θ . El problema (19) resulta un problema QP o LP (dependiendo de la función objetivo) con un número infinito de restricciones. Como las restricciones son afines, si las restricciones se cumplen en los vértices del poliedro se cumplen para todos los puntos en su interior, por lo que el número infinito de restricciones se transforma en un número finito (aunque muy grande). El problema se puede expresar como,

$$\min_{\mathbf{u}} J(x(t), \mathbf{u}) \quad (22)$$

$$\text{s.a.} \quad (23)$$

$$\left. \begin{aligned} \mathbf{R}\mathbf{u} \leq \mathbf{r}_0 + \mathbf{R}_\theta \theta_i + \mathbf{V}x(t) \\ \mathbf{R}_T(\mathbf{g}_{u_N} \mathbf{u} + \mathbf{g}_{\theta_N} \theta_i + \mathbf{f}_{x_N} x(t)) \leq \mathbf{r}_T \end{aligned} \right\} \forall \theta_i \in \varepsilon$$

donde ε es el conjunto finito de puntos extremos (vértices) de Θ .

5.2 Predicciones en Bucle Cerrado

Al resolver el problema min-max la solución tiene que satisfacer las restricciones para todos los valores posibles de las incertidumbres. En muchos casos estas condiciones plantean problemas de no-factibilidad porque el tamaño de la banda de las predicciones crece de tal forma que es imposible satisfacer las restricciones. Esta forma

de operar no tiene en cuenta hasta sus últimas consecuencias el concepto de horizonte deslizante. La secuencia de control $u(t)$, $u(t+1) \dots u(t+N-1)$ se calcula para todo el horizonte sin ninguna información sobre las incertidumbres. Sin embargo en la realidad cuando haya que calcular $u(t+j)$ el controlador ya tiene información sobre las incertidumbres hasta ese instante de tiempo.

El problema es muy difícil de resolver, una forma que permite una solución en ciertos casos es utilizando Programación Dinámica. Considérese el sistema descrito por: $x(t+1) = f(x(t), u(t), \theta(t))$ y el problema:

$$\min_{u(t)} [\max_{\theta(t)} [\min_{u(t+1)} \dots [\max_{\theta(t+N-1)} J(\cdot)]] \dots] \quad (24)$$

La idea fundamental de la Programación Dinámica es resolver el problema (24) desde el corchete más interior hacia el exterior. Es decir desde el problema (J_{t+N-1}) definido por:

$$\min_{u(t+N-1)} \bar{J}_{t+N-1}(x(t+N-1), u(t+N-1)) \quad (25)$$

con $\bar{J}_{t+N-1}(x(t+N-1), u(t+N-1)) = \max_{\theta(t+N-1)} L(x(t+N-1), u(t+N-1)) + F(x(t+N))$.

Note que $F(x(t+N))$ mide el mérito de la última posición y de esta forma se evita entrar en un bucle infinito. Supóngase que se puede resolver el problema (25) explícitamente, i.e., determinando $J_{t+N-1}^*(x(t+N-1))$ como una función de $x(t+N-1)$. Esto es, el coste de ir desde $x(t+N-1)$ al final. A partir de este coste se podría obtener $J_{t+N-2}^*(x(t+N-2))$ y se puede llegar a:

$$J_t^*(x(t)) = \min_{u(t)} \bar{J}_t(x(t), u(t)) \quad (26)$$

con

$$\begin{aligned} \bar{J}_t(x(t), u(t)) &= \max_{\theta(t)} L(x(t), u(t)) \\ &+ J_{t+1}^*(f(x(t), u(t), \theta(t))) \end{aligned}$$

El control predictivo min-max en bucle cerrado consiste en calcular los movimientos de control $u^*(t)$ minimizando $\bar{J}_t(x(t), u(t))$. El factor clave de la Programación Dinámica es encontrar la función $J_{t+j}^*(x(t+j))$.

Se ha podido demostrar (Bemporad *et al.*, 2003) que si el sistema es lineal $x(t+1) = \mathbf{A}(\omega(t))x(t) + \mathbf{B}(\omega(t))u(t) + E v(t)$, con el vector de incertidumbres $\theta(t)^T = [\omega(t)^T v(t)^T]$, y la función de coste de un paso $L(x(t+j), u(t+j)) = \|\mathbf{Q}x(t+j)\|_p + \|\mathbf{R}u(t+j)\|_p$ con el coste terminal $J_{t+N}^*(x(t+N)) = \|\mathbf{P}x(t+N)\|_p$, la solución es afín a trozos y se puede encontrar utilizando programación multi-paramétrica.

Otra estrategia utilizada es aproximar la función $J_t^*(x(t+j))$ en una rejilla de puntos como se sugiere en (Lee y Yu, 1997). La idea es calcular

$J_{t+N-1}^*(x(t+N-1))$ en todos los puntos de la rejilla en el primer paso para en un segundo paso calcular la función $J_{t+N-2}^*(x(t+N-2))$ utilizando una interpolación de $J_{t+N-1}^*(x(t+N-1))$ cuando $x(t+N-1) = f(x(t+N-2), u(t+N-2), \theta(t+N-2))$ no coincide con uno de los puntos de la rejilla.

Una forma de reducir la banda de incertidumbre de la predicción es utilizar una realimentación lineal (Bemporad, 1998), (Chisci *et al.*, 2001) del vector de estado antes de aplicar el control predictivo. Es decir, hacer:

$$u(t+k) = -Kx(t+k | t) + v(t+k) \quad (27)$$

donde K es una realimentación que estabiliza al sistema y la variable auxiliar $v(t)$ es la referencia para el bucle interno y la variable manipulable para el controlador predictivo. De esta forma se consigue disminuir el tamaño de la banda de incertidumbre de la predicción.

