Robust Recursive Kalman–Filtering

Peter Ruckdeschel

UNIVERSITÄT BAYREUTH

Mathematics VII

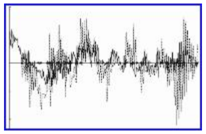
Magdeburger Stochastik–Tage 2002

at Otto–von–Guericke–Universität

Magdeburg

March 20th 2002

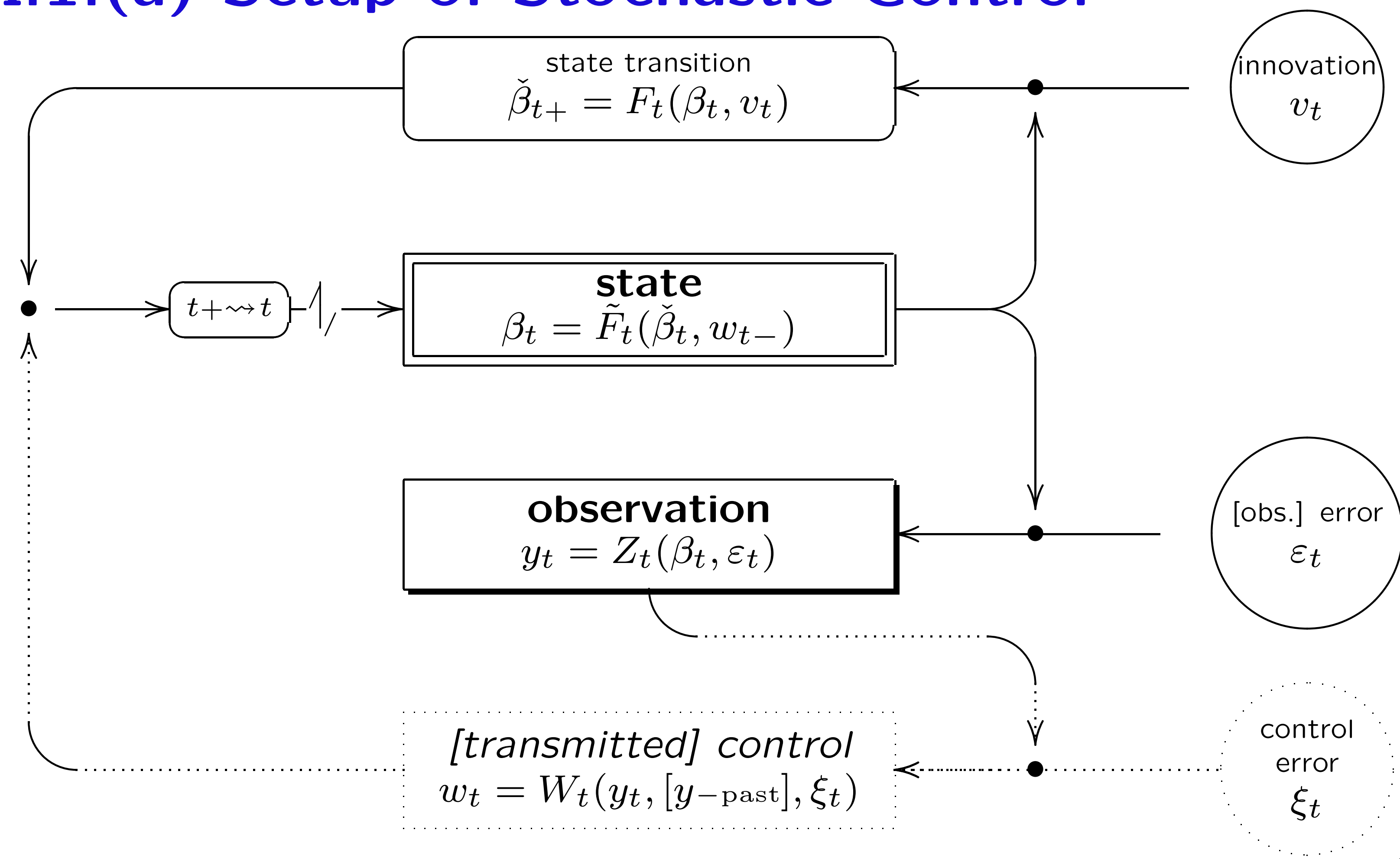
E-mail: `peter.ruckdeschel@uni-bayreuth.de`

