

Fundamentos de modelo CEGH de procesos estocásticos multivariados.

SimSEE - Reporte técnico / IIE-FING-UDELAR

Ruben Chaer

Diciembre de 2011, rev. Junio 2013

Montevideo – Uruguay.

1. Introducción.

Uno de las mayores dificultades que aparece al intentar hacer un simulador de un sistema dado, es la representación de los procesos estocásticos involucrados en la realidad concreta (el sistema) que se está intentando simular.

A modo de ejemplo, supongamos un sistema formado por un aerogenerador un banco de baterías y como carga o demanda del sistema el consumo de una vivienda rural. Para simular en forma adecuada ese sistema, es necesario modelar de alguna forma la velocidad del viento y la demanda de energía como los dos procesos estocásticos de mayor importancia, pero también la disponibilidad del aerogenerador, la batería y la electrónica involucrada.

En algunos casos basta con disponer de un conjunto de realizaciones posibles de los procesos estocásticos para observar el comportamiento del sistema mediante simulaciones imponiendo dichos valores como “entradas al sistema”. En esos casos la evolución del sistema no tiene un lazo de control que intente compensar las consecuencias de las trayectorias previsibles de los procesos estocásticos y los procesos son modelados “sin estado” sin cometer por ello un error en los resultados de la simulación.

En otros casos, el sistema es capaz de reaccionar frente a la previsión (o pronóstico) de variación de las realizaciones de los procesos estocásticos. En estos casos, no basta con disponer de un conjunto de realizaciones de los procesos, es necesario tener un MODELO que permita evaluar las probabilidades de las realizaciones posibles futuras para poder calcular cuál será la evolución del sistema. En estos casos, el MODELO deberá incluir la definición del estado de los procesos estocásticos en base al cuál se pueda condicionar las trayectorias que seguirán a partir de un instante dado las realizaciones futuras.

Para fijar ideas, en el ejemplo de la instalación rural, la demanda del sistema puede tomar acciones que contrarresten las evoluciones de la generación. Si el modelo que se tenga de la velocidad de viento permite asignar una probabilidad alta a un día de poco viento, seguramente la demanda restringirá el uso en lo posible de energía durante el día para tener buena carga en la batería para la noche y a la inversa, si el modelo permite asignar alta probabilidad a un día de mucho viento, entonces los usuarios serán más liberales con el uso de la energía durante el día a riesgo de tener que prender velas en la noche. La operación de los sistemas con capacidad de almacenar energía es siempre un compromiso entre usar parte de la energía almacenada en el presente o guardarla para el futuro sustituyendo su uso en el presente por energías de otras fuentes (por ej. prender velas en el ejemplo de instalación rural).

Los sistemas, en menor o mayor grado, tienen inercias que hacen que la operación sea posible en forma continua. Por ejemplo, es muy poco probable que el promedio de caudal a una central hidroeléctrica tenga variaciones importantes en una hora para la siguiente. A mayor inercia en el proceso representado mayor podrá ser el paso de tiempo de simulación que permite observar la realización del proceso como una evolución suave y continua. A mayor inercia, mayor será el impacto del pasado del sistema en su futuro. En la jerga de modelado y control de sistemas se dice que un sistema es dinámico o tiene dinámica cuando el pasado del sistema es relevante para calcular

la evolución del sistema en el futuro. Es decir que además de la información de las entradas futuras es necesario conocer información del pasado para calcular la evolución futura. Decimos que disponemos de un modelo del sistema o de su dinámica cuando podemos tener una ecuación como la ec. 1

$$\dot{x} = f(x, r, u, t)$$

ec.(1) Modelo del sistema.
Ecuación de evolución del estado.

Dónde:

$x \in R^n$ es el vector de estado del sistema y por definición captura la información necesaria del pasado para poder calcular la evolución.

$r \in R^p$ es un vector de entradas no-controlables. Es información externa, sobre la que no se tiene control y que se supone independiente del valor de x

$u \in R^m$ es el vector de control. Son las variables sobre las que el Operador puede actuar. Por ejemplo, la potencia despachada de cada generador.

t Es el "tiempo".

La ec. 1 establece la derivada del estado como una función del estado actual, de las entradas no controlables, de las entradas de control y del tiempo actual. La evolución del vector de estado es entonces calculable a partir del conocimiento del estado x_0 en un tiempo inicial t_0 y de las entradas r y u para $t \geq t_0$ mediante la integración de la ec. 1 por la integral de la ec. 2.

$$x = x_0 + \int_{\xi=t_0}^{\xi=t} f(x(\xi), r(\xi), u(\xi), \xi) d\xi$$

ec.(2) Simulación del sistema

Para fijar rápidamente la idea de lo importante que es una buena representación de los procesos estocásticos téngase en mente el caso de la simulación del sistema de generación eléctrica del Uruguay. El Uruguay, tiene cuatro centrales hidroeléctricas ubicadas en dos ríos, con una potencia instalada de aproximadamente 1500 MW. Dependiendo de las lluvias del año, el cubrimiento de la demanda eléctrica se realizará con más o menos energía de origen hidráulico debiéndose cubrir el resto con generación en base a combustibles fósiles o con importación de energía. Los costos de éstas dos últimas fuentes rondan los 200 USD/MWh. La variabilidad de la generación hidráulica anual es importante lo que tiene como consecuencia que el Costo de Abastecimiento de la Demanda (CAD) tenga también un variabilidad importante. En la fig.1 se muestra la generación hidráulica que tendría lugar en una simulación del año 2010 suponiendo que se vuelven a repetir las lluvias de los últimos 100 años, ordenando de los años más lluviosos a los más secos de izquierda a derecha respectivamente.

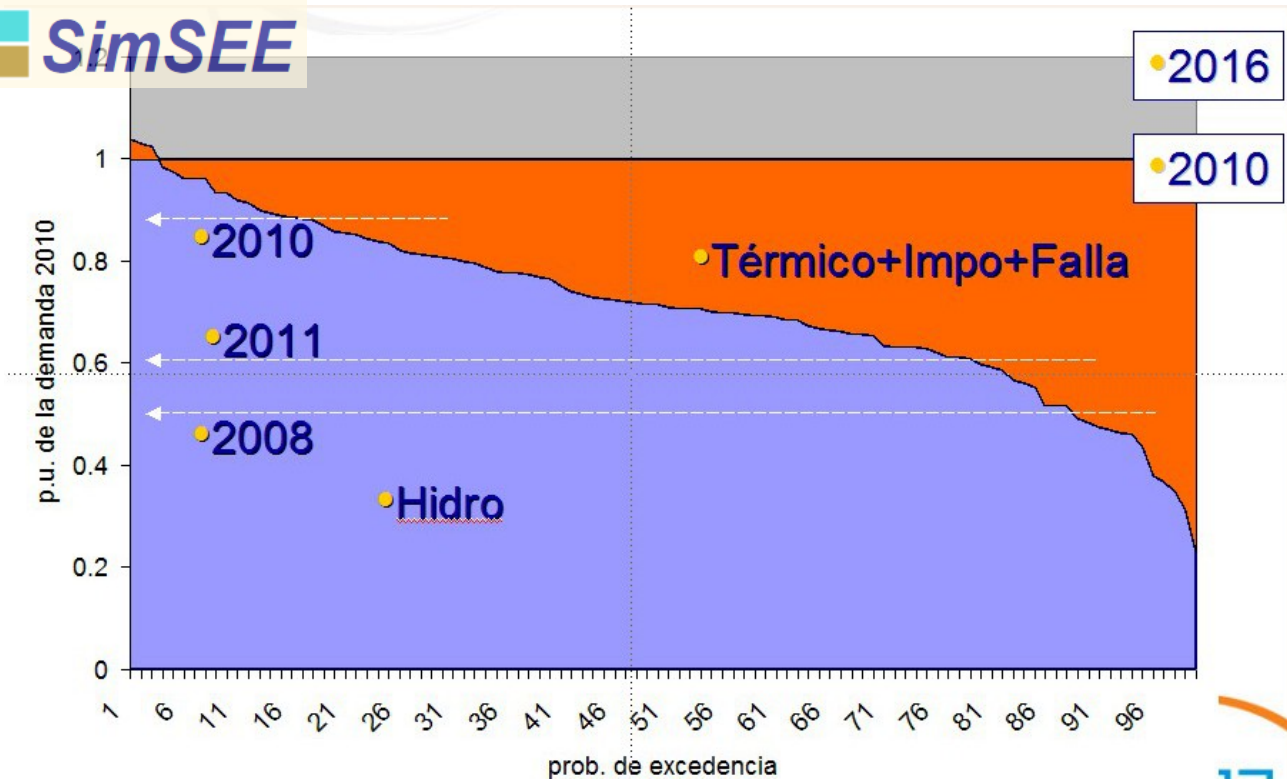


Fig. 1: Variabilidad de la generación hidráulica.

Insistiendo sobre el ejemplo de la generación hidroeléctrica (por ser la fuente de mayor volatilidad de los costos de generación en Uruguay), notemos que la generación hidroeléctrica del conjunto de las cuatro represas dependerá de los caudales afluentes a sus cuencas y de la capacidad de embalsar el agua de cada hidroeléctrica. La capacidad de embalsar el agua, debe pensarse como la capacidad de almacenar un recurso del sistema para su uso futuro. Almacenar tiene sentido si se sabe que ese recurso usado en el futuro evitará costos mayores que usado en el presente. Para tener la capacidad de “adivinar” cuánto valdrá el recurso en el futuro es necesario que los procesos estocásticos involucrados estén modelados de forma de representar lo mejor posible las “inercias” ocultas en los sub-sistemas asociados a dichos procesos. No menos importante es representar adecuadamente las correlaciones entre las diferentes variables estocásticas. A modo de ejemplo, dada la superficie del Uruguay, existe una correlación positiva entre los aportes a las represas la cual debe modelarse para no inducir al error de sobrevalorar la disponibilidad del recurso en una de las hidroeléctricas dado que cuando esté disponible en una existe por la correlación positiva una chance importante de que también esté en las demás. Otro ejemplo, es la velocidad de viento en los parques eólicos distribuidos en el territorio nacional. Si por simplicidad se modelaran todos los parques como instalados en un único sitio, todas las potencias de los parques variarían simultáneamente mostrando mayor variabilidad que la real, dado que en la medida en que las instalaciones se distancian geográficamente la correlación entre la velocidad de viento disminuye.

El propósito de este capítulo es mostrar una forma de analizar series temporales de datos que representen la salida de un proceso estocástico e identificar en base a ese análisis un modelo de Correlaciones en Espacio Gaussiano con Histograma - CEGH que puede ser utilizado en la representación de un sistema dinámico. En particular nuestro principal interés es el uso de estos modelos en la plataforma SimSEE. El modelo identificado tiene la propiedad fundamental de tener una estructura que facilita su inclusión en el modelado de la dinámica del sistema (ec. 1) y de generar series sintéticas que conservan las auto-correlaciones y las correlaciones cruzadas (en un espacio transformado que se describe más adelante) y los histogramas de amplitud en el espacio real. La conservación de los histogramas de amplitud es importante pues el sistema de generación de energía eléctrica tiene componente altamente no-lineales y por consiguiente la energía de los procesos no se conservaría si no se respetan las amplitudes.