6. CONTROL PREDICTIVO Y SISTEMAS HÍBRIDOS

En la mayor parte de los procesos no sólo hay variables continuas sino que también variables que toman valores discretos. Durante mucho tiempo, la forma de modelar y controlar uno y otro proceso se realizó como algo totalmente distinto. Por una parte grafos de transición de estados, redes de Petri etc. para describir a los procesos con variables discretas y por el otro ecuaciones diferenciales, funciones de transferencia etc. para describir procesos continuos. Desde el comienzo de los 90 ha habido un gran interés por el estudio de procesos que tienen parte continua y discreta: los sistemas híbridos. El control predictivo puede ser también aplicado a sistemas híbridos (ver capítulo 10 de (Camacho y Bordons, 2004)). Los sistemas híbridos se han modelado de distintas formas: bien como un grafo de transición de estados (con dinámica continua dentro de cada estado) o bien como un conjunto de ecuaciones diferenciales, o en diferencias, con variables discretas. Esta última es la aproximación más usual en el campo del control predictivo.

6.1 Sistemas Dinámicos con variables lógicas

Este tipo de sistemas ("Mixed Logical Dynamical Systems", MLD) propuesto en (Bemporad y Morari, 1999) se describe por:

$$\begin{aligned} x(t+1) &= Ax(t) + B_1U(t) + B_2\delta(t) + B_3z(t) \\ y(t) &= Cx(t) + D_1u(t) + D_2\delta(t) + D_3z(t) \quad (28) \\ E_1x(t) + E_2u(t) + E_3\delta(t) + E_4z(t) &\leq g \end{aligned}$$

donde $x(t) = [x_r^T(t) \ x_b^T(t)]$ siendo $x_r(t) \in \mathbb{R}^n$ la parte continua del vector de estado y $x_b(t) \in \{0, 1\}^{n_b}$ la parte discreta. La salida del sistema tiene una estructura similar $y(t) = [y_r^T(t) \ y_b^T(t)]$ donde $y_r(t) \in \mathbb{R}^m$ es la parte continua y la parte discreta viene dada por $y_b(t) \in \{0, 1\}^{m_b}$. El vector de entrada $u(t) = [u_r^T(t) \ u_b^T(t)]$ está compuesto de una parte continua $u_r(t) \in \mathbb{R}^l$ y otra discreta $u_b(t) \in \{0, 1\}^{l_b}$. La descripción suele requerir variables auxiliares continuas $z(t) \in \mathbb{R}^r$ y discretas $\delta(t) \in \{0, 1\}^{r_b}$.

La idea clave para obtener esta descripción a partir de las expresiones lógicas de la parte discreta de un sistema híbrido es que las expresiones lógicas se pueden expresar fácilmente (ver (Camacho y Bordons, 2004), capítulo 10) en restricciones. Por ejemplo supongamos dos variables lógicas L_1 y L_2 que pueden tomar los valores lógicos 0 ó 1. La expresión lógica " L_1 OR L_2 " se transforma en la restricción $L_1 + L_2 \geq 1$. La expresión lógica " L_1 AND L_2 " se transforma en $L_1 + L_2 \geq 2$ con ambas variables pertenecientes a $[0, 1]$.

6.2 Control Predictivo de sistemas MLD

El problema de control predictivo se puede formular como:

$$\begin{aligned} \mathbf{u}^* = \arg \min_{\mathbf{u}} & \|\mathbf{x} - \mathbf{r}_x\|_{\mathbf{Q}_x}^p + \|\mathbf{u} - \mathbf{r}_u\|_{\mathbf{Q}_u}^p + \\ & \|\delta - \mathbf{r}_\delta\|_{\mathbf{Q}_\delta}^p + \|\mathbf{z} - \mathbf{r}_z\|_{\mathbf{Q}_z}^p \end{aligned} \quad (29)$$

sujeto a las ecuaciones (28), donde $\|x\|_Q^p$ denota $x^T Q x$ cuando $p = 2$ y $Q\|x\|_p$ para $p = 1$ ó $p = \infty$ y \mathbf{Q}_x , \mathbf{Q}_u , \mathbf{Q}_δ , y \mathbf{Q}_z son matrices de peso de dimensiones apropiadas y todas las señales se predicen con la información disponible hasta el instante t en la forma usual del control predictivo. Los vectores \mathbf{x} , \mathbf{u} , δ , \mathbf{z} , \mathbf{r}_x , \mathbf{r}_u , \mathbf{r}_δ y \mathbf{r}_z son los vectores de estados futuros predichos, movimientos de control, variables lógicas auxiliares, variables reales auxiliares, y sus referencias futuras correspondientes.

El control predictivo resultante resulta ser un problema de optimización con un conjunto de restricciones lineales y con variables de decisión reales y enteras. Este tipo de problemas se conoce con el nombre de problemas de programación mixta (real y entera). Son problemas mucho más difíciles de resolver que los problemas LP o QP (ver (Floudas, 1995)).

