Autocalibrated parameter choice for non-smooth noise models

Christian Clason

(based on joint work with Kazufumi Ito, Bangti Jin, Karl Kunisch)

Institute for Mathematics and Scientific Computing, Karl-Franzens-Universität Graz

Applied Inverse Problems Conference
College Station, Texas, May 25, 2011

Problem formulation

Tikhonov functional

$$\min_{x\in X} \varphi(\mathcal{F}(Kx, y^\delta)) + \alpha \|x\|_X$$

- $K : X \to Y$ (not necessarily linear), X, Y Banach spaces
- $y^\delta \in Y$ noisy data
- $\mathcal{F} : Y \times Y \to \mathbb{R}$ discrepancy appropriate for noise model
- $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ monotone, continuous

- $x_\alpha \in X$ minimizer
- $x^\dagger \in X$ exact solution

Discrepancy: Examples

- Gaussian noise:

$$\mathcal{F}(x,y) = \frac{1}{2}\|x-y\|_{L^2}^2$$

- Impulsive noise:

$$\mathcal{F}(x,y) = \|x-y\|_{L^1}$$

- Uniform noise:

$$\mathcal{F}(x,y) = \|x-y\|_{L^\infty}$$

- Poisson noise:

$$\mathcal{F}(x,y) = \int (x - y\log x)$$

Parameter choice

Ivanov discrepancy principle

Choose α^* such that

$$\alpha^* \|x_{\alpha^*}\|_X \approx \tau\delta$$

holds for $\tau > 0$ and noise level

$$\delta := \mathcal{F}(Kx^\dagger, y^\delta)$$

(τ depends on K, X, but not noise)

But: requires knowledge of noise level

Heuristic parameter choice

Calibration: Assume that

$$\mathcal{F}(Kx, y^\delta) \approx \mathcal{F}(Kx^\dagger, y^\delta) = \delta$$

for a reasonable range of α

$\rightsquigarrow$ Assumption on noise structure vs. range of K

$\rightsquigarrow$ Find α^* satisfying

Balancing equation

$$\alpha^* \|x_{\alpha^*}\|_X = \tau \mathcal{F}(Kx_{\alpha^*}, y^\delta)$$

Automatic parameter choice

Fixed point iteration

$$\alpha_{k+1} = \tau \frac{\mathcal{F}(Kx_{\alpha_k}, y^\delta)}{\|x_{\alpha_k}\|_X}$$

Theorem

If starting value α_0 satisfies

$$\tau \mathcal{F}(Kx_{\alpha_0}, y^\delta) < \alpha_0 \|x_{\alpha_0}\|_X$$

then $\{\alpha_k\}$

- *is monotonically decreasing*
- *converges to solution of balancing equation*

L^1 fitting

$$\min_{x \in L^2} \|Kx - y^\delta\|_{L^1} + \frac{\alpha}{2}\|x\|_{L^2}^2$$

- $K : L^2(\Omega) \to L^2(\Omega)$ bounded linear operator
- $\Omega \subset \mathbb{R}^n$ bounded domain
- $y^\delta \in L^\infty(\Omega)$ noisy measurement

- appropriate for impulsive noise (e. g., salt & pepper)

- solution by Fenchel duality and semi-smooth Newton method: [C/Jin/Kunisch, SIAM J Imaging Sci 2010]

L^∞ fitting

$$\min_{x \in L^2} \|Kx - y^\delta\|_{L^\infty} + \frac{\alpha}{2}\|x\|_{L^2}^2$$

- $K : L^2(\Omega) \to L^\infty(\Omega)$ bounded linear operator
- $\Omega \subset \mathbb{R}^n$ bounded domain
- $y^\delta \in L^\infty(\Omega)$ noisy measurement
- Appropriate for uniformly distributed noise

L^∞ fitting

$$\min_{c\in\mathbb{R},x\in\mathrm{L}^2} \quad c + \frac{\alpha}{2}\|x\|_{\mathrm{L}^2}^2 \quad \text{s.t.} \quad \|Kx - y^\delta\|_{\mathrm{L}^\infty} \leq c$$