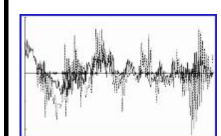


I. A Robust Filter with H. Rieder

I.1. State Space Models (SSM's) and Outliers

I.1.(a) Setup of Stochastic Control





I.1.(b) Definitions and Assumptions: Linear, Time–Discrete, Euclidean Setup

ideal model:

$$y_t = Z_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}_q(0, V_t), \tag{1}$$

$$\beta_t = F_t\beta_{t-1} + v_t, \qquad v_t \sim \mathcal{N}_p(0, Q_t), \tag{2}$$

$$\beta_0 \sim \mathcal{N}_p(a_0, Q_0) \tag{3}$$

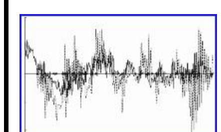
hyper-parameters: F_t, Z_t, Q_t, V_t, a_0

I.1.(c) Types of Outliers

$$\text{AO} \quad :: \quad \varepsilon_t^{\text{real}} \sim (1 - r_{\text{AO}})\mathcal{N}_q(0, V_t) + r_{\text{AO}}\mathcal{L}(\varepsilon_t^{\text{cont}}) \tag{4}$$

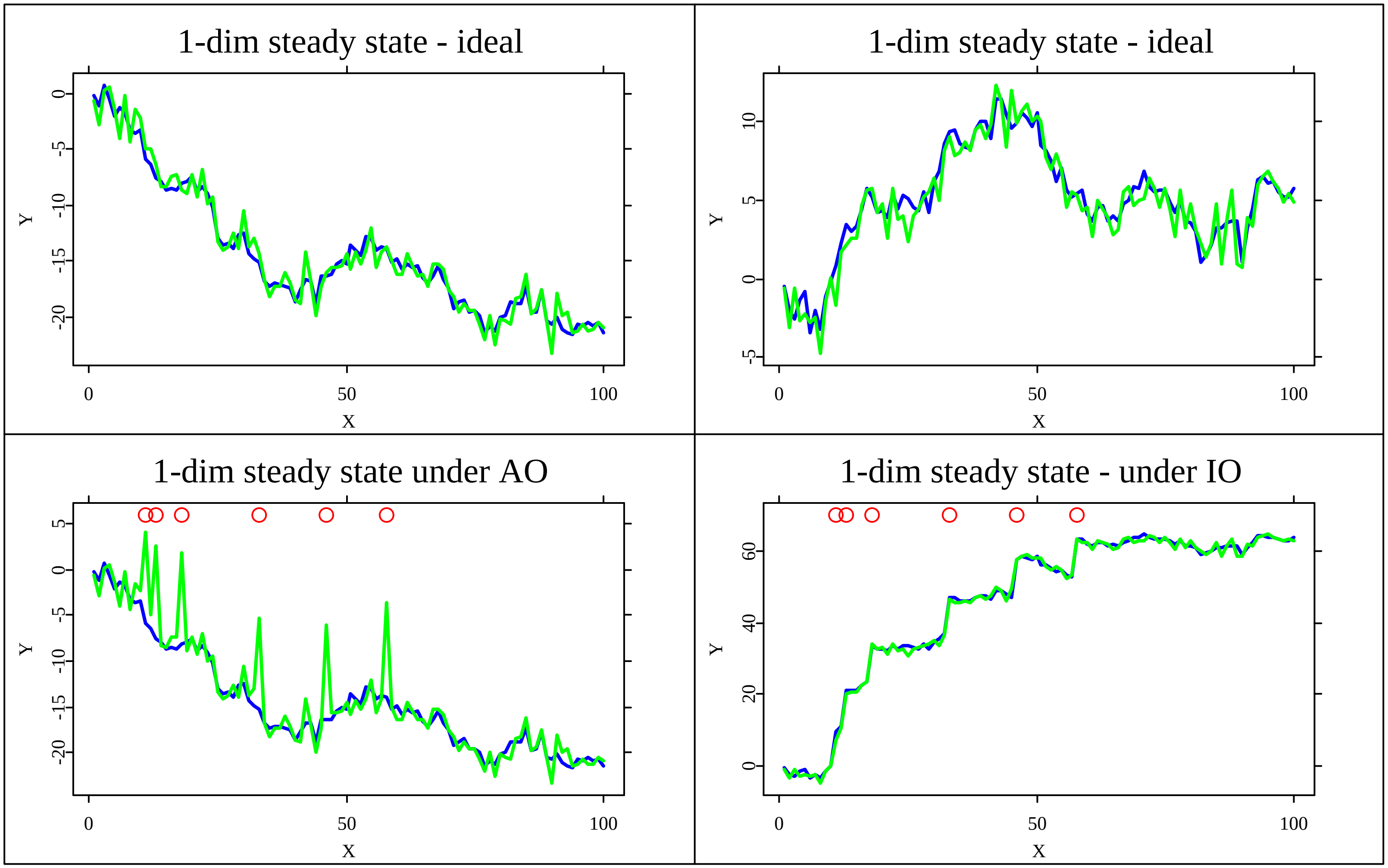
$$\text{SO} \quad :: \quad y_t^{\text{real}} \sim (1 - r_{\text{SO}})\mathcal{L}(y_t^{\text{id}}) + r_{\text{SO}}\mathcal{L}(y_t^{\text{cont}}) \tag{5}$$