6.3 Sistemas afines a trozos

Otra forma de modelar sistemas híbridos es mediante sistemas afines a trozos ("piecewise affine systems", PWA). De hecho, se ha probado (Heemels *et al.*, 2001) que un sistema MLD (y muchas otras

descripciones de sistemas híbridos) se pueden describir como sistemas PWA. Los sistemas afines a trozos tienen también la ventaja de permitir aproximar sistemas no lineales con un grado de precisión arbitrario. Un sistema afín a trozos se puede describir por:

$$\begin{aligned} x(t+1) &= A^i x(t) + B^i u(t) + f^i \\ y(t) &= C^i x(t) + g^i \end{aligned} \quad \text{for } \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \in \mathcal{X}_i$$

donde $\{\mathcal{X}_i\}_{i=1}^s$ es una partición poliédrica del espacio de estado y entrada. Cada \mathcal{X}_i viene dado por

$$\mathcal{X}_i = \left\{ \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \mid \mathbf{R}^i \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \leq \mathbf{r}^i \right\}$$

Cada subsistema \mathbb{S}^i está definido por la 7-tupla $(A^i, B^i, C^i, f^i, g^i, \mathbf{R}^i, \mathbf{r}^i)$, $i \in \{1, 2, \dots, s\}$. $A^i \in \mathbb{R}^{n \times n}$, $B^i \in \mathbb{R}^{n \times m}$, and (A^i, B^i) es un par controlable. $C^i \in \mathbb{R}^{r \times n}$ and $\mathbf{R}^i \in \mathbb{R}^{p_i \times (n+m)}$ y f^i, g^i, \mathbf{r}^i son vectores constantes apropiados. La dimensiones de los vectores de estado, de entradas de salida son n , m y r respectivamente. Finalmente, p_i es el número de hiperplanos que definen al poliedro i .

Asumiendo que todo el estado está disponible (un estimador del estado sería necesario en caso contrario), la formulación del control predictivo es:

$$\begin{aligned} \mathbf{u} &= \arg(\min_{\mathbf{u}} J) & (30) \\ \text{s.a. : } J &= \sum_{i=1}^N q_{ii} (y(t+i|t) - w(t+i))^2 + \\ & \sum_{i=0}^{N-1} r_{ii} u(t+i)^2 \\ u_{\min} &\leq u(t+i) \leq u_{\max} \quad i = 0, \dots, N-1 \end{aligned}$$

Considérese el problema de predicción: el subsistema que describe el proceso en el instante t es conocido pero el siguiente subsistema donde el proceso estará en el instante de tiempo siguiente depende de la señal de control aplicada. En general, una secuencia de subsistemas $I = \{I(t) \ I(t+1) \ \dots \ I(t+N)\}$ se pueden "activar" a lo largo de la trayectoria. Sólo el valor inicial $I(t) = I(t)(x(t))$ de esta secuencia es conocido. Nótese que si no hay restricciones, el número posible de secuencias para un horizonte de N es s^{N-1} , número que puede ser muy elevado. El problema de optimización resultante se puede expresar como:

$$\mathbf{u}^* = \arg(\min_{\mathbf{u}, I} J) \quad (31)$$

donde hay que añadir restricciones para considerar las dependencias entre los subsistemas y sus ecuaciones correspondientes, i.e.:

$$\mathbf{R}^{I(t+j)} x(t+j) \leq \mathbf{r}^{I(t+j)}, \quad j = \{1, \dots, N\} \quad (32)$$

Este problema podría ser resuelto encontrando el óptimo para cada posibles secuencia de I , i.e.

$$\mathbf{u}^* = \arg \left(\min_I \left(\min_{\mathbf{u}} \left(J_{\mathbf{R}^{IU} u \leq \mathbf{r}^{IU}} \right) \right) \right) \quad (33)$$

donde $\mathbf{R}^{IU} u \leq \mathbf{r}^{IU}$ son las restricciones debidas a las dependencias entre I y U .

Los problemas MILP o MIQP resultante son mucho más difíciles de resolver en tiempo real que los problemas QP o LP de los controladores predictivos nominales.

7. MÉTODOS RÁPIDOS PARA IMPLEMENTAR CONTROLADORES PREDICTIVOS

Una de las desventajas de los controladores predictivos es que requieren, en algunos casos, tiempos de cálculo demasiado grandes para aplicaciones de tiempo real. Esto ocurre por ejemplo en el caso de presencia de restricciones, cambios en los parámetros del proceso, control predictivo robusto o control predictivo no lineal. Han aparecido algoritmos en la literatura que permiten la implementación fácil de controladores predictivos. Tal es el caso de la estructura propuesta (Bordons y Camacho, 1998) para sistemas que puedan ser descritos por el método de la curva de reacción o la utilización de redes neuronales para aproximar controladores predictivos no lineales (Gómez Ortega y Camacho, 1994).

Recientemente se ha demostrado (Bemporad *et al.*, 2002) que un controlador predictivo se puede expresar como un programa multiparamétrico cuya solución resulta en un controlador que es afín a trozos en el espacio de estado. La idea es simple y fue sugerida por primera vez en (Zafiriou, 1990): el óptimo de un problema QP se alcanza en un conjunto de restricciones activas (el conjunto puede ser vacío) y para todos los puntos del espacio que tengan el mismo conjunto de restricciones activas, la solución es afín.

7.1 Ejemplo Ilustrativo: el doble Integrator

Considere el doble integrador descrito por la ecuación siguiente:

$$Y(s) = \frac{1}{s^2} U(s)$$

Si el proceso se muestrea con un tiempo de muestreo de una unidad y suponiendo un muestreador y mantenedor entre ambos integradores, la descripción discreta del sistema es:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad C = [1 \ 0]$$