2. Elementos de estadística.

La intención de esta sección es resumir las bases de estadística necesarias para el desarrollo del modelo y refrescar al lector las definiciones y notación utilizadas.

2.1. Función de distribución acumulada de probabilidad.

Dada una variable aleatoria $X \in R$ su función de distribución acumulada de probabilidad $F_X(x): R \rightarrow R$ es por definición: $F_X(x) = P(X \leq x)$, siendo $P(X \leq x)$ la probabilidad de que la variable aleatoria X tome valores inferiores o iguales al valor dado x .

Se puede probar que con esa definición, $F_X(x)$ es monótona creciente y va de 0 a 1.

2.2. Función de densidad de probabilidad.

Dado un vector de variables aleatorias $x \in R^n$, podemos definir la función de densidad de probabilidad $p_x(x): R^n \rightarrow R$ tal que la probabilidad de que el vector x pertenezca a un volumen $\Omega \in R^n$ está dado por la integral en el volumen: $\int_{\zeta \in \Omega} p_x(\zeta) dV_\zeta$ siendo $dV_\zeta = d\zeta_1 \cdot d\zeta_2 \dots d\zeta_n$ el elemento de volumen.

2.3. Algunos resultados para aleatorias univariadas.

A continuación se enumeran algunos resultados útiles para variables aleatorias univariadas.

a) Relación entre las funciones de distribución y de densidad para una variable aleatoria univariada.

Por definición de la función de distribución acumulada y de densidad de probabilidad de las secciones 2.1 y 2.2 se pueden establecer las relaciones siguientes:

$$F_X(x) = P(X \leq x) = \int_{\zeta = -\infty}^{\zeta = x} p_X(\zeta) d\zeta$$

ec.(3) Relación entre las funciones de Distribución y la Densidad de Probabilidad.

o lo que es equivalente:

$$p_X(x) = \frac{d}{dx} F_X(x)$$

b) Transformación de una variable aleatoria.

Dada una variable aleatoria y su funciones características $F(x)$ y $p_x(x)$ y dada una transformación $T: R \rightarrow R$ monótona consideremos la variable aleatoria $y = T(x)$.

Por definición la Distribución de la nueva variable será:

$$F_Y(y) = P(Y \leq y) = \int_{\zeta = -\infty}^{\zeta = y} p_Y(\zeta) d\zeta$$

Si $y=T(x)$ es monótona creciente, haciendo el cambio de variable en la integral se puede escribir:

$$F_Y(y)=P(Y\leq y)=\int_{\xi=-\infty}^{\xi=y} p_Y(\xi) d\xi = \int_{\beta=-\infty}^{\beta=x} p_Y(T(\beta)) \frac{\partial T(\beta)}{\partial x} d\beta \quad \text{donde } x=T^{-1}(y)$$

El integrando de la última integral es identificable como $p_X(x)$ con lo que se puede escribir:

$$\text{si } y=T(x) \text{ entonces: } p_Y(y)=p_X(x) \cdot \left(\frac{\partial T(x)}{\partial x}\right)^{-1} \quad \text{ec.(4) Cambio de variable de una aleatoria univariada.}$$

c) Generador de números aleatorios con una función de densidad de probabilidad dada.

Dada una función de densidad de probabilidad $p_X(x)$ es inmediato disponer de la distribución acumulada de probabilidad $F_X(x)$ y viceversa a partir de la ec. 3.

La función de distribución acumulada $F_X(x)$ es monótona creciente desde el valor 0 (cero) para $x=-\infty$ hasta el valor 1 (uno) para $x=+\infty$ y por lo tanto es invertible. Sea $F_X^{-1}(u)$ la inversa de la función de distribución. Por construcción $F_X^{-1}(u)$ está definida del intervalo $[0,1]$ en R .

Sea u una variable aleatoria con distribución uniforme en $[0,1]$. Típicamente los diferentes lenguajes de programación tienen una función para generar números pseudo-aleatorios con distribución uniforme. Por ejemplo, la función *random* de Pascal.

$$\text{Por definición } p_U(u)=\begin{cases} 1; \forall u \in [0,1] \\ 0; \forall u \notin [0,1] \end{cases}$$

Consideremos el cambio de variable $\xi=T(u)=F_X^{-1}(u)$ y utilizando la relación de la ec. 4, podemos escribir que:

$$p_\xi(\xi)=p_U(u) \cdot \left(\frac{\partial T(u)}{\partial u}\right)^{-1}$$

$$\text{como: } \frac{\partial T(u)}{\partial u} = \frac{\partial F_X^{-1}(u)}{\partial u} = \left(\frac{\partial F_X(\xi)}{\partial x}\right)^{-1} = (p_X(\xi))^{-1}, \text{ sustituyendo } \frac{\partial T(u)}{\partial u} \text{ en la ec. anterior}$$

tenemos:

$p_\xi(\xi)=p_U(u) \cdot p_X(x)$ y como $\xi=T(u)=F_X^{-1}(u)$ mapea el intervalo $[0,1]$ en R para cualquier valor de ξ , podemos asegurar que $p_U(u)=1$ y por tanto hemos encontrado un mecanismo para generar una variable aleatoria ξ cuya función densidad de probabilidad coincide con una dada: $p_\xi(\xi)=p_X(x)$.

Este es el mecanismo utilizado en la plataforma SimSEE para generar variables aleatorias con funciones características dadas a partir de un generador de números pseudo aleatorios.

d) Deformador Uniformizante.

En la sección anterior se mostró como generar una variable aleatoria con una distribución dada a partir de una variable con distribución uniforme mediante el cambio de variable de la uniforme usando a inversa de la función de distribución acumulada de la distribución objetivo. Utilizando la inversa de dicha transformación es posible pasar de una variable con distribución dada a una variable aleatoria con distribución uniforme.

Es decir, si se tiene una variable aleatoria X con distribución $F_X(x)$ entonces, la variable

aleatoria obtenida mediante la transformación $u = F_X(x)$ tendrá distribución uniforme.

En la jerga del modelado CEGH decimos que la distribución $F_X(x)$ es un “deformador uniformizante” de la variable aleatoria X dada una serie de realizaciones de X si aplicamos “deformador uniformizante” a cada punto de la serie se obtiene una serie de números entre 0 y 1 con distribución uniforme.

2.4. Independencia estadística.

Se dice que dos conjuntos de variables aleatorias x e y son independientes (o tienen independencia estadística) si el conocimiento de los valores que tome uno de los conjuntos no condiciona los valores que toma el otro.

Si dos variables aleatorias son independientes, entonces la función densidad de probabilidad del conjunto de ambas variables es el producto de las funciones de densidad de cada una.

$$p_{x,y}(x, y) = p_x(x) p_y(y) \quad \text{ec.(5) Proceso conjunto de variables independientes.}$$

Si las variables no son independientes, el conocimiento de los valores x condiciona los valores que puede tomar y y viceversa. Para cada valor de x será entonces posible construir la función densidad de probabilidad de suceso condicionado y/x (se lee “ y dado x ”) y la función densidad de probabilidad de conjunto y, x podrá calcularse como se muestra en la ec. 6 con la relación conocida como teorema de Bayes.

$$p_{x,y}(x, y) = p_x(x) p_{y/x}(x, y) = p_y(y) p_{x/y}(x, y) \quad \text{ec.(6) Teorema de Bayes.}$$

2.5. Medidas de probabilidad.

La forma de realizar medidas de probabilidad es mediante la integral de una función $m(x): R^n \rightarrow R^m$ ponderada por la función densidad de probabilidad como se muestra en la ec. 7.

$$\langle m(x) \rangle_x = \int_{\xi \in R^n} m(\xi) \cdot p_x(\xi) dV_\xi \quad \text{ec.(7) Medida de probabilidad.}$$

Para realizar entonces cualquier cuantificación probabilística es necesario conocer la función de densidad de probabilidad $p_x(\xi)$.

Ejemplos de medidas de probabilidad son el Valor Esperado y la matriz de Covarianzas definidas por las ec. 8 y ec. 9 respectivamente.

$$E(x) = \langle x \rangle_x = \int_{\xi \in R^n} x \cdot p_x(\xi) dV_\xi \quad \text{ec.(8) Valor Esperado}$$

$$\Sigma_{xx} = \langle (x - E(X))(x - E(X))^T \rangle_x \quad \text{ec.(9) Matriz de Covarianzas.}$$

Dos procesos aleatorios $x \in R^n$ e $y \in R^n$ con iguales funciones de densidad de probabilidad $p_x(x) = p_y(x) = p(x)$ son equivalentes desde el punto de vista de cualquier medida de

probabilidad del tipo definido en la ec. 7. Por ejemplo, tendrán igual valor esperado e igual matriz de covarianzas.

Observar que si dos procesos coinciden en valor esperado y en la matriz de covarianzas no necesariamente tienen la misma función de densidad de probabilidad y por tanto podrán diferir en otras medidas de probabilidad.

En resumen, lo importante a los efectos de caracterizar un proceso estocástico es obtener una representación de su función de densidad de probabilidad. Sólo así se está en condiciones de reproducir cualquier medida de probabilidad.

2.6. Transformaciones compactas.

Cualquier transformación biunívoca de $T: R^n \rightarrow R^n$ que transforme conjuntos compactos del espacio de salida en volúmenes compactos del espacio de llegada mantiene las equivalencias entre los procesos estocásticos. Esto es sencillo de demostrar, pues cualquier integral de medida de probabilidad (ec. 7) en el espacio de llegada, puede expresarse en el espacio de salida pues para cada elemento de volumen dV_ξ del espacio de llegada se puede identificar el correspondiente elemento de volumen dV_β del espacio de salida y viceversa.

Dado el proceso transformado, $y=T(x)$ siendo T una transformación biunívoca es posible escribir $x=T^{-1}(y)$. Las integrales de probabilidad del proceso transformado se pueden expresar como se muestra en la ec.

$$\int_{\xi \in R^n} m(\xi) \cdot p_y(\xi) dV_\xi = \int_{\beta \in R^n} m(\beta) \cdot p_x(\beta) dV_\beta \quad \text{ec.(10) Correspondencia de las medidas en una transformación biunívoca.}$$

Las integrales de la ec. 10 son integrales múltiples relacionadas por el cambio de variable $y=T(x)$. Esto permite establecer que las funciones de densidad de probabilidad están relacionadas por la ec. 11.

$$p_x(x) = |J_T(T(x))| \cdot p_y(T(x)) \quad \text{ec.(11) Transformación de la densidad de probabilidad.}$$

dónde $|J_T(T(x))|$ es el determinante de la matriz Jacobiana de la transformación $y=T(x)$

2.7. Procesos estocásticos.

Un proceso estocástico genera una serie (o secuencia) de variables aleatorias definidas de "el tiempo" en un conjunto dado. La expresión "el tiempo" significa los números reales para procesos de tiempo continuo o el conjunto de los números enteros para procesos discretos. En ambos casos con un sentido definido que dado un punto ordena el espacio en "pasado" y "futuro" en referencia a dicho punto. Por comodidad y porque la implementación de modelado CEGH es para procesos que serán tratados como discretos (en general son procesos continuos muestreados a intervalos de tiempo constante) usaremos la variable entera k para representar el tiempo.