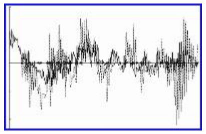
$$\text{IO} \quad :: \quad v_t^{\text{real}} \sim (1 - r_{\text{IO}})\mathcal{N}_p(0, Q_t) + r_{\text{IO}}\mathcal{L}(v_t^{\text{cont}}) \tag{6}$$



I.1.(d) Example: Model under AO and IO

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel





I.2. Classical Method: Kalman–Filter

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

Filter problem

$$\mathrm{E}\left|\beta_t - f_t(y_{1:t})\right|^2 = \min_{f_t}!, \qquad y_{1:t} = (y_1, \ldots, y_t), y_{1:0} := \emptyset \tag{7}$$

General solution: $\mathrm{E}[\beta_t | y_{1:t}]$

LS-solution among linear filters: Kalman–filter (Kalman[/Bucy] [60/61])

Initialization: $\beta_{0|0} = a_0, \qquad \Sigma_{0|0} = Q_0$ (8)

Prediction: $\beta_{t|t-1} = F_t \beta_{t-1|t-1}$ (9)

$$\Sigma_{t|t-1} = F_t \Sigma_{t-1|t-1} F_t^\tau + Q_t = \mathrm{Cov}(\Delta\beta_t) \tag{10}$$

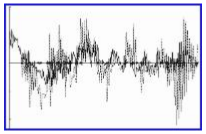
with $\Delta\beta_t = \beta_t - \beta_{t|t-1}$ [state innovation] (11)

Correction: $\beta_{t|t} = \beta_{t|t-1} + \hat{M}_t(y_t - Z\beta_{t|t-1}) = \beta_{t|t-1} + \hat{M}_t \Delta y_t$ (12)

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \hat{M}_t Z_t \Sigma_{t|t-1} = \mathrm{Cov}(\beta_t - \beta_{t|t}) \tag{13}$$

with $\hat{M}_t = \Sigma_{t|t-1} Z_t^\tau [Z_t \Sigma_{t|t-1} Z_t^\tau + V_t]^{-1}$ [Kalman–Gain] (14)

$\Delta y_t = y_t - Z_t \beta_{t|t-1} = Z_t \Delta\beta_t + \varepsilon_t$ [obs. innov.] (15)



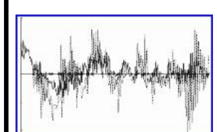
I.3. Robustification Approaches for SSM's

I.3.(a) State of the Art

- already 209 References to that subject in Kassam/Poor[85]; many different notions of robustness
- here: robustness w.r.t. AO/SO–distributional deviations
- key features: **recursivity** and **bounded correction step**

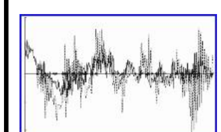
I.3.(b) Various "Robustnesses"

- in Control Theory, c.f. $\mathcal{H}^\infty/\mathcal{H}^2$–approach e.g. Başar/Bernhard [91], Rotea/Khargonekar [95]
- by Hard Rejection, e.g. Meyr/Spies [84]
- by "Fat Tails"
  - Bayesian Approach: e.g. West [81–85],
  - Posterior Mode, e.g. Künstler/Fahrmeir/Kaufmann [91–99]



- by Analogy:
  - M–Estimators for Regression e.g. Boncelet[/Dickinson] [83–85], Cipra/Romera [91]
  - L–Estimators: numerous examples in image processing; an initial example: 3R–smoother by Tukey [77]
- Non–Recursive Robustness
  - without sampling a.o. Pupeikis [98], Schick [89], Birmiwal/Shen [93]
  - with MCMC–methods: Carlin [92], Carter/Kohn[94]
- Minmax-Robustness:
  - in the frequency domain: e.g. Kassam/Lim [77], Franke [85], Franke/Poor [84]
  - ACM-[type]-filter: Martin/Masreliez [77-79]
  - SO-optimal filter in one dimension: Birmiwal/Shen [93]

