Variational Methods in Image Denoising

Jamylle Carter

Postdoctoral Fellow

Mathematical Sciences Research Institute (MSRI)

MSRI Workshop for Women in Mathematics:
Introduction to Image Analysis
22 January 2005

Collaborators

- **Tony F. Chan (University of California, Los Angeles [UCLA])**
  - Professor, Department of Mathematics
  - Dean, Division of Physical Sciences, College of Letters and Science
- **Lieven Vandenberghe (UCLA)**
  - Associate Professor, Department of Electrical Engineering

Outline

1. Edges in image processing
2. Insufficiency of classical tools
3. Total Variation (TV) model of Rudin-Osher-Fatemi
   (a) **Variational methods**
      i. [BHKU02] Short course on variational methods
      ii. Total variation (TV) regularization
4. Computational challenges of primal TV
5. Dual TV method
   (a) Derivation of problem
   (b) Previous and current work
6. Acknowledgments

Digital Images

Digital Image An image $f(x, y)$ discretized in both spatial coordinates and in brightness

Gray scale Digital Image = Matrix

- row and column indices = point in image
- matrix element value = gray level at that point
- **pixel** = element of digital array or **picture element**

Example

- Square
- Surface Plot of Square
- Edges in 1D

Digital Image Processing

The set of techniques for the manipulation, correction, and enhancement of digital images

Methods in Image Processing

- Fourier/wavelet transforms
- Stochastic/statistical methods
- **Partial differential equations (PDEs) and differential/geometric models**
  - Systematic treatment of geometric features of images (shape, contour, curvature)
  - Wealth of techniques for PDEs and computational fluid dynamics

Gibbs Phenomenon

- Fourier series approximation of a square wave
- Another view
- Animation

Back to previous page

Partial Differential Equations

Computational Fluid Dynamics

- Think of a fluid as as a collection of discrete particles (molecules).
- Too complicated, so represent fluid as a continuum.
- To solve the continuous problem, use numerics to re-discretize the fluid in the right way.

Image Processing

- A digital image is a collection of discrete particles (pixels).
- Hard to manipulate each individual pixel, so represent the image as a continuous function.
- To solve the continuous problem, use numerics to re-discretize the image in the right way.

Linear Diffusion [Wit83]

Proposal: Filter an original image $u^0(x, y)$ with Gaussian kernels of variance $2t$.

Result: A one-parameter family of images $u(t, x, y)$ Witkin referred to as "a scale space."

Limitation: Linear Gaussian smoothing blurs and displaces images. Example

Insight: The family of images $u(t, x, y)$ is the solution of the linear heat equation with $u^0(x, y)$ as the initial data:

$$u_t = u_{xx} + u_{yy}$$
$$u^0(x, y) = u^0(x, y)$$

New Idea: Process images by evolving **nonlinear** PDEs $\rightarrow$ using appropriate function spaces

The Fundamental Solution of the Heat Equation

$$u_t = u_{xx} + u_{yy}$$
$$u^0(x, y) = u^0(x, y)$$

- Parabolic partial differential equation
- Describes diffusion of
  - Heat in a region Ω
  - Dye or other substance in a still fluid
- At a microscopic level, it results from random processes.

Outline

1. Edges in image processing
2. Insufficiency of classical tools
3. Total Variation (TV) model of Rudin-Osher-Fatemi
   (a) **Variational methods**
      i. [BHKU02] Short course on variational methods
      ii. Total variation (TV) regularization
4. Computational challenges of primal TV
5. Dual TV method
   (a) Derivation of problem
   (b) Previous and current work
6. Acknowledgments

Examples of Digital Image Processing

Smoothing Removing bad data.

Sharpening Highlighting **edges** (discontinuities).

Restoration Determination of unknown original image from given noisy image.

- Usually involves prior knowledge about the noise process (*e.g.*., Gaussian noise).
- Ill-conditioned inverse problem.
- No unique solution.
- **Regularization techniques** impose **desirable properties** on the solution by **restricting the solution space**.

How To Describe Desired Image?

- Contains **edges** (discontinuities).
- In other words, preserves high gradient in a geometric setting.
- **Qualitative description.**

How To Obtain Desired Image?

- Create a **functional**, which upon being optimized, achieves the stated goal of an image with edges.
- Choose the right **solution space**.
- Called the **variational approach**.
- Formulated in continuous domain which has many analytical tools.

Total Variation

- Original image u has simple geometric description.