Se llama "realización del proceso estocástico" a cualquier serie $\{x_k\}$ de las posibles que puede generar el proceso. El vector x_k representa para el instante k los valores que toma el conjunto de salidas del proceso. Se llama "ensamble de realizaciones" al conjunto de todas las realizaciones posibles del proceso. Cuando se está hablando de procesos del mundo real, existe una

realización que se llama "la histórica" o también Crónica Histórica que es la realización que ha tenido lugar en el pasado para ese proceso. A modo de ejemplo, si se observan los números de una ruleta en el casino, es posible identificar por ejemplo los últimos 100 números salidos en forma ordenada como la crónica de las últimas 100 jugadas. Cuando se mira hacia el futuro, el proceso generará una realización de un conjunto de realizaciones posibles, pero no es posible adelantar cuál. En el caso de la ruleta, el siguiente número será uno entre 0 y 36 todos con igual probabilidad. En este ejemplo, el pasado no condiciona las probabilidades de la realización del futuro. Es entonces un ejemplo de proceso estocástico "sin estado". En este ejemplo, de una sola variable aleatoria resulta intuitivo, que para describir la función densidad de probabilidad de los números de la siguiente jugada (esto es tener la descripción del ensamble de realizaciones posibles para la siguiente jugada) es posible hacerlo en base a la realización histórica. Basta simplemente armar un histograma a partir de ordenar los números y salvo defectos mecánicos del sistema se deberá obtener que la probabilidad de cada número es $1/37$. El que sea posible describir el ensamble a partir de una realización del proceso, que resulta intuitivo para el ejemplo de la ruleta, no se cumple en todos los casos. Por ejemplo, si la variable estocástica es el precio del barril de petróleo, como por simple observación es un precio en "expansión" no es posible determinar por construcción del histograma de los valores históricos las características del ensamble posible de realizaciones futuras.

a) Procesos en estado estacionarios.

Se dice que un proceso está en estado estacionario cuando las propiedades estadísticas en el ensamble de cualquier subconjunto de su serie de salida $x_k, x_{k-1}, \dots, x_{k-n}$ son independientemente del tiempo k . El ejemplo de los números de la ruleta es un proceso en estado estacionario y el ejemplo del precio del barril de petróleo es claramente un proceso no estacionario.

b) Procesos ergódicos.

Decimos que un proceso es ergódico para una cierta función medible $m(x_k, x_{k-1}, \dots, x_{k-n})$ definida sobre subconjuntos de la serie de salidas del proceso si la integral en el ensamble del subconjunto $x_k, x_{k-1}, \dots, x_{k-n}$ puede sustituirse por una integral de medida en el tiempo a partir de una realización dada como se muestra en la ec. 12.

$$\int_{\xi_k \in R^{n(j-j+1)}} m(\xi_k) \cdot p_y(\xi_k) dV_{\xi_k} = \lim_{H \rightarrow +\infty} \left(\frac{1}{2H} \sum_{j=-H}^{j=+H} m(\beta_j) \right)$$

ec.(12) Estimación de la medida en el ensamble a partir de una realización dada en un proceso ergódico.

dónde: ξ_k es un punto del ensamble del conjunto de variables $x_k, x_{k-1}, \dots, x_{k-n}$ referido al tiempo k y $\beta_j = x_j, x_{j-1}, \dots, x_{j-n}$ es un punto de la realización considerada para la estimación en el tiempo j .

Observar que en la ec. 12, el límite del lado derecho no depende el instante de tiempo k al que se refiere la medida del lado izquierdo y por lo tanto, la medida del lado izquierdo debe ser independiente de k lo que implica que el proceso está en estado estacionario para dicha media. Entonces, para que un proceso sea ergódico es necesario que sea estacionario.

Un ejemplo sencillo de proceso ergódico es el del juego de lotería previamente mencionado.

2.8. Procesos Gaussianos.

Se dice que un conjunto de variables aleatorias son gaussianas cuando su función densidad de probabilidad puede escribirse de la forma:

$$p_x(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-E(x))^T \Sigma^{-1}(x-E(x))} \quad \text{ec.(13) Densidad de probabilidad gaussiana multivariable.}$$

Un proceso es gaussiano si cualquier subconjunto $x_k, x_{k-1}, \dots, x_{k-n}$ de sus salidas tiene distribución gaussiana.

Entre las propiedades interesantes de los procesos gaussianos, destaca que su función de densidad de probabilidad queda totalmente definida por el conocimiento del valor esperado y la matriz de covarianzas. Si estamos hablando de una sola variable aleatoria, el conocimiento del valor esperado y de la varianza determinan entonces completamente la función densidad de probabilidad si la variable responde a un proceso gaussiano.

Otra propiedad relevante para los propósitos de este trabajo es que la combinación lineal de variables aleatorias gaussianas independientes es gaussiana. Esto implica ni más ni menos que si un sistema lineal es atacado por vector de ruidos blancos (por ruido blanco se entiende una señal “sin memoria” o sea que el presente tiene independencia estadística del pasado) gaussianos independientes, sus salidas son gaussianas pues la salida será una combinación de las entradas y sus pasados que al suponerse entradas “sin memoria” son todas variables independientes.

Entonces, las dos propiedades de los procesos gaussianos que se utilizarán fuertemente en el modelado CEGH son:

1. La función densidad de probabilidad queda totalmente determinada por el vector de valor esperado y por la matriz de covarianzas.
2. La combinación lineal de variables aleatorias gaussianas independientes es una variable aleatoria gaussiana.

3. Definición del Modelado CEGH

Como se mencionó en la introducción el objetivo del modelado CEGH es tener una representación de un proceso estocástico que permita su incorporación en el modelado de un sistema dinámico.

Para la identificación del modelo se dispone generalmente de la realización histórica del conjunto de variables que forma la salida del proceso estocástico a modelar. El desafío del modelado, es lograr crear un modelo que “capte” la estructura de la dinámica del proceso estocástico de forma de poder inferir en todo momento “el

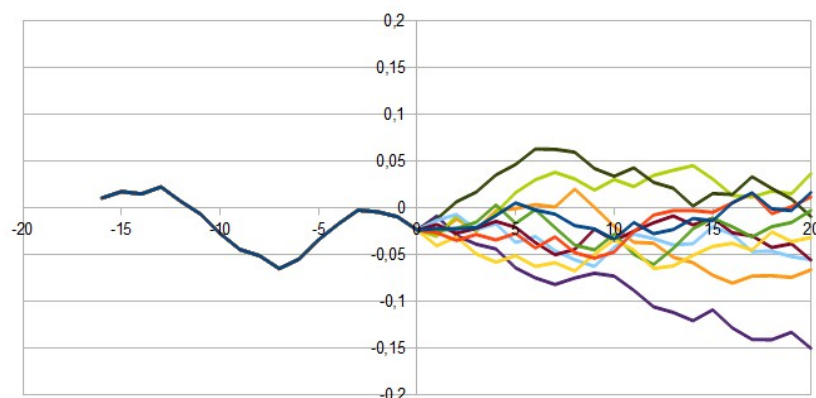


Fig. 2: Salida de un proceso estocástico con inercia

cono de las salidas” del proceso. Si el proceso no tuviera dinámica, la salida siguiente no estaría condicionada por el pasado y podría modelarse simplemente como un sorteo con un determinado histograma. Por ejemplo, en el caso del La Ruleta, como proceso estocástico, el modelo sería simplemente un sorteo uniforme entre 0 y 36 sin ninguna dependencia con el pasado.

En el caso de los procesos que intervienen en la simulación de sistemas de energía, como pueden ser las lluvias, la radiación solar, el viento, la demanda, etc., los valores tienen cierta continuidad que permite pensar en “el cono del futuro” como ramilletes de realizaciones posibles a partir de un “presente conocido”. Este comportamiento de “continuidad” que permite dibujar las salidas como un cono (ver Fig. 2), implica que la salida del proceso tiene dependencia estadística con su pasado y por lo tanto es pensable en obtener para el proceso una representación del tipo de la ec. 1 con la identificación de un “vector de estado” que almacena la información relevante del pasado.

La idea detrás del modelado CEGH, es obtener un modelo del tipo de la ec. 1; a partir de la realización histórica de las salidas de un proceso estocástico; de forma de que el modelo obtenido, si es usado para generar series sintéticas (o sea generar realizaciones), tenga la misma función de densidad de probabilidad para cualquier subconjunto $x_k, x_{k-1}, \dots, x_{k-n}$ de la serie de realizaciones. Observar que el índice k identifica una “posición temporal” en el ensamble de realizaciones y los índices $k, k-1, \dots, k-n$ identifican una ventana de $n+1$ posiciones del ensamble de realizaciones que se desplaza al variar la posición temporal k del subconjunto.

Lo primero a recordar, es que el modelado se realiza a partir de la única realización histórica disponible con lo cual las estimaciones NUNCA son en el ensamble de lo posible sino que son sobre la realización histórica. Esto lleva que sea necesario intentar “ergodizar” el proceso mediante transformaciones más o menos automáticas y que pueden contener información adicional propia del conocimiento del proceso.

Por ejemplo, si se está modelando la radiación solar que recibe un panel solar sobre la superficie de La Tierra, el conocimiento de la geometría estelar y del ciclo anual puede utilizarse para modificar la serie de radiación a otra serie que no tenga la estacionalidad diaria y anual previo a intentar el modelado.

Si se considera el subconjunto en que $n=0$ (esto es la ventana en el ensamble selecciona solo un casillero temporal), que el modelo genere series sintéticas con igual función densidad de probabilidad que el proceso original (el de la serie usada para identificar el modelo) significa que el histograma de amplitud de las series sintéticas coincide con el histograma de amplitudes de la realización histórica.

Si se considera una ventana que involucre n elementos temporales; el modelo logra reproducir la función de densidad de probabilidad conjunta de $x_k, x_{k-1}, \dots, x_{k-n}$ y entonces el modelo será capaz de captar la dependencia estadística de las salidas del proceso con su pasado hasta n pasos de tiempos. En la frase anterior, se priorizó al hablar de la relación de dependencia un sentido que es “el pasado influyendo sobre el presente”. Si bien puede no importar a la formalidad matemática, en la práctica la hipótesis de trabajo es que la “causalidad” existe y que el futuro no puede afectar al pasado.

