Optimizaton and Kalman-Bucy Smoothing

Aleksandr Y. Aravkin
University of Washington
sasha.aravkin@gmail.com

Bradley Bell
University of Washington
bradbell@uw.edu

James V. Burke
University of Washington
jvburke@uw.edu

Gianluigi Pillonetto
University of Padova
giapi@dei.unipd.it

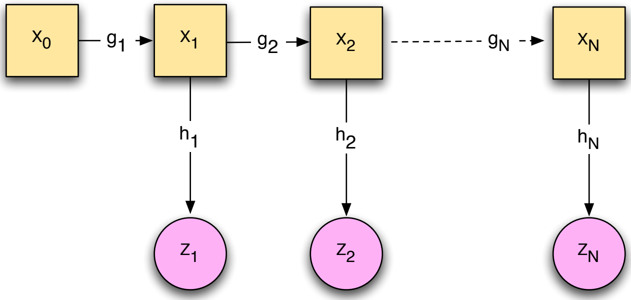
The Chinese University of Hong Kong,
March 4, 2016

Kalman Smoothing Framework

$$\begin{aligned} \mathbf{x_1} &= g_1(x_0) + \mathbf{w_1}, & & \text{(Initialization)} \\ \mathbf{x_k} &= g_k(\mathbf{x_{k-1}}) + \mathbf{w_k} & k = 2, \ldots, N, & \quad \text{(State Transition Dynamics)} \\ \mathbf{z_k} &= h_k(\mathbf{x_k}) + \mathbf{v_k} & k = 1, \ldots, N\ , & \quad \text{(Observations)} \end{aligned}$$

- g_k, h_k known (nonlinear) process and measurement functions
- $\mathbf{w_k} \sim N(0, Q_k)$, $\mathbf{v_k} \sim N(0, R_k)$ mutually independent
- $Q_k \in \mathcal{S}^n_{++}$, $R_k \in \mathcal{S}^{m(k)}_{++}$ known covariance matrices
- $\mathbf{x_k} \in \mathbb{R}^n$ unknown states
- $\mathbf{z_k} \in \mathbb{R}^{m(k)}$ observed measurements (known)

Kalman Smoothing Framework: Graphical Illustration



Modeling and Estimation of Dynamics under Uncertainty

- Navigation
- tracking
- healthcare
- finance
- weather
- imaging

Classical Approach for Linear Systems

$$
\left.
\begin{array}{rcllr}
\mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\
\mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\
\mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k)
\end{array}
\right\}
\text{mutually independent}
$$

Classical Approach for Linear Systems

$$
\left.\begin{array}{rcllr}
\mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\
\mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\
\mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k)
\end{array}\right\} \text{mutually independent}
$$

Kalman Filter

Classical Approach for Linear Systems

$$\left.\begin{array}{rcllr} \mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\ \mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\ \mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k) \end{array}\right\} \text{mutually independent}$$

Kalman Filter

Time Update $\qquad \mathbf{x_{k|k-1}} = G_k \mathbf{x_{k-1|k-1}} + \mathbf{w_k}$

Classical Approach for Linear Systems

$$\left.\begin{array}{rcl} \mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, \qquad\qquad\qquad\qquad\quad \mathbf{w_1} \sim N(0, Q_1) \\ \mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} \quad k = 2, \dots, N, \quad \mathbf{w_k} \sim N(0, Q_k) \\ \mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} \qquad k = 1, \dots, N\,, \quad \mathbf{v_k} \sim N(0, R_k) \end{array}\right\} \begin{array}{c}\text{mutually}\\ \text{independent}\end{array}$$

Kalman Filter

Time Update	$\mathbf{x_{k\|k-1}} = G_k \mathbf{x_{k-1\|k-1}} + \mathbf{w_k}$
Measurement Update	$\mathbf{z_{k\|k-1}} = H_k \mathbf{x_{k\|k-1}} + \mathbf{v_k}$

Classical Approach for Linear Systems

$$\left.\begin{array}{rclll} \mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\ \mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\ \mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k) \end{array}\right\} \text{mutually independent}$$

Kalman Filter

Time Update	$\mathbf{x_{k\|k-1}} = G_k \mathbf{x_{k-1\|k-1}} + \mathbf{w_k}$
Measurement Update	$\mathbf{z_{k\|k-1}} = H_k \mathbf{x_{k\|k-1}} + \mathbf{v_k}$
Joint Distr.	$P(\mathbf{x_{k\|k-1}}, \mathbf{z_{k\|k-1}})$

Classical Approach for Linear Systems

$$
\left.\begin{array}{rclll}
\mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\
\mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\
\mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k)
\end{array}\right\} \begin{array}{c}\text{mutually}\\ \text{independent}\end{array}
$$

Kalman Filter

Time Update	$\mathbf{x_{k\|k-1}} = G_k \mathbf{x_{k-1\|k-1}} + \mathbf{w_k}$
Measurement Update	$\mathbf{z_{k\|k-1}} = H_k \mathbf{x_{k\|k-1}} + \mathbf{v_k}$
Joint Distr.	$P(\mathbf{x_{k\|k-1}}, \mathbf{z_{k\|k-1}})$
Cond. Distr.	$P(\mathbf{x_{k\|k}} \mid \mathbf{z_{k\|k-1}} = z_k)$

Classical Approach for Linear Systems

$$
\left.\begin{array}{rclll}
\mathbf{x_1} & = & G_1 x_0 + \mathbf{w_1}, & & \mathbf{w_1} \sim N(0, Q_1) \\
\mathbf{x_k} & = & G_k \mathbf{x_{k-1}} + \mathbf{w_k} & k = 2, \ldots, N, & \mathbf{w_k} \sim N(0, Q_k) \\
\mathbf{z_k} & = & H_k \mathbf{x_k} + \mathbf{v_k} & k = 1, \ldots, N\ , & \mathbf{v_k} \sim N(0, R_k)
\end{array}\right\} \text{mutually independent}
$$

Kalman Filter

Time Update	$\mathbf{x_{k\|k-1}} = G_k \mathbf{x_{k-1\|k-1}} + \mathbf{w_k}$
Measurement Update	$\mathbf{z_{k\|k-1}} = H_k \mathbf{x_{k\|k-1}} + \mathbf{v_k}$
Joint Distr.	$P(\mathbf{x_{k\|k-1}}, \mathbf{z_{k\|k-1}})$
Cond. Distr.	$P(\mathbf{x_{k\|k}} \| \mathbf{z_{k\|k-1}} = z_k)$

Compute the maximum likelihood estimator $x_{k|k}$ at time k.

