Estimación de parámetros estocásticos en modelos

Germán Ariel Torres
(torres@famaf.unc.edu.ar)

Facultad de Ciencias Exactas y Naturales y Agrimensura (FACENA)
Universidad Nacional del Nordeste
IMIT (CONICET)

Objetivos

Objetivos

- Se desea estimar parámetros estocásticos en modelos

Objetivos

- Se desea estimar parámetros estocásticos en modelos
- Al mismo tiempo se desea estimar el estado de un sistema con los parámetros mejorados

Objetivos

- Se desea estimar parámetros estocásticos en modelos
- Al mismo tiempo se desea estimar el estado de un sistema con los parámetros mejorados
- Estos parámetros suelen aparecer especialmente en modelos meteorológicos que utilizan parametrizaciones para representar procesos físicos a una escala pequeña, pero afectan la gran escala

Algunos ejemplos

Algunos ejemplos

- Parámetros que aparecen en parametrizaciones físicas en modelos meteorológicos

Algunos ejemplos

- Parámetros que aparecen en parametrizaciones físicas en modelos meteorológicos
- Parámetros que aparecen en modelos de crecimiento de tumores (por ejemplo efectividad de una droga)

Algunos ejemplos

- Parámetros que aparecen en parametrizaciones físicas en modelos meteorológicos
- Parámetros que aparecen en modelos de crecimiento de tumores (por ejemplo efectividad de una droga)
- Estimación de ERD (event related desynchronization). Es un parámetro para describir el aumento o descenso en la actividad de la señal de un EEG.

Algunos ejemplos

- Parámetros que aparecen en parametrizaciones físicas en modelos meteorológicos
- Parámetros que aparecen en modelos de crecimiento de tumores (por ejemplo efectividad de una droga)
- Estimación de ERD (event related desynchronization). Es un parámetro para describir el aumento o descenso en la actividad de la señal de un EEG.
- Error del modelo: el error del modelo es la diferencia entre el resultado del modelo y la realidad. Es desconocido y se asume que tiene una distribución gaussiana. ¡Esto no siempre es cierto!

Algunos ejemplos

- Parámetros que aparecen en parametrizaciones físicas en modelos meteorológicos
- Parámetros que aparecen en modelos de crecimiento de tumores (por ejemplo efectividad de una droga)
- Estimación de ERD (event related desynchronization). Es un parámetro para describir el aumento o descenso en la actividad de la señal de un EEG.
- Error del modelo: el error del modelo es la diferencia entre el resultado del modelo y la realidad. Es desconocido y se asume que tiene una distribución gaussiana. ¡Esto no siempre es cierto!
- Necesitamos la matriz de covarianza del error del modelo.

Técnicas actuales

Técnicas actuales

- Se suele representar el estado aumentado del sistema

$$\begin{pmatrix} x \\ p \end{pmatrix}$$

El estado evoluciona en el tiempo con un modelo de predicción, y se define una evolución para los parámetros.

Técnicas actuales

- Se suele representar el estado aumentado del sistema

$$\begin{pmatrix} x \\ p \end{pmatrix}$$

  El estado evoluciona en el tiempo con un modelo de predicción, y se define una evolución para los parámetros.
- Se aplica un filtro de Kalman a este estado aumentado y se estiman los parámetros
  [Anderson, 2001, Zupanski and Zupanski, 2006]

Técnicas actuales

- Se suele representar el estado aumentado del sistema

$$\begin{pmatrix} x \\ p \end{pmatrix}$$

  El estado evoluciona en el tiempo con un modelo de predicción, y se define una evolución para los parámetros.
- Se aplica un filtro de Kalman a este estado aumentado y se estiman los parámetros
  [Anderson, 2001, Zupanski and Zupanski, 2006]
- El filtro de Kalman es una de las técnicas de asimilación de datos

Técnicas actuales

- Se suele representar el estado aumentado del sistema

$$\begin{pmatrix} x \\ p \end{pmatrix}$$

  El estado evoluciona en el tiempo con un modelo de predicción, y se define una evolución para los parámetros.
- Se aplica un filtro de Kalman a este estado aumentado y se estiman los parámetros
  [Anderson, 2001, Zupanski and Zupanski, 2006]
- El filtro de Kalman es una de las técnicas de asimilación de datos
- También están las técnicas variacionales (ECMWF,NCEP)

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.
- Los ingredientes son

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.
- Los ingredientes son
  - el estado del sistema (no observable)

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.
- Los ingredientes son
  - el estado del sistema (no observable)
  - observaciones del estado del sistema (observable)

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.
- Los ingredientes son
  - el estado del sistema (no observable)
  - observaciones del estado del sistema (observable)
  - un modelo de propagación

Filtro de Kalman

- En el artículo [Kalman, 1960] Rudolph Kalman presentó una técnica que permite calcular el estado oculto (o no medible) de un sistema.
- Los ingredientes son
  - el estado del sistema (no observable)
  - observaciones del estado del sistema (observable)
  - un modelo de propagación
  - un operador de observación

Formulación

El estado del sistema se representa de la siguiente manera:

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

donde

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

donde

- A es una matriz que hace evolucionar el estado

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

donde

- A es una matriz que hace evolucionar el estado
- H_t es un operador de observación

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

donde

- A es una matriz que hace evolucionar el estado
- H_t es un operador de observación
- $v_t \sim \mathcal{N}(0, R)$