- $K : \mathrm{L}^2(\Omega) \to \mathrm{L}^\infty(\Omega)$ bounded linear operator
- $\Omega \subset \mathbb{R}^n$ bounded domain
- $y^\delta \in \mathrm{L}^\infty(\Omega)$ noisy measurement
- Appropriate for uniformly distributed noise
- Equivalent reformulation

L^∞ fitting

$$\min_{c\in\mathbb{R},x\in L^2} \frac{1}{2}c^2 + \frac{\alpha}{2}\|x\|_{L^2}^2 \quad \text{s. t.} \quad \|Kx - y^\delta\|_{L^\infty} \le c$$

- $K : L^2(\Omega) \to L^\infty(\Omega)$ bounded linear operator
- $\Omega \subset \mathbb{R}^n$ bounded domain
- $y^\delta \in L^\infty(\Omega)$ noisy measurement

- Appropriate for uniformly distributed noise

- Unique solution (c^*, x^*), semi-smooth Newton method
- Approach follows [C/Ito/Kunisch, ESAIM:M2AN 2010]

Regularization

Moreau–Yosida penalization of box constraint: For $\gamma > 0$, consider

$$\min_{c\in\mathbb{R},x\in L^2} \frac{1}{2}c^2 + \frac{\alpha}{2}\|x\|_{L^2}^2 + \frac{\gamma}{2}\|\max(0, Kx - y^\delta - c)\|_{L^2}^2 + \frac{\gamma}{2}\|\min(0, Kx - y^\delta + c)\|_{L^2}^2$$

(max, min pointwise in Ω)

- Existence, uniqueness of solution (c_γ, x_γ),
- Convergence to (c^*, x^*) as $\gamma \to \infty$

$\rightsquigarrow$ Continuation in $\gamma \to \infty$

Optimality system: Lagrangian approach

Saddle point problem

$$\min_{x,c,y} \max_{p} \mathcal{L}(x, c, y, p)$$

for $\mathcal{L} : \mathrm{L}^2 \times \mathbb{R} \times \mathrm{L}^2 \times \mathrm{L}^2 \to \mathbb{R}$,

$$\begin{aligned}\mathcal{L}(x, c, y, p) = \frac{1}{2}c^2 + \frac{\alpha}{2}\|x\|^2_{\mathrm{L}^2} &+ \frac{\gamma}{2}\|\max(0, y - y^\delta - c)\|^2_{\mathrm{L}^2} \\ &+ \frac{\gamma}{2}\|\min(0, y - y^\delta + c)\|^2_{\mathrm{L}^2} \\ &+ \langle p, Kx - y\rangle_{\mathrm{L}^2}\end{aligned}$$

Optimality system

Solution (x^*, c^*, y^*, p^*) of saddle point problem satisfies:

$$
\begin{cases}
0 = \alpha x^* + K^* p^*, \\
0 = c^* - \gamma \int \max(0, y^* - y^\delta - c^*) + \gamma \int \min(0, y^* - y^\delta + c^*), \\
0 = -p^* + \gamma \max(0, y^* - y^\delta - c^*) + \gamma \min(0, y^* - y^\delta + c^*), \\
0 = Kx^* - y^*
\end{cases}
$$

(Necessary and sufficient conditions)

Semi-smooth Newton method

Pointwise max, min Newton-differentiable from $L^p \to L^q$ if $p > q$,
Newton derivative

$$D_N(\max(0, x))h = \begin{cases} h(t) & x(t) > 0, \\ 0 & \text{else.} \end{cases}$$

$\rightsquigarrow$ Optimality system is

- semi-smooth with respect to c: embedding $\mathbb{R} \to L^\infty(\Omega)$
- semi-smooth with respect to y: operator $K : L^2(\Omega) \to L^\infty(\Omega)$

