State-Space Models - Analysis of time series (Theory)

(CIV6540 - Probabilistic Machine Learning for Civil Engineers)

Professor: James-A. Goulet

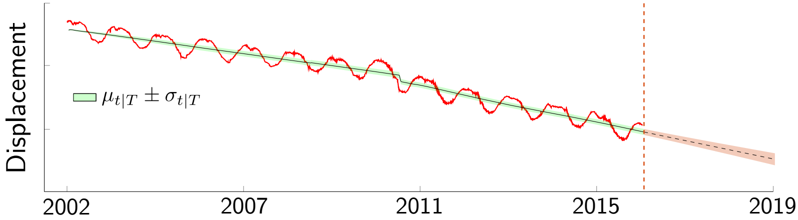
Département des génies civil, géologique et des mines
Polytechnique Montréal

Chapter 12 – Goulet (2020)
Probabilistic Machine Learning for Civil Engineers
MIT Press

Chapter 18 – Murphy, K. (2012)
Machine learning: A probabilistic perspective.
MIT Press.

Time-series

How do we interpret data when it is **time-dependent**?

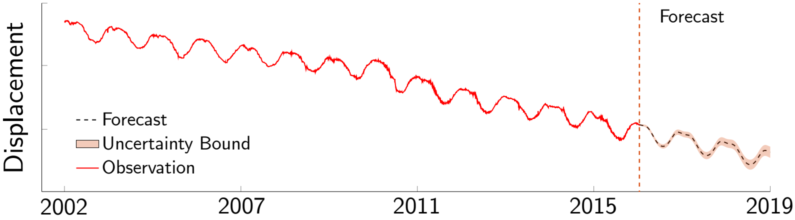


Possible tasks

1. Filter the observation noise & patterns
2. Forecast future states or observations
3. Estimate hidden state variables
4. Detect anomalies

Time-series

How do we interpret data when it is **time-dependent**?

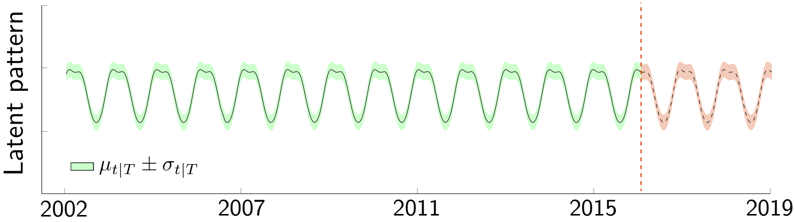


Possible tasks

1. Filter the observation noise & patterns
2. Forecast future states or observations
3. Estimate hidden state variables
4. Detect anomalies

Time-series

How do we interpret data when it is **time-dependent**?

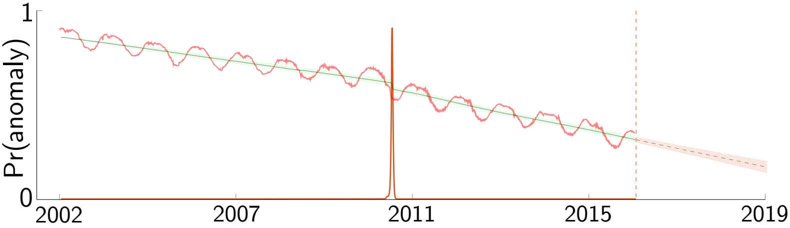


Possible tasks

1. Filter the observation noise & patterns
2. Forecast future states or observations
3. Estimate hidden state variables
4. Detect anomalies

Time-series

How do we interpret data when it is **time-dependent**?



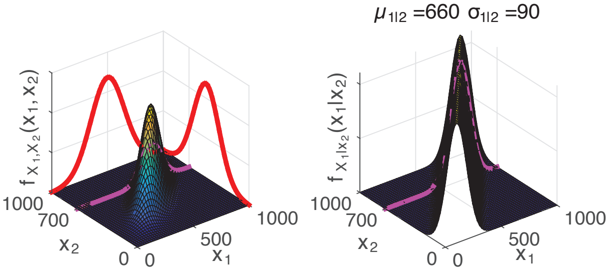
Possible tasks

1. Filter the observation noise & patterns
2. Forecast future states or observations
3. Estimate hidden state variables
4. Detect anomalies

Methodology overview

- **Machine Learning**: State-Space Models
- **Control theory**: Kalman Filter
- **Statistics**: Dynamic Linear Models

These methods are all based on **Gaussian conditionals**



Kalman filter – Origin

Developped for NASA as the guiding system for the **Appolo mission** (1960)

Still employed today in modern rockets...

"Rocket science" applied to civil engineering

[Wikipedia, SpaceX]

Module #7 Outline

Revision $\mathcal{N}(\mu, \Sigma)$

Theory

$t+n|t, t|T$

θ estimation

SKF

Limitations

Topics organization

Background	1	Revision probability & linear algebra
	2	Probability distributions
Machine Learning Basics	0	Introduction
	3	Bayesian Estimation $p(A|B) = \frac{p(B|A)p(A)}{p(B)}$
	4	MCMC sampling & Newton
Supervised learning	5	Regression
	6	Classification
	7	LSTM networks for time series
Unsupervised learning	7	State-space model for time-series
Decision Making & RL	8	Decision Theory
	9	AI & Sequential decision problems

Section Outline

Revision $\mathcal{N}(\mu, \Sigma)$

1.1 Fonctions of Gaussian Random Variables

1.2 Transformation Rule

1.3 Conditional Gaussian PDF

Fonctions of Gaussian Random Variables – $g(\mathbf{X})$ (Revision)

Given $\mathbf{X} = [X_1, X_2, \cdots, X_n]^\intercal$ defined by the PDF $f_{\mathbf{X}}(\mathbf{x})$, and given $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_m]^\intercal$ obtained from a **linear** function

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}) = \begin{bmatrix} g_1(\mathbf{X}) \\ g_2(\mathbf{X}) \\ \vdots \\ g_n(\mathbf{X}) \end{bmatrix}$$

Question: what is the PDF $f_{\mathbf{Y}}(\mathbf{y})$,

2 cases:

1. $m = n = 1$
2. $m, n \geq 1$

We consider only
linear fonctions
i.e. $\mathbf{Y} = \mathbf{g}(\mathbf{X}) = \mathbf{A}\mathbf{X} + \mathbf{b}$

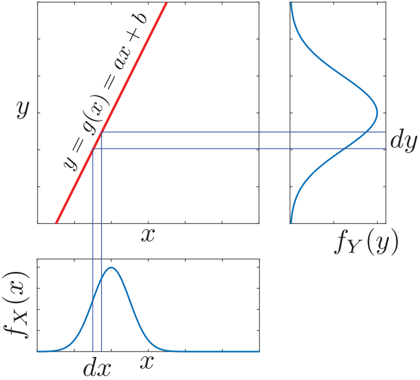
$Y = g(X)$: Case 1, $m = n = 1$

$$f_Y(y)dy = f_X(x)dx$$

$$f_Y(y) = \underbrace{f_X(x)\left|\frac{dx}{dy}\right|}_{f_X(x)\geq 0}$$

For $Y = aX + b$

$$\left|\frac{dy}{dx}\right| = |a|$$

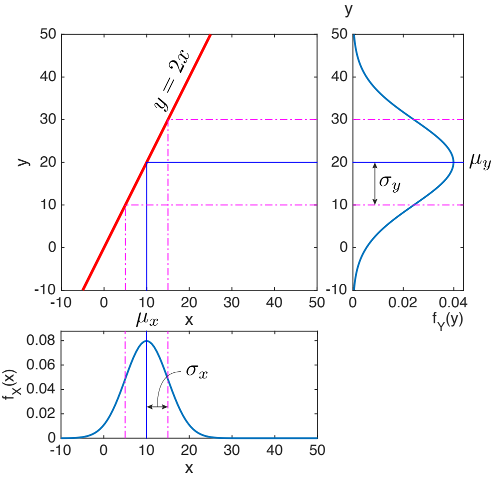


$$X \sim \mathcal{N}(x;\mu_X,\sigma_X^2) \rightarrow Y \sim \mathcal{N}\left(y; a\mu_X + b, (a\sigma_X)^2\right)$$

Example: Case 1, $Y = 2X$ [ ]

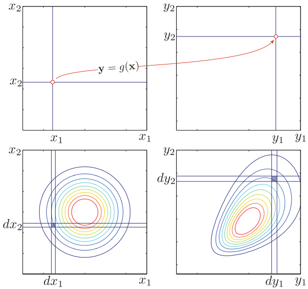
$$X \sim \mathcal{N}(\underbrace{10}_{\mu_X}, \underbrace{5^2}_{\sigma_X^2})$$

$$Y \sim \mathcal{N}(\underbrace{2\mu_X}_{\mu_Y}, (\underbrace{2\sigma_X}_{\sigma_Y})^2)$$



[CIVML/linearisation.py]

$\mathbf{Y} = \mathbf{g}(\mathbf{X})$: Case 2, $m, n > 1$



$$f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}$$

$$f_{\mathbf{X}}(\mathbf{x})\frac{d\mathbf{x}}{d\mathbf{y}} = f_{\mathbf{Y}}(\mathbf{y})$$

$$\boxed{J_{k,i} = J_{y_k,x_i} = \frac{\partial y_k}{\partial x_i}}$$

Jacobian ($\mathbf{J}_{\mathbf{y},\mathbf{x}}$): size of the neighbourhood from $d\mathbf{y}$ to $d\mathbf{x}$

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} = \begin{bmatrix} J_{y_1,x_1} & \cdots & J_{y_1,x_n} \\ & \ddots & \vdots \\ \text{sym.} & & J_{y_m,x_n} \end{bmatrix} = \begin{bmatrix} \nabla g_1(\mathbf{x}) \\ \vdots \\ \nabla g_n(\mathbf{x}) \end{bmatrix}$$

$$\frac{d\mathbf{x}}{d\mathbf{y}} = |\det \mathbf{J}_{\mathbf{y},\mathbf{x}}|^{-1}$$

$$\frac{d\mathbf{y}}{d\mathbf{x}} = |\det \mathbf{J}_{\mathbf{y},\mathbf{x}}|$$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\,|\det \mathbf{J}_{\mathbf{y},\mathbf{x}}|^{-1}$$

$\mathbf{Y}=\mathbf{g}(\mathbf{X})$: Case 2, $m,n>1$ (cont.)

Given a linear function $\mathbf{Y}=\mathbf{g}(\mathbf{X})=\mathbf{AX}+\mathbf{b}$ where

$$\mathbf{X}=\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},\quad \boldsymbol{\mu}_\mathbf{X}=\begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{bmatrix},\quad \boldsymbol{\Sigma}_\mathbf{X}=\begin{bmatrix} \sigma^2_{X_1} & \cdots & \rho_{1n}\sigma_{X_1}\sigma_{X_n} \\ & \ddots & \vdots \\ \text{sym.} & & \sigma^2_{X_n} \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \quad \end{bmatrix}_{m\times 1}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} \quad\quad \end{bmatrix}_{m\times n}}_{\mathbf{A}=\mathbf{J}_{\mathbf{y},\mathbf{x}}} \times \underbrace{\begin{bmatrix} \quad \end{bmatrix}_{n\times 1}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} \quad \end{bmatrix}_{m\times 1}}_{\mathbf{b}}$$

$$\boxed{\begin{aligned} \boldsymbol{\mu}_\mathbf{Y} &= \mathbf{A}\boldsymbol{\mu}_\mathbf{X}+\mathbf{b} \\ \boldsymbol{\Sigma}_\mathbf{Y} &= \mathbf{A}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{A}^\intercal \\ \boldsymbol{\Sigma}_\mathbf{YX} &= \mathbf{A}\boldsymbol{\Sigma}_\mathbf{X} \end{aligned}} \left.\vphantom{\begin{aligned}a\\b\\c\end{aligned}}\right\} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix},\ \begin{bmatrix} \boldsymbol{\mu}_\mathbf{X} \\ \boldsymbol{\mu}_\mathbf{Y} \end{bmatrix},\ \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{X} & \boldsymbol{\Sigma}_\mathbf{XY} \\ \boldsymbol{\Sigma}_\mathbf{YX} & \boldsymbol{\Sigma}_\mathbf{Y} \end{bmatrix}$$

Gaussian Conditional PDF (Revision)

Conditional distributions of a multivariate Gaussian are Gaussian

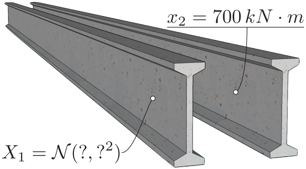
$$\mathbf{X} = \begin{Bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{Bmatrix}, \ \boldsymbol{\mu} = \begin{Bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{Bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$$

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\underbrace{\mathbf{X}_2 = \mathbf{x}_2}_{\text{observations}}) = \frac{f_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_2}(\mathbf{x}_2)} = \mathcal{N}(\mathbf{x}_1; \boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$$

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \quad \boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

$$\text{2D: } \mu_{1|2} = \mu_1 + \rho\sigma_1 \frac{x_2 - \mu_2}{\sigma_2}, \quad \sigma_{1|2} = \sigma_1\sqrt{1 - \rho^2}$$

Example – Multivariate Gaussian Conditional



The prior knowledge of the true resistance (X_1, X_2) of two beams is

$$\begin{array}{lcll} X_1 & \sim & \mathcal{N}(x_1; 500, 150^2) & [kN \cdot m] \\ X_2 & \sim & \mathcal{N}(x_2; 500, 150^2) & [kN \cdot m] \end{array}$$

Knowing that the resistance of beams is correlated: $\rho_{12} = 0.8$,
what is the resistance of X_1,
if we observe that $x_2 = 700\,kN \cdot m$?

$$f_{X_1|x_2}(x_1|x_2) = \mathcal{N}(x_1; \mu_{1|2}, \sigma^2_{1|2}) = \mathcal{N}(x_1; 660, 90^2)$$

[Radio-Canada]

Example – Multivariate Gaussian Conditional (cont.)

$$f_{X_1X_2}(x_1,x_2) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}}) \begin{cases} \boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} 500 \\ 500 \end{bmatrix} \\ \boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 150^2 & 0.8\cdot150^2 \\ 0.8\cdot150^2 & 150^2 \end{bmatrix} \end{cases}$$

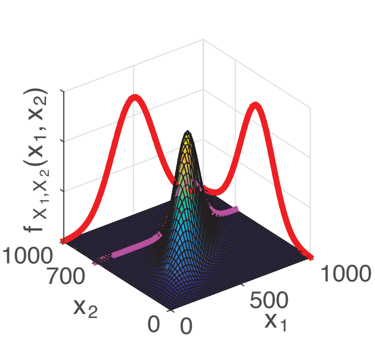
$$f_{X_1|x_2}(x_1|x_2) = \frac{f_{X_1X_2}(x_1,x_2)}{f_{X_2}(x_2)} = \mathcal{N}(\mu_{1|2}, \sigma^2_{1|2}) \begin{cases} \mu_{1|2} = \mu_1 + \rho\sigma_1 \frac{x_2-\mu_2}{\sigma_2} \\ \sigma_{1|2} = \sigma_1\sqrt{1-\rho^2} \end{cases}$$

$$\mu_{1|2} = 500 + 0.8\times150\frac{\overbrace{700}^{\text{observation}}-500}{150} = 660\,kN\cdot m$$

$$\sigma_{1|2} = 150\sqrt{1-0.8^2} = 90\,kN\cdot m$$

Example – Multivariate Gaussian Conditional (cont.) [ ]

$$f_{X_1|x_2}(x_1|x_2) = \frac{f_{X_1X_2}(x_1,x_2)}{f_{X_2}(x_2)} = \mathcal{N}(\mu_{1|2},\sigma^2_{1|2}) \begin{cases} \mu_{1|2} = \mu_1 + \rho\sigma_1\frac{x_2-\mu_2}{\sigma_2} \\ \sigma_{1|2} = \sigma_1\sqrt{1-\rho^2} \end{cases}$$



[CIVML/poutres_conditionnelles.py]

Summary – Multivariate Gaussian

For $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{X} \sim \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_\mathbf{X}, \boldsymbol{\Sigma}_\mathbf{X})$, and a vector of observations $\mathbf{y}$

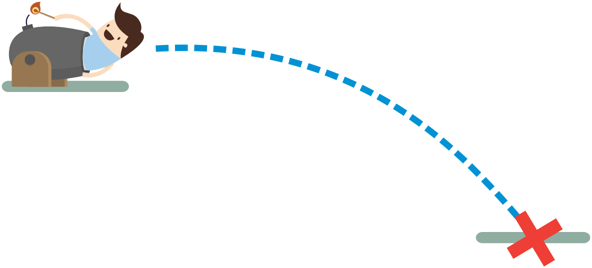
$$\left.\begin{array}{rcl} \boldsymbol{\mu}_\mathbf{Y} & = & \mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b} \\ \boldsymbol{\Sigma}_\mathbf{Y} & = & \mathbf{A}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{A}^\intercal \\ \boldsymbol{\Sigma}_\mathbf{XY} & = & \boldsymbol{\Sigma}_\mathbf{X}\mathbf{A}^\intercal \end{array}\right\} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\mu}_\mathbf{X} \\ \boldsymbol{\mu}_\mathbf{Y} \end{bmatrix}, \begin{bmatrix} \Sigma_\mathbf{X} & \Sigma_\mathbf{XY} \\ \Sigma_\mathbf{YX} & \Sigma_\mathbf{Y} \end{bmatrix}$$

$$\overbrace{f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\underbrace{\mathbf{Y}=\mathbf{y}}_{\text{observations}})}^{\text{posterior knowledge}} = \frac{\overbrace{f_{\mathbf{XY}}(\mathbf{x},\mathbf{y})}^{\text{prior knowledge}}}{f_\mathbf{Y}(\mathbf{y})} = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}}, \boldsymbol{\Sigma}_{\mathbf{X}|\mathbf{Y}})$$

$$\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}} = \boldsymbol{\mu}_\mathbf{X} + \boldsymbol{\Sigma}_\mathbf{XY}\boldsymbol{\Sigma}_\mathbf{Y}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y}), \quad \boldsymbol{\Sigma}_{\mathbf{X}|\mathbf{Y}} = \boldsymbol{\Sigma}_\mathbf{X} - \boldsymbol{\Sigma}_\mathbf{XY}\boldsymbol{\Sigma}_\mathbf{Y}^{-1}\boldsymbol{\Sigma}_\mathbf{YX}$$