Formulación

El estado del sistema se representa de la siguiente manera:

$$\begin{aligned} y_t &= H_t x_t + v_t, \quad \text{measurement equation,} \\ x_{t+1} &= A x_t + w_t, \quad \text{state equation.} \end{aligned}$$

donde

- A es una matriz que hace evolucionar el estado
- H_t es un operador de observación
- $v_t \sim \mathcal{N}(0, R)$
- $w_t \sim \mathcal{N}(0, Q)$

Algoritmo

Llamemos

$$\begin{aligned} \hat{x}_{t|s} &= E\left(x_t | Y_{1:s}\right), \\ P_{t|s} &= E\left(\left(x_t - \hat{x}_{t|s}\right)\left(x_t - \hat{x}_{t|s}\right)^T | Y_{1:s}\right). \end{aligned}$$

Algoritmo

Llamemos

$$\begin{aligned} \hat{x}_{t|s} &= E\left(x_t | Y_{1:s}\right), \\ P_{t|s} &= E\left(\left(x_t - \hat{x}_{t|s}\right)\left(x_t - \hat{x}_{t|s}\right)^T | Y_{1:s}\right). \end{aligned}$$

La estimación del estado $\left(\hat{x}_{t|t}, P_{t|t}\right)$ se puede obtener de manera recursiva por el siguiente algoritmo.

Algoritmo

Algoritmo

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1},$$

Algoritmo

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado}$$

Algoritmo

$$\begin{aligned} \hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\ P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \end{aligned}$$

Algoritmo

$$\begin{aligned} \hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\ P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \end{aligned}$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t \left(HP_{t|t-1}H^T + R\right)^{-1},
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia}
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia} \\
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\left(y_t - H_t\hat{x}_{t|t-1}\right),
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia} \\
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\left(y_t - H_t\hat{x}_{t|t-1}\right), \quad \text{análisis}
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia} \\
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\left(y_t - H_t\hat{x}_{t|t-1}\right), \quad \text{análisis} \\
P_{t|t} &= \left(I - KH_t\right)P_{t|t-1},
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia} \\
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\left(y_t - H_t\hat{x}_{t|t-1}\right), \quad \text{análisis} \\
P_{t|t} &= \left(I - KH_t\right)P_{t|t-1}, \quad \text{matriz de covarianza del análisis}
\end{aligned}
$$

Algoritmo

$$
\begin{aligned}
\hat{x}_{t|t-1} &= A\hat{x}_{t-1|t-1}, \quad \text{forecast del estado} \\
P_{t|t-1} &= AP_{t-1|t-1}A^T + Q, \quad \text{matriz de covarianza del forecast} \\
K_t &= P_{t|t-1}H_t\left(HP_{t|t-1}H^T + R\right)^{-1}, \quad \text{matriz de ganancia} \\
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\left(y_t - H_t\hat{x}_{t|t-1}\right), \quad \text{análisis} \\
P_{t|t} &= \left(I - KH_t\right)P_{t|t-1}, \quad \text{matriz de covarianza del análisis}
\end{aligned}
$$

donde $x_{1|0} = \mu_0$ y $P_{1|0} = \Sigma_0$.

Características del filtro de Kalman

Características del filtro de Kalman

- Fue diseñado sólo para modelos lineales

Características del filtro de Kalman

- Fue diseñado sólo para modelos lineales
- La hipótesis de gaussianidad puede no ser cierta en algunos modelos

Características del filtro de Kalman

- Fue diseñado sólo para modelos lineales
- La hipótesis de gaussianidad puede no ser cierta en algunos modelos
- Cuando el vector de estado es muy grande el filtro es inaplicable

Características del filtro de Kalman

- Fue diseñado sólo para modelos lineales
- La hipótesis de gaussianidad puede no ser cierta en algunos modelos
- Cuando el vector de estado es muy grande el filtro es inaplicable
- El filtro minimiza la traza de la matriz de covarianza del análisis

Características del filtro de Kalman

- Fue diseñado sólo para modelos lineales
- La hipótesis de gaussianidad puede no ser cierta en algunos modelos
- Cuando el vector de estado es muy grande el filtro es inaplicable
- El filtro minimiza la traza de la matriz de covarianza del análisis
- Se corre on-line

Modificaciones

Modificaciones

- Se puede generalizar el filtro de Kalman cuando el modelo es casi lineal

Modificaciones

- Se puede generalizar el filtro de Kalman cuando el modelo es casi lineal
- En este caso debe conocerse la derivada del modelo (el modelo tangente) y del operador de observación

Modificaciones

- Se puede generalizar el filtro de Kalman cuando el modelo es casi lineal
- En este caso debe conocerse la derivada del modelo (el modelo tangente) y del operador de observación
- Funciona cuando el modelo no es tan no lineal

Modificaciones

- Se puede generalizar el filtro de Kalman cuando el modelo es casi lineal
- En este caso debe conocerse la derivada del modelo (el modelo tangente) y del operador de observación
- Funciona cuando el modelo no es tan no lineal
- Las matrices de covarianza suelen recortarse a través de alguna factorización

Modificaciones

Modificaciones

- En lugar de usar la matriz completa se usa un conjunto de estados (llamado ensemble) [Evensen, 1997]

Modificaciones

- En lugar de usar la matriz completa se usa un conjunto de estados (llamado ensemble) [Evensen, 1997]
- Esta formulación es útil cuando el modelo es no lineal!