$\rightsquigarrow$ Generalized Newton method $D_N F(x^k)\delta x = -F(x^k)$ converges locally superlinearly for $\gamma > 0$

Test problem: Inverse heat conduction

- K is Volterra integral operator of the first kind,

$$(Kx)(t) = \int_0^t k(s,t)x(s)\,ds,$$

$$k(s,t) = \frac{(s-t)^{-\frac{3}{2}}}{2\sqrt{\pi}} e^{-\frac{1}{4(s-t)}}$$

- Discretization: collocation and midpoint-rule ($N = 300$ nodes)
- Exponentially ill-posed, condition number $\approx 10^{81}$
- Comparison with optimal α from sampling
- Comparison with L^2 fitting (α from sampling):

$$x_{L^2} = (K^*K + \alpha I)^{-1}(K^*y)$$

Noise models

1. L^1 fitting: impulsive Gaussian noise,

$$y^\delta = \begin{cases} Kx^\dagger + \|Kx^\dagger\|_{\mathrm{L}^\infty}\xi, & \text{with probability } d, \\ Kx^\dagger, & \text{otherwise,} \end{cases}$$

ξ normally distributed with mean 0 and standard deviation 1

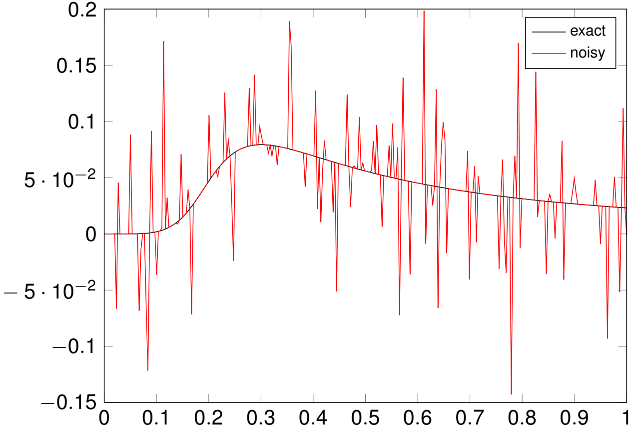
2. L^∞ fitting: uniform noise,

$$y^\delta = Kx^\dagger + \|Kx^\dagger\|_{\mathrm{L}^\infty}\xi$$

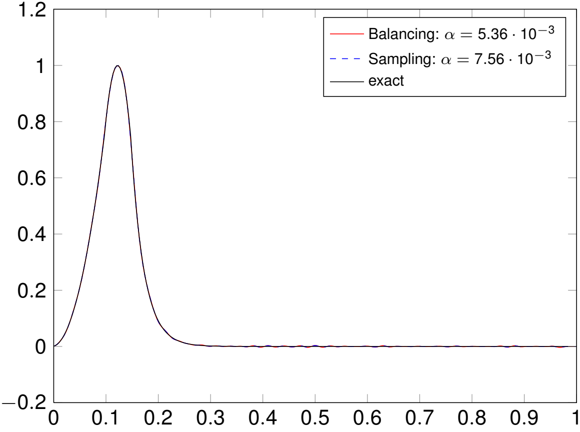
ξ uniformly distributed between $\left[-\frac{d}{2}, \frac{d}{2}\right]$

Both models: $\tau = 10^{-2}$

L^1 fitting: Data ($d = 0.3$)

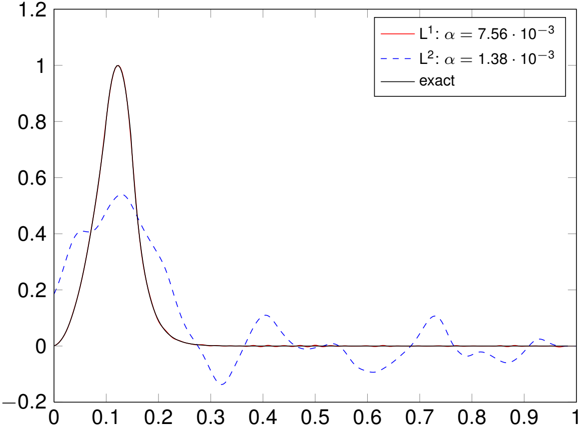


L^1 fitting: Reconstruction ($d = 0.3$)



