A homotopy method in regularization of total variation denoising problems

Luis A. Melara
Colorado College

Joint work with:

A. J. Kearsley (NIST), R.A. Tapia (Rice University)

MCSD, NIST, GAITHERSBURG, MARYLAND USA
Tuesday 30 August 2005

OUTLINE

- Motivation
- Previous Work
- Problem Formulation
- Computational Method
- Results
- Conclusions

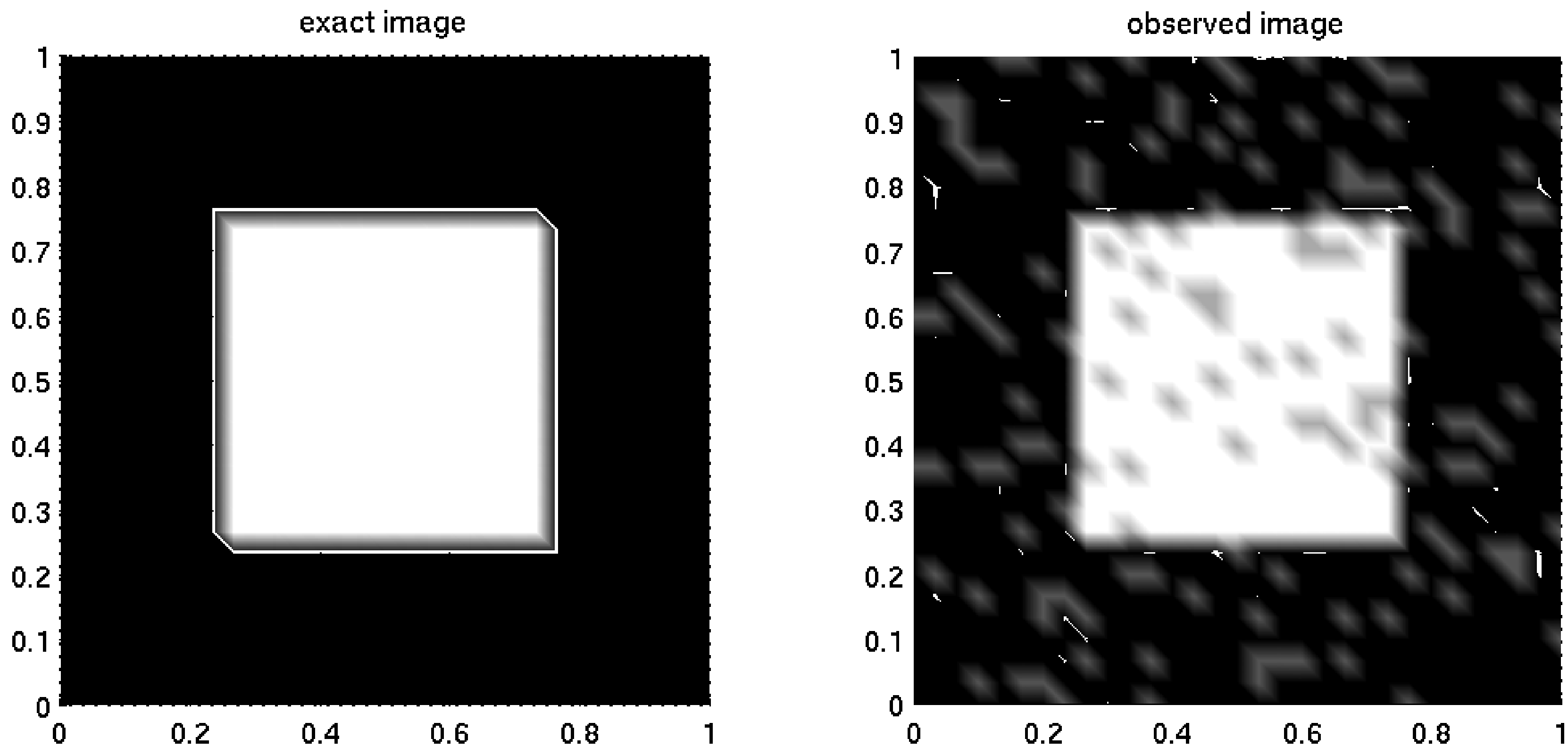

exact image
observed image
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Image Restoration

Three categories:

- Statistical methods.
- Transform-based methods.
- Optimization-based methods.

Optimization-based methods

- Images containing sharp edges.

- Image solves constrained optimization problem.

Bounded Variational Seminorm

Let $\mathbf{x} \in \Omega \subset \mathbb{R}^2$. Let $u : \Omega \to \mathbb{R}^2$.

$$\int_\Omega \|\nabla u\|_2 = \int_\Omega \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2} d\mathbf{x}$$

Contaminated image $\mathbf{u}_{obs}$

Observed image u_{obs}:

$$u_{obs}(\mathbf{x}) = u_{exact}(\mathbf{x}) + \eta(\mathbf{x}).$$

- $\eta \in H^1(\Omega)$ is noise.

- u_{exact} is exact image.

