Robustness issues in Kalman filtering revisited

Peter Ruckdeschel

Fraunhofer ITWM, Abteilung Finanzmathematik, Fraunhofer-Platz 1, D-67663 Kaiserslautern

Peter.Ruckdeschel@itwm.fraunhofer.de

Limassol, October 30, 2009

State Space Models (SSM's)

Linear, Time–Discrete, Euclidean Setup

ideal model:

$$x_t = F_t x_{t-1} + v_t, \qquad v_t \overset{\text{indep.}}{\sim} (0, Q_t),$$
$$y_t = Z_t x_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{indep.}}{\sim} (0, V_t),$$
$$x_0 \sim (a_0, Q_0),$$

$\{v_t\}, \{\varepsilon_t\}, x_0$ indep. as processes

(hyper–parameters F_t, Z_t, Q_t, V_t, a_0 known)

Generalizations also covered

- using Markov kernels: Dynamic Bayesian / Hidden Markov Models
- using SDE's: linear continuous time SSM's
- user-set/determined control

Types of Outliers

exogenous outliers affecting only singular observations

$$\text{AO} \;::\; \varepsilon_t^{\text{re}} \sim (1 - r_{\text{AO}})\mathcal{L}(\varepsilon_t^{\text{id}}) + r_{\text{AO}}\mathcal{L}(\varepsilon_t^{\text{di}})$$
$$\text{SO} \;::\; y_t^{\text{re}} \sim (1 - r_{\text{SO}})\mathcal{L}(y_t^{\text{id}}) + r_{\text{SO}}\mathcal{L}(y_t^{\text{di}})$$

endogenous outliers / structural changes

$$\text{IO} \;::\; v_t^{\text{re}} \sim (1 - r_{\text{IO}})\mathcal{L}(v_t^{\text{id}}) + r_{\text{IO}}\mathcal{L}(v_t^{\text{di}})$$

but also trends, level shifts

Different and competing goals

A/SO attenuation of "false alarms"

IO tracking: detect structural changes as fast as possible
or: recover situation without structural changes

both identification problem:
simultaneous treatment only possible with delay

Classical Method: Kalman–Filter

Filter Problem

$\mathrm{E}\,\left|x_t - f_t(y_{1:t})\right|^2 = \min_{f_t}!,$ with $y_{1:t} = (y_1, \ldots, y_t)$, $y_{1:0} := \emptyset$

Kalman–Filter

optimal solution among linear filters — Kalman[/Bucy] [60/61]:

Initialization: $x_{0|0} = a_0$

Prediction: $x_{t|t-1} = F_t x_{t-1|t-1}, \qquad [\Delta x_t = x_t - x_{t|t-1}]$

Correction: $x_{t|t} = x_{t|t-1} + M_t^0 \Delta y_t, \qquad [\Delta y_t = y_t - Z_t x_{t|t-1}]$

+ corresp. recursions for prediction/filtering error cov.'s $\Sigma_{t|t[-1]}$ and Kalman gain M_t^0

robustifying recursive Least Squares: rLS$_{[\text{AO}]}$

in correction step replace $M_t^0 \Delta y_t$ by

$$H_{b_t}(M_t^0 \Delta y_t) = M_t^0 \Delta y_t \min\{1, b_t / \left|M_t^0 \Delta y_t\right|\}$$

Optimality Results

A simplified, but general model

(ideal) unobservable interesting signal $X \sim P^X(dx)$, $\mathrm{E}\,|X|^2 < \infty$
obs. Y; ideal transition prob's have densities:

$$P^{Y|X=x}(dy) = \pi(y,x)\,\mu(dy)$$

range of X s.t. MSE risk makes sense

(real) $\hat{Y} = (1-U)Y + U\tilde{Y}, \qquad U \sim \mathrm{Bin}(1,r) \qquad$ (SO)

SO-nbd. $\mathcal{U}(r) := \{\, \mathcal{L}(\hat{Y}) \mid \hat{Y} \text{ acc. to (SO)} \,\}$

Optimality Problems: find reconstruction f_0 for X s.t.

$\max_{\mathcal{U}} \mathrm{E}_{\mathrm{re}}\,|X - f(\hat{Y})|^2 = \min_f!$ [minmax-SO]

$\mathrm{E}_{\mathrm{id}}\,|X - f(Y)|^2 = \min_f!$ s.t. $\sup_{\mathcal{U}} \big|\mathrm{E}_{\mathrm{re}}\, f(\hat{Y})\big| \le b$ [Lem5-SO]

Fraunhofer ITWM

Optimality Results (cont.)

