Primal Dual Algorithms for Convex Optimization in Imaging Science

Ernie Esser

UC Irvine

MAIPCV2011 Winter School, Tohoku University

11-25-2011

Variational Models (1)

A powerful modeling tool in image processing and computer vision is to construct a functional such that its minimizer(s) are good solutions to the problem of interest.

A general form:

$$\min_{u \in \mathbb{R}^m} J(u) \qquad \text{such that} \qquad Au = b$$

Common difficulties:

- Image and video problems are large with many variables
- Models often involve nonsmooth objective functions

Variational Models (2)

$$\min_{u \in \mathbb{R}^m} J(u) \qquad \text{such that} \qquad Au = b$$

Common advantages:

- Separable structure:

$$J(u) = \sum_{i=1}^{n} J_i(u_i) \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} \qquad N \geq n$$

  with simple functions J_i

- **Convex:** $J((1-s)u + sv) \leq (1-s)J(u) + sJ(v)$ for all u, v $s \in (0,1)$

Goal: Study and develop practical algorithms for minimizing large, nonsmooth convex functions with separable structure.

Convexity

In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.

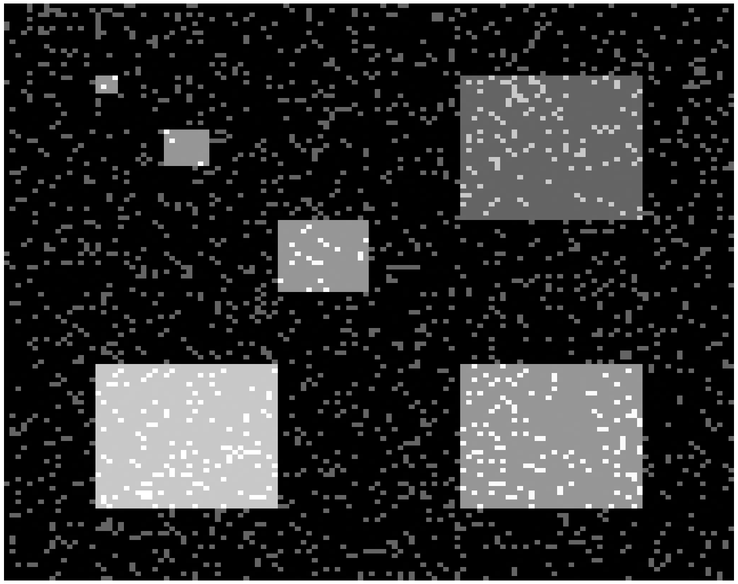
- R. T. Rockafellar

Let us first look at some illustrative examples of large, nonlinear, nonsmooth convex models in image processing which can be effectively solved using the primal-dual algorithms which will be discussed later:

TVL1 Denoising (1)

$$\min_u \|u\|_{TV} + \lambda \|u - f\|_1$$

$\|u\|_{TV}$ is a discretization of the total variation of $u \approx \int |\nabla u|$ (details later)

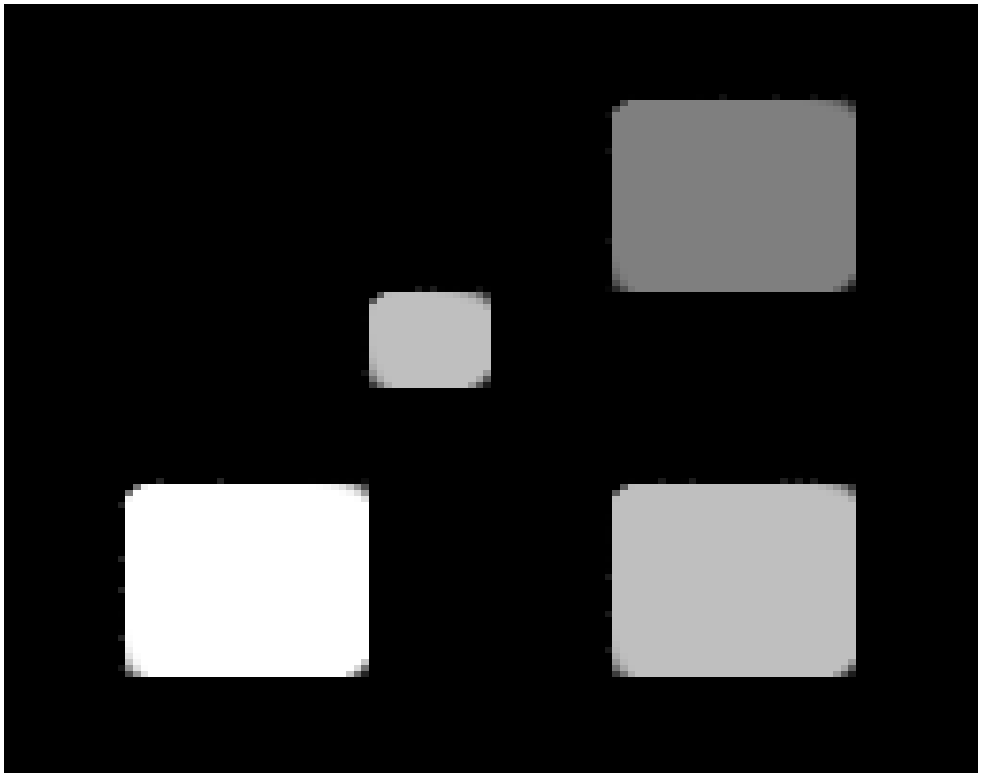


Noisy Image

L. Rudin, S. Osher, and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, 60, 1992, pp. 259-268.

T.F. Chan and S. Esedoglu, *Aspects of Total Variation Regularized L^1 Function Approximation*, 2004.

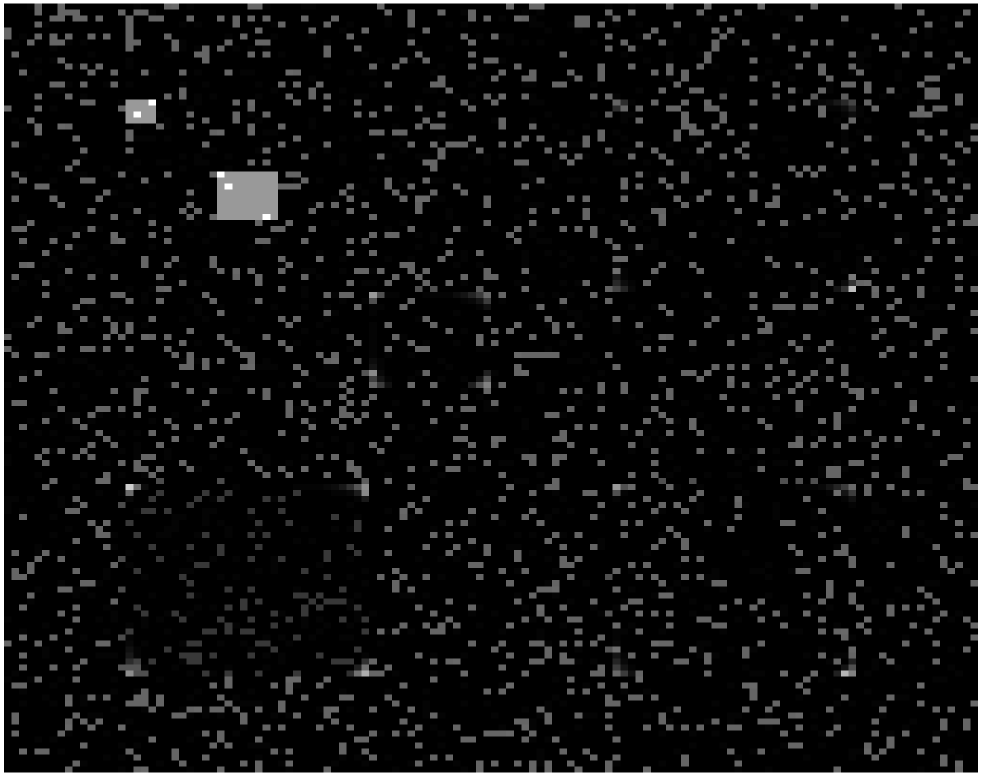
TVL1 Denoising (2)



Recovered Image

TVL1 Denoising (3)

f−u



Sparse Error

Constrained TVL2 Deblurring

$$\min_u \|u\|_{TV} \qquad \text{such that} \qquad \|k * u - f\|_2 \leq \epsilon$$

Original, blurry/noisy and image recovered from 300 iterations

E. ESSER, X. ZHANG, AND T. F. CHAN, *A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science*, SIAM J. Imaging Sci. Volume 3, Issue 4, 2010.