Define

$$\sigma = \left(\int_\Omega |u_{exact} - u_{obs}|^2 d\mathbf{x} \right)^{1/2} > 0.$$

Equality Constrained Optimization Problem (EC)

Find $u \in H^1(\Omega)$ s.t.

$$\begin{aligned}
\min_{u \in H^1(\Omega)} \quad & \int_\Omega \|\nabla u\|_2 \, d\mathbf{x} \\
\text{s.t.} \quad & 1/2 \left(\|u - u_{obs}\|^2_{L^2(\Omega)} - \sigma^2 \right) = 0.
\end{aligned}$$

Lagrangian Functional

To solve EC, we write Lagrangian functional:

$$\ell(u,\lambda) = \int_\Omega \|\nabla u\|_2 \, d\mathbf{x} + \lambda/2 \left(\|u - u_{obs}\|^2_{L^2(\Omega)} - \sigma^2 \right),$$

where $\lambda \in \mathbb{R}$ is Lagrange multiplier.

Euler-Lagrange Equations

Find $u \in H^1(\Omega)$ s.t.

$$\mathcal{M}(u, \lambda) = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|_2} \right) + \lambda (u - u_{obs}) = 0,$$

for $\mathbf{x} \in \Omega$ and

$$\frac{\partial u}{\partial n} = 0, \quad \mathbf{x} \in \Gamma,$$

Γ is boundary of Ω.

<u>Previous Work</u>

Fatemi, Osher and Rudin (1992).

$$\frac{\partial u}{\partial s} = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|_2} \right) + \lambda(u - u_{obs}),$$

- $s > 0$, $\mathbf{x} \in \Omega$,
- $u(x, y, 0)$ given and
- $\partial u / \partial n = 0$ for $\mathbf{x} \in \Gamma$.

At steady state solution:

$$\lambda = \frac{1}{2\sigma^2} \int_\Omega \|\nabla u\|_2 - \frac{\nabla u_{obs} \cdot \nabla u}{\|\nabla u\|_2} d\mathbf{x}.$$

Regularization

If $\nabla u = 0 \quad \Rightarrow \quad \mathcal{M}(u, \lambda)$ is undefined.

Regularized Lagrangian functional:

$$\begin{aligned} \ell_\epsilon(u, \lambda) &= \int_\Omega \sqrt{\|\nabla u\|_2^2 + \epsilon^2}\, d\mathbf{x} \\ &+ \lambda/2 \left(\|u - u_{obs}\|_{L^2(\Omega)}^2 - \sigma^2 \right). \end{aligned}$$

Associated Euler-Lagrange Equations

Find $u \in H^1(\Omega)$ s.t.

$$\mathcal{M}_\epsilon(u,\lambda) = -\nabla \cdot \left(\frac{\nabla u}{\sqrt{\|\nabla u\|_2^2 + \epsilon^2}} \right) + \lambda(u - u_{obs}) = 0,$$

for $\mathbf{x} \in \Omega$ with

$$\frac{\partial u}{\partial n} = 0, \quad \mathbf{x} \in \Gamma$$

For $\epsilon > 0$, $\mathcal{M}_\epsilon(u,\lambda)$ is well-defined.

Related Previous Work

Dobson, Omen and Vogel (1996).

Solve

$$\min_{u \in H^1(\Omega)} \frac{1}{2} \|u - u_{obs}\|^2_{L^2(\Omega)} + \gamma \int_\Omega \sqrt{\|\nabla u\|_2^2 + \epsilon^2},$$

where $\gamma \in \mathbb{R}$ is small penalization parameter (Majava, 2001).

Newton's method is used.

Hessian of Lagrangian

Hessian A of Lagrangian given by:

$$\langle Am, m \rangle =$$

$$\int_\Omega \left[-\nabla \cdot \left(\frac{\nabla m}{\sqrt{\|\nabla u\|_2^2 + \epsilon^2}} \right) + \nabla \cdot \left(\frac{(\nabla u \cdot \nabla m) \nabla u}{\left(\sqrt{\|\nabla u\|_2^2 + \epsilon^2} \right)^3} \right) \right] m$$

$$+ \lambda\, m^2 \, d\mathbf{x},$$

for $m \in H^1(\Omega)$.

Denote $J_\epsilon(u; m, m) = \langle Am, m \rangle$.

Idea

Newton's method $\Rightarrow$ Kantorovich Theorem.

Show direct relationship between regularizaton parameter ϵ and radius of Kantorovich ball.

Proposition 1

Given $m, p \in H^1(\Omega)$, then the following inequality holds,

$$|||J_\epsilon(m;\,\cdot\,,\,\cdot\,) - J_\epsilon(p;\,\cdot\,,\,\cdot\,)||| \leq c_1(\epsilon)||m-p||_{H^1(\Omega)}, \qquad (1)$$

with

$$c_1(\epsilon) = 3N(Q)\epsilon^2/h \cdot \max_{Q_k \subset \Omega} \left\{ \frac{1}{D_1^2}, \frac{1}{D_1 \cdot D_2}, \frac{1}{D_2^2} \right\},$$

where

$$D_1 = \|\nabla m\|_2^2 + \epsilon^2,$$

and

$$D_2 = \|\nabla p\|_2^2 + \epsilon^2.$$

Proposition 2

Let $u \in H^1(\Omega)$. If the Jacobian $J_\epsilon(u; \cdot, \cdot)$ is symmetric positive definite, then it has a bounded inverse with bound,

$$|||J_\epsilon(u; \cdot, \cdot)^{-1}||| \leq c_2(\epsilon) = \left(\min_{Q_k \subset \Omega} \left\{ \frac{\epsilon^2}{(D_3)^3}, \lambda \right\} \right)^{-1},$$

where

$$D_3 = \sqrt{\|\nabla u\|_2^2 + \epsilon^2}.$$

Definition

Let

$$B_p(u^{(0)}, r) = \{u : \|u - u^{(0)}\|_p < r\},$$

be $p-$ball of radius r centered at $u^{(0)}$.

