Inverse problems
Total Variation Regularization

Mark van Kraaij

Casa seminar 23 May 2007

Introduction

Fredholm first kind integral equation of convolution type in one space dimension:

$$\begin{aligned} g(x) &= \int_0^1 k(x-x')f(x')dx' = (\mathcal{K}f)(x), \qquad 0 < x < 1, \\ k(x) &= C\,\mathrm{exp}(-x^2/2\gamma^2). \qquad \text{(Gaussian kernel)} \end{aligned}$$

Introduction

Fredholm first kind integral equation of convolution type in one space dimension:

$$\begin{aligned} g(x) &= \int_0^1 k(x-x')f(x')dx' = (\mathcal{K}f)(x), \qquad 0 < x < 1, \\ k(x) &= C\,\mathrm{exp}(-x^2/2\gamma^2). \qquad \text{(Gaussian kernel)} \end{aligned}$$

Direct problem:

Given the source f and the kernel k, determine the blurred image g.

For a piecewise smooth source and smooth kernel, an accurate approximation of $g = \mathcal{K}f$ is found using standard numerical quadrature.

Introduction

Fredholm first kind integral equation of convolution type in one space dimension:

$$\begin{aligned} g(x) &= \int_0^1 k(x-x')f(x')dx' = (\mathcal{K}f)(x), \qquad 0 < x < 1, \\ k(x) &= C\exp(-x^2/2\gamma^2). \qquad \text{(Gaussian kernel)} \end{aligned}$$

Inverse problem:

Given the blurred image g and the kernel k, determine the source f.

Discretize equation using collocation in the independent variable x and quadrature in x' to obtain a discrete linear system $\mathbf{Kf} = \mathbf{d}$.

Introduction

Composite midpoint quadrature with $h = 1/n$:

$$\mathbf{f} \; = \; \mathbf{K}^{-1}\mathbf{d}, \quad \text{where } \mathbf{K}_{ij} = h\,C\exp\Big(-\frac{((i-j)h)^2}{2\gamma^2}\Big).$$

To obtain an accurate quadrature approximation, n must be relatively large. Unfortunately, matrix $\mathbf{K}$ becomes increasingly ill-conditioned for large n.

Introduction

Composite midpoint quadrature with $h = 1/n$:

$$\mathbf{f} \ = \ \mathbf{K}^{-1}\mathbf{d}, \quad \text{where } \mathbf{K}_{ij} \ = h\,C\exp\Big(-\frac{((i-j)h)^2}{2\gamma^2}\Big).$$

To obtain an accurate quadrature approximation, n must be relatively large. Unfortunately, matrix $\mathbf{K}$ becomes increasingly ill-conditioned for large n.

Errors due to quadrature can be controlled, errors in $\mathbf{d}$ may be amplified!

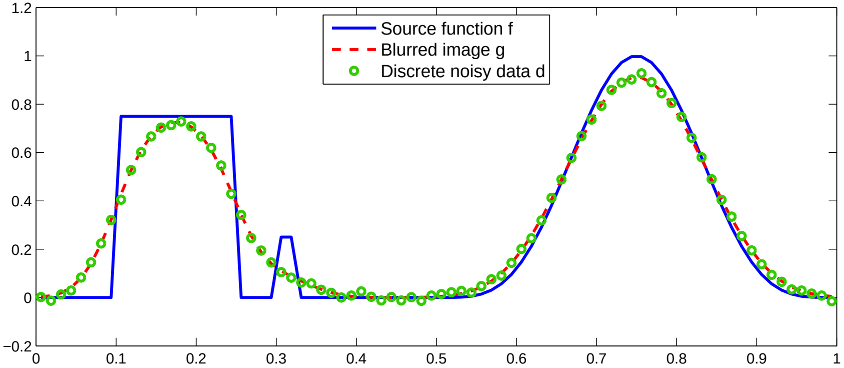


Figure 1: $\gamma = 0.05, C = \frac{1}{\gamma\sqrt{2\pi}}, n = 80$

Regularization by filtering

Assume a discrete data model for the discrete linear system $\mathbf{K}\mathbf{f} = \mathbf{d}$, i.e.

$$\mathbf{d} \;=\; \mathbf{K}\mathbf{f}_{\text{true}} + \boldsymbol{\eta}, \qquad \text{with } \; \delta := ||\boldsymbol{\eta}|| > 0 \; \text{ the error level.}$$

