THEORY AND NUMERICAL PRACTICE FOR OPTIMIZATION PROBLEMS INVOLVING $\ell^p$-FUNCTIONALS, WITH $p \in [0, 1)$

Daria Ghilli, K.Kunisch
Institute of Mathematics and Scientific Computing, University of Graz

PGMO Days
November, 2017
1. **SPARSE OPTIMIZATION PROBLEMS OF THE FORM:**

\[
\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} |Ax - b|^2 + \beta |\Lambda x|^p,
\]

where \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, \Lambda \in \mathbb{R}^{r \times n}, \beta \in \mathbb{R}^+, \ 0 \leq p < 1. \)

\[|x|^p = \left( \sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}}: \]

- quasi-norm for \( 0 < p < 1. \)
- \( |x|_0 = \sum_{k=1}^{\infty} |x_k|^0 = \) number of nonzero elements of \( x. \)

2. **SPECIFICITIES:**

- \( 0 \leq p < 1: \) nonconvex, nonsmooth.
- \( \Lambda \in \mathbb{R}^{r \times n}. \)
APPLICATIONS

1. **provides sparsity**, e.g.:
   - **image restoration**:
     - \( \Lambda = \| \cdot \|_1 \): sparsity of solutions.
     - \( \Lambda = \nabla_h \): corresponds to the total variation, provides better edges preservation.
   - **optimal control**:
     - \( \Lambda = \| \cdot \|_1 \): sparsity of solutions.
     - \( \Lambda = \nabla_h \): enhances piecewise constant solutions.
   - **compressed sensing**: \( |x|_p^p, 0 < p < 1 \) allows a smaller number of measurements.

2. **modelling purposes**: **fracture mechanics**.
   - \( \Lambda = \nabla_h \) and \( 0 < p < 1 \) models the formation of the crack opening
Outline

1. AIMS:
   - Existence.
   - Convenient optimality conditions.
   - Numerical schemes for (P).

2. METHODS:
   - Monotone primal-dual active set schemes to solve the opt. cond.

3. NUMERICAL EXPERIENCE:
   - Optimal control problems
   - Fracture mechanics
   - Microscopy imaging

Remark

More general framework:

\[
\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} |Ax - b|^2 + \sum_{i=1}^{r} \phi_{\beta, p}(\Lambda x_i)
\]

\(\phi_{\beta, p}: \mathbb{R} \to \mathbb{R}^+\)

SCAD: MCP
Outline

1. AIMS:
   - Existence.
   - Convenient optimality conditions.
   - Numerical schemes for (P).

2. METHODS:
   - Monotone primal-dual active set schemes to solve the opt. cond.

3. NUMERICAL EXPERIENCE:
   - optimal control problems
   - fracture mechanics
   - microscopy imaging

Remark
More general framework:

\[
\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} |Ax - b|^2 + \sum_{i=1}^{r} \phi_{\beta, p}(\Lambda x)_i, \quad \phi_{\beta, p} : \mathbb{R} \rightarrow \mathbb{R}^+ \text{ s.t.}
\]

ℓ^p: \hspace{1cm} SCAD: \hspace{1cm} MCP:
Remember

K. Ito, K. Kunisch (2014): problem (P) with

- $\Lambda = I$ in infinite dimensional sequence spaces $\ell^p$, $p \in [0, 1)$.
- main focus on the $\ell^0$ case (numerically).

Case $0 < p < 1$ never tested.

**Theorem (Existence $0 < p \leq 1$, K. Ito, K. Kunisch)**

Consider

$$\min J(x) = \frac{1}{2} |Ax - b|^2_2 + \beta |x|^p_p$$

For any $\beta > 0$, and $A \in \mathcal{L}(\ell^2)$ there exists a solution $\bar{x} \in \ell^p$ to (P).