Modificaciones

- En lugar de usar la matriz completa se usa un conjunto de estados (llamado ensemble) [Evensen, 1997]
- Esta formulación es útil cuando el modelo es no lineal!
- Hay varias familias de métodos que tratan de simplificar el planteo original: EnSKF, RRSQRT, LETKF, etc.

Modificaciones

- En lugar de usar la matriz completa se usa un conjunto de estados (llamado ensemble) [Evensen, 1997]
- Esta formulación es útil cuando el modelo es no lineal!
- Hay varias familias de métodos que tratan de simplificar el planteo original: EnSKF, RRSQRT, LETKF, etc.
- En [Torres, 2010] se presentó un paquete de subrutinas que toma los aspectos comunes de la asimilación de datos y se independiza un poco más del modelo incluyendo métodos económicos y paralelización

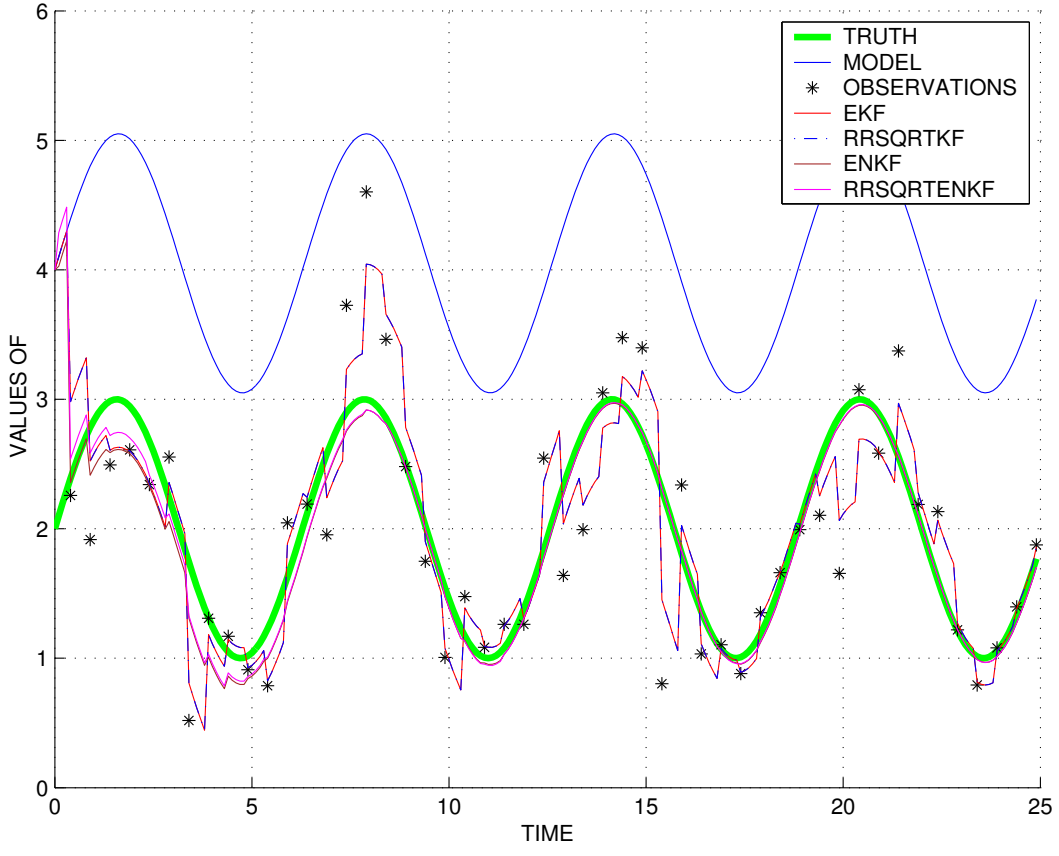
Modificaciones

Modificaciones

- El filtro rompe el balance físico de las variables y “tarda” en recuperarlo

Modificaciones

- El filtro rompe el balance físico de las variables y “tarda” en recuperarlo



Smoother

Smoother

Suponiendo que tenemos observaciones futuras disponibles, podemos utilizarlas para mejorar la estimación.

Smoother

Suponiendo que tenemos observaciones futuras disponibles, podemos utilizarlas para mejorar la estimación.
El algoritmo es:

Smoother

Suponiendo que tenemos observaciones futuras disponibles, podemos utilizarlas para mejorar la estimación.
El algoritmo es:

$$J_t = P_{t|t} A^T P^{-1}_{t+1|t},$$

Smoother

Suponiendo que tenemos observaciones futuras disponibles, podemos utilizarlas para mejorar la estimación.
El algoritmo es:

$$\begin{aligned} J_t &= P_{t|t} A^T P^{-1}_{t+1|t}, \\ \hat{x}_{t|T} &= \hat{x}_{t|t} + J_t \left(\hat{x}_{t+1|T} - \hat{x}_{t+1|t} \right), \end{aligned}$$

Smoother

Suponiendo que tenemos observaciones futuras disponibles, podemos utilizarlas para mejorar la estimación.
El algoritmo es:

$$\begin{aligned} J_t &= P_{t|t}A^T P_{t+1|t}^{-1}, \\ \hat{x}_{t|T} &= \hat{x}_{t|t} + J_t\left(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}\right), \\ P_{t|T} &= P_{t|t} + J_t\left(P_{t+1|T} - P_{t+1|t}\right)J_t^T. \end{aligned}$$