Assuming $\mathbf{K}$ is invertible the SVD ($\mathbf{K} = \mathbf{U}\,\mathrm{diag}(s_i)\mathbf{V}^T$) gives

$$\mathbf{K}^{-1}\mathbf{d} \;=\; \mathbf{f}_{\text{true}} + \sum_{i=1}^{n} s_i^{-1}(\mathbf{u}_i^T\boldsymbol{\eta})\mathbf{v}_i.$$

Regularization by filtering

Assume a discrete data model for the discrete linear system $\mathbf{K}\mathbf{f} = \mathbf{d}$, i.e.

$$\mathbf{d} \;=\; \mathbf{K}\mathbf{f}_{\mathrm{true}} + \boldsymbol{\eta}, \qquad \text{with } \; \delta := ||\boldsymbol{\eta}|| > 0 \; \text{ the error level.}$$

Assuming $\mathbf{K}$ is invertible the SVD ($\mathbf{K} = \mathbf{U}\,\mathrm{diag}(s_i)\mathbf{V}^T$) gives

$$\mathbf{K}^{-1}\mathbf{d} \;=\; \mathbf{f}_{\mathrm{true}} + \sum_{i=1}^{n} s_i^{-1}(\mathbf{u}_i^T\boldsymbol{\eta})\mathbf{v}_i.$$

Instabilities arise due to division by small singular values. Use regularizing filter function $w_\alpha(s_i^2)$ for which the product $w_\alpha(s^2)s^{-1} \longrightarrow 0$ as $s \longrightarrow 0$.

Approximate solution

$$\mathbf{f}_\alpha \;=\; \sum_{i=1}^{n} w_\alpha(s_i^2)s_i^{-1}(\mathbf{u}_i^T\mathbf{d})\mathbf{v}_i.$$

Regularization by filtering

TSVD	**Tikhonov**
$w_\alpha(s^2) = \begin{cases} 1 & \text{if } s^2 > \alpha, \\ 0 & \text{if } s^2 \leq \alpha. \end{cases}$	$w_\alpha(s^2) = \dfrac{s^2}{s^2 + \alpha}$
$\mathbf{f}_\alpha = \sum_{s_i^2 > \alpha} s_i^{-1} (\mathbf{u}_i^T \mathbf{d}) \mathbf{v}_i$	$\mathbf{f}_\alpha = (\mathbf{K}^T\mathbf{K} + \alpha \mathbf{I})^{-1} \mathbf{K}^T \mathbf{d}$