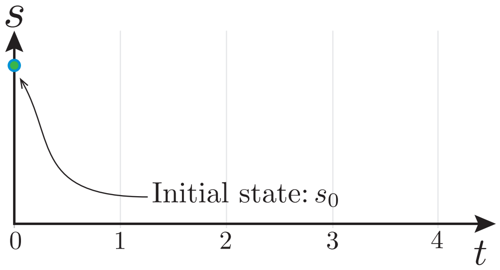
Conceptual introduction

$$\overbrace{\begin{array}{rcll} y_1 = s_{t+1} + v & = & s_t + \dot{s}_t \Delta t + \frac{1}{2}\ddot{s}_t \Delta t^2 + w + v & \text{(Position)} \\ \dot{s}_{t+1} & = & \dot{s}_t + \ddot{s}_t \Delta t + \dot{w} & \text{(Speed)} \\ \ddot{s}_{t+1} & = & \ddot{s}_t + \ddot{w} & \text{(Acceleration)} \end{array}}^{\text{Kinematic equations}}$$



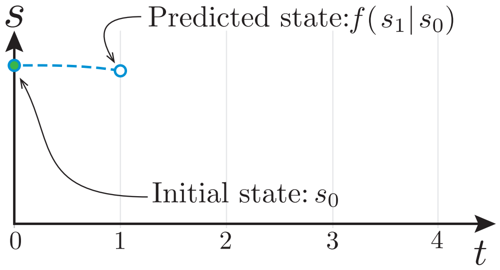
Conceptual introduction

$$\overbrace{\begin{array}{rcll} y_1 = s_{t+1} + v & = & s_t + \dot{s}_t \Delta t + \frac{1}{2}\ddot{s}_t \Delta t^2 + w + v & \text{(Position)} \\ \dot{s}_{t+1} & = & \dot{s}_t + \ddot{s}_t \Delta t + \dot{w} & \text{(Speed)} \\ \ddot{s}_{t+1} & = & \ddot{s}_t + \ddot{w} & \text{(Acceleration)} \end{array}}^{\text{Kinematic equations}}$$



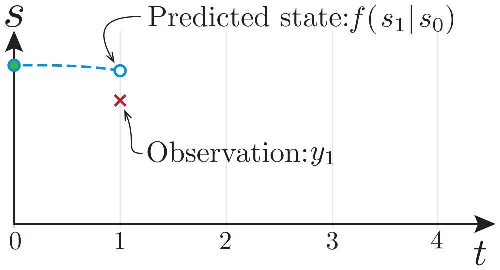
Conceptual introduction

$$\overbrace{\begin{array}{rcll} y_1 = s_{t+1} + v & = & s_t + \dot{s}_t \Delta t + \frac{1}{2}\ddot{s}_t \Delta t^2 + w + v & \text{(Position)} \\ \dot{s}_{t+1} & = & \dot{s}_t + \ddot{s}_t \Delta t + \dot{w} & \text{(Speed)} \\ \ddot{s}_{t+1} & = & \ddot{s}_t + \ddot{w} & \text{(Acceleration)} \end{array}}^{\text{Kinematic equations}}$$



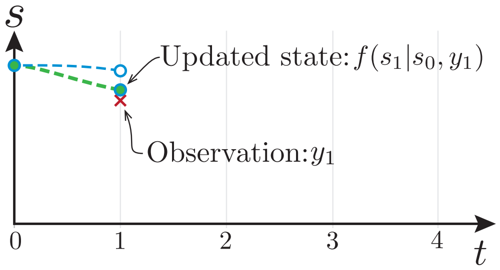
Conceptual introduction

$$\overbrace{\begin{array}{rcll} y_1 = s_{t+1} + v & = & s_t + \dot{s}_t\Delta t + \frac{1}{2}\ddot{s}_t\Delta t^2 + w + v & \text{(Position)}\\ \dot{s}_{t+1} & = & \dot{s}_t + \ddot{s}_t\Delta t + \dot{w} & \text{(Speed)}\\ \ddot{s}_{t+1} & = & \ddot{s}_t + \ddot{w} & \text{(Acceleration)} \end{array}}^{\text{Kinematic equations}}$$



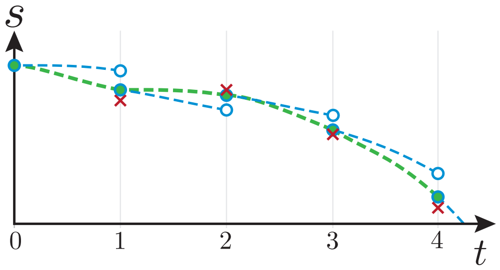
Conceptual introduction

$$\overbrace{\begin{aligned} y_1 = s_{t+1} + v &= s_t + \dot{s}_t\Delta t + \tfrac{1}{2}\ddot{s}_t\Delta t^2 + w + v && \text{(Position)}\\ \dot{s}_{t+1} &= \dot{s}_t + \ddot{s}_t\Delta t + \dot{w} && \text{(Speed)}\\ \ddot{s}_{t+1} &= \ddot{s}_t + \ddot{w} && \text{(Acceleration)}\end{aligned}}^{\text{Kinematic equations}}$$



Conceptual introduction

$$\overbrace{\begin{array}{rcll} y_1 = s_{t+1} + v & = & s_t + \dot{s}_t \Delta t + \frac{1}{2}\ddot{s}_t \Delta t^2 + w + v & \text{(Position)} \\ \dot{s}_{t+1} & = & \dot{s}_t + \ddot{s}_t \Delta t + \dot{w} & \text{(Speed)} \\ \ddot{s}_{t+1} & = & \ddot{s}_t + \ddot{w} & \text{(Acceleration)} \end{array}}^{\text{Kinematic equations}}$$

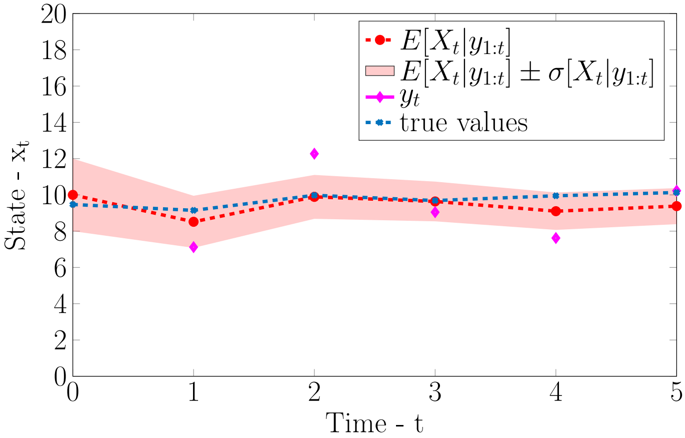


Section Outline

Theory

2.1 Conceptual introduction
2.3 Basic problem setup
2.4 Notation
2.5 Prediction and update steps
2.6 Example #1
2.7 General formulation

Example #1 – Temperature time-series



Basic problem setup

$t = 0, 1, \cdots, T$: Discrete time stamps
x_t: State of a system at a time step t (deterministic, yet unknown)
y_t: Imperfect observation of x_t at a time step t

$$\underbrace{y_t = x_t + v_t}_{\text{observation model}}, \quad \overbrace{v_t : V \sim \mathcal{N}(v; 0, \sigma_V^2)}^{\text{observation error}}$$

$$\underbrace{x_t = x_{t-1} + w_t}_{\text{transition model}}, \quad \overbrace{w_t : W \sim \mathcal{N}(w; 0, \sigma_W^2)}^{\text{model error}}$$

Our knowledge of x_t conditional on the all available data $y_{1:t}$ is

$$f_{X_t|y_{1:t}}(x_t|y_{1:t}) = \mathcal{N}(x_t; \mathbb{E}[X_t|y_{1:t}],\ \text{Var}[X_t|y_{1:t}])$$

Compact notation

Long form	$\equiv$	Short form
$\{y_1, y_2, \cdots, y_t\}$	$\equiv$	$y_{1:t}$
$\mathbb{E}[X_t \vert y_{1:t}]$	$\equiv$	$\mu_{t\vert t}$
$\mathbb{E}[X_t \vert y_{1:t-1}]$	$\equiv$	$\mu_{t\vert t-1}$
$\mathrm{Var}[X_t \vert y_{1:t}]$	$\equiv$	$\sigma^2_{t\vert t}$

Prediction step: $t = 0 \rightarrow t = 1$

At time $t = 0$ only our prior knowledge $\mathbb{E}[X_0] = \mu_0$ and $\text{Var}[X_0] = \sigma_0^2$ is available. Note: no knowledge, $\text{Var}[X_0] \rightarrow \infty$

We compute our **prior knowledge** at $t = 1$ from our **knowledge** at $t = 0$.

Transition model: $\underline{x_1 = x_0 + w_1}$, $w_1 : W \sim \mathcal{N}(0, \sigma_W^2)$

Observation model: $\underline{y_1 = x_1 + v_1}$, $v_1 : V \sim \mathcal{N}(0, \sigma_V^2)$

$$
\begin{aligned}
\mathbb{E}[X_1|X_0] &\equiv \mu_{1|0} &&= \mu_0 \\
\text{Var}[X_1|X_0] &= \sigma_{1|0}^2 &&= \sigma_0^2 + \sigma_W^2 \\
\mathbb{E}[Y_1|X_0] &= \mu_{1|0} &&= \mu_0 \\
\text{Var}[Y_1|X_0] &= \sigma_{1|0}^2 + \sigma_V^2 &&= \sigma_0^2 + \sigma_W^2 + \sigma_V^2 \\
\text{Cov}[X_1, Y_1|X_0] &= \sigma_{1|0}^2 &&= \sigma_0^2 + \sigma_W^2
\end{aligned}
$$

Update step: $t = 1$

Joint prior knowledge: $f(x_1, y_1|x_0) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_1|X_0] \\ \mathbb{E}[Y_1|X_0] \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} = \begin{bmatrix} \mathrm{Var}[X_1|X_0] & \mathrm{Cov}[X_1, Y_1|X_0] \\ \mathrm{Cov}[X_1, Y_1|X_0] & \mathrm{Var}[Y_1|X_0] \end{bmatrix}$$

Posterior knowledge:

$$f(x_1|x_0, y_1) = \frac{f(x_1, y_1|x_0)}{f(y_1|x_0)} = \mathcal{N}(x_1; \mu_{1|1}, \sigma^2_{1|1})$$

Posterior mean and variance:

$$\begin{aligned} \mu_{1|1} &= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y_1 - \mu_Y) \\ \sigma^2_{1|1} &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} \end{aligned}$$

Prediction step: $t-1 \rightarrow t$

From our **posterior knowledge** at $t-1$, the **transition model** $\underline{x_t = x_{t-1} + w_t}$ and the **observation model** $\underline{y_t = x_t + v_t}$, we can compute our **prior knowledge** at t

$$
\begin{aligned}
\mathbb{E}[X_t|y_{1:t-1}] &\equiv \mu_{t|t-1} &&= \mu_{t-1|t-1}\\
\mathrm{Var}[X_t|y_{1:t-1}] &= \sigma^2_{t|t-1} &&= \sigma^2_{t-1|t-1} + \sigma^2_W\\
\mathbb{E}[Y_t|y_{1:t-1}] &= \mu_{t|t-1} &&\\
\mathrm{Var}[Y_t|y_{1:t-1}] &= \underbrace{\sigma^2_{t-1|t-1} + \sigma^2_W}_{\sigma^2_{t|t-1}} + \sigma^2_V &&\\
\mathrm{Cov}[X_t, Y_t|y_{1:t-1}] &= \overbrace{\sigma^2_{t-1|t-1} + \sigma^2_W} &&
\end{aligned}
$$

Update step: $t-1 \rightarrow t$

Joint prior knowledge: $f(x_t, y_t | y_{1:t-1}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_t|y_{1:t-1}] \\ \mathbb{E}[Y_t|y_{1:t-1}] \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} = \begin{bmatrix} \text{Var}[X_t|y_{1:t-1}] & \text{Cov}[X_t, Y_t|y_{1:t-1}] \\ \text{Cov}[X_t, Y_t|y_{1:t-1}] & \text{Var}[Y_t|y_{1:t-1}] \end{bmatrix}$$

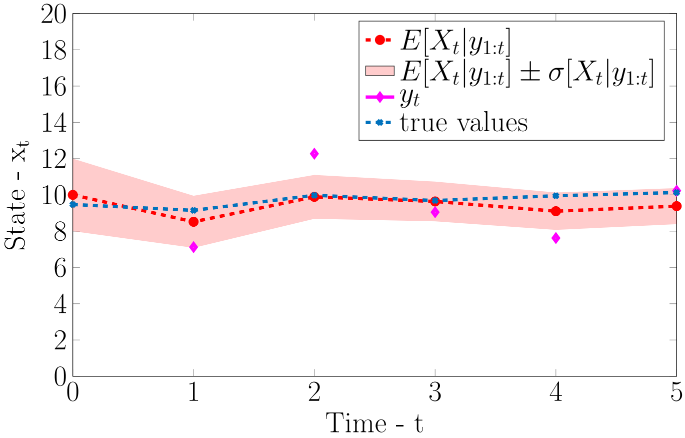
Posterior knowledge:

$$f(x_t|y_{1:t}) = \frac{f(x_t, y_t|y_{1:t-1})}{f(y_t|y_{1:t-1})} = \mathcal{N}(x_t; \mu_{t|t}, \sigma^2_{t|t})$$

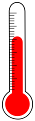
Posterior mean and variance:

$$\begin{aligned} \mu_{t|t} &= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y_t - \mu_Y) \\ \sigma^2_{t|t} &= \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} \end{aligned}$$

Example #1 – Temperature time-series



Example #1 – Temperature time-series



We want to estimate the temperature x_t over a period of 5 hours using imprecise measurements made every hour, $t = 0:5$.

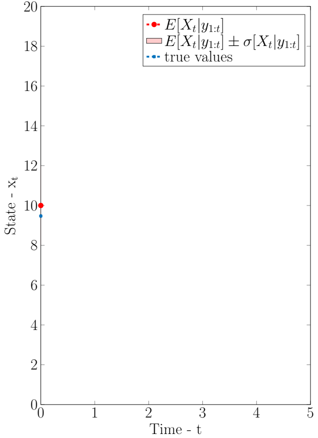
Our prior knowledge is that $\mathbb{E}[X_0] = 10\,°C$ and $\sigma[X_0] = 2\,°C$.

$$\underbrace{y_t = x_t + v_t}_{\text{observation model}}, \quad \overbrace{v_t : V \sim \mathcal{N}(0, 2^2)}^{\text{observation error}}$$

$$\underbrace{x_t = x_{t-1} + w_t}_{\text{transition model}}, \quad \overbrace{w_t : W \sim \mathcal{N}(0, 0.5^2)}^{\text{model error}}$$

Example #1 – Temperature time-series [.m]

$t = 1 \mid y_t = 7.1$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 10$
$Var[X_{t-1|t-1}] = 4$



[CIV_ML/KF_example_LL.m]

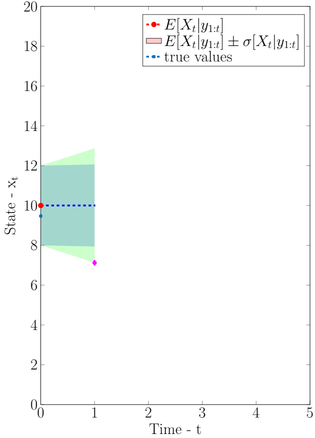
Example #1 – Temperature time-series [.m]

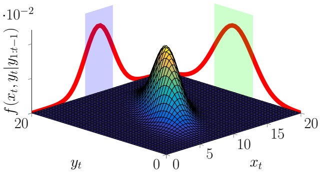
$t = 1 \mid y_t = 7.1$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 10$
$Var[X_{t-1|t-1}] = 4$

Prediction step:

$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 10$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 10$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 4.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 8.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 4.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.72$





[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 1 \mid y_t = 7.1$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 10$
$Var[X_{t-1|t-1}] = 4$

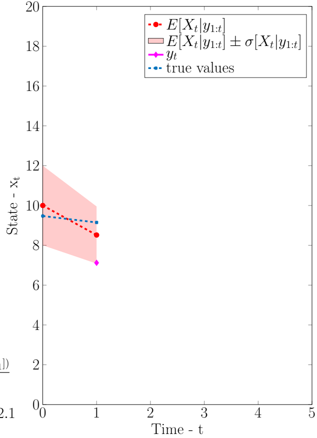
Prediction step:

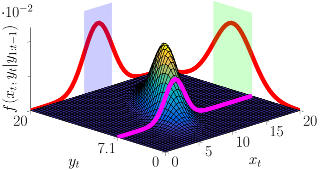
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 10$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 10$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 4.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 8.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 4.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.72$

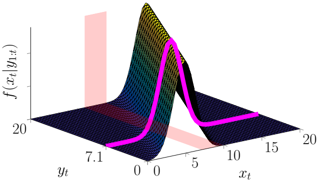
Update step:

$$E[X_{t|t}] = E[X_{t|t-1}] + \frac{Cov[X_{t|t-1}, Y_{t|t-1}](y_t - E[Y_{t|t-1}])}{Var[Y_{t|t-1}]} = 8.5$$
$$Var[X_{t|t}] = Var[X_{t|t-1}] - \frac{Cov[X_{t|t-1}, Y_{t|t-1}]^2}{Var[Y_{t|t-1}]} = 2.1$$