Características

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman
- El smoother se ejecuta desde atrás para adelante, a la inversa del filtro de Kalman

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman
- El smoother se ejecuta desde atrás para adelante, a la inversa del filtro de Kalman
- El costo computacional es comparable al filtro de Kalman

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman
- El smoother se ejecuta desde atrás para adelante, a la inversa del filtro de Kalman
- El costo computacional es comparable al filtro de Kalman
- Se espera que el smoother obtenga una estimación mejor que el filtro

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman
- El smoother se ejecuta desde atrás para adelante, a la inversa del filtro de Kalman
- El costo computacional es comparable al filtro de Kalman
- Se espera que el smoother obtenga una estimación mejor que el filtro
- Cuando el filtro convergió hay ejemplos en que la mejora del smoother no es significativa

Características

- Se pueden utilizar los datos obtenidos por el filtro de Kalman
- El smoother se ejecuta desde atrás para adelante, a la inversa del filtro de Kalman
- El costo computacional es comparable al filtro de Kalman
- Se espera que el smoother obtenga una estimación mejor que el filtro
- Cuando el filtro convergió hay ejemplos en que la mejora del smoother no es significativa
- También puede usarse un ensemble y algunas simplificaciones para poder tratar las matrices de covarianza

Ilustración

Ilustración

Imaginemos un auto de carreras

Ilustración

Imaginemos un auto de carreras

Ilustración

Para simplificar:

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante
- el piloto dispone de un modelo que calcula la velocidad teórica en cada instante de tiempo

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante
- el piloto dispone de un modelo que calcula la velocidad teórica en cada instante de tiempo
- hay radares que miden la velocidad cada 100 metros

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante
- el piloto dispone de un modelo que calcula la velocidad teórica en cada instante de tiempo
- hay radares que miden la velocidad cada 100 metros

Cuando el vehículo pasa por un puesto de control, el piloto dispone de dos datos:

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante
- el piloto dispone de un modelo que calcula la velocidad teórica en cada instante de tiempo
- hay radares que miden la velocidad cada 100 metros

Cuando el vehículo pasa por un puesto de control, el piloto dispone de dos datos:

- la velocidad teórica: dada por un modelo (por ejemplo, utilizando la ley de Newton)

Ilustración

Para simplificar:

- supongamos que el auto puede moverse con aceleración constante
- el piloto dispone de un modelo que calcula la velocidad teórica en cada instante de tiempo
- hay radares que miden la velocidad cada 100 metros

Cuando el vehículo pasa por un puesto de control, el piloto dispone de dos datos:

- la velocidad teórica: dada por un modelo (por ejemplo, utilizando la ley de Newton)
- la velocidad dada por el radar

Ilustración

Ilustración

Ninguno de esos dos datos contienen la realidad, puesto que ambos datos están afectados por error

Ilustración

Ninguno de esos dos datos contienen la realidad, puesto que ambos datos están afectados por error

- el modelo de la velocidad podría no considerar: rozamientos, vientos, limitaciones del motor, errores en el parámetro de aceleración, etc.

Ilustración

Ninguno de esos dos datos contienen la realidad, puesto que ambos datos están afectados por error

- el modelo de la velocidad podría no considerar: rozamientos, vientos, limitaciones del motor, errores en el parámetro de aceleración, etc.
- el dato del radar está afectado por el error del instrumental de medición

Ilustración

Ninguno de esos dos datos contienen la realidad, puesto que ambos datos están afectados por error

- el modelo de la velocidad podría no considerar: rozamientos, vientos, limitaciones del motor, errores en el parámetro de aceleración, etc.
- el dato del radar está afectado por el error del instrumental de medición

¿A quién le creemos?

Ilustración

Ilustración

El piloto, en cada puesto de control, puede obtener una estimación más precisa de su velocidad realizando un filtro de Kalman de manera on-line e ir corrigiendo las estimaciones paso a paso modificando las condiciones iniciales entre puesto y puesto

Ilustración

El piloto, en cada puesto de control, puede obtener una estimación más precisa de su velocidad realizando un filtro de Kalman de manera on-line e ir corrigiendo las estimaciones paso a paso modificando las condiciones iniciales entre puesto y puesto

El piloto, al finalizar la carrera, puede ir reproduciendo paso a paso la carrera, y hacer uso de las observaciones futuras cuando está pasando por un puesto de control.