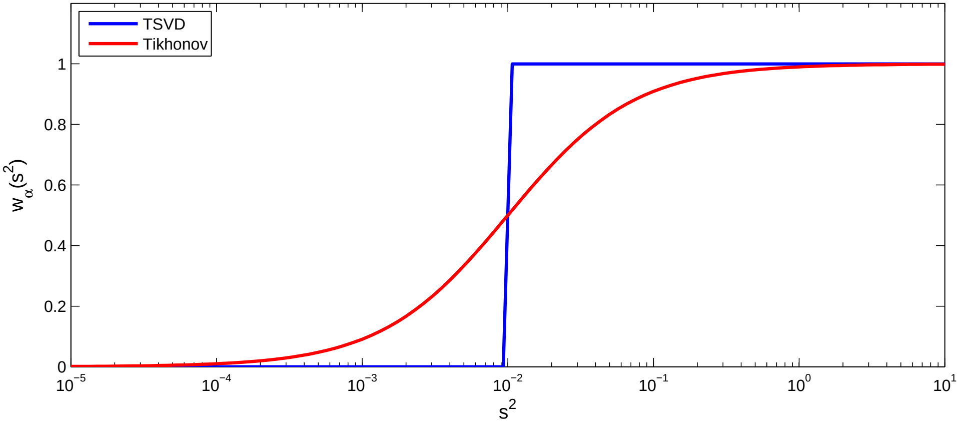
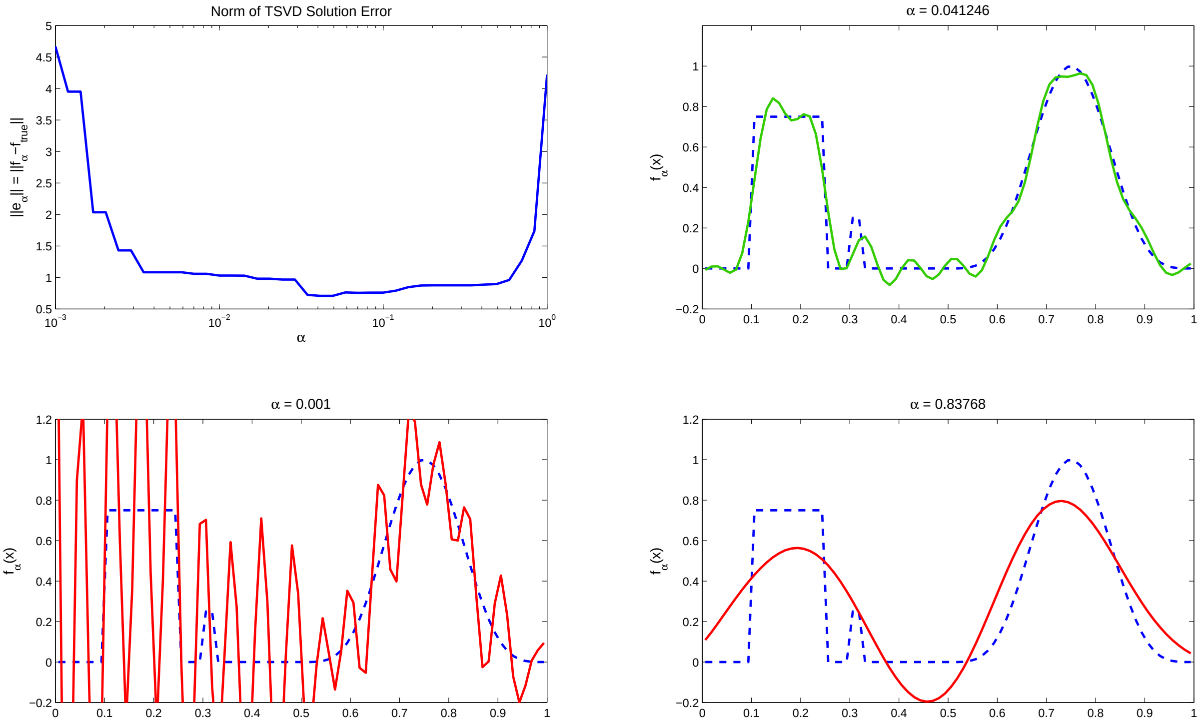


Figure 2: Regularization parameter $\alpha = 10^{-2}$

Regularization by filtering


Norm of TSVD Solution Error

α = 0.041246

α = 0.001

α = 0.83768

Figure 3: TSVD regularized solutions with 2% error level

Regularization by filtering

Choose regularization parameter α such that

$$\mathbf{e}_\alpha = \mathbf{e}_\alpha^{\text{trunc}} + \mathbf{e}_\alpha^{\text{noise}} \rightarrow 0 \qquad \text{for } \delta \rightarrow 0.$$

For both TSVD and Tikhonov filter

$$\begin{aligned} ||\mathbf{e}_\alpha^{\text{noise}}|| \leq \alpha^{-\frac{1}{2}}\delta &\rightarrow 0, \qquad \text{for } \alpha = \delta^p \text{ and } p < 2, \\ ||\mathbf{e}_\alpha^{\text{trunc}}|| &\rightarrow 0, \qquad \text{for } \alpha \rightarrow 0 \text{ thus } p > 0. \end{aligned}$$

Regularization by filtering

Choose regularization parameter α such that

$$\mathbf{e}_\alpha \;=\; \mathbf{e}_\alpha^{\text{trunc}} + \mathbf{e}_\alpha^{\text{noise}} \to 0 \qquad \text{for } \delta \to 0.$$

For both TSVD and Tikhonov filter

$$\begin{aligned} ||\mathbf{e}_\alpha^{\text{noise}}|| \le \alpha^{-\frac{1}{2}}\delta \;&\to\; 0, \qquad \text{for } \alpha = \delta^p \text{ and } p < 2, \\ ||\mathbf{e}_\alpha^{\text{trunc}}|| \;&\to\; 0, \qquad \text{for } \alpha \to 0 \text{ thus } p > 0. \end{aligned}$$

For TSVD assume $\delta \ge s_{\min}$ and $\mathbf{f}_{\text{true}} = \mathbf{K}^T\mathbf{z}$ for $\mathbf{z} \in \mathbb{R}^n$ then

$$||\mathbf{e}_\alpha|| \;\le\; \alpha^{\frac{1}{2}}||\mathbf{z}|| + \alpha^{-\frac{1}{2}}\delta.$$

Minimizing w.r.t. α gives

$$\alpha \;=\; \delta/||\mathbf{z}|| \qquad \text{and} \qquad ||\mathbf{e}_\alpha|| \;\le\; 2||\mathbf{z}||^{\frac{1}{2}}\delta^{\frac{1}{2}} \;=\; \mathcal{O}(\sqrt{\delta}).$$

Variational regularization methods

For very large ill-conditioned systems regularization by filtering is impractical since it requires the SVD of a large matrix.