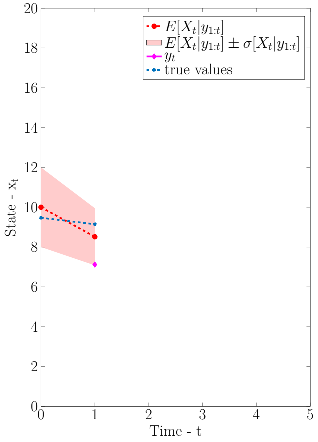
$E[X_t|y_{1:t}]$ =8.5 $\sigma[X_t|y_{1:t}]$ =1.4



[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 2 \mid y_t = 12.3$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 8.5$
$Var[X_{t-1|t-1}] = 2.1$



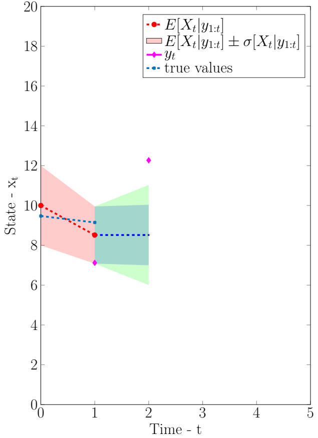
[CIV_ML/KF_example_LL.m]

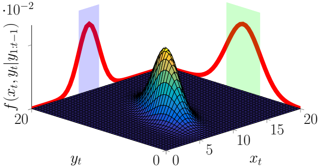
Example #1 – Temperature time-series [.m]

$t = 2 \mid y_t = 12.3$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 8.5$
$Var[X_{t-1|t-1}] = 2.1$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 8.5$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 8.5$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 2.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 6.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 2.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.61$





[CIV_ML/KF_example_LL.m]

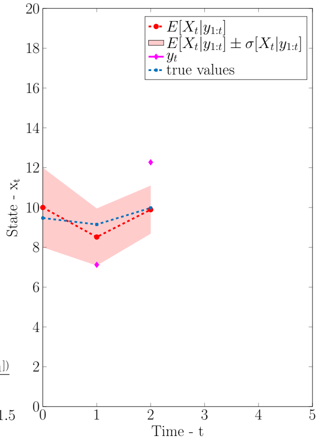
Example #1 – Temperature time-series [.m]

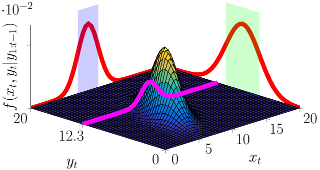
$t=2 \mid y_t = 12.3$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 8.5$
$Var[X_{t-1|t-1}] = 2.1$

Prediction step:
$x_t = x_{t-1} + w,\ \ w: W \sim \mathcal{N}(0,\sigma_W^2)$
$y_t = x_t + v,\ \ v: V \sim \mathcal{N}(0,\sigma_V^2)$

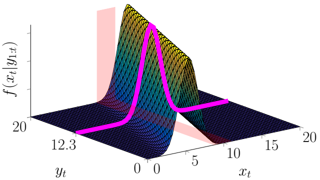
$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 8.5$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 8.5$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 2.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 6.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 2.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1},Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.61$

Update step:
$E[X_{t|t}] = E[X_{t|t-1}] + \frac{Cov[X_{t|t-1},Y_{t|t-1}](y_t - E[Y_{t|t-1}])}{Var[Y_{t|t-1}]}$
$= 9.9$
$Var[X_{t|t}] = Var[X_{t|t-1}] - \frac{Cov[X_{t|t-1},Y_{t|t-1}]^2}{Var[Y_{t|t-1}]} = 1.5$





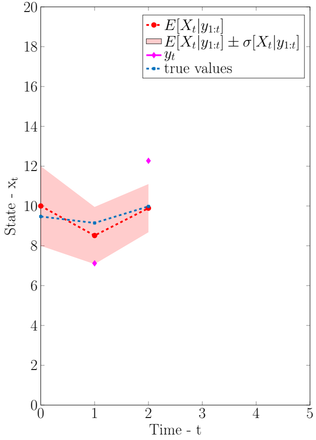
$E[X_t|y_{1:t}]$ =9.9 $\sigma[X_t|y_{1:t}]$ =1.2



[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 3 \mid y_t = 9$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.9$
$Var[X_{t-1|t-1}] = 1.5$



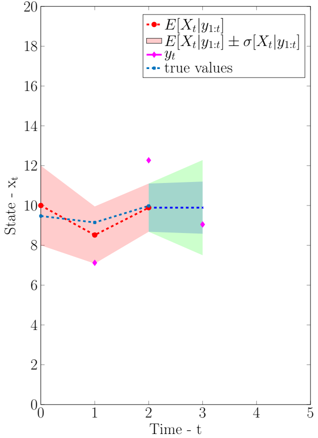
[CIV_ML/KF_example_LL.m]

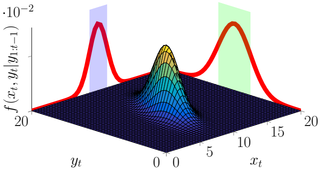
Example #1 – Temperature time-series [.m]

$t = 3 \mid y_t = 9$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.9$
$Var[X_{t-1|t-1}] = 1.5$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.9$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.9$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.7$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.7$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.7$
$\rho_{XY} = \frac{Cov[X_{t|t-1},Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.55$





[CIV_ML/KF_example_LL.m]

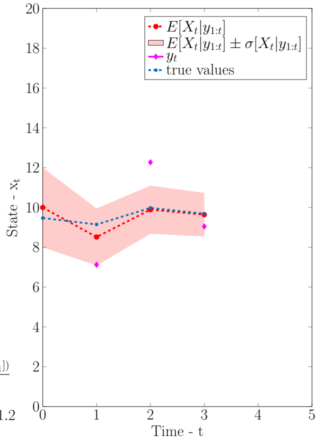
Example #1 – Temperature time-series [.m]

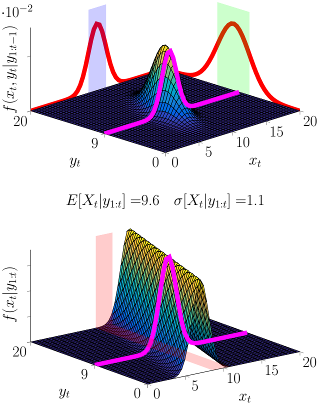
$t=3 \mid y_t = 9$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.9$
$Var[X_{t-1|t-1}] = 1.5$

Prediction step:
$x_t = x_{t-1} + w,\ w : W \sim \mathcal{N}(0,\sigma_W^2)$
$y_t = x_t + v,\ v : V \sim \mathcal{N}(0,\sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.9$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.9$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.7$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.7$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.7$
$\rho_{XY} = \frac{Cov[X_{t|t-1},Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.55$

Update step:
$E[X_{t|t}] = E[X_{t|t-1}] + \frac{Cov[X_{t|t-1},Y_{t|t-1}](y_t - E[Y_{t|t-1}])}{Var[Y_{t|t-1}]}$
$= 9.6$
$Var[X_{t|t}] = Var[X_{t|t-1}] - \frac{Cov[X_{t|t-1},Y_{t|t-1}]^2}{Var[Y_{t|t-1}]} = 1.2$

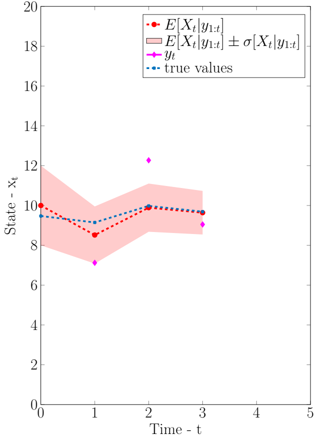




[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 4 \mid y_t = 7.6$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.6$
$Var[X_{t-1|t-1}] = 1.2$



[CIV_ML/KF_example_LL.m]

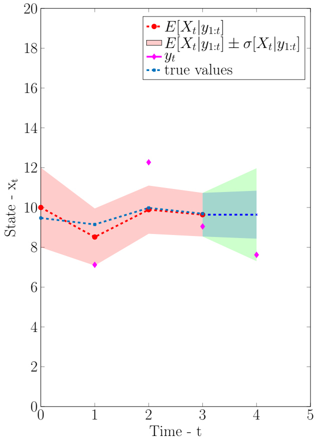
Example #1 – Temperature time-series [.m]

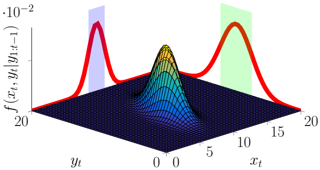
$t = 4 \mid y_t = 7.6$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.6$
$Var[X_{t-1|t-1}] = 1.2$

Prediction step:

$x_t = x_{t-1} + w,\ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.6$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.6$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.5$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.5$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.5$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.52$





[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 4 \mid y_t = 7.6$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.6$
$Var[X_{t-1|t-1}] = 1.2$

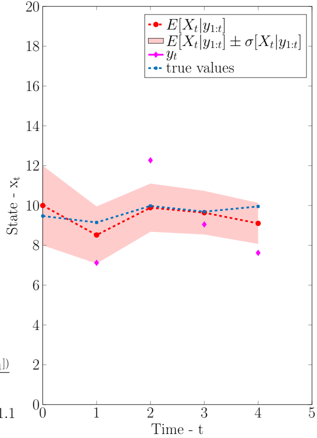
Prediction step:

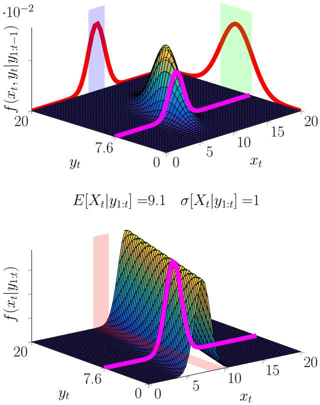
$x_t = x_{t-1} + w, \;\; w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v, \;\; v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.6$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.6$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.5$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.5$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.5$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.52$

Update step:

$E[X_{t|t}] = E[X_{t|t-1}] + \frac{Cov[X_{t|t-1}, Y_{t|t-1}](y_t - E[Y_{t|t-1}])}{Var[Y_{t|t-1}]}$
$= 9.1$
$Var[X_{t|t}] = Var[X_{t|t-1}] - \frac{Cov[X_{t|t-1}, Y_{t|t-1}]^2}{Var[Y_{t|t-1}]} = 1.1$

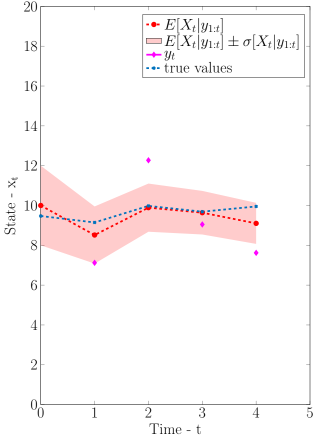




[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 5 \mid y_t = 10.2$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.1$
$Var[X_{t-1|t-1}] = 1.1$



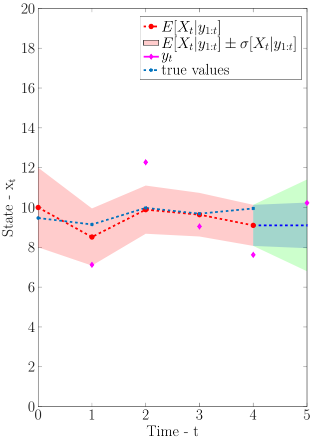
[CIV_ML/KF_example_LL.m]

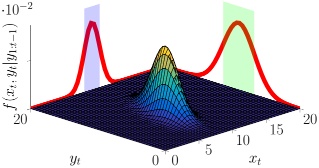
Example #1 – Temperature time-series [.m]

$t = 5 \mid y_t = 10.2$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.1$
$Var[X_{t-1|t-1}] = 1.1$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.1$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.1$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.5$





[CIV_ML/KF_example_LL.m]

Example #1 – Temperature time-series [.m]

$t = 5 \mid y_t = 10.2$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 9.1$
$Var[X_{t-1|t-1}] = 1.1$

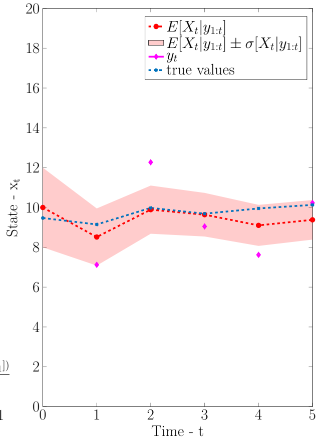
Prediction step:

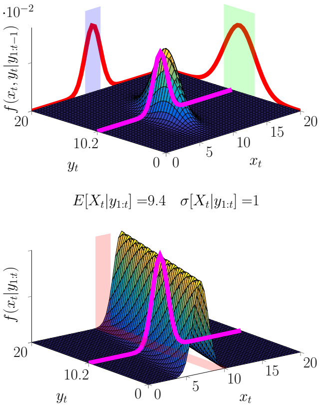
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 9.1$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 9.1$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 5.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1},Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.5$

Update step:

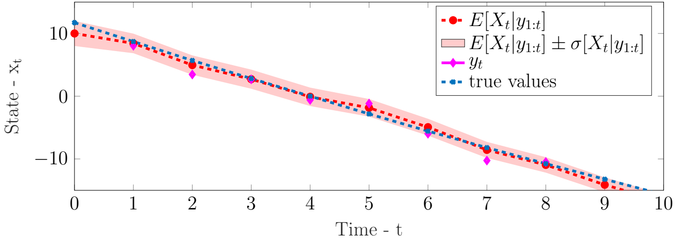
$E[X_{t|t}] = E[X_{t|t-1}] + \frac{Cov[X_{t|t-1},Y_{t|t-1}](y_t - E[Y_{t|t-1}])}{Var[Y_{t|t-1}]}$
$= 9.4$
$Var[X_{t|t}] = Var[X_{t|t-1}] - \frac{Cov[X_{t|t-1},Y_{t|t-1}]^2}{Var[Y_{t|t-1}]} = 1$

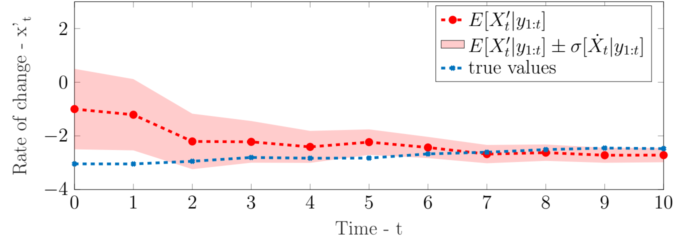




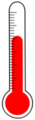
[CIV_ML/KF_example_LL.m]

Example #2 – Temperature time-series (constant speed)





Example #2 – Temperature time-series (constant speed)



We want to estimate $\mathbf{x}_t = [x_t, \dot{x}_t]^\intercal$, the temperature and its rate of change over a period of 5 hours using imprecise measurements made every hour, $t = 0:5$.

$$\underbrace{y_t = x_t + v_t \equiv \overbrace{\mathbf{C}}^{\mathbf{C}=[1,0]}\mathbf{x}_t + v_t}_{\text{observation model}}$$

$$\underbrace{\left.\begin{array}{rcr} x_t & = & x_{t-1} + \Delta t \dot{x}_{t-1} + w_t \\ \dot{x}_t & = & \dot{x}_{t-1} + \dot{w}_t \end{array}\right\} \equiv \mathbf{x}_t = \overbrace{\mathbf{A}}^{\mathbf{A}=\left[\begin{array}{cc} 1 & \Delta t \\ 0 & 1 \end{array}\right]}\mathbf{x}_{t-1} + \overbrace{\mathbf{w}_t}^{\mathbf{w}_t=\left[\begin{array}{c} w_t \\ \dot{w}_t \end{array}\right]}}_{\text{transition model}}$$

We need a **general formulation** applicable for **any linear observation/transition model**

$$\begin{array}{rcl} \mathbf{y}_t & = & \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \\ \mathbf{x}_t & = & \mathbf{A}\mathbf{x}_{t-1} + \mathbf{w}_t \end{array}$$

General formulation – Prediction step: $t-1 \rightarrow t$

We compute our **prior knowledge** at t from our **posterior knowledge** at $t-1$.