Classical Approach for Linear Systems

Maximum Likelihood Computation under Gaussian Assumptions

Time Update: $\hat{x}_{0|0} = x_0$ and $P_{0|0} = 0$

$$\begin{aligned}
\hat{x}_{k|k-1} &= G_k \hat{x}_{k-1|k-1} \\
P_{k|k-1} &= G_k P_{k-1|k-1} G_k^T + Q_k
\end{aligned}$$

Measurment Update:

$$\begin{aligned}
K_k &= P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}) \\
P_{k|k} &= P_{k|k-1} - K_k H_k P_{k|k-1}
\end{aligned}$$

An Optimization Approach

Maximum *a posteriori* Formulation

$$P\left(\{x_k\}\middle|\{z_k\}\right) \propto P\left(\{z_k\}\middle|\{x_k\}\right) P\left(\{x_k\}\right) = \prod_{k=1}^{N} P(\{v_k\})P(\{w_k\})$$

$$\propto \prod_{k=1}^{N} \exp\left(-\frac{1}{2}(z_k - h_k(x_k))^T R_k^{-1}(z_k - h_k(x_k)) - \frac{1}{2}(x_k - g_k(x_{k-1}))^T Q_k^{-1}(x_k - g_k(x_{k-1}))\right).$$

An Optimization Approach

Maximum *a posteriori* Formulation

$$P\left(\{x_k\}\middle|\{z_k\}\right) \propto P\left(\{z_k\}\middle|\{x_k\}\right) P\left(\{x_k\}\right) = \prod_{k=1}^{N} P(\{v_k\})P(\{w_k\})$$

$$\propto \prod_{k=1}^{N} \exp\left(-\frac{1}{2}(z_k - h_k(x_k))^T R_k^{-1}(z_k - h_k(x_k)) - \frac{1}{2}(x_k - g_k(x_{k-1}))^T Q_k^{-1}(x_k - g_k(x_{k-1}))\right).$$

The maximum likelihood problem for the MAP estimate is

$$\min_{\{x_k\}} \sum_{k=1}^{N} \frac{1}{2}(z_k - h_k(x_k))^\top R_k^{-1}(z_k - h_k(x_k)) + \frac{1}{2}(x_k - g_k(x_{k-1}))^\top Q_k^{-1}(x_k - g_k(x_{k-1}))$$

An Optimization Approach

Maximum *a posteriori* Formulation

$$P\left(\{x_k\}\middle|\{z_k\}\right) \propto P\left(\{z_k\}\middle|\{x_k\}\right)P\left(\{x_k\}\right) = \prod_{k=1}^{N} P(\{v_k\})P(\{w_k\})$$

$$\propto \prod_{k=1}^{N} \exp\left(-\frac{1}{2}(z_k - h_k(x_k))^T R_k^{-1}(z_k - h_k(x_k)) - \frac{1}{2}(x_k - g_k(x_{k-1}))^T Q_k^{-1}(x_k - g_k(x_{k-1}))\right).$$

The maximum likelihood problem for the MAP estimate is

$$\min_{\{x_k\}} \sum_{k=1}^{N} \frac{1}{2}(z_k - h_k(x_k))^\top R_k^{-1}(z_k - h_k(x_k)) + \frac{1}{2}(x_k - g_k(x_{k-1}))^\top Q_k^{-1}(x_k - g_k(x_{k-1}))$$

which can be written as

$$\min_{\{x_k\}} \frac{1}{2} \sum_{k=1}^{N} \|z_k - h_k(x_k)\|^2_{R_k^{-1}} + \|x_k - g_k(x_{k-1})\|^2_{Q_k^{-1}}.$$

Notational Simplification

$$g(x) = \begin{bmatrix} x_1 \\ x_2 - g_2(x_1) \\ \vdots \\ x_N - g_N(x_{N-1}) \end{bmatrix} \qquad h(x) = \begin{bmatrix} h_1(x_1) \\ h_2(x_2) \\ \vdots \\ h_N(x_N) \end{bmatrix}$$

$$\begin{aligned} R &= \mathrm{diag}(\{R_k\}) \\ Q &= \mathrm{diag}(\{Q_k\}) \end{aligned} \qquad \begin{aligned} x &= \mathrm{vec}(\{x_k\}) \\ w &= \mathrm{vec}(\{g_0, 0, \ldots, 0\}) \\ z &= \mathrm{vec}(\{z_1, z_2, \ldots, z_N\}) \end{aligned}$$

With this notation, the MAP problem becomes

$$\min_x f(x) = \tfrac{1}{2}\|g(x) - w\|^2_{Q^{-1}} + \tfrac{1}{2}\|h(x) - z\|^2_{R^{-1}}.$$

Tri-Diagonal Structure in the Linear Case

Consider the linear case: $g_k(x_{k-1}) = G_k x_{k-1} \quad h_k(x_k) = H_k x_k$

Then set $H := \mathrm{diag}(\{H_k\})$ and $G := \begin{bmatrix} \mathrm{I} & 0 & & \\ -G_2 & \mathrm{I} & \ddots & \\ & \ddots & \ddots & 0 \\ & & -G_N & \mathrm{I} \end{bmatrix}$.