Peter Ruckdeschel



I.4. robustifying recursive Least Squares: rLS

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

I.4.(a) Idea and Definition

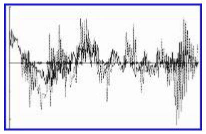
[from now on: omitting t where possible]

- restriction to AO/SO's:
  For arbitrary IO's the problem is not well–posed!
- no AO/SO's in the prediction step
- in the correction step: instead of $M\Delta y$ we use

$$H_b(M\Delta y) = M\Delta y \min\{1, \frac{b}{|M\Delta y|}\} \qquad (16)$$

- So recursions (8),(9),(12) are transformed to

$$\beta_{0|0} = a_0, \qquad (17)$$
$$\beta_{t|t-1} = F_t \beta_{t-1|t-1} \qquad (18)$$
$$\beta_{t|t} = \beta_{t|t-1} + H_b(M_t(y_t - Z\beta_{t|t-1})) \qquad (19)$$

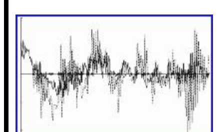


I.4.(b) Properties

- no rotation as in [Masreliez/]Martin ACM [77/79]
- if $\mathrm{E}[\Delta\beta|\Delta y]$ is linear in Δy, then
  - the optimal M is $\hat{M}_t$ (Kalman Gain)
  - rLS is SO-optimal (see part II)
- strict normality gets lost during the history of $\beta_{t|t}$ for growing t
- $\beta_{t|t}$ is “nearly” normal and $\hat{M}_t$ cannot be improved significantly

I.4.(c) Availability/Implementation

- `XploRe`
  - c.f. `http://www.xplore-stat.de`
  - rLS realized in the `XploRe`–quantlib `kalman`
  - documentation: XploRe Application Guide
- `ISP`: macros available on demand
- `S-Plus/R`: not yet

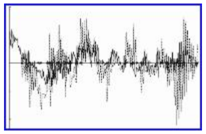


I.4.(d) Calibration

Choice of b: Anscombe–Critrerium

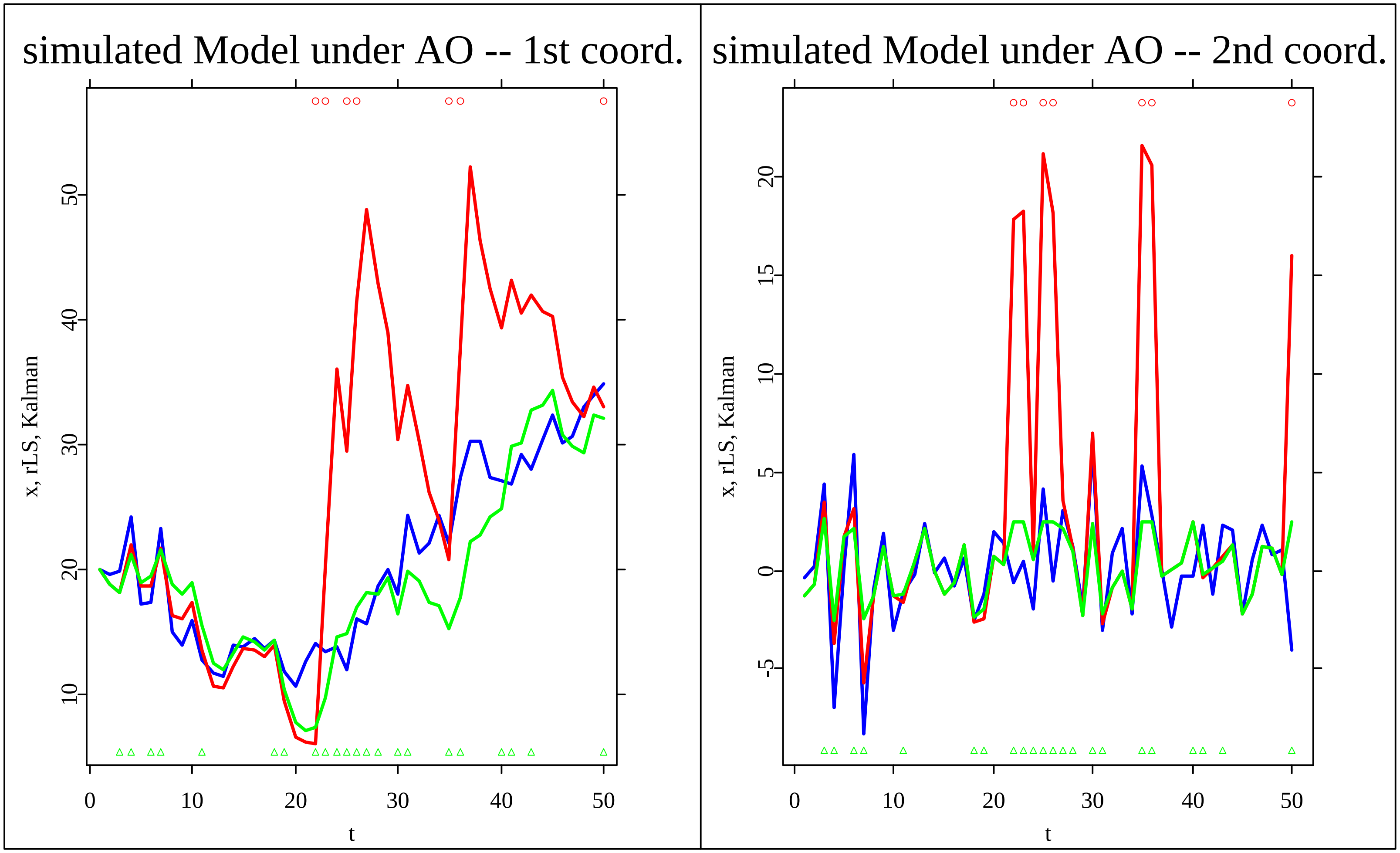
$$\mathrm{E}\left|\Delta\beta - H_b(\hat{M}\Delta y)\right|^2 \overset{!}{=} (1+\delta)\,\mathrm{tr}\,\Sigma_{t|t} \tag{20}$$