Transition model: $\underline{\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{w}_t}$, $\mathbf{w}_t : \mathbf{W}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$

Observation model: $\underline{\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t}$, $\mathbf{v}_t : \mathbf{V}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$

$$
\begin{aligned}
\boldsymbol{\mu}_{t|t-1} &= \mathbf{A}\boldsymbol{\mu}_{t-1|t-1} \\
\boldsymbol{\Sigma}_{t|t-1} &= \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^{\intercal} + \mathbf{Q} \\
\mathbb{E}[\mathbf{Y}_t|y_{1:t-1}] &= \mathbf{C}\boldsymbol{\mu}_{t|t-1} \\
\mathrm{Cov}[\mathbf{Y}_t|y_{1:t-1}] &= \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\intercal} + \mathbf{R} \\
\mathrm{Cov}[\mathbf{X}_t, \mathbf{Y}_t|y_{1:t-1}] &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\intercal}
\end{aligned}
$$

General formulation – Update step: $t-1 \rightarrow t$

Joint prior knowledge: $f(\mathbf{x}_t, \mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{t|t-1} \\ \mathbb{E}[\mathbf{Y}_t|\mathbf{y}_{1:t-1}] \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{XY}} \\ \boldsymbol{\Sigma}_{\mathbf{YX}} & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{t|t-1} & \mathrm{Cov}[\mathbf{X}_t, \mathbf{Y}_t|\mathbf{y}_{1:t-1}] \\ \mathrm{Cov}[\mathbf{X}_t, \mathbf{Y}_t|\mathbf{y}_{1:t-1}]^{\intercal} & \mathrm{Cov}[\mathbf{Y}_t|\mathbf{y}_{1:t-1}] \end{bmatrix}$$

Posterior knowledge:

$$f(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{f(\mathbf{x}_t, \mathbf{y}_t|\mathbf{y}_{1:t-1})}{f(\mathbf{y}_t|\mathbf{y}_{1:t-1})} = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$$

Posterior mean and variance:

$$\begin{aligned} \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{Y}}) \\ \boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}} \end{aligned}$$

General formulation – Kalman Filter classic equations

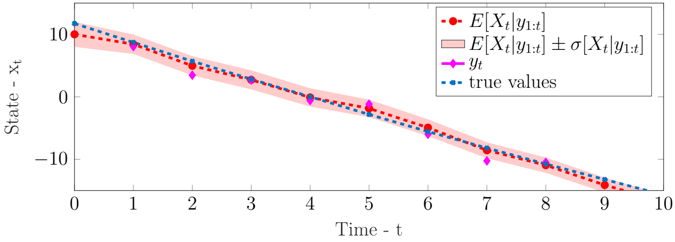
Prediction step

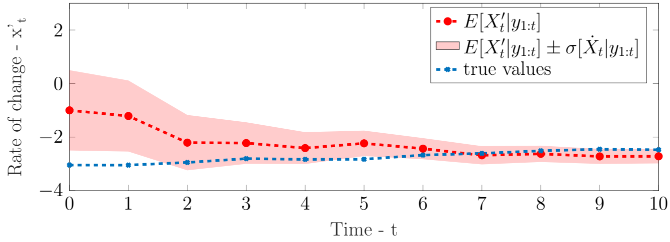
$$
\begin{aligned}
\boldsymbol{\mu}_{t|t-1} &= \mathbf{A}\boldsymbol{\mu}_{t-1|t-1}\\
\boldsymbol{\Sigma}_{t|t-1} &= \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^{\intercal} + \mathbf{Q}
\end{aligned}
$$

Measurement step

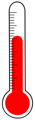
$$
\begin{aligned}
f(\mathbf{x}_t|\mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) & \\
\boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\mathbf{r}_t & \text{Posterior expected value}\\
\boldsymbol{\Sigma}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{C})\boldsymbol{\Sigma}_{t|t-1} & \text{Posterior covariance}\\
\mathbf{r}_t &= \mathbf{y}_t - \hat{\mathbf{y}}_t & \text{Innovation vector}\\
\hat{\mathbf{y}}_t &= \mathbf{C}\boldsymbol{\mu}_{t|t-1} & \text{Prediction observations vector}\\
\mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\intercal}\mathbf{G}_t^{-1} & \text{Kalman gain matrix}\\
\mathbf{G}_t &= \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\intercal} + \mathbf{R} & \text{Innovation covariance matrix}
\end{aligned}
$$

Example #2 – Temperature time-series (constant speed)





Example #2 – Temperature time-series (constant speed)



We want to estimate $\mathbf{x}_t = [x_t, \dot{x}_t]^\intercal$, the temperature and its rate of change over a period of 5 hours using imprecise measurements made every hour, $t = 0:5$.

$$\underbrace{y_t = x_t + v_t \equiv \overbrace{\mathbf{C}}^{[1,0]}\mathbf{x}_t + \overbrace{v_t}^{V\sim\mathcal{N}(0,\mathbf{R})}}_{\text{observation model}}$$

$$\underbrace{\left.\begin{array}{rcr} x_t & = & x_{t-1} + \Delta t \dot{x}_{t-1} + w_t \\ \dot{x}_t & = & \dot{x}_{t-1} + \dot{w}_t \end{array}\right\} \equiv \overbrace{\mathbf{x}_t}^{\left[\begin{array}{c} x_t \\ \dot{x}_t \end{array}\right]} = \overbrace{\mathbf{A}}^{\left[\begin{array}{cc} 1 & \Delta t \\ 0 & 1 \end{array}\right]}\mathbf{x}_{t-1} + \overbrace{\mathbf{w}_t}^{\overbrace{\left[\begin{array}{c} w_t \\ \dot{w}_t \end{array}\right]}^{\mathbf{W}\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}}}_{\text{transition model}}$$

Example #2 – Temperature time-series (constant speed)

$$\underbrace{y_t = \overbrace{\mathbf{C}}^{[1,0]}\mathbf{x}_t + \overbrace{v_t}^{V\sim\mathcal{N}(0,\mathbf{R})}}_{\text{observation model}}$$

$$\underbrace{\overbrace{\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix}}^{}\!\!\mathbf{x}_t = \overbrace{\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}}^{}\!\!\mathbf{A}\,\mathbf{x}_{t-1} + \overbrace{\begin{bmatrix} w_t \\ \dot{w}_t \end{bmatrix}}^{\mathbf{W}\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}\!\!\mathbf{w}_t}_{\text{transition model}}$$

Prior knowledge

$\mathbb{E}[X_0] = 10\,°C, \sigma[X_0] = 2\,°C$

$\mathbb{E}[\dot{X}_0] = -1\,°C, \sigma[\dot{X}_0] = 0.5\,°C$

$X_0 \perp\!\!\!\perp \dot{X}_0 \rightarrow \rho_{X_0\dot{X}_0} = 0$

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 10 \\ -1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 2^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}$$

Observation errors

$$\mathbf{R} = \sigma_V^2 = 2^2$$

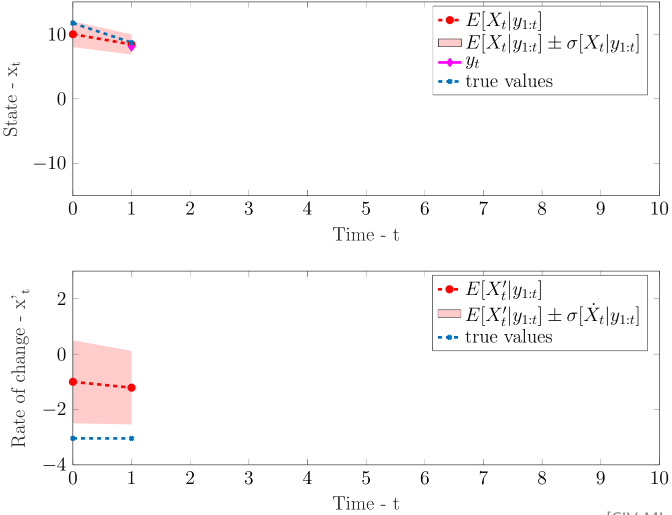
Model errors

$$\mathbf{Q} = \sigma_w^2 \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \quad \sigma_w^2 = 0.1$$

Example #2 – Temperature time-series (const. speed) []

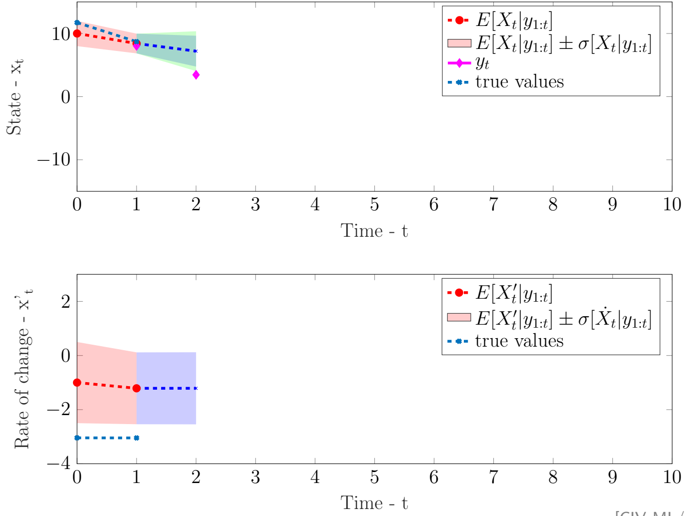
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



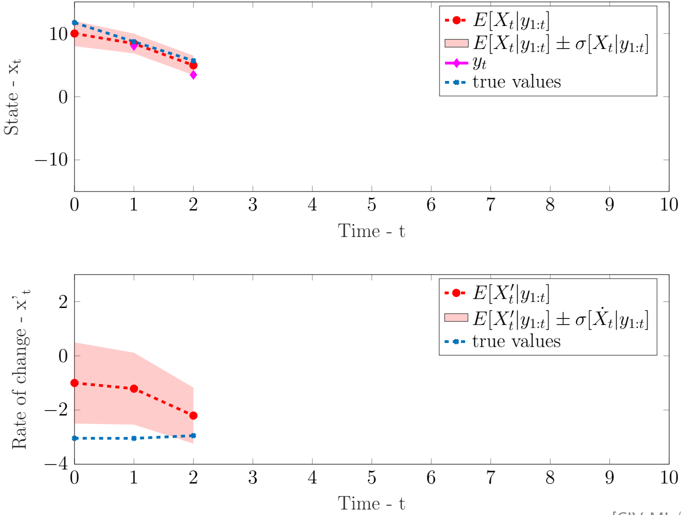
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



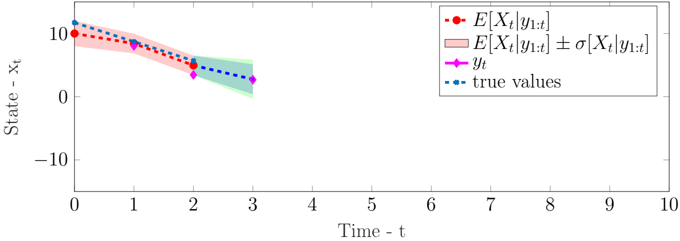
[CIV_ML/KF_example_LT.py]

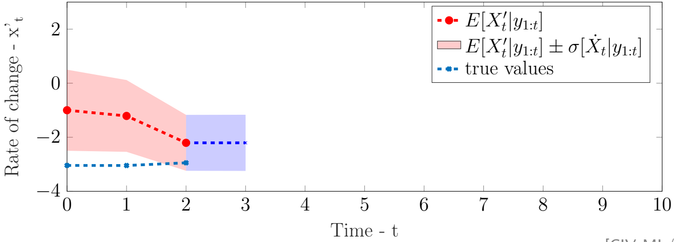
Example #2 – Temperature time-series (const. speed) [ ]



[CIV_ML/KF_example_LT.py]

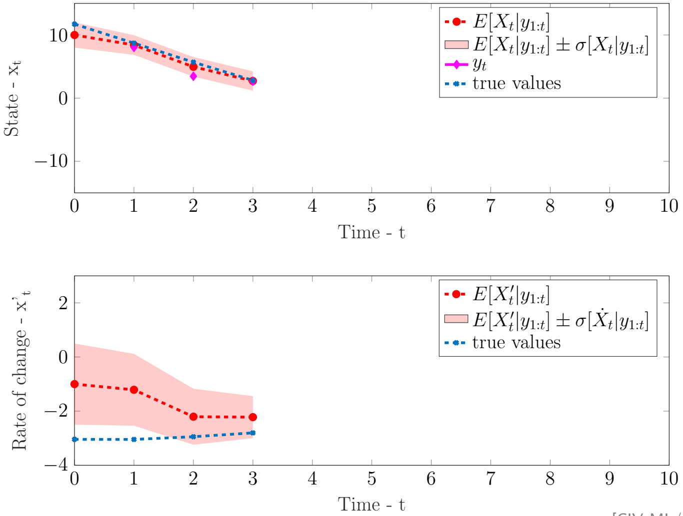
Example #2 – Temperature time-series (const. speed) [ ]





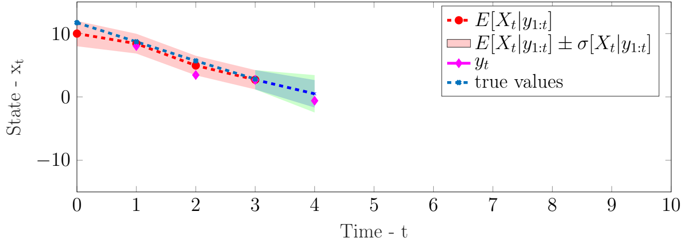
[CIV_ML/KF_example_LT.py]

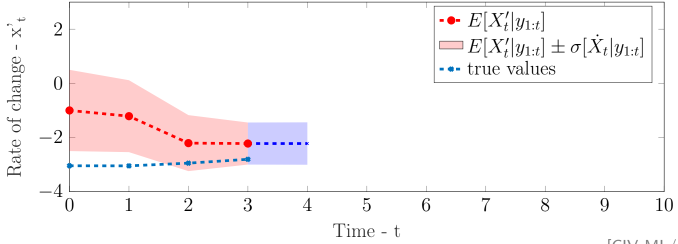
Example #2 – Temperature time-series (const. speed) [ ]



[CIV_ML/KF_example_LT.py]

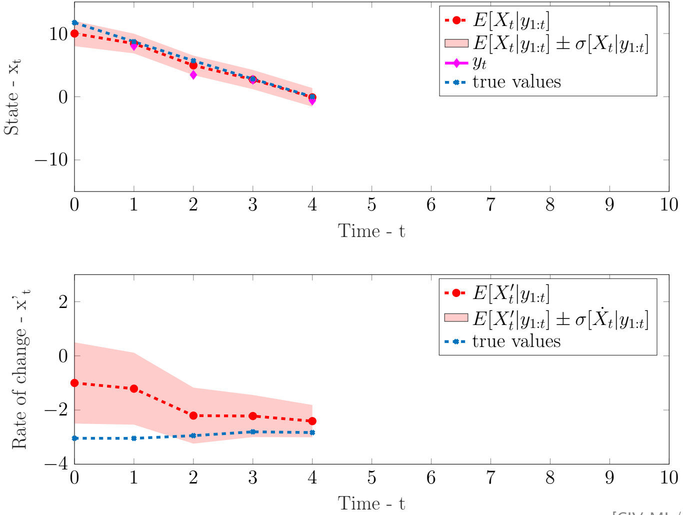
Example #2 – Temperature time-series (const. speed) [ ]





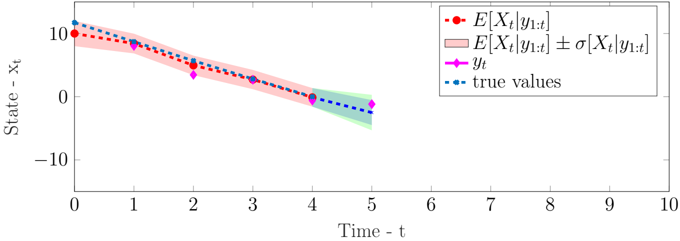
[CIV_ML/KF_example_LT.py]

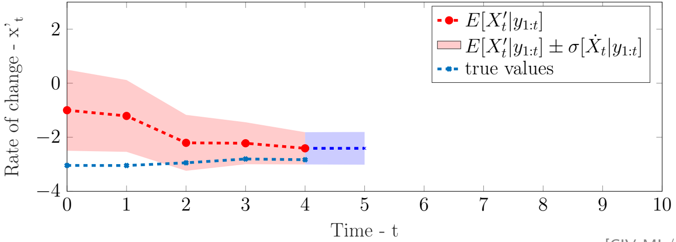
Example #2 – Temperature time-series (const. speed) [ ]



[CIV_ML/KF_example_LT.py]

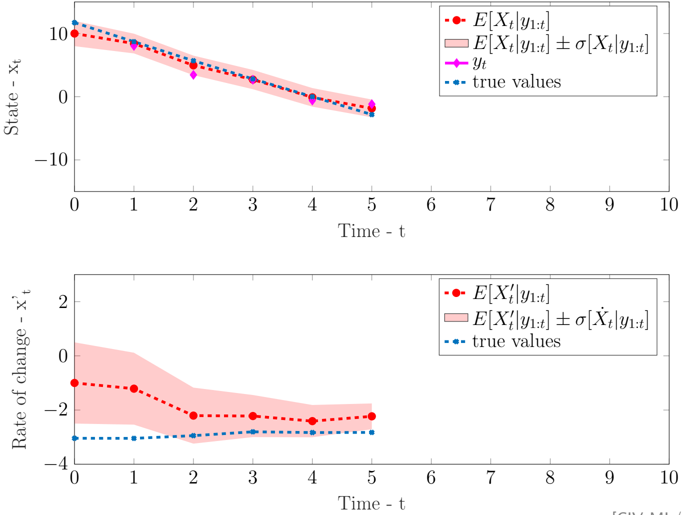
Example #2 – Temperature time-series (const. speed) [ ]





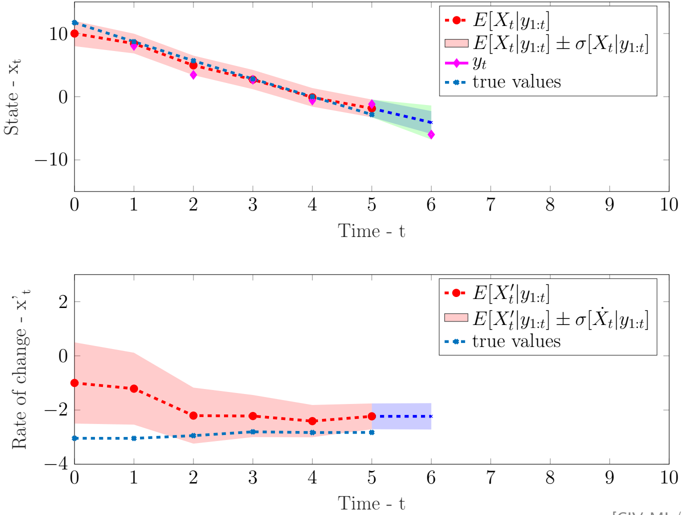
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



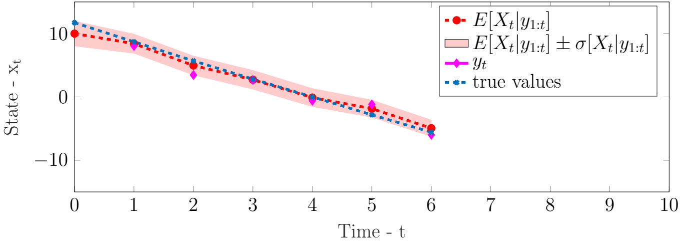
[CIV_ML/KF_example_LT.py]

