

Lecture IX. Variations in Sparse Signal Reconstruction

Hojin Kim

Department of Radiation Oncology, Yonsei Cancer Center, Severance Hospital, Yonsei University College of Medicine





YCC PHYSICS

In Lecture VII

- Review of Iterative Methods
 - Features of under-determined problem
 - Sparse signal recovery with L₁-norm minimization
 - Basic algorithm for L₁-norm minimization with proximity operator

Today, Let's take a look at further details about
 Variations in sparse signal recovery (a.k.a. Transform Sparsity)



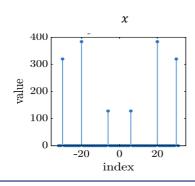
Transform Sparsity



YCC PHYSICS

Real-World Application

• In reality: Image / Fluence-map has many non-zero elements







32412 non-zero / 65536 (49.5%)

minimize
$$f(x) = ||x||_1 = \sum_{i=1}^{n} |x_i|$$

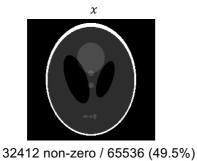
subject to $Ax = b$

$$\Rightarrow \min_{x} \min f(x) = \|Ax - b\|_{2}^{2} + \lambda \|x\|_{1}$$



Real-World Application -- Transform Sparsity

• Transform sparsity: Making something sparse by transformation





Transform (D)

Total-Variation (TV) $Dx = \mid x_{u-1,v} - x_{u,v} \mid + \mid x_{u,v-1} - x_{u,v} \mid$

2184 non-zero / 65536 (3.3%)

• Minimizing transform sparsity (Total-variation, Wavelet, Discrete cosine, etc.)

minimize
$$f(x) = ||Ax - b||_2^2 + \lambda ||x||_1$$
 \Rightarrow minimize $f(x) = ||Ax - b||_2^2 + \lambda ||Dx||_1$



YCC PHYSICS

Limit of Basic L₁-minimization Algorithm

Soft-thresholding operator

minimize
$$f(x) = ||Ax - b||_2^2 + \lambda ||x||_1$$

= $g(x) + h(x)$ \Rightarrow $g(x) = ||Ax - b||_2^2$
 $h(x) = \lambda ||x||_1$

* Proximity operator h(x) deals with very simple L₁-norm operator

$$prox_h(z,t) = \underset{x}{\operatorname{argmin}} \left[\frac{1}{2t} \|x - z\|_{2}^{2} + \|x\|_{1} \right] = SoftThreshold(z,t)$$

→ Not sufficient for the real-world applications



Algorithms for Transform Sparsity

Alternating Direction Method of Multipliers (ADMM) Chambolle-Pock Algorithm



YCC PHYSICS

Publications for ADMM (Split-Bregman Iterative Method)

SIAM J. IMAGING SCIENCES Vol. 2, No. 2, pp. 323-343

© 2009 Society for Industrial and Applied Mathematics

The Split Bregman Method for L1-Regularized Problems*

Tom Goldstein[†] and Stanley Osher[†]

Abstract. The class of L1-regularized optimization problems has received much attention recently because of the introduction of "compressed sensing," which allows images and signals to be reconstructed from small amounts of data. Despite this recent attention, many L1-regularized problems still remain difficult to solve, or require techniques that are very problem-specific. In this paper, we show that Bregman iteration can be used to solve a wide variety of constrained optimization problems. Using this technique, we propose a "split Bregman" method, which can solve a very broad class of L1-regularized problems. We apply this technique to the Rudin-Osher-Fatemi functional for image denoising and to a compressed sensing problem that arises in magnetic resonance imaging.

 $\textbf{Key words.} \ \ \text{constrained optimization, L1-regularization, compressed sensing, total variation denoising}$

AMS subject classification. 65K05

DOI. 10.1137/080725891

Foundations and Trends[®] in Machine Learning Vol. 3, No. 1 (2010) 1–122 © 2011 S. Boyd, N. Parikh, E. Chu, B. Peleato



Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers

Stephen Boyd¹, Neal Parikh², Eric Chu³ Borja Peleato⁴ and Jonathan Eckstein⁵

- ¹ Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA, boyd@stanford.edu
- 94309, U.S.A., objects stanford.edu Chiversity, Stanford, CA 94305, USA, npparikh@cs.stanford.edu

 Blectrical Engineering Department, Stanford University, Stanford, CA
- 94305, USA, cchu508@stanford.edu