Considérese la función objetivo:

$$J = \sum_{j=1}^{\infty} (y(t+j)^2 + \lambda u(t+j-1)^2)$$

Nótese que utilizando una función objetivo con horizonte infinito garantizamos la estabilidad del caso nominal cuando no hay restricciones. Cuando no hay restricciones, este problema no puede ser resuelto ya que el número de variables de decisión para el algoritmo numérico es infinito. Si suponemos que las restricciones van a afectar a la primera parte de la respuesta, y descomponiendo la función objetivo en:

$$\begin{aligned} J &= \lambda u(t)^2 \\ &+ \sum_{j=1}^{N-1} (x(t+j)^T Q x(t+j) + u(t+j)^T R u(t+j)) \\ &+ \sum_{j=N}^{\infty} (x(t+j)^T Q x(t+j) + u(t+j)^T R u(t+j)) \end{aligned} \quad (34)$$

con $R = \lambda$ y $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ para el ejemplo. Si la última parte de la función de costo (34) no es afectada por las restricciones, se puede encontrar su valor óptimo resolviendo la ecuación de Riccati correspondiente. Para $N = 2$ y $\lambda = 0.1$, la ley de control resultante (obtenida al resolver la ecuación de Riccati) es $u(t) = Kx(t)$ con $K = [-0.8166 \quad -1.7499]$. El valor de la función objetivo cuando se aplica esta ley de control (para el caso nominal sin ruido ni restricciones) es una función del estado que viene dada por $J(x) = x^T P x$, donde P se obtiene al resolver la ecuación de Riccati (ver (Camacho y Bordons, 2004) que en este caso resulta:

$$P = \begin{bmatrix} 2.1429 & 1.2246 \\ 1.2246 & 1.3996 \end{bmatrix}$$

Si la ley de control óptima se aplica para $k > N$, la función de coste resulta:

$$\begin{aligned} J &= \sum_{j=1}^N (x(t+j)^T Q_j x(t+j) \\ &+ u(t+j-1)^T R u(t+j-1)) \end{aligned} \quad (35)$$

con $Q_j = Q$ for $j < N$ y $Q_N = P$. Si los movimientos de control están restringidos por $-1 \leq u(t) \leq 1$ y $-1 \leq u(t+1) \leq 1$, hay nueve regiones descritas, con sus controladores correspondientes, por expresiones del tipo:

Region 1:

$$\text{Si } \begin{bmatrix} -0.8166 & -1.7499 \\ 0.6124 & 0.4957 \\ 0.8166 & 1.7499 \\ -0.6124 & -0.4957 \end{bmatrix} x(t) \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{ entonces}$$

$$u(t) = [-0.8166 \quad -1.7499]x(t)$$

Region 2:

$$\text{Si } \begin{bmatrix} 2.4491 & 5.2482 \\ -0.8166 & -2.5665 \\ 0.8166 & 2.5665 \end{bmatrix} x(t) \leq \begin{bmatrix} -2.9991 \\ 2.7499 \\ -0.7499 \end{bmatrix}$$

entonces $u(t) = 1$

Region 3:

:

Region 9:

$$\text{Si } \begin{bmatrix} -6.7349 & -18.7181 \\ -2.4491 & -7.6974 \end{bmatrix} x(t) \leq \begin{bmatrix} -17.4314 \\ -8.2474 \end{bmatrix}$$

entonces $u(t) = -1$

La ley de control resultante no está restringida para los puntos de la región 1 y coincide, como era de esperar, con la solución de la ecuación de Riccati. Las regiones pueden verse en la figura 9, que también muestra la evolución del estado (cuando se aplica la ley de control afín descrita previamente) con un estado inicial dado por $x(t)^T = [-5 \quad 5]$.

Nótese que el controlador minimiza la expresión (35) con $N = 2$, que minimizará la función con horizonte infinito cuando la señal de control no está afectada por restricciones después para horizontes mayores que 2. Nótese que la parte a implementar del controlador en línea consiste en: a) leer o estimar el estado, b) decidir en que región está el sistema y c) aplicar la ley de control correspondiente. En este ejemplo el número de operaciones para determinar la región es de 48 multiplicaciones, 24 sumas y 24 comparaciones. El cálculo de la señal de control requiere como máximo dos multiplicaciones y una suma. Es decir un número muy reducido de operaciones. Nótese que el mayor tiempo de cálculo corresponde a la determinación de la región, este tiempo depende del número de regiones y dimensión del vector de estado y puede ser considerable. En (Tøndel y Johansen, 2002) se sugiere un método para organizar las restricciones de forma que se minimice el número de operaciones necesarias para determinar la región.

7.2 Implementaciones rápidas de MPC robustos

Cuando hay incertidumbres, es necesario resolver un problema min-max y satisfacer un conjunto de restricciones para cualquier realización posible de las incertidumbres. Se ha demostrado que la programación multiparamétrica puede ser extendida a este problema. En el caso de normas ∞ (o norma 1) la solución es afín a trozos (Bemporad *et al.*, 2001). Este hecho puede deducirse fácilmente porque en estos casos el problema min-max se puede expresar como un problema LP (ver capítulo

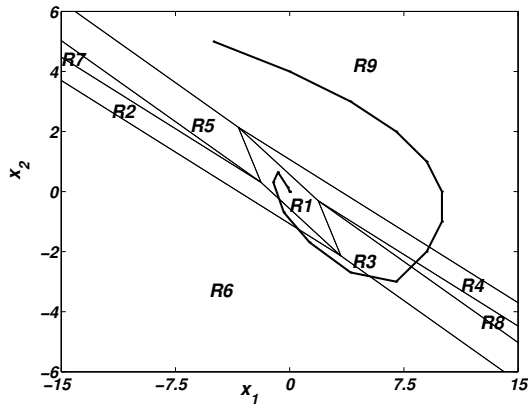


Fig. 9. Regiones de control afines para el doble integrador.

8 de (Camacho y Bordons, 2004)) y que para estos problemas la solución es lineal a trozos. Para el caso de funciones objetivos cuadráticas, se ha demostrado (Ramírez y Camacho, 2001) que el min-max MPC es también una ley de control afín a trozos.