Thm. ([Minmax-SO], [Lem5-SO], (R.[01,09]))

1. *[minmax-SO]: ∃ a* **saddlepoint** $(f_0, \tilde{P}_0^Y)$ *for*

$$f_0(y) := \mathrm{E}[X] + D(y)\min\{1, \rho/|D(y)|\}$$
$$\tilde{P}_0^Y(dy) := \tfrac{1-r}{r}\big(|D(y)|/\rho - 1\big)_+ \, P^{Y^{\mathrm{id}}}(dy)$$

where $D(y) = \mathrm{E}_{\mathrm{id}}[X|Y=y] - \mathrm{E}[X]$
and $\rho > 0$ *assures that* $\int \tilde{P}_0^Y(dy) = 1$.

2. *[Lem5-SO]: solved by* f_0 *for* $b = \rho$.

3. *Relation to rLS:*
If $\mathrm{E}_{\mathrm{id}}[X|Y] = MY$, *necessarily* $M = M^0$ *(class. Kalman gain)*
and f_0 *is rLS.*

Fraunhofer ITWM

Back in the Δx Model for $t > 1$

- SO-optimality Thm: $\mathrm{E}_{\mathrm{id}}[\Delta x|\Delta y]$ is *linear* $\Longrightarrow$ rLS optimal

Prop. (Linearity for $\mathrm{E}_{\mathrm{id}}[\Delta x|\Delta y]$; R.[01/09])

Assume $\mathcal{L}_{\mathrm{id}}(\varepsilon)$ *normal. Then* $\mathrm{E}_{\mathrm{id}}[\Delta x|\Delta y]$ *is linear*
$\Longleftrightarrow$ $\mathcal{L}(\Delta x)$ *is normal*
$\Longleftrightarrow$ $\mathrm{E}[(e^\tau \Delta x)^3|\Delta y] = 0\ \forall e \in \mathbb{R}^p$ *("conditionally unskewed")*

Way out: An Even Larger SO-Model: the eSO–model

- allow deviations in X: $(\hat{X}, \hat{Y}) := (1-U)(X, X+\varepsilon) + U(\tilde{X}, \tilde{Y})$
- BUT $\mathrm{E}[\tilde{X}] = \mathrm{E}[X]$, $\qquad \mathrm{E}[|\tilde{X}|^2] \le G$

Thm. (minimax-eSO; R.[01])

The pair $(f_0, \tilde{P}_0^Y)$ *from minimax-SO-Thm also is a saddlepoint in the Minimax-eSO-Problem to the same radius .*

Fraunhofer ITWM

Consequences of minimax-eSO-Thm

Normal setup: $\{\varepsilon_t\}, \{v_t\}, x_0$ Gaussian in id. model; Δx_t stemming from rLS-past

Replacing the center of the nbd

- no longer SO-nbd around $\mathcal{L}(\Delta x_t)$
- instead: **eSO** -nbd around *fictive* $\Delta x_t^{\mathcal{N}}$, $\Delta x_t^{\mathcal{N}} \sim \mathcal{N}_p(0, \Sigma_{t|t-1})$
  s.t. $\Delta x_t \mathrel{\hat{=}}$ **eSO**-distortion $\tilde{X}$ to "ideal" $\Delta x_t^{\mathcal{N}}$

Consequence:

$\Delta x^{\mathcal{N}} \sim \mathcal{N}_p(0, \Sigma)$ exists $\Longrightarrow$ rLS is exactly minimax for each t.

- explains good empirical results
- existence of $\Delta x^{\mathcal{N}} \sim \mathcal{N}_p(0, \Sigma)$ not yet proved
- BUT for $p = 1$ checked numerically, for $p > 1$ research in progress

Fraunhofer ITWM

IO Robustness for Tracking

- well-known: Kalman filter is inert in presence of IO's
- need faster tracking $\rightsquigarrow$ *"hysteric"* filter

Approach

- assume additive observation error, $y_t = Z_t x_t + \varepsilon_t$
- idea: $[Z_t]x_t \rightleftharpoons \varepsilon_t$:
  - estimate ε_t in a[n] (optimally) robust way $\rightsquigarrow$ apply rLS-type filter $g(Y_t)$
  - estimate x_t by $Z_t^-(y_t - g(Y_t)) \rightsquigarrow \mathrm{rLS_{IO}}$
  - invertibility issue if $\mathrm{rk}\, Z_t < p$
- rLS-optimality translates correspondingly