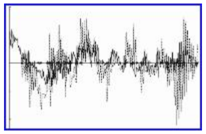
- with known hyper-parameters, calibration can be done beforehand!
- simplifications for implementation of (20):
  - assuming strict normality,
  - for $n = 1$ analytic terms,
  - for $n > 1$ MC-Simulation
- alternatives:
  - simulation of a bundle of paths and then MC-integration
  - numerical integration



I.4.(e) Example: rLS for Simulated Data

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel





II. Robust Optimalities

II.1. Robust Optimization Problems

II.1.(a) Reduction to a Simple Model

- ideal model:

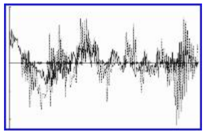
$$Y := X + \varepsilon, \qquad X \sim P^X, \;\; \varepsilon \sim P^\varepsilon \text{ indep.} \tag{21}$$

$$P^X, P^\varepsilon \in \mathcal{M}_1(\mathbb{B}^q)$$

$$\Big[\mathrm{E}_{\mathrm{id}}\, |\, X\,|^2, \quad \mathrm{E}_{\mathrm{id}}\, |\varepsilon|^2 < \infty\Big]$$

- Identifications: innovation representation

$$\mathrm{rLS}: \quad X \;\; \hat{=} \;\; [Z_t]\Delta\beta_t, \quad \varepsilon \;\; \hat{=} \;\; \varepsilon_t$$



UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

II.1.(b) Types of Outliers / Neighborhoods

Types of Outliers

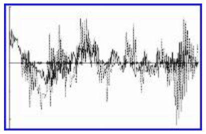
$$\begin{aligned} \text{SO} \quad :: \quad \hat{Y} \sim \hat{P}^Y = &\quad (1 - r_{\text{SO}}) P^Y + r_{\text{SO}} \tilde{P}^Y, \\ & P^Y = P^X * P^\varepsilon \end{aligned} \tag{22}$$

$$\begin{aligned} \text{AO} \quad :: \quad \hat{\varepsilon} \sim \hat{P}^\varepsilon = &\quad (1 - r_{\text{AO}}) P^\varepsilon + r_{\text{AO}} \tilde{P}^\varepsilon, \\ \Rightarrow \hat{Y} \sim \hat{P}^Y = &\quad (1 - r_{\text{AO}}) P^Y + r_{\text{AO}} \tilde{P}^Y, \\ & \hat{Y} = X + \hat{\varepsilon}, \qquad \tilde{P}^Y = P^X * \tilde{P}^\varepsilon \end{aligned} \tag{23}$$

Neighborhoods

$$\text{SO} \quad :: \quad \mathcal{U}_r := \{\mathcal{L}(X, \hat{Y}) : \; X \sim P^X, \hat{Y} \sim \hat{P}^Y, \; \hat{P}^Y \text{acc. to (24)}\} \tag{24}$$

$$\text{AO} \quad :: \quad \mathcal{V}_r := \{\mathcal{L}(X, \hat{Y}) : \; X \sim P^X, \hat{Y} \sim \hat{P}^Y, \; \hat{P}^Y \text{acc. to (25)}\} \tag{25}$$