Maximum likelihood estimation

Maximum likelihood estimation

Supongamos que tenemos observaciones $x_1, \ldots, x_n$ de variables aleatorias independientes e idénticamente distribuidas $X_1, \ldots, X_n$

Maximum likelihood estimation

Supongamos que tenemos observaciones $x_1, \ldots, x_n$ de variables aleatorias independientes e idénticamente distribuidas $X_1, \ldots, X_n$

La función densidad conjunta pertenece a una familia de distribuciones que dependen de un parámetro (desconocido)

$$f(x_1, \ldots, x_n; \theta) = f(x_1; \theta) \cdot \ldots \cdot f(x_n; \theta).$$

Maximum likelihood estimation

Supongamos que tenemos observaciones $x_1, \ldots, x_n$ de variables aleatorias independientes e idénticamente distribuidas $X_1, \ldots, X_n$

La función densidad conjunta pertenece a una familia de distribuciones que dependen de un parámetro (desconocido)

$$f(x_1, \ldots, x_n; \theta) = f(x_1; \theta) \cdot \ldots \cdot f(x_n; \theta).$$

La función likelihood se define como

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Maximum likelihood estimation

Maximum likelihood estimation

La función log-likelihood se define como:

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i;\theta)).$$

Maximum likelihood estimation

La función log-likelihood se define como:

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i;\theta)).$$

La idea es encontrar un parámetro que maximice la función densidad, o lo que es lo mismo, maximice la función log-likelihood

Expectation-Maximization

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Pero la idea es mejorar no sólo los parámetros respecto de las observaciones, sino también con respecto al estado

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Pero la idea es mejorar no sólo los parámetros respecto de las observaciones, sino también con respecto al estado

El algoritmo se divide en dos etapas

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Pero la idea es mejorar no sólo los parámetros respecto de las observaciones, sino también con respecto al estado

El algoritmo se divide en dos etapas

- Calcular una esperanza condicional

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Pero la idea es mejorar no sólo los parámetros respecto de las observaciones, sino también con respecto al estado

El algoritmo se divide en dos etapas

- Calcular una esperanza condicional
- Resolver un problema de optimización

Expectation-Maximization

El objetivo es estimar un conjunto de parámetros dada una sucesión de observaciones

Pero la idea es mejorar no sólo los parámetros respecto de las observaciones, sino también con respecto al estado

El algoritmo se divide en dos etapas

- Calcular una esperanza condicional
- Resolver un problema de optimización

El método fue propuesto por [Dempster et al., 1977].

Expectation-Maximization

Expectation-Maximization

Paso E: Calcular

$$Q\left(\theta, \theta^{(t)}\right) = E_{Y|X,\theta^{(t)}}\left(\log L\left(X, Y; \theta\right)\right).$$

Expectation-Maximization

Paso E: Calcular

$$Q\left(\theta, \theta^{(t)}\right) = E_{Y|X,\theta^{(t)}}\left(\log L\left(X, Y; \theta\right)\right).$$

Paso M: Resolver

$$\theta^{(t+1)} = \arg\max_{\theta} Q\left(\theta, \theta^{(t)}\right).$$

Gaussian linear state-space models

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

- M_t es el operador de observación

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

- M_t es el operador de observación
- ϕ es una matriz de transición

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

- M_t es el operador de observación
- ϕ es una matriz de transición
- $v_t \sim \mathcal{N}(0, R)$

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

- M_t es el operador de observación
- ϕ es una matriz de transición
- $v_t \sim \mathcal{N}(0, R)$
- $w_t \sim \mathcal{N}(0, Q)$

Gaussian linear state-space models

Supongamos un modelo lineal [Shumway and Stoffer, 1982]:

$$\begin{aligned} y_t &= M_t x_t + v_t, \quad t = 1, \ldots, n, \\ x_t &= \phi x_{t-1} + w_t, \quad t = 1, \ldots, n. \end{aligned}$$

donde

- M_t es el operador de observación
- ϕ es una matriz de transición
- $v_t \sim \mathcal{N}(0, R)$
- $w_t \sim \mathcal{N}(0, Q)$
- x_0 tiene esperanza μ y matriz de covarianza Σ

Gaussian linear state-space models

Gaussian linear state-space models

Deseamos estimar los parámetros

$$\mu, \Sigma, \phi, Q, R$$

a partir de este tipo de modelo y de observaciones

Gaussian linear state-space models

Gaussian linear state-space models

La función log-likelihood de todos los datos es:

$$\begin{aligned}
\log L &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_0-\mu)^T\Sigma^{-1}(x_0-\mu) - \frac{n}{2}\log|Q| \\
&- \frac{1}{2}\sum_{t=1}^{n}(x_t-\phi x_{t-1})^T Q^{-1}(x_t-\phi x_{t-1}) - \frac{n}{2}\log|R| \\
&- \frac{1}{2}\sum_{t=1}^{n}(y_t-M_t x_t)^T R^{-1}(y_t-M_t x_t)\,.
\end{aligned}$$

Gaussian linear state-space models

Gaussian linear state-space models

El paso de expectation consiste en calcular

$$
\begin{aligned}
E\left(\log L|y_1,\ldots,y_n\right) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}tr\left\{\Sigma^{-1}\left(P_0^n + \left(x_0^n-\mu\right)\left(x_0^n-\mu\right)^T\right)\right\} \\
&- \frac{n}{2}\log|Q| - \frac{1}{2}tr\left\{Q^{-1}\left(C - B\phi^T - \phi B^T + \phi A\phi^T\right)\right\} \\
&- \frac{n}{2}\log|R| \\
&- \frac{1}{2}tr\left\{R^{-1}\sum_{t=1}^{n}\left[\left(y_t - M_t x_t^n\right)\left(y_t - M_t x_t^n\right)^T + M_t P_t^n M_t^T\right]\right\}
\end{aligned}
$$

Gaussian linear state-space models

El paso de expectation consiste en calcular

$$
\begin{aligned}
E\left(\log L|y_1,\ldots,y_n\right) &= -\frac{1}{2}\log|\Sigma| - \frac{1}{2}tr\left\{\Sigma^{-1}\left(P_0^n + \left(x_0^n-\mu\right)\left(x_0^n-\mu\right)^T\right)\right\} \\
&- \frac{n}{2}\log|Q| - \frac{1}{2}tr\left\{Q^{-1}\left(C - B\phi^T - \phi B^T + \phi A\phi^T\right)\right\} \\
&- \frac{n}{2}\log|R| \\
&- \frac{1}{2}tr\left\{R^{-1}\sum_{t=1}^{n}\left[\left(y_t - M_t x_t^n\right)\left(y_t - M_t x_t^n\right)^T + M_t P_t^n M_t^T\right]\right\}
\end{aligned}
$$

Las cantidades x_0^n, P_0^n, x_t^n, etc. son el resultado de aplicar un smoother.