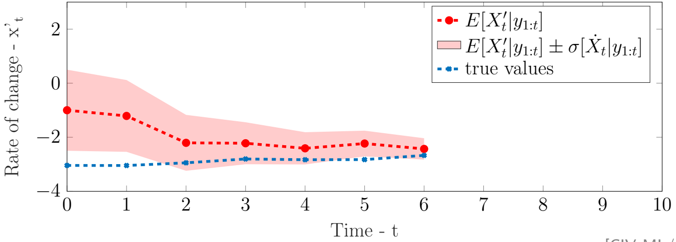
Example #2 – Temperature time-series (const. speed) [ ]



[CIV_ML/KF_example_LT.py]

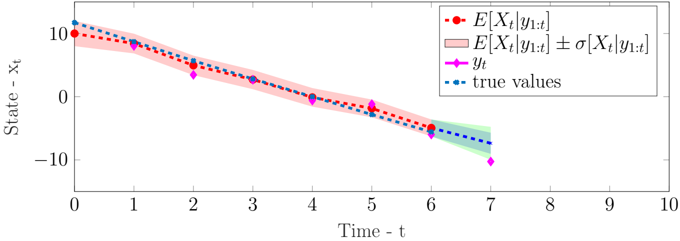
Example #2 – Temperature time-series (const. speed) [🐍]

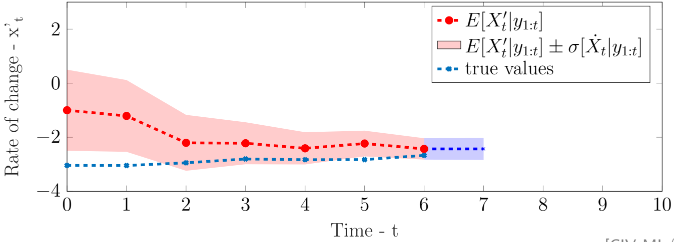




[CIV_ML/KF_example_LT.py]

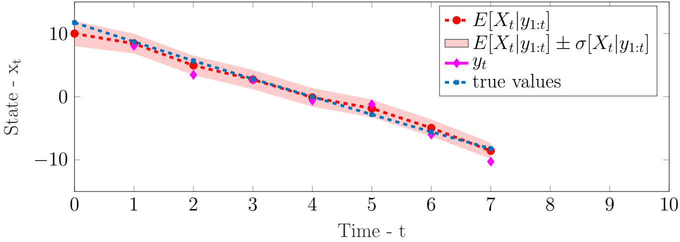
Example #2 – Temperature time-series (const. speed) [ ]

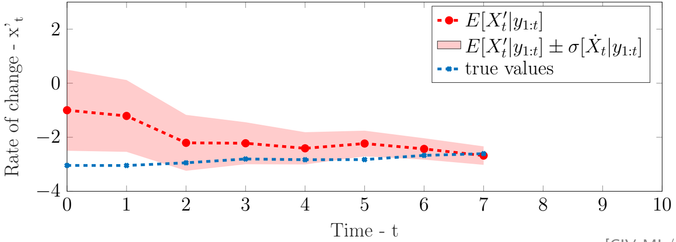




[CIV_ML/KF_example_LT.py]

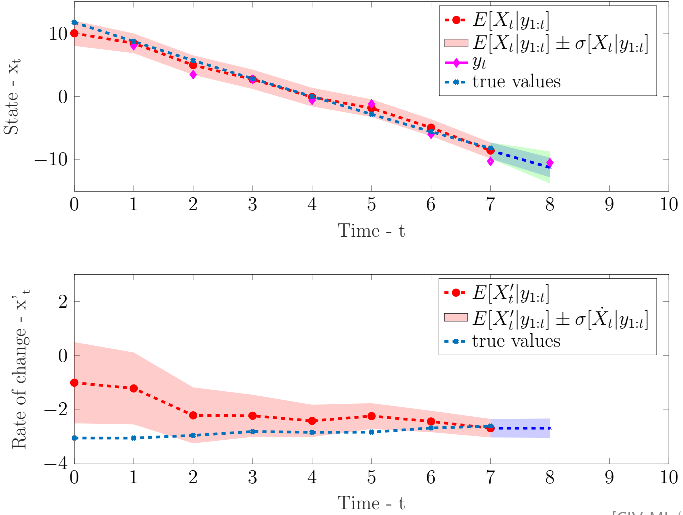
Example #2 – Temperature time-series (const. speed) [ ]





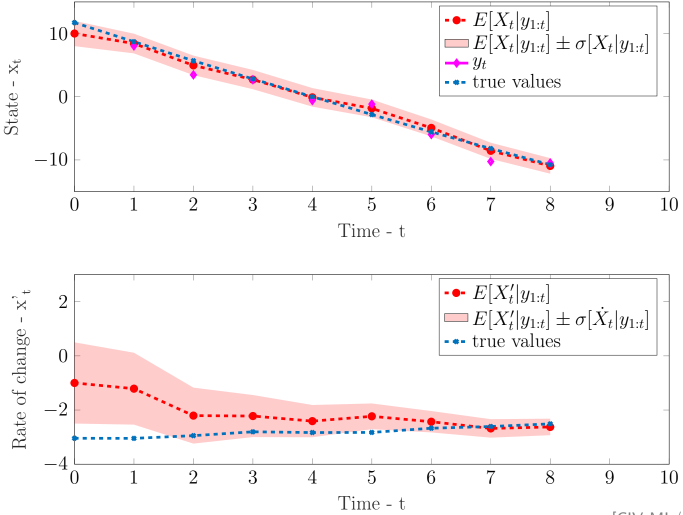
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



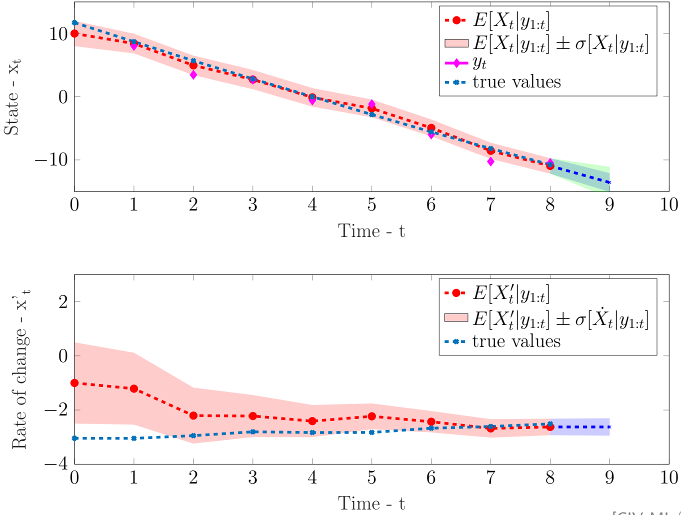
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) []



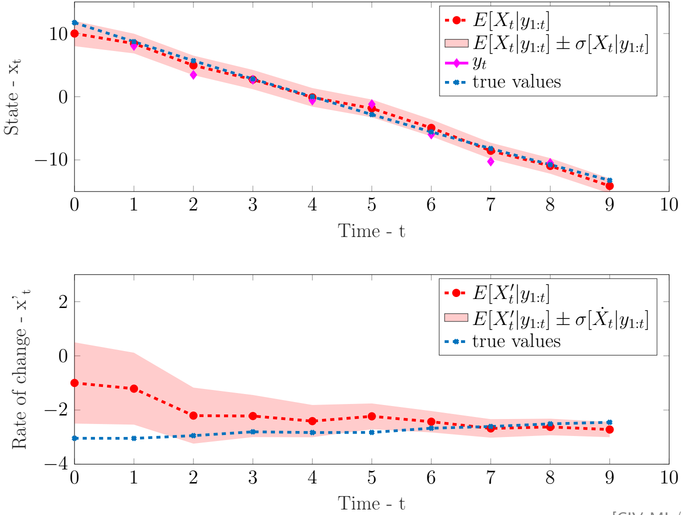
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



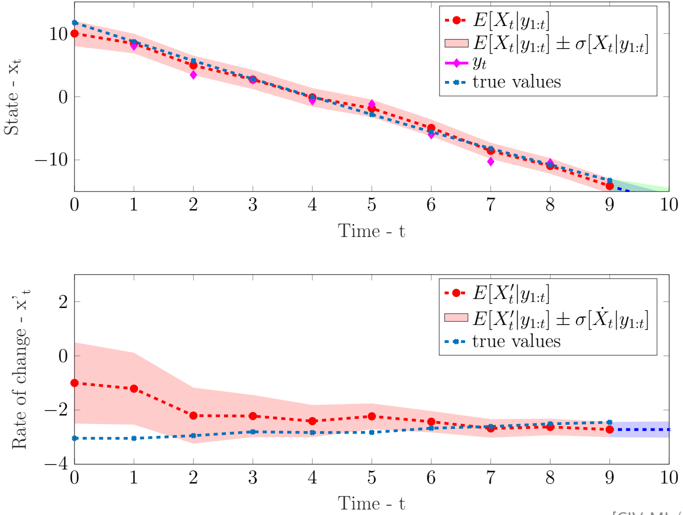
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



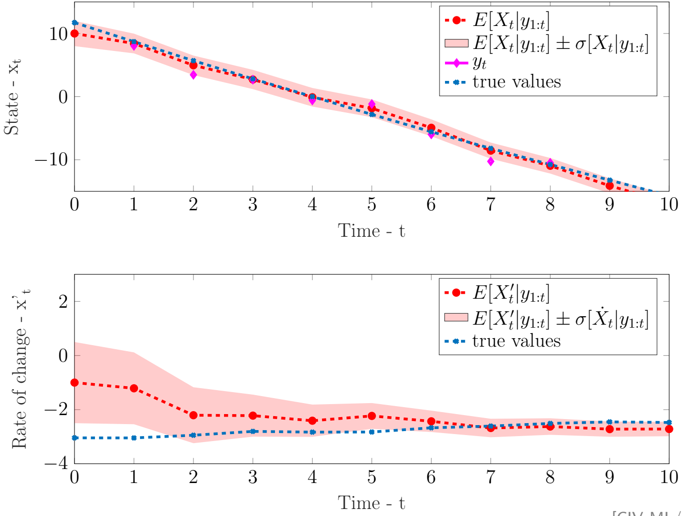
[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [🐍]



[CIV_ML/KF_example_LT.py]

Example #2 – Temperature time-series (const. speed) [ ]



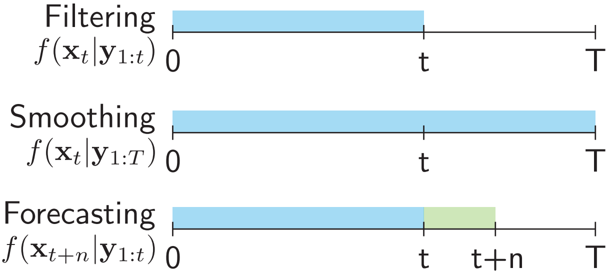
[CIV_ML/KF_example_LT.py]

Section Outline

$t+n|t, t|T$

3.1 Introduction to forecasting and smoothing
3.2 Forecasting: $f(\mathbf{x}_{t+n}|\mathbf{y}_{1:t})$
3.3 Smoothing: $f(\mathbf{x}_t|\mathbf{y}_{1:T})$

Filtering/Smoothing/Forecasting



Forecasting: $f(\mathbf{x}_{t+n}|\mathbf{y}_{1:t})$

Prediction step

$$
\begin{aligned}
f(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}) \\
\boldsymbol{\mu}_{t+1|t} &= \mathbf{A}\boldsymbol{\mu}_{t|t} \\
\boldsymbol{\Sigma}_{t+1|t} &= \mathbf{A}\boldsymbol{\Sigma}_{t|t}\mathbf{A}^{\intercal} + \mathbf{Q}
\end{aligned}
$$

When **forecasting**, there is no update step

When **data is missing** at a time step,
the new estimate $\equiv$ **forecasting**

Example #2 (cont.) – Forecasting temperature

$$\underbrace{y_t = \overbrace{\mathbf{C}}^{[1,0]}\mathbf{x}_t + \overbrace{v_t}^{V\sim\mathcal{N}(0,\mathbf{R})}}_{\text{observation model}}$$

$$\underbrace{\overbrace{\mathbf{x}_t}^{\left[\begin{array}{c}x_t\\ \dot{x}_t\end{array}\right]} = \overbrace{\mathbf{A}}^{\left[\begin{array}{cc}1 & \Delta t\\ 0 & 1\end{array}\right]}\mathbf{x}_{t-1} + \overbrace{\mathbf{w}_t}^{\overbrace{\left[\begin{array}{c}w_t\\ \dot{w}_t\end{array}\right]}^{\mathbf{W}\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}}}_{\text{transition model}}$$

Model errors

$$\mathbf{Q} = \sigma_w^2\left[\begin{array}{cc}\frac{\Delta t^3}{3} & \frac{\Delta t^2}{2}\\ \frac{\Delta t^2}{2} & \Delta t\end{array}\right], \quad \sigma_w^2 = 0.1$$

Prior knowledge

$$\mathbb{E}[X_0] = 10\,^\circ C, \sigma[X_0] = 2\,^\circ C$$
$$\mathbb{E}[\dot{X}_0] = 0\,^\circ C, \sigma[\dot{X}_0] = 0.5\,^\circ C$$
$$X_0 \perp\!\!\!\perp \dot{X}_0 \rightarrow \rho_{X_0\dot{X}_0} = 0$$

$$\boldsymbol{\mu}_0 = \left[\begin{array}{c}10\\ 0\end{array}\right], \quad \boldsymbol{\Sigma}_0 = \left[\begin{array}{cc}2^2 & 0\\ 0 & 0.5^2\end{array}\right]$$

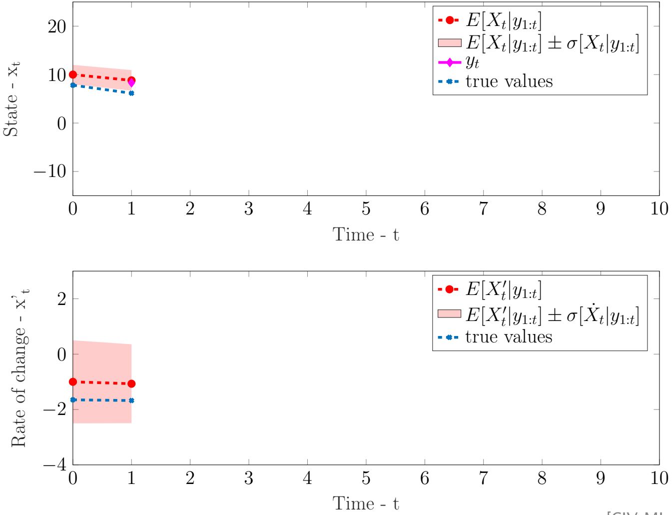
Observation errors

$$\mathbf{R} = \sigma_V^2 = 2^2$$

Example #2 (cont.) – Forecasting temperature []

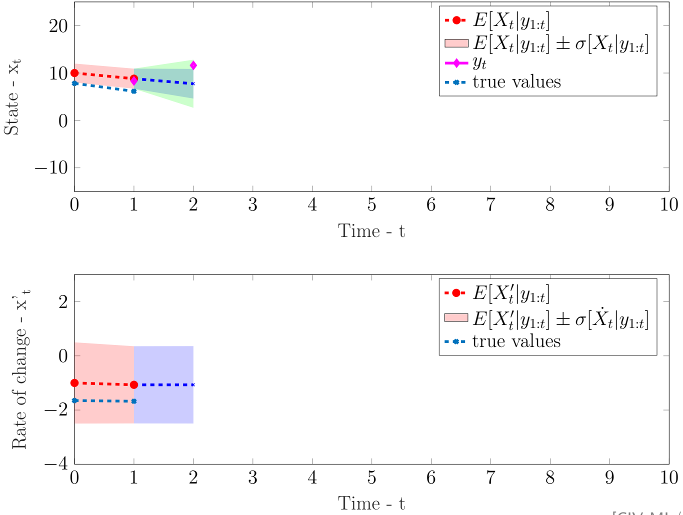
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



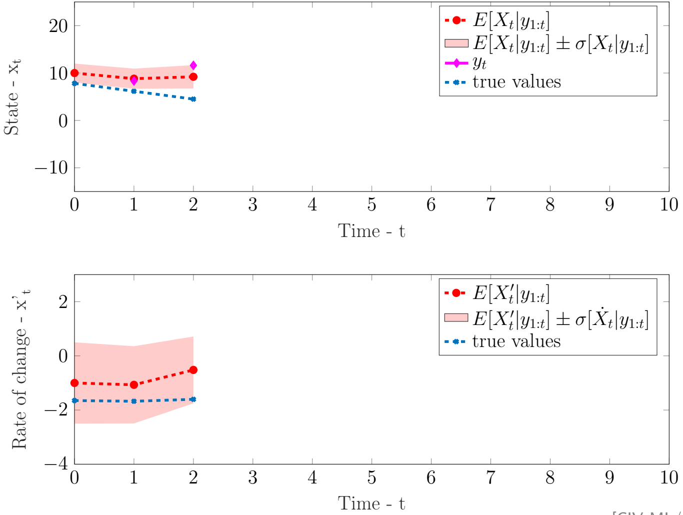
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



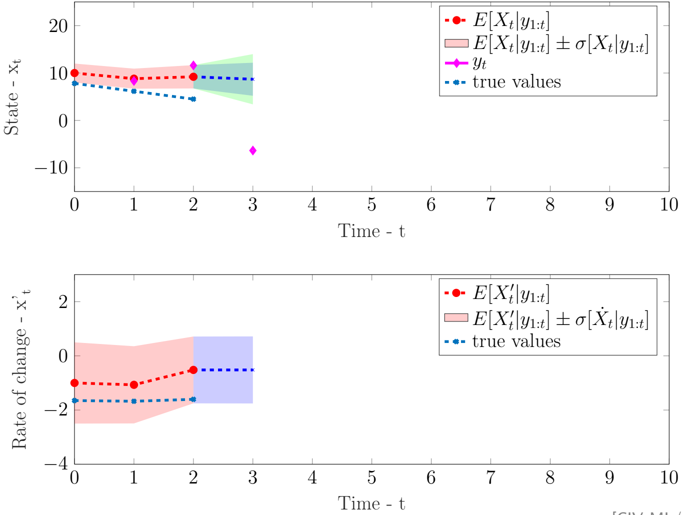
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