$\rightsquigarrow$ still recursive (!) and non iterative

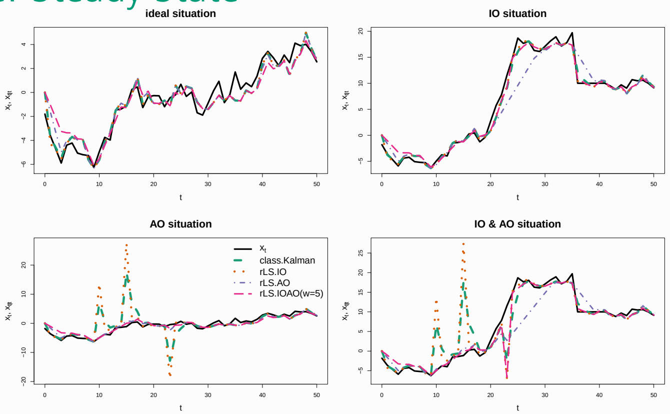
Simultaneous AO & IO Robustness

- decision whether AO or IO can only be taken with delay
- $\rightsquigarrow$ decision lag w

A hybrid filter: $\mathrm{rLS_{IOAO}}$

- run IO-robust and AO-robust filter in parallel
- monitor $|\Delta y_t^{\mathrm{AO}}|$
- when run of length w "large" $|\Delta y_t^{\mathrm{AO}}|$'s switch from AO-robust to IO-robust filter
- despite w-delayed decision as to IO / AO: no smoother is used, just filters
- later on: could be enhanced by *robust smoothers*

Example: Steady State



hyper-parameters: $p = q = 1$, $F_t = Z_t = 1$, ideal model: $v_t, \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$,
AO's in obs. 10,15,23, IO's in obs. 20–25 (trend) and 37–42 (level).

Example: Steady State

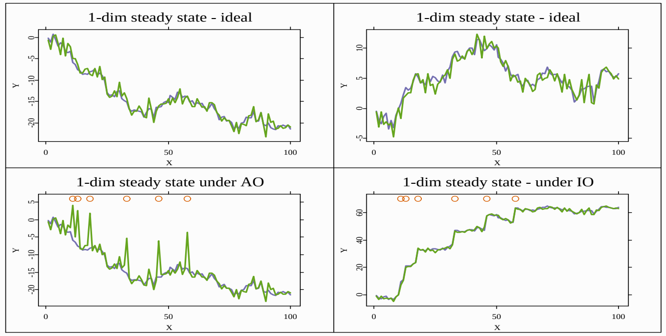
empirical MSE							
Sit.	Type	Kalm	$\mathrm{rLS_{IO}}$	$\mathrm{rLS_{AO}}$	$\mathrm{rLS_{IOAO}}$	ACM	hybr
ideal	filter	**0.59**	0.60	0.75	1.08	0.77	1.41
	pred	1.69	**1.67**	1.96	2.26	2.01	
IO	filter	1.04	**0.83**	6.54	1.36	25.19	1.36
	pred	5.28	**4.71**	12.17	5.42	32.16	
AO	filter	15.25	30.38	0.91	1.16	**0.82**	1.79
	pred	15.15	29.68	2.00	2.25	**2.05**	
IO&AO	filter	17.00	30.52	12.89	**7.78**	28.76	**1.53**
	pred	21.94	34.56	19.23	**13.87**	36.08	

Conclusion

- the *very* flexible class of dynamic models of SSMs can be robustified in a unified way
- contrary to common prejudice:
  simultaneous IO & AO treatment is possible in SSM's
  — albeit with minor delay
- our filters are
  - ± model based:
    - – need model specification (estimation of hyper-param's)
    - + can be more precise in ideal model
  - + recursive, hence fast
  - + valid for higher dimensions

(Extra-slide:) Example: Steady State

Model under AO and IO



hyper-parameters: $p = q = 1$, $F_t = Z_t = 1$, ideal model: $v_t, \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$,
$r_{\mathrm{IO}} = r_{\mathrm{AO}} = 0.1$, $\tilde{v}_t, \tilde{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(10, 0.1)$.