Sparse/Low Rank Decomposition (1)

$$\min_{u,e} \|u\|_* + \lambda \|e\|_1 \qquad \text{such that} \qquad f = u + e$$

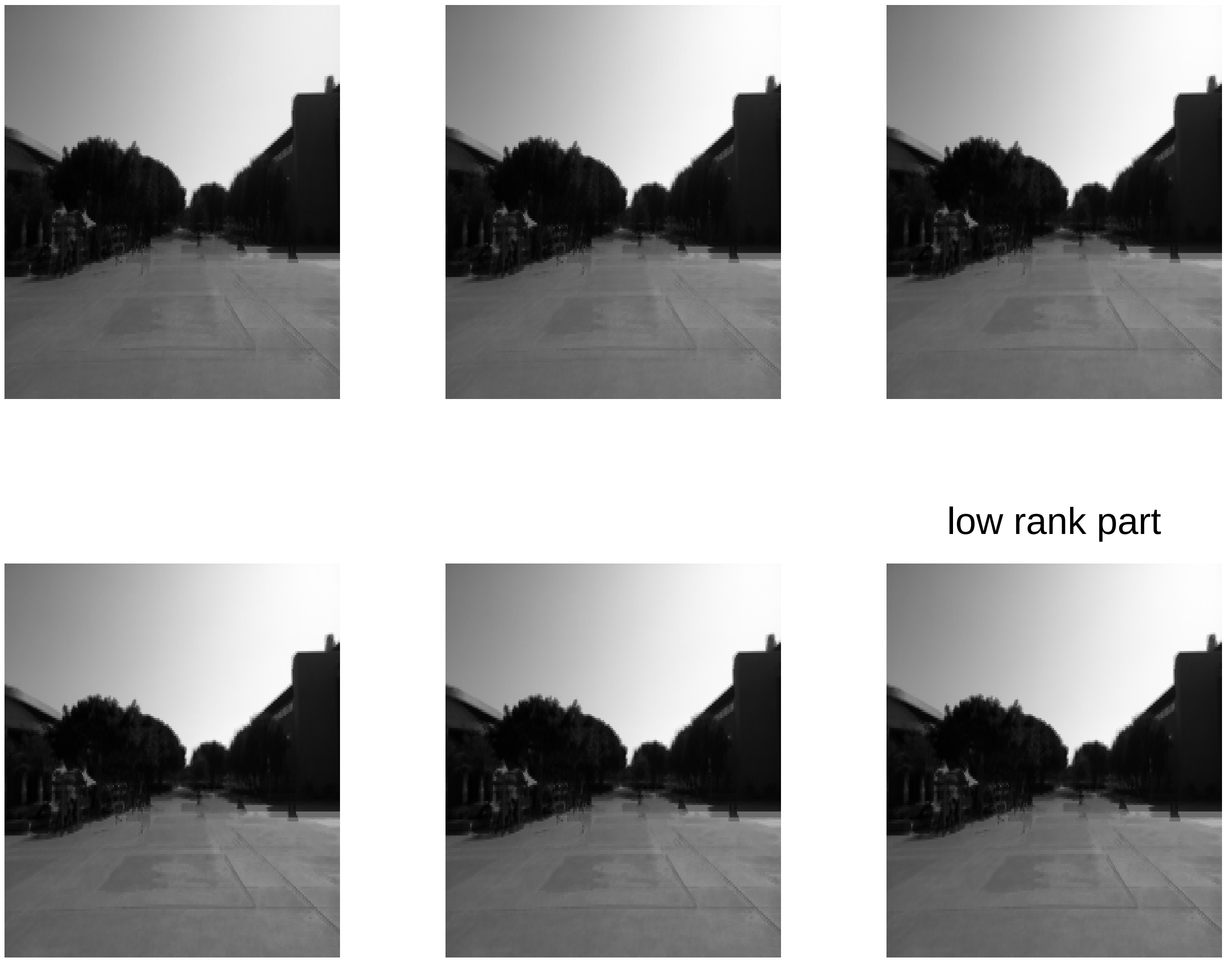
Here, $\|u\|_*$ denotes the nuclear norm, which is the sum of the singular values of u.

original movie

Original Video

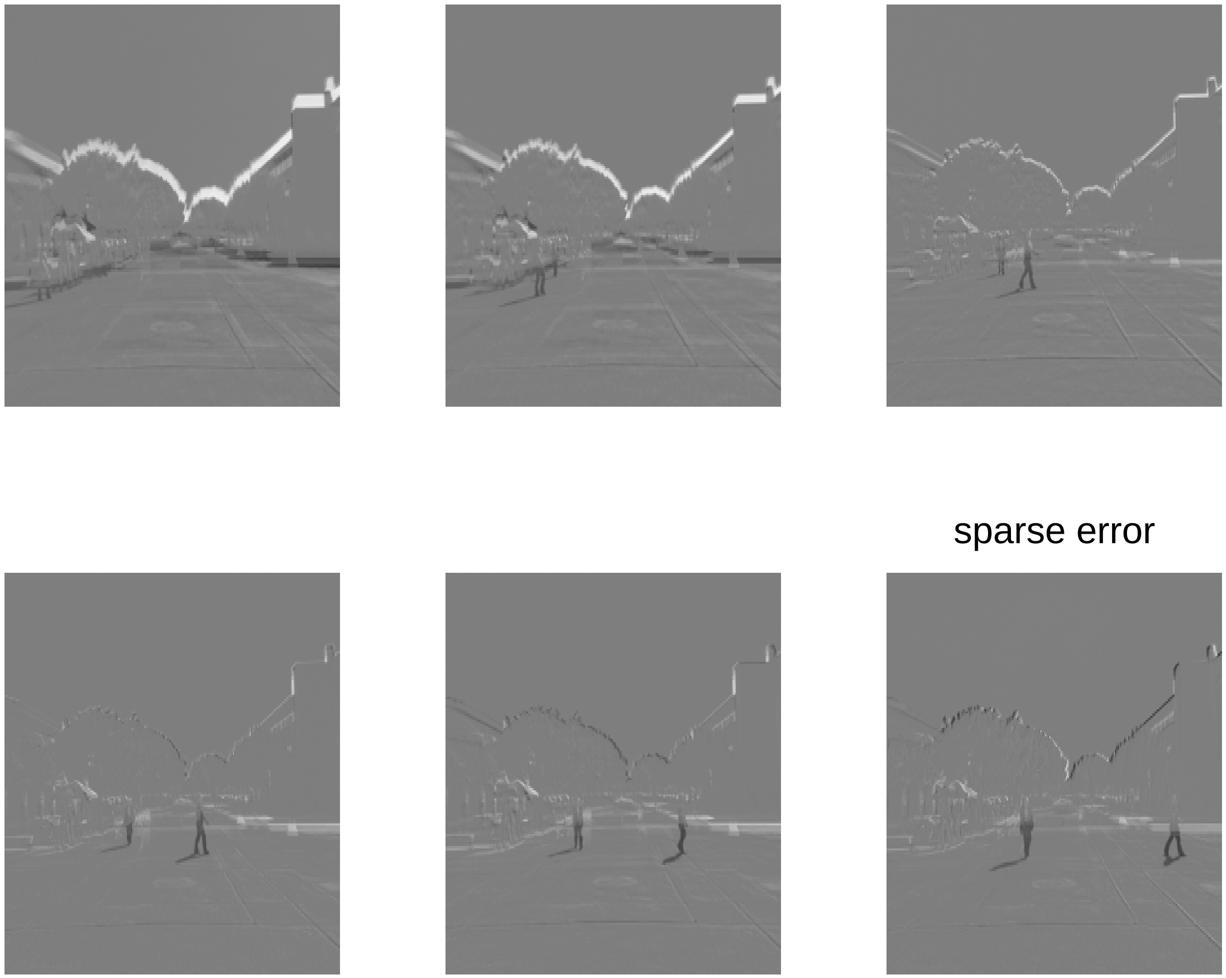
E. Candes, X. Li, Y. Ma and J. Wright, *Robust Principal Component Analysis*, 2009.