II.1.(c) Problems to be Solved

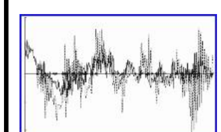
UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

	"Lemma 5"	Minimax
SO	$\mathrm{E}_{\mathrm{id}}[\lvert X - f(\hat{Y})\rvert^2] = \min_f!$ s.t. $\lvert \mathrm{E}_{\mathrm{real}}[X - f(\hat{Y})]\rvert \le b$ $\mathrm{E}_{\mathrm{id}}[X] = \mathrm{E}_{\mathrm{id}}[f(\hat{Y})]$	$\mathrm{E}_P \lvert X - f(\hat{Y})\rvert^2 = \min_f \max_{P \in \mathcal{U}_r}!$
AO	$\mathrm{E}_{\mathrm{id}}[\lvert X - f(\hat{Y})\rvert^2] = \min_f!$ s.t. $\lvert \mathrm{E}_{\mathrm{real}}[X - f(\hat{Y}) \mid \varepsilon]\rvert \le b$ $\mathrm{E}_{\mathrm{id}}[X] = \mathrm{E}_{\mathrm{id}}[f(\hat{Y})]$	$\mathrm{E}_P \lvert X - f(\hat{Y})\rvert^2 = \min_f \max_{P \in \mathcal{V}_r}!$

Equivalences under $\mathrm{E}_{\mathrm{id}}[X] = \mathrm{E}_{\mathrm{id}}[f(\hat{Y})]$

$$\left|\mathrm{E}_{\mathrm{real}}[X - f(\hat{Y})]\right| \le b \;\; \forall P \in \mathcal{U}_r \qquad \Longleftrightarrow \qquad \left|f(\hat{Y}) - \mathrm{E}_{\mathrm{id}}[X]\right| \le b/r$$

$$\left|\mathrm{E}_{\mathrm{real}}[X - f(\hat{Y}) \mid \varepsilon]\right| \le b \;\; \forall P \in \mathcal{V}_r \qquad \Longleftrightarrow \qquad \left|\mathrm{E}_{\mathrm{id}}[f(\hat{Y}) \mid \varepsilon] - \mathrm{E}_{\mathrm{id}}[X]\right| \le b/r$$



II.2. Solution in the SO-Case

II.2.(a) Solution to Problem “Lemma 5”-SO

Setting $D(Y) := \mathrm{E}_{\mathrm{id}}[X|Y] - \mathrm{E}_{\mathrm{id}}[X]$ and $b' = b/r$, we get

$$\hat{f}(Y) := \mathrm{E}_{\mathrm{id}}[X] + D(Y)\min\{1, \frac{b'}{|D(Y)|}\} \tag{26}$$

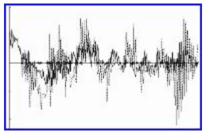
Proof :

$$\mathrm{E}_{\mathrm{id}}[|X - f(Y)|^2] = \mathrm{E}_{\mathrm{id}}[|X - \mathrm{E}_{\mathrm{id}}[X|Y]|^2] + \mathrm{E}_{\mathrm{id}}[|\,\mathrm{E}_{\mathrm{id}}[X|Y] - f(Y)|^2] =$$
$$= \mathrm{const} + \mathrm{E}_{\mathrm{id}}[|D(Y) - (f(Y) - \mathrm{E}_{\mathrm{id}}[X])|^2]$$

pointwise minimization in Y subject to $|f(Y) - \mathrm{E}_{\mathrm{id}}[X]| \le b'$ gives the result. ////

If $\mathrm{E}_{\mathrm{id}}[X|Y] = MY$ for some M, necessarily $M = \hat{M}$ and $\hat{f}(Y)$ is rLS.

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel



II.2.(b) Solution to Problem Minimax-SO

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

- Birmival/Shen [93]:
  - for $q = 1$
  - only Lebesgue-densities for both id. and cont. distr.
  - applying Minimax-Thm without giving justification
- here:
  - $q \geq 1$
  - arbitrary cont. distr.
  - assumption in the ideal model only:

$$(A) \qquad \exists P \in \mathcal{M}_1(\mathbb{B}^q): \quad \text{for } t \in \text{supp}(P^X), \qquad P^\varepsilon(\cdot - t) \ll P.$$

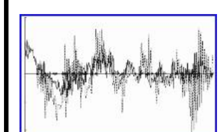
  - Minimax-Thm justified by Franke/Poor [84]