(Extra-slide:) **Calibration**

Two proposals for the choice of *b*

- Anscombe–Criterium

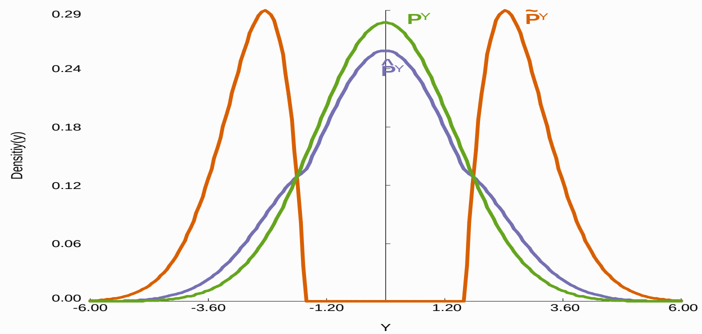
$$\mathrm{E}_{\mathrm{id}} \left|\Delta x - H_b(M^0 \Delta y)\right|^2 \overset{!}{=} (1+\delta)\, \mathrm{E}_{\mathrm{id}} \left|\Delta x - M^0 \Delta y\right|^2$$

- radius–criterium

$$\mathrm{E}_{\mathrm{id}}(|M^0 \Delta y| - b)_+ \overset{!}{=} rb/(1-r)$$

(Extra-slide:) Example: SO-least favorable densities

$P^{Y^{\mathrm{id}}} = P^Y$, $P_0^{Y^{\mathrm{re}}} = \hat{P}^Y$, $P_0^{Y^{\mathrm{di}}} = \tilde{P}^Y$ for $P^X = P^\varepsilon = \mathcal{N}(0,1)$, $r = 0.1$:
tails $\propto |x| e^{-x^2/2}$ instead of $\propto e^{-|x|}$



(Extra-slide:) Example: Steady State

empirical MSE —without obs. 23								
Sit.	Type	Kalm	rLS_{IO}	rLS_{AO}	rLS_{IOAO}	ACM	hybr
ideal	filter	**0.59**	0.60	0.75	1.10	0.78	1.43
	pred	1.71	**1.69**	1.99	2.29	2.03	
IO	filter	0.94	**0.74**	6.08	1.26	24.48	1.38
	pred	5.59	**4.98**	12.05	5.73	31.66	
AO	filter	12.16	24.07	0.86	1.15	**0.84**	1.83
	pred	12.18	23.05	1.94	2.25	**2.10**	
IO&AO	filter	13.28	24.21	11.58	**1.31**	27.93	1.56
	pred	17.01	26.34	17.80	**5.63**	35.38	

Fraunhofer ITWM

References

Birmiwal, K. and Shen, J. (1993) : *Optimal robust filtering.* Stat. Decis., **11**(2): 101–119.

Fried, R. and Schettlinger, K. (2008) : R-package `robfilter`: Robust Time Series Filters. `http://cran.r-project.org/web/packages/robfilter`.

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

Kalman, R.E. and Bucy, R. (1961) : New results in filtering and prediction theory. *Journal of Basic Engineering—Transactions of the ASME*, **83**: 95–108.

Martin, D. (1979) : *Approximate conditional-mean type smoothers and interpolators.* In *Smoothing techniques for curve estimation.* Proc. Workshop Heidelberg 1979. Lect. Notes Math. 757, p. 117-143

Masreliez C.J. and Martin R. (1977) : Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. Autom. Control*, **AC-22**: 361–371.

Rieder, H., Kohl, M., and Ruckdeschel, P. (2008) : The cost of not knowing the radius. *Stat. Meth. & Appl.*, **17**, 13–40.

Fraunhofer ITWM

References (cont.)

Ruckdeschel, P. (2001) : *Ansätze zur Robustifizierung des Kalman Filters.* Bayreuther Mathematische Schriften, Vol. 64.

R Development Core Team (2009) : R*: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`

R-Forge Administration and Development Team (2008) : *R-Forge User's Manual,* BETA. SVN revision: 47, August, 12 2008. `http://r-forge.r-project.org/R-Forge_Manual.pdf`

Schick, I.C. (1989) : *Robust recursive estimation of a discrete–time stochastic linear dynamic system in the presence of heavy-tailed observation noise*. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Schick I.C. and Mitter S.K. (1994) : Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Stat.*, **22**(2): 1045–1080.

Spangl, B. (2008) : *On Robust Spectral Density Estimation*. PhD Thesis at Technical University, Vienna.

Fraunhofer ITWM