Parameter identification problems with non-Gaussian noise

Christian Clason

Institut für Mathematik und wissenschaftliches Rechnen
Karl-Franzens-Universität Graz

MFO Workshop Computational Inverse Problems
Oberwolfach, October 26, 2012

Motivation

Tikhonov functional

$$\min_{u \in U} \mathcal{F}(S(u), y^\delta) + \alpha \mathcal{R}(u)$$

- $S : U \subset X \to Y$ (not necessarily linear); X,Y Banach spaces
- $y^\delta \in Y$ noisy data

- $\mathcal{R} : X \to \mathbb{R}$ regularization term: a priori information on solution
- $\mathcal{F} : Y \times Y \to \mathbb{R}$ discrepancy term: a priori information on noise

Discrepancy: Examples

- Gaussian noise:

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

- Impulsive noise:

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

- Uniform noise:

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

- Poisson noise:

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

Why Newton methods?

Motivation: parameter identification in PDEs

- S (at least) twice Fréchet differentiable
- Directional derivatives computable as solutions to adjoint, linearized (adjoint) equations
- S has smoothing properties

Can this be exploited for superlinear convergence if $\mathcal{F}$ is not smooth?

1 Motivation

2 Semi-smooth Newton method
- Semi-smoothness in finite dimensions
- Semi-smoothness in function spaces

3 Parameter identification with non-Gaussian noise
- L^1 data fitting
- L^∞ data fitting

Motivation: Generalized Newton method

Consider $F : X \to Y$, X, Y Banach spaces

Generalized Newton method for $F(u) = 0$

- Choose $u^0 \in X$ (close to solution u^*)
- For $k = 0, 1, \dots$
  1. Choose invertible $M_k \in \mathcal{L}(X, Y)$
  2. Solve for s^k in
  $$M_k s^k = -F(u^k)$$
  3. Set $u^{k+1} = u^k + s^k$.

Motivation: Generalized Newton method

Generalized Newton method for $F(u) = 0$

- Choose $u^0 \in X$ (close to solution u^*)
- For $k = 0, 1, \ldots$
  1. Choose invertible $M_k \in \mathcal{L}(X, Y)$
  2. Solve for s^k in
  $$M_k s^k = -F(u^k)$$
  3. Set $u^{k+1} = u^k + s^k$.

- When does the method converge, i.e., $\left\|u^k - u^*\right\|_X \to 0$?
- When is the convergence superlinear, i.e., $\left\|u^{k+1} - u^*\right\|_X / \left\|u^k - u^*\right\|_X \to 0$?

Convergence of generalized Newton method

Let $d^k = u^k - u^*$. Then:

$$\left\| u^{k+1} - u^* \right\|_X = \left\| M_k^{-1}(F(u^* + d^k) - F(u^*) - M_k d^k) \right\|_X$$

⇝ Superlinear convergence under

1. Regularity condition

$$\|M_k^{-1}\|_{\mathcal{L}(Y,X)} \leq C \text{ for all } k$$

2. Approximation condition

(A) $$\lim_{\|d^k\|_X \to 0} \frac{\|F(u^* + d^k) - F(u^*) - M_k d^k\|_Y}{\|d^k\|_X} = 0$$

Semi-smooth Newton methods

Idea: if F is non-smooth, choose $M_k \in \partial F(u^k)$ (subdifferential)

$\rightsquigarrow$ Semi-smooth Newton method if (A) holds

- in finite dimensions: ∂F Clarke's subdifferential
  [Mifflin 1977, Kummer 1992, Qi/Sun 1993]
- in infinite dimensions: Clarke's subdifferential not available
  $\rightsquigarrow$ directly require condition (A) and uniform boundedness
  [Chen/Nashed/Qi 2000, Hintermüller/Ito/Kunisch 2002]
- Connection: semi-smooth functions on $\mathbb{R}$ define semi-smooth superposition operators on $L^p(\Omega)$
  [Ulbrich 2002/03/11, Schiela 2008]

Semi-smoothness in finite dimensions

Consider $f : \mathbb{R}^m \to \mathbb{R}^n$ locally Lipschitz.