Otro concepto del mundo real, es el de “El Olvido” en el sentido de que la dependencia estadística tiene que ir perdiendo importancia en la medida en que el tiempo transcurre. Si bien el primer principio de la termodinámica, nos puede llevar a pensar que al haber un balance permanente de energía y materia, todo lo que no está en un lado está en otro y por lo tanto no hay olvido, el segundo principio de la termodinámica implica un aumento permanente de la entropía que “borronea” la información del pasado.

Notación. En lo que sigue usaremos la letra y para referirnos a la serie de medidas (la realización histórica) o la salida del proceso y reservaremos la letra x para referirnos a la misma serie en un espacio transformado en que las series son gaussianas. Es así que diremos que x es la imagen de y en el Espacio Gaussiano y que y es la imagen de x en el Espacio Real (o espacio de la señal original).

En la práctica, dado que la información disponible en la única realización histórica es finita, no será posible identificar un modelo que asegure reproducir la función de densidad de probabilidad para cualquier valor de la ventana n que determina los subconjuntos $x_k, x_{k-1}, \dots, x_{k-n}$. A mayor valor de n que se pretenda abarcar, mayor será la cantidad de parámetros del modelo con el mismo sustento estadístico (la información de la realización histórica).

Supongamos que se logra encontrar una transformación compacta que transforma el proceso original a un espacio G donde el mismo es ergódico. (nadie está afirmando que dicha transformación exista.) en ese caso, las funciones de densidad de probabilidad de los subconjuntos $y_k, y_{k-1}, \dots, y_{k-n}$ transformados de los subconjuntos $x_k, x_{k-1}, \dots, x_{k-n}$ son independientes de la posición temporal k .

El modelado CEGH, implica buscar una transformación

$y_k, y_{k-1}, \dots, y_{k-n} = G(x_k, x_{k-1}, \dots, x_{k-n}, k)$ que “ergodice” y “gaussianice” el proceso y lo transforme en un proceso ergódico y gaussiano y luego en el espacio transformado buscar un sistema lineal que al ser atacado por ruido

blanco gaussiano reproduzca las funciones de densidad de probabilidad de los subconjuntos $y_k, y_{k-1}, \dots, y_{k-n}$. A los efectos de que el modelado en el espacio transformado sea útil, la transformación $G(\cdot, k)$ debe ser compacta (esto es transformar en unívocamente volúmenes compactos del espacio de salida en volúmenes compactos del espacio de llegada). A la inversa de esta transformación la denotaremos como $H(\cdot, k) = G^{-1}(\cdot, k)$.

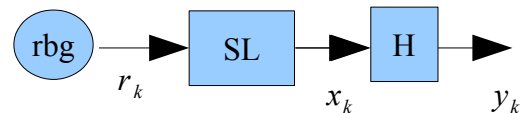


Fig. 3: Modelo CEGH

Una vez obtenido el modelo CEGH, el mismo puede representarse por el esquema de la Fig. 3, en el que de izquierda a derecha se tiene: Una fuente de ruidos blancos gaussianos (rbg) independientes que genera la serie de vectores $\{r_k\}$ que alimenta la entrada del Sistema Lineal (SL) el que genera a la su salida una serie de vectores $\{x_k\}$ cuyos elementos son gaussianas y que si el modelo fue exitosamente identificado a partir de la información histórica mantienen la “misma estructura estadística” que el proceso que intenta modelar. En este contexto, se entiende por “misma estructura estadística” a que las integrales de medidas del estilo de las definidas por la ec. 7 se aproximen a la misma integral calculada sobre la serie (ergodizada y gaussianizada) del proceso original. Las integrales de medida utilizadas para realizar este chequeo pueden variar de acuerdo al objetivo del modelado. Generalmente las medidas de interés son aquellas que guarden cierta relación con la capacidad del sistema en el que se va a utilizar el modelo para “filtrar” variaciones. Por último, el modelo transforma la serie $\{x_k\}$ del “espacio gaussiano” a una serie en el “espacio real” mediante la transformación $H(\cdot, k)$.

No tenemos certeza de que la transformación que $G(\cdot, k)$ exista ni un método genérico para calcularla, pero en la práctica se ha desarrollado una aproximación que es la que se desarrolla en la sección 3.2 Primer aproximación al modelado de procesos multi variados. Transformaciones “por paso” y “por señal”. que ha mostrado ser útil para los casos de modelado de algunos procesos relacionados con variables de la Naturaleza como se los aportes a las centrales hidroeléctricas, la

radiación solar y la velocidad de viento.

3.1. Proceso aleatorio ergódico univariado.

Dado proceso aleatorio con una salida de dimensión 1 (uno) y_k . Las variables aleatorias Y_k tendrán una cierta distribución $F_{Y_k}(y)$. Si el proceso es ergódico, la distribución no dependerá del tiempo k y se puede escribir por tanto $F_{Y_k}(y) = F_Y(y)$.

Supondremos que para estimar la distribución $F_Y(y)$ se dispone de la realización histórica $\{y_k\}$ de las Y_k .

En la búsqueda de una transformación $G(\cdot, k)$, observar que utilizando los resultados de las secciones 2.3.c y 2.3.d es posible a partir de una distribución conocida de una variable aleatoria, transformarla en otra variable aleatoria cuya distribución sea gaussiana. Para ello basta con “uniformizar” la variable original con el resultado de 2.3.d y luego gaussianizar la variable uniforme con el resultado de la 2.3.c. Esto nos asegura que la variable aleatoria tiene distribución gaussiana, pero no necesariamente asegura que el proceso sea gaussiano.

El primer paso es estimar $F_Y(y)$ a partir de la realización histórica $\{y_k\}$ lo que es posible por haber supuesto que el proceso es ergódico. Una forma sencilla de realizar esta estimación es re-ordenar los elementos de la realización histórica en forma creciente obteniendo así una serie monótona creciente con todos los valores históricos.

Sea $F_n(y)$ la función de distribución de probabilidad acumulada de la distribución normal (gaussiana de valor medio nulo y varianza unidad). Entonces, se pueden calcular las transformaciones G y H como: $x = G(y, k) = F_n^{-1}(F_Y(y))$ y $y = H(x, k) = F_Y^{-1}(F_n(x))$ respectivamente. Como se puede apreciar en este caso al suponer que el proceso es ergódico para la construcción de $F_Y(y)$ se pierde la dependencia del tiempo k .

3.2. Primer aproximación al modelado de procesos multi variados. Transformaciones “por paso” y “por señal”.

En esta sección se muestra un enfoque posible para la construcción de las funciones de transformación de un conjunto de N señales a partir de sus realizaciones históricas representadas por la serie vectorial $\{y_k\}$, siendo el índice k el identificador del “tiempo” en las series de medidas e y_k el vector de los N valores de la realización histórica del conjunto de señales en el instante k . El sentido de la expresión “por paso” no es otro que el que expresa la dependencia en k de las transformaciones $x = G(y, k)$ y su inversa $y = G(x, k)$.

En la sección 3.1 se mostró cómo construir $G(\cdot, k)$ para un proceso ergódico uni-variado. La aproximación propuesta en esta sección para el cálculo de una $G(\cdot, k)$ para un ramillete de series es tratar cada señal como si fuese independiente al momento de identificar las transformaciones

$G(\cdot, k)$ y agregar la dependencia en el modelo lineal en espacio gaussiano. El que esa sea una buena o mala aproximación a una transformación $G(\cdot, k)$ que permita considerar al proceso de las N en el espacio transformado como ergódico y gaussiano depende de la dependencia entre las señales. Generalmente, mediante cambios de variables adecuados previos al cálculo de las transformaciones por señal es posible tener una buena aproximación pero hay casos de dependencia entre las señales que no es posible identificar esos cambios de variables y entonces, el método de buscar las transformaciones “por señal” no es aplicable.

Por construcción, al realizar las transformaciones “por señal”, se asegura que si se observa cada señal por separado, el modelo “es bueno” y eso ya es un “marco” de aproximación al proceso del conjunto de señales.

Dependiendo del proceso al que corresponden las series habrá cierto comportamiento estacional (por ejemplo la radiación solar sobre un punto de la tierra tiene un comportamiento claramente de estacionalidad anual) o podría haber también un comportamiento tendencial como el que tienen típicamente las series que representan precios en una moneda dada.

Lo primero es detectar las estacionalidades y/o tendencias y quitarlas de las series tratando de transformarlas en series que puedan estar asociadas a un proceso en estado estacionario. En esta etapa juega un papel fuerte el conocimiento que se tenga del fenómeno asociado al proceso estocástico. Otra herramienta útil puede ser realizar un análisis de la distribución de potencias en el espectro de frecuencias para determinar si el mismo tiene potencia concentrada en algunas frecuencias que estén marcando la existencia de una señal periódica. El ejemplo de identificación de la radiación solar es un ejemplo en el que el conocimiento del sistema solar ayuda en intentar proponer un modelo de la radiación recibida en plano horizontal por un panel como el producto de la radiación solar extraterrestre representada como una senoide multiplicada por un índice de nubosidad.

Sea $\{y_k^i\}$ la serie correspondiente a la señal i que está representada por la componente i -ésima del vector y_k . Formas de transformar las $\{y_k^i\}$ para lograr series que por lo menos parezcan salidas de un proceso ergódico hay muchas y posiblemente es en esta etapa del proceso de identificación en que el “investigador” (o sea usted) puede agregar información del conocimiento físico del proceso.

En término del análisis espectral, para poder identificar una periodicidad, en la práctica, la misma debe ser de frecuencia tal que por lo menos tenga 5 (cinco) ciclos completos en la serie de datos analizados. Por ejemplo, en el caso de aportes hidráulicos a las represas de Uruguay, se dispone de 100 años de aportes medios semanales, con lo cual el límite para identificar un ciclo a partir del análisis espectral 20 años. Es decir que ciclos con períodos de repetición superiores a 20 años deberán analizarse por otros métodos.

A los efectos de la identificación del modelo CEGH, se supone entonces que los comportamientos tendenciales y/o variaciones de muy baja frecuencias fueron quitados previamente de las señales (en base a conocimiento del fenómeno u otras técnicas) y que se está trabajando sobre series que parecen corresponder a un proceso en estado estacionario aunque pueda tener estacionalidades marcadas. Bajo este supuesto, las funciones $G(\cdot, k)$ tendrán la dependencia en el tiempo k que permite captar las estacionalidades. Lo típico es poder describir ciclos de estacionalidad y para ello podemos especificar el largo del período estacional con un parámetro que llamaremos NPP (Número de Puntos por Período) y buscar funciones $G(\cdot, k)$ tal que $G(\cdot, k) = G(\cdot, k + h \cdot NPP)$ para cualquier entero h . La idea es que el comportamiento estadístico tienen una estacionalidad que se repite cada NPP puntos de la serie de datos. Para el cálculo de las NPP funciones $G(\cdot, k)$ de cada señal, habrá que generar NPP histogramas de amplitudes de la variable para describir la variable según el momento dentro del ciclo estacional. Por ejemplo si el proceso en cuestión es la radiación solar, y dividimos el año en cuatro estaciones podemos construir los histogramas de la radiación recibida para cada una de las cuatro estaciones.