Theorem(Kantorovich)

Consider a function, $L_\epsilon : \mathbb{R}^n \to \mathbb{R}^n$ that is defined on a convex set $C \subseteq \mathbb{R}^n$. Let the Jacobian operator of L_ϵ be J_ϵ and further assume that J_ϵ is a Lipschitz function with Lipschitz constant α_L. Assume that $u^{(0)}$ is some starting point selected from C and that the following are all satisfied for some $u^{(0)} \in C$,

1. $|||J_\epsilon(u;\ \cdot,\ \cdot) - J_\epsilon(v;\ \cdot,\ \cdot)||| \le \alpha_L \|u - v\|,\ \forall u, v \in C$,
2. $|||J_\epsilon(u^{(0)};\ \cdot,\ \cdot)^{-1}||| \le \alpha_1$,
3. $|||J_\epsilon(u^{(0)};\ \cdot,\ \cdot)^{-1} L_\epsilon(u^{(0)};\ \cdot)||| \le \alpha_2$.

If $\delta = \alpha_L \alpha_1 \alpha_2 \le 1/2$, and if $B_p(u^{(0)}, r) \subseteq C$ with

$$r = \alpha_2(1 - \sqrt{1 - 2\delta})/\delta,$$

then the Newton sequence $\{u^{(n)}\}$ given by

$$u^{(n+1)} = u^{(n)} - J_\epsilon(u^{(n)};\ \cdot,\ \cdot)^{-1} L_\epsilon(u^{(n)};\ \cdot),$$

is well-defined, remains in the ball $B_p(u^{(0)}, r)$, and converges to the unique solution of $L_\epsilon(u^*; \cdot) = 0$ inside $B_p(u^{(0)}, r)$.

Summary

Large ϵ $\Rightarrow$ nice convergence but undesirable solution u^*.

Small ϵ $\Rightarrow$ bad convergence but desirable solution u^*.

From Proposition 1, Proposition 2 and Kantorovich Theorem:

$$\text{Large } \epsilon_0 \;\Rightarrow\; u^{(n)}(u^{(0)}) \longrightarrow u_0^*.$$

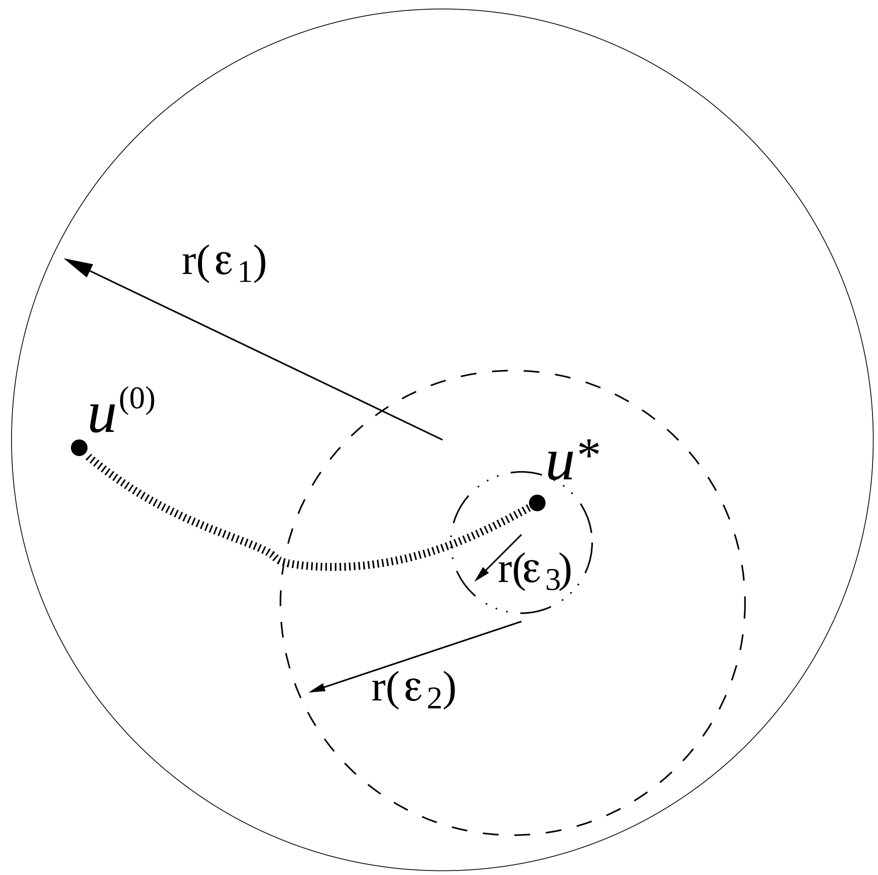
$$\text{Reduce } \epsilon_0 \;\Rightarrow\; u^{(n)}(u_0^*) \longrightarrow u_1^*.$$

$$\text{Reduce } \epsilon_0 \;\Rightarrow\; u^{(n)}(u_1^*) \longrightarrow u_2^*.$$

$$\vdots$$

$$\vdots$$

Homotopy



From Propositions 1 and 2: α_L and α_1 depend on ϵ

$\Rightarrow$ radius r of $B_p(u, r)$ depends on ϵ.