Gaussian linear state-space models

Gaussian linear state-space models

El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

Gaussian linear state-space models

El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

$$\mu(r+1) \quad = \quad x_0^n$$

Gaussian linear state-space models

El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

$$\begin{aligned} \mu(r+1) &= x_0^n \\ \Sigma(r+1) &= P_0^n \end{aligned}$$

Gaussian linear state-space models

El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

$$\begin{aligned} \mu(r+1) &= x_0^n \\ \Sigma(r+1) &= P_0^n \\ \phi(r+1) &= BA^{-1} \end{aligned}$$

Gaussian linear state-space models

El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

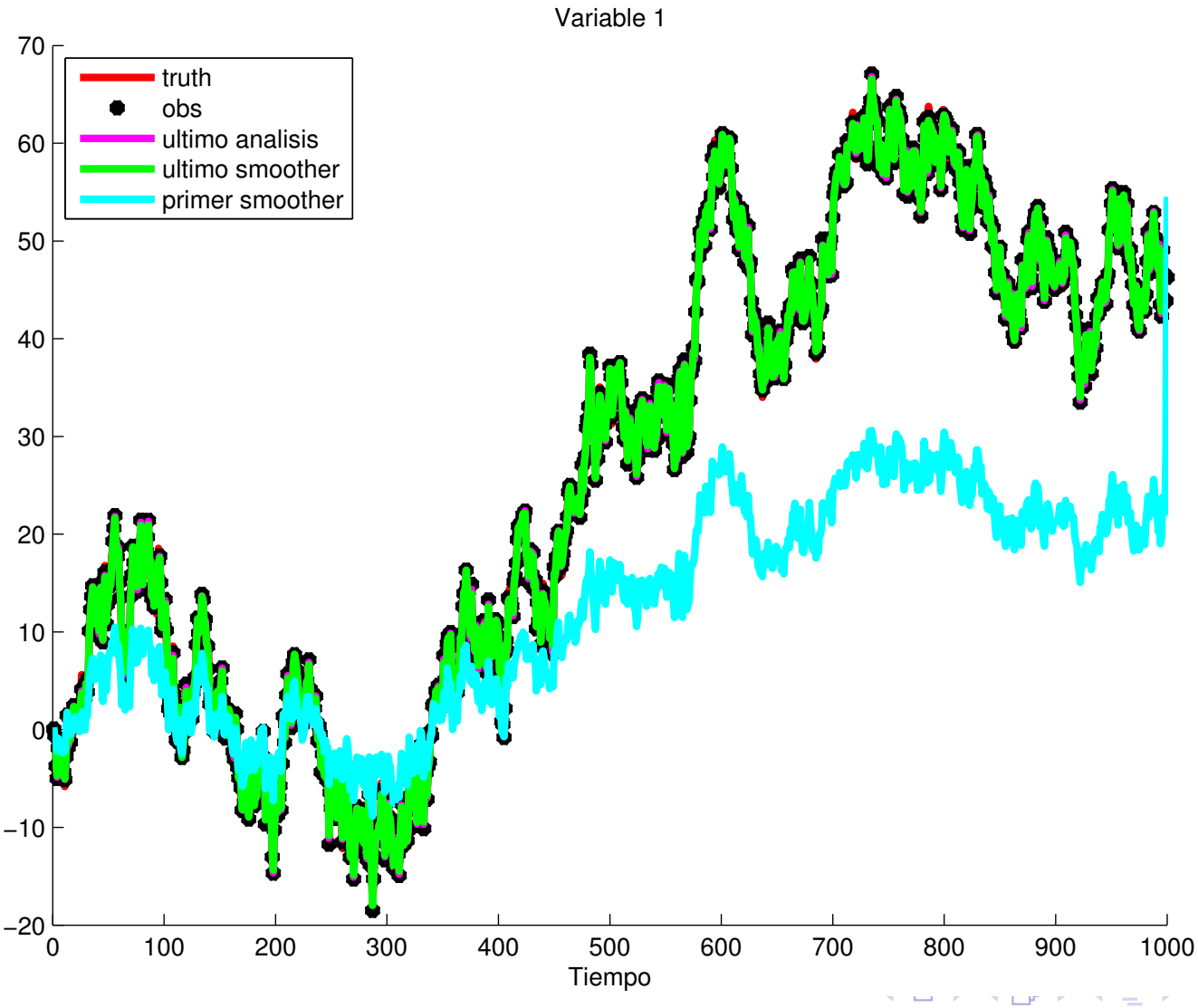
$$\begin{aligned}
\mu(r+1) &= x_0^n \\
\Sigma(r+1) &= P_0^n \\
\phi(r+1) &= BA^{-1} \\
Q(r+1) &= n^{-1}\left(C - BA^{-1}B^T\right)
\end{aligned}$$