En la práctica, para el análisis de las series, el parámetro NPP determina que es suficiente con crear NPP funciones $G(\cdot, k)$ etiquetables con $k = 1, \dots, NPP$. Para la confección de estos histogramas de amplitudes, el largo de la serie de datos se ve dividido por NPP disponiéndose por lo tanto de menos información para la confección de cada histograma estacional que si se construyera un único histograma. Esta falta de información puede llevar a tener una variación artificial de los histogramas de un paso a otro dentro del período estacional por causa de

la población reducida de datos con que fueron obtenidos. Para evitar dar un peso exagerado a la ocurrencia de un valor particular de esa población escasa, es posible considerar que la precisión de un dato dentro del período estacional puede ser relativa.

Puesto en palabras y sobre un ejemplo, si en la semana del 17/8/2011 hubo un aporte muy importante de lluvias a una represa, en la realidad la precisión de esa fecha no es información relevante (y se vuelve desinformación en una población escasa) y razonablemente se podría decir que es esa fecha más menos 2 semanas en cuanto a la relevancia estadística del dato. Esto lleva a la definición de un nuevo parámetro NPO que llamaremos número de pasos de Overlapping o solapamiento, de forma tal que el dato y_k^i participe de la formación de los histogramas de amplitudes asociados a los pasos $k_h = (k \text{ mod } NPP) + j$ con j entero tal que $-NPO \leq j \leq NPO$ y siendo “mod” la operación residuo.

Si $NPO=0$ el dato y_k^i tiene peso estadístico en el histograma $k_h = k \text{ mod } NPP$ y solo en ese histograma.

Si $NPO=2$ el dato y_k^i tiene peso estadístico en el histogramas $k_h = k \text{ mod } NPP$ y en los dos siguientes y en los dos anteriores dentro del período estacional.

En el caso del análisis de las velocidades de viento, surgió la necesidad de definir un comportamiento estacional diario además de la estacionalidad anual. En este caso hay un “miniciclo” dentro de un un ciclo. Si las series son horarias, tendremos una estacionalidad de 8760 horas correspondiente al ciclo anual y unos miniciclos de 24 horas. Para poder definir los miniciclos se agregaron dos parámetros que son $NPMC$ que es el número de puntos por miniciclo y NPT que es el número de puntos de *Traslapping*. En el ejemplo de la velocidades de viento, una definición posible sería $NPP=8760; NPO=3; NPMC=24; NPT=2$ lo que implicaría la generación de 8760 histogramas en base a la serie de datos, para lo cual cada muestra se considerará válida en el instante a la que corresponde en las tres horas anteriores y siguientes y ese esquema (de 6 horas) se considera que aporta información para el día al que corresponde el dato (miniciclo del dato) y en los dos miniciclos siguientes y anteriores. Con esta configuración, un mismo dato aporta información a $(1+2*NPO)*(1+2*NPT)$ de los NPP histogramas que se crearán. La selección de los parámetros es un compromiso entre el conocimiento de los procesos físicos y el largo de las series de datos disponibles que condiciona la población con que se realizara la construcción de cada histograma. Si las series son muy largas, no es relevante preocuparse de los parámetros NPO y NPT .

3.3. Identificación de sistema lineal a partir de series de entrada y salida.

En esta sección se supondrá que se trabaja sobre series numéricas que son gaussianas de valor medio nulo y varianza unitaria. Es decir supondremos que estamos trabajando siempre en el espacio gaussiano.

Sea $\{X_k\}$ la serie del vector salidas del filtro lineal a identificar y $\{Z_k\}$ la serie del vector de entradas conocidas al filtro. Dado la cantidad de retardos N_r del filtro (cantidad de retardos a considerar en su memoria) la ecuación del filtro que nos proponemos identificar es la siguiente:

$$X_{k+1} = \sum_{h=0}^{h=N_r-1} A_h \cdot X_{k-h} + C \cdot Z_k + B \cdot R_k$$

ec.(14) Modelo del proceso.

Supongamos que se dispone de una serie de medidas $\{X_k\}$ de la salida de un proceso estocástico que lo queremos modelar como una caja negra que recibe a su entrada la serie $\{Z_k\}$ que es conocida y que se quiere identificar las matrices A_h , B y C que mejor modelen la relación entre las entradas y las salidas siendo $\{R_k\}$ una serie de ruidos desconocidos. El modelo se identificará para minimizar la varianza de $\{R_k\}$.

Suponemos que las series $\{X_k\}$ y $\{Z_k\}$ son gaussianas y de varianza unidad y que corresponden a un proceso ergódico y estacionario.

Se supone que $\{R_k\}$ y $\{X_k\}$ son independientes. Esto es por la “no anticipación” del proceso. El ruido que “ataca” en el instante k sólo puede tener efectos desde ese instante en adelante.

La ecuación 15 puede escribirse aplicada sobre las series resultando en la ecuación 15 a continuación.

$$\{X_{k+1}\} = \sum_{h=0}^{h=N_r-1} A_h \cdot \{X_{k-h}\} + C \cdot \{Y_k\} + B \cdot \{R_k\}$$

ec.(15) Modelo aplicado a las series.

La ecuación 15 puede escribirse de forma de dejar explícita una matriz con las matrices A_h y C agrupadas como se muestra a continuación:

$$\{X_{k+1}\} = [A_0, A_1, \dots, A_{N_r-1}, C] \cdot \begin{bmatrix} \{X_{k-0}\} \\ \{X_{k-1}\} \\ \vdots \\ \{X_{k-(N_r-1)}\} \\ \{Y_k\} \end{bmatrix} + B \cdot \{R_k\}$$

ec.(16) Modelo matricial.

La solución de mínimos cuadrados que reduce el error de predicción $B \cdot \{R_k\}$ es la solución del sistema:

$$\{X_{k+1}\}^* \begin{bmatrix} \{X_{k-0}\} \\ \{X_{k-1}\} \\ \vdots \\ \{X_{k-(N_r-1)}\} \\ \{Z_k\} \end{bmatrix}^T = [A_0, A_1, \dots, A_{N_r-1}, C] \cdot \begin{bmatrix} \{X_{k-0}\} \\ \{X_{k-1}\} \\ \vdots \\ \{X_{k-(N_r-1)}\} \\ \{Z_k\} \end{bmatrix} \cdot \begin{bmatrix} \{X_{k-0}\} \\ \{X_{k-1}\} \\ \vdots \\ \{X_{k-(N_r-1)}\} \\ \{Z_k\} \end{bmatrix}^T$$

Que puede escribirse como: $[R_{xx}(1), R_{xx}(2), \dots, R_{xx}(N_r), R_{xz}(1)] = [A_0, A_1, \dots, A_{N_r-1}, C] \cdot M$

En dónde la matriz de correlaciones $R_{xx}(j)$ de retardo j , entre las series de vectores $\{Z_k\}$ y $\{X_k\}$ se define como:

$$R_{xx}(j) = \langle \{Z_k\} \{X_{k-j}\}^T \rangle \approx \frac{1}{N} \sum_{k=1}^{k=N} Z_k X_{k-j}^T$$

y

$$M = \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots & R_{xx}(N_r-1) & R_{xz}(0) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(N_r-2) & R_{xz}(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{xx}(N_r-1) & R_{xx}(N_r-2) & \cdots & R_{xx}(0) & R_{xz}(N_r-1) \\ R_{zx}(0) & R_{zx}(1) & \cdots & R_{zx}(N_r-1) & R_{zz}(0) \end{bmatrix}$$

Resolviendo

$$[A_0, A_1, \dots, A_{N_r-1}, C] = [R_{xx}(1), R_{xx}(2), \dots, R_{xx}(N_r), R_{xy}(1)] \cdot M^{-1}$$

Para calcular B multiplicamos ambos lados de la ecuación 16 por su transpuesto y tomamos el valor esperado obteniendo:

$$\langle \{X_{k+1}\} \{X_{k+1}\}^T \rangle = [A_0, A_1, \dots, A_{N_r-1}, C] M [A_0, A_1, \dots, A_{N_r-1}, C]^T + BB^T$$

En dónde se ha usado que $\langle \{R_k\} \{R_k\}^T \rangle = I$

Despejando BB^T se tiene:

$$BB^T = \langle \{X_{k+1}\} \{X_{k+1}\}^T \rangle - [A_0, A_1, \dots, A_{N_r-1}, C] M [A_0, A_1, \dots, A_{N_r-1}, C]^T \quad \text{ec.(17)}$$

Para el cálculo de B a partir del conocimiento de BB^T se puede utilizar la descomposición de Cholesky si BB^T es definida positiva. En el caso en que sea semi-definida positiva, es posible calcular B aplicando SVD (Singular Value Descomposición).

3.4. Incorporación de pronósticos a partir de sesgos y atenuadores.

La ecuación 15 representa la parte lineal en el espacio gaussiano de un sintetizador CEGH de SimSEE. Como ya se describió, ese modelo lineal es obtenido mediante un proceso de identificación. Este tipo de identificación captura el comportamiento estadístico de las series de entrada reduciendo el error de pronóstico cuando la única información que se utiliza para pronosticar es la contenida en el pasado de las propias series de datos. En las aplicaciones reales, se utiliza más información para realizar el pronóstico de la evolución de las variables de interés que aquella contenida en la realización histórica de las propias variables.

A modo de ejemplo, para el pronóstico de los aportes de los siguientes 10 días a las represas se utiliza información para evaluar las probabilidades de lluvias y sus volúmenes estimados en las cuencas basadas por ejemplo en imágenes satelitales información que no está en las series históricas. Generalmente, esta información adicional es útil en el corto plazo luego del cual el proceso se rige por la estadística de las series históricas.

Si nos imaginamos la salida de un proceso estocástico con dinámica como un cono del futuro posible, el conocimiento de información adicional tiene valor, si permite “afinar” el cono de la evolución futura de una variable reduciendo su apertura y posiblemente sesgando su guía. En la

fig.4 se muestra lo que podría ser el cono de las trayectorias de una variable $x(t)$ a partir de su valor conocido en el instante $t=t_0$ en el vértice del cono. En la figura se muestra un cono rojo que contendrá las realizaciones que generará el modelo CEGH si no se dispone información adicional para el pronóstico. En la misma figura en color naranja se muestra cómo podría verse modificado el cono que contiene las realizaciones futuras si se dispone de más información. La curva celeste dibujada en el centro del cono naranja es lo que llamamos “la guía del cono” y es la trayectoria de probabilidad de excedencia 50%. Es decir que dada una realización $x(t)$ sintetizada por el modelo CEGH, para cada tiempo t el valor de $x(t)$ tiene igual probabilidad de estar por encima que por debajo de la guía.