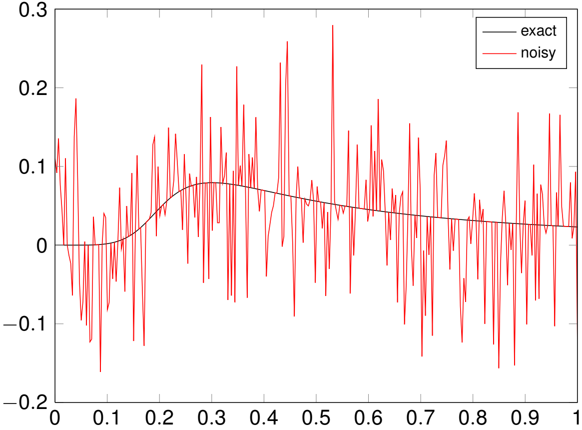
(a) comparison with sampling

L^1 fitting: Reconstruction ($d = 0.3$)

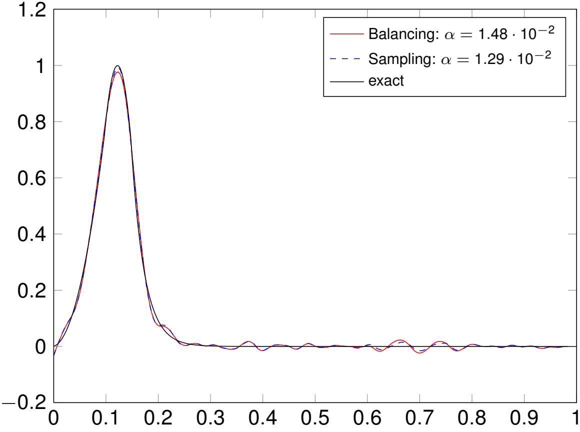


(b) comparison with L^2 fitting

L^1 fitting: Data ($d = 0.6$)

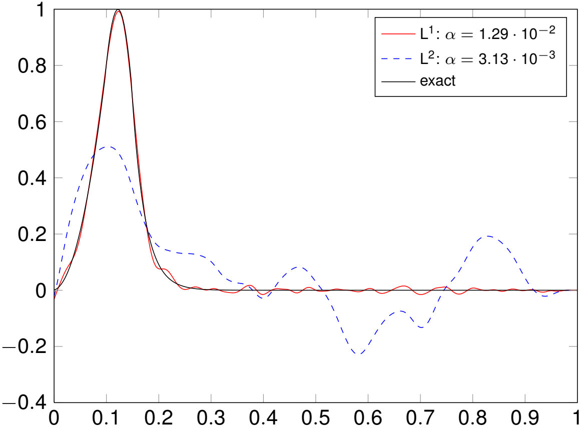


L^1 fitting: Reconstruction ($d = 0.6$)



(a) comparison with sampling

L^1 fitting: Reconstruction ($d = 0.6$)