  Objects: Set of connected sets.

  Edges: Smooth contours of objects.

- Image is smooth inside the objects but has jumps across the boundaries.

- The functional space modeling these properties is $BV(\Omega)$, the space of integrable functions with finite total variation

$$TV(u) = \int_\Omega |\nabla u|,$$

where Ω denotes the image domain (for instance, the computer screen) and is usually a rectangle.

Total Variation

$$
\begin{aligned}
TV(u) &= \int_\Omega |\nabla u| \\
&= \int_a^b \left| \frac{du}{dx} \right| \, dx \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \left| \frac{du(x_k)}{dx} \right| \Delta x \\
&\approx \sum_{k=1}^{n} \left| \frac{u(x_{k+1}) - u(x_k)}{\Delta x} \right| \Delta x \\
&\approx \sum_{k=1}^{n} |u(x_{k+1}) - u(x_k)|
\end{aligned}
$$

Bounded Variation

- A function $f(x)$ is said to have **bounded variation** if, over the closed interval $x \in [a, b]$, there exists an M such that

$$|f(x_1) - f(a)| + |f(x_2) - f(x_1)| + \cdots + |f(b) - f(x_{n-1})| \leq M$$

for all $a < x_1 < x_2 < \cdots < x_{n-1} < b$.

- The total variation (TV) is the vertical component of the arc-length of the graph of f.
- Example
- For smooth images, the TV norm is equivalent to the L^1 norm of the derivative.
- TV norm is some measure of the amount of oscillation found in the function $u(x)$.

Basic Properties of BV Space

- Provides regularity of solutions.
- Allows sharp discontinuities (edges).

Total Variation Integral

$$TV(u) = \int_\Omega |\nabla u| \, dx$$
$$= \sup \left\{ \int_\Omega u(\nabla \cdot \mathbf{w}) \, d\mathbf{x} : \ \mathbf{w} \in C_c^1(\Omega, \mathbb{R}^2), \ |\mathbf{w}(\mathbf{x})| \leq 1 \, \forall \mathbf{x} \in \Omega \right\}$$

- $C_c^1(\Omega, \mathbb{R}^2)$ is the set of functions with continuous first derivatives and compact support on Ω.
- $u \in L^1(\Omega)$
- $\Omega \subset \mathbb{R}^2$ is a bounded open set.

New Definition of BV Space

$$BV(\Omega) = \left\{ u \in L^1(\Omega) \; : \; \int_\Omega |\nabla u| < \infty \right\}$$

Features

- BV functions are L^1 functions with bounded TV semi-norm.
- u need not be differentiable
- Discontinuities allowed
- Derivatives considered in the weak sense

Functions of bounded variation need not be differentiable!

Total Variation Minimization [ROF92]

$$\min_{u \in BV(\Omega)} \int_\Omega \left(\alpha |\nabla u| + \frac{1}{2} |u - z|^2 \right) dx$$

α**:** Regularization parameter (user-chosen scalar).

Unknown image: A real-valued function $u : \Omega \to \mathbb{R}$.

Noisy image: A real-valued function $z : \Omega \to \mathbb{R}$.

Ω**:** Nonempty, bounded, open set in $\mathbb{R}^2$ (usually a **square**).

Strictly Convex Functional: Admits unique minimum.

Intuition Behind Total Variation Minimization

$$\boxed{\min_u \int_\Omega \alpha|\nabla u| + \frac{1}{2}(u-z)^2 \, dx}$$

- $\min_u |\nabla u|$
  - $|\nabla u| = 0$
  - u constant
- $\min_u \|u - z\|$
  - $u = z$
  - u noisy image
- Weighted sum between constant image and noisy image

Calculus of Variations

- A branch of mathematics that is a sort of generalization of calculus.
  - Deals with functions of functions (**functionals**) as opposed to functions of numbers.
  - Functionals can be formed as integrals involving an unknown function $y(x)$ and its derivatives.
- Calculus of variations seeks to find the path, curve, surface, etc., $y(x)$ for which a given functional has a stationary value (which, in physical problems, is usually a minimum or maximum).

Calculus of Variations

Extremal functions: Those functions $y(x)$ making the functional attain a maximum or minimum value.

Variational: Used of all extremal functional questions.