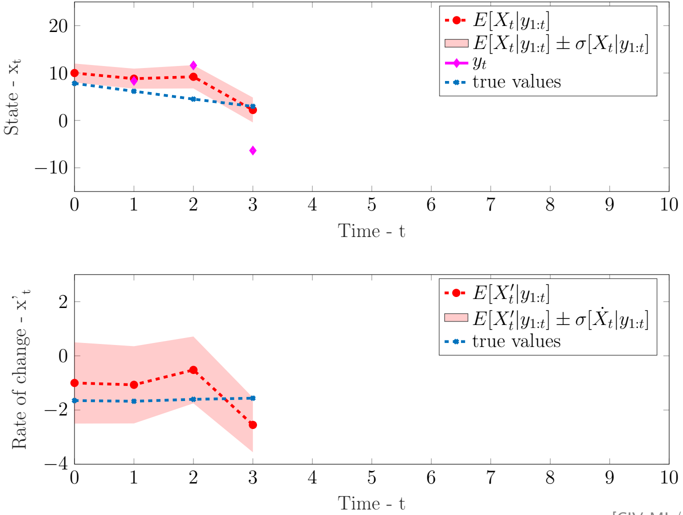
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



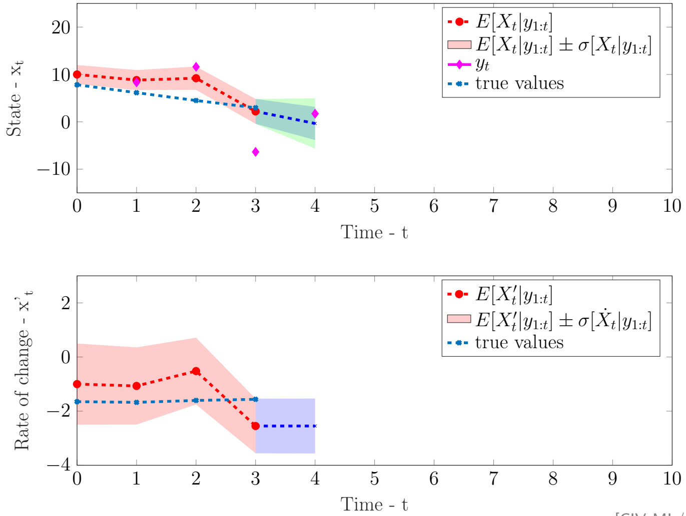
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



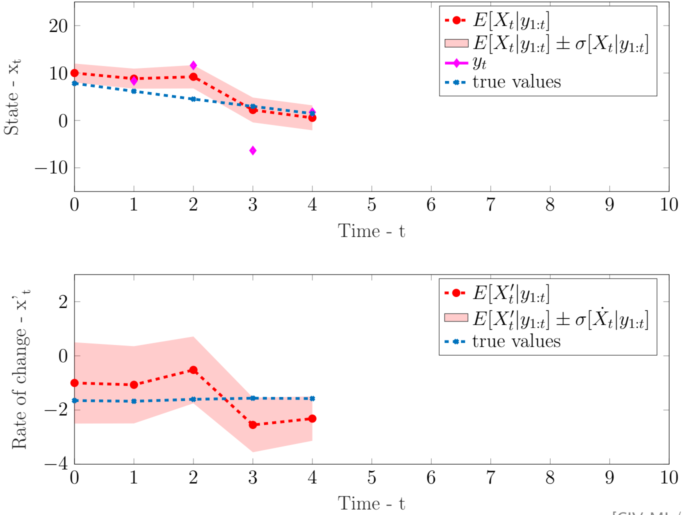
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



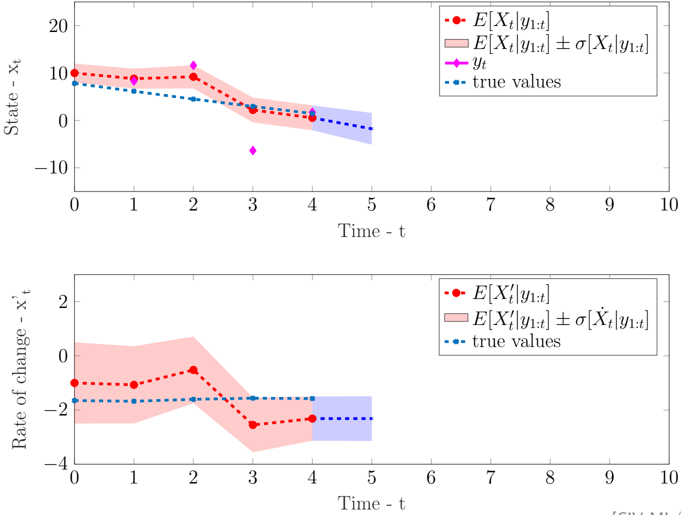
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



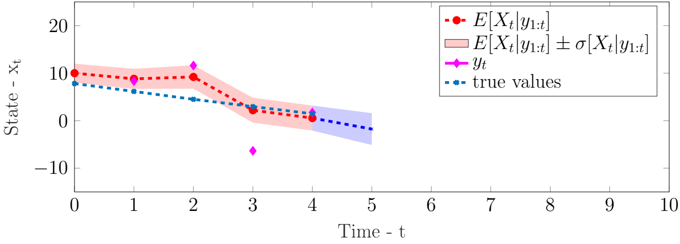
[CIV_ML/KF_example_LT.py]

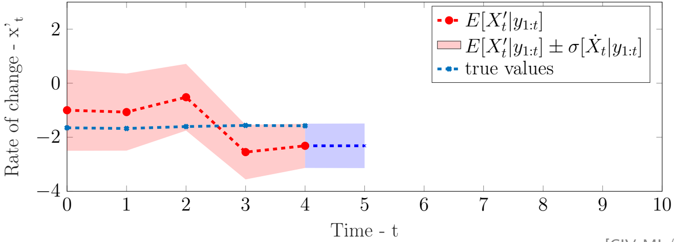
Example #2 (cont.) – Forecasting temperature [ ]



[CIV_ML/KF_example_LT.py]

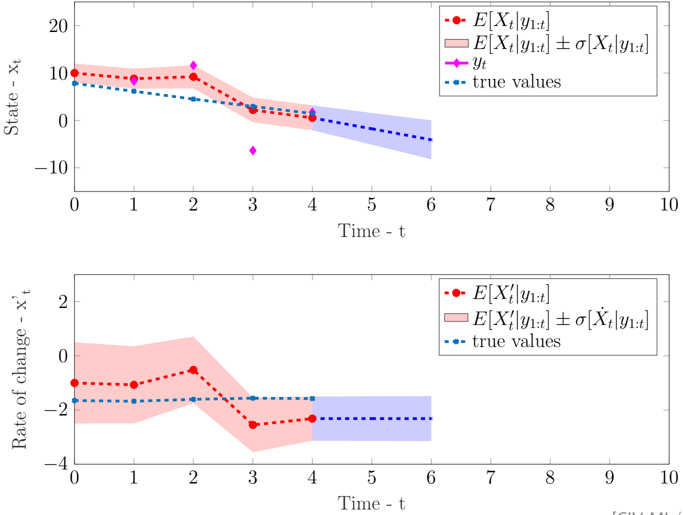
Example #2 (cont.) – Forecasting temperature [ ]





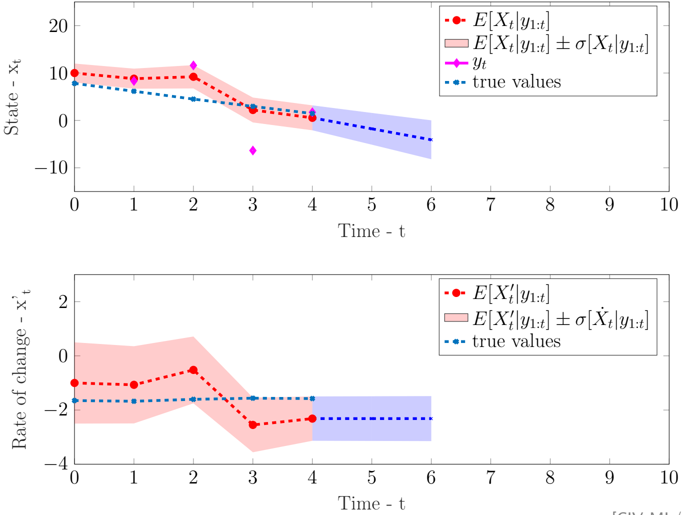
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [Python]



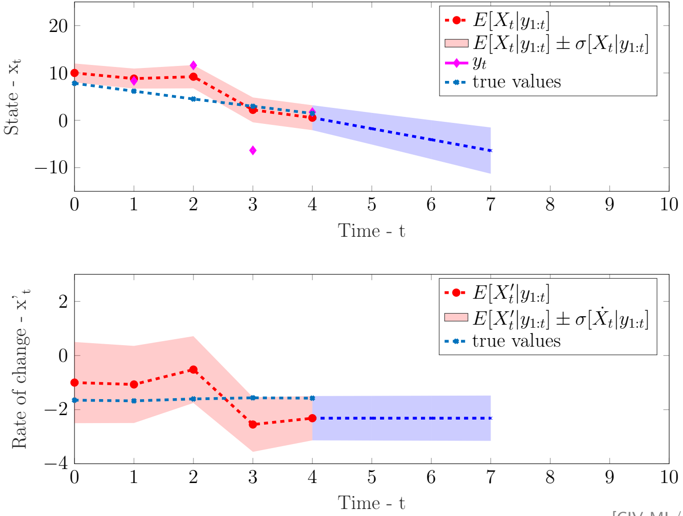
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



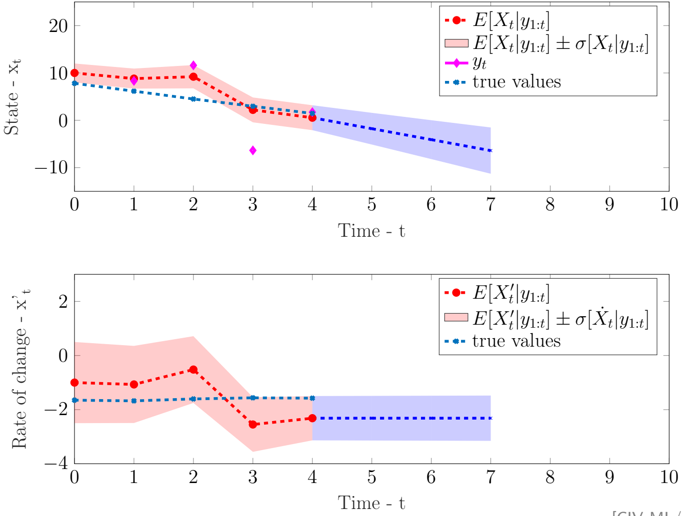
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



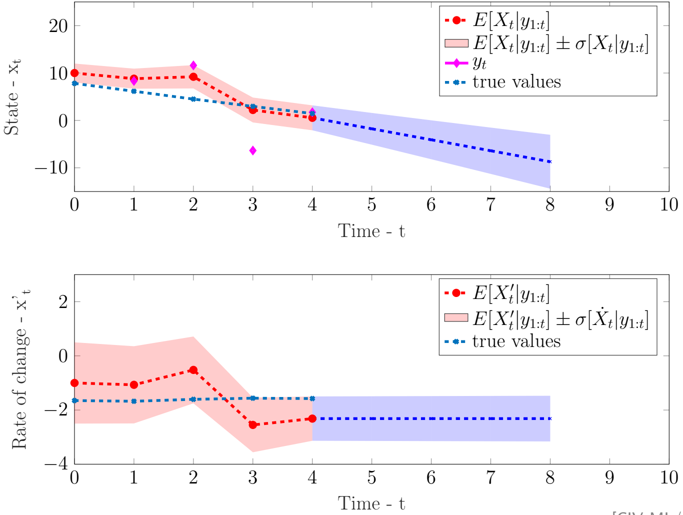
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



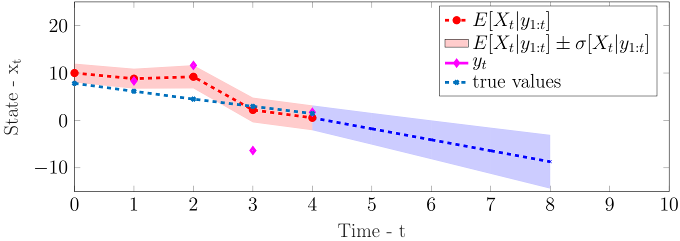
[CIV_ML/KF_example_LT.py]

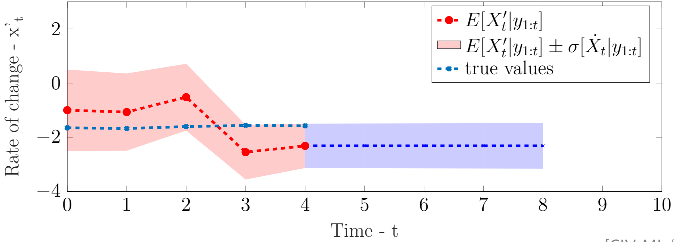
Example #2 (cont.) – Forecasting temperature [ ]



[CIV_ML/KF_example_LT.py]

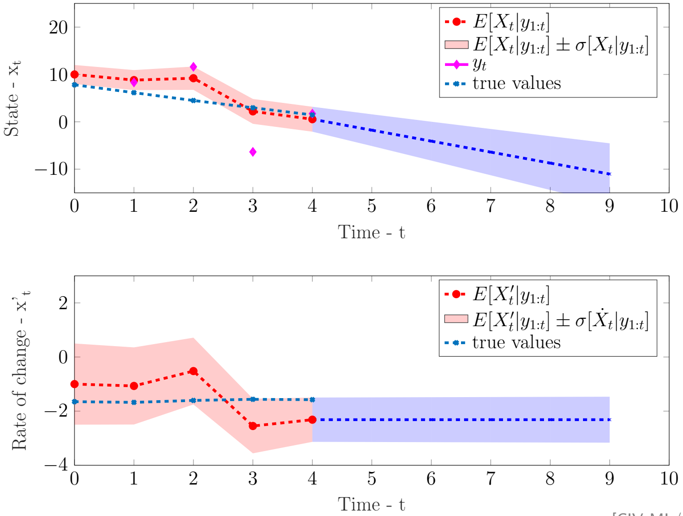
Example #2 (cont.) – Forecasting temperature [ ]





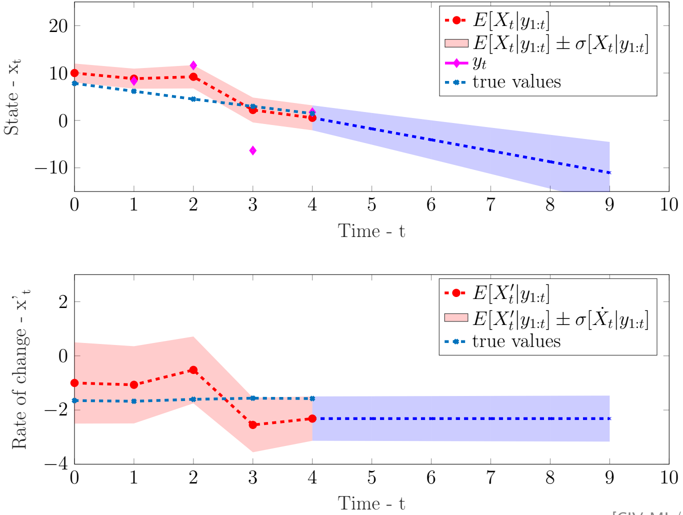
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



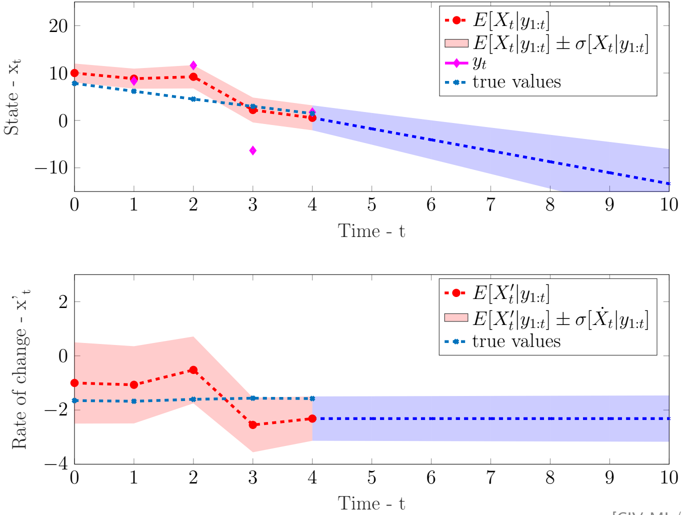
[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



[CIV_ML/KF_example_LT.py]

Example #2 (cont.) – Forecasting temperature [ ]



[CIV_ML/KF_example_LT.py]

Smoothing: $f(\mathbf{x}_t|\mathbf{y}_{1:T})$

$$\begin{array}{lccccccccccc}
t & 0 & & 1 & & & & t & & & & T-1 & & T \\
\text{Filter:} & \begin{matrix}\boldsymbol{\mu}_0\\ \boldsymbol{\Sigma}_0\end{matrix} & \rightarrow & \begin{matrix}\boldsymbol{\mu}_{1|1}\\ \boldsymbol{\Sigma}_{1|1}\end{matrix} & \rightarrow & \cdots & \rightarrow & \begin{matrix}\boldsymbol{\mu}_{t|t}\\ \boldsymbol{\Sigma}_{t|t}\end{matrix} & \rightarrow & \cdots & \rightarrow & \begin{matrix}\boldsymbol{\mu}_{T-1|T-1}\\ \boldsymbol{\Sigma}_{T-1|T-1}\end{matrix} & \rightarrow & \begin{matrix}\boldsymbol{\mu}_{T|T}\\ \boldsymbol{\Sigma}_{T|T}\end{matrix} \\
 & & & & & & & & & & & & \swarrow & \\
\text{Smooth:} & \begin{matrix}\boldsymbol{\mu}_{0|T}\\ \boldsymbol{\Sigma}_{0|T}\end{matrix} & \leftarrow & \begin{matrix}\boldsymbol{\mu}_{1|T}\\ \boldsymbol{\Sigma}_{1|T}\end{matrix} & \leftarrow & \cdots & \leftarrow & \begin{matrix}\boldsymbol{\mu}_{t|T}\\ \boldsymbol{\Sigma}_{t|T}\end{matrix} & \leftarrow & \cdots & \leftarrow & \begin{matrix}\boldsymbol{\mu}_{T-1|T}\\ \boldsymbol{\Sigma}_{T-1|T}\end{matrix} & &
\end{array}$$