"Proof"

$\gamma : \ell^2 \to \ell^p : \quad \gamma(y)_i = |y_i|^\frac{2}{p} \text{ sgn}(y_i)$ is homeomorphism

$\gamma : \ell^2 \to \ell^2$ is weakly sequentially continuous

$$\min_{y \in \ell^2} \frac{1}{2} |A\gamma(y) - b|^2_2 + \beta |y|^2_2$$

has a solution $\bar{y}$, hence $\bar{x} = \gamma(\bar{y})$ is a solution to (P).
**General operator Λ**

**Theorem (Existence, 0 < p ≤ 1)**

Consider

\[
\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} |Ax - b|^2_2 + \beta |Λx|^p_p
\]

(P)

For any \( \beta > 0 \), and \( A \in \mathbb{R}^{m \times n}, Λ \in \mathbb{R}^{r \times n} \) such that

\[
\text{Ker}A \cap \text{Ker}Λ = \{0\},
\]

(K)

there exists a solution \( \bar{x} \in \mathbb{R}^n \) to (P).

"Proof": assumption (K) ensures coercivity, from which existence follows.
MONOTONE CONVERGENT ALGORITHM

Regularized version of (P), for $\varepsilon > 0$:

$$
\min_{x \in \mathbb{R}^n} J_\varepsilon(x) = \frac{1}{2} |Ax - b|^2 + \beta \psi_\varepsilon(|\Lambda x|^2), \quad (P_\varepsilon)
$$

where for $t \geq 0$

$$
\psi_\varepsilon(t) = \begin{cases} 
\frac{p}{2} \frac{t}{\varepsilon^2 - p} + (1 - \frac{p}{2})\varepsilon^p & \text{for } 0 \leq t \leq \varepsilon^2 \\
\frac{p}{t^{\frac{p}{2}}} & \text{for } t \geq \varepsilon^2,
\end{cases}
$$

Note that

$$
\psi'_\varepsilon(t) = \frac{p}{2 \max(\varepsilon^{2-p}, t^{\frac{2-p}{2}})}, \quad \text{for } t \geq 0,
$$

hence $\psi_\varepsilon$ is $C^1$ and concave on $[0, \infty)$. Also $t \to \psi_\varepsilon(t^2) \in C^1(-\infty, \infty)$. 
Necessary optimality condition for \((P_e)\):

\[
A^* A x + \Lambda^* \frac{\beta p}{\max(\varepsilon^{2-p}, |\Lambda x|^{2-p})} \Lambda x = A^* b, \quad (E_e)
\]

Algorithm 1 Monotone iterative algorithm

1: Initialize \(x^0\), \(\varepsilon^0\), and set \(y^0 = \Lambda x^0\). Set \(k = 0\);
2: repeat
3: \hspace{1em} Solve for \(x^{k+1}\)
   \[
   A^* A x^{k+1} + \Lambda^* \frac{\beta p}{\max(\varepsilon^{2-p}, |\Lambda x^k|^{2-p})} \Lambda x^{k+1} = A^* b.
   \]
4: \hspace{1em} Set \(k = k + 1\).
5: until the stopping criterion is fulfilled.
6: Reduce \(\varepsilon\) and repeat 2.
**Theorem (Convergence result)**

**Convergence:** Every cluster point of $x^k$, of which there exists at least one, is a solution of $(E_e)$.

**Monotonicity:** $J_\varepsilon(x^k)$ is strictly monotonically decreasing unless there exists some $k$ such that $x^k = x^{k+1}$ and $x^k$ satisfies $(E_e)$.

**Sketch of the proof, monotonicity:**
By the concavity of $\Psi_\varepsilon$ and some algebraic manipulations, we get for some constant $\kappa > 0$

$$J_\varepsilon(x^{k+1}) + \frac{1}{2} |A(x^{k+1} - x^k)|_2^2 + \kappa |\Lambda(x^{k+1} - x^k)|_2^2 \leq J_\varepsilon(x^k),$$

from which the monotonicity follows. $\square$
Consider

\[-\Delta y = f \text{ in } \Omega, \quad y = 0 \text{ on } \partial\Omega.\]

We study

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} |Ax - b|^2 + \beta |\Lambda x|^p,
\]

(M)

where

- \(A^*A\) is an \(M\) matrix coinciding with the discretized Laplacian
- \(\Lambda\) is the backward finite difference gradient

For \(\beta > 0\), (M) gives a solution piecewise constant enhancing behaviour.
$\beta = 0.05$

$\beta = 0.08$

$\beta = 0.3$

$p = .1$, mesh size $h = \frac{1}{64}$.