The MAP problem becomes

$$\min_x f(x) = \tfrac{1}{2}\|Hx - z\|^2_{R^{-1}} + \tfrac{1}{2}\|Gx - w\|^2_{Q^{-1}}.$$

The (smoothing) estimate for x is the solution of the linear system

$$(H^\top R^{-1} H + G^\top Q^{-1} G)\, x = H^\top R^{-1} z + G^\top Q^{-1} w \ .$$

The Tri-Diagonal Structure

$$C = (H^\top R^{-1} H + G^\top Q^{-1} G) = \begin{bmatrix} C_1 & A_2^{\mathrm{T}} & 0 & & \\ A_2 & C_2 & A_3^{\mathrm{T}} & 0 & \\ 0 & \ddots & \ddots & \ddots & \\ & 0 & A_N & C_N & \end{bmatrix},$$

with $A_k \in \mathbb{R}^{n\times n}$ and $C_k \in \mathbb{R}^{n\times n}$ defined as follows:

$$\begin{aligned} A_k &= -Q_k^{-1} G_k \ , \\ C_k &= Q_k^{-1} + G_{k+1}^\top Q_{k+1}^{-1} G_{k+1} + H_k^\top R_k^{-1} H_k \ . \end{aligned}$$

Note that one forward solve gives the optimal estimate for $\hat{x}_{N|N}$.

The back solve gives the MAP estimates for $x_{N-1}, x_{N-2}, \ldots$.

The single forward solve to obtain $\hat{x}_{N|N}$ is the *Kalman Filter*.

The back solve is the *Kalman-Bucy smoother*.

Block Tri-Diagonal Solvers

Block Tri-Diagonal Solvers:

- Rauch-Tung-Striebel = forward-backward block tridiagonal (FBT) algorithm = Thomas algorithm
- Mayne-Fraser = forward/backward block tridiagonal (FBT) algorithm
- Mayne's algorithm A = backward-forward block tridiagonal (FBT) algorithm

Block Tri-Diagonal Solvers

Block Tri-Diagonal Solvers:

- Rauch-Tung-Striebel = forward-backward block tridiagonal (FBT) algorithm = Thomas algorithm
- Mayne-Fraser = forward/backward block tridiagonal (FBT) algorithm
- Mayne's algorithm A = backward-forward block tridiagonal (FBT) algorithm
- Burn at both ends solver
- Divide and Conquer
- Twisted

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Euler approximation: $x_{k+1} = x_k + \dot{x}_k \Delta t + w_k$.

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Euler approximation: $x_{k+1} = x_k + \dot{x}_k \Delta t + w_k$.

Linear Process model: $\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \Delta t & I \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix} + \begin{bmatrix} \dot{w}_k \\ w_k \end{bmatrix}$

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Euler approximation: $x_{k+1} = x_k + \dot{x}_k \Delta t + w_k$.

Linear Process model: $\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \Delta t & I \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix} + \begin{bmatrix} \dot{w}_k \\ w_k \end{bmatrix}$

Stochastic calculus gives $Q_k = \sigma^2 \begin{bmatrix} \Delta t & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t^3/3 \end{bmatrix}$ $(\sigma^2 = 1)$

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Euler approximation: $x_{k+1} = x_k + \dot{x}_k \Delta t + w_k$.

Linear Process model: $\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \Delta t & I \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix} + \begin{bmatrix} \dot{w}_k \\ w_k \end{bmatrix}$

Stochastic calculus gives $Q_k = \sigma^2 \begin{bmatrix} \Delta t & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t^3/3 \end{bmatrix}$ $(\sigma^2 = 1)$

Observation model: $z_k = H_k x_k + v_k$, $\quad H_k = \begin{bmatrix} 0 & 1 \end{bmatrix}$ $(R_k = .35^2)$

Example: Smoothing Processes

Model a smooth signal as integrated Brownian motion.

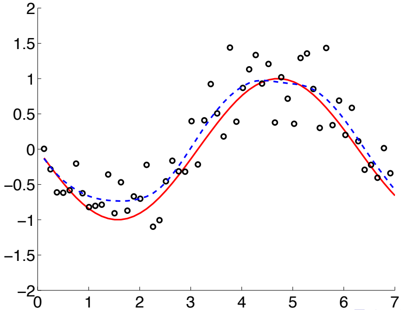
Smooth process: $\dot{x}_{k+1} = \dot{x}_k + \dot{w}_k$

Euler approximation: $x_{k+1} = x_k + \dot{x}_k \Delta t + w_k$.

Linear Process model: $\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \Delta t & I \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix} + \begin{bmatrix} \dot{w}_k \\ w_k \end{bmatrix}$

Stochastic calculus gives $Q_k = \sigma^2 \begin{bmatrix} \Delta t & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t^3/3 \end{bmatrix}$ $(\sigma^2 = 1)$

Observation model: $z_k = H_k x_k + v_k$, $\quad H_k = \begin{bmatrix} 0 & 1 \end{bmatrix}$ $(R_k = .35^2)$



Application Motivated Model Extensions

Application Motivated Model Extensions

- Inclusion of state constraints, especially, inequalities and bounds

Application Motivated Model Extensions

- Inclusion of state constraints, especially, inequalities and bounds

- Non-Gaussian densities:
  - in observations for robustness
  - in states for rapid trend shifts
  - in priors on the state for sparseness and/or roughness

Application Motivated Model Extensions

- Inclusion of state constraints, especially, inequalities and bounds

- Non-Gaussian densities:
  - in observations for robustness
  - in states for rapid trend shifts
  - in priors on the state for sparseness and/or roughness

- State dependent covariance matrices

Application Motivated Model Extensions

- Inclusion of state constraints, especially, inequalities and bounds

- Non-Gaussian densities:
  - in observations for robustness
  - in states for rapid trend shifts
  - in priors on the state for sparseness and/or roughness

- State dependent covariance matrices

- learning parameters for hyperpriors

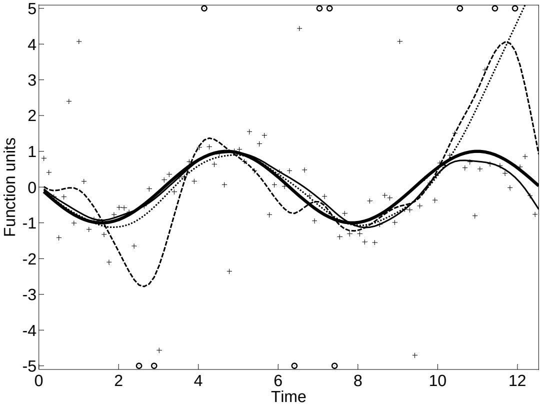
Alternative Approaches

Bayesian methods based on simulation and sampling

- unscented filters
- sigma-point methods
- ensemble filters
- particle filters

Example: Outliers in the observations

Robust Smoothing Process for $x(t) = -sin(t)$



Simulation: measurements (+), outliers (o) (absolute residuals more than three standard deviations), true function (thick line), ℓ_1-Laplace estimate (thin line), Gaussian estimate (dashed line), Gaussian outlier removal estimate (dotted line)