Directional derivative

$$f'(x;h) = \lim_{t \to 0^+} \frac{f(x+th) - f(x)}{t}$$

Clarke's subdifferential

$$\partial f(x) = \text{co} \left\{ \lim_{y \to x} f'(y) \right\}$$

(Rademacher's Theorem: f Lipschitz $\Rightarrow$ f differentiable a.e.)

Semi-smoothness in finite dimensions

$f : \mathbb{R}^m \to \mathbb{R}^n$ is semi-smooth at x, iff

1. f is Lipschitz near x,
2. $f'(x, h)$ exists for all h,
3. for all $M \in \partial f(x + h)$,

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - Mh\|}{\|h\|} \to 0$$

(equivalent to original definition: $\lim\limits_{\substack{M \in \partial f(x+th') \\ h' \to h, t \to 0^+}} \{Mh'\}$ exists for all h)

Examples of semi-smooth functions

- f continuously differentiable, $\partial f(x) = \{f'(x)\}$
- f continuous, piecewise continuously differentiable:

$$f(x) \in \{f^1(x), \dots, f^N(x)\} \subset C^1(\mathbb{R}^m) \quad \text{for all } x$$

with

$$\partial f(x) = \text{co}\left\{(f^i)'(x) : f(x) = f^i(x)\right\}$$

($\rightsquigarrow$ differentiate piecewise, convex hull at connection)

- Sum, composition of semi-smooth functions

Examples of semi-smooth functions

Specifically: $f : \mathbb{R} \to \mathbb{R}$,

- $f(t) = \max(0, t)$,

$$\partial f(t) = \begin{cases} \{0\} & t < 0 \\ \{1\} & t > 0 \\ [0, 1] & t = 0 \end{cases}$$

- $f(t) = |t|$,

$$\partial f(t) = \begin{cases} \{-1\} & t < 0 \\ \{1\} & t > 0 \\ [-1, 1] & t = 0 \end{cases}$$

Examples of semi-smooth functions

Specifically: $f : \mathbb{R} \to \mathbb{R}$,

- $f(t) = \operatorname{sign}_\beta(t) := \begin{cases} 1 & t > \beta \\ -1 & t < -\beta \\ \frac{1}{\beta} t & |t| \leq \beta \end{cases}$

$$\partial f(t) = \begin{cases} \{0\} & |t| > \beta \\ \{\beta^{-1}\} & |t| \leq \beta \\ [0, \beta^{-1}] & |t| = \beta \end{cases}$$

($f(t) = \operatorname{sign}(t)$ is not semi-smooth)

Semi-smoothness in infinite dimensions

- $F : X \to Y$ is Newton differentiable at u if there exists neighborhood $N(u)$ and mapping $G : N(u) \to \mathcal{L}(X, Y)$ with

$$\lim_{\|h\|\to 0} \frac{\|F(u+h) - F(u) - G(u+h)h\|_Y}{\|h\|_X} \to 0$$

  Newton derivative at u is $D_N F \in \{G(s) : s \in N(u)\}$

- $F : X \to Y$ is semi-smooth at u if N-differentiable and

$$\lim_{t\to 0^+} G(u+th)h$$

  exists uniformly for $\|h\|_X = 1$

Examples of semi-smooth operators

- F continuously Fréchet differentiable, $D_N F(u) = F'(u)$
- $H := \alpha F + \beta G$ for semi-smooth F, G,

$$D_N H = \alpha D_N F + \beta D_N G \quad \text{(sum rule)}$$

- $F \circ H$ with $H : X \to Y$ continuously F-differentiable, $F : Y \to Z$ N-differentiable at $H(u)$,

$$D_N(F \circ H)(u + h) = D_N F(H(u + h))H'(u + h)$$

is N-derivative at u for h sufficiently small (chain rule)

Examples of semi-smooth operators

If $\psi : \mathbb{R} \to \mathbb{R}$ semi-smooth, $1 \leq p < q \leq \infty$, then

