Solving Bilevel Optimization Problems in Image Processing via Inexact Trust Region Methods

David Villacís

Ecuadorian Research Center on Mathematical Modelling
Escuela Politécnica Nacional de Quito

Joint work with: Juan Carlos De Los Reyes

SVAN 2016
Rio de Janeiro - Brazil, 2016

Contents

Introduction

Original Image

Noisy Image

Introduction

- We will be referring specifically to the problem of image denoising.

Introduction

- We will be referring specifically to the problem of image denoising.
- Image De-noising problems have been succesfully treated using variational models that use the TV semi-norm (Rudin-Osher-Fatemi Model).

Introduction

Mode Mat

- We will be referring specifically to the problem of image denoising.
- Image De-noising problems have been succesfully treated using variational models that use the TV semi-norm (Rudin-Osher-Fatemi Model).
- These models include a series of balance parameters that allow some control on the denoising task.

Parameter Learning Framework

We will using a bilevel optimization approach for the determination of the balance of the regulaization parameter in TV image denoising.

Parameter Learning Framework

We will using a bilevel optimization approach for the determination of the balance of the regulaization parameter in TV image denoising.

$$\min_{\lambda_i \geq 0, i=1,\ldots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2_X \qquad (1)$$

Parameter Learning Framework

We will using a bilevel optimization approach for the determination of the balance of the regulaization parameter in TV image denoising.

$$\min_{\lambda_i \geq 0, i=1,\dots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2_X \tag{1}$$

subject to:

Parameter Learning Framework

We will using a bilevel optimization approach for the determination of the balance of the regulaization parameter in TV image denoising.

$$\min_{\lambda_i \geq 0, i=1,\dots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|_{L^2(\Omega)}^2 + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2 \tag{1}$$

subject to:

$$\min_u \left(|Du|(\Omega) + \sum_{i=1}^{d} \int_\Omega \lambda_i \phi_i(u,f) dx \right) \tag{2}$$

Optimization Problem in $H_0^1(\Omega)$

$$\min_{\lambda_i \geq 0, i=1,\ldots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2_X \tag{3}$$

Optimization Problem in $H_0^1(\Omega)$

$$\min_{\lambda_i \geq 0, i=1,\ldots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2_X \tag{3}$$

subject to:

Optimization Problem in $H_0^1(\Omega)$

$$\min_{\lambda_i \geq 0, i=1,\ldots,d} \sum_{k=1}^{N} \|\overline{u}_k - u_k\|^2_{L^2(\Omega)} + \beta \sum_{i=1}^{d} \|\lambda_i\|^2_X \tag{3}$$

subject to:

$$\min_u \left(\frac{\epsilon}{2} \|Du\|^2_{L^2} + |Du|(\Omega) + \sum_{i=1}^{d} \int_\Omega \lambda_i \phi_i(u,f)\,dx \right) \tag{4}$$

Optimization Problem in $H_0^1(\Omega)$

ModeMat

Considering just Gaussian noise $\phi = |u - f|^2$

Optimization Problem in $H_0^1(\Omega)$

Considering just Gaussian noise $\phi = |u - f|^2$
A necessary and sufficient condition for (4) is given by the following elliptic variational inequality:

Optimization Problem in $H_0^1(\Omega)$

Considering just Gaussian noise $\phi = |u-f|^2$
A necessary and sufficient condition for (4) is given by the following elliptic variational inequality:

$$\epsilon(Du, D(v-u))_{L^2} + \int_\Omega \lambda(u-f)(v-u)dx + \int_\Omega |Dv|dx - \int_\Omega |Du|dx \geq 0 \ \forall v \in H_0^1(\Omega) \tag{5}$$

Optimization Problem in $H_0^1(\Omega)$

Considering just Gaussian noise $\phi = |u - f|^2$
A necessary and sufficient condition for (4) is given by the following elliptic variational inequality:

$$\begin{aligned} &\epsilon(Du, D(v-u))_{L^2} + \int_\Omega \lambda(u-f)(v-u)dx \\ &+ \int_\Omega |Dv|dx - \int_\Omega |Du|dx \geq 0 \; \forall v \in H_0^1(\Omega) \end{aligned} \tag{5}$$

This problem corresponds to an optimization problem governed by a variational inequality of the second kind.