Sparse/Low Rank Decomposition (2)



Recovered Background u

Sparse/Low Rank Decomposition (3)



Sparse Error e

Multiphase Segmentation (1)

Many problems deal with the normalization constraint $c \in C$, where

$$C = \{c = (c_1, ..., c_W) : c_w \in \mathbb{R}^M, \sum_{w=1}^{W} c_w = 1, c_w \geq 0\}$$

Example: Convex relaxation of multiphase segmentation

Goal: Segment a given image, $h \in \mathbb{R}^M$, into W regions where the intensities in the w^{th} region are close to given intensities $z_w \in \mathbb{R}$ and the lengths of the boundaries between regions are not too long.

$$\min_{c \in C} \sum_{w=1}^{W} \left(\|c_w\|_{TV} + \frac{\lambda}{2} \langle c_w, (h - z_w)^2 \rangle \right)$$

This is a convex approximation of the related nonconvex functional which additionally requires the labels, c, to only take on the values zero and one.

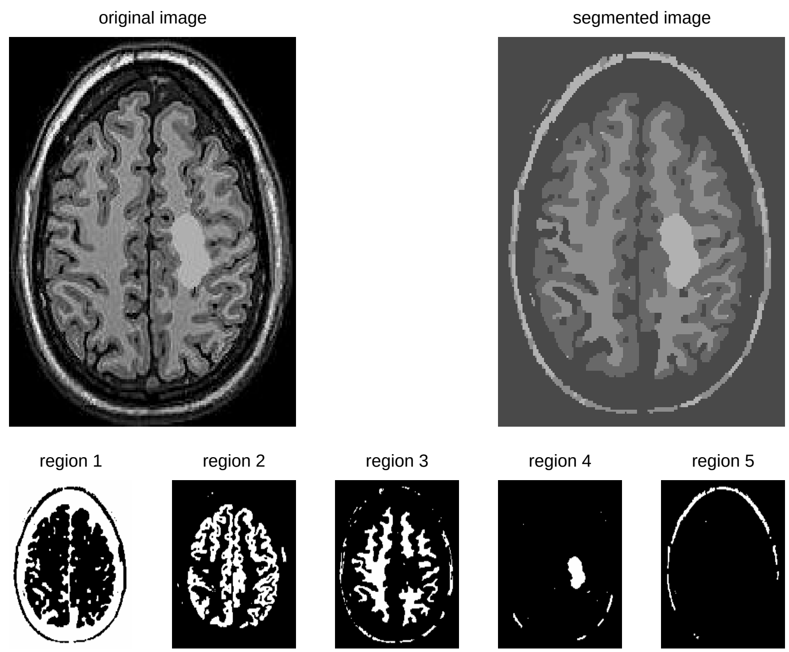
E. BAE, J. YUAN, AND X. TAI, *Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach*, UCLA CAM Report [09-75], 2009.

C. ZACH, D. GALLUP, J.-M. FRAHM, AND M. NIETHAMMER, *Fast global labeling for real-time stereo using multiple plane sweeps*, VMV, 2008.

Multiphase Segmentation (2)

$$\lambda = .0025 \qquad z = \begin{bmatrix} 75 & 105 & 142 & 178 & 180 \end{bmatrix}$$

Threshold c when each $\|c_w^{k+1} - c_w^k\|_\infty < .01$ (150 iterations)



Segmentation of Brain Image Into 5 Regions

Modifications: We can also add μ_w parameters to regularize differently the lengths of the boundaries of each region and alternately update the averages z when they are not known beforehand.

Reductions to Standard Form

All the previous illustrative examples can be rewritten in the form

$$\min_u \sum_i J_i(u_i) \qquad \text{such that} \qquad Au = b \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}$$

Sometimes this requires introducing additional variables and constraints.

Constrained TVL2 Deblurring Example

$$\min_u \|u\|_{TV} \qquad \text{such that} \qquad \|k * u - f\|_2 \leq \epsilon$$

Let D be a discrete gradient and $\|\cdot\|_E$ an l_1-like norm such that $\|Du\|_E$ corrseponds to our discretization of $\|u\|_{TV}$ (details later).

Introduce $w = Du$ and $z = k * u - f$. An equivalent problem is:

$$\min_{u,w,z} \|w\|_E + g_B(z) \qquad \text{s.t.} \qquad w - Du = 0 \;\text{ and }\; z - k * u = -f$$

$$g_B(z) = \begin{cases} 0 & \text{if } \|z\|_2 \leq \epsilon \\ \infty & \text{otherwise} \end{cases}$$ is a convex indicator function.

Convex Models for Nonconvex Problems

Convex optimization is still important for many nonconvex problems:

- Convex relaxation
  Basis Pursuit Example:

$$\min_{Au=b} \|u\|_0 \qquad \to \qquad \min_{Au=b} \|u\|_1$$

- Exact convex relaxation
  - Functional lifting
  - 2 phase segmentation $\qquad c \in \{0,1\} \to c \in [0,1]$
- Convex subproblems for nonconvex problems
  - Alternating minimization for problems like blind deconvolution
  - Global branch and bound methods rely on convex subproblems

M. Burger and M. Hintermller, *Projected Gradient Flows for BV / Level Set Relaxation*, 2005.

T. Goldstein, X. Bresson and S. Osher, *Global Minimization of Markov Random Fields with Applications to Optical Flow*, 2009.

Outline for the Rest of the Talk

- Convex analysis background
- Connections between primal, dual and saddle point problem formulations
- A practically useful class of primal dual methods that are simple, require few assumptions and can take advantage of separable structure
- Algorithm variants that accelerate the convergence rate and generalize applicability
- Applications and implementation details

Some Convex Optimization References

- D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, 1996.
- D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Second Edition. 1999.
- D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*, Prentice Hall, 1989.
- S. Boyd and L. Vandenberghe, *Convex Analysis*, Cambridge University Press, 2006.
- P.L. Combettes, *Proximal Splitting Methods in Signal Processing*, 2011.
- I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, SIAM, Classics in Applied Mathematics, 28, 1999.
- R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- R.T. Rockafellar and R. Wets, *Variational Analysis*, Springer, 1998.
- L. Vandenberghe, *Optimization Methods for Large-Scale Systems*, Course Notes: http://www.ee.ucla.edu/˜vandenbe/ee236c.html

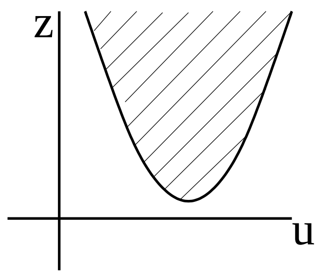
Closed Proper Convex

Assume we are working with closed, proper convex functions of the form $J : \mathbb{R}^n \to (-\infty, \infty]$

Convex: $J((1-s)u + sv) \leq (1-s)J(u) + sJ(v)$ for all u, v $s \in (0, 1)$

Proper: J is not identically equal to ∞

Closed: The epigraph $\operatorname{Epi} J := \{(u, z) : u \in \mathbb{R}^n, z \in \mathbb{R}, z \geq J(u)\}$ is closed.



This is equivalent to lower semicontinuity of J.

Note: We could also define convexity of J in terms of $\operatorname{Epi} J$ being a convex set.

$\operatorname{dom}(J) = \{u \in \mathbb{R}^n : J(u) < \infty\}$ is also the projection of $\operatorname{Epi} J$ onto $\mathbb{R}^n$

Subgradients

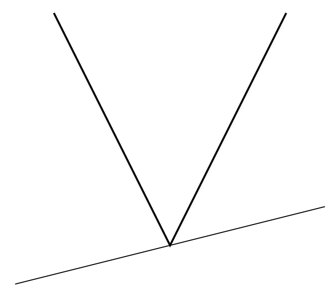
Subgradients of J define non vertical supporting hyperplanes to $\mathrm{Epi}\, J$.

The **subdifferential** $\partial J(u)$ is the set of all subgradients of J at u.