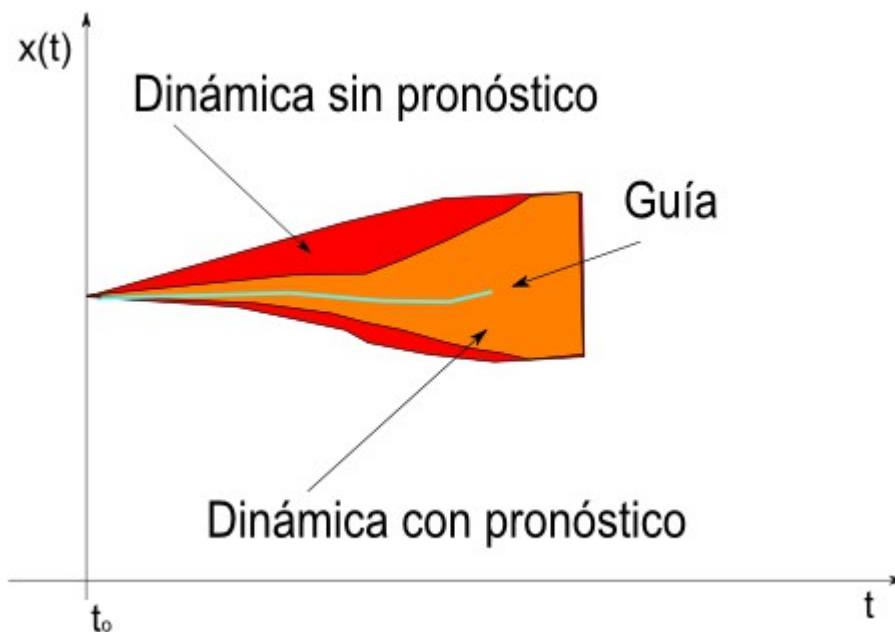


Fig. 4: Conos de dinámica con y sin pronósticos de corto plazo.

a) Implementación en los modelos CEGH.

Para la introducción de la información de los pronósticos se agregó en los modelos CEGH la posibilidad de especificar una serie de sesgos $\{S_k\}$ y de factores de apertura $\{P_k\}$ que actúan como se muestra en la ec. 18 a los efectos de cambiar el cono de la dinámica de acuerdo a la información disponible de corto plazo (el pronóstico).

$$X_{k+1} = \sum_{h=0}^{h=N_x-1} A_h \cdot X_{k-h} + C \cdot Y_k + S_k + P_k \cdot B \cdot R_k$$

ec.(18) Modelo de proceso con sesgos y factores de apertura del ruido.

Donde:

- el vector $\{X_k\}$ representa el estado de las salidas sintetizador en el espacio gaussiano.
- Las matrices A_h forman el filtro recursivo que aporta la información del pasado del

- estado de las salidas, que tienen influencia en la determinación de la siguiente salida.
- Las matrices C y B son las matrices que multiplican las entradas deterministas Y_k y las aleatorias R_k .
 - El estado del filtro lineal está dado por el conjunto de vectores $[X_k, X_{k-1}, \dots, X_{k-(N_r-1)}]$ que es necesario conocer, además de las entradas Y_k y R_k para poder calcular X_{k+1} .
 - El vector de entradas aleatorias R_k es un vector de ruidos blancos independientes con distribución normal. La transformación lineal B transforma el vector R_k en otro vector del mismo espacio que el del vector de estado X_k . Esta transformación tiene el efecto de “mezclar” los canales de ruido. Por ejemplo en el caso de los aportes a las represas capta comportamientos tales como el que si dos represas están cercanas, hay cierta probabilidad de que cuando llueve en la cuenca de una también suceda en la de la otra.
 - Las matrices P_k , son matrices diagonales con “factores de apertura” por canal. Multiplican las salidas de los canales de ruido (resultantes de la mezcla realizada por B) por un “factor de apertura” que permite regular el nivel de ruido y controlar por tanto la apertura del cono de pronósticos a la salida del sintetizador. Si esos valores se mantienen en CERO no se inyecta ruido y por lo tanto la salida del sintetizador es determinística. Cada señal de apertura permite ir de una situación determinística (apertura = 0) a la situación de máxima apertura (apertura=1) en cada canal de ruido en forma independiente.
 - La serie de vectores S_k , es la serie de vectores de sesgos que permite sesgar la salida del filtro lineal. Cambiando la guía del cono de cada canal.

La forma de introducir la información de los pronósticos es entonces suministrando la serie de vectores de sesgos $\{S_k\}$ y la serie de matrices de apertura $\{P_k\}$.

b) Determinación de los sesgos a partir del pronóstico.

En la práctica, dado un punto de partida de una simulación en el presente, existen “pronósticos” que permiten reducir la varianza en lo inmediato de lo que es esperable de un proceso estocástico respecto a lo que daría la salida del CEGH si nos limitáramos a fijar el estado inicial y simplemente simuláramos suponiendo $S_k=0$ y $P_k=I$ en la ec. 18.

La información de los pronósticos viene dada generalmente por una especificación del cono de la dinámica de las variables que es esperable se verifique en el futuro inmediato. Parte de ese cono puede coincidir con el cono de la dinámica “sin pronósticos”. La diferencia es en definitiva la información agregada por el pronóstico sobre la dinámica del proceso contenida en la información histórica de sus realizaciones.

Para fijar ideas, para la Programación Semanal (actividad que se realiza una vez por semana para programar el despacho de las diferentes centrales de generación de acuerdo a los recursos disponibles para la siguiente semana) es fundamental disponer de cierto pronóstico de cuales serán los caudales de aportes hidráulicos a las represas. En la actualidad (2012) existen medidas y modelos de previsión que brindan buenos pronósticos para los siguientes 7 días y con más error hasta unos 15 días. Esos pronósticos se expresan con curvas como las que se muestran en la fig.5 en la que se muestra el cono de aportes esperados para los siguientes veinte días en la central hidroeléctrica de Salto Grande. La guía del cono corresponde a la probabilidad 50% y los laterales inferior y superior a las probabilidades de excedencia de 95% y 5% respectivamente.

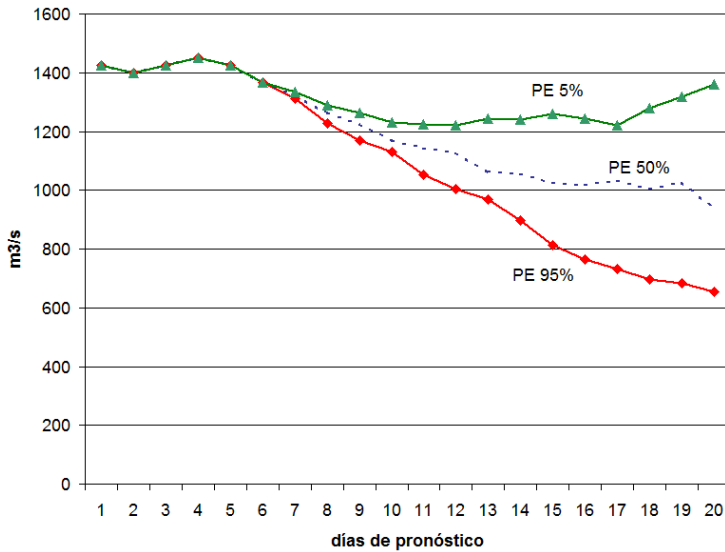


Fig. 5: Ejemplo de pronóstico de aportes a Salto Grande.

gaussiano cada componente del vector R_k tiene distribución normal por lo que el punto que deja 50% de valores de un lado y del otro es el CERO. Igualmente el vector BR_k es también un conjunto de fuentes aleatorias gaussianas de valor esperado nulo por ser combinación lineal de gaussianas de valor esperado nulo. En cada fila de la ecuación matricial ec.18 se está calculando el valor siguiente de las componentes del estado X a partir de los valores anteriores (estado del filtro) y de los valores determinísticos Y_k .

Es así que el aporte de BR_k puede verse como la componente ruidosa alrededor de la proyección determinística del estado. Si anulamos esa componente ruidosa, como es simétrica se tiene que la proyección determinística del estado corresponde a la trayectoria de 50% de probabilidad.

Si partimos de $k=0$ como punto vértice del cono, tendremos que en ese punto el estado es conocido así como las entradas determinísticas Y_0 y podemos calcular por tanto la proyección determinística del estado en $k=1$ como se muestra en la ec.19. Dónde el valor S_0 debe ser calculado para que el vector X_1 coincida con el valor de las guías de probabilidad 50% \mathcal{Y}_1

$$X_1 = \sum_{h=0}^{N_r-1} A_h \cdot X_{0-h} + C \cdot Y_0 + S_0 \quad \text{ec.(19) Proyección determinística del primer paso.}$$

Despejando se tiene: $S_0 = X_1 - \sum_{h=0}^{N_r-1} A_h \cdot X_{0-h} + C \cdot Y_0$ y aplicando el mismo razonamiento para el paso k se puede escribir:

$$S_k = X_{k+1} - \sum_{h=0}^{N_r-1} A_h \cdot X_{k-h} + C \cdot Y_k \quad \text{ec.(20) Cálculo de los vectores de sesgos para seguir la guía.}$$

En el editor de SimSEE el formulario que permite editar los parámetros de las instancias de modelos CEGH tiene la facilidad de introducción de pronósticos. La fig. 6 muestra el panel para ingreso de pronósticos del formulario de edición de parámetros de una instancia CEGH en SimSEE para un ejemplo en el que el CEGH tiene tres “canales” (o salidas) correspondientes a los aportes medios semanales de caudales a las represas de Bonete, Palmar y Salto. El formulario tiene la

Dada la guía de probabilidad 50% en el espacio real, mostraremos ahora como a partir de la ec.18 es posible obtener el los sesgos $\{S_k\}$ a aplicar en el espacio gaussiano.

El primer paso es pasar la GUÍA del espacio real al espacio gaussiano. Para eso basta con utilizar las transformaciones no lineales del modelo CEGH. llamemos \mathcal{Y}_k a la GUÍA de probabilidad 50% en el espacio gaussiano. Ahora hay que calcular los S_k para que $prob(X_k > \mathcal{Y}_k) = 0.5$

Observar que en el espacio

posibilidad de incluir un pronóstico determinístico (Columna “valores iniciales y guía del pronóstico en la fig. 6), especificando los valores en el espacio real de la guía del cono. También se puede seleccionar el Número de Pasos de Control del Cono (NPCC) en que esos valores determinísticos deben ser considerados para la generación de sesgos, el Número de Pasos de Liberación del Cono (NPLC) en que el sesgo pasará del valor en que haya quedado en el último paso de control a cero. Para controlar la apertura del cono se incluyeron dos parámetros, Número de Pasos Sin Apertura (NPSA) y Número de Pasos de Apertura el Cono (NPAC). El parámetro NPSA corresponde al número de pasos desde el inicio en que se considerará que la guía es determinística y por tanto los factores de apertura del cono son nulos. Transcurridos los NPSA pasos con factores de apertura nulos, el parámetro NPAC regula la cantidad de pasos en que los factores de apertura pasarán de 0 (Cero) a 1.