- Mathematically, this involves finding stationary values $y(x)$ of integrals of the form

$$I = \int_a^b f(y, y', x)\, dx$$

- I has an extremum only if the Euler-Lagrange differential equation is satisfied, i.e., if

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0$$

- Brachistochrone curve

Euler-Lagrange Equation (PDE)

- 1st-order necessary condition for the minimizer u

$$\min_u \int_\Omega \alpha|\nabla u| + \frac{1}{2}(u-z)^2\,dx$$

- **Theory**

$$-\alpha\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right) + u - z = 0$$

- Degenerate **nonlinear elliptic PDE** when $|\nabla u| = 0$
- **Practice**

$$\boxed{-\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right) + u - z = 0}$$

small $\beta > 0$

Previous Work

$$-\alpha \nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta}} \right) + u - z = 0$$

Rudin-Osher-Fatemi 1992 Time marching to steady state with gradient descent. Improvement in **Marquina-Osher 1999**.

Chan-Chan-Zhou 1995 Continuation procedure on β.

Vogel-Oman 1996 Fixed point iteration.

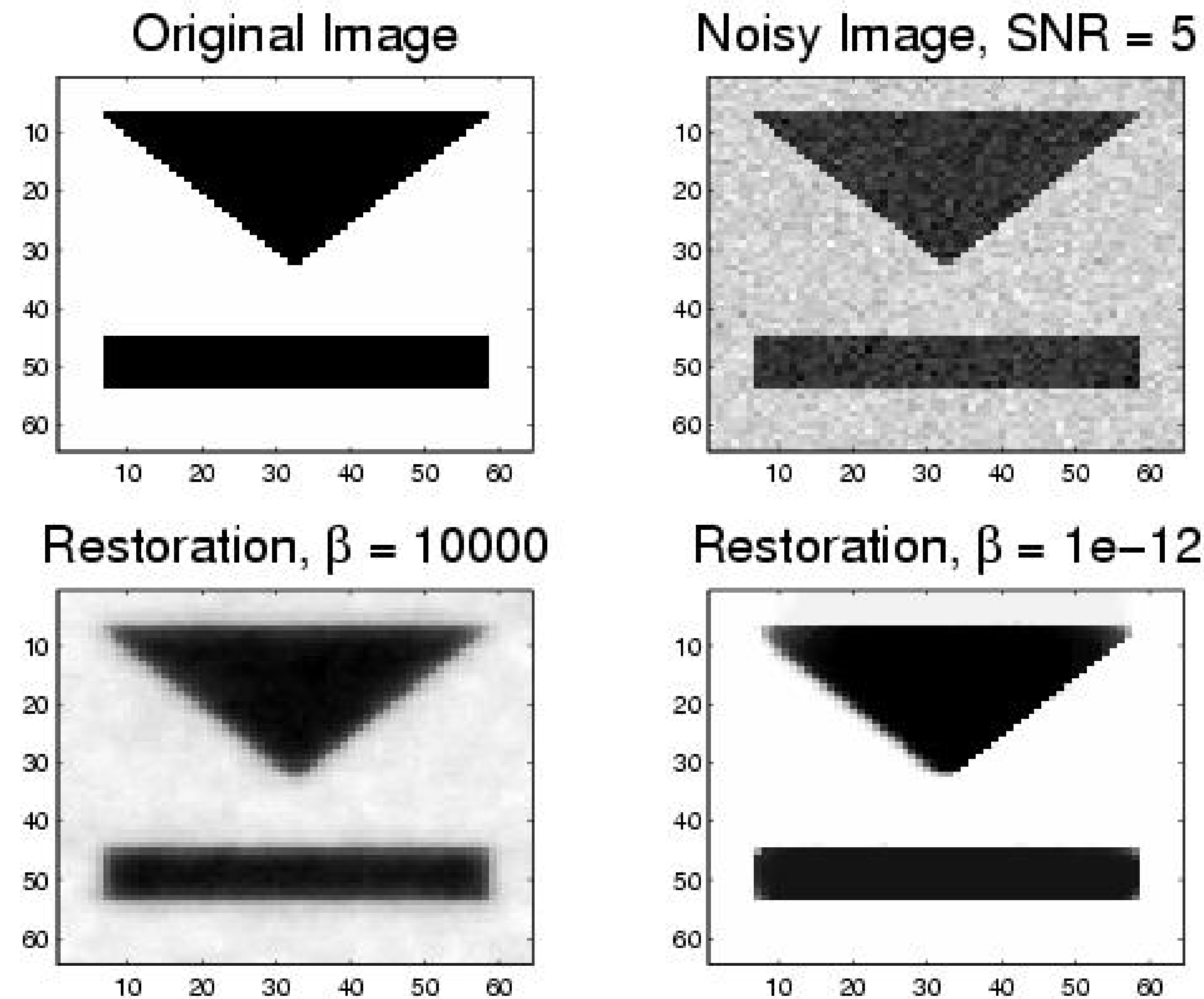
Dependence on β

$$\boxed{-\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)+u-z=0} \tag{1}$$

β **large** Smeared edges.