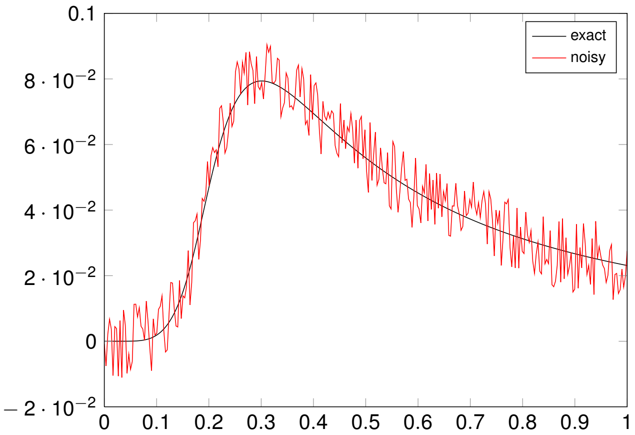
(b) comparison with L^2 fitting

L^1 fitting: Comparison

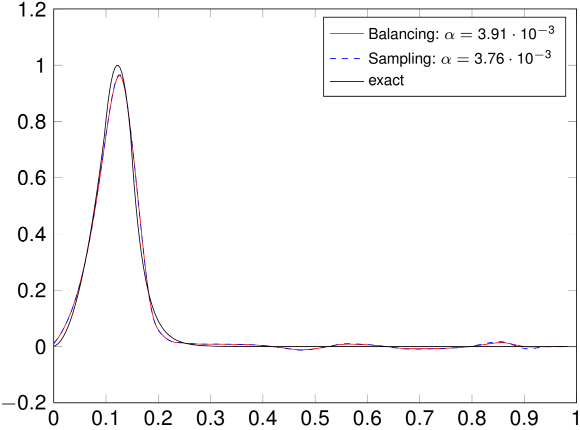
d	δ	δ_b	α_s	α_b	e_s	e_b
0.2	1.31e-2	1.31e-2	3.85e-3	4.34e-3	1.58e-3	1.59e-3
0.3	1.94e-2	1.94e-2	7.39e-3	6.41e-3	1.91e-3	3.29e-3
0.4	2.51e-2	2.51e-2	1.71e-2	8.28e-3	1.09e-2	1.59e-2
0.5	3.38e-2	3.37e-2	1.26e-2	1.13e-2	3.06e-2	3.08e-2
0.6	3.63e-2	3.62e-2	3.85e-2	1.15e-2	5.34e-2	8.89e-2
0.7	4.38e-2	4.38e-2	1.26e-2	1.57e-2	5.84e-2	6.06e-2
0.8	4.68e-2	4.66e-2	1.83e-2	1.33e-2	8.93e-2	1.02e-1
0.9	5.62e-2	5.49e-2	3.59e-2	1.26e-3	1.25e-1	8.68e-1
1.0	6.74e-2	6.55e-2	1.87e-1	2.69e-2	1.90e-1	2.58e-1

δ: L^1 noise level, e: L^2 error; $\cdot_s$: sampling choice, $\cdot_b$: balancing

L^∞ fitting: Data ($d = 0.3$)

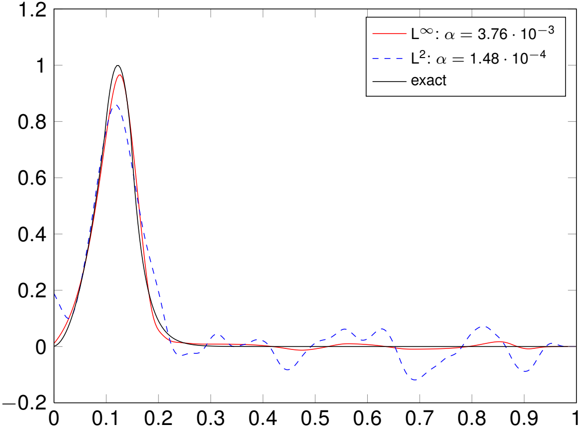


L^∞ fitting: Reconstruction ($d = 0.3$)



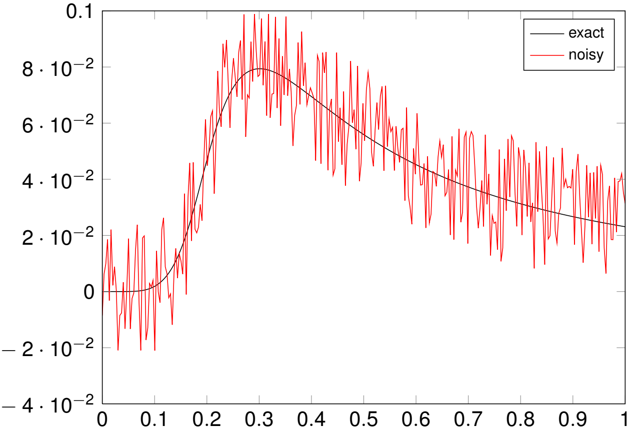
(a) comparison with sampling

L^∞ fitting: Reconstruction ($d = 0.3$)