$p \in \partial J(u)$ means $J(v) - J(u) - \langle p, v-u \rangle \geq 0$ for all v

Example: $f(x) = |x|$

$$\partial f(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \\ [-1, 1] & x = 0 \end{cases}$$



Condition for minimizer: $0 \in \partial J(u) \Leftrightarrow J(u) \leq J(v) \qquad \forall v$

Note: If J is differentiable, then $\partial J(u) = \nabla J(u)$

Indicator Function and Normal Cone

Let C be a convex set. Define the indicator function for C by

$$g_C(u) = \begin{cases} 0 & u \in C \\ \infty & \text{otherwise} \end{cases}$$

g_C is convex and its subdifferential is the normal cone $N_C(u)$

$$\begin{aligned} N_C(u) &= \partial g_C(u) \\ &= \{p : -\langle p, v - u\rangle \geq 0 \qquad \forall v \in C\} \\ &= \{p : \langle p, u - v\rangle \geq 0 \qquad \forall v \in C\} \end{aligned}$$

Convex Conjugate

(This is also called the Legendre-Fenchel Transform)

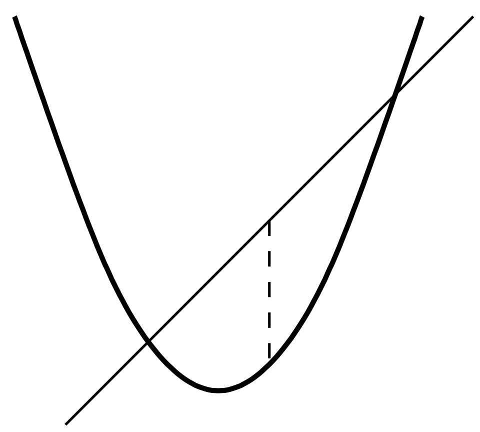
$$J^*(p) = \sup_u \langle u, p \rangle - J(u)$$

We will see that this can be thought of as a dual way of representing a convex function as a pointwise supremum of affine functions.

(If J is differentiable, then $p = \nabla J(u^*)$, where the supremum is attained at u^*, the point where the hyperplane is tangent to J.)

Useful Properties:

- J^* is convex
- $J^{**} = J$
  (still assuming J is closed proper convex)
- $p \in \partial J(u) \Leftrightarrow u \in \partial J^*(p)$



Convexity of Convex Conjugate

$$J^*(p) = \sup_u \langle u, p \rangle - J(u)$$

J^* is convex because it is a $\sup$ of affine functions, namely

$$J^*(p) = \sup_{(u,z) \in \mathrm{Epi}\, J} \langle u, p \rangle - z$$

$J^{**} = J$

Consider the set of all $(p, q), p \in \mathbb{R}^n, q \in \mathbb{R}$ such that

$$J(u) \geq \langle u, p \rangle - q \qquad \forall u$$

Equivalently,

$$q \geq \langle u, p \rangle - J(u) \qquad \forall u$$

$$q \geq \sup_u \langle u, p \rangle - J(u) = J^*(p) \text{ by definition}$$

Since $q \geq J^*(p)$, $\{(p, q) : J(u) \geq \langle u, p \rangle - q \; \forall u\} = \mathrm{Epi}\, J^*$.

$$\begin{aligned} J(u) &= \sup_{(p,q) \in \mathrm{Epi}\, J^*} \langle u, p \rangle - q \text{ by convexity of } J \\ &= \sup_p \langle u, p \rangle - J^*(p) = J^{**}(p) \text{ by definition} \end{aligned}$$

R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

"Inverse" of ∂J

$$p \in \partial J(u)$$

$$J(v) - J(u) - \langle p, v - u \rangle \geq 0 \qquad \forall v \text{ by definition}$$

$$\langle u, p \rangle - J(u) \geq \sup_v \langle v, p \rangle - J(v) = J^*(p)$$

Fenchel Inequality: $\langle u, p \rangle \geq J(u) + J^*(p)$

$$\langle u, p \rangle - J^*(p) \geq J^{**}(u)$$

$$\langle u, p \rangle - J^*(p) \geq \sup_q \langle u, q \rangle - J^*(q)$$

$$J^*(q) - J^*(p) - \langle u, q - p \rangle \geq 0 \qquad \forall q$$

$$u \in \partial J^*(p)$$

Example: Convex Conjugate of Norm

Suppose $J(u) = \|u\|$. Then

$$J^*(p) = \sup_u \langle u, p \rangle - \|u\|$$

$$= \begin{cases} 0 & \text{if } \langle u, p \rangle \leq \|u\| \;\; \forall u \\ \infty & \text{ otherwise} \end{cases}$$

$$= \begin{cases} 0 & \text{if } \sup_{\|u\| \leq 1} \langle u, p \rangle \leq 1 \\ \infty & \text{ otherwise} \end{cases}$$

$$= \begin{cases} 0 & \text{if } \|p\|_* \leq 1 \text{ by dual norm definition} \\ \infty & \text{ otherwise} \end{cases}$$

So the Legendre-Fenchel transform of a norm is the indicator function for the unit ball under the dual norm.

Support Function

Dual norms are related to support functions.

Let C be a convex set.

Recall indicator function $g_C(u) = \begin{cases} 0 & u \in C \\ \infty & \text{otherwise} \end{cases}$

By definition, $g_C^*(u) = \sup_{p \in C} \langle p, u \rangle :=$ support function $\sigma_C(u)$.

From the previous example, if $J(u) = \|u\|$ and $C = \{p : \|p\|_* \leq 1\}$, then

$$\sigma_C(u) = \|u\|_{**} = J^{**}(u) = J(u) = \|u\|$$

General Moreau Decomposition

Let $f \in \mathbb{R}^m$, J a closed proper convex function on $\mathbb{R}^n$, and $A \in \mathbb{R}^{n\times m}$.

$$f = \arg \min_{u\in R^m} J(Au) + \frac{1}{2\alpha} \|u - f\|_2^2 + \alpha A^T \arg \min_{p\in\mathbb{R}^n} J^*(p) + \frac{\alpha}{2} \left\| A^T p - \frac{f}{\alpha} \right\|_2^2$$

Proof: Let p^* be a minimizer of $J^*(p) + \frac{\alpha}{2}\|A^T p - \frac{f}{\alpha}\|_2^2$

Then $0 \in \partial J^*(p^*) + \alpha A(A^T p^* - \frac{f}{\alpha})$

Let $u^* = f - \alpha A^T p^*$, which implies $Au^* \in \partial J^*(p^*)$

Then $p^* \in \partial J(Au^*)$, which implies $A^T p^* \in A^T \partial J(Au^*)$

Thus $0 \in A^T \partial J(Au^*) + \frac{u^* - f}{\alpha}$, which means

$u^* = \arg\min_u J(Au) + \frac{1}{2\alpha}\|u - f\|^2$

J. J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93, 1965.

P. COMBETTES, AND W. WAJS, *Signal Recovery by Proximal Forward-Backward Splitting*, Multiscale Model. Simul., 2006.

Orthogonal Projection

Let C be a convex set. The orthogonal projection of z onto C is

$$\Pi_C(z) = \arg\min_{u \in C} \frac{1}{2}\|u - z\|_2^2 = \arg\min_u g_C(u) + \frac{1}{2}\|u - z\|_2^2$$

Example: A familiar case of the Moreau decomposition is writing a vector as a sum of projections onto orthogonal subspaces

Let L be a subspace of $\mathbb{R}^n$ and g_L the indicator function for L. Then

$$g_L^*(p) = \sup_{u \in L} \langle u, p \rangle = \begin{cases} 0 & p \in L^\perp \\ \infty & \text{otherwise} \end{cases} = g_{L^\perp}(p)$$

So by the Moreau decomposition, we can write a vector $f \in \mathbb{R}^n$ as

$$f = \Pi_L(f) + \Pi_{L^\perp}(f)$$

Soft Thresholding

Minimization of l_1-like norms leads to soft thresholding.

$$S_\alpha(z) = \arg\min_u \|u\|_1 + \frac{1}{2\alpha}\|u - z\|^2$$

By the Moreau decomposition and the fact that the dual norm of $\|\cdot\|_1$ is $\|\cdot\|_\infty$,

$$\begin{aligned} z &= S_\alpha(z) + \alpha \arg\min_{\|p\|_\infty \le 1} \frac{\alpha}{2}\|p - \frac{z}{\alpha}\|^2 \\ &= S_\alpha(z) + \alpha \Pi_{\{p:\|p\|_\infty \le 1\}}(\frac{z}{\alpha}) \\ &= S_\alpha(z) + \Pi_{\{p:\|p\|_\infty \le \alpha\}}(z) \end{aligned}$$

Componentwise,

$$S_\alpha(z)_i = \begin{cases} z_i - \alpha\, \mathrm{sign}(z_i) & |z_i| > \alpha \\ 0 & \text{otherwise} \end{cases}$$

Subdifferential Calculus

When can we use Fermat's rule for simplifying sums of convex functions composed with linear operators?