Numerical Implementation

Let

$$\widehat{Q} = (0,1) \times (0,1), \quad \Omega = \bigcup_{k=1}^{N(Q)} Q_k,$$

$N(Q) =$ number of elements in Ω.

Define affine map F:

$$F(\hat{\mathbf{x}}) = \mathbf{D}\hat{\mathbf{x}} + \mathbf{d} = \mathbf{x},$$

$\hat{\mathbf{x}} \in \widehat{Q}$ and $\mathbf{x} \in Q_k$, with

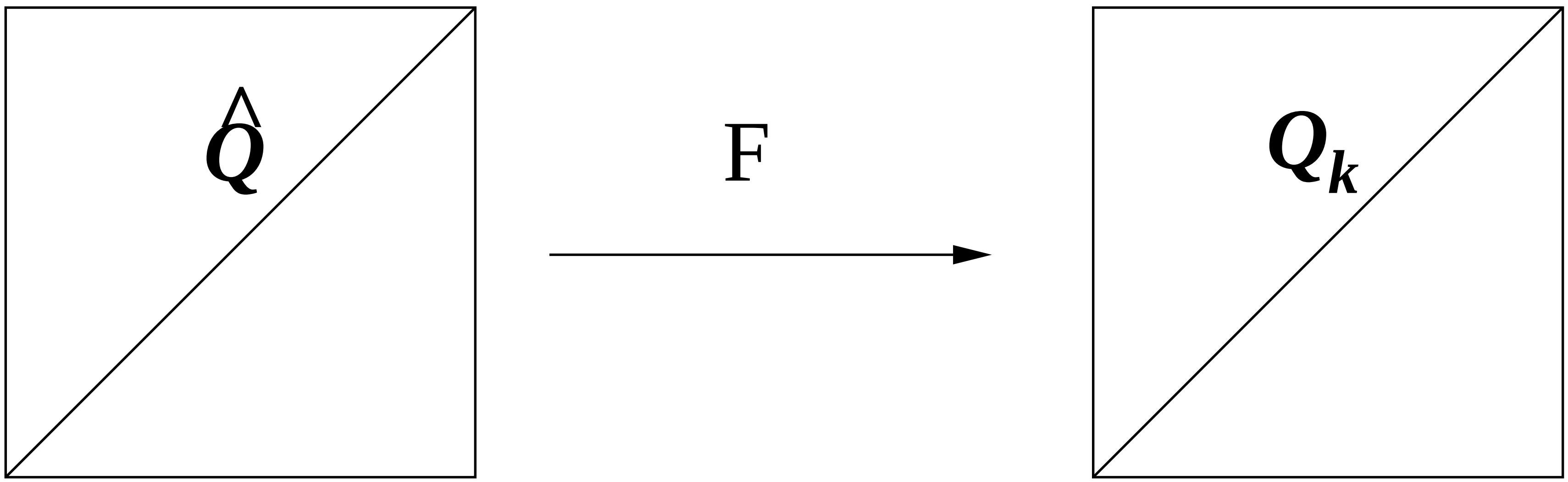
$$\mathbf{D} = \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix}, \ \mathbf{d} = \begin{pmatrix} ih \\ jh \end{pmatrix},$$

where $i, j = 1, \ldots, N$.

$N =$ number of partitions along $x-$ and $y-$axes.

$h = 1/N$ is mesh size along $x-$ and $y-$axes.

Visual Description



where

$$\hat{Q} = (0,1) \times (0,1), \quad \text{and}$$

$$Q_k = (ih, jh) \times ((i+1)h, (j+1)h) \subset \Omega$$

Function composition

Use affine map F,

$$u(\mathbf{x}) = \left(\widehat{u} \circ F^{-1}\right)(\mathbf{x}) = \widehat{u}(F^{-1}(\mathbf{x})).$$

By chain rule:

$$\nabla u(\mathbf{x}) = \nabla F^{-1} \cdot \widehat{\nabla} \widehat{u} \left(F^{-1}(\mathbf{x})\right),$$

with

$$\nabla = (\partial/\partial x, \partial/\partial y), \quad \text{and} \quad \widehat{\nabla} = (\partial/\partial \widehat{x}, \partial/\partial \widehat{y}).$$

Function Approximation

Approximation of u:

$$u = \sum_{k=1}^{N(h)} \mathbf{u}_k \varphi_k,$$

where $\mathbf{u}_k = u(ih, jh)$, for some i, j,

$N(h) =$ number of nodes in Ω.