Estos resultados han hecho cambiar la concepción de que estos controladores sólo pueden ser aplicados a procesos lentos y requiriendo gran capacidad de cálculo. Las realizaciones rápidas permiten aplicar controladores predictivos con restricciones o controladores predictivos min-max a procesos rápidos sin requerir grandes procesadores.

8. MPC NO LINEAL

En general, los procesos industriales son no lineales, pero aún así, la mayoría de aplicaciones del MPC están basadas en el uso de modelos lineales. Existen dos importantes razones para ello: por un lado, la identificación de un modelo lineal a partir de datos de proceso es relativamente sencilla y por otro, los modelos lineales proporcionan buenos resultados cuando la planta opera en las cercanías del punto de operación. Además, el uso de un modelo lineal junto con una función de coste cuadrática da lugar a un problema convexo (Programación Cuadrática), cuya solución está bien estudiada y existen numerosos productos comerciales disponibles.

Sin embargo, la respuesta dinámica de los controladores lineales que resultan es inaceptable cuando se aplican a procesos que son no lineales con distinto grado de severidad. Aunque el número de aplicaciones de Nonlinear Model Predictive Control (NMPC) es aún limitada (ver (Badgwell y Qin, 2001), (Qin y Badgwell, 1998)), su potencial es realmente grande y acabará abriéndose camino en aquellas áreas donde las no-linealidades son severas y la demanda de mercado exige frecuentes cambios en el régimen de operación.

En esta sección se muestran algunos de los desarrollos más recientes y las nuevas tendencias en los aspectos tanto teóricos como prácticos de NMPC. Ambos aspectos son importantes ya que hay temas teóricos tales como modelado o estabilidad que poseen tanta importancia como aquéllos de índole práctica como identificación o complejidad computacional. Se puede considerar que los temas más relevantes en este campo son el desarrollo de modelos que sean capaces de capturar la dinámica no lineal del proceso y la obtención de una solución factible en tiempo real al problema de la minimización.

El desarrollo de modelos no lineales adecuados puede ser una tarea muy complicada y en general no existe una forma claramente apropiada para representar este tipo de sistemas. Gran parte del éxito del MPC se ha debido a la relativa facilidad con que se pueden obtener modelos de convolución (impulso o escalón) o funciones de transferencia de bajo orden. Los modelos no lineales son más difíciles de construir, bien por correlaciones de datos de entrada-salida o bien por el uso de primeros principios a partir de ecuaciones de balance de masa y energía.

Hay tres tipos de modelos que se usan en las formulaciones de NMPC:

- Modelos empíricos, que se obtienen de datos reales y pueden tomar la forma de modelos de entrada-salida (tales como NARX, modelos de Volterra, Hammerstein o bilineales y redes neuronales) o modelos no lineales en el espacio de estados.
- Modelos fundamentales, que provienen de ecuaciones de balance, también llamados de primeros principios. Las ecuaciones se obtienen por el conocimiento del proceso, aplicando ecuaciones de balance de masa, energía momento. En este caso, la predicción se hace como una simulación de las ecuaciones no lineales que describen la dinámica del proceso. Para procesos industriales complejos, este tipo de modelos es difícil y costoso de construir, ya que necesita conocimiento muy experto.
- Modelos de caja gris, que se desarrollan combinando los enfoques empírico y fundamental, haciendo uso de las ventajas de cada tipo de modelo. En este enfoque híbrido, la información de los primeros principios se ve enriquecida con datos empíricos.

El otro asunto crucial en NMPC es la solución del problema. Requiere la consideración (y al menos la resolución parcial) de un problema no lineal no convexo (NLP) que da lugar a una serie de dificultades computacionales relacionadas con el coste y la fiabilidad de la resolución del problema NLP en línea.

Normalmente el problema se resuelve haciendo uso de la Programación Cuadrática Secuencial (Sequential Quadratic Programming, SQP), que son extensiones de métodos de tipo Newton para lograr la convergencia a las condiciones de Karush-Kuhn-Tucker (KKT) del problema de optimización. El método debe garantizar convergencia rápida y debe tratar problemas de mal acondicionamiento y no linealidades extremas. SQP es una técnica iterativa en la cual la solución en cada paso se obtiene mediante la resolución de una aproximación al problema no lineal en la que el objetivo es sustituido por una aproximación cuadrática y las restricciones no lineales por aproximaciones que sí los son.

La solución exacta del problema de optimización en cada instante de muestreo es una tarea ardua. Por ello, en los últimos años ha aparecido una serie de formulaciones que pretenden evitar los problemas asociados a la optimización no convexa. Estas formulaciones deben abordar la estabilidad y la factibilidad de la solución que debe calcularse durante el periodo de muestreo. Una revisión de las técnicas existentes se puede encontrar en (Camacho y Bordons, 2004), donde formulaciones como MPC lineal extendido, modelos locales, NMPC subóptimo, uso de horizontes cortos, descomposición de la secuencia de control, linealización por realimentación, MPC basado en modelos de Volterra o NMPC con redes neuronales son descritas en detalle.

9. ESTABILIDAD

La solución eficiente del problema del control óptimo es importante para cualquier aplicación a procesos reales, pero la estabilidad del bucle cerrado es también de crucial importancia. Para sistemas lineales sin restricciones, la estabilidad se puede analizar con herramientas convencionales de teoría de sistemas lineales, pero si aparecen restricciones o el sistema es no lineal, la ley de control se convierte en no lineal y deben usarse otras herramientas. Se trata de un campo donde han aparecido resultados significativos recientemente.