Example: Consider $\partial[G(u) + H(u)]$ for closed proper convex G, H and $G(u) = J(Au)$

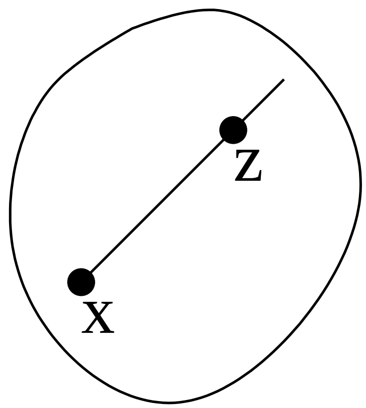
$$\partial[G(u) + H(u)] \supset \partial G(u) + \partial H(u)$$

$$\partial G(u) \supset A^T \partial J(Au)$$

These inclusions are equalities if some technical conditions hold, usually satisfied in practice.

Relative Interior

Let C be a convex set. $z \in \operatorname{ri} C$ iff for every $x \in C$ there exists $\mu > 1$ such that $(1-\mu)x + \mu z \in C$.



Example: Let $D = \{(x, y) \in \mathbb{R}^2 : y = 0,\ x \in [0, 1]\}$

Let g_D be the indicator function for D, this segment of the x-axis.

$\operatorname{dom} g_D = D$

$\operatorname{int}(\operatorname{dom} g_D)$ is empty

$\operatorname{ri} \operatorname{dom} g_D = \{(x, y) : y = 0, x \in (0, 1)\}$

Chain Rule

If $\operatorname{ri}\operatorname{dom} G$ and $\operatorname{ri}\operatorname{dom} H$ have a point in common, then

$$\partial[G(u) + H(u)] = \partial G(u) + \partial H(u)$$

Recall $G(u) = J(Au)$. If the image of A contains a point of $\operatorname{ri}\operatorname{dom} J$, then

$$\partial G(u) = A^T \partial J(Au)$$

These requirements can be weakened whenever G, H or J is polyhedral, (ie: when the epigraph is the intersection of finitely many closed half spaces), in which case we can replace $\operatorname{ri}\operatorname{dom}$ with dom in the above conditions.

R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

Strict/Strong Convexity and Differentiability

Sketch of main properties:

- J is differentiable at u iff J has a unique subgradient at u
  ie: J is differentiable iff ∂J is single valued
- J is differentiable iff J^* is strictly convex
- ∇J is Lipschitz continuous with constant $\frac{1}{\alpha}$ iff J^* is strongly convex with modulus α

The Lipschitz condition means $\|\nabla J(u_1) - \nabla J(u_2)\|_2 \leq \frac{1}{\alpha}\|u_1 - u_2\|_2$

The strong convexity condition means $J^* - \frac{\alpha}{2}\|\cdot\|^2$ is convex

Infimal Convolution

$$(J_1 \Box J_2)(y) = \inf_u J_1(u) + J_2(y - u)$$

Still assume J_1 and J_2 are closed proper convex functions.

If $\mathrm{ri}\,\mathrm{dom}\,J_1^*$ and $\mathrm{ri}\,\mathrm{dom}\,J_2^*$ have a point in common and either J_1 or J_2 is differentiable, then $J_1 \Box J_2$ is differentiable.

Use the fact that

$$J_1 \Box J_2 = (J_1^* + J_2^*)^*$$

Then since either J_1^* or J_2^* is strictly convex, so is the sum, and thus $J_1 \Box J_2$ is differentiable.

Moreau-Yosida Regularization

Moreau envelope of J:

$$E_\alpha(y) = \min_u J(u) + \frac{1}{2\alpha}\|u - y\|^2$$

Let $\bar{u}(y) = \arg\min_u J(u) + \frac{1}{2\alpha}\|u - y\|^2$.

Properties of Moreau-Yosida regularized J:

- $E_\alpha(y)$ is differentiable
- $\nabla E_\alpha(y) = \frac{1}{\alpha}(y - \bar{u}(y))$
- $\nabla E_\alpha(y)$ is Lipschitz continuous with constant $\frac{1}{\alpha}$

Note that $0 \in \partial J(\bar{u}(y)) + \frac{1}{\alpha}(\bar{u}(y) - y)$

So $\nabla E_\alpha(u^*) = 0 \Leftrightarrow 0 \in \partial J(u^*)$.

Proximal Point Method

Compute $u^* = \arg\min_u J(u)$ by iterating

$$u^{k+1} = \arg\min_u J(u) + \frac{1}{2\alpha}\|u - u^k\|^2$$

If a solution exists, $\{u^k\}$ converges to a solution.

The proximal point method is related to gradient descent on E_α,

$$u^{k+1} = u^k - \delta \nabla E_\alpha(u^k) \text{ for } \delta = \alpha$$

Note this gradient scheme converges for $\delta \in (0, 2\alpha)$.

(In practice however, α can depend on k as long as $\liminf_{k\to\infty} > 0$.)

Outline for Algorithm Comparisons

Transition to primal dual methods for linearly constrained models of the form

$$\min_u J(u) \qquad \text{such that} \qquad Au = b$$

- Method of multipliers and Bregman iteration
- Bregman operator splitting and linearized method of multipliers
- Alternating direction method of multipliers (ADMM) and split Bregman
- Split inexact Uzawa and linearized ADMM special cases
- Proximal forward backward splitting
- Primal dual hybrid gradient (PDHG) and modified variants
- Comparisons
- Accelerated and generalized variants

Primal and Dual Problems

Primal Problem: $\min_u J(u)$ such that $Au = b$ $(P0)$

Lagrangian: $L(u, p) = J(u) + \langle p, b - Au \rangle$

p can be thought of as a Lagrange multiplier or a dual variable

Augmented Lagrangian: $L_\delta(u, p) = J(u) + \langle p, b - Au \rangle + \frac{\delta}{2} \|Au - b\|^2$

Dual Functions:

$$q(p) = \inf_u L(u, p) = \langle p, b \rangle - \sup_u \langle A^T p, u \rangle - J(u) = \langle p, b \rangle - J^*(A^T p)$$

$$q_\delta(p) = \inf_u L_\delta(u, p)$$

Dual Problem: $\max_p q(p)$ (Q0)

Assuming (P0) has a solution, the maximums of q and q_δ are attained and equal.

Saddle Point Characterizations

Strong Duality: Assuming a solution u^* to (P0) exists, a solution p^* to (Q0) exists and $J(u^*) = q(p^*)$.

Saddle Point Characterization: u^* solves (P0) and p^* solves (Q0) iff (u^*, p^*) is a saddle point of L

$$L(u^*, p) \leq L(u^*, p^*) \leq L(u, p^*) \qquad \forall u, p$$

Optimality Conditions:

$$Au^* = b$$

$$A^T p^* \in \partial J(u^*) \Leftrightarrow u^* \in \partial J^*(A^T p^*)$$

Method of Multipliers

$$u^{k+1} = \arg\min_u L_\delta(u, p^k) = J(u) + \langle p^k, b - Au \rangle + \frac{\delta}{2}\|Au - b\|^2$$
$$p^{k+1} = p^k + \delta(b - Au^{k+1})$$

Note that (u^{k+1}, p^{k+1}) is a saddle point of $L(u,p) - \frac{1}{2\delta}\|p - p^k\|^2$ because

$$p^{k+1} = \arg\max_p L(u^{k+1}, p) - \frac{1}{2\delta}\|p - p^k\|^2$$
$$u^{k+1} = \arg\min_u L(u, p^{k+1}) - \frac{1}{2\delta}\|p^{k+1} - p^k\|^2$$

For $\delta > 0$, $\{p^k\}$ converges to a solution of (Q0) by analogy to the proximal point method.

Any limit point of $\{u^k\}$ solves the primal problem (P0).