Superposition operator

$$\Psi : L^q(\Omega) \to L^p(\Omega), \quad (\Psi(u))(x) := \psi(u(x))$$

is semi-smooth with respect to u,

Newton derivative at u

$$D_N(\Psi(u))(x) \in \partial(\psi(u(x)))$$

Superposition operators: examples

- $\psi(t) = \max(0, t)$, $\Psi : u(x) \mapsto \max(0, u(x))$,

$$[D_N\Psi(u)h](x) = \begin{cases} 0 & u(x) < 0 \\ h(x) & u(x) > 0 \\ \delta\, h(x) & u(x) = 0 \end{cases}$$

  for any $\delta \in [0, 1]$ (and similarly $u(x) \mapsto \min(0, u(x))$)

- $\psi(t) = \text{sign}_\beta(t)$, $\Psi : u(x) \mapsto \text{sign}_\beta(u(x))$,

$$[D_N\Psi(u)h](x) = \begin{cases} 0 & |u(x)| \geq \beta \\ \frac{1}{\beta}h(x) & |u(x)| < \beta \end{cases}$$

Semi-smooth Newton method

Consider $F : X \to Y$, $F(u^*) = 0$, where

- $F : X \to Y$ is Newton differentiable with N-derivative $D_N F$
- for all $\|s - u^*\|_X$ sufficiently small,

$$\left\|(D_N F)(s)^{-1}\right\|_{\mathcal{L}(Y,X)} \leq C$$

- $\left\|u^0 - u^*\right\|_X$ is sufficiently small

Then, iterates

$$u^{k+1} = u^k - D_N F(u^k)^{-1} F(u^k)$$

converge (locally) superlinearly to u^*

Application to non-smooth optimization

Goal: Newton-type method for minimizing $\|\cdot\|_{L^1}$, $\|\cdot\|_{L^\infty}$

Challenges:

1. Not twice (Newton-)differentiable

   $\rightsquigarrow$ smoothing approach to obtain N-differentiable F-derivative

2. Local convergence of Newton method

   $\rightsquigarrow$ continuation strategy in smoothing parameter

Application: $\|\cdot\|_{L^1}$ minimization

- Huber smoothing of $\mathcal{F}(u) = \|u\|_{L^1}$:

$$\mathcal{F}_\beta(u) = \int |u(x)|_\beta \, dx, \quad |t|_\beta = \begin{cases} t - \frac{\beta}{2} & t > \beta \\ -t - \frac{\beta}{2} & t < -\beta \\ \frac{1}{2\beta} t^2 & |t| \le \beta \end{cases}$$

$$\mathcal{F}_\beta'(u) = \text{sign}_\beta(u)$$

$\rightsquigarrow$ $\mathcal{F}_\beta'$ semi-smooth from L^q to L^p, $q > p$

$\rightsquigarrow$ convergence $\mathcal{F}_\beta(u) \to \mathcal{F}(u)$ for $\beta \to 0$

Application: $\|\cdot\|_{L^\infty}$ minimization

- Reformulation: $\min_u J(u) + \|u\|_{L^\infty}$ equivalent to

$$\min_{u,c} J(u) + c \quad \text{subject to} \quad \|u\|_{L^\infty} \leq c$$

- Moreau–Yosida regularization of $\mathcal{F}(u) = I_{\{\|\cdot\|_{L^\infty} \leq c\}}(u)$:

$$\mathcal{F}_\gamma(u) = \frac{\gamma}{2}\left[\|\max(0, u - c)\|_{L^2}^2 + \|\min(0, u + c)\|_{L^2}^2\right],$$

$$\mathcal{F}_\gamma'(u) = \gamma\left[\max(0, u - c) + \min(0, u + c)\right]$$

$\rightsquigarrow$ $\mathcal{F}_\gamma'$ semi-smooth from L^q to L^p, $q > p$

$\rightsquigarrow$ convergence $\mathcal{F}_\gamma(u) \to \mathcal{F}(u)$ for $\gamma \to \infty$