$p = .05$, mesh size $h = \frac{1}{64}$. 
**Optimal control problem**

We consider the linear control system

\[
\frac{d}{dt}y(t) = Ay(t) + Bu(t) \quad y(0) = 0,
\]

that is,

\[
y(T) = \int_0^T e^{A(T-s)} Bu(s)ds.
\]

One dimensional controlled heat equation \( x \in (0, 1), y(0) = y(1) = 0: \)

\[
y_t = y_{xx} + d_1(x)u_1(t) + d_2(x)u_2(t),
\]

After discretization in space and time

\[
A\vec{u} = \sum_{k=1}^m e^{A(T-t_k-\frac{\Delta t}{2})}(B\vec{u})_k \Delta t, \quad \vec{u} = (u_1^1, \cdots, u_m^1, u_1^2, \cdots, u_m^2).
\]

Sparsity formulation:

\[
\min_{u \in \mathbb{R}^{2m}} J(u) = \frac{1}{2} \|Au - b\|^2_2 + \beta \|\Lambda u\|^p_p,
\] \hspace{1cm} (TC)

\( \Lambda \) is the backward finite difference operator acting on each component of the control.
Numerical results

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$10^{-3}$</th>
<th>$10^{-2}$</th>
<th>$10^{-1}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>iterations</td>
<td>630</td>
<td>635</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>$</td>
<td>\Lambda u</td>
<td>_0$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>\Lambda u</td>
<td>_p$</td>
<td>158</td>
<td>16.7</td>
</tr>
<tr>
<td>Residue</td>
<td>$3 \times 10^{-3}$</td>
<td>$2 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-3}$</td>
<td>$2.5 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Optimal control problem, $p = .5$, mesh size $h = \frac{1}{50}$.

- **Sparsity:** $|\Lambda u|_0, |\Lambda u|_p$ decrease for $\beta$ increasing.
- **Tested for values of $p$ near to 1** (e.g. for $p = .9$): less piecewise constant behaviour compared to $p = .5$. 
Quasi-static evolutions of cohesive fractures

$u : \Omega \rightarrow \mathbb{R}$ is the displacement function

minimization of the energy functional

$$I(u) = \frac{1}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 \, dx + \beta \theta(\jump{u}_{\Gamma})$$

where

$\Gamma$ is the set of cracks;

$\jump{u}_{\Gamma} = u^+_{\Gamma} - u^-_{\Gamma}$ denotes the jump of the displacement on $\Gamma$;

$\beta > 0$ is a material parameter;

external force $g : \Omega \times [0, T] \rightarrow \mathbb{R}$. We require

$$u|_{\partial \Omega} = g|_{\partial \Omega}.$$
Cohesive fracture energy

- $\theta$ is a monotonic non-decreasing function of $[u]$: 
- $\theta'$ decreasing $\rightarrow$ $\theta$ nonconvex/nonsmooth.

Dugdale model (1960)
Barenblatt model (1962).

- In our model $\theta([u]_\Gamma) = |[u]_\Gamma|^p$, nonconvex, not differentiable in zero.

Numerical experiments
After discretization of $\Omega$, solve the unconstrained minimization problem:

$$\min |Au_h - g|^2_2 + \beta |[u_h]_\Gamma|^p,$$
Nonhomogeneous material, nonconstant B.C.

\[ t = 1.5 \]
\[ t = 2 \]
\[ t = 3 \]

Constant (in \( y \)) boundary condition, nonhomogeneous (in \( y \)) material

\[ t = 0.2 \]
\[ t = 1.5 \]
\[ t = 3 \]

Nonconstant boundary condition \( g = \frac{1}{100} \cos(2(y - 0.5))(x - 0.5) \), nonhomogeneous (in \( y \)) material (as above)
**Comparison with GIST**