Incluso en el caso de que el algoritmo de optimización encuentre una solución, esto no garantiza la estabilidad del bucle cerrado (incluso sin incertidumbres). El uso de penalización o restricción terminal, funciones de Lyapunov o conjuntos invariantes ha dado lugar a una amplia familia de técnicas que son capaces de garantizar la estabilidad del sistema controlado.

Este problema ha sido abordado desde distintos puntos de vista, dando lugar a una serie de contribuciones que siempre analizan el problema del regulador (llevar el estado al reposo) en el

marco del espacio de estados. Las formulaciones de estabilidad garantizada están recogidas en la publicación de Mayne *et al.* (Mayne *et al.*, 2000). En esta referencia, los autores presentan las condiciones suficientes para el diseño de un controlador predictivo estabilizante con restricciones.

Los ingredientes clave de un MPC estabilizante son el conjunto terminal y el coste terminal. El estado terminal denota el estado predicho del sistema al final del horizonte de predicción. Se fuerza a que este estado terminal alcance un conjunto terminal que contenga el estado en régimen estacionario y tiene un coste asociado que se denomina coste terminal, que se añade como un nuevo término a la función de coste.

Se supone que el sistema es localmente estabilizable mediante una ley de control de la forma $u = h(x)$, que debe satisfacer las siguientes condiciones:

- Existe una región Ω tal que para todo $x(t) \in \Omega$, entonces $h(x(t)) \in U$ (conjunto de acciones de control admisibles) y el estado del sistema en el siguiente tiempo de muestreo $x(t+1) \in \Omega$.
- Para todo $x(t) \in \Omega$, existe una función de Lyapunov $V(x)$ tal que

$$V(x(t)) - V(x(t+1)) \geq x(t)^T R x(t) + h(x(t))^T S h(x(t))$$

Si se verifican estas condiciones, entonces al considerar Ω como conjunto terminal y $V(x)$ como coste terminal, el controlador MPC (con iguales valores de horizonte de control y de predicción) estabiliza asintóticamente todos los estados iniciales que sean factible. Por tanto, si el estado inicial es tal que el problema de optimización tiene solución, entonces el sistema es llevado al régimen permanente asintóticamente y satisface las restricciones durante su evolución.

La condición que se impone sobre Ω asegura la satisfacción de las restricciones y la segunda condición asegura que el coste óptimo es una función de Lyapunov.

Como el problema de optimización que se debe resolver en cada periodo de muestreo puede no ser convexo, la solución óptima puede no ser única y puede ser muy difícil de obtener. Así, se han propuesto diferentes enfoques para relajar este hecho. La principal contribución en este tema es la prueba de estabilidad asintótica en el caso de soluciones sub-óptimas (Scokaert *et al.*, 1999): basta considerar cualquier solución factible con coste asociado estrictamente menor que el del tiempo de muestreo anterior. En efecto, cualquier solución factible asegura factibilidad, y por tanto cumplimiento de las restricciones, y el coste estrictamente decreciente garantiza estabilidad asintótica. Es

conveniente remarcar que la sub-optimalidad no es deseable, ya que implica una pérdida de prestaciones.

Otra técnica para reducir la carga computacional del problema de optimización es la eliminación de la restricción terminal. Es especialmente interesante cuando no existen restricciones en el estado. En este caso, la carga computacional no tiene por qué ser incrementada para introducir restricciones terminales del estado por razones de estabilidad. Este tema se ha analizado en (Hu y Linnemann, 2002), (Jadbabaie *et al.*, 2001) y (Limon *et al.*, 2003).

El horizonte de predicción es un parámetro de diseño del MPC de gran importancia. Si se aumenta su valor el dominio de atracción del controlador se hace mayor y se mejora el comportamiento. Sin embargo, el número de variables de decisión también es mayor y por tanto se incrementa la complejidad del problema de optimización. La condición necesaria que se debe considerar para elegir el horizonte de predicción es la factibilidad del estado inicial. Así, el horizonte se puede reducir aumentando la región terminal.

Si el análisis de estabilidad es una tarea compleja, el de robustez (es decir, cuando aparecen errores de modelado) es lógicamente peor. Los resultados de estabilidad mostrados previamente son válidos sólo en el caso de modelos perfectos, lo que no es cierto en la práctica. Este tema puede ser considerado como abierto y con resultados sólo preliminares. Se han propuesto formulaciones en la forma de problema min-max o H_∞ -NMPC aunque los requisitos de cálculo son prohibitivos.

10. TEMAS ABIERTOS

El control predictivo se considera un controlador maduro en el caso de procesos lineales y es utilizado con bastante éxito en la industria. En cualquier caso, no se puede considerar al control predictivo como una disciplina ya cerrada a la investigación, sino que, al contrario, es una rama del control de gran efervescencia. Esto se puede constatar por el hecho de que en todos los congresos de control de estos últimos años siempre hay sesiones específicas de control predictivo y raro es el número de las revistas de control donde no aparezca algún artículo sobre este tema.

Los temas todavía no del todo resueltos y que son objeto de investigación se pueden a *grosso modo* clasificar en problemas de implementación y problemas de análisis y diseño. El control predictivo es muy difícil de implementar en tiempo real para procesos no lineales, procesos híbridos o procesos muy rápidos. En el caso de sistemas no lineales e híbridos han aparecido soluciones

para casos particulares y normalmente de pequeña dimensión, pero no existen soluciones generales. Se ha demostrado recientemente que la estructura de control predictivo para procesos lineales es afín a trozos y por lo tanto el controlador puede calcularse previamente. Esto ha permitido que el control predictivo pueda ser aplicado a procesos más rápidos, pero, desafortunadamente, la técnica sólo es aplicable a problemas de pequeña dimensión (incluyendo horizontes de control pequeños).