Dual Interpretation

$$\max_p L(u,p) - \frac{1}{2\delta}\|p - p^k\|^2 = L_\delta(u, p^k)$$

Since $q_\delta(p) = \min_u L_\delta(u,p)$,

$$\begin{aligned} q_\delta(p^k) &= \min_u \max_p L(u,p) - \frac{1}{2\delta}\|p - p^k\|^2 \\ &= \max_p \inf_u L(u,p) - \frac{1}{2\delta}\|p - p^k\|^2 \\ &= \max_p q(p) - \frac{1}{2\delta}\|p - p^k\|^2 \end{aligned}$$

The max is attained at p^{k+1}

So $p^{k+1} = \arg\max_p q(p) - \frac{1}{2\delta}\|p - p^k\|^2$.

This is the proximal point method for maximizing q.

Gradient Interpretation

The Lagrange multiplier update can also be understood as a gradient ascent step for maximizing q_δ.

Since $q_\delta(p^k) = \max_p q(p) - \frac{1}{2\delta}\|p - p^k\|^2$ is the Moreau envelope of q,

$$\nabla q_\delta(p^k) = \frac{1}{\delta}(p^{k+1} - p^k) = b - Au^{k+1}$$

Therefore the multiplier update

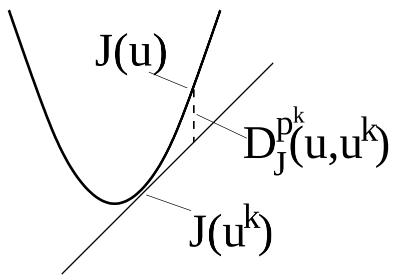
$$p^{k+1} = p^k + \delta(b - Au^{k+1})$$

can be interpreted as the gradient ascent step

$$p^{k+1} = p^k + \delta \nabla q_\delta(p^k)$$

Bregman Iteration

Bregman Distance: $D_J^{p^k}(u,u^k) = J(u) - J(u^k) - \langle p^k, u - u^k\rangle$ where $p^k \in \partial J(u^k)$.



Bregman iteration for solving (P0):

$$u^{k+1} = \arg\min_u D_J^{p^k}(u,u^k) + \frac{\delta}{2}\|Au-b\|^2$$
$$p^{k+1} = p^k + \delta A^T(b - Au^{k+1}) \in \partial J(u^{k+1})$$

Equivalent u^{k+1} update:

$$u^{k+1} = \arg\min_u J(u) - \langle p^k, u\rangle + \frac{\delta}{2}\|Au-b\|^2$$

Initialization: $p^0 = 0$, u^0 arbitrary

S. Osher, M. Burger, D. Goldfarb, J. Xu, *An iterated regularization method for total variation based image restoration*, 2005.

Bregman / Method of Multipliers

Bregman iteration for (P0)

$$u^{k+1} = \arg\min_u J(u) - \langle p^k, u\rangle + \frac{\delta}{2}\|Au - b\|^2$$
$$p^{k+1} = p^k + \delta A^T(b - Au^{k+1}),\ p^0 = 0$$

Equivalent to method of multipliers:

$$u^{k+1} = \arg\min_u J(u) + \langle \lambda^k, b - Au\rangle + \frac{\delta}{2}\|Au - b\|^2$$
$$\lambda^{k+1} = \lambda^k + \delta(b - Au^{k+1}),\ \lambda^0 = 0$$

with $p^k = A^T\lambda^k \qquad \forall k.$

It was from the Bregman interpretation that this was shown to be very useful for l_1 minimization problems

W. Yin, S. Osher, D. Goldfarb and J. Darbon, *Bregman Iterative Algorithms for l_1-Minimization with Applications to Compressed Sensing*, 2007.

Decoupling Variables

Given a step of the method of multipliers algorithm of the form

$$u^{k+1} = \arg\min_u J(u) + \langle p^k, b - Au \rangle + \frac{\delta}{2}\|Au - b\|^2$$

modify the objective functional by adding

$$\frac{1}{2}\left\langle u - u^k, \left(\frac{1}{\alpha} - \delta A^T A\right)(u - u^k)\right\rangle,$$

where α is chosen such that $0 < \alpha < \frac{1}{\delta\|A^T A\|}$.

Modified update is given by

$$u^{k+1} = \arg\min_u J(u) + \langle p^k, b - Au \rangle + \frac{1}{2\alpha}\|u - u^k + \alpha\delta A^T(Au^k - b)\|^2.$$

Bregman Operator Splitting (BOS)

The strategy of linearizing the quadratic penalty can be interpreted from the Bregman perspective or considered as a linearized variant of the method of multipliers.

Full BOS algorithm:

$$u^{k+1} = \arg\min_u J(u) + \frac{1}{2\alpha}\|u - u^k + \alpha\delta A^T(Au^k - b - \frac{p^k}{\delta})\|^2$$

$$p^{k+1} = p^k + \delta(b - Au^{k+1})$$

If $\alpha, \delta > 0$ and $\alpha\delta < \frac{1}{\|A\|^2}$, then limit points of $\{u^k\}$ solve (P0)

Ref: X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*, UCLA CAM Report [09-03] 2009.

General Proximal Point Interpretation

(u^{k+1}, p^{k+1}) as defined by the BOS iteration is a saddle point of

$$\min_u \max_p J(u) + \langle p, b - Au \rangle - \frac{1}{2\delta} \|p - p^k\|^2 + \frac{1}{2} \|u - u^k\|_D^2$$

with $D = \frac{1}{\alpha} - \delta A^T A$ and $\alpha, \delta > 0$ chosen to ensure D is positive definite.

Pros and Cons of BOS

- BOS takes full advantage of separable structure of the problem. If $J(u) = \sum_i J_i(u_i)$, each minimization step decouples into simple proximal minimizations of the form

$$u_i^{k+1} = \arg\min_{u_i} J_i(u) + \frac{1}{2\alpha}\|u_i - \text{stuff}\|^2$$

  These often have closed form solutions or are easy to solve.

- The rate of convergence can be slow, especially if A is poorly conditioned.

Reformulation for Split Bregman/ADMM

In imaging applications, combining operator splitting with constrained optimization techniques was a breakthrough for solving total variation regularized problems.

- FTVd $\rightarrow$ Quadratic penalty method with alternating minimization

  Y. WANG, J. YANG, W. YIN AND Y. ZHANG, *A New Alternating Minimization Algorithm for Total Variation Image Reconstruction*, 2007.

- Split Bregman $\rightarrow$ Alternating minimization variant of Bregman iteration

  T. GOLDSTEIN AND S. OSHER, *The Split Bregman Algorithm for L1 Regularized Problems*, 2008.

Split Bregman and related methods often require considering the objective to be a sum of two convex functions, so we change the primal problem notation to

$$\min_{\substack{z \in \mathbb{R}^n, u \in \mathbb{R}^m \\ Bz + Au = b}} \quad F(z) + H(u) \qquad \text{such that } Bz + Au = b$$

where F and H are closed proper convex functions.

Split Bregman

$$z^{k+1} = \arg\min_z F(z) - F(z^k) - \langle p_z^k, z - z^k \rangle + \frac{\delta}{2}\|b - Au^k - Bz\|^2$$

$$u^{k+1} = \arg\min_u H(u) - H(u^k) - \langle p_u^k, u - u^k \rangle + \frac{\delta}{2}\|b - Au - Bz^{k+1}\|^2$$

$$p_z^{k+1} = p_z^k + \delta B^T(b - Au^{k+1} - Bz^{k+1})$$

$$p_u^{k+1} = p_u^k + \delta A^T(b - Au^{k+1} - Bz^{k+1})$$

$$p_z^0 = 0 \qquad p_u^0 = 0 \qquad p_u^k \in \partial H(u^k) \qquad p_z^k \in \partial F(z^k)$$

We will see this converges for $\delta > 0$ by comparing it to the Alternating Direction Method of Multipliers (ADMM)

ADMM

$$L_\delta(z,u,\lambda^k) = F(z) + H(u) + \langle \lambda^k, b - Au - Bz\rangle + \frac{\delta}{2}\|b - Au - Bz\|^2$$

$$\begin{aligned} z^{k+1} &= \arg\min_z L_\delta(z, u^k, \lambda^k) \\ u^{k+1} &= \arg\min_u L_\delta(z^{k+1}, u, \lambda^k) \\ \lambda^{k+1} &= \lambda^k + \delta(b - Au^{k+1} - Bz^{k+1}) \end{aligned}$$

Equivalence to Split Bregman with

$$p_z^k = B^T\lambda^k \qquad p_u^k = A^T\lambda^k \qquad \lambda^0 = 0$$

D. GABAY, AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comp. Math. Appl., 2 1976, pp. 17-40.

R. GLOWINSKI, AND A. MARROCCO, *Sur lapproximation par elements finis dordre un, et la resolution par penalisation-dualite dune classe de problemes de Dirichlet nonlineaires, Rev. Francaise dAut.*, 1975.