- GIST "General Iterative Shrinkage and Thresholding", toolbox by Tuia, Rémi Flamary, Michel Barlaud, 2016\(^1\).
- GIST solves (P) \((\Lambda = I)\) by generating \(\{x^k\}\) via

\[
x^{k+1} = \arg\min_x \frac{1}{2} \|x - u^k\|^2 + \frac{1}{t^k} \|x\|_p^p, \text{ where } u^k = x^k - \frac{Ax^k - b}{t^k}
\]

<table>
<thead>
<tr>
<th>(\beta)</th>
<th>0.001</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>itermon</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>time(_{mon})</td>
<td>0.085</td>
<td>0.082</td>
<td>0.39</td>
<td>0.387</td>
<td>0.478</td>
<td>0.673</td>
</tr>
<tr>
<td>time(_{GIST})</td>
<td>0.445</td>
<td>0.563</td>
<td>0.701</td>
<td>0.444</td>
<td>0.468</td>
<td>0.461</td>
</tr>
</tbody>
</table>

Time, iterations, value of J to which the Monotone algorithm overcome GIST’s.

**Remark**

The time of the Monotone is almost always smaller than GIST’s

\(^{1}\)https://github.com/rflamary/nonconvex-optimization
Next aim: find a solution of the original problem:

\[
\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} |Ax - b|^2 + \beta |\Lambda x|^p
\]  

(P)

Denote by \( B_i = \left| (A\Lambda^{-1})_i \right|^2 \), \( i \)-th column.

**Theorem (Necessary optimality conditions)**

Let \( \bar{x} \) be a global minimizer of (P). Denote \( \bar{y} = \Lambda \bar{x} \). Then:

**Case** \( 0 < p < 1 \):

\[
\begin{cases}
A^* (A\bar{x} - b) + \Lambda^* \lambda = 0 \\
(\Lambda \bar{x})_i = 0 \\
|(|\Lambda \bar{x}|)_i > 0 \text{ and } \lambda_i = \frac{\beta p (|\Lambda \bar{x}|)_i}{(|\Lambda \bar{x}|)_i^{2-p}} 
\end{cases}
\]

if \( |B_i \bar{y}_i + \lambda_i| < \mu_i \)

\[|(|\Lambda \bar{x}|)_i > 0 \text{ and } \lambda_i = \frac{\beta p (|\Lambda \bar{x}|)_i}{(|\Lambda \bar{x}|)_i^{2-p}} \]

if \( |B_i \bar{y}_i + \lambda_i| > \mu_i \),

where \( \mu_i = \beta \frac{1}{2-p} (2 - p)(2(1 - p))^{-\frac{1}{2-p}} B_i^{1 - \frac{2-p}{2}} \). If \( |B_i \bar{y}_i + \lambda_i| = \mu_i \), then

\( (\Lambda \bar{x})_i = 0 \) or \( (\Lambda \bar{x})_i = \left( \frac{2\beta(1-p)}{B_i} \right)^{\frac{1}{2-p}} \text{ sgn} (B_i \bar{y}_i + \lambda_i) \).
Algorithm 2 Primal-dual active set monotone (PDASM)-general structure

1:

▶ OUTER LOOP: Primal dual active set strategy.

- active indexes $A_n = \{i : |B_i y^n_i + \lambda^n_i| \leq \mu_i\}$; inactive ones $I_n = A_n^c$.
- solve in $x^{n+1}, \lambda^{n+1}$

$$
\begin{cases}
A^*(Ax^{n+1} - b) + \Lambda^* \lambda^{n+1} = 0 \\
(\Lambda x^{n+1})_i = 0 \\
\lambda_i^{n+1} = \frac{\beta \rho (\Lambda x^{n+1})_i}{\max(\epsilon^{2-\rho}, |(\Lambda x^{n+1})_i|^{2-\rho})} & \text{if } i \in A_n \\
\end{cases}
\quad (O)
$$

▶ INNER LOOP: Monotone algorithm to solve the nonlinear part of (O) (used at each iteration of the active set strategy).

2: Reduce $\epsilon$ and repeat.
**Theorem (Uniqueness)**

Under a diagonal dominance condition (DD) and a strict complementarity condition (SC), there exists at most one solution \( x, \lambda \) to the necessary optimality condition.