Alternative, Tikhonov variational representation

$$\mathbf{f}_\alpha = \arg\min_{\mathbf{f}\in\mathbb{R}^n} ||\mathbf{K}\mathbf{f} - \mathbf{d}||^2 + \alpha||\mathbf{f}||^2,$$

- might be easier to compute
- incorporate constraints, e.g. nonnegativity of f
- replace least squares term $||\mathbf{K}\mathbf{f} - \mathbf{d}||^2$ by other fit-to-data functionals
- replace penalty term $||\mathbf{f}||^2$ by other functionals incorporating a priori information → **Total Variation Regularization**

Variational regularization methods

For very large ill-conditioned systems regularization by filtering is impractical since it requires the SVD of a large matrix.

Alternative, Tikhonov variational representation

$$\mathbf{f}_\alpha = \arg\min_{\mathbf{f}\in\mathbb{R}^n} ||\mathbf{Kf} - \mathbf{d}||^2 + \alpha \mathrm{TV}(\mathbf{f}),$$

with the discrete one-dimensional total variation function

$$\begin{aligned} \mathrm{TV}(\mathbf{f}) &= \sum_{i=1}^{n-1} |f_{i+1} - f_i| \\ &= \sum_{i=1}^{n-1} \left|\frac{f_{i+1} - f_i}{\Delta x}\right| \Delta x. \end{aligned}$$

This penalizes highly oscillatory solutions while allowing jumps in the regularized solution.

Variational regularization methods

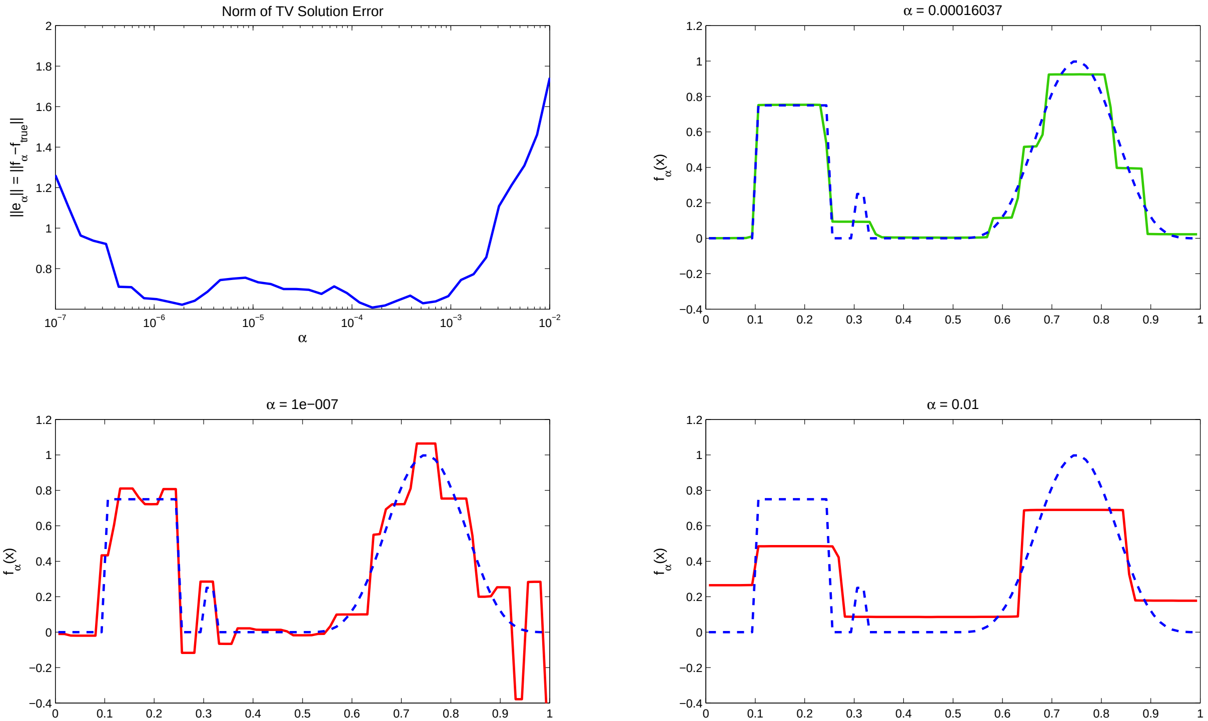


Figure 4: TV regularized solutions with 2% error level

Definition total variation

Definition. The total variation of a function $f \in L^1(\Omega)$ is defined by

$$\mathrm{TV}(f) \;=\; \sup_{\mathbf{v} \in \mathcal{V}} \int_\Omega f \,\mathrm{div}\, \mathbf{v} \, dx,$$