$\boldsymbol{\mu}_{T|T}$ and $\boldsymbol{\Sigma}_{T|T}$ is treated as an **observation**
to **update** $\boldsymbol{\mu}_{T-1|T-1}$ and $\boldsymbol{\Sigma}_{T-1|T-1}$
leading to the **smoothed estimate** $\boldsymbol{\mu}_{T-1|T}$ and $\boldsymbol{\Sigma}_{T-1|T}$

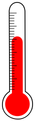
General formulation – RTS Kalman Smoother equations

$$\begin{array}{lcccccccccc}
t & 0 & & 1 & & t & & T-1 & & T \\
\text{Filter:} & \begin{matrix}\boldsymbol{\mu}_0\\ \boldsymbol{\Sigma}_0\end{matrix} & \longrightarrow & \begin{matrix}\boldsymbol{\mu}_{1|1}\\ \boldsymbol{\Sigma}_{1|1}\end{matrix} & \rightarrow \cdots \rightarrow & \begin{matrix}\boldsymbol{\mu}_{t|t}\\ \boldsymbol{\Sigma}_{t|t}\end{matrix} & \rightarrow \cdots \rightarrow & \begin{matrix}\boldsymbol{\mu}_{T-1|T-1}\\ \boldsymbol{\Sigma}_{T-1|T-1}\end{matrix} & \longrightarrow & \begin{matrix}\boldsymbol{\mu}_{T|T}\\ \boldsymbol{\Sigma}_{T|T}\end{matrix} \\
\text{Smooth:} & \begin{matrix}\boldsymbol{\mu}_{0|T}\\ \boldsymbol{\Sigma}_{0|T}\end{matrix} & \longleftarrow & \begin{matrix}\boldsymbol{\mu}_{1|T}\\ \boldsymbol{\Sigma}_{1|T}\end{matrix} & \leftarrow \cdots \leftarrow & \begin{matrix}\boldsymbol{\mu}_{t|T}\\ \boldsymbol{\Sigma}_{t|T}\end{matrix} & \leftarrow \cdots \leftarrow & \begin{matrix}\boldsymbol{\mu}_{T-1|T}\\ \boldsymbol{\Sigma}_{T-1|T}\end{matrix} & \swarrow &
\end{array}$$

RTS Kalman Smoother

$$\begin{aligned}
f(\mathbf{x}_t|\mathbf{y}_{1:T}) &= \mathcal{N}(\boldsymbol{\mu}_{t|T}, \boldsymbol{\Sigma}_{t|T}) \\
\boldsymbol{\mu}_{t|T} &= \boldsymbol{\mu}_{t|t} + \mathbf{J}_t\left(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}\right) \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t} + \mathbf{J}_t\left(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{t+1|t}\right)\mathbf{J}_t^{\intercal} \\
\mathbf{J}_t &\triangleq \boldsymbol{\Sigma}_{t|t}\mathbf{A}^{\intercal}\boldsymbol{\Sigma}_{t+1|t}^{-1}
\end{aligned}$$

Example #1 (cont.) – Smoothing temperature



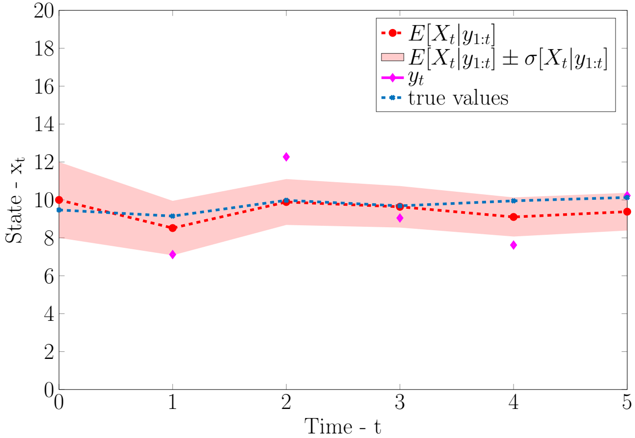
We want to estimate the temperature x_t over a period of 5 hours using imprecise measurements made every hour, $t = 0 : 5$.

Our prior knowledge is that $\mathbb{E}[X_0] = 10\,°C$ and $\sigma[X_0] = 7\,°C$.

$$\underbrace{y_t = x_t + v_t}_{\text{observation model}}, \quad \overbrace{v_t : V \sim \mathcal{N}(0, 1^2)}^{\text{observation error}}$$

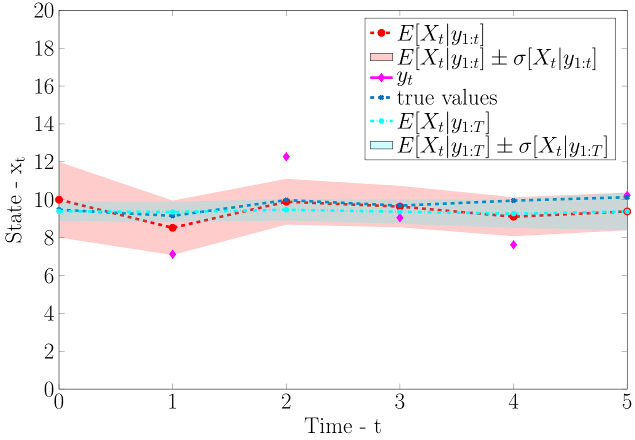
$$\underbrace{x_t = x_{t-1} + w_t}_{\text{transition model}}, \quad \overbrace{w_t : W \sim \mathcal{N}(0, 0.5^2)}^{\text{model error}}$$

Example #1 (cont.) – Smoothing temperature [ ]



[CIV_ML/KF_example_LT.py]

Example #1 (cont.) – Smoothing temperature [ ]



[CIV_ML/KF_example_LT.py]

Section Outline

θ estimation

4.1 Measuring model's performance: Log-likelihood
4.2 Newton-Raphson algorithm

Measuring a model performance

How do we measure the performance of a model?

With the **marginal prior probability of each observation**.

$$f(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{y}_t; \mathbf{C}\boldsymbol{\mu}_{t|t-1}, \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^\intercal + \mathbf{R})$$

Likelihood

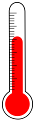
$$f(\mathbf{y}_{1:T}) = \prod_{t=1}^{T} f(\mathbf{y}_t|\mathbf{y}_{1:t-1}) \qquad [\text{note: } \mathbf{y}_i \perp\!\!\!\perp \mathbf{y}_j, \forall i \neq j]$$

Since $f(\mathbf{y}_t|\mathbf{y}_{1:t-1}) \ll 1$, $\prod_{t=1}^{T} f(\mathbf{y}_t|\mathbf{y}_{1:t-1}) \approx 0$ ⚠

Log-likelihood

$$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{1:t-1})$$ ✓

Example #1 (cont.) – Log-likelihood



We want to estimate the temperature x_t over a period of 5 hours using imprecise measurements made every hour, $t = 0:5$.

Our prior knowledge is that $\mathbb{E}[X_0] = 10\,°C$ and $\sigma[X_0] = 7\,°C$.

$$\underbrace{y_t = x_t + v_t}_{\text{observation model}}, \quad \overbrace{v_t : V \sim \mathcal{N}(0, 1^2)}^{\text{observation error}}$$

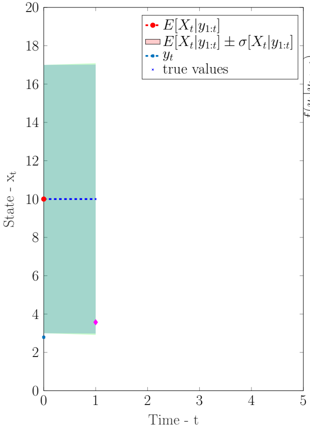
$$\underbrace{x_t = x_{t-1} + w_t}_{\text{transition model}}, \quad \overbrace{w_t : W \sim \mathcal{N}(0, 0.5^2)}^{\text{model error}}$$

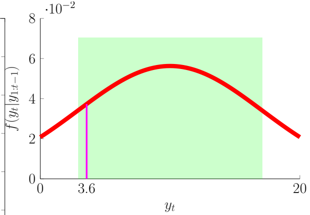
Example #1 (cont.) – Log-likelihood [.m]

$t = 1 \mid y_t = 3.6$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 10$
$Var[X_{t-1|t-1}] = 49$

Prediction step:
$x_t = x_{t-1} + w, \;\; w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v, \;\; v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 10$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 10$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 49.3$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 50.3$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 49.3$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.99$





Log-likelihood:
$f(\mathbf{y}_t|\mathbf{y}_{t-1}) = 0.04$
$\log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -3.29$
$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -3.29$

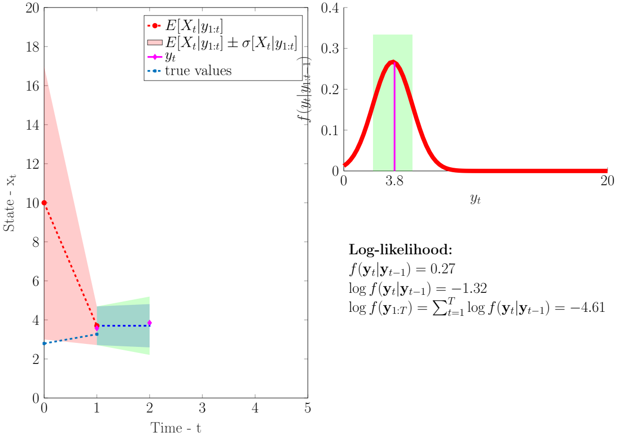
[CIV_ML/KF_example_LL.m]

Example #1 (cont.) – Log-likelihood [.m]

$t = 2 \mid y_t = 3.8$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 3.7$
$Var[X_{t-1|t-1}] = 1$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 3.7$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 3.7$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 1.2$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 2.2$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 1.2$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.74$



Log-likelihood:
$f(\mathbf{y}_t|\mathbf{y}_{t-1}) = 0.27$
$\log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -1.32$
$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -4.61$

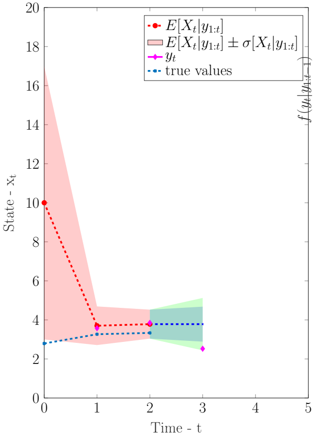
[CIV_ML/KF_example_LL.m]

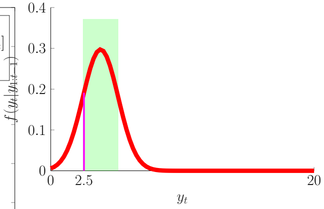
Example #1 (cont.) – Log-likelihood [.m]

$t = 3 \mid y_t = 2.5$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 3.8$
$Var[X_{t-1|t-1}] = 0.6$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 3.8$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 3.8$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 0.8$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 1.8$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 0.8$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.67$





Log-likelihood:
$f(\mathbf{y}_t|\mathbf{y}_{t-1}) = 0.19$
$\log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -1.65$
$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -6.27$

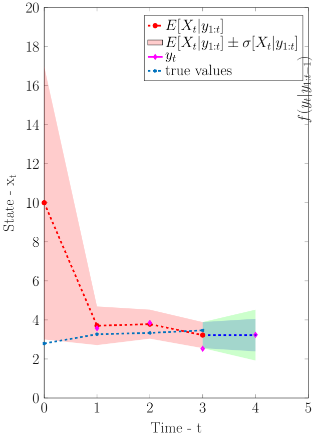
[CIV_ML/KF_example_LL.m]

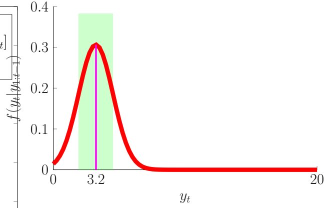
Example #1 (cont.) – Log-likelihood [.m]

$t = 4 \mid y_t = 3.2$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 3.2$
$Var[X_{t-1|t-1}] = 0.4$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 3.2$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 3.2$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 0.7$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 1.7$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 0.7$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.64$





Log-likelihood:
$f(\mathbf{y}_t|\mathbf{y}_{t-1}) = 0.31$
$\log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -1.18$
$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -7.45$

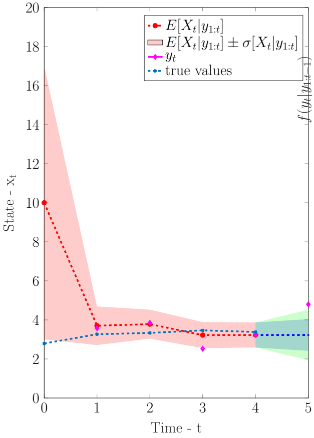
[CIV_ML/KF_example_LL.m]

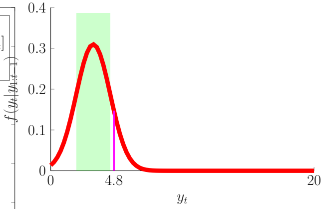
Example #1 (cont.) – Log-likelihood [.m]

$t = 5 \mid y_t = 4.8$
$E[X_{t-1|t-1}] = E[X_{t-1}|y_{1:t-1}] = 3.2$
$Var[X_{t-1|t-1}] = 0.4$

Prediction step:
$x_t = x_{t-1} + w,\ \ w : W \sim \mathcal{N}(0, \sigma_W^2)$
$y_t = x_t + v,\ \ v : V \sim \mathcal{N}(0, \sigma_V^2)$

$E[X_{t|t-1}] = E[X_{t-1|t-1}] = 3.2$
$E[Y_{t|t-1}] = E[X_{t|t-1}] = 3.2$
$Var[X_{t|t-1}] = Var[X_{t-1|t-1}] + \sigma_W^2 = 0.7$
$Var[Y_{t|t-1}] = Var[X_{t|t-1}] + \sigma_V^2 = 1.7$
$Cov[X_{t|t-1}, Y_{t|t-1}] = Var[X_{t|t-1}] = 0.7$
$\rho_{XY} = \frac{Cov[X_{t|t-1}, Y_{t|t-1}]}{\sigma[X_{t|t-1}]\sigma[Y_{t|t-1}])} = 0.63$





Log-likelihood:
$f(\mathbf{y}_t|\mathbf{y}_{t-1}) = 0.15$
$\log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -1.92$
$\log f(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\mathbf{y}_{t-1}) = -9.37$

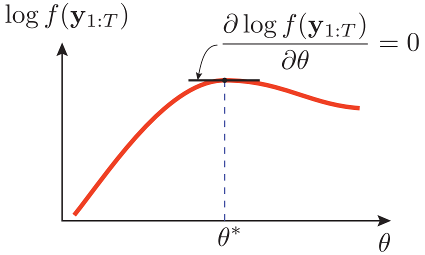
[CIV_ML/KF_example_LL.m]

Learning model parameters $\boldsymbol{\theta}$

How do we learn the parameters $\boldsymbol{\theta} = [\sigma_W, \sigma_V, \ldots]^\intercal$ of a model?

When a large amount of data is available
$\rightarrow$ **Maximum Likelihood Estimate (MLE)**

$$\begin{aligned}\boldsymbol{\theta}^* &= \arg\max_{\boldsymbol{\theta}} \overbrace{\log f(\mathbf{y}_{1:T})}^{\text{log-likelihood}} \\ &\equiv \arg\max_{\boldsymbol{\theta}} \log f(\mathbf{y}_{1:T}|\boldsymbol{\theta})\end{aligned}$$



$$\frac{d}{d\theta_j} \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}) = 0 \rightarrow \textbf{Newton-Raphson}$$

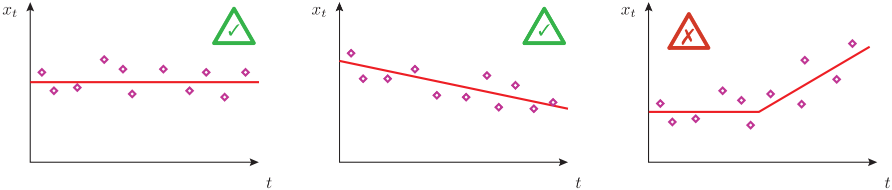
Section Outline

SKF

5.1 Introduction
5.2 Switching Kalman Filter
5.3 SKF example

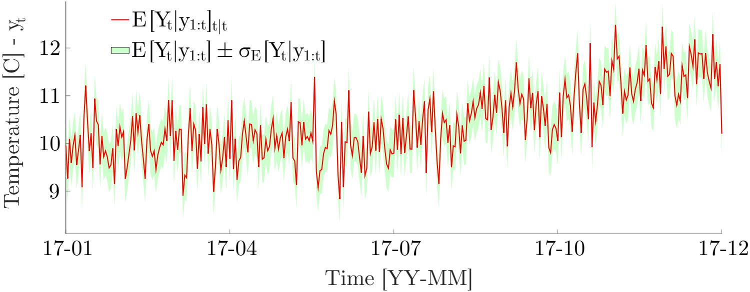
Non-stationary time series

The **Kalman filter and smoother** are applicable when the time-serie analyzed is **stationary** or **trend-stationary**

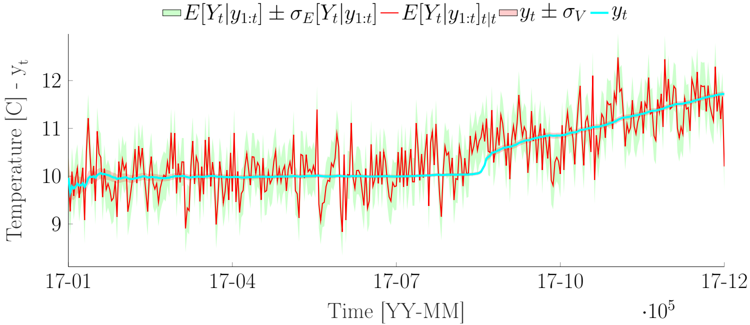


Switching Kalman Filter: Estimate at each time step t, what is the probability of each model structure.