Optimization Problem in Finite Dimensional Spac

The finite dimensional version of this variational inequality corresponds to:

Optimization Problem in Finite Dimensional Spac

The finite dimensional version of this variational inequality corresponds to:

$$\langle Au, v-u\rangle + \langle \lambda(u-f), v-u\rangle + \sum_{j=1}^{m} (|(\mathbb{K}v)_j| - |(\mathbb{K}u)_j|) \geq 0 \ \forall v \in \mathbb{R}^n \quad (6)$$

Optimization Problem in Finite Dimensional Spac

The finite dimensional version of this variational inequality corresponds to:

$$\langle Au, v-u\rangle + \langle \lambda(u-f), v-u\rangle + \sum_{j=1}^{m} (|(\mathbb{K}v)_j| - |(\mathbb{K}u)_j|) \geq 0 \ \forall v \in \mathbb{R}^n \quad (6)$$

- $A \in^{n\times n}$ symmetric positive definite (discrete Laplacian)
- $K^{(i)} \in^{m\times n}$, $i = 1, ..., d$, discrete i-th partial derivative,

$$\mathbb{K} : \mathbb{R}^n \to \mathbb{R}^{m\times d}, \quad \mathbb{K}u = (K^{(1)}u, ..., K^{(d)}u)$$

- $(\mathbb{K}u)_j \in \mathbb{R}^d$, $j = 1, ..., m$, j-th row of u, corresponds to discrete gradient at element j

Directional Differentiability of the Solution Operat

If we consider the following perturbed problem, given $h \in \mathbb{R}, t > 0$:

$$Au^t + (\lambda + th)(u - f) - \mathbb{K}^* q^t = 0 \tag{7}$$

$$\langle q_j^t, (\mathbb{K}u^t)_j \rangle = |(\mathbb{K}u^t)_j|, |q_j^t| \leq 1 \tag{8}$$

Theorem (Directional Differentiability)

The solution operator of (6) is directionally differentiable. Its directional derivative at $\lambda \in \mathbb{R}$ in the direction $h \in \mathbb{R}$ is the unique solution $\eta \in \mathbb{R}$ of the VI of the first kind:

$$\begin{aligned} & \eta \in \mathcal{K}(u), \\ & \langle A\eta, v-\eta\rangle + \langle h(\eta - f), v-\eta\rangle + \\ \sum_{j\in\mathcal{I}(u)} \Big\langle \frac{(\mathbb{K}\eta)_j}{|(\mathbb{K}u)_j|} - \langle(\mathbb{K}u)_j, (\mathbb{K}\eta)_j\rangle \frac{(\mathbb{K}u)_j}{|(\mathbb{K}u)_j|^3}, (\mathbb{K}v)_j - (\mathbb{K}\eta)_j \Big\rangle & \geq 0, \\ & \forall\, v \in \mathcal{K}(u) \end{aligned} \tag{9}$$

with

$$\begin{aligned} \mathcal{I}(u) &:= \{j \in \{1,...,m\} : (\mathbb{K}u)_j \neq 0\}, \\ \mathcal{A}(u) &:= \{1,...,m\} \setminus \mathcal{I}(u), \\ \mathcal{K}(u) &:= \{v \in^n\colon\ \langle q_j, (\mathbb{K}v)_j\rangle \geq |(\mathbb{K}v)_j| \quad \forall j \in \mathcal{A}(u)\}. \end{aligned} \tag{10}$$

ModeMat

Lemma

$\mathcal{K}(u) =$

$$\left\{ v \in \mathbb{R}^n : \langle (\lambda - A)u - \lambda f, v \rangle \geq \sum_{j \in \mathcal{I}(u)} \left\langle \frac{(\mathbb{K}u)_j}{|(\mathbb{K}u)_j|}, (\mathbb{K}v)_j \right\rangle + \sum_{j \in \mathcal{A}(y)} |(\mathbb{K}v)_j| \right\}$$

- $\mathcal{K}(u)$ is unique
- η is unique
- $S : \mathbb{R} \to \mathbb{R}^n$ is direccionally differentiable.

Inexact Trust Region Method

We will propose an inexact trust region algorithm for the solution of the following finite dimensional optimization problem:

Inexact Trust Region Method

We will propose an inexact trust region algorithm for the solution of the following finite dimensional optimization problem:

$$\min J(u, \lambda) = \frac{1}{2}|u - z_d|^2 + \frac{\beta}{2}|\lambda|^2 \tag{11}$$

subject to: $$\langle Au, v-u\rangle + \langle \lambda(u-f), v-u\rangle + \sum_{j=1}^{m}(|(\mathbb{K}v)_j| - |(\mathbb{K}u)_j|) \geq 0 \tag{12}$$

Inexact Trust Region Method

We will propose an inexact trust region algorithm for the solution of the following finite dimensional optimization problem:

$$\min J(u,\lambda) = \frac{1}{2}|u - z_d|^2 + \frac{\beta}{2}|\lambda|^2 \tag{11}$$

subject to: $$\langle Au, v-u\rangle + \langle \lambda(u-f), v-u\rangle + \sum_{j=1}^{m}(|(\mathbb{K}v)_j| - |(\mathbb{K}u)_j|) \geq 0 \tag{12}$$