Valores Iniciales para Simulación (por paso de sorteo) y conos de PRONOSTICOS:

Borne	Valores iniciales y guía del pronóstico.	NPCC	NPLC	NPSA	NPAC	pe[p.u.]	NRet.
Bonete	100.0; 80.0; 60.0; 50.0	3	4	3	4	0.5	1
Palmar	59.0; 40.0; 30.0; 10.0	2	4	3	4	0.5	1
Salto	3500.0; 3150.0; 2900.0; 2500.0; 2400.0; 2300.0; 2200.0	6	1	6	1	0.5	1

Calibrar Cono

Número de pasos a graficar

Fig. 6: Panel de introducción de pronósticos en modelo CEGH de SimSEE.

La misma interface permite generar los sesgos y los factores de aperturas a partir de los valores iniciales, la guía de pronósticos los valores NPCC, NPAC, NPSA y NPLC de cada canal mediante la aplicación de la ec. 20 y las definiciones antes dadas a los parámetros.

Presionar el botón “Calibrar Cono”, se realizan los cálculos y como verificación se realiza una simulación de 100 realizaciones y se grafica para cada canal un gráfico como el que se muestra en la fig. 6 para el canal Palmar. En la gráfica se puede apreciar, el vértice del cono (a la izquierda) partiendo del valor inicial $59.0 \text{ m}^3/\text{s}$ seguido por tres valores sin dispersión (dado que $NPSA=3$).

De estos tres valores, los dos primeros son $40.0 \text{ m}^3/\text{s}$ y $30.0 \text{ m}^3/\text{s}$ obligados por ser los valores especificados para la guía y el parámetro $NPCC=2$. Del quinto valor en adelante (es decir pasado el primero que es el valor inicial y los siguientes 3 sin apertura del cono) comienza a aparecer el ruido y por eso el cono de pronóstico se abre. En la figura se muestran en el centro del cono la trayectoria de probabilidad 50% y la trayectoria del valor esperado. La inferior corresponde a aquella que tiene valores inferiores con probabilidad 5% y la curva superior corresponde a la que tiene valores superiores con probabilidad 5%. Entre la curva inferior y superior definen “el cono” de probabilidad 90%.

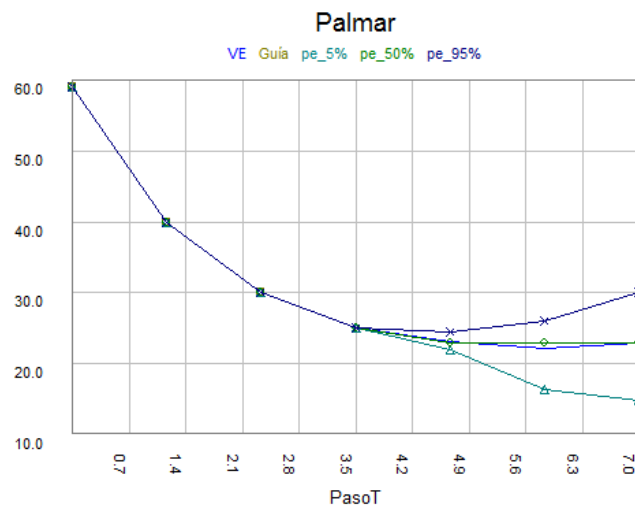


Fig. 7: Ejemplo de pronóstico de aportes a Palmar

4. Reducción de la dimensión del estado de modelos CEGH en la Programación Dinámica Estocástica.

Ruben Chaer – Julio 2013

4.1. Introducción.

Los modelos CEGH (Correlaciones en Espacio Gaussiano con Histograma) son utilizados intensivamente en la plataforma SimSEE para el modelado de los procesos estocásticos.

En pocas palabras, el modelado CEGH implica primero identificar el modelo a partir de un conjunto de series históricas de valores del vector de variables de interés. Por ejemplo, si el proceso a modelar son los aportes a las represas de Bonete, Palmar y Salto Grande, el vector de valores tendrá dimensión 3 y la serie histórica de ese vector podrá ser por ejemplo los aportes medios semanales expresados en m^3/s a las tres represas. Sea $\{Y_k\}$ la serie de vectores que representan los valores históricos. En el modelado CEGH, se identifica una transformación del espacio de los valores reales a un espacio transformado en que las series “se parecen” a series gaussianas en estado estacionario. A los efectos de introducir la notación, identifiquemos esta transformación como:

$$x_k = TNL(y_k, k) \quad \text{ec.(21) Transformación al espacio gaussiano.}$$

$$x_k, y_k \in R^n$$

A partir de las series $\{x_k\}$ (que se supone representan un proceso gaussiano conjunto y que cada componente del vector por separado es una gaussiana con valor esperado nulo y varianza unidad) se identifica un filtro lineal de forma que dicho filtro, cuando es alimentado por fuentes de ruido blanco gaussiano genera en su salida un vector que reproduce las matrices de covarianza $\langle x_k x_k^t \rangle, \langle x_k x_{k-1}^t \rangle, \dots, \langle x_k x_{k-n_r}^t \rangle$.

La estructura del filtro lineal identificado puede ser cualquiera. En particular, la herramienta AnalisisSerial [???] que forma parte de la plataforma SimSEE identifica un filtro lineal recursivo puro con el formato:

$$x_{k+1} = \sum_{h=0}^{n_r-1} A_{k-h} \cdot x_{k-h} + B \cdot r_k \quad \text{ec.(22) Estructura del filtro lineal identificado por la herramienta AnalisisSerial.}$$

Al reproducir las matrices de covarianza para diferentes retardos $0, 1, \dots, n_r$ el modelo así identificado tiene para el conjunto de variables aleatorias dado por las componentes del vector

$$z = \begin{bmatrix} x_k \\ x_{k-1} \\ \dots \\ x_{k-n_r} \end{bmatrix} \quad \text{cuya matriz de covarianzas se puede calcular como:}$$

$$\Sigma = \begin{bmatrix} \langle x_k x_k^t \rangle & \langle x_k x_{k-1}^t \rangle & \dots & \langle x_k x_{k-n_r}^t \rangle \\ \langle x_{k-1} x_k^t \rangle & \langle x_{k-1} x_{k-1}^t \rangle & \dots & \langle x_{k-1} x_{k-n_r}^t \rangle \\ \dots & \dots & \dots & \dots \\ \langle x_{k-n_r} x_k^t \rangle & \langle x_{k-n_r} x_{k-1}^t \rangle & \dots & \langle x_{k-n_r} x_{k-n_r}^t \rangle \end{bmatrix}$$

ec.(23) Matriz de covarianzas del proceso conjunto de los vectores

$$z = \begin{bmatrix} x_k \\ x_{k-1} \\ \dots \\ x_{k-n_r} \end{bmatrix}$$

Dada la matriz de covarianza de un proceso gaussiano conjunto, la función de densidad de probabilidad se puede calcular como:

$$p_z(z) = e^{-\frac{1}{2}(z \Sigma^{-1} z)}$$

ec.(24) Función densidad de probabilidad del proceso conjunto Z

Observar que bajo el supuesto de que existe la TNL de la ec. 21 el sistema lineal identificado en el espacio gaussiano, puede teóricamente sintetizar series que tienen la misma función densidad de probabilidad que el proceso usado para la identificación. Por lo tanto, las series sintetizadas cumplen cualquier medida de probabilidad que se realice con la integral sobre el espacio del proceso usando como ponderador la $p_z(z)$. Como la TNL de la ec. 21 es invertible y transforma todo volumen compacto del espacio original en un volumen compacto del espacio gaussiano; es posible calcular la función de densidad de probabilidad del proceso en el espacio original a partir de $p_z(z)$ multiplicando dicha función por el jacobiano de TNL^{-1} .

Hasta aquí se realizó un resumen de los fundamentos de los modelos CEGH. A los efectos de la consideración de un modelo CEGH en un algoritmo de Programación Dinámica Estocástica (PDE), es necesario representar la función de Costo Futuro (CF) (también conocida como función de Bellman) definida de el espacio de estados del sistema incluyendo los procesos estocásticos en los reales. El CEGH con filtro lineal como el de 22 tiene un estado de dimensión igual a la dimensión de x por la cantidad de retardos n_r que se consideren. Como es sabido los algoritmos de PDE sufren de “La Maldición de la Dimensionalidad de Bellman” [??] que básicamente se trata de la explosión combinatoria que implica el producto cartesiano de las discretizaciones que se consideren sobre cada dimensión del espacio de estado. Para fijar ideas, si se tiene un sistema con un vector de estado en $x \in R^5$ y para representar la función $CF(x): R^5 \rightarrow R$ se fijan los rangos de variación de cada componente de x y se discretiza cada rango en 10 puntos, se tendrá que el total de puntos sobre los que hay que calcular $CF(x)$ es 10^5 . Por este motivo, para poder llevar a cabo la PDE, en muchas ocasiones se recurre a simplificar la representación del estado del sistema. Claro está que dicha simplificación siempre significa pérdida de información y por lo tanto el resultado será una función $CF(x)$ con menos información. Dado que el gradiente de $\nabla_x CF(x)$ contiene la información de la Política de Operación Óptima del sistema, la pérdida de información se traduce en la obtención de una política sub-óptima.

En este trabajo, se analiza como impacta sobre los modelos CEGH optar por reducir la dimensión del espacio de estado del filtro lineal y cómo considerar en la PDE dicha pérdida de información.

4.2. Reducción del espacio de estados en el espacio gaussiano de un modelo CEGH.

El filtro lineal de la ec. 22 puede escribirse como un filtro recursivo de orden 1 como se muestra en la ec. 25

$$z_{k+1} = \bar{A} \cdot z_k + \bar{B} \cdot r_k \quad \text{ec.(25) Sistema lineal de orden 1.}$$

Donde las matrices \bar{A} y \bar{B} son calculables fácilmente a partir de las matrices A_h y B de la ec. 22.

Entonces, sin pérdida de generalidad es posible trabajar sobre la estructura de filtro lineal de la ec. 25.

Para la representación del filtro lineal en el algoritmo de Programación Dinámica Estocástica, se decide utilizar una “reducción” del espacio de estado dada por una transformación $y = M_r z$. Por simplicidad supondremos que las filas de esa matriz tiene filas ortogonales. Si $z \in R^n$ y $y \in R^m$ la operación de reducción implica que $m < n$.

En principio, si se quiere tener una transformación de “amplificación” M_a se tendría que cumplir que si se aplica la amplificación sobre un punto $y \in R^m$ se obtenga un punto $z \in R^n$ tal que si luego se aplica la transformación reductora M_r sobre z se vuelva a obtener el punto y . Por lo tanto se debe cumplir la relación de la ec. 26.

$$M_r \cdot M_a = I \quad \text{ec.(26) La cadena de Amplificación y Reducción tiene que ser la identidad.}$$

Si se supone que las filas de M_r son ortonormales, $M_a = M_r^t$ es solución de la ec. 26.