 Electrical Engineering Department, Stanford University, Stanford, CA
 94305, USA, peleato@stanford.edu
- 5 Management Science and Information Systems Department and RUTCOR, Rutgers University, Piscataway, NJ 08854, USA, jeckstei@rci.rutgers.edu

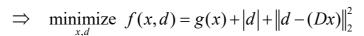
YCC PHYSICS

Alternating Direction Method of Multiplier (ADMM) a.k.a Split-Bregman

Concept of ADMM

$$\underset{x}{\text{minimize}} \ f(x) = g(x) + h(x) \quad \Rightarrow \quad \begin{array}{c} g(x) \colon \text{Convex, Differentiable} \\ h(x) = \|Dx\|_1 \colon \text{Convex, Non-differentiable} \end{array}$$

$$h(x) = ||(Dx)||_1 \implies d \triangleq (Dx)$$



$$\Rightarrow \text{ minimize } f(x,d,q) = g(x) + \left| d \right| + \frac{\lambda}{2} \left\| d - (Dx) - q \right\|_{2}^{2}$$

YONSEI UNIVERSITY

YCC PHYSICS

Pseudo-Code of ADMM

minimize
$$f(x,d,q) = g(x) + |d| + \frac{\lambda}{2} ||d - (Dx) - q||_2^2$$

Step 1. Update *x*: minimize
$$g(x) + \frac{\lambda}{2} ||d - (Dx) - q||_2^2$$

Step 2. Update
$$d$$
: minimize $|d| + \frac{\lambda}{2} ||d - (Dx) - q||_2^2 \implies d_{k+1} = SoftThreshold\left((Dx_{k+1}) + q_k, \frac{1}{\lambda}\right)$

Step 3. Update
$$q$$
: minimize $\frac{\lambda}{2} \| d - (Dx) - q \|_2^2$ \Rightarrow $q_{k+1} = q_k + (Dx_{k+1}) - d_{k+1}$



Example of ADMM -- Image Denoising

· Denoising with TV min.

Step 1. Update
$$x$$
: minimize $\frac{\mu}{2} \|x - f\|_2^2 + \frac{\lambda}{2} \|d_h - D_h x - q_h\|_2^2 + \frac{\lambda}{2} \|d_v - D_v x - q_v\|_2^2$

$$x_{i,j}^{k+1} = \frac{\mu f}{\mu + 4\lambda} + \frac{\lambda}{\mu + 4\lambda} \left[x_{i+1,j}^{k+1} + x_{i-1,j}^{k+1} + x_{i,j+1}^{k+1} + x_{i,j-1}^{k+1} + D_h^T (d_{h,(i,j)}^k - q_{h,(i,j)}^k) + D_v^T (d_{v,(i,j)}^k - q_{v,(i,j)}^k) \right]$$

$$x_{i,j}^{k+1} = \frac{\mu f}{\mu + 4\lambda} + \frac{\lambda}{\mu + 4\lambda} \left[x_{i+1,j}^{k+1} + x_{i-1,j}^{k+1} + x_{i,j+1}^{k+1} + x_{i,j-1}^{k+1} + x_{i,j-1}^{k+1}$$



YCC PHYSICS

Example of ADMM -- Image Denoising

· Denoising with TV min.

minimize
$$f(x) = g(x) + h_h(x) + h_v(x) = \frac{\mu}{2} ||x - f||_2^2 + ||D_h x||_1 + ||D_v x||_1$$

$$\Rightarrow \quad \underset{x, d_h, d_v, b_h, b_v}{\text{minimize}} \quad \frac{\mu}{2} ||x - f||_2^2 + |d_h| + |d_v| + \frac{\lambda}{2} ||d_h - D_h x - q_h||_2^2 + \frac{\lambda}{2} ||d_v - D_v x - q_v||_2^2$$

Step 2. Update d_u , d_v :

Step 3. Update
$$q_h, q_v$$
: minimize $\frac{\lambda}{2} \| d_h - D_h x - q_h \|_2^2 \implies q_h^{k+1} = q_h^k + (D_h x^{k+1} - d_h^{k+1})$
minimize $\frac{\lambda}{2} \| d_v - D_v x - q_v \|_2^2 \implies q_v^{k+1} = q_v^k + (D_v x^{k+1} - d_v^{k+1})$