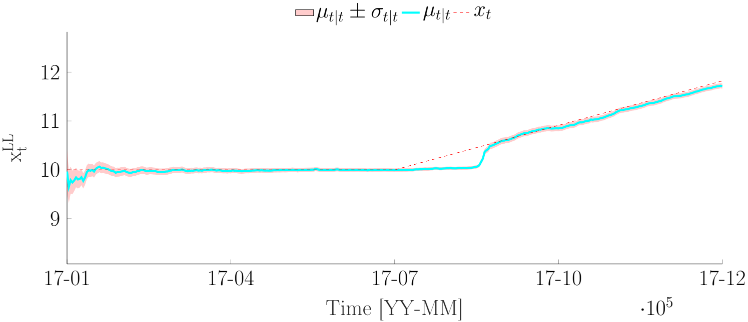
Switching Kalman Filter



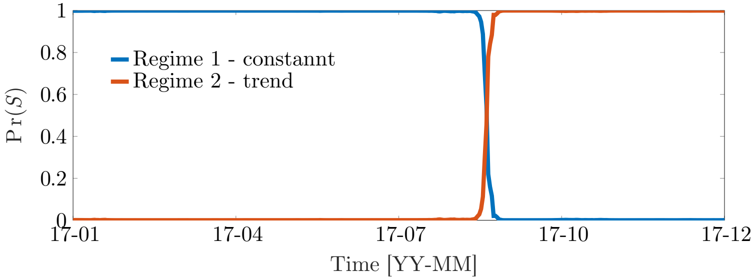
Switching Kalman Filter



Switching Kalman Filter



Switching Kalman Filter



Switching Kalman Filter – Algorithm

t-1

model 1

$$\begin{aligned}
\boldsymbol{\mu}^{1}_{t-1|t-1} &= \mathbb{E}[\mathbf{X}_{t-1}|\mathbf{y}_{1:t-1}, S_{t-1}=1]\\
\boldsymbol{\Sigma}^{1}_{t-1|t-1} &= \mathrm{Cov}[\mathbf{X}_{t-1}\mathbf{X}_{t-1}|\mathbf{y}_{1:t-1}, S_{t-1}=1]\\
\pi^{1}_{t-1|t-1} &= \Pr(S_{t-1}=1|\mathbf{y}_{1:t-1})
\end{aligned}$$

Switching Kalman Filter – Algorithm

t-1

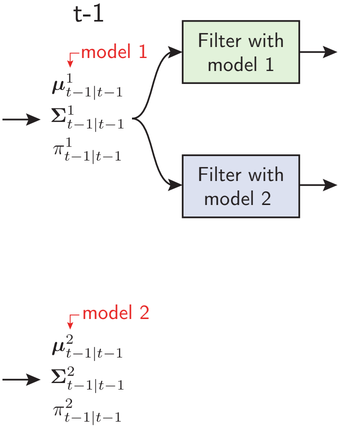
model 1

$$\rightarrow \begin{array}{l} \boldsymbol{\mu}^1_{t-1|t-1} \\ \boldsymbol{\Sigma}^1_{t-1|t-1} \\ \pi^1_{t-1|t-1} \end{array}$$

model 2

$$\rightarrow \begin{array}{l} \boldsymbol{\mu}^2_{t-1|t-1} \\ \boldsymbol{\Sigma}^2_{t-1|t-1} \\ \pi^2_{t-1|t-1} \end{array}$$

Switching Kalman Filter – Algorithm



$$
\begin{aligned}
\boldsymbol{\mu}^{1,1}_{t|t} &= \mathbb{E}[\mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=1] \\
\boldsymbol{\Sigma}^{1,1}_{t|t} &= \mathrm{Cov}[\mathbf{X}_t, \mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=1] \\
\mathcal{L}^{1,1}_{t|t} &= \Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, S_{t-1}=1, S_t=1)
\end{aligned}
$$

Switching Kalman Filter – Algorithm

t-1

model 1

$\boldsymbol{\mu}^1_{t-1|t-1}$
$\boldsymbol{\Sigma}^1_{t-1|t-1}$
$\pi^1_{t-1|t-1}$

Filter with model 1

$$
\begin{aligned}
\boldsymbol{\mu}^{1,1}_{t|t} &= \mathbb{E}[\mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=1]\\
\boldsymbol{\Sigma}^{1,1}_{t|t} &= \mathrm{Cov}[\mathbf{X}_t, \mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=1]\\
\mathcal{L}^{1,1}_{t|t} &= \Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, S_{t-1}=1, S_t=1)
\end{aligned}
$$

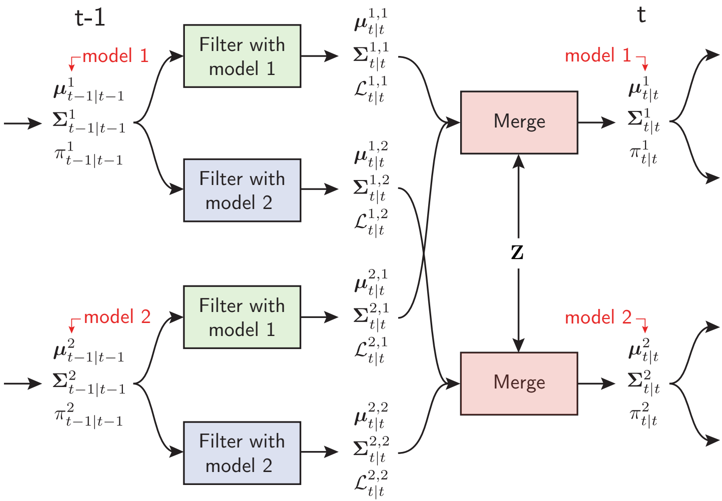
Filter with model 2

$$
\begin{aligned}
\boldsymbol{\mu}^{1,2}_{t|t} &= \mathbb{E}[\mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=2]\\
\boldsymbol{\Sigma}^{1,2}_{t|t} &= \mathrm{Cov}[\mathbf{X}_t, \mathbf{X}_t|\mathbf{y}_{1:t}, S_{t-1}=1, S_t=2]\\
\mathcal{L}^{1,2}_{t|t} &= \Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, S_{t-1}=1, S_t=2)
\end{aligned}
$$

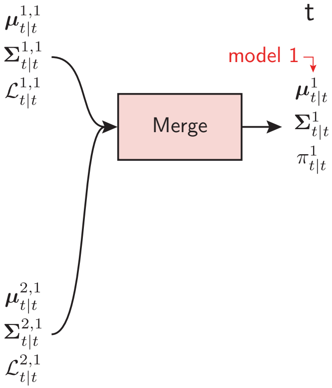
model 2

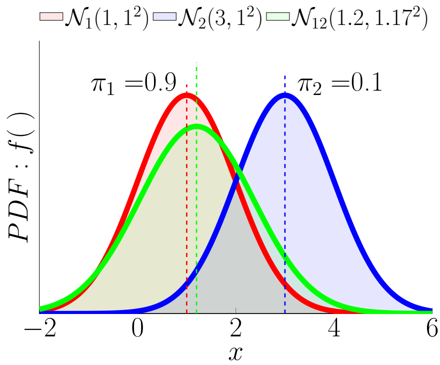
$\boldsymbol{\mu}^2_{t-1|t-1}$
$\boldsymbol{\Sigma}^2_{t-1|t-1}$
$\pi^2_{t-1|t-1}$

Switching Kalman Filter – Algorithm



Switching Kalman Filter – Merge step [ ]





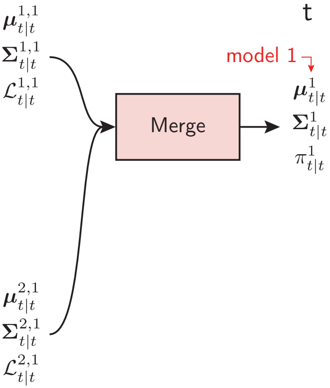
The **merge step** consists in **mixing** several Gaussian random variables

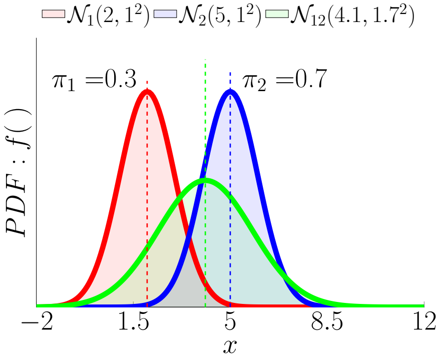
$$\mu_{12} = \pi_1\mu_1 + \pi_2\mu_2$$

$$\sigma_{12}^2 = \pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1 - \mu_2)^2$$

[Murphy (2012), §11.2.1 | CIV_ML/Gaussian_mixture_demo.py]

Switching Kalman Filter – Merge step [Python]



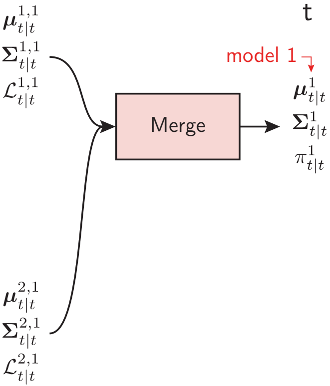


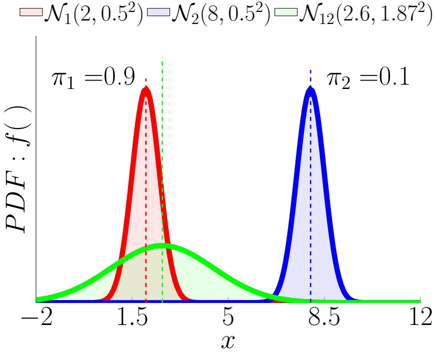
The **merge step** consists in **mixing** several Gaussian random variables

$$
\begin{aligned}
\mu_{12} &= \pi_1\mu_1 + \pi_2\mu_2 \\
\sigma^2_{12} &= \pi_1\sigma^2_1 + \pi_2\sigma^2_2 + \pi_1\pi_2(\mu_1 - \mu_2)^2
\end{aligned}
$$

[Murphy (2012), §11.2.1 | CIV_ML/Gaussian_mixture_demo.py]

Switching Kalman Filter – Merge step [ ]



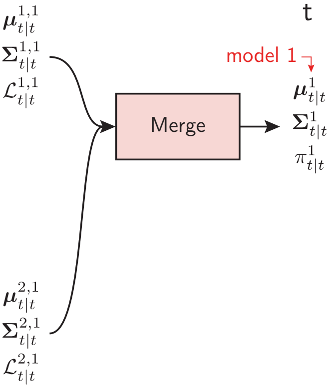


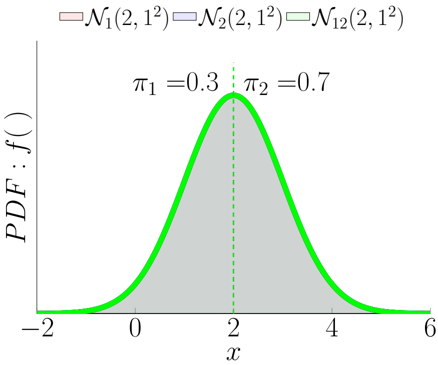
$$
\begin{aligned}
\mu_{12} &= \pi_1\mu_1 + \pi_2\mu_2\\
\sigma_{12}^2 &= \pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1-\mu_2)^2
\end{aligned}
$$

The **merge step** consists in **mixing** several Gaussian random variables

[Murphy (2012), §11.2.1 | CIV_ML/Gaussian_mixture_demo.py]

Switching Kalman Filter – Merge step [🐍]



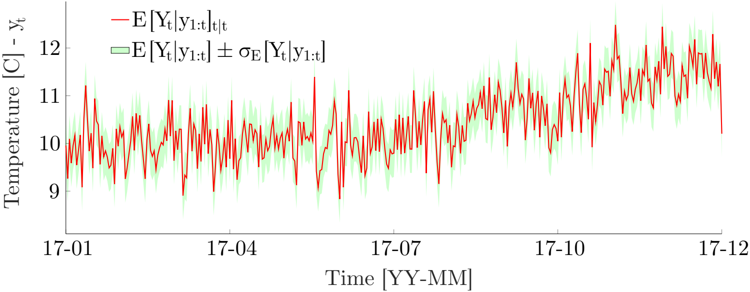


$$
\begin{aligned}
\mu_{12} &= \pi_1\mu_1 + \pi_2\mu_2 \\
\sigma_{12}^2 &= \pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1 - \mu_2)^2
\end{aligned}
$$

The **merge step** consists in **mixing** several Gaussian random variables

[Murphy (2012), §11.2.1 | CIV_ML/Gaussian_mixture_demo.py]

Example #4 – non-stationary temperature data



Example #4 – non-stationary data

$$\underbrace{y_t = \overbrace{\mathbf{C}}^{[1,0]}\mathbf{x}_t + \overbrace{v_t}^{V\sim\mathcal{N}(0,0.5^2)}}_{\text{observation model}}$$

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 10 \\ 0 \end{bmatrix}, \ \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1^2 & 0 \\ 0 & 0.1^2 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 0.99999 & 0.00001 \\ 0.00001 & 0.99999 \end{bmatrix}$$

Model 1: Constant

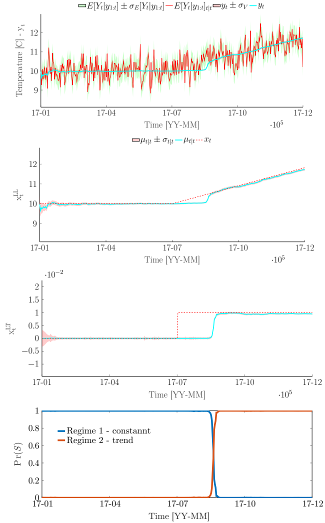
$$\underbrace{\overbrace{\mathbf{x}_t}^{\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix}} = \overbrace{\mathbf{A}}^{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}\mathbf{x}_{t-1} + \overbrace{\mathbf{w}_t}^{\overbrace{\begin{bmatrix} w_t \\ \dot{w}_t \end{bmatrix}}^{\mathbf{W}\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}}}_{\text{transition model}}$$

Model 2: Non-constant

$$\underbrace{\overbrace{\mathbf{x}_t}^{\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix}} = \overbrace{\mathbf{A}}^{\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}}\mathbf{x}_{t-1} + \overbrace{\mathbf{w}_t}^{\overbrace{\begin{bmatrix} w_t \\ \dot{w}_t \end{bmatrix}}^{\mathbf{W}\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}}}_{\text{transition model}}$$

$$\mathbf{Q}_{11} = Q_{22} = \mathbf{Q}_{21} = \mathbf{0}^*, \ \mathbf{Q}_{12} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_W^2 \end{bmatrix}$$

Example #4 – non-stationary data



Limitations

MLE Parameter calibration:

- Sensitive to $\boldsymbol{\theta}_0$ (Several local maxima)
- Does not account for uncertainty in parameter estimates $\rightarrow$ you need a lot of data

Numerical errors issues:

- The covariance is rapidly reduced in the measurement step (e.g. accurate measurements are used / after a period with missing data
- There is a large difference between the variance of several state variables contributing to the same observation.
- Alternative approaches:
  - **UD Filter**
  - **Square-Root Filter**

Limitations

Linear models:

- The Kalman Filter and Smoother is only applicable for linear models
- Alternative approaches:
  - **Particle Filter**
  - **Unscented Kalman Filter**
  - **Extended Kalman Filter**

Non-constant time-steps:

- If times steps are not constant across the dataset, parameter values must be time-step-size dependent.

$$\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$$

Summary

Transition model

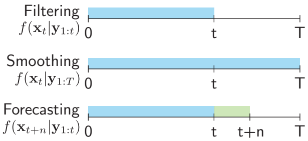
$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \underbrace{\mathbf{w}_t}_{W\sim\mathcal{N}(\mathbf{0},\mathbf{Q})}$$

Observation model

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \underbrace{\mathbf{v}_t}_{V\sim\mathcal{N}(\mathbf{0},\mathbf{R})}$$

- **A**: Model transition matrix
- **C**: Model observation matrix
- **Q**: Model errors covariance matrix
- **R**: Measurement errors covariance matrix

Filtering/Smoothing/Forecasting



$$(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}, \mathcal{L}_t) = \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1}, y_t; \mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R})$$

Prediction step

$$\begin{aligned}\boldsymbol{\mu}_{t|t-1} &= \mathbf{A}\boldsymbol{\mu}_{t-1|t-1}\\ \boldsymbol{\Sigma}_{t|t-1} &= \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^{\intercal} + \mathbf{Q}\end{aligned}$$

Measurement step

$$\begin{aligned}f(\mathbf{x}_t|\mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})\\ \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\mathbf{r}_t && \text{Posterior expected value}\\ \boldsymbol{\Sigma}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{C})\boldsymbol{\Sigma}_{t|t-1} && \text{Posterior covariance}\end{aligned}$$

Optimal parameter $\boldsymbol{\theta}^*$ must be identified using a maximization algorithm (e.g. Newton-Raphson)

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \ln \mathcal{L}_t$$

Marginal prior probability of each observation

$$\mathcal{L}_t = f(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{y}_t; \mathbf{C}\boldsymbol{\mu}_{t|t-1}, \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\intercal} + \mathbf{R})$$