where

- Ω denotes a simply connected, nonempty, open subset of $\mathbb{R}^d$, $d = 1, 2, \ldots$, with Lipschitz continuous boundary.
- the space of test functions

$$\mathcal{V} \;=\; \{\mathbf{v} \in C_0^1(\Omega; \mathbb{R}^d) \,|\, |\mathbf{v}(x)| \leq 1 \text{ for all } x \in \Omega\}.$$

- $C_0^1(\Omega; \mathbb{R}^d)$ denotes the space of vector valued functions $\mathbf{v} = (v_1, \ldots, v_d)$ whose component functions v_i are each continuously differentiable and compactly supported on Ω, i.e., each v_i vanishes outside some compact subset of Ω.

Definition total variation

Example. Let $\Omega = [0,1] \subset \mathbb{R}$, and define

$$f(x) = \begin{cases} f_0, & x < \frac{1}{2}, \\ f_1 & x > \frac{1}{2}, \end{cases}$$

where f_0, f_1 are constants. For any $v \in C_0^1[0,1]$,

$$\begin{aligned} \int_0^1 f(x)v'(x)\,dx &= \int_0^{1/2} f(x)v'(x)\,dx + \int_{1/2}^1 f(x)v'(x)\,dx \\ &= (f_0 - f_1)v(1/2). \end{aligned}$$

This quantity is maximized over all $v \in \mathcal{V}$ when $v(1/2) = \mathrm{sign}(f_0 - f_1)$, thus $\mathrm{TV}(f) = |f_1 - f_0|$.

Definition total variation

Definition. The total variation of a function f defined on $[0, 1]$ is defined by

$$\mathrm{TV}(f) = \sup \sum_i |f(x_i) - f(x_{i-1})|,$$

where the supremum is taken over all partitions $0 = x_0 < \cdots < x_n = 1$.

Proposition. If f is smooth ($f \in W^{1,1}(\Omega)$) then

$$\mathrm{TV}(f) \; = \; \int_\Omega |\nabla f| \, dx.$$

Definition total variation

Definition. The total variation of a function f defined on $[0,1]$ is defined by

$$\mathrm{TV}(f) = \sup \sum_i |f(x_i) - f(x_{i-1})|,$$

where the supremum is taken over all partitions $0 = x_0 < \cdots < x_n = 1$.

Proposition. If f is smooth ($f \in W^{1,1}(\Omega)$) then

$$\mathrm{TV}(f) \ = \ \int_\Omega |\nabla f| \, dx.$$

$\mathrm{TV}(f)$ can be interpreted geometrically as the lateral surface area of the graph of f. If f has many large amplitude oscillations

$\longrightarrow$ f has large lateral surface area

$\longrightarrow$ $\mathrm{TV}(f)$ is large

Numerical methods for total variation

Find regularized solutions to operator equations $\mathcal{K}f = g$ by minimizing Tikhonov-TV functional

$$T_\alpha(f) = \frac{1}{2}||\mathcal{K}f - g||^2 + \alpha \mathrm{TV}(f),$$

where

$$\begin{aligned} \mathrm{TV}(f) &= \int_0^1 \left|\frac{df}{dx}\right| dx, && \text{in 1D,} \\ \mathrm{TV}(f) &= \int_0^1 \int_0^1 |\nabla f| \, dx \, dy, && \text{in 2D.} \end{aligned}$$

Standard methods requiring **gradient** and/or **hessian** info (e.g. Steepest descent, Newton's method) are not suitable due to the non-differentiability of the Euclidean norm at the origin.