Example of ADMM -- Compressed Sensing

Total-variation Minimization

minimize
$$f(x) = g(x) + h_h(x) + h_v(x) = \frac{\mu}{2} ||Ax - b||_2^2 + ||D_h x||_1 + ||D_v x||_1$$

where $|Dx(i,j)| = |x(i-1,j) - x(i,j)| + |x(i,j-1) - x(i,j)| = |D_h x(i,j)| + |D_v x(i,j)|$

$$\Rightarrow \underset{x,d_h,d_v,b_h,b_v}{\text{minimize}} \frac{\mu}{2} ||Ax - b||_2^2 + |d_h| + |d_v| + \frac{\lambda}{2} ||d_h - D_h x - q_h||_2^2 + \frac{\lambda}{2} ||d_v - D_v x - q_v||_2^2$$

Step 1. Update
$$x$$
: minimize $\frac{\mu}{2} \|Ax - b\|_2^2 + \frac{\lambda}{2} \|d_h - D_h x - q_h\|_2^2 + \frac{\lambda}{2} \|d_v - D_v x - q_v\|_2^2$

$$x_{i,j}^{k+1} = \frac{\mu A^T b}{\mu A^T A + 4\lambda} + \frac{\lambda}{\mu A^T A + 4\lambda} [x_{i+1,j}^{k+1} + x_{i-1,j}^{k+1} + x_{i,j+1}^{k+1} + x_{i,j+1}^{k+1} + x_{i,j-1}^{k+1} + d_{h,(i-1,j)}^k - d_{h,(i,j)}^k + d_{v,(i,j-1)}^k - d_{v,(i,j)}^k - q_{h,(i-1,j)}^k + q_{h,(i,j)}^k - q_{v,(i,j-1)}^k + q_{v,(i,j)}^k]$$

YONSEI

YCC PHYSICS

Publication for Chambolle-Pock Algorithm

A first-order primal-dual algorithm for convex problems with applications to imaging

> Antonin Chambolle, and Thomas Pock[†] June 9, 2010

Abstract

In this paper we study a first-order primal-dual algorithm for convex optimization problems with known saddle-point structure. We prove convergence to a saddle-point with rate O(1/N) in finite dimensions, which is optimal for the complete class of non-smooth problems we are considering in this paper. We further show accelerations of the proposed algorithm to yield optimal rates on easier problems. In particular we show that we can achieve $O(1/N^2)$ convergence on problems, where the primal or the dual objective is uniformly convex, and we can show linear convergence, i.e. $O(1/e^N)$ on problems where both are uniformly convex. The wide applicability of the proposed algorithm is demonstrated on several imaging problems such as image denoising, image deconvolution, image inpainting, motion estimation and image segmentation.



Chambolle-Pock Algorithm

Pseudo-code of Chambolle-Pock algorithm

$$y_{l}^{n+1} = prox_{\sigma F^{*}}(y_{l}^{n} + \sigma \cdot K\tilde{x}^{n})$$

$$x^{n+1} = prox_{\tau G}(x^{n} - \tau \sum_{l} K^{T} y_{l}^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^{n})$$

NOTE:

*Reminding 'Proximity Operator': $prox_h(z,t) = \underset{x}{\operatorname{argmin}} \left[\frac{1}{2t} \|x - z\|_2^2 + h(x) \right]$

Dual norm (denoted by F)

Assuming that F is L_p -norm, and dual norm of F (F*) is L_q -norm: $\frac{1}{p} + \frac{1}{q} = 1$

*Hyper-parameters: $\sigma \cdot \tau \cdot L^2 \le 1$



YCC PHYSICS

Chambolle-Pock Algorithm

• Pseudo-code of Chambolle-Pock algorithm

$$y_l^{n+1} = prox_{\sigma F^*}(y_l^n + \sigma \cdot K\tilde{x}^n)$$

$$x^{n+1} = prox_{\tau G}(x^n - \tau \sum_l K^T y_l^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$$

Acceleration

$$\sigma_{0} \cdot \tau_{0} \cdot L^{2} \leq 1 \qquad \qquad y_{l}^{n+1} = prox_{\sigma F^{*}}(y_{l}^{n} + \sigma_{n} \cdot K\tilde{x}^{n})$$

$$\theta_{n} = \frac{1}{\sqrt{1+2\mu}}, \tau_{n+1} = \theta_{n}\tau_{n}, \sigma_{n+1} = \frac{\sigma_{n}}{\theta_{n}} \qquad \Rightarrow \qquad x^{n+1} = prox_{\tau G}(x^{n} - \tau_{n}\sum_{l}K^{T}y_{l}^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta_{n}(x^{n+1} - x^{n})$$