**Theorem (Convergence)**

Under (DD), if \( \bar{x}, \bar{\lambda} \) is a solution to \( (O_e) \) satisfying (SC) then the sets

\[
S^n = \left\{ i \in I(\bar{x}, \bar{\lambda}) : \lambda_i^n = \frac{\beta p x_i^n}{\max(\varepsilon^{2-p}, |(\Lambda x^n)_i|^{2-p})} \right\}, \quad T^n = \{ i \in A(\bar{x}, \bar{\lambda}) : (\Lambda x^n)_i = 0 \}
\]

are monotonically nondecreasing. As soon as \( S^n = S^{n+1} \) and \( T^n = T^{n+1} \), then for some \( n \), we have \( (x^n, \lambda^n) = (\bar{x}, \bar{\lambda}) \).

**Remark**

Numerical experience-comparison with the Monotone:

1. same values for the \( \ell^0 \)-term for corresponding values of \( \beta \).

2. considerably smaller residue within a significantly fewer number of inner iterations.
**Theorem (Uniqueness)**

Under a diagonal dominance condition (DD) and a strict complementarity condition (SC), there exists at most one solution \( x, \lambda \) to the necessary optimality condition.

**Theorem (Convergence)**

Under (DD), if \( \bar{x}, \bar{\lambda} \) is a solution to \((O_e)\) satisfying (SC) then the sets

\[
S^n = \left\{ i \in I(\bar{x}, \bar{\lambda}) : \lambda_i^n = \frac{\beta p x_i^n}{\max(\varepsilon^2 - p, |(\Lambda x^n)_i|^2 - p)} \right\},
\]

\[
T^n = \left\{ i \in A(\bar{x}, \bar{\lambda}) : (\Lambda x^n)_i = 0 \right\}
\]

are monotonically nondecreasing. As soon as \( S^n = S^{n+1} \) and \( T^n = T^{n+1} \), then for some \( n \), we have \( (x^n, \lambda^n) = (\bar{x}, \bar{\lambda}) \).

**Remark**

Numerical experience-comparison with the Monotone:

1. same values for the \( \ell^0 \)-term for corresponding values of \( \beta \).

2. considerably smaller residue within a significantly fewer number of inner iterations.
**Microscopy Image Reconstruction**

*Single molecule detection-based techniques. STORM (stochastic optical reconstruction microscopy) method:*

- **Advantages:** sub-diffraction-limit spatial resolution
- **Disadvantages:** time, limited maximum density of emitters that can be accurately localized per frame.
- **Aim:** better resolution and higher emitter density
Compressed sensing methods based on $l^1$ techniques:

$$\min_{x \in \mathbb{R}^n} |x|^1 \quad \text{s.t. } |Ax - b|^2_2 \leq \varepsilon,$$

formulated as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} |Ax - b|^2_2 + \beta |x|^1.$$

see e.g. L. Zhu, W. Zhang, D. Elnatan, and B. Huang, H. P. Babcock, J. R. Moffitt, Y. Cao, X. Zhuang

We test our algorithm for $l^p$, $p < 1$ functionals:

$$\min_{x \in \mathbb{R}^n} |x|^p \quad \text{s.t. } |Ax - b|^2_2 \leq \varepsilon,$$

formulated as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} |Ax - b|^2_2 + \beta |x|^p.$$

**Remark**

Compressed sensing approach to microscopy image:

- $x \in \mathbb{R}^n$ is the high resolution reconstructed image
- $b \in \mathbb{R}^m, m < n$ is the lower resolution measured image
- solve an underdetermined system where the solution is sparse
**Numerical Results**

- Different resolution: $16 \times 16$ pixel conventional image vs $128 \times 128$ true image.
- Tested with 0-1 cross shape image and standard phantom image. Here we focus on the cross image.

PDASM $p = .1$
NUMERICAL RESULTS

- Different resolution: 16 × 16 pixel conventional image vs 128 × 128 true image.
- Tested with 0-1 cross shape image and standard phantom image. Here we focus on the cross image.