Model problems

1. Potential problem: $S : U \subset L^2(\Omega) \to H^1(\Omega), u \mapsto y,$

$$\langle \nabla y, \nabla v \rangle_{L^2} + \langle uy, v \rangle_{L^2} = \langle f, v \rangle_{L^2} \qquad \text{for all } v \in H^1(\Omega)$$

2. Robin problem: $S : U \subset L^2(\Gamma_i) \to H^{1/2}(\Gamma_c), u \mapsto y|_{\Gamma_c},$

$$\langle \nabla y, \nabla v \rangle_{L^2} + \langle uy, v \rangle_{L^2(\Gamma_i)} = \langle f, v \rangle_{L^2(\Gamma_c)} \qquad \text{for all } v \in H^1(\Omega)$$

3. Conductivity problem: $S : U \subset H^1(\Omega) \to H^1_0(\Omega), u \mapsto y,$

$$\langle u\nabla y, \nabla v \rangle_{L^2} = \langle f, v \rangle_{L^2} \qquad \text{for all } v \in H^1_0(\Omega)$$

Properties

- S uniformly bounded in $U \subset X$; $u_n \rightharpoonup u$ in X implies

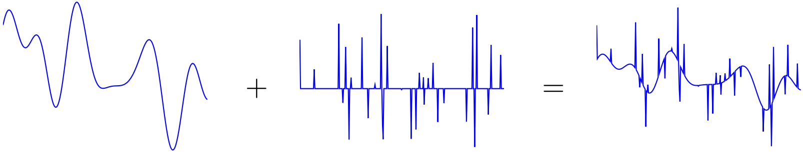
$$S(u_n) \to S(u) \quad \text{in } \mathrm{L}^2(\Omega)$$

- S twice Fréchet differentiable with uniformly bounded derivatives
(Directional derivatives given by solution of linearized equations, computable using formal Lagrangian technique)

- $S(U) \subset \mathrm{L}^q(\Omega), q > 2$

L^1 data fitting: Motivation

Here: data subject to impulsive noise:



data noise measurement

- Appears in digital image acquisition, processing (hardware defects, cosmic rays, …)
- Characterization: noise is "sparse" $\rightsquigarrow$ L^1 minimization

L^1 data fitting: Problem

$$\min_{u \in X} \|S(u) - y^\delta\|_{L^1} + \frac{\alpha}{2} \|u\|_X^2$$

- $y^\delta \in L^q(\Omega)$, $q > 2$, data with impulsive noise
- X Hilbert space (e.g. $L^2(D)$, $H^1(D)$)
- Y has compact embedding in $L^q(\Omega)$, $q > 2$

- Regularization properties:
  [Hofmann/Kaltenbacher/Pöschl/Scherzer 2007]
  [Scherzer/Grasmair/Grossauer/Haltmeier/Lenzen 2009]

L^1 data fitting: Approximation

Huber smoothing

$$\min_{u\in X} \|S(u) - y^\delta\|_{L^1_\beta} + \frac{\alpha}{2}\|u\|_X^2$$

Optimality conditions

$$(OS_\beta) \qquad \alpha u_\beta + S'(u_\beta)^*(\operatorname{sign}_\beta(S(u_\beta) - y^\delta)) = 0$$

Semi-smooth Newton method

Consider (OS_β) as $F(u) = 0$ for $F : X \to X^*$,

$$F(u) = \alpha u + S'(u)^*(\operatorname{sign}_\beta(S(u) - y^\delta))$$

$S(u) \in L^q, q > 2 \rightsquigarrow P(u) = \operatorname{sign}_\beta(S(u) - y^\delta)$ semi-smooth

Newton derivative

$$\begin{aligned} D_N P(u)h &= \chi_{\mathcal{I}} \beta^{-1}(S'(u)h) \\ &= \begin{cases} \beta^{-1}(S'(u)h) & \text{if } |(S(u) - y^\delta)| \leq \beta \\ 0 & \text{else} \end{cases} \end{aligned}$$