Example of Chambolle-Pock Algorithm

$$y_l^{n+1} = prox_{\sigma F^*}(y_l^n + \sigma \cdot K\tilde{x}^n)$$

$$x^{n+1} = prox_{\tau G}(x^n - \tau \sum_{l} K^T y_l^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$$

Applying Chambolle-Pock algorithm to Total-Variation Minimization

minimize
$$F(Kx) + G(x) = \frac{\mu}{2} ||Ax - b||_2^2 + ||D_h x||_1 + ||D_v x||_1 (x \ge 0)$$

$$\Rightarrow F(Kx) = F\left(\begin{bmatrix} A \\ D_h \\ D_v \end{bmatrix} x\right) = f_1(Ax) + f_2(D_h x) + f_3(D_v x), \quad G(x) = [x]_+$$

$$\Rightarrow f_1(y_1) = \frac{\mu}{2} ||y_1 - b||_2^2 \quad (y_1 = Ax), \quad f_2(y_2) = ||y_2||_1 \quad (y_2 = D_h x), \quad f_3(y_3) = ||y_3||_1 \quad (y_3 = D_v x)$$

YONSEI UNIVERSITY

YCC PHYSICS

Example of Chambolle-Pock Algorithm

$$y_l^{n+1} = prox_{\sigma F^*}(y_l^n + \sigma \cdot K\tilde{x}^n)$$
minimize $F(Kx) + G(x)$ \Rightarrow
$$x^{n+1} = prox_{\tau G}(x^n - \tau \sum_l K^T y_l^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$$

• Applying Chambolle-Pock algorithm to Total-Variation Minimization

$$\min_{x} |F(Kx) - G(x)| = \frac{1}{2} ||Ax - b||_{2}^{2} + \lambda ||D_{h}x||_{1} + \lambda ||D_{v}x||_{1} (x \ge 0)$$

$$\Rightarrow |F(Kx) - F(A ||D_{h}||x||_{1}) = f_{1}(Ax) + f_{2}(D_{h}x) + f_{3}(D_{v}x), \quad G(x) = [x]_{+}$$

$$\Rightarrow |f_{1}(y_{1}) - \frac{1}{2} ||y_{1} - b||_{2}^{2} (y_{1} = Ax), \quad f_{2}(y_{2}) = \lambda ||y_{2}||_{1} (y_{2} = D_{h}x), \quad f_{3}(y_{3}) = \lambda ||y_{3}||_{1} (y_{3} = D_{v}x)$$



Example of Chambolle-Pock Algorithm (Cont'd)

$$F(Kx) = f_{1}(Ax) + f_{2}(D_{h}x) + f_{3}(D_{v}x), G(x) = [x]_{+}$$

$$f_{1}(y_{1}) = \frac{\mu}{2} \|y_{1} - b\|_{2}^{2} (y_{1} = Ax),$$

$$f_{2}(y_{2}) = \|y_{2}\|_{1} (y_{2} = D_{h}x),$$

$$f_{3}(y_{3}) = \|y_{3}\|_{1} (y_{3} = D_{v}x)$$

$$\tilde{x}^{n+1} = prox_{\sigma F^{*}}(y_{l}^{n} + \sigma \cdot K\tilde{x}^{n})$$

$$x^{n+1} = prox_{\tau G}(x^{n} - \tau \sum_{l} K^{T} y_{l}^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^{n})$$

Step 1. Update y's:

$$y_{1}^{n+1} = prox_{\sigma f_{1}^{*}}(y_{1}^{n} + \sigma \cdot A\tilde{x}^{n}) = \frac{(y_{1}^{n} + \sigma \cdot A\tilde{x}^{n}) + \sigma \mu b}{1 + \sigma \mu}$$

$$y_{2}^{n+1} = prox_{\sigma f_{2}^{*}}(y_{2}^{n} + \sigma \cdot D_{h}\tilde{x}^{n}) = Truncate(y_{2}^{n} + \sigma \cdot D_{h}\tilde{x}^{n}, \sigma)$$

$$= \begin{cases} y_{2}^{n} + \sigma \cdot D_{h}\tilde{x}^{n} & (|y_{2}^{n} + \sigma \cdot D_{h}\tilde{x}^{n}| < \sigma) \\ \sigma & (|y_{2}^{n} + \sigma \cdot D_{h}\tilde{x}^{n}| < \sigma) \end{cases}$$