PDASM $p = .1$ compared to FISTA, $\beta = 10^{-6}$
Comparison with FISTA

Error (surplus emitters): better with PDASM
Concluding remarks

1. **Monotone scheme:**
   - efficient and competitive method to solve nonsmooth nonconvex problems.
   - convergence under no specific conditions.

2. **Primal-dual active set monotone scheme:**
   - efficient to find a solution of the unregularized problem.
   - less number of iterations and smaller residue w.r.t the monotone.
   - Convergence proved under a diagonal dominance condition.

Main references:

- *On monotone and primal-dual active set schemes for $\ell^p$-type problems,* $0 < p \leq 1$, submitted, D. Ghilli, K. Kunisch

- *A monotone scheme for sparsity optimization in $\ell^p$ spaces with $p \in (0, 1]$, IFAC Proceedings,* D. Ghilli, K, Kunisch

Some other references

- **I.R.L.S.**

- **P.D.A.S.**
▶ Miscellanea

Thank you for the attention
Remark: when $\Lambda$ is the unidirectional discretized gradient:

\[(c) \quad \beta = 0.01 \quad (d) \quad \beta = 0.1 \quad (e) \quad \beta = 0.3\]

**Figure:** Solution of the M-matrix problem, $p = .1$, mesh size $h = \frac{1}{64}$.

- Comparing the graphs, e.g., for $\beta = 0.3$: now we find subdomains where the solution is only unidirectionally piecewise constant (in the previous slide was piecewise constant).
- The number of iterations, $|\Lambda x|_0^c$, $|\Lambda x|_p^p$ and the residue are comparable.
Elliptic control problem

Consider

$$\inf \frac{1}{2} |y - y_d|^2 + \beta |\nabla u|^p, \quad p \in (0, 1],$$

where we minimize over $u \in L^p(\Omega)$ such that $\nabla u \in L^p(\Omega)$, $\Omega$ is the unit square, $y_d \in L^2(\Omega)$ is a given target function, and $y \in L^2(\Omega)$ satisfies

$$\begin{cases} 
-\Delta y = u & \text{in } \Omega \\
y = 0 & \text{in } \partial \Omega.
\end{cases}$$

We solve the following discretized problem:

$$\min_{u \in \mathbb{R}^n^2} \frac{1}{2} |Eu - b|^2 + \beta |\Lambda u|^p, \quad (0.1)$$

where $E = (A^*A)^{-1}$, $A$ is such that $A^*A$ is the 5-point star discretization of the Laplacian with Dirichlet boundary condition).

$\Lambda$ is the two-directional forward finite different discretization of the gradient.
Table: Sparsity in an elliptic control problem, $p = .1$, mesh size $h = \frac{1}{64}$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$10^{-3}$</th>
<th>$10^{-2}$</th>
<th>$10^{-1}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of iterates</td>
<td>102</td>
<td>119</td>
<td>5204</td>
<td>10440</td>
</tr>
<tr>
<td>$</td>
<td>\Lambda u</td>
<td>_0$</td>
<td>3170</td>
<td>2483</td>
</tr>
<tr>
<td>$</td>
<td>\Lambda u</td>
<td>_p$</td>
<td>$3.2 \times 10^4$</td>
<td>$2.6 \times 10^4$</td>
</tr>
<tr>
<td>Residue</td>
<td>$1.6 \times 10^{-5}$</td>
<td>$2.4 \times 10^{-4}$</td>
<td>$2 \times 10^{-3}$</td>
<td>$7 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Figure: Solution of the elliptic control problem, $p = .1$, mesh size $h = \frac{1}{64}$.

(A) $\beta = 0.01$  
(B) $\beta = 0.1$  
(C) $\beta = 1$
Proof (Necessary optimality conditions).

1. Note that if $\bar{x}$ is a global minimizer of $(mJ)$, then $\bar{y} = \Lambda \bar{x}$ is a global minimizer of

$$
\min_{y \in \mathbb{R}^n} \frac{1}{2} |\tilde{A} y - b|^2 + \beta |y|^p,
$$

(MJI)

where $\tilde{A} = A \Lambda^{-1}$.