Robust Smoothers for Observations with Outliers

Classical Smoother:

$$\min_x f(x) = \tfrac{1}{2}\|Hx - z\|^2_{R^{-1}} + \tfrac{1}{2}\|Gx - w\|^2_{Q^{-1}}.$$

Robust smoother with Laplace density on the observations:

$$\underset{x \in \mathbf{R}^{Nn}}{\text{minimize}} \;\; f(x) = \sqrt{2}\left\| R^{-1/2}(h(x) - z) \right\|_1 + \tfrac{1}{2}\left\|g(x) - w\right\|_{Q^{-1}}$$

Robust Smoothers

Laplace density on the observations:

$$\underset{x \in \mathbf{R}^{Nn}}{\text{minimize}} \;\; f(x) = \sqrt{2} \left\| R^{-1/2}(h(x) - z) \right\|_1 + \tfrac{1}{2} \left\| g(x) - w \right\|_{Q^{-1}}^2$$

Robust Smoothers

Laplace density on the observations:

$$\underset{x\in\mathbf{R}^{Nn}}{\text{minimize}}\;\; f(x) = \sqrt{2}\left\|R^{-1/2}(h(x)-z)\right\|_1 + \tfrac{1}{2}\|g(x)-w\|^2_{Q^{-1}}$$

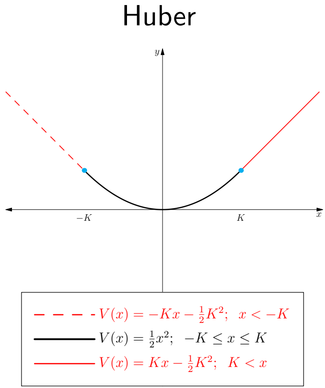
Convex piecewise linear-quadratic (PLQ) penalties:

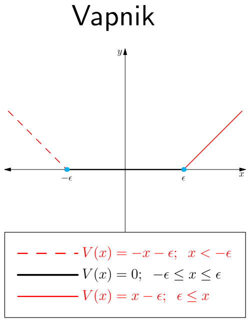
$$\underset{x\in\mathbf{R}^{Nn}}{\text{minimize}}\;\; \sum_{k=1}^{N} V_k\left(h(x_k)-z_k;R_k\right) + J_k\left(x_k - g(x_{k-1});Q_k\right)\;,$$

where the penalties V_k and J_k are associated with log-concave densities of the form

$$p_{v,k}(z) \propto \exp\left(-V_k(z:R_k)\right)\right) \quad \text{and} \quad p_{w,k}(x) \propto \exp\left(-J_k(x;Q_k)\right)$$

Quadratic Support (QS) Functionals and PLQ Densities





$$\rho(U, M, b, B; y) = \sup_{u \in U} \left\{ \langle u, b + By \rangle - \frac{1}{2}\langle u, Mu \rangle \right\}$$

U convex polyhedron, M psd, B injective, $[B^{\mathrm{T}}\mathrm{cone}(U)]^\circ = \{0\}$

Other QS Functions: $\rho(U, M, b, B; \cdot)$

1. Norms, Gauges and Support Functions.

2. The Huber function.
Take $M = I$, $B = I$, and $b = 0$.
Given $\kappa > 0$ set $U = \kappa\mathbb{B}_\infty$.
Then ρ is the multivariate Huber function.

3. Generalized Huber functions.
Let $M \in \mathbb{S}^n_{++}$, $B = I$, and $b = 0$.
Let $\|y\|_M = \sqrt{y^T M y}$ and $\|y\|_{M^{-1}} = \sqrt{y^T M^{-1} y}$.
Set $U = \kappa\mathbb{B}_M = \{\kappa u \mid \|u\|_M \leq 1\}$.
Then,

$$\rho(y) = \begin{cases} \frac{1}{2}\|y\|^2_{M^{-1}} & , \text{if } \|y\|_{M^{-1}} \leq \kappa \\ \kappa\|y\|_{M^{-1}} - \frac{\kappa^2}{2} & , \text{if } \|y\|_{M^{-1}} > \kappa \end{cases} .$$

Other QS Densities: $\rho(U, M, b, B; \cdot)$

3 Order intervals and Vapnik loss functions. Let $\|\cdot\|$ be a norm with closed unit ball $\mathbb{B}$, let $K \subset \mathbb{R}^n$ be a non-empty symmetric convex cone ($K^\circ = -K$), and let $w <_K v$ ($v - w \in \operatorname{intr}(K)$). Set

$$U = (\mathbb{B}^\circ \cap K) \times (\mathbb{B}^\circ \cap K^\circ), \quad M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad b = -\begin{pmatrix} v \\ w \end{pmatrix}, \quad \text{and} \quad B = \begin{bmatrix} I \\ I \end{bmatrix}.$$

Then $\rho(y) = \operatorname{dist}\left(y \,|w \leq_K z \leq_K v\right).$

$\{y \mid w \leq_K y \leq_K v\}$ is an "order interval".

If we take $w = -v$, then $\{y \mid -v \leq_K y \leq_K v\}$ is a symmetric neighborhood of the origin.
By taking $\|\cdot\| = \|\cdot\|_1$, $K = \mathbb{R}^n_+$, and $v = \epsilon\mathbf{1}$=-w, we recover the multivariate Vapnik loss function. Further examples of symmetric cones are $\mathbb{S}^n_+$ and the Lorentz "ice cream" cone (ℓ^2-cone).