Semi-smooth Newton method

Semi-smooth Newton step

$$D_N F(u^k)\delta u = \alpha\delta u + (S''(u^k)\delta u)^* P(u^k) + \tfrac{1}{\beta} S'(u^k)^*(\chi_{\mathcal{I}^k} S'(u^k)\delta u) = -F(u^k)$$

Action of $D_N F(u^k)$ on given δu can be calculated by solving linearized state, adjoint equation (obtained by formal Lagrangian approach)

$\rightsquigarrow$ solve using matrix-free Krylov-method (GMRES, BiCGStab)

Semi-smooth Newton method

But: superlinear convergence requires regularity condition;
S nonlinear, functional non-convex $\rightsquigarrow$ assume for $c > 0$

Local coercivity condition

$$\langle S''(u_\beta)(h,h), P(u_\beta)\rangle_{L^2} + \alpha\|h\|_X^2 \geq c\|h\|_X^2 \quad \text{for all } h \in X$$

(cf. sufficient second order conditions in optimization)

Here: satisfied for

- large α (for large noise)
- large β or sparse residual (for small noise) ($\Rightarrow P(u_\beta)$ small)

Implies regularity condition, thus local superlinear convergence of semi-smooth Newton method

Numerical example

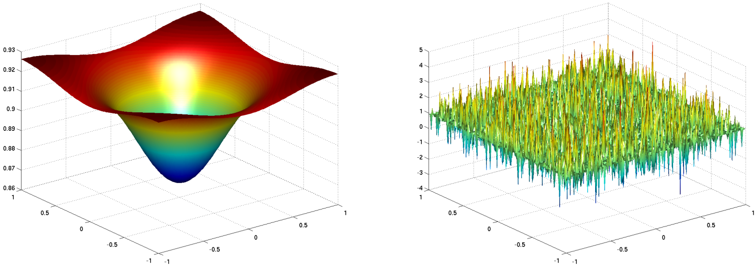
- Potential problem: $y = S(u)$ solves $\langle \nabla y, \nabla v\rangle + \langle uy, v\rangle = \langle f, v\rangle$
- Discretization with linear finite elements, $N = 128 \times 128$ nodes
- Random impulsive noise:

$$y^\delta(x) = \begin{cases} y^\dagger(x) + \|y^\dagger\|_{L^\infty}\xi(x), & \text{with probability } r \\ y^\dagger(x), & \text{else} \end{cases}$$

  $y^\dagger = S(u^\dagger)$, $\xi(x)$ normally distributed random variable

- Choice of α by fixed point iteration (2–4 iterations)
- Termination of continuation at $\beta \approx 10^{-7}$

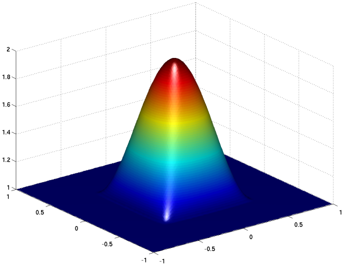
Results: potential problem, $r = 0.3$



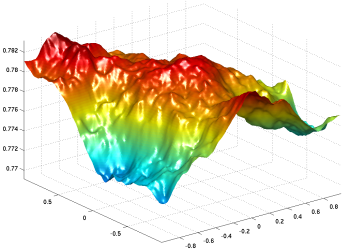
(a) exact data $y^\dagger$

(b) noisy data y^δ

Results: potential problem, $r = 0.3$

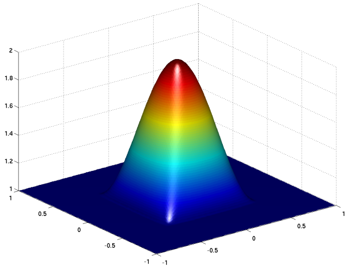


(c) exact solution $u^{\dagger}$

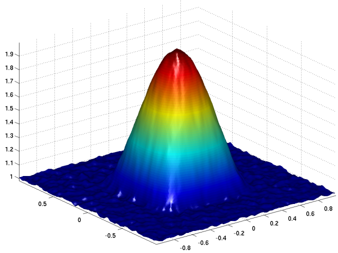


(d) reconstruction u_{α} (L^2)