Numerical methods for total variation

Find regularized solutions to operator equations $\mathcal{K}f = g$ by minimizing approximate Tikhonov-TV functional

$$\tilde{T}_\alpha(f) = \frac{1}{2}||\mathcal{K}f - g||^2 + \alpha J_\beta(f),$$

where

$$J_\beta(f) = \int_0^1 \sqrt{\left(\frac{df}{dx}\right)^2 + \beta^2}\, dx, \qquad \text{in 1D,}$$

$$J_\beta(f) = \int_0^1 \int_0^1 \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \beta^2}\, dx\, dy, \qquad \text{in 2D.}$$

Now the standard techniques can be used to minimize the discretized approximate Tikhonov-TV functional

$$\tilde{T}(\mathbf{f}) = \frac{1}{2}||\mathbf{K}\mathbf{f} - \mathbf{d}||^2 + \alpha J(\mathbf{f}).$$

Numerical methods for total variation

A one-dimensional discretization

Using a

- composite midpoint quadrature,
- central difference approximation for the derivative,

the approximation to the one-dimensional TV functional becomes

$$J(\mathbf{f}) \;=\; \frac{1}{2}\sum_{i=1}^{n} \psi\big((\mathbf{D}_i\mathbf{f})^2\big)\,\Delta x,$$

where

$$\begin{aligned}
\mathbf{f} &= (f_0,\ldots,f_n)^T, \qquad \text{with } f_i = f(x_i),\; x_i = i\Delta x,\; \Delta x = 1/n,\\
\mathbf{D}_i &= [0,\ldots,0,-1/\Delta x, 1/\Delta x, 0,\ldots,0]_{1\times(n+1)},\\
\psi(t) &= 2\sqrt{t+\beta^2}.
\end{aligned}$$

Numerical methods for total variation

From the directional derivative the **gradient** is derived

$$\begin{aligned}
\frac{d}{d\tau} J(\mathbf{f} + \tau \mathbf{v})\Big|_{\tau=0} &= \sum_{i=1}^{n} \psi'\big((\mathbf{D}_i \mathbf{f})^2\big)\ (\mathbf{D}_i \mathbf{f})(\mathbf{D}_i \mathbf{v}) \Delta x \\
&= \Delta x\ (\mathbf{D}\mathbf{v})^T \operatorname{diag}(\psi'(\mathbf{f}))\ (\mathbf{D}\mathbf{f}) \\
&= \langle \Delta x\ \mathbf{D}^T \operatorname{diag}(\psi'(\mathbf{f}))\ \mathbf{D}\mathbf{f}, \mathbf{v} \rangle \\
&= \langle \operatorname{grad} J(\mathbf{f}), \mathbf{v} \rangle \quad =: \quad \langle \mathbf{L}(\mathbf{f})\mathbf{f}, \mathbf{v} \rangle,
\end{aligned}$$

where

$$\begin{aligned}
[\operatorname{diag}(\psi'(\mathbf{f}))]_{i,i} &= \psi'\big((\mathbf{D}_i \mathbf{f})^2\big), \\
\mathbf{D} &= [\mathbf{D}_1; \ldots; \mathbf{D}_n]_{n \times (n+1)}.
\end{aligned}$$

The matrix $\mathbf{L}(\mathbf{f})$ is symmetric and positive semidefinite.

Numerical methods for total variation

In a similar way the **hessian** is derived

$$\text{Hess}\, J(\mathbf{f}) \;=\; \mathbf{L}(\mathbf{f}) + \mathbf{L}'(\mathbf{f})\mathbf{f},$$

where

$$\begin{aligned}\mathbf{L}'(\mathbf{f})\mathbf{f} &= \Delta x\, \mathbf{D}^T \operatorname{diag}(2(\mathbf{D}\mathbf{f})^2\psi''(\mathbf{f}))\mathbf{D}, \\ [\operatorname{diag}(2(\mathbf{D}\mathbf{f})^2\psi''(\mathbf{f}))]_{i,i} &= 2(\mathbf{D}_i\mathbf{f})^2\psi''((\mathbf{D}_i\mathbf{f})^2).\end{aligned}$$

Numerical methods for total variation

In a similar way the **hessian** is derived

$$\text{Hess } J(\mathbf{f}) \;=\; \mathbf{L}(\mathbf{f}) + \mathbf{L}'(\mathbf{f})\mathbf{f},$$

where

$$\begin{aligned}\mathbf{L}'(\mathbf{f})\mathbf{f} &= \Delta x\, \mathbf{D}^T \operatorname{diag}(2(\mathbf{D}\mathbf{f})^2\psi''(\mathbf{f}))\mathbf{D},\\ [\operatorname{diag}(2(\mathbf{D}\mathbf{f})^2\psi''(\mathbf{f}))]_{i,i} &= 2(\mathbf{D}_i\mathbf{f})^2\psi''((\mathbf{D}_i\mathbf{f})^2).\end{aligned}$$