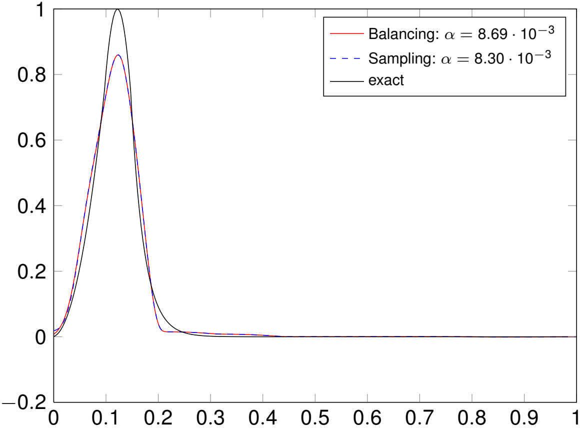


(b) comparison with L^2 fitting

L^∞ fitting: Data ($d = 0.6$)

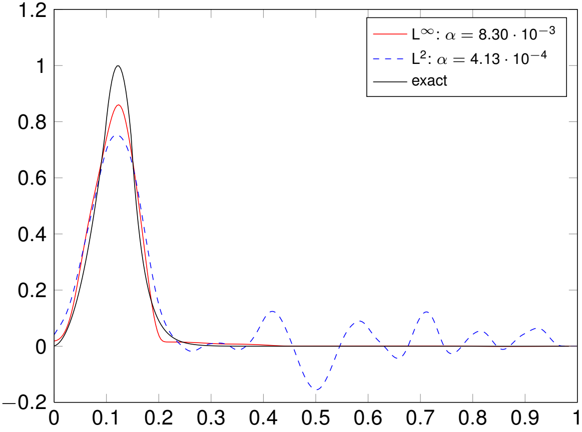


L^∞ fitting: Reconstruction ($d = 0.6$)



(a) comparison with sampling

L^∞ fitting: Reconstruction ($d = 0.6$)



(b) comparison with L^2 fitting

L^∞ fitting: Comparison

d	δ	δ_b	α_s	α_b	e_s	e_b
0.2	7.87e-3	7.74e-3	8.70e-4	2.83e-3	4.19e-2	4.19e-2
0.3	1.18e-2	1.18e-2	6.14e-3	4.15e-3	4.89e-2	5.15e-2
0.4	1.58e-2	1.57e-2	2.01e-3	6.60e-3	7.28e-2	7.29e-2
0.5	1.98e-2	1.96e-2	5.59e-3	7.87e-3	6.32e-2	6.34e-2
0.6	2.38e-2	2.38e-2	5.59e-3	9.53e-3	5.79e-2	6.93e-2
0.7	2.77e-2	2.69e-2	1.18e-2	9.80e-3	6.24e-2	6.26e-2
0.8	3.17e-2	3.09e-2	1.79e-3	1.27e-2	3.47e-2	7.87e-2
0.9	3.56e-2	3.52e-2	1.38e-2	1.56e-2	7.54e-2	7.54e-2
1.0	3.94e-2	3.94e-2	1.05e-2	1.48e-2	7.53e-2	7.54e-2

δ: L^∞ noise level, e: L^2 error; $\cdot_s$: sampling choice, $\cdot_b$: balancing

Conclusion

For non-smooth noise models (and smooth data):

- Noise structure more important than noise level
- Ivanov discrepancy principle useful for heuristic parameter choice
- Fixed point iteration for effective automatic parameter choice

Preprints, MATLAB code:

`http://www.uni-graz.at/~clason/publications.html`