Results: potential problem, $r = 0.3$



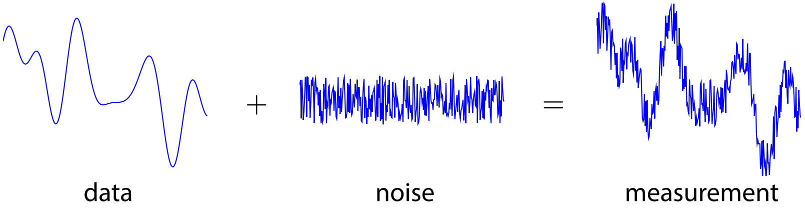
(e) exact solution $u^\dagger$



(f) reconstruction u_α (L^1)

L^∞ data fitting: Motivation

Here: data subject to uniform noise:



- Appears in digital acquisition, processing (quantization errors)
- Maximum likelihood estimate $\rightsquigarrow$ L^∞ minimization

L^∞ data fitting: Problem

$$\min_{u \in X} \|S(u) - y^\delta\|_{L^\infty} + \frac{\alpha}{2}\|u\|_X^2$$

- $y^\delta \in L^\infty(\Omega)$ data with uniform noise
- X Hilbert space (e.g. $L^2(D)$, $H^1(D)$)
- Y has compact embedding in $L^\infty(\Omega)$

- Regularization properties:
  [Hofmann/Kaltenbacher/Pöschl/Scherzer 2007]
  [Scherzer/Grasmair/Grossauer/Haltmeier/Lenzen 2009]

L^∞ data fitting: Approximation

Reformulation

$$\min_{(u,c)\in X\times \mathbb{R}} c + \frac{\alpha}{2}\|u\|_X^2 \quad \text{subject to} \quad \|S(u) - y^\delta\|_{L^\infty(\Omega)} \leq c$$

Moreau–Yosida approximation

$$\begin{aligned}\min_{(u,c)\in X\times \mathbb{R}} c + \frac{\alpha}{2}\|u\|_X^2 &+ \frac{\gamma}{2}\|\max(0, S(u) - y^\delta - c)\|_{L^2(\Omega)}^2 \\ &+ \frac{\gamma}{2}\|\min(0, S(u) - y^\delta + c)\|_{L^2(\Omega)}^2\end{aligned}$$

Optimality conditions

Optimality system

$$\begin{cases} \alpha u_\gamma + S'(u)^* \left((S(u) - y^\delta - c)^+ + (S(u) - y^\delta + c)^- \right) = 0, \\ 1 + \int_\Omega (-(S(u) - y^\delta - c)^+ + (S(u) - y^\delta + c)^- \, dx = 0. \end{cases}$$

with $(\cdot)^+ = \max(0, \cdot), \quad (\cdot)^- = \min(0, \cdot)$

$\rightsquigarrow$ consider as $F(u, c) = 0$ for $F : X \times \mathbb{R} \to X^* \times \mathbb{R}$

Semi-smooth Newton method

Smoothing properties of S, embedding $c \in \mathbb{R} \hookrightarrow L^\infty$
$\rightsquigarrow F(u,c)$ semi-smooth in u and c

Newton derivatives

$$D_{N,u}(S(u) - y^\delta - c)^+ h = \chi_\mathcal{A}(S'(u)h)$$

$$= \begin{cases} S'(u)h & \text{if } S(u) - y^\delta \geq c \\ 0 & \text{else} \end{cases}$$

$$D_{N,c}(S(u) - y^\delta - c)^+ h = -h \int_\Omega \chi_\mathcal{A}(x)\, dx$$

Semi-smooth Newton method

Semi-smooth Newton step

$$\begin{pmatrix} D_{N,u}F_1(u^k,c^k) & D_{N,c}F_1(u^k,c^k) \\ D_{N,u}F_2(u^k,c^k) & D_{N,c}F_2(u^k,c^k) \end{pmatrix} \begin{pmatrix} \delta u \\ \delta c \end{pmatrix} = - \begin{pmatrix} F_1(u^k,c^k) \\ F_2(u^k,c^k) \end{pmatrix}$$