The **gradient** and **hessian** of the discretized approximate Tikhonov-TV functional are now simply given by

$$\begin{aligned}\text{grad } \tilde{T}(\mathbf{f}) &= \mathbf{K}^T(\mathbf{K}\mathbf{f} - \mathbf{d}) + \alpha\mathbf{L}(\mathbf{f})\mathbf{f},\\ \text{Hess } \tilde{T}(\mathbf{f}) &= \mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}(\mathbf{f}) + \alpha\mathbf{L}'(\mathbf{f})\mathbf{f}.\end{aligned}$$

Numerical methods for total variation

1. Steepest descent with line search for total variation

$$\mathbf{f}_{\nu+1} \;=\; \mathbf{f}_\nu - \tau \,\mathrm{grad}\, \tilde{T}(\mathbf{f}_\nu)$$

Algorithm:

```
$\nu := 0$;
$\mathbf{f}_0 :=$ initial guess;
while no convergence
    $\mathbf{g}_\nu := \mathbf{K}^T(\mathbf{K}\mathbf{f}_\nu - \mathbf{d}) + \alpha \mathbf{L}(\mathbf{f}_\nu)\mathbf{f}_\nu$;   % gradient
    $\tau_\nu := \arg\min_{\tau>0} \tilde{T}(\mathbf{f}_\nu - \tau \mathbf{g}_\nu)$;      % line search
    $\mathbf{f}_{\nu+1} := \mathbf{f}_\nu - \tau_\nu \mathbf{g}_\nu$;                    % update approximate solution
    $\nu := \nu + 1$;
```

Numerical methods for total variation

2. Newton's method with line search for total variation

$$\mathbf{f}_{\nu+1} \;=\; \mathbf{f}_\nu - \tau\,(\mathrm{Hess}\,\tilde{T}(\mathbf{f}_\nu))^{-1}\,\mathrm{grad}\,\tilde{T}(\mathbf{f}_\nu)$$

Algorithm:

```
$\nu := 0$;
$\mathbf{f}_0 :=$ initial guess;
while no convergence
    $\mathbf{g}_\nu := \mathbf{K}^T(\mathbf{K}\mathbf{f}_\nu - \mathbf{d}) + \alpha\mathbf{L}(\mathbf{f}_\nu)\mathbf{f}_\nu$;   % gradient
    $\mathbf{H}_J := \mathbf{L}(\mathbf{f}_\nu) + \mathbf{L}'(\mathbf{f}_\nu)\mathbf{f}_\nu$;               % hessian of penalty functional
    $\mathbf{H} := \mathbf{K}^T\mathbf{K} + \alpha\mathbf{H}_J$;                         % hessian of cost functional
    $\mathbf{s}_\nu := -\mathbf{H}^{-1}\mathbf{g}_\nu$;                              % newton step
    $\tau_\nu := \arg\min_{\tau>0} \tilde{T}(\mathbf{f}_\nu + \tau\mathbf{s}_\nu)$;        % line search
    $\mathbf{f}_{\nu+1} := \mathbf{f}_\nu + \tau_\nu\mathbf{s}_\nu$;                      % update approximate solution
    $\nu := \nu + 1$;
```

Numerical methods for total variation

3. Lagged diffusivity fixed point iteration

$$\begin{aligned}\mathbf{f}_{\nu+1} &= (\mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}(\mathbf{f}_\nu))^{-1}\mathbf{K}^T\mathbf{d} \\ &= \mathbf{f}_\nu - (\mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}(\mathbf{f}_\nu))^{-1}\mathrm{grad}\,\tilde{T}(\mathbf{f}_\nu).\end{aligned}$$

- Fixed point form can be derived by setting $\mathrm{grad}\,\tilde{T}(\mathbf{f}) = \mathbf{0}$.
- The matrix $\mathbf{L}(\mathbf{f})$ can be viewed as a discretization of a steady-state diffusion operator and is evaluated at $\mathbf{f}_\nu$, hence the name.
- The equivalent quasi-Newton iteration can also be derived by dropping the term $\alpha\mathbf{L}'(\mathbf{f})\mathbf{f}$ from the Hessian.
- The quasi Newton form tends to be less sensitive to roundoff error.