Gaussian linear state-space models

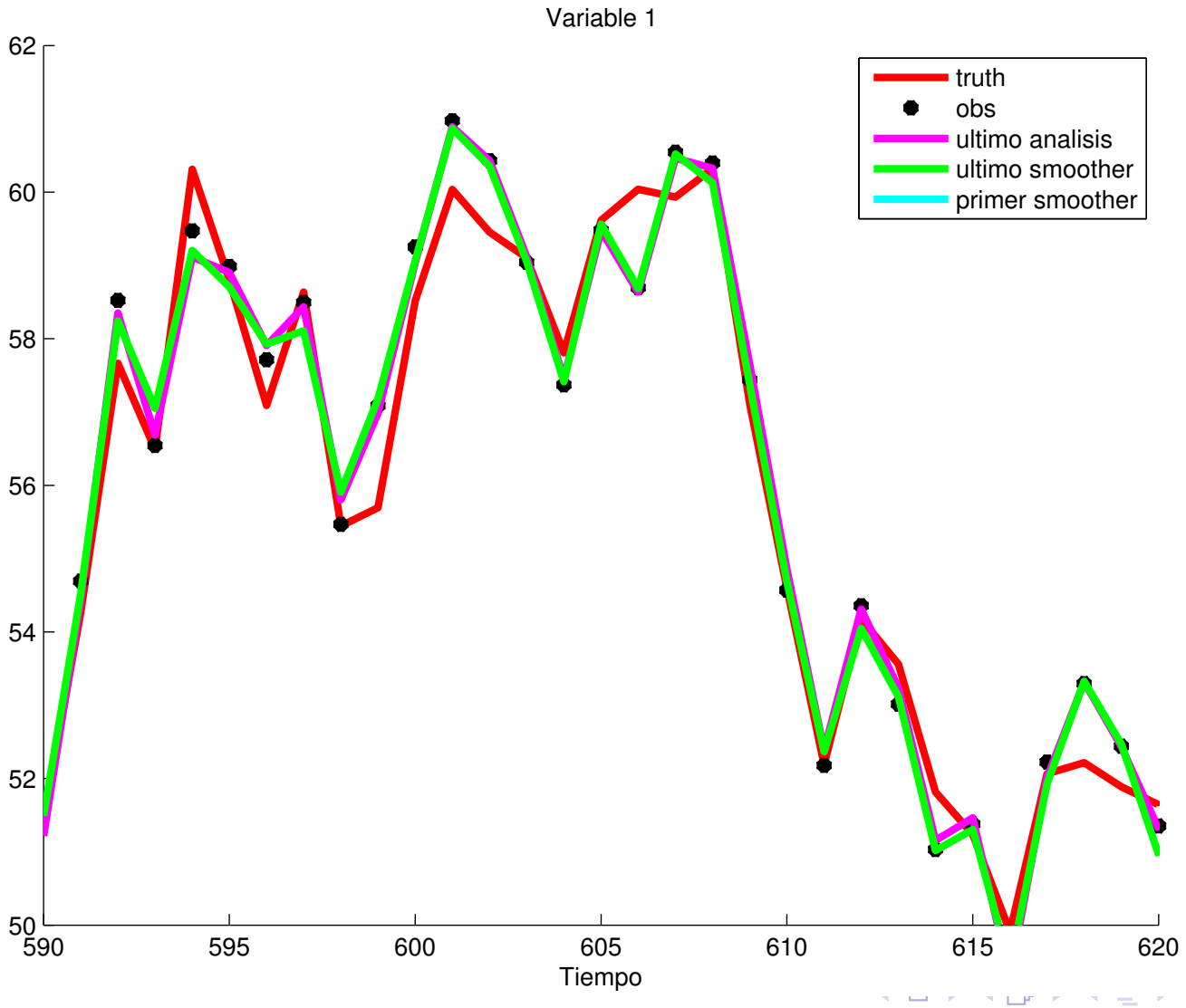
El paso de maximización consiste en igualar el gradiente de la función anterior a cero respecto de los parámetros y despejarlos.

$$
\begin{aligned}
\mu(r+1) &= x_0^n \\
\Sigma(r+1) &= P_0^n \\
\phi(r+1) &= BA^{-1} \\
Q(r+1) &= n^{-1}\left(C - BA^{-1}B^T\right) \\
R(r+1) &= n^{-1}\sum_{t=1}^{n}\left[\left(y_t - M_t x_t^n\right)\left(y_t - M_t x_t^n\right)^T + M_t P_t^n M_t^T\right]
\end{aligned}
$$