Action on given $\delta u, \delta c$ can be calculated by solving linearized state, adjoint equation (obtained by formal Lagrange approach)

$\rightsquigarrow$ solve using matrix-free Krylov-method (GMRES, BiCGStab)

Semi-smooth Newton method

Local coercivity condition

$$\langle S''(u_\gamma)(h,h), \gamma(S(u_\gamma) - y^\delta - c_\gamma)^+ + \gamma(S(u_\gamma) - y^\delta + c_\gamma)^- \rangle_{L^2} + \alpha\|h\|_X^2 \geq c\|h\|_X^2 \quad \text{for all } h \in X$$

Here: satisfied for

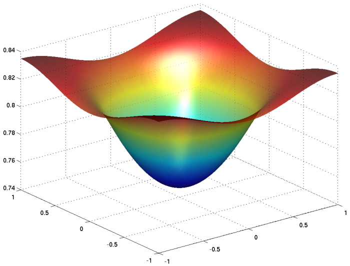
- large α (for large noise)
- small γ or small residual (for small noise)

- Implies regularity condition, superlinear convergence
- Continuation in $\gamma \to \infty$ for globalization

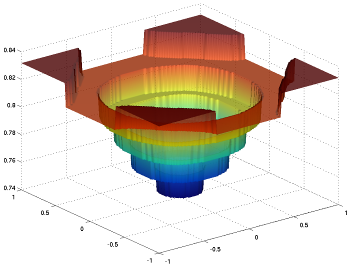
Numerical example

- Potential problem: $y = S(u)$ solves $\langle \nabla y, \nabla v\rangle + \langle uy, v\rangle = \langle f, v\rangle$
- Discretization with linear finite elements, $N = 128 \times 128$ nodes
- Quantization noise: round $y^\dagger = S(u^\dagger)$ to n_b nearest values
- Choice of α by fixed point iteration (7 iterations)
- Termination of continuation at $\gamma \approx 10^{10}$

Results: potential problem, $n_b = 6$

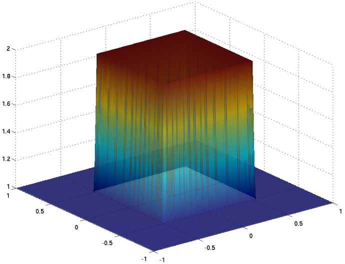


(a) exact data $y^\dagger$

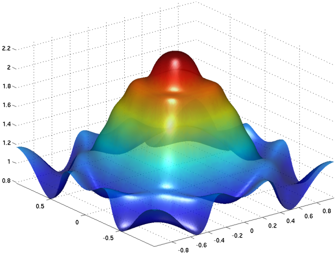


(b) noisy data y^δ

Results: potential problem, $n_b = 6$

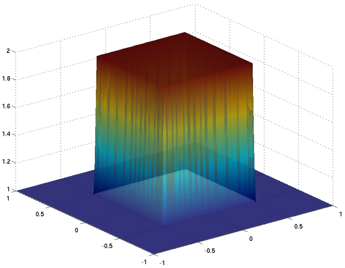


(c) exact solution $u^\dagger$

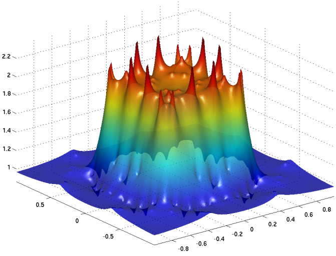


(d) reconstruction u_α (L^2)

Results: potential problem, $n_b = 6$



(e) exact solution $u^\dagger$



(f) reconstruction u_α (L^∞)

Conclusion

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

- Noise structure more important than noise level
- Semi-smooth Newton methods allow solution of non-smooth problems

Outlook:

- Mixed noise (impulsive+Gaussian, Cauchy, Rician)
- Banach space regularization
- Applications

Preprints, MATLAB/Python codes:

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