Optimization with PLQ Penalties

QS Functions closed wrt addition, pointwise max, infimal convolution, and affine composition

$$\min_{y \in \mathbb{R}^n} \rho(U, M, b, B; y) = \min_{y \in \mathbb{R}^n} \sup_{u \in U} \left\{ \langle u, b + By \rangle - \tfrac{1}{2} \langle u, Mu \rangle \right\},$$

where $U = \{u : A^{\mathrm{T}} u \leq a\}$.

Optimization with PLQ Penalties

QS Functions closed wrt addition, pointwise max, infimal convolution, and affine composition

$$\min_{y\in\mathbb{R}^n} \rho(U, M, b, B; y) = \min_{y\in\mathbb{R}^n} \sup_{u\in U} \left\{ \langle u, b + By\rangle - \tfrac{1}{2}\langle u, Mu\rangle \right\},$$

where $U = \{u : A^{\mathrm{T}} u \le a\}$.

The KKT conditions are

$$\begin{aligned}
0 &= B^{\mathrm{T}} u \\
0 &= b + By - Mu - Aq \\
0 &= A^{\mathrm{T}} u + s - a \\
0 &= q_i s_i \ , \ \ i = 1, \ldots, \ell \ , \ \ q, s \ge 0 \ ,
\end{aligned}$$

Optimization with PLQ Penalties

QS Functions closed wrt addition, pointwise max, infimal convolution, and affine composition

$$\min_{y \in \mathbb{R}^n} \rho(U, M, b, B; y) = \min_{y \in \mathbb{R}^n} \sup_{u \in U} \left\{ \langle u, b + By \rangle - \tfrac{1}{2} \langle u, Mu \rangle \right\} ,$$

where $U = \{u : A^{\mathrm{T}} u \leq a\}$.

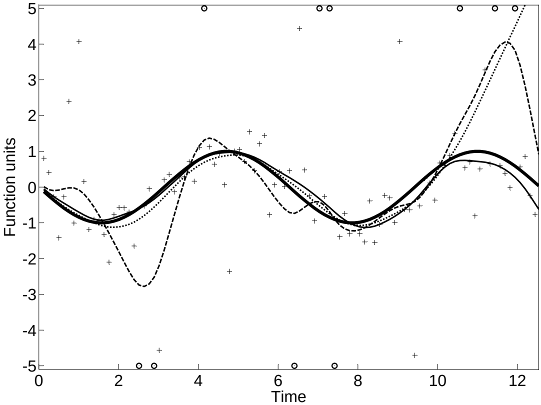
The KKT conditions are

$$\begin{aligned} 0 &= B^{\mathrm{T}} u \\ 0 &= b + By - Mu - Aq \\ 0 &= A^{\mathrm{T}} u + s - a \\ 0 &= q_i s_i \ , \ \ i = 1, \ldots, \ell \ , \ \ q, s \geq 0 \ , \end{aligned}$$

In the case of Kalman smoothing, IP algorithms yield tridiagonal systems at each iteration.

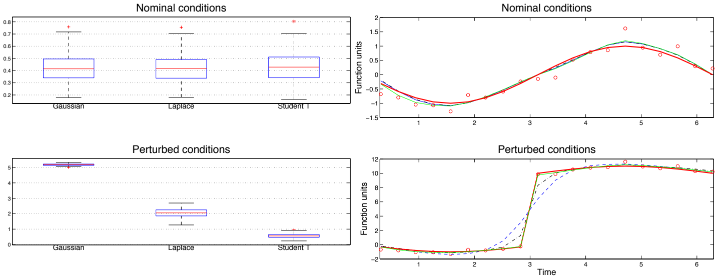
Example: Outliers in the observations

Robust Smoothing Process for $x(t) = sin(t)$



Simulation: measurements (+), outliers (o) (absolute residuals more than three standard deviations), true function (thick line), ℓ_1-Laplace estimate (thin line), Gaussian estimate (dashed line), Gaussian outlier removal estimate (dotted line)

Example: Trend Filtering



Reconstruction of a sudden change in state obtained by ℓ_2, ℓ_1, and T-Trend smoothers. **Left:** Boxplot of reconstruction errors under nominal (top) and perturbed (bottom) conditions. **Right**: Reconstructions obtained using ℓ_2 (dashed), ℓ_1 (dashdot) and T-Trend (thin line) smoother. The thick line is the true state.

Nonlinearities: Convex-Composite Optimization

$$\underset{x \in \mathbf{R}^{Nn}}{\text{minimize}} \; f(x) = \sum_{k=1}^{N} V_k\left(h(x_k) - z_k; R_k\right) + J_k\left(x_k - g(x_{k-1}); Q_k\right) \; ,$$

$$f(x) = \rho(F(x)) \; , \quad \text{where}$$

$$\rho\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sum_{k=1}^{N} V_k\left(y_{1k}; R_k\right) + J_k\left(y_{2k}; Q_k\right) \; , \quad F(x) = \begin{bmatrix} g(x) - w \\ h(x) - z \end{bmatrix} \; .$$

ρ is convex (PLQ) and F is smooth.

Gauss-Newton Method for Convex-Composite Optimization

Gauss-Newton Search Direction:

$$d^\nu = \arg\min_d \tilde{f}(d) := \rho\left(F(x^\nu) + \nabla F(x^\nu)^\top d\right)$$

Update: $x^{\nu+1} := x^\nu + \gamma^\nu d^\nu$

Sufficient Decrease: $0 < \kappa < 1$

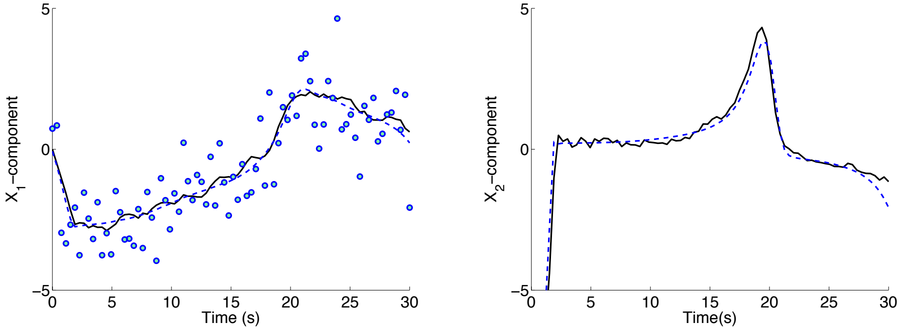
$$f(x^\nu + \gamma^\nu d^\nu) \leq f(x^\nu) + \kappa\gamma^\nu \tilde{\Delta} f(x^\nu),$$