The kth basis function $\varphi_k \in \mathcal{P}_1$ where

$$\mathcal{P}_1 = \text{ span}\{1, x, y\}$$

Objective Function

$$\int_{\Omega} \sqrt{\|\nabla u\|_2^2 + \epsilon^2}\, d\mathbf{x} \;=\; \sum_{k=1}^{N(Q)} \int_{Q_k} \sqrt{\|\nabla u\|_2^2 + \epsilon^2}\, d\mathbf{x}.$$

Use F to obtain in each Q_k:

$$\int_{Q_k} \sqrt{\|\nabla u\|_2^2 + \epsilon^2}\, d\mathbf{x} = \int_{\widehat{Q}} \sqrt{\|\nabla F^{-1}\, \widehat{\nabla} \hat{u}\|_2^2 + \epsilon^2}\, \text{det}\, \widehat{\nabla} F d\hat{\mathbf{x}}.$$

Equality Constraint

Similarly,

$$\frac{1}{2}\left(\int_\Omega |u-u_{obs}|^2-\sigma^2\right)=\frac{1}{2}\left(\sum_{k=1}^{N(Q)}\int_{Q_k}|u-u_{obs}|^2-\sigma^2\right).$$

Again, use mapping F.

Idea:

$$\int_{Q_k}|u|^2 d\mathbf{x}=\int_{\hat{Q}}|\hat{u}|^2 \det \hat{\nabla} F\, d\hat{\mathbf{x}}.$$

Optimization Method

Let $u \in H^1(\Omega)$.

Let

$$f(u) = \int_\Omega \sqrt{\|\nabla u\|_2^2 + \epsilon^2}\, d\mathbf{x},$$

and

$$g(u) = 1/2 \left(\int_\Omega |u - u_{obs}|^2 - \sigma^2 \right).$$

Augmented Lagrangian $\mathcal{L}(u, \lambda, \rho)$:

$$\mathcal{L}(u, \lambda, \rho) = f(u) + \lambda\, g(u) + (\rho/2)\, g(u)^2,$$

where $\lambda \in \mathbb{R}$ and $\rho \in \mathbb{R}$.

Notation: Discrete Approximations

- Denote $f(u)$ by $f(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{N(h)}$,

- Denote $g(u)$ by $g(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{N(h)}$,

- Simplify $\mathcal{L}(u, \lambda, \rho)$ and denote:

$$\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{u}, \lambda, \rho),\ \mathbf{u} \in \mathbb{R}^{N(h)},$$

$N(h) =$ number of nodes in Ω.

Gradient of $\mathcal{L}(\mathbf{u})$

Find $\mathbf{u} \in \mathbb{R}^{N(h)}$ s.t.

$$\nabla\mathcal{L}(\mathbf{u}) = 0, \qquad \nabla\mathcal{L}(\mathbf{u}) \in \mathbb{R}^{N(h)},$$

with

$$(\nabla\mathcal{L}(\mathbf{u}))_i = (\nabla f(\mathbf{u}))_i + \lambda\,(\nabla g(\mathbf{u}))_i + 2\rho\, g(\mathbf{u})\ (\nabla g(\mathbf{u}))_i$$

$$= \partial f(\mathbf{u})/\partial \mathbf{u}_i + \lambda \partial g(\mathbf{u})/\partial \mathbf{u}_i + 2\rho\, g(\mathbf{u})\, \partial g(\mathbf{u})/\partial \mathbf{u}_i.$$

for $i = 1, \dots, N(h)$

Hessian of $\mathcal{L}(\mathbf{u})$

The ijth component of Hessian $\mathbf{A} \in \mathbb{R}^{N(h)\times N(h)}$ of $\mathcal{L}(\mathbf{u})$ is:

$$\mathbf{A}_{ij} = \frac{\partial^2 \mathcal{L}(\mathbf{u})}{\partial \mathbf{u}_i \partial \mathbf{u}_j},$$

where $i, j = 1, \ldots, N(h)$.

Newton's Method

Given $\mathbf{u}^{(0)}$, Newton sequence $\{\mathbf{u}^{(n)}\}$ is generated by:

$$\mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \alpha\, \mathbf{s}^{(n)},$$

where $\alpha \in \mathbb{R}$ and $\mathbf{s}^{(n)} \in \mathbb{R}^{N(h)}$ solves

$$\mathbf{A}^{(n)}\, \mathbf{s}^{(n)} = -\nabla \mathcal{L}(\mathbf{u}^{(n)}).$$

Sufficient Decrease Criteria

Set $\alpha := 1$.

If inequality,

$$\mathcal{L}(\mathbf{u}^{(n)} + \alpha \mathbf{s}^{(n)}) < \mathcal{L}(\mathbf{u}^{(n)}) + 10^{-4}\, \alpha\, \nabla\mathcal{L}(\mathbf{u}^{(n)}) \cdot \mathbf{s}^{(n)},$$

not met, then

$$\alpha := \alpha/2.$$

Otherwise, update:

$$\mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \alpha \mathbf{s}^{(n)}, \quad \text{and}$$

reset $\alpha := 1$.

Lagrange multiplier update

Least Squares update formula:

$$\lambda = -\frac{\nabla f(\mathbf{u}) \cdot \nabla g(\mathbf{u})}{\nabla g(\mathbf{u}) \cdot \nabla g(\mathbf{u})}.$$

Diagonalized multiplier method, convergence properties follow from Tapia (1977).