β **small** PDE nearly degenerate.

No β if we rewrite (1) in terms of new variable...


Original Image
Noisy Image, SNR = 5
Restoration, β = 10000
Restoration, β = 1e−12
10 20 30 40 50 60

Introduce Dual Variable w

Francfort/Chan-Golub-Mulet 1995 [CGM99]
Giusti 1984 [Giu84]

Total Variation Integral

$$TV(u) = \int_\Omega |\nabla u|\,dx$$
$$= \sup\left\{ \int_\Omega u(\nabla \cdot \mathbf{w})\,d\mathbf{x} : \ \mathbf{w} \in C_c^1(\Omega, \mathbb{R}^2), \ |\mathbf{w}(\mathbf{x})| \leq 1\,\forall \mathbf{x} \in \Omega \right\}$$

- $C_c^1(\Omega, \mathbb{R}^2)$ is the set of functions with continuous first derivatives and compact support on Ω.

Dot Product

$$a \cdot b = |a||b| \cos\theta$$

$$\max_{|b| \leq 1} a \cdot b = \max_{|b| \leq 1} |a||b| \cos\theta$$

$$\cos\theta = 1 \iff \theta = 0$$

$$\Rightarrow b \text{ coincident with } a$$

$$\Rightarrow b = \frac{a}{\|a\|}$$

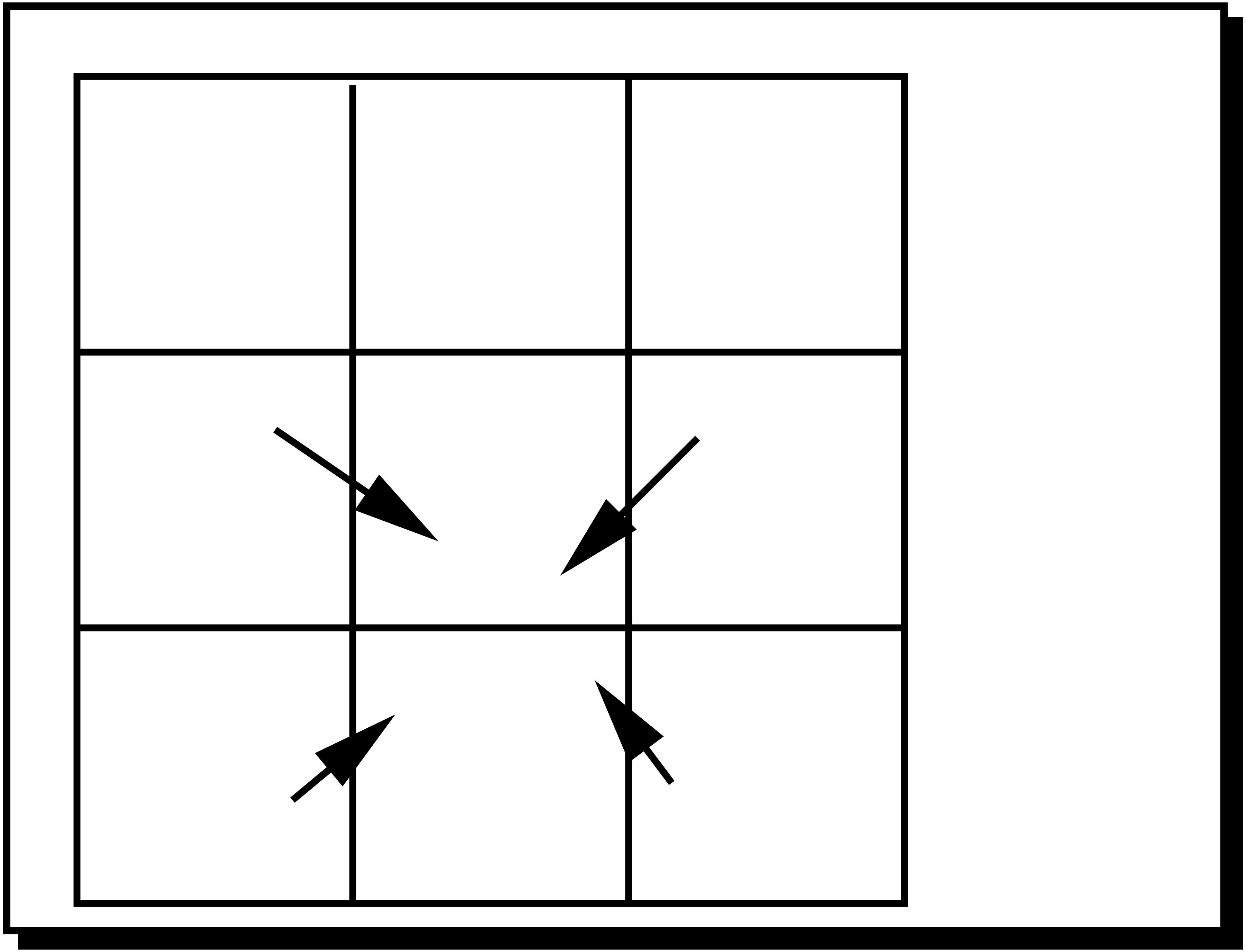
$$\boxed{\max_{|b| \leq 1} a \cdot b = \|a\|}$$