Por otro lado, por facilidad de la representación de las probabilidades en el espacio de estado, es deseable que las componentes del vector de estado reducido tengan varianza unidad (esto es $\langle (y^i)^2 \rangle = 1; i = 1 \dots m$, siendo y^i la componente i -ésima del vector y). Esto simplifica la especificación de los estados como “bandas de probabilidad”. Para lograr este objetivo, hay que escalar las filas de M_r para que la matriz de covarianzas $\langle y \cdot y^t \rangle$ tenga unos en la diagonal.

$$\langle y \cdot y^t \rangle = M_r \langle z \cdot z^t \rangle M_r^t \quad \text{ec.(27) Matriz de covarianzas del vector de espacio reducidos.}$$

Para lograr cumplir con la ec. 26 y que la matriz $\langle y \cdot y^t \rangle$ definida como en la ec. 27 tenga 1s (unos) en su diagonal, dada una Propuesta de Transformación Reductora \check{M}_r de filas ortonormales y la matriz de covarianzas $\langle z \cdot z^t \rangle$ procedemos de la siguiente forma:

- 1) Calculamos $\langle y \cdot y^t \rangle = \check{M}_r \langle z \cdot z^t \rangle \check{M}_r^t$ y llamamos $\zeta_1^2, \zeta_2^2, \dots, \zeta_m^2$ los elementos de su diagonal.
- 2) Calculamos M_r dividiendo las filas de \check{M}_r por los valores $\zeta_1, \zeta_2, \dots, \zeta_m$ como se expresa en la ec. 28.
- 3) Calculamos una Propuesta de Transformación Amplificadora como $\check{M}_a = \check{M}_r^t$ que por ser \check{M}_r ortonormal se asegura el cumplimiento de la ec. 26.
- 4) Por último, para mantener el cumplimiento de la ec. 26 calculamos M_a escalando las

columnas de \check{M}_a con los multiplicadores $\zeta_1, \zeta_2, \dots, \zeta_m$ como se muestra en la ec. 29

$$M_r = \begin{bmatrix} \frac{1}{\zeta_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\zeta_2} & \dots & 0 \\ 0 & 0 & \dots & \frac{1}{\zeta_m} \end{bmatrix} \cdot \check{M}_r$$

ec.(28) Escalamiento de la propuesta de transformación reductora para asegurar varianza unitaria de las componentes del vector de estado reducido.

$$M_a = \check{M}_a \cdot \begin{bmatrix} \zeta_1 & 0 & \dots & 0 \\ 0 & \zeta_2 & \dots & 0 \\ 0 & 0 & \dots & \zeta_m \end{bmatrix}$$

ec.(29) Escalamiento de la propuesta de transformación aplicadora para mantener el equilibrio de la ec. 26

4.3. Transformación Amplificadora en el espacio gaussiano de un modelo CEGH.

En la sección 4.2 se mostró las características deseables que debieran tener las transformaciones de reducción de la dimensión del estado M_r y de amplificación de la dimensión del estado M_a .

La reducción del espacio de estado, es usado durante el algoritmo de Programación Dinámica Estocástica (PDE) para reducir la dimensión del espacio de estado (espacio sobre el que se calcula la función de Costo Futuro (CF)) como forma de disminuir el impacto de la Maldición de la Dimensionalidad de Bellman. En la ejecución del algoritmo de PDE, se realiza un recorrido de las etapas temporales (pasos de tiempo) en reversa en el tiempo (desde el último paso de tiempo al primero) y para cada paso de tiempo se realiza un “barrido” de todos los puntos de la discretización del espacio de estado sobre el que se quiere calcular la función CF. Es precisamente ese conjunto de puntos que se obtiene por el producto cartesiano de las discretizaciones sobre cada dimensión del espacio de estado que produce una de las explosiones combinatorias que da lugar a la Maldición de Bellman.

Al trabajar con un estado reducido $y \in R^m$ en un espacio de menor dimensión que el original, se estará “perdiendo información” respecto al estado real del sistema. Hay un conjunto de puntos del espacio sin reducir que transformados por M_r dan el mismo y por lo tanto dado un y es imprecisa la posición del sistema en el espacio original.

Un problema de optimización adicional es determinar la transformación M_r que lleva a la menor pérdida de información relevante para el operador. Este problema de optimización NO es objeto de este trabajo. En este trabajo se supone que la transformación reductora es dada y a lo sumo se recomienda escalar las filas de la matriz para que cada variable por separado tenga varianza unidad en el espacio reducido con el solo propósito de facilitar la descripción de las variables.

Durante la resolución del algoritmo de PDE, al calcular el mínimo entre el costo de la etapa, partiendo de un estado reducido y dado, más el costo futuro del estado y' alcanzado al final

de la etapa, es necesario para poder utilizar el modelo del sistema (ecuaciones que representan las restricciones incluyendo la dinámica) asignar un valor del estado no reducido del sistema. Dado que hay muchos valores posibles de ese estado que corresponden al estado reducido y , se tiene que la reducción de estado, agrega al algoritmo una incertidumbre sobre el estado. En el caso de la implementación propuesta para SimSEE, esta incertidumbre se refleja mediante la suma de ruido sobre el valor $z = M_a \cdot y$ en la forma que muestra la ec.

$$z = M_a \cdot y + B_a \cdot w_a$$

ec.(30) Inyección de Ruido en la transformación de amplificación de la dimensión del estado.

El principal objetivo de este trabajo es el cálculo de la matriz B_a de la ec. 30 de forma tal que se logre el efecto de reproducir la matriz de covarianzas $\langle z \cdot z' \rangle / y$ del vector de estado conocido el valor del vector de estado reducido y .

Si no se conociera el vector de estado reducido y el vector de estado z tiene una matriz de covarianzas $\langle z \cdot z' \rangle$ conocida que se puede calcular a partir de las matrices del filtro lineal 25.

Expresando la covarianza de ambos lado de la ec. 25 podemos escribir:

$$\langle z_{k+1} \cdot z_{k+1}' \rangle = \bar{A} \cdot \langle z_k \cdot z_k' \rangle \cdot \bar{A}' + \bar{B} \cdot \langle r_k \cdot r_k' \rangle \cdot \bar{B}'$$

ec.(31) Calculando la matriz de covarianzas del estado z

Teniendo en cuenta que el proceso es ergódico, se puede escribir:

$\langle z_{k+1} \cdot z_{k+1}' \rangle = \langle z_k \cdot z_k' \rangle = \langle z \cdot z' \rangle$. Además, por construcción r_k es una serie de vectores de ruidos blancos independientes por lo que $\langle r_k \cdot r_k' \rangle = I$. Utilizando estas relaciones en la ec. 31 se obtiene la ec. 32.

$$\langle z \cdot z' \rangle = \bar{A} \cdot \langle z \cdot z' \rangle \cdot \bar{A}' + \bar{B} \cdot \bar{B}'$$

ec.(32) Sistema que determina la matriz de covarianzas de z en función de las matrices del filtro.

Para resolver el sistema de la ec. 32 se observa que es un sistema lineal en $\langle z \cdot z' \rangle$ y que se puede expresar como un sistema lineal invertible vectorizando (creando vectores con las columnas concatenadas de las matrices) ambos lados de la expresión y teniendo en cuenta que el vectorizado del producto de tres matrices cumple: $\text{vect}(A \cdot X \cdot B) = \text{Kron}(B', A) \cdot \text{vec}(X)$ dónde la operación “vec” corresponde a crear un vector a partir de una matriz concatenando sus columnas y la operación “Kron” es el Producto de Kronecker. (ver. Petersen & Pedersen, The Matrix Cookbook, Version: November 14, 2008, Page 60).

$$\langle z \cdot z' \rangle = \text{reshape}((I - \text{Kron}(\bar{A}, \bar{A}))^{-1} \cdot \text{vec}(\bar{B} \cdot \bar{B}')),$$

dónde la operación “reshape” es la inversa de la operación “vec”.

ec.(33) Matriz de covarianzas del estado del filtro lineal.

Si hubiera que adivinar la posición del sistema de la ec. 25 sin ninguna información disponible, se podría representar el estado como: $z = H \cdot w$ siendo H cualquier matriz que cumpla $H \cdot H' = \langle z \cdot z' \rangle$ y w un vector de ruidos blancos gaussianos normales.

Observar que dada cualquier matriz ortonormal F , se cumple $H \cdot F \cdot F' \cdot H' = \langle z, z' \rangle$.

Utilizando esta representación del vector de estado y la descomposición de la ec. 30 se puede escribir:

$$z = M_a \cdot y + B_a \cdot w_a = H \cdot F \cdot w \quad \text{ec.(34)}$$

Multiplicando la ec. 34 por la matriz reductora M_r se puede escribir:

$$M_r \cdot z = M_r \cdot M_a \cdot y + M_r \cdot B_a \cdot w_a = M_r \cdot H \cdot F \cdot w \quad \text{ec.(35)}$$

Como $M_r \cdot z = y$ y $M_r \cdot M_a = I$ en la ec. 35 se deberá cumplir que $M_r \cdot B_a \cdot w_a = 0$ y $y = M_r \cdot H \cdot F \cdot w$.

Para lograr calcular B_a consistente con estas igualdades, se observa que las columnas de la matriz $M_r \cdot H$ expanden el espacio R^m del estado reducido y . Como la matriz $M_r \cdot H$ tiene por construcción m filas y n columnas con $m < n$, la transformación $M_r \cdot H$ tiene rango m y el núcleo de la transformación $M_r \cdot H$ tiene dimensión $n - m$.

Si identificamos una base ortonormal F_a del núcleo de $M_r \cdot H$ y una base ortonormal F_b del subespacio ortogonal al núcleo, y construimos la matriz F cuyas primeras $n - m$ columnas son las de F_a y las siguientes m columnas son las de F_b y separamos el vector w de la ec. 35 en el vector w_a con las primeras $n - m$ componentes de w y el vector w_b con las siguientes m componentes de w , como la matriz F así construida es ortonormal, se puede reescribir la ec. 35 como:

$$M_r \cdot z = M_r \cdot M_a \cdot y + M_r \cdot B_a \cdot w_a = M_r \cdot H \cdot [F_a, F_b] \cdot \begin{bmatrix} w_a \\ w_b \end{bmatrix} \quad \text{ec.(36)}$$

Y operando se puede escribir:

$$M_r \cdot z = M_r \cdot M_a \cdot y + M_r \cdot B_a \cdot w_a = M_r \cdot H \cdot F_a \cdot w_a + M_r \cdot H \cdot F_b \cdot w_b \quad \text{ec.(37)}$$

La ec. 37 se puede descomponer en las ec. 38 y ec. 39

$$y = M_r \cdot H \cdot F_b \cdot w_b \quad \text{ec.(38)}$$

$$0 = M_r \cdot B_a \cdot w_a = M_r \cdot H \cdot F_a \cdot w_a \quad \text{ec.(39)}$$

De la ec. 39 se obtiene una matriz $B_a = H \cdot F_a$ que es compatible con representar el estado por el modelo de la ec. 34.