where $\tilde{\Delta} f(x^\nu) = \tilde{f}(d^\nu) - f(x^\nu)$

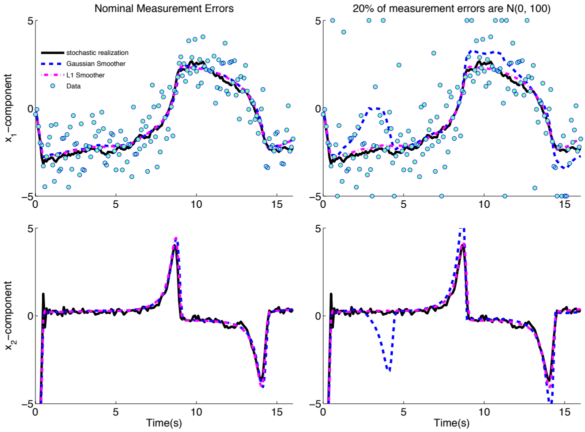
Example: Nonlinear Processes

Van der Pol oscillator:

$\dot{X}_1(t) = X_2(t)$ and $\dot{X}_2(t) = \mu[1 - X_1(t)^2]X_2(t) - X_1(t)$



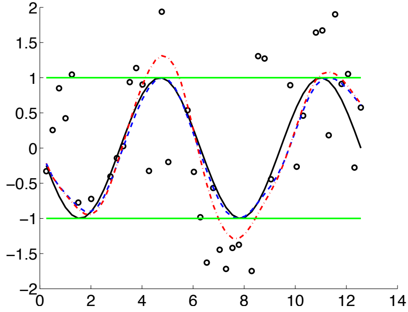
Example: Robust Smoothing Process for Van der Pol



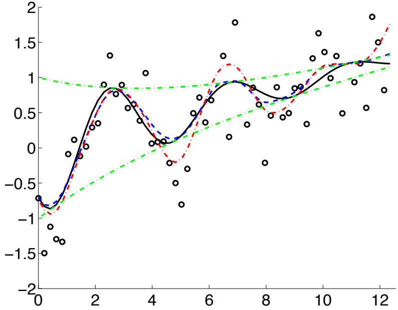
The left two panels show estimation of x_1, (top) and x_2 (bottom) with errors from the nominal model. The stochastic realization is represented by a thick black line; the Gaussian smoother is the blue dashed line, and the ℓ_1-smoother is the magenta dash-dotted line. Right two panels show the same stochastic realization but with large measurement errors. Outliers appear on the top and bottom boundary in the top right panel.

Example: State Bounds

$x(y) = \sin t$



$x(t) = \exp(-\alpha t)\sin(\beta t) + .1t$



Black solid line is true signal, magenta dash-dot line is unconstrained Kalman smoother, and blue dashed line is the constrained Kalman smoother. Measurements are displayed as circles, and bounds are shown as green lines.

Covariance State Dependence

$$
\begin{aligned}
\mathbf{x_1} &= g_1(x_0) + \mathbf{w_1}, & & \text{(Initialization)}\\
\mathbf{x_k} &= g_k(\mathbf{x_{k-1}}) + \mathbf{w_k} & k = 2, \ldots, N, & \quad \text{(State Transition Dynamics)}\\
\mathbf{z_k} &= h_k(\mathbf{x_k}) + \mathbf{v_k} & k = 1, \ldots, N\ , & \quad \text{(Observations)}
\end{aligned}
$$

- g_k, h_k known (nonlinear) process and measurement functions
- $\mathbf{w_k} \sim N(0, Q_k(x_k))$, $\mathbf{v_k} \sim N(0, R_k(x_k))$ mutually independent
- $Q_k(x_k) \in \mathcal{S}^n_{++}$, $R_k(x_k) \in \mathcal{S}^{m(k)}_{++}$ covariance matrices
- $\mathbf{x_k} \in \mathbb{R}^n$ unknown states
- $\mathbf{z_k} \in \mathbb{R}^{m(k)}$ observed measurements

Applications

Tracer Kinetics

p_0 plasma injection

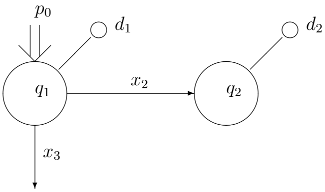
q_1 plasma compartment mass

q_2 urine compartment mass

x_1 plasma volume

x_2 transfer rate from q_1 to q_2

x_3 transfer rate out of system



Compartmental model.

Applications

Tracer Kinetics

p_0 plasma injection

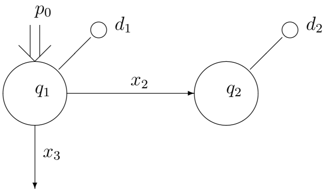
q_1 plasma compartment mass

q_2 urine compartment mass

x_1 plasma volume

x_2 transfer rate from q_1 to q_2

x_3 transfer rate out of system



Compartmental model.

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_{k+1} = (1 + \mathbf{w_k}) \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k + \mathbf{w_k} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k$$

Applications

Tracer Kinetics

p_0 plasma injection

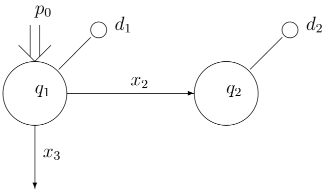
q_1 plasma compartment mass

q_2 urine compartment mass

x_1 plasma volume

x_2 transfer rate from q_1 to q_2

x_3 transfer rate out of system



Compartmental model.