Ejemplo en dos dimensiones



Ejemplo en dos dimensiones



Desafíos

Desafíos

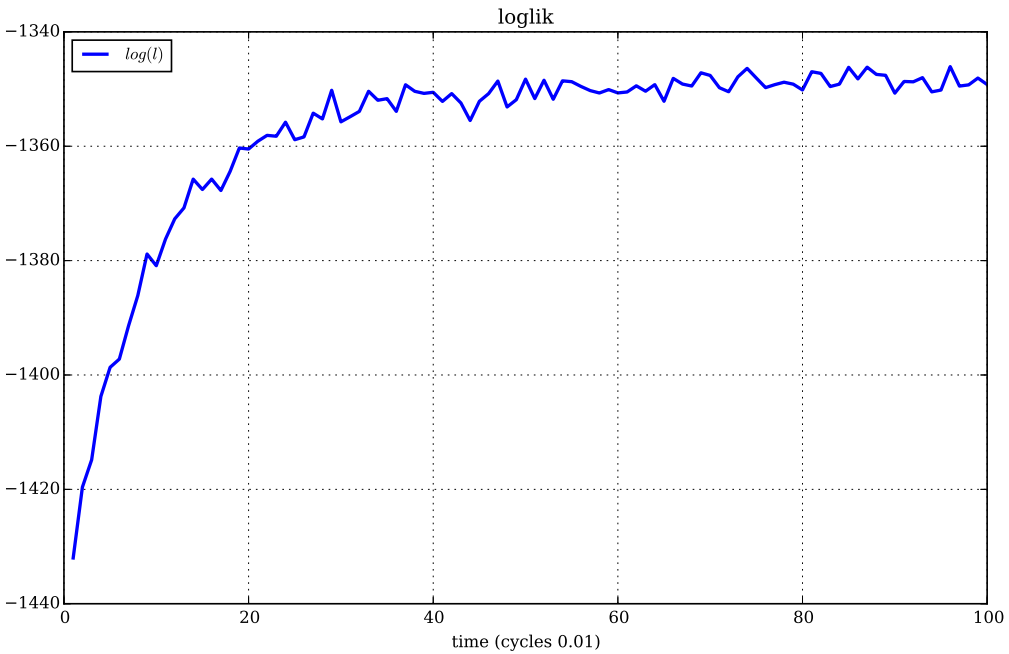
- Deducir el algoritmo EM cuando se tiene un ruido no aditivo en alguno de los parámetros

Desafíos

- Deducir el algoritmo EM cuando se tiene un ruido no aditivo en alguno de los parámetros
- Cuando se realiza el algoritmo EM utilizando un ensemble para el filtro y el smoother, debe dar resultados satisfactorios.

Desafíos

- Para el caso del modelo AR1: $x_{t+1} = \phi x_t + \sigma\xi$, con $\xi \sim \mathcal{N}(0,1)$ la estrategia del estado aumentado no da resultados [Santitissadeekorn and Jones, 2015], por lo que se puede usar el algoritmo EM



Desafíos

Desafíos

- Algunos modelos como el de Lorenz63 o Lorenz96 pueden ser usados como estudios preliminares antes de pasar a un modelo más complejo (como los modelos meteorológicos)

Desafíos

- Algunos modelos como el de Lorenz63 o Lorenz96 pueden ser usados como estudios preliminares antes de pasar a un modelo más complejo (como los modelos meteorológicos)
- El objetivo final es implementar este algoritmo para obtener parámetros estocásticos en modelos meteorológicos a pesar de su alto costo computacional

Desafíos

- Algunos modelos como el de Lorenz63 o Lorenz96 pueden ser usados como estudios preliminares antes de pasar a un modelo más complejo (como los modelos meteorológicos)
- El objetivo final es implementar este algoritmo para obtener parámetros estocásticos en modelos meteorológicos a pesar de su alto costo computacional
- Se incorporaron rutinas preliminares de EM a un paquete que implementa la versión LETKF del filtro de Kalman y el smoother

Desafíos

- Algunos modelos como el de Lorenz63 o Lorenz96 pueden ser usados como estudios preliminares antes de pasar a un modelo más complejo (como los modelos meteorológicos)
- El objetivo final es implementar este algoritmo para obtener parámetros estocásticos en modelos meteorológicos a pesar de su alto costo computacional
- Se incorporaron rutinas preliminares de EM a un paquete que implementa la versión LETKF del filtro de Kalman y el smoother
- Este paquete está diseñado de manera tal que es muy “similar” a los modelos meteorológicos pues el destino final es un modelo meteorológico de los que se encuentran operativos

¡MUCHAS GRACIAS!

Bibliografía I

Anderson, J. L. (2001).
An ensemble adjustment Kalman filter for data assimilation.
Monthly Weather Review, 129(12):2884–2903.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977).
Maximum likelihood from incomplete data via the EM algorithm.

Journal of the Royal Statistical Society. Series B (methodological), pages 1–38.

Evensen, G. (1997).
Advanced data assimilation for strongly nonlinear dynamics.
Monthly Weather Review, 125(6):1342–1354.

Bibliografía II

Kalman, R. E. (1960).
A new approach to linear filtering and prediction problems.
Journal of Basic Engineering, 82(1):35–45.

Santitissadeekorn, N. and Jones, C. (2015).
Two-stage filtering for joint state-parameter estimation.
Monthly Weather Review, 143:2028–2042.

Shumway, R. H. and Stoffer, D. S. (1982).
An approach to time series smoothing and forecasting using the EM algorithm.
Journal of Time Series Analysis, 3(4):253–264.

Bibliografía III

Torres, G. (2010).
Algorithm 900: a discrete time Kalman filter package for large scale problems.
ACM Transactions on Mathematical Software (TOMS), 37(1):11.

Zupanski, D. and Zupanski, M. (2006).
Model error estimation employing an ensemble data assimilation approach.
Monthly Weather Review, 134(5):1337–1354.