2. We use similar arguments to Ito, Kunisch (2013) to get optimality condition for $(mJ)$. After introducing the multiplier $\lambda$, we get

$$
\begin{cases}
\tilde{A}^* (\tilde{A} y - b) + \lambda = 0 \\
\bar{y}_i = 0 \quad \text{if} \quad \left| \tilde{A}_i \bar{y}_i + \lambda_i \right| < \mu_i \\
|y_i| > 0 \quad \text{and} \quad \lambda_i = \frac{\beta p y_i}{|y_i|^{2-p}} \quad \text{if} \quad \left| \tilde{A}_i \bar{y}_i + \lambda_i \right| > \mu_i.
\end{cases}
$$

3. Then the optimality conditions follows with $\bar{y} = \Lambda \bar{x}$.
Algorithm 3 Primal-dual active set monotone scheme

1: Initialize $x^0, y^0, \lambda^0, \varepsilon$. Set $n = 0$.
2: repeat {OUTER LOOP}
3: \hspace{1em} Let $A_n = \{ i : |B_i y_i^n + \lambda_i^n| \leq \mu_i \}$, $\mathcal{I}_n = \mathcal{A}_n^c$.
4: \hspace{1em} Initialize $x^{0,n+1} = x^n$, $\lambda^{0,n+1} = \lambda^n$, $y^{0,n+1} = \Lambda x^{0,n}$. Set $k = 0$.
5: \hspace{1em} repeat {INNER LOOP}
6: \hspace{2em} Solve for $x^{k+1,n+1}, \lambda_{A_n}^{k+1,n+1}$
7: \hspace{2em} \hspace{1em} $\left\{ \begin{array}{l}
(A^* A + \Lambda_{\mathcal{I}_n}^* N_{\mathcal{I}_n}^{k,n+1} \Lambda_{\mathcal{I}_n} + \eta P) x^{k+1,n+1} + \Lambda_{A_n}^* \lambda_{A_n}^{k+1,n+1} = A^* b \\
\Lambda_{A_n} x^{k+1,n+1} = 0
\end{array} \right.$
8: \hspace{2em} Set $y^{k+1,n+1} = \Lambda x^{k+1,n+1}$, $\lambda_{\mathcal{I}_n}^{k+1,n+1} = \frac{\beta p y_{\mathcal{I}_n}^{k+1,n+1}}{\max(\varepsilon^{2-p}, |y_{\mathcal{I}_n}^{k+1,n+1}|^{2-p})}$.
9: \hspace{2em} If $y_{\mathcal{I}_n}^{k+1,n+1}$ is a singular point, go to 9.
10: \hspace{1em} Set $k = k + 1$.
11: until the stopping criteria for the inner loop is fulfilled.
12: Set $n = n + 1$;
13: until the stopping criteria for the outer loop is fulfilled.
14: Reduce $\varepsilon$ and go to 3.
Each single frame reconstruction is achieved by solving

$$\min_{x \in \mathbb{R}^n} |x|^p \quad \text{such that} \quad |Ax - b|^2_2 \leq \varepsilon,$$

$x$ reconstructed image, $b$ experimentally observed image, $A$ impulse response (PSF) (of size $m \times n$, where $m$ and $n$ are the numbers of pixels in $b$ and $x$, respectively).

**Optimal reconstructed image:**

- fewest number of fluorophores
- reproduces the measured image to a given accuracy (when convolved with the optical impulse response).

We reformulate the problem as:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}|Ax - b|^2_2 + \beta|x|^p.$$
A STORM reconstruction from a 16x16 pixel image-mon. active, $p = .1, \beta = 10^{-9}$
A STORM reconstruction from a 16x6 pixel image-fista, $p = .1, \beta = 10^{-9}$
A STORM reconstruction (same resolution)-mon. active, \( p = .1, \beta = 10^{-5} \)
A STORM reconstruction (same resolution)-mon. active, $p = .1, \beta = 10^{-5}$
A STORM reconstruction (same resolution)-fista, $p = 0.1$, $\beta = 10^{-5}$
A STORM reconstruction (same resolution), fista, $p = .1, \beta = 10^{-5}$