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_{k+1} = (1 + \mathbf{w_k}) \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k + \mathbf{w_k} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}_k$$

Other Applications

- Radar position errors and aspect angle and turning rate
- Optical wavefront reconstruction from diversity images

The MAP Objective J

The MAP object for the Kalman smoother is

$$
\begin{aligned}
J(x) &= -\log\det\left(Q^{-1/2}(x)\right) - \log\det\left(R^{-1/2}(x)\right) \\
&\quad + \tfrac{1}{2}\|Q^{-1/2}(x)(g(x)-w)\|_2^2 + \tfrac{1}{2}\|R^{-1/2}(x)(h(x)-z)\|_2^2 \\
&= \tfrac{1}{2}\log\det(W(x)) + \tfrac{1}{2}c(x)^T W(x)^{-1} c(x),
\end{aligned}
$$

where $c : \mathbb{R}^{nN} \to \mathbb{R}^{M+nN}$ and $W : \mathbb{R}^{nN} \to \mathcal{S}_{++}^{M+nN}$.

J as a Composition with a DC Function

$$J(x) := \tfrac{1}{2}c(x)^T W(x)^{-1}c(x) + \tfrac{1}{2}\log\det(W(x))$$

Then $J = \hat{\rho} \circ \hat{F}$ with

$$\hat{\rho}(c, W) := \tfrac{1}{2}c^T W^{-1}c - \left(-\tfrac{1}{2}\log\det(W)\right)$$

and

$$\hat{F}(x) = (c(x), W(x)).$$

$\hat{\rho}$ is the difference of two convex functions (DC function), the matrix-fractional function and the negative log-determinant.

A Convex-Composite Reformulation

Change of variables: $V = W^{-1/2}$

$$J(x) = \tfrac{1}{2}c(x)^T W(x)^{-1} c(x) + \tfrac{1}{2}\log\det(W(x)) \quad = \hat{\rho} \circ \hat{F}(x)$$

$$K(x) = \tfrac{1}{2}\|V(x)c(x)\|_2^2 - \log \circ \det[V(x)] \quad = \rho \circ F(x) \ ,$$

where

$$\rho(u, v) \quad = \quad \tfrac{1}{2}u^{\mathrm{T}}u - \sum_i \log[v_i] \ , \quad F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \end{bmatrix} = \begin{bmatrix} V(x)c(x) \\ \mathrm{vec}[\{V_{ii}(x)\}] \end{bmatrix} \ .$$

A Convex-Composite Reformulation

Change of variables: $V = W^{-1/2}$

$$J(x) = \tfrac{1}{2} c(x)^T W(x)^{-1} c(x) + \tfrac{1}{2} \log\det(W(x)) \quad = \hat{\rho} \circ \hat{F}(x)$$

$$K(x) = \tfrac{1}{2} \| V(x) c(x) \|_2^2 - \log \circ \det[V(x)] \quad = \rho \circ F(x) ,$$

where

$$\rho(u, v) \;=\; \tfrac{1}{2} u^{\mathrm{T}} u - \sum_i \log[v_i] , \quad F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \end{bmatrix} = \begin{bmatrix} V(x) c(x) \\ \mathrm{vec}[\{ V_{ii}(x) \}] \end{bmatrix} .$$

Modeling Assumption:
The Cholesky factors for $Q_k^{-1}(x_k)$ and $R_k^{-1}(x_k)$ are given to us as explicit functions of the state.

Numerical Example: Gauss-Newton $\rho(F(x^\nu) + F'(x^\nu)d)$

"Ground Truth" $\quad x(t) = \begin{bmatrix} 1 - 2\cos(t) \\ t - 2\sin(t) \end{bmatrix}$

Dynamics $\quad g_k(x_{k-1}) = \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} x_{k-1}, \; Q_k = \begin{bmatrix} \Delta t & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t^3/3 \end{bmatrix}$

Observations $\quad h_k(x_k) = x_{2,k}$

Measurement Variance $\quad R_k(x_k) = (3 - x_{1,k})^{-2}$

Data

Two full periods of the time series $x(t)$, with $N = 100$ discrete time points equally spaced over the interval $[0, 4\pi]$, and with noise sampled from $N(0, R_k(x_k))$.

A Numerical Example

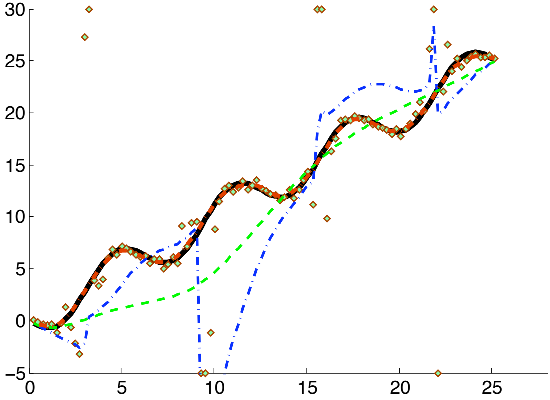


Figure : True state x_1 (black curve), Extended Smoother estimate (thick red dash-dot), Kalman filter estimate (blue dash-dot) and Kalman Smoother estimate (green dashed curve). Measurements are diamonds, and those outside the axis range are displayed on the figure boundary.

An Underwater Tracking Application

In this experiment, an object was tracked using ocean floor transponders. The object was hung on a steel cable approximately 200 meters below a ship. The ship was pitching and rolling on the surface of the ocean and the pilot of the ship was attempting to 'hold station'; i.e., stay at a specific latitude and longitude. A pressure sensor was mounted on the object and it recorded pressure measurements at an approximate rate of once per second. Four acoustic transponders were mounted on the bottom of the ocean and their locations were determined to sub-meter accuracy prior to this experiment. The acoustic travel time between these bottom mounted transponders and the object at the end of the cable was measured at approximately sixteen second intervals. The acoustic travel times, and the pressure measurements, were used to estimate the location of the object at the end of the cable.
The goal is to track the object.

An Underwater Tracking Application: The State

N is the total number of time points at which we have tracking data.

The state vector at time t_k is defined by

$$x_k = (e_k, n_k, d_k, \dot{e}_k, \dot{n}_k, \dot{d}_k)^{\mathrm{T}}$$

where

$$(e_k, n_k, d_k) \;=\; (\text{ east, north, depth })$$

the location of the object (in meters from the origin), and $(\dot{e}_k, \dot{n}_k, \dot{d}_k)$ is the time derivative of this location.