THM 1:(R. [01]) Under (A) there is a saddlepoint $(f_0, \tilde{P}_0^Y)$ with

$$f_0(Y) \quad := \quad \mathrm{E}_{\mathrm{id}}[X] + D(Y)\min\{1, \frac{\tilde{\rho}}{\big|D(Y)\big|}\} \tag{27}$$

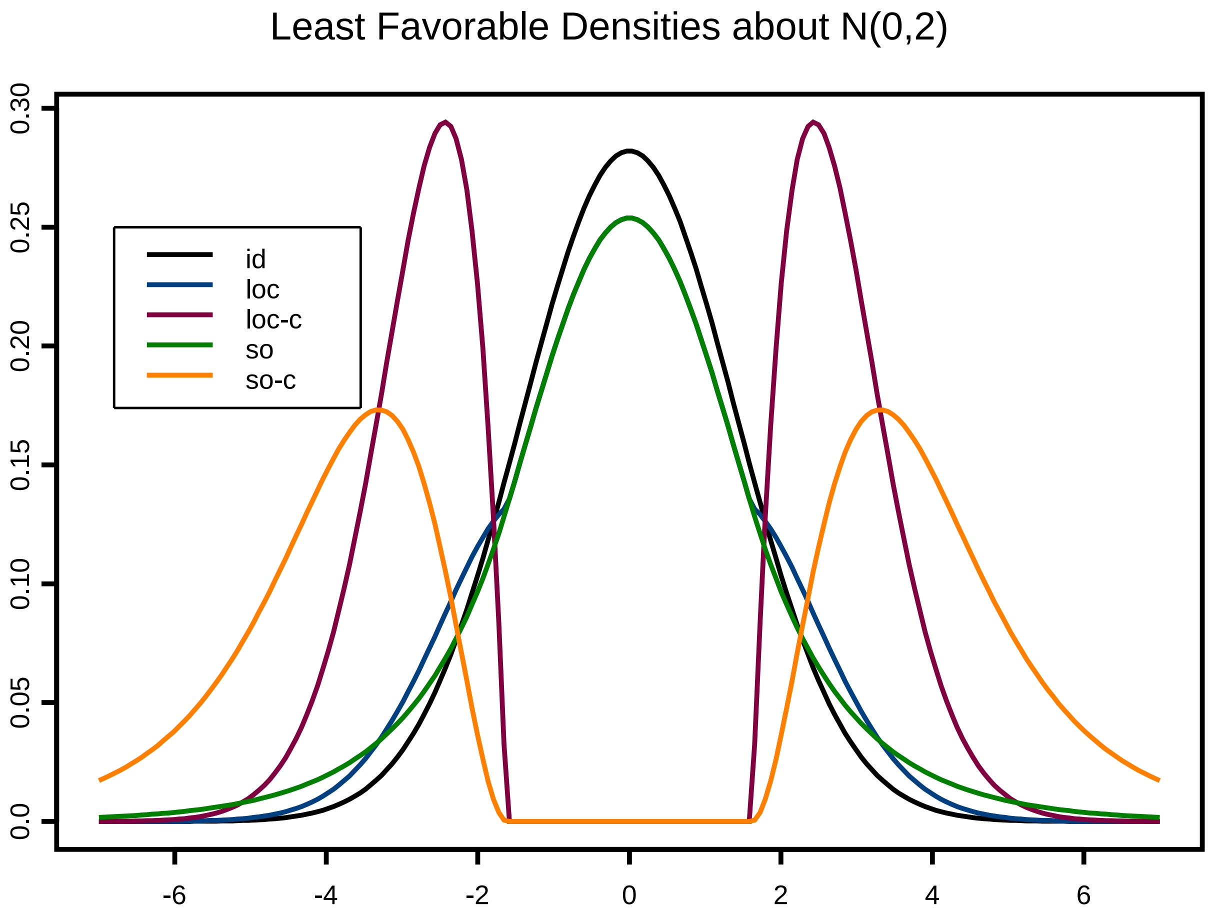
$$\tilde{P}_0(dy) \quad := \quad \frac{1 - r_{\mathrm{SO}}}{r_{\mathrm{SO}}}(1/\tilde{\rho}\big|D(y)\big| - 1)_+ P^Y(dy) \tag{28}$$

with $\tilde{\rho} > 0$ assuring that $\int_{\mathbb{R}^q} \tilde{P}_0(dy) = 1$.