El otro gran problema todavía no resuelto es el problema del análisis de estabilidad y robustez de estos controladores. Es decir, aún en el caso de que los controladores sean implementables, analizar cómo se puede asegurar su estabilidad en el caso nominal o en el caso de que el modelo no sea exacto. Para el caso de restricciones, el análisis de estabilidad parecía ser un problema demasiado complicado de resolver. Aún en el caso de que el optimizador fuera capaz de encontrar una solución, no está garantizada la estabilidad del bucle cerrado. La utilización de penalizaciones terminales y/o restricciones, funciones de Lyapunov, o conjuntos invariantes han dado lugar a una familia de técnicas que garantizan la estabilidad del sistema. Este problema ha sido atacado de distintos puntos de vista y han aparecido numerosas contribuciones en años recientes, siempre analizando el problema del regulador (llevar estado al reposo) y normalmente en el espacio de estados. Se han obtenido resultados que utilizan técnicas de control robusto en el contexto de controladores predictivos. La idea básica es tener en cuenta las incertidumbres sobre el proceso de una manera explícita y diseñar el controlador predictivo para optimizar la función objetivo ante la peor situación posible de las incertidumbres. En cualquier caso, estos resultados exigen el cómputo de regiones invariantes que salvo en el caso de sistemas lineales son difícilmente calculables.

Estos resultados prometedores permiten pensar que el control predictivo experimentará una mayor diseminación tanto en el mundo académico como en la comunidad industrial en los próximos años.

11. RECONOCIMIENTOS

Los autores quieren reconocer el apoyo del Ministerio de Ciencia y Tecnología a través de los proyectos DPI2001-2380-C02-01 y DPI200-4375-C03-01 y los comentarios de los componentes del Grupo de Control Predictivo de la Universidad de Sevilla.

REFERENCIAS

- Alvarez, T. y C. Prada (1997). Handling Infeasibility in Predictive Control. *Computers and Chemical Engineering* **21**, 577–582.
- Arruda, L.V.R., R. Lüders, W.C. Amaral y F.A.C. Gomide (1994). An Object-oriented Environment for Control Systems in Oil Industry. In: *Proceedings of the 3rd Conference on Control Applications, Glasgow, UK*. pp. 1353–1358.
- Aström, K.J. y B. Wittenmark (1984). *Computer Controlled Systems. Theory and Design*. Prentice-Hall. Englewood Cliffs, NJ.
- Badgwell, T.A. y S.J. Qin (2001). *Nonlinear Predictive Control*. Chap. Review of Nonlinear Model Predictive Control Applications. IEE Control Engineering series.
- Bazaraa, M.S. y C.M. Shetty (1979). *Nonlinear Programming*. Wiley.
- Bemporad, A. (1998). Reducing Conservativeness in Predictive Control of Constrained Systems with Disturbances. In: *IEEE Conference on Decision and Control*.
- Bemporad, A., F. Borrelli y M. Morari (2001). Robust Model Predictive Control: Piecewise Linear Explicit Solution. In: *Proc. European Control Conference, ECC'01*.
- Bemporad, A., F. Borrelli y M. Morari (2003). Min-max Control of Constrained Uncertain Discrete-time Linear Systems. *IEEE Tran. on Automatic Control* **48**(9), 1600 – 1606.
- Bemporad, A., M. Morari, V. Dua y E.N. Pistikopoulos (2002). The Explicit Linear Quadratic Regulator for Constrained Systems. *Automatica* **38**(1), 3–20.
- Bemporad, A. y M. Morari (1999). Control of Systems Integrating Logic, Dynamics and Constraints. *Automatica* **35**(3), 407–427.
- Bordons, C. y E.F. Camacho (1998). Generalized Predictive Controller for a Wide Class of Industrial Process. *IEEE Transaction on Control Systems Technology* **6**(3), 372–387.
- Bordons, C. y J.R. Cueli (2004). Predictive Controller with Estimation of Measurable Disturbances. Application to an Olive Oil Mill. *Journal of Process Control* **14**(3), 305–315.
- Camacho, E.F. y C. Bordons (2004). *Model Predictive Control, Second Edition*. Springer-Verlag.
- Chia, T.L. y C.B. Brosilow (1991). Modular Multivariable Control of a Fractionator. *Hydrocarbon Processing* pp. 61–66.
- Chisci, L., J. A. Rossiter y G. Zappa (2001). Systems with persistent disturbances: Predictive control with restricted constraints. *Automatica* **37**, 1019–1028.
- Clarke, D.W. (1988). Application of Generalized Predictive Control to Industrial Processes. *IEEE Control Systems Magazine* **122**, 49–55.
- Clarke, D.W., C. Mohtadi y P.S. Tuffs (1987). Generalized Predictive Control. Part I. The Basic Algorithm. *Automatica* **23**(2), 137–148.
- Clarke, D.W. y P.J. Gawthrop (1979). Self-tuning Control. *Proceedings IEEE* **123**, 633–640.
- Clarke, D.W. y R. Scattolini (1991). Constrained Receding-horizon Predictive Control. *Proceedings IEE* **138**(4), 347–354.
- Cutler, C.R. y B.C. Ramaker (1980). Dynamic Matrix Control- A Computer Control Algorithm. In: *Automatic Control Conference, San Francisco*.
- Floudas, C.A. (1995). *Non-Linear and Mixed Integer Optimization*. Oxford Academic Press.
- García, C.E., D.M. Prett y M. Morari (1989). Model Predictive Control: Theory and Practice-a Survey. *Automatica* **25**(3), 335–348.
- Gómez Ortega, J. y E.F. Camacho (1994). Neural Network MBPC for Mobile Robots Path Tracking. *Robotics and Computer Integrated Manufacturing Journal* **11**(4), 271–278.
- Gómez Ortega, J. y E.F. Camacho (1996). Mobile Robot Navigation in a Partially Structured Environment using Neural Predictive Control. *Control Engineering Practice* **4**, 1669–1679.
- Greco, C., G. Menga, E. Mosca y G. Zappa (1984). Performance Improvement of Self Tuning Controllers by Multistep Horizons: The MUSMAR Approach. *Automatica* **20**, 681–700.
- Heemels, W.P.M.H., B. De Schutter y A. Bemporad (2001). Equivalence of Hybrid Dynamical Models. *Automatica* **37**(7), 1085–1091.
- Hu, B. y A. Linnemann (2002). Towards Infinite-horizon Optimality in Nonlinear Model Predictive Control. *IEEE Transactions on Automatic Control* **47**(4), 679–682.
- Jadbabaie, A., J. Yu y J. Hauser (2001). Unconstrained Receding-Horizon Control of Nonlinear Systems. *IEEE Transactions on Automatic Control* **46**(5), 776–783.
- Keyser, R.M.C. De y A.R. Van Cuawenberghe (1985). Extended Prediction Self-adaptive Control. In: *IFAC Symposium on Identification and System Parameter Estimation, York, UK*. pp. 1317–1322.
- Lee, J.H. y Z. Yu (1997). Worst-case Formulations of Model Predictive Control for Systems with Bounded Parameters. *Automatica* **33**(5), 763–781.
- Limon, D., T. Alamo y E.F. Camacho (2003). Stable Constrained MPC without Terminal Constraint. In: *Proceedings of the American Control Conference*.
- Linkers, D.A. y M. Mahfonf (1994). *Advances in Model-Based Predictive Control*. Chap. Generalized Predictive Control in Clinical Anaesthesia. Oxford University Press.