An Underwater Tracking Application: The Measurement

The the first 4 components of the measurement vector, $z_k \in \mathbb{R}^5$, at time t_k are the range measurements to the 4 corresponding bottom mounted transponders and the last measurement component is the depth corresponding to the pressure measurement.
For $j = 1, \ldots, 4$, the model for the mean of the corresponding range measurements was

$$h_{j,k}(x_k) = \|(e_k, n_k, d_k) - b_j\|_2 - \Delta r_j.$$

The functions $h_{j,k}$ are nonlinear due to the presence of the norm. These measurements were modeled as independent and having a standard deviation of 3 meters. The model for the mean of the pressure measurement was $h_{5,k}(x_k) = d_k$. These measurements were modeled as having a standard deviation of 0.05 meters (the pressure sensor was much more accurate than the range measurements).

An Underwater Tracking Application: GPS Validation

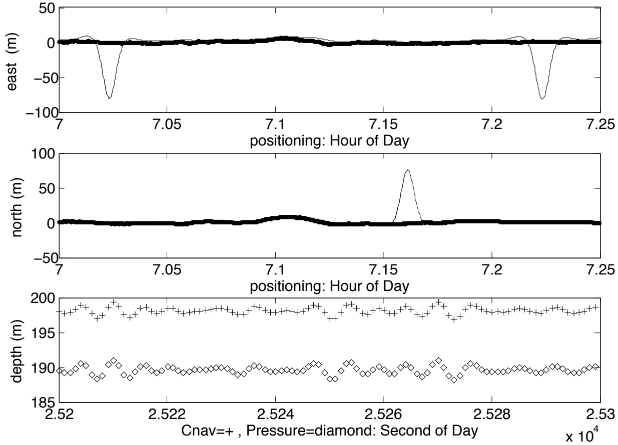
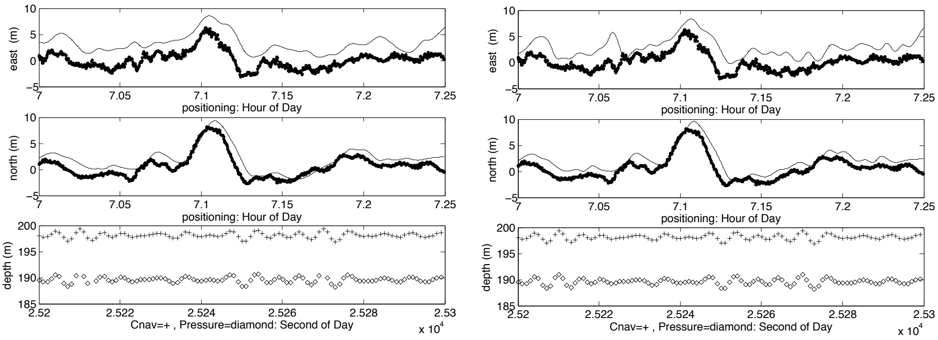


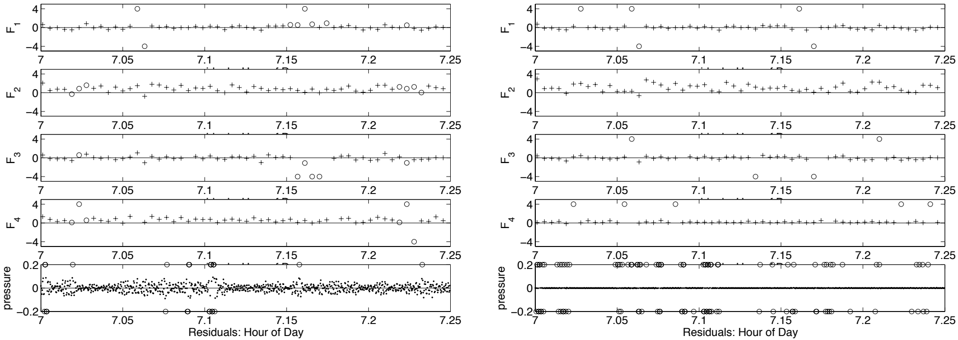
Figure : Track: Independent GPS verification (thick line and +), Iterated Gaussian smoother estimate (thin line).

An Underwater Tracking: Smoother Comparison



Track: Independent GPS verification (thick line and $+$)
Left: Gaussian smoother with outlier removal (thin line)
Right: ℓ_1-Laplace smoother (thin line).

An Underwater Tracking: Residuals Comparison



Left: Gaussian smoother with outlier removal
Right: ℓ_1-Laplace smoother.
In these plots, transponders are labeled by their frequencies in KHz; $F_1 = 11.25$, $F_2 = 11.75$, $F_3 = 12.25$, and $F_4 = 12.75$. All residuals are in meters.

Thank You!

- "An Inequality Constrained Kalman-Bucy Smoother by Interior Point Likelihood Maximization", with Bradley Bell and Gianluigi Pillonetto, *Automatica*, **45**(2009)25-33.

- "An ℓ_1-Laplace Robust Kalman Smoother" , *IEEE Transactions on Automatic Control*, **56**(2011) 2898–2911, with Aleksandr Aravkin, Bradley Bell and Gianluigi Pillonetto.

- "Sparse/Robust Estimation and Kalman Smoothing with Nonsmooth Log-Concave Densities: Modeling, Computation, and Theory." with A.Y.Aravkin and G.Pillonetto. *Journal of Machine Learning Research*, **14**(2013) 2689-2728.

- "Optimization viewpoint on Kalman smoothing, with applications to robust and sparse estimation." with A.Y.Aravkin and G.Pillonetto. In Compressed Sensing & Sparse Filtering, eds., A. Carmi, L. Mihaylova, and S. Godsill. Springer. pp. 237-281, 2014.

- "Robust and Trend-following Student's t-Kalman Smoothers." with A.Y. Aravkin and G. Pillonetto. SIAM J. Control Optim. **52**(2014): 2891-2916.

- "Smoothing dynamical systems with state-dependent covariance matrices." with A.Y. Aravkin. (CDC), 2014.