ADMM Convergence

Theorem 1 *(Eckstein, Bertsekas) Consider the primal problem where F and H are closed proper convex functions, $F(z) + \|Bz\|^2$ is strictly convex and $H(u) + \|Au\|^2$ is strictly convex. Let $\lambda^0 \in \mathbb{R}^d$ and $u^0 \in \mathbb{R}^m$ be arbitrary and let $\alpha > 0$. Suppose we are also given sequences $\{\mu_k\}$ and $\{\nu_k\}$ such that $\mu_k \geq 0$, $\nu_k \geq 0$, $\sum_{k=0}^{\infty} \mu_k < \infty$ and $\sum_{k=0}^{\infty} \nu_k < \infty$. Suppose that*

$$\|z^{k+1} - \arg\min_{z \in \mathbb{R}^n} F(z) + \langle \lambda^k, -Bz \rangle + \frac{\delta}{2}\|b - Au^k - Bz\|^2\| \leq \mu_k$$

$$\|u^{k+1} - \arg\min_{u \in \mathbb{R}^m} H(u) + \langle \lambda^k, -Au \rangle + \frac{\delta}{2}\|b - Au - Bz^{k+1}\|^2\| \leq \nu_k$$

$$\lambda^{k+1} = \lambda^k + \delta(b - Au^{k+1} - Bz^{k+1}).$$

If there exists a saddle point of $L(z, u, \lambda)$, then $z^k \to z^*$, $u^k \to u^*$ and $\lambda^k \to \lambda^*$, where (z^*, u^*, λ^*) is such a saddle point. On the other hand, if no such saddle point exists, then at least one of the sequences $\{u^k\}$ or $\{\lambda^k\}$ must be unbounded.

J. ECKSTEIN AND D. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, North-Holland, 1992.

Dual Interpretations

$$L(z,u,\lambda) = F(z) + H(u) + \langle \lambda, b - Au - Bz \rangle$$

$$q(\lambda) = \inf_{u \in \mathbb{R}^m, z \in \mathbb{R}^n} L(z,u,\lambda) = -F^*(B^T\lambda) - H^*(A^T\lambda) + \langle \lambda, b \rangle$$

Dual Problem: $\max_{\lambda \in \mathbb{R}^d} q(\lambda)$

Strong Duality: $F(z^*) + H(u^*) = q(\lambda^*)$

Saddle Point Characterization: (z^*, u^*) solves primal problem , λ^* solves dual iff

$$L(z^*, u^*, \lambda) \leq L(z^*, u^*, \lambda^*) \leq L(z, u, \lambda^*) \qquad \forall\ z, u, \lambda$$

Optimality Conditions $Au^* + Bz^* = b$

$$B^T\lambda^* \in \partial F(z^*) \qquad z^* \in \partial F^*(B^T\lambda^*)$$

$$A^T\lambda^* \in \partial H(u^*) \qquad u^* \in \partial H^*(A^T\lambda^*)$$

Douglas Rachford Splitting

Define: $\Psi(\lambda) = B\partial F^*(B^T\lambda) - b \qquad \phi(\lambda) = A\partial H^*(A^T\lambda)$

An Approach for solving dual: Find $0 \in \Psi(\lambda) + \phi(\lambda)$

Formal Douglas Rachford Splitting:

$$0 \in \frac{\hat{\lambda}^k - \lambda^k}{\delta} + \Psi(\hat{\lambda}^k) + \phi(\lambda^k),$$

$$0 \in \frac{\lambda^{k+1} - \lambda^k}{\delta} + \Psi(\hat{\lambda}^k) + \phi(\lambda^{k+1}).$$

ADMM Equivalent Version (Derived using Moreau decomposition)

$$\hat{\lambda}^k = \arg\min_{\hat{\lambda}} F^*(B^T\hat{\lambda}) - \langle \hat{\lambda}, b\rangle + \frac{1}{2\delta}\|\hat{\lambda} - (2\lambda^k - y^k)\|^2$$

$$\lambda^{k+1} = \arg\min_{\lambda} H^*(A^T\lambda) + \frac{1}{2\delta}\|\lambda - (y^k - \lambda^k + \hat{\lambda}^k)\|^2$$

$$y^{k+1} = y^k + \hat{\lambda}^k - \lambda^k$$

S. Setzer, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, 2009.

Split Inexact Uzawa

We can apply the same linearization techniques to ADMM that we applied to the method of multipliers when deriving BOS.

$$
\begin{aligned}
z^{k+1} &= \arg\min_z L_\delta(z, u^k, \lambda^k) \\
u^{k+1} &= \arg\min_u L_\delta(z^{k+1}, u, \lambda^k) \\
\lambda^{k+1} &= \lambda^k + \delta(b - Au^{k+1} - Bz^{k+1})
\end{aligned}
$$

In general, we can add $\frac{1}{2}\|z - z^k\|_{Q_z}^2$ to the z^{k+1} objective and/or $\frac{1}{2}\|u - u^k\|_{Q_u}^2$ to the u^{k+1} objective if Q_z and Q_u are positive definite.

The quadratic terms can be linearized for example by choosing $Q_z = \frac{1}{\alpha} - \delta B^T B$ with $\alpha, \delta > 0$ chosen to ensure Q_z is positive definite.

X. Zhang, M. Burger, and S. Osher, *A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration*, UCLA CAM Report [09-99], 2009.

A Simpler Problem to Show Connections

We will see that many popular primal-dual methods are actually quite similar to each other. Consider

$$\min_{u \in \mathbb{R}^m} J(Au) + H(u) \qquad (P)$$

J, H closed proper convex

$H : R^m \to (-\infty, \infty]$

$J : R^n \to (-\infty, \infty]$

$A \in \mathbb{R}^{n \times m}$

Assume there exists an optimal solution u^* to (P)

So we can use Fenchel duality later, also assume there exists $u \in \mathrm{ri}(\mathrm{dom}\, H)$ such that $Au \in \mathrm{ri}(\mathrm{dom}\, J)$ (almost always true in practice)

Saddle Point Form via Legendre Transform

Since $J^{**} = J$ for arbitrary closed proper convex J, we can use this to define a saddle point version of (P).

$$J(Au) = J^{**}(Au) = \sup_p \langle p, Au \rangle - J^*(p)$$

Primal Function	$F_P(u) = J(Au) + H(u)$
Saddle Function	$L_{PD}(u, p) = \langle p, Au \rangle - J^*(p) + H(u)$

Saddle Point Problem

$$\min_u \sup_p -J^*(p) + \langle p, Au \rangle + H(u) \qquad (PD)$$

Dual Problem and Strong Duality

The dual problem is

$$\max_{p \in \mathbb{R}^n} F_D(p) \qquad (D)$$

where the dual functional $F_D(p)$ is a concave function defined by

$$F_D(p) = \inf_{u \in \mathbb{R}^m} L_{PD}(u, p) = \inf_{u \in \mathbb{R}^m} -J^*(p) + \langle p, Au \rangle + H(u) = -J^*(p) - H^*(-A^T p)$$

- By Fenchel duality there exists an optimal solution p^* to (D)
- Strong duality holds, meaning $F_P(u^*) = F_D(p^*)$
- u^* solves (P) and p^* solves (D) iff (u^*, p^*) is saddle point of L_{PD}