Intuition Behind Dual Variable w

$$\boxed{|\nabla u| = \max_{|w|\leq 1} (\nabla u \cdot w)}$$

Interpretation

$$w = \begin{cases} \frac{\nabla u}{|\nabla u|} & u \text{ smooth and } |\nabla u| \neq 0 \\ \text{not unique} & \text{otherwise} \end{cases}$$



Deriving Dual Formulation

$$\boxed{\min_u \int_\Omega \alpha|\nabla u| + \frac{1}{2}(u-z)^2\,dx}$$

- Dual definition of Total Variation

$$= \min_u \int_\Omega \alpha \max_{|w|\le 1} -u(\nabla\cdot w) + \frac{1}{2}(u-z)^2\,dx$$

- Interchange max and min

$$= \max_{|w|\le 1} \min_u \underbrace{\int_\Omega -\alpha\, u(\nabla\cdot w) + \frac{1}{2}(u-z)^2\,dx}_{\Psi(u)}$$

- Quadratic function of u

$$\nabla\Psi(u) = \vec{0} \Longleftrightarrow u = z + \alpha(\nabla\cdot w)$$

Deriving Dual Formulation

- Write u in terms of w

$$\max_{|w|\le 1} \int_\Omega -\alpha \underbrace{(z+\alpha(\nabla\cdot w))}_{u}(\nabla\cdot w) + \frac{1}{2}(\underbrace{z+\alpha(\nabla\cdot w)}_{u} - z)^2 \, dx$$

- Dual max problem

$$\max_{|w|\le 1} \int_\Omega -\alpha z(\nabla\cdot w) - \frac{\alpha^2}{2}(\nabla\cdot w)^2 \, dx$$

Deriving Dual Formulation

- Dual max problem

$$\max_{|w|\leq 1} \int_\Omega -\alpha z(\nabla \cdot w) - \frac{\alpha^2}{2}(\nabla \cdot w)^2 \, dx$$

- Dual min problem

$$-\min_{|w|\leq 1} \int_\Omega \alpha z(\nabla \cdot w) + \frac{\alpha^2}{2}(\nabla \cdot w)^2 \, dx$$

- Standard form

$$\boxed{\min_{|w|\leq 1} \int_\Omega z(\nabla \cdot w) + \frac{\alpha^2}{2}(\nabla \cdot w)^2 \, dx}$$

Dual Total Variation Regularization

$$\boxed{\min_{|w|\le 1} \int_\Omega z(\nabla\cdot w) + \frac{\alpha^2}{2}(\nabla\cdot w)^2\,dx}$$

- **Advantages**
  - Quadratic objective function in $(\nabla\cdot w)$
  - No need for perturbation parameter β
  - $u = z + \alpha(\nabla\cdot w)$
  - w unique at edges
  - w not unique at flat regions but no information lost
- **Disadvantages**
  - Constrained optimization problem in w
  - One constraint per pixel