Recall that the inactive set is given by:

$$\mathcal{I} := \{j \in \{1,\ldots,m\} : (\mathbb{K}u)_j \neq 0\} \tag{13}$$

Inexact Trust Region Method

Linearized system in case of strict complementarity

$$
\begin{aligned}
(A+\lambda)\eta + \mathbb{K}^*\xi &= h(f-u) \\
\xi_j - \frac{(\mathbb{K}\eta)_j}{|(\mathbb{K}u)_j|} + \langle (\mathbb{K}u)_j, (\mathbb{K}\eta)_j \rangle \frac{(\mathbb{K}y)_j}{|(\mathbb{K}u)_j|^3} &= 0 \qquad \forall j \in \mathcal{I}(u) \\
(\mathbb{K}\eta)_j &= 0 \qquad \forall j \in \mathcal{A}(u)
\end{aligned}
$$

Inexact Trust Region Method

Linearized system in case of strict complementarity

$$
\begin{aligned}
(A+\lambda)\eta + \mathbb{K}^*\xi &= h(f-u) \\
\xi_j - \frac{(\mathbb{K}\eta)_j}{|(\mathbb{K}u)_j|} + \langle (\mathbb{K}u)_j, (\mathbb{K}\eta)_j \rangle \frac{(\mathbb{K}y)_j}{|(\mathbb{K}u)_j|^3} &= 0 \qquad \forall j \in \mathcal{I}(u) \\
(\mathbb{K}\eta)_j &= 0 \qquad \forall j \in \mathcal{A}(u)
\end{aligned}
$$

Strict complementarity adjoint

$$
\begin{aligned}
(A+\lambda)p + \mathbb{K}^*\xi &= u - z_d \\
\xi_j - \frac{(\mathbb{K}p)_j}{|(\mathbb{K}u)_j|} + \langle (\mathbb{K}u)_j, (\mathbb{K}p)_j \rangle \frac{(\mathbb{K}u)_j}{|(\mathbb{K}u)_j|^3} &= 0 \qquad \forall j \in \mathcal{I}(u) \\
(\mathbb{K}p)_j &= 0 \qquad \forall j \in \mathcal{A}(u)
\end{aligned}
$$

Inexact Trust Region Method

Derivative candidate

$$j'(\lambda)h = \beta\langle\lambda, h\rangle - h\langle(u-f), p\rangle = 0$$

The quadratic model of the reduced cost function is given by

$$q_k(s) = j(\lambda_k) + g_k^T s + \frac{1}{2} s^T H_k s,$$

where $g_k = j'(\lambda_k)$ and H_k is a matrix with second order information, obtained with the BFGS method.

Inexact Trust Region Method

Actual and predicted reductions

$$ared_k(s^k) := j(\lambda_k) - j(\lambda_k + s^k) \quad \text{and} \quad pred_k(s^k) = j(\lambda_k) - q_k(s^k).$$

The quality indicator is computed by

$$\rho_k(s^k) = \frac{ared_k(s^k)}{pred_k(s^k)}.$$

Numerical Experiment

- We solved the low level problem based on a Newton Semi-Smooth Method

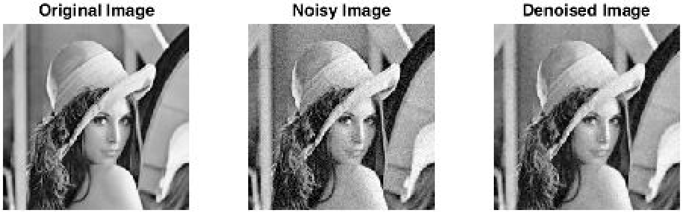
Numerical Experiment

- We solved the low level problem based on a Newton Semi-Smooth Method
- We implemented the trust region algorithm using the previously discussed forms for the adjoint and gradient calculation.

Numerical Experiment

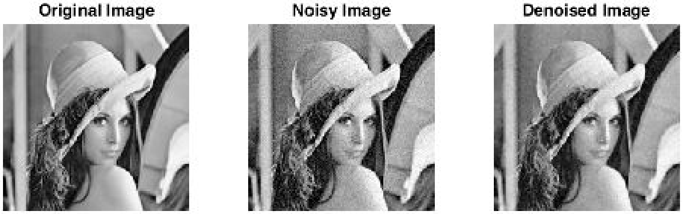
- We solved the low level problem based on a Newton Semi-Smooth Method
- We implemented the trust region algorithm using the previously discussed forms for the adjoint and gradient calculation.