Algorithm

Initialize λ, ρ.

do i=1, ITERS1

    do j=1, ITERS2

        Solve $\nabla \mathcal{L}(\mathbf{u}) = 0$ (Newton's method)

        $\lambda_+ :=$ Least Squares Update

    end do

    $\epsilon_+ = 10^{-i}$

    Reinitialize λ_+

end do

Numerical Examples

Let $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$.

True image $\mathbf{u}_{exact}$.

Component $(\mathbf{u}_{obs})_i$ given by:

$$(\mathbf{u}_{obs})_i = (\mathbf{u}_{exact})_i + \text{ rand}(1) * C,$$

where $C \sim 10^{-1}$.

`rand.m` produces uniformly distributed random numbers on $(0.00, 1.00)$, (Matlab built-in function).

Numerical Example 1

- Initial iterate $\mathbf{u}^{(0)} = \mathbf{u}_{obs}$:

$$(\mathbf{u}_{obs})_i = (\mathbf{u}_{exact})_i + \text{ rand}(1) * 4.00e - 1,$$

- $N = 30$ partitions along $x-$ and $y-$axes.
- $N(Q) = 30^2 = 900$ elements in Ω.
- $N(h) = 31^2 = 961$ nodes in Ω.
- $\rho = 300$, fixed.
- Initial $\lambda_0 = 10$.

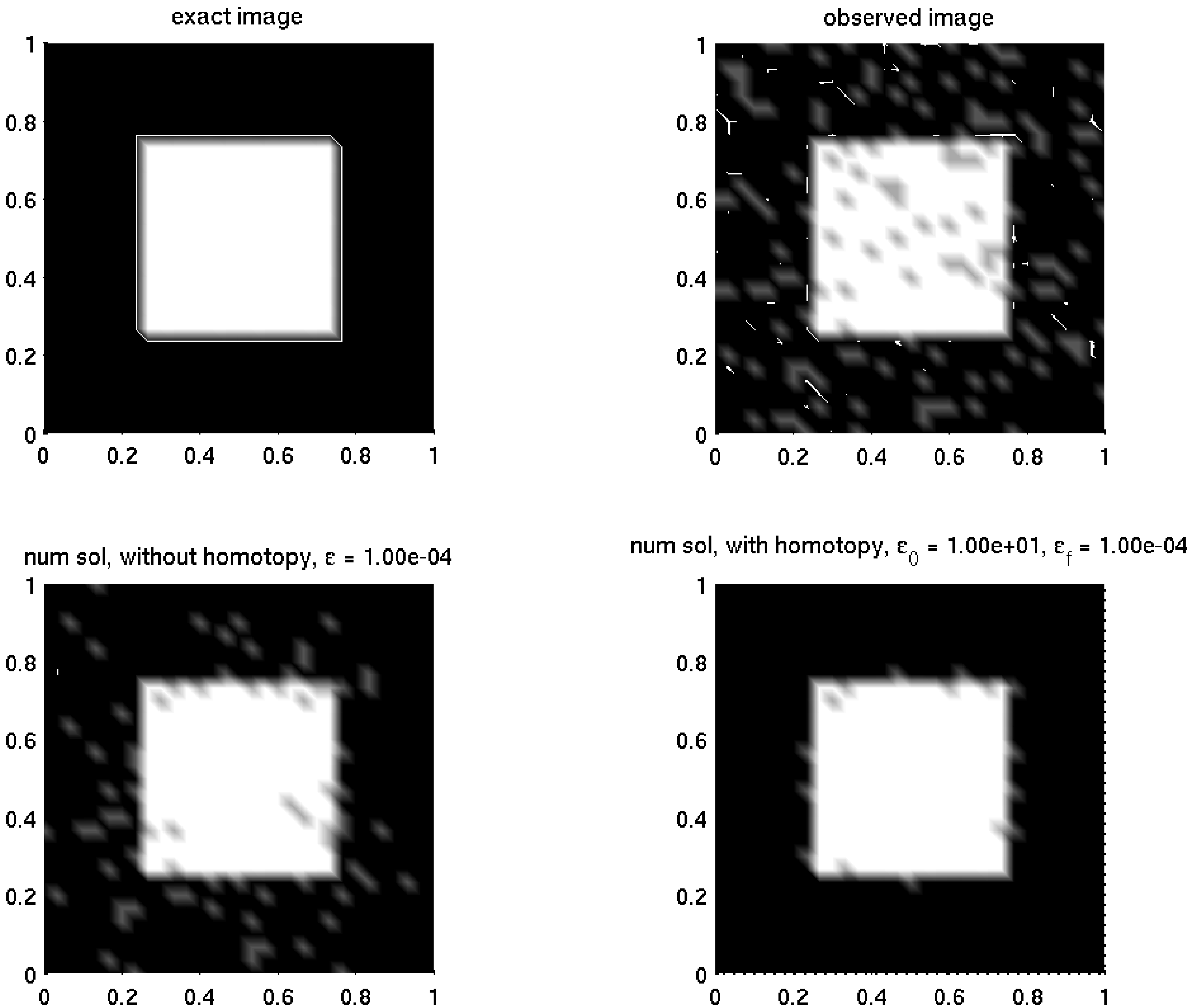
Two cases:

- No homotopy: $\epsilon = 1.00d - 04$.