Outline

1. Edges in image processing
2. Insufficiency of classical tools
3. Total Variation (TV) model of Rudin-Osher-Fatemi
   (a) Variational methods
      i. [BHKU02] Short course on variational methods
      ii. Total variation (TV) regularization
4. Computational challenges of primal TV
5. Dual TV method
   (a) Derivation of problem
   (b) **Previous and current work**
6. Acknowledgments

Previous Work on Dual TV [CGM99]

$$\boxed{-\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)+u-z=0} \tag{2}$$

- Difficulty is the linearization of the highly nonlinear term $-\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)$
- Introduce

$$w=\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}$$

- Replace (2) by equivalent system of nonlinear PDEs:

$$-\alpha\nabla\cdot w+u-z=0$$
$$w\sqrt{|\nabla u|^2+\beta}-\nabla u=0$$

- Linearize this (u,w) system by Newton's method.

Previous Work on Dual TV [Cha04]

$$\boxed{\min_{|w|\leq 1} \int_\Omega z(\nabla \cdot w) + \frac{\alpha^2}{2}(\nabla \cdot w)^2 \, dx}$$

$$\boxed{u = z + \alpha(\nabla \cdot w)}$$

- For an $N \times N$ image

$$-\alpha(\nabla \cdot w) = \min\{\|\alpha \text{div } w - z\|^2 : |w_{ij}|^2 - 1 \leq 0 \, \forall i, j = 1, \ldots, N\}$$

- Karush-Kuhn-Tucker conditions provide explicit expression for Lagrange multiplier $\lambda_{ij} \geq 0$.

$$\lambda_{ij} = |(\nabla(\alpha \text{div } w - z))_{ij}|$$

- Semi-implicit gradient descent algorithm.

$$w_{ij}^{n+1} = w_{ij}^{n} + \tau(\left(\nabla\left(\text{div } w^n - \frac{z}{\alpha}\right)\right)_{ij} - \left|\left(\nabla\left(\text{div } w^n - \frac{z}{\alpha}\right)\right)_{ij}\right| w_{ij}^{n+1}) \quad (3)$$

- Converges for $\tau \leq \frac{1}{8}$.

Current Work on Dual TV
C.–Chan–Vandenberghe (in progress)

1D:

$$\boxed{\min_{|w|\le 1} \int_\Omega z\frac{dw}{dx} + \frac{\alpha}{2}\left(\frac{dw}{dx}\right)^2 \, dx}$$

Nonlinear Gauss-Seidel algorithm with projection.

2D:

$$\boxed{\min_{\|\mathbf{w}\|_\infty \le 1} \int_\Omega z(\nabla\cdot\mathbf{w}) + \frac{\alpha}{2}(\nabla\cdot\mathbf{w})^2 \, d\mathbf{x}}$$

Linear algebra structure of [CGM99] applied to Newton equations of interior-point method.

Thank you!

References

[AK02] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer, 2002.

[BHKU02] A. Ben Hamza, Hamid Krim, and Gozde B. Unal. Unifying probabilistic and variational estimation. *IEEE Signal Processing Magazine*, pages 37–47, 2002.

[BT89] Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[Car01] Jamylle Laurice Carter. *Dual Methods for Total*

Variation-Based Image Restoration. PhD thesis, University of California, Los Angeles, 2001.

[CGM99] Tony F. Chan, Gene H. Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.

[Cha04] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[CP04] P.L. Combettes and J.-C. Pesquet. Image restoration subject to a total variation constraint. *IEEE Transactions on Image Processing*, 13(9):1213–1222, September 2004.

[DV97] David C. Dobson and Curtis R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM*

Journal on Numerical Analysis, 34(5):1779–1791, 1997.

[Giu84] Enrico Giusti. *Minimal surfaces and functions of bounded variation*. Birkhäuser, 1984.

[GS00] L. Grippo and M. Sciandrone. On the Convergence of the Block Nonlinear Gauss-Seidel Method Under Convex Constraints. *Operations Research Letters*, 26(3):127–136, April 2000.

[Lan70] Cornelius Lanczos. *The Variational Principles of Mechanics*. University of Toronto, fourth edition, 1970.

[Mal02] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Transactions on Image Processing*, 11(12):1450–1456, December 2002.

[OSV03] Stanley Osher, Andres Sole, and Luminita Vese. Image decomposition and restoration using total variation minimization and the h^{-1} norm. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 1(3):349–370, 2003.

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[ROF92] Leonid I. Rudin, Stanley J. Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[Wit83] A.P. Witkin. Scale space filtering. In A. Bundy, editor, *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 1019–1023, 1983.