Numerical methods for total variation

3. Lagged diffusivity fixed point iteration

Algorithm:

```
$\nu := 0$;
$\mathbf{f}_0 :=$ initial guess;
while no convergence
    $\mathbf{g}_\nu := \mathbf{K}^T(\mathbf{K}\mathbf{f}_\nu - \mathbf{d}) + \alpha\mathbf{L}(\mathbf{f}_\nu)\mathbf{f}_\nu$;   % gradient
    $\mathbf{H} := \mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}(\mathbf{f}_\nu)$;   % approximate hessian of cost functional
    $\mathbf{s}_\nu := -\mathbf{H}^{-1}\mathbf{g}_\nu$;   % quasi-newton step
    $\mathbf{f}_{\nu+1} := \mathbf{f}_\nu + \mathbf{s}_\nu$;   % update approximate solution
    $\nu := \nu + 1$;
```

If $\mathbf{K}^T\mathbf{K}$ is positive definite, then this fixed point iteration converges globally and no line-search is needed.

Because the Hessian is approximated, only linear convergence is expected.

Numerical methods for total variation

Results for one-dimensional test problem

- All three methods give essentially the same reconstruction.
- Measure for numerical performance is relative iterative solution error

$$e_\alpha^\nu = \frac{||\mathbf{f}_\alpha^\nu - \mathbf{f}_\alpha||}{||\mathbf{f}_\alpha||},$$

with $\mathbf{f}_\alpha$ the minimizer of the approximate Tikhonov-TV functional (use accurate approximation obtained with Primal-Dual Newton method).

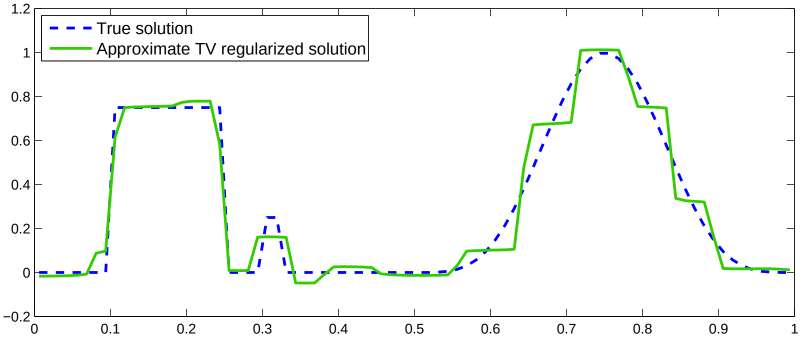
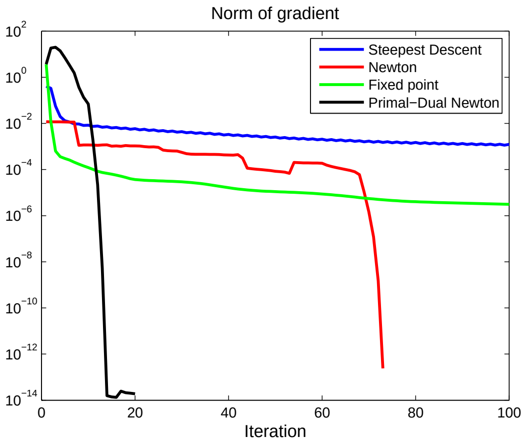


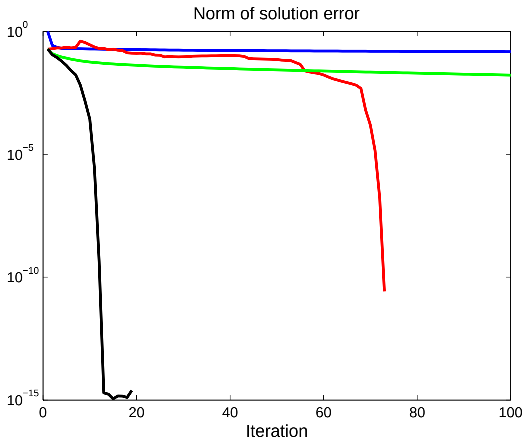
Figure 5: $\alpha = 0.0001, \beta = 0.1$

Numerical methods for total variation

Results for one-dimensional test problem

- Steepest Descent: Hessian ill-conditioned → slow (linear) convergence
- Newton: Line search restricts stepsize → local quadratic convergence attained late (α small, β small)
- Cost per iteration is about the same for Newton, Fixed point and Primal-Dual Newton





Summary

- Variational representation of Tikhonov functional
- TV is a penalty functional which penalizes oscillatory solutions
- Approximate TV functional needed for standard optimization tools

References

[Vog02] Curtis R. Vogel. *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics, 2002.