- Homotopy:

Initial $\epsilon_0 = 1.00d + 01 \Rightarrow \epsilon_f = 1.00d - 04$.

Numerical Results


exact image
observed image
num sol, without homotopy, ε = 1.00e-04
num sol, with homotopy, ε0 = 1.00e+01, εf = 1.00e-04
0
0.2
0.4
0.6
0.8
1

Numerical Example 2

- Initial iterate $\mathbf{u}^{(0)} = \mathbf{u}_{obs}$:

$$(\mathbf{u}_{obs})_i = (\mathbf{u}_{exact})_i + \text{ rand}(1) * 3.80e - 1,$$

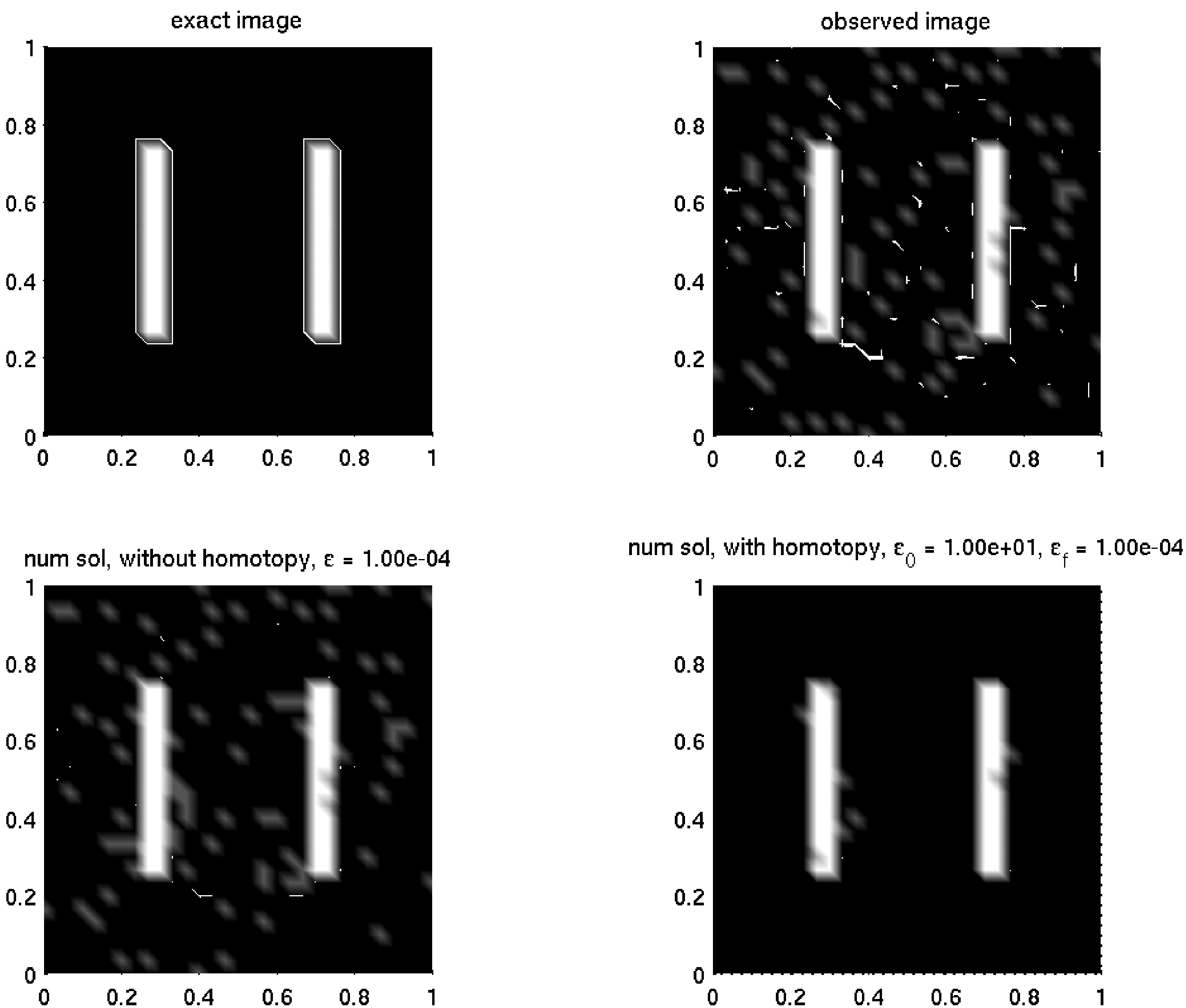
- $N = 30$ partitions along $x-$ and $y-$axes.
- $N(Q) = 30^2 = 900$ elements in Ω.
- $N(h) = 31^2 = 961$ nodes in Ω.
- $\rho = 300$, fixed.
- Initial $\lambda_0 = 10$.

Two cases:

• No homotopy: $\epsilon = 1.00d - 04$.

• Homotopy:
Initial $\epsilon_0 = 1.00d + 01 \Rightarrow \epsilon_f = 1.00d - 04$.