Numerical Experiment



(b) Gaussian Noise $\sigma = 0.002$

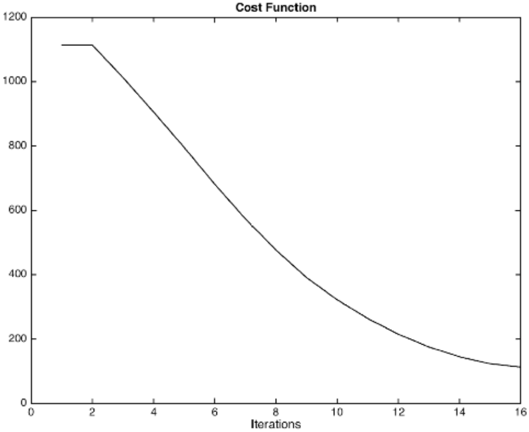
Numerical Experiment



(c) Gaussian Noise $\sigma = 0.002$

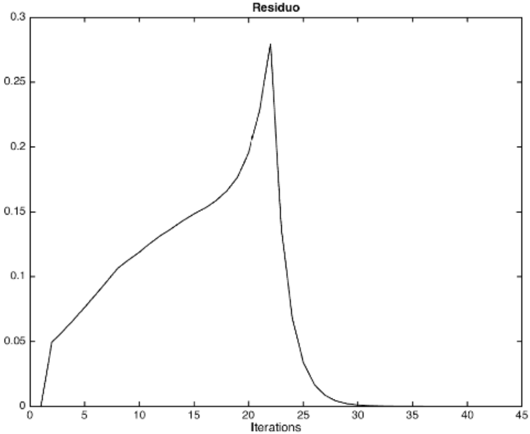
Here the optimal parameter was calculated at $\lambda = 10762$.

Numerical Experiment



(d) Evolution of the cost function $\sigma = 0.002$

Numerical Experiment



(e) Evolution of the residue $\sigma = 0.002$

Numerical Experiment

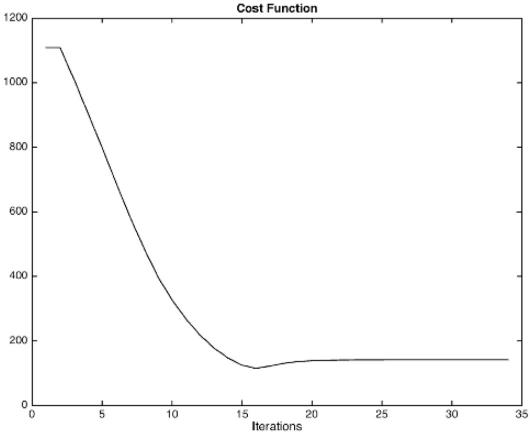
(f) Gaussian Noise $\sigma = 0.02$

Numerical Experiment

(g) Gaussian Noise $\sigma = 0.02$

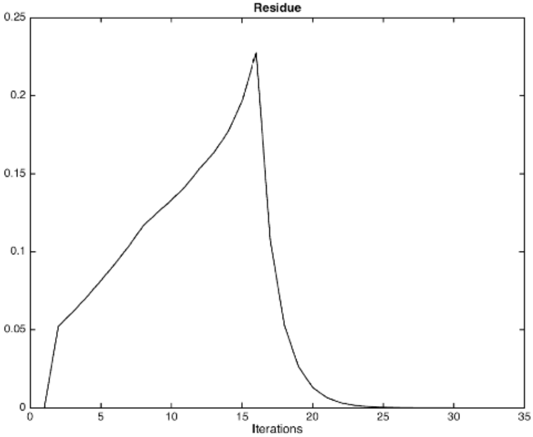
Here the optimal parameter was calculated at $\lambda = 2786$.

Numerical Experiment



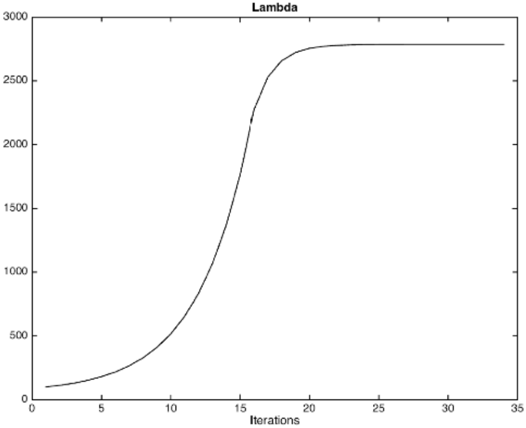
(h) Evolution of the cost function $\sigma = 0.02$

Numerical Experiment



(i) Evolution of the residue $\sigma = 0.02$

Numerical Experiment



(j) Evolution of the residue $\sigma = 0.02$

Remarks

- Instead of using a regularization technique for handling the non smooth term, we proposed a method that takes advantage of the differentiability of the solution operator.

Remarks

- Instead of using a regularization technique for handling the non smooth term, we proposed a method that takes advantage of the differentiability of the solution operator.
- Since the information generated does not necessarily correspond to an element of the sub differential, that is the case of a non-empty bi-active set, we can see that an inexact trust region algorithm provides robust iterates.

THANKS!

@davillacis

david.villacis01@epn.edu.ec

ModeMat