- Maciejowski, J.M. (2001). *Predictive Control with Constraints*. Prentice Hall. Harlow.
- Mayne, D.Q., J.B. Rawlings, C.V. Rao y P.O.M. Scokaert (2000). Constrained Model Predictive Control: Stability and Optimality. *Automatica* **36**, 789–814.
- Morari, M. (1994). *Advances in Model-Based Predictive Control*. Chap. Model Predictive Control: Multivariable Control Technique of Choice in the 1990s? Oxford University Press.
- Mosca, E., J.M. Lemos y J. Zhang (1990). Stabilizing I/O Receding Horizon Control. In: *IEEE Conference on Decision and Control*.
- Peterka, V. (1984). Predictor-based Self-tuning Control. *Automatica* **20**(1), 39–50.
- Propoi, A.I. (1963). Use of LP Methods for Synthesizing Sampled-data Automatic Systems. *Automatic Remote Control*.
- Qin, S.J. y T.A. Badgwell (1997). An Overview of Industrial Model Predictive Control Technology. In *Chemical Process Control: Assessment and New Directions for Research*. In: *AIChE Symposium Series 316, 93*. Jeffrey C. Kantor, Carlos E. Garcia and Brice Carnahan Eds. 232–256.
- Qin, S.J. y T.A. Badgwell (1998). An Overview of Nonlinear Model Predictive Control Applications. In: *IFAC Workshop on Nonlinear Model Predictive Control. Assessment and Future Directions*. Ascona (Switzerland).
- Ramírez, D.R. y E.F. Camacho (2001). On the piecewise linear nature of Min-Max Model Predictive Control with bounded uncertainties. In: *Proc. 40th Conference on Decision and Control, CDC'2001*.
- Rawlings, J. (2000). Tutorial Overview of Model Predictive Control. *IEEE Control Systems Magazine* pp. 38–52.
- Richalet, J. (1993). Industrial Applications of Model Based Predictive Control. *Automatica* **29**(5), 1251–1274.
- Richalet, J., A. Rault, J.L. Testud y J. Papon (1976). Algorithmic Control of Industrial Processes. In: *4th IFAC Symposium on Identification and System Parameter Estimation*. Tbilisi USSR.
- Richalet, J., A. Rault, J.L. Testud y J. Papon (1978). Model Predictive Heuristic Control: Application to Industrial Processes. *Automatica* **14**(2), 413–428.
- Richalet, J., S. Abu el Ata-Doss, C. Arber, H.B. Kuntze, A. Jacobash y W. Schill (1987). Predictive Functional Control. Application to Fast and Accurate Robots. In: *Proc. 10th IFAC Congress, Munich*.
- Scokaert, P.O.M., D.Q. Mayne y J.B. Rawlings. (1999). Suboptimal model predictive control (feasibility implies stability). *IEEE Transactions on Automatic Control* **44**(3), 648–654.
- Tøndel, P. y T.A. Johansen (2002). *Proc. of the IFAC World Congress, B'02, Editors: E.F. Camacho, L. Basanez and J.A. de la Puente*. Chap. Complexity Reduction in Explicit Linear Model Predictive Control. Elsevier Science.
- Tyler, M.L. y M. Morari (1996). Propositional Logic in Control and Monitoring Problems. *Technical Report AUT96-15, Institut für Automatik, ETH- Swiss Federal Institute of Technology, Zurich, Switzerland*.
- Ydstie, B.E. (1984). Extended Horizon Adaptive Control. In: *Proc. 9th IFAC World Congress, Budapest, Hungary*.
- Zadeh, L.A. y B.H. Whalen (1962). On Optimal Control and Linear Programming. *IRE Trans. Automatic Control*.
- Zafiriou, E. (1990). Robust Model Predictive Control of Processes with Hard Constraints. *Computers and Chemical Engineering* **14**(4/5), 359–371.