Numerical Results


exact image
observed image
num sol, without homotopy, ε = 1.00e-04
num sol, with homotopy, ε0 = 1.00e+01, εf = 1.00e-04
0
0.2
0.4
0.6
0.8
1

Numerical Example 3

- Initial iterate $\mathbf{u}^{(0)} = \mathbf{u}_{obs}$:

$$(\mathbf{u}_{obs})_i = (\mathbf{u}_{exact})_i + \text{ rand}(1) * 3.80e - 1,$$

- $N = 30$ partitions along $x-$ and $y-$axes.

- $N(Q) = 30^2 = 900$ elements in Ω.

- $N(h) = 31^2 = 961$ nodes in Ω.

- $\rho = 300$, fixed.

- Initial $\lambda_0 = 10$.

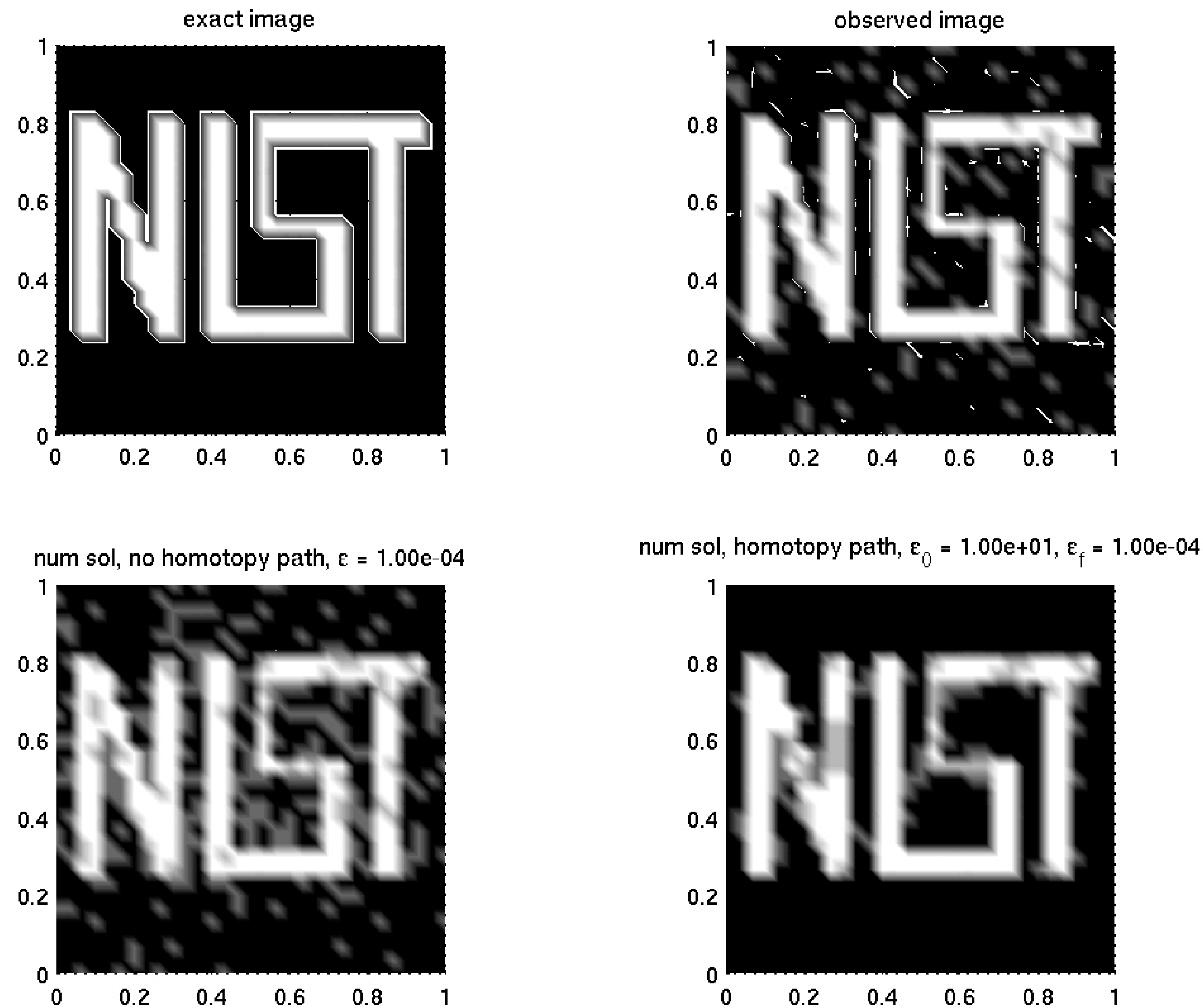
Two cases:

- No homotopy: $\epsilon = 1.00d - 04$.

- Homotopy:

Initial $\epsilon_0 = 1.00d + 01 \Rightarrow \epsilon_f = 1.00d - 04$.

Numerical Results



exact image
observed image
num sol, no homotopy path, ε = 1.00e-04
num sol, homotopy path, ε0 = 1.00e+01, εf = 1.00e-04

Conclusion

- Developed numerical technique and strong theoretical justification for total variational image denoising problems.

- Articulated effect of regularization parameter on Kantorovich estimates.

- To appear in JOTA.