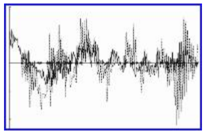


II.2.(c) Example:

Densities of P^Y, $\hat{P}^Y$, $\tilde{P}^Y$ for $P^X = P^\varepsilon = \mathcal{N}(0,1)$, $r = 0.1$



UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

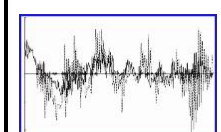


II.3. Back in the $\Delta\beta$ Model for $t > 1$

UNIVERSITÄT BAYREUTH
Peter Ruckdeschel

II.3.(a) Approaches up to Now

- [Masreliez/]Martin [77/79] assume $\mathcal{L}(\Delta\beta)$ *normal*. BUT:
  - if correction step is bounded, $\mathcal{L}(\Delta\beta)$ cannot be normal (R. [01]: as. version of Cramér–Lévy–Theorem)
- rLS is optimal in both "Lemma 5" and minimax sense if $\mathrm{E}_{\mathrm{id}}[\Delta\beta|\Delta y]$ is *linear*. BUT:
  - if $\mathcal{L}_{\mathrm{id}}(\varepsilon)$ is normal, $\mathrm{E}_{\mathrm{id}}[\Delta\beta|\Delta y]$ is linear iff $\mathcal{L}(\Delta\beta)$ is normal (R. [01]: ODE for Fourier transforms of $\mathcal{L}_{\mathrm{id}}(\varepsilon)$ and $\mathcal{L}(\Delta\beta)$.).
- Schick[/Mitter] [89/94] work with a Taylor-expansion for a non-normal $\mathcal{L}(\Delta\beta)$. BUT:
  - stochastic error terms??
  - come up with a bank of (at least t) Kalman–Filters — not very operational
- Birmiwal/Shen [93] work with exact $\mathcal{L}(\Delta\beta)$. BUT:
  - splitting up the history of outlier occurrences yields 2^t different terms — not very operational either



II.3.(b) An Even Larger SO-Model

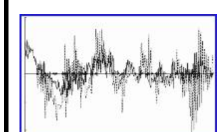
Consider the following outlier model:

- $X \sim P^X$, $\tilde X \sim \tilde P^X$,
  $\varepsilon \sim P^\varepsilon$, $\tilde Y \sim \tilde P^Y$,
  $U \sim \mathrm{Bin}(1, r_{\mathrm{eSO}})$ all sto. indep.
- Observation:

$$(\hat X, \hat Y) := (1-U)(X, X+\varepsilon) + U(\tilde X, \tilde Y). \qquad (29)$$

- $P^X, P^\varepsilon, r_{\mathrm{eSO}}$ known, $\tilde P^X, \tilde P^Y$ unknown /arbitrary,
- but: $\mathrm{E}[\tilde X] = \mathrm{E}[X]$, $\mathrm{E}[|\tilde X|^2] \le G$ for some known $0 < G < \infty$.

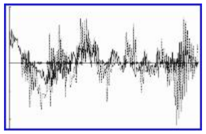
THM 2:(R. [01]) Under (A) $(f_0, \tilde P_0^Y)$ from THM 1 still form a saddlepoint in the larger eSO–model to the same radius — $\tilde P^X$ being arbitrary with $\mathrm{E}[\tilde X] = \mathrm{E}[X]$, $\mathrm{E}[|\tilde X|^2] = G$



II.3.(c) Consequences of THM 2

Instead of regarding the saddlepoint solution to the $\mathcal{U}_r$-nbd around $\mathcal{L}(\Delta\beta)$ we assume that for each t there is a r.v. $\Delta\beta^{\mathcal{N}} \sim \mathcal{N}_p(0,\Sigma)$ s.t. $\Delta\beta$ can be considered a $\tilde{X}$ in the corresponding eSO-nbd around $\Delta\beta^{\mathcal{N}} \sim \mathcal{N}_p(0,\Sigma)$ with the given radius

- in this setup the rLS is exactly minimax for each t
- explains good results
- no analytic proof for the existence of $\Delta\beta^{\mathcal{N}} \sim \mathcal{N}_p(0,\Sigma)$
- BUT for $p = 1$ in a large number of models numerical — not simulational ! — proof



III More Addressed Problems

- AO–problem: both Lemma 5– and Minimax–approach
- Stationarity of the rLS– (and rIC–filter)
- Estimation of Hyper–Parameters:
  - Embedding into LAN–Theory — L_2–differentiability of this model
  - Concept of a Robust One–Step–EM–Algorithm

For questions and comments, as well as for
a detailed outline and a list of references
you please contact me by E-mail.

Also, the slides of this talk are available upon request in `-pdf-`–format

peter.ruckdeschel@uni-bayreuth.de