YON SEI UNIVERSITY

YCC PHYSICS

Example of Chambolle-Pock Algorithm (Cont'd)

$$F(Kx) = f_{1}(Ax) + f_{2}(D_{h}x) + f_{3}(D_{v}x), G(x) = [x]_{+}$$

$$f_{1}(y_{1}) = \frac{\mu}{2} \|y_{1} - b\|_{2}^{2} (y_{1} = Ax),$$

$$f_{2}(y_{2}) = \|y_{2}\|_{1} (y_{2} = D_{h}x),$$

$$f_{3}(y_{3}) = \|y_{3}\|_{1} (y_{3} = D_{v}x)$$

$$\tilde{x}^{n+1} = prox_{\tau G}(x^{n} - \tau \sum_{l} K^{T} y_{l}^{n+1})$$

$$\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^{n})$$

Step 2. Update
$$x$$
:
$$x^{n+1} = prox_{\tau G}(x^n - \tau \sum_{l} K^T y_l^{n+1})$$
$$= \left[x^n - \tau (A^T y_1^{n+1} + D_h^T y_2^{n+1} + D_v^T y_3^{n+1}) \right]_{\perp}$$

Step 3. Update
$$\tilde{x}$$
: $\tilde{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$ (Acceleration)

Beyond Total-Variation (TV) Transform

YONSEI

I. BACKGROUND

Replacing TV-transform

Wavelet transform:

minimize
$$f(x) = ||Ax - p||_2^2 + \sum_{l=1}^{\infty} \lambda_l ||W_l x||_1$$



Non-local TV transform:

minimize
$$f(x) = ||Ax - p||_2^2 + \lambda ||\nabla_{NLTV}x||_k \quad (k \le 1)$$

 $||\nabla_{NLTV}x||_1 = \sum_i \sqrt{\sum_{j \in \Omega} (u_j - u_i)^2 w_{ij}}, w_{ij} = \exp\left(-\sum_{k=-a}^a G(k) \cdot \left|u_{j+k} - u_{i+k}\right|^2 / 2h_0^2\right)$

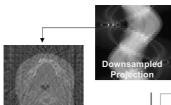
Block-Matching & 3D filtering:

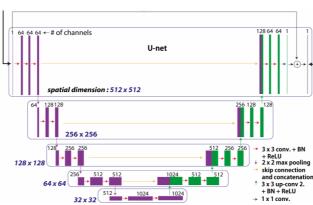


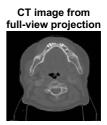
I. BACKGROUND

Replacing TV-transform (Cont'd)

Deep learning







23

YONSEI UNIVERSITY

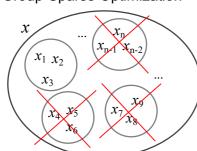
YCC PHYSICS

Notion of Group-Sparse Optimization

Sparse Signal Optimization

Lk-norm:	$ x _k = (x_1 ^k + x_2 ^k + x_3 ^k + + x_n ^k)^{\frac{1}{k}}$
L ₂ -norm:	$ x _2 = \sqrt{ x_1 ^2 + x_2 ^2 + x_3 ^2 + \dots + x_n ^2}$
L₁-norm:	$ x _1 = x_1 + x_2 + x_3 + + x_n $
L ₀ -norm:	$ x _0 = \sum_j 1 \{x_j \neq 0\}$

• Group-Sparse Optimization



- 'Group is sparse': Entire x is divided into sub-groups
 - → Very few groups contribute to the optimization

$$\min_{x} \| Ax - d \|_{2}^{2} + \lambda \| x \|_{2}^{p} (p \le 1)$$

$$L_{2,p}\text{-norm}$$

- For treatment plans? → Beam Angle, Energy Layer!!!



Summary

- Understanding **Necessity of Trasform Sparsity** in real-world applications
- → Transforming 'Dense signal' into 'Sparse signal', so that L1-min. is applicable
- Important algorithms to solve transform sparsity: ADMM, C.P.
- → Both have potential for the various applications
- Besides TV transform, the other options would be available