R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

More Saddle Point Formulations

Introduce the constraint $w = Au$ in (P) and form the Lagrangian

$$L_P(u, w, p) = J(w) + H(u) + \langle p, Au - w \rangle$$

The corresponding saddle point problem is

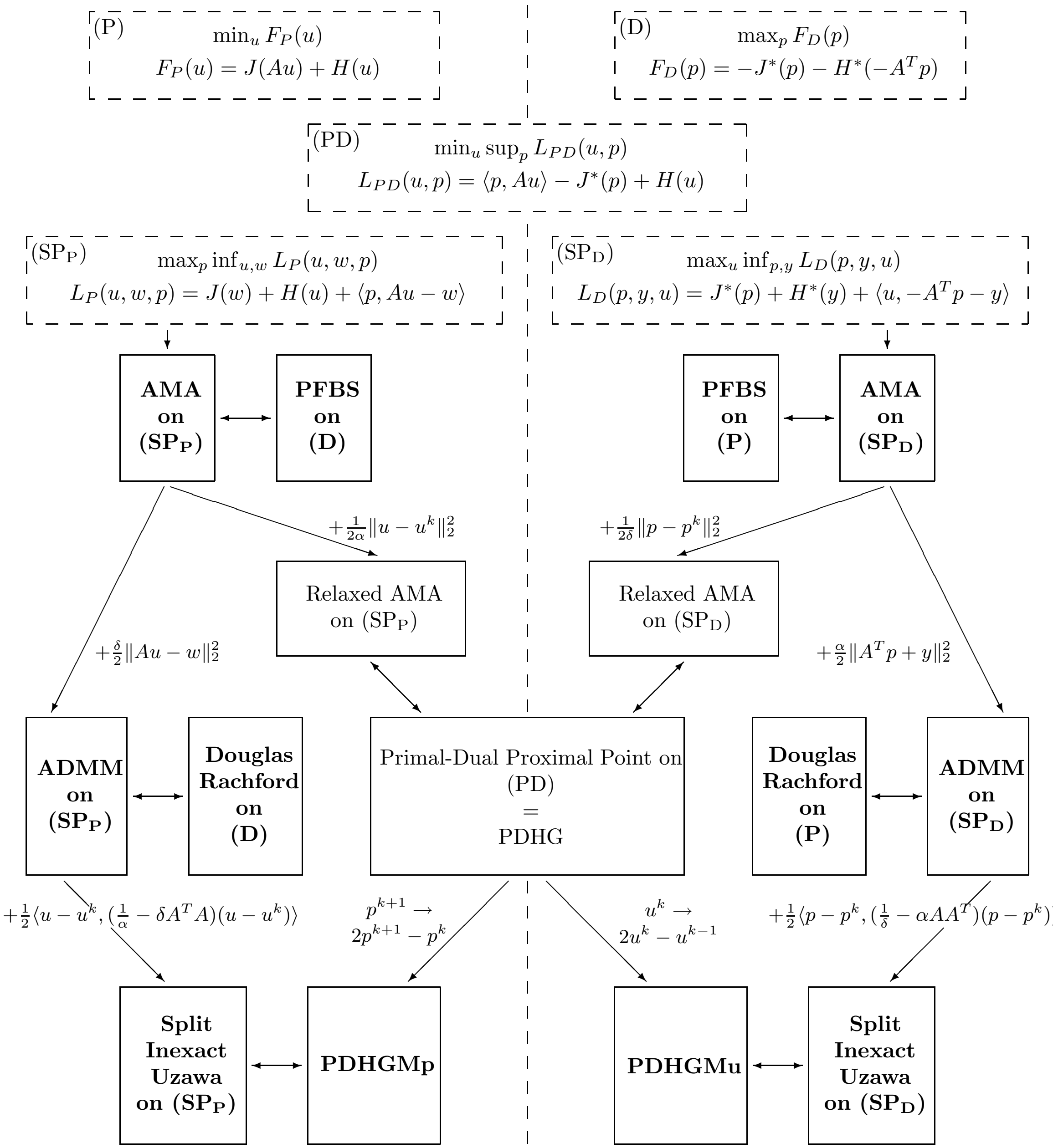
$$\max_{p \in \mathbb{R}^n} \inf_{u \in \mathbb{R}^m, w \in \mathbb{R}^n} L_P(u, w, p) \qquad (SPP)$$

Introduce the constraint $y = -A^T p$ in (D) and form the Lagrangian

$$L_D(p, y, u) = J^*(p) + H^*(y) + \langle u, -A^T p - y \rangle$$

Obtain yet another saddle point problem,

$$\max_{u \in \mathbb{R}^m} \inf_{p \in \mathbb{R}^n, y \in \mathbb{R}^m} L_D(p, y, u) \qquad (SPD)$$



Legend:
(P): Primal
(D): Dual
(PD): Primal-Dual
($\mathrm{SP_P}$): Split Primal
($\mathrm{SP_D}$): Split Dual

AMA: Alternating Minimization Algorithm (4.2.1)
PFBS: Proximal Forward Backward Splitting (4.2.1)
ADMM: Alternating Direction Method of Multipliers (4.2.2)
PDHG: Primal Dual Hybrid Gradient (4.2)
PDHGM: Modified PDHG (4.2.3)
Bold: Well Understood Convergence Properties

Primal Dual Hybrid Gradient (PDHG)

Interpret PDHG as a primal-dual proximal point method for finding a saddle point of

$$\min_{u\in\mathbb{R}^m} \sup_{p\in\mathbb{R}^n} -J^*(p) + \langle p, Au\rangle + H(u) \qquad (PD)$$

PDHG iterations:

$$p^{k+1} = \arg\max_{p\in\mathbb{R}^n} -J^*(p) + \langle p, Au^k\rangle - \frac{1}{2\delta_k}\|p - p^k\|_2^2$$

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^m} H(u) + \langle A^T p^{k+1}, u\rangle + \frac{1}{2\alpha_k}\|u - u^k\|_2^2$$

M. Zhu, and T. F. Chan, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, UCLA CAM Report [08-34], May 2008.

B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for total variation image restoration*, 2010.

Proximal Forward Backward Splitting

PFBS alternates a gradient descent step with a proximal step:

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) + \frac{1}{2\delta_k} \|p - (p^k + \delta_k A u^{k+1})\|_2^2,$$

where $u^{k+1} = \nabla H^*(-A^T p^k)$.

Since $u^{k+1} = \nabla H^*(-A^T p^k) \Leftrightarrow -A^T p^k \in \partial H(u^{k+1})$, which is equivalent to

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle,$$

PFBS on (D) can be rewritten as

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle$$

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) + \langle p, -Au^{k+1} \rangle + \frac{1}{2\delta_k} \|p - p^k\|_2^2$$

Converges if we assume $\nabla(H^*(-A^T))$ is Lipschitz continuous and the product of δ_k and its Lipschitz constant is in $(0, 2)$.

P. Combettes and W. Wajs, *Signal Recovery by Proximal Forward-Backward Splitting*, 2006.

AMA on Split Primal

AMA applied to (SPP) alternately minimizes first the Lagrangian $L_P(u, w, p)$ with respect to u and then the augmented Lagrangian $L_P + \frac{\delta_k}{2}\|Au - w\|_2^2$ with respect to w before updating the Lagrange multiplier p.

$$
\begin{aligned}
u^{k+1} &= \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle \\
w^{k+1} &= \arg\min_{w \in \mathbb{R}^n} J(w) - \langle p^k, w \rangle + \frac{\delta_k}{2}\|Au^{k+1} - w\|_2^2 \\
p^{k+1} &= p^k + \delta_k(Au^{k+1} - w^{k+1})
\end{aligned}
$$

- Can show equivalence to PFBS on (D) by a direct application of Moreau's decomposition

P. TSENG, *Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities*, SIAM J. Control Optim., Vol. 29, No. 1, 1991, pp. 119-138.

Equivalence by Moreau Decomposition

AMA applied to (SPP):

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle$$

$$w^{k+1} = \arg\min_{w \in \mathbb{R}^n} J(w) - \langle p^k, w \rangle + \frac{\delta_k}{2} \|Au^{k+1} - w\|_2^2$$

$$p^{k+1} = p^k + \delta_k (Au^{k+1} - w^{k+1})$$

The rewritten PFBS on (D) and AMA on (SPP) have the same first step.

Combining the last two steps of AMA yields

$$p^{k+1} = (p^k + \delta_k A u^{k+1}) - \delta_k \arg\min_w J(w) + \frac{\delta_k}{2} \left\| w - \frac{(p^k + \delta_k A u^{k+1})}{\delta_k} \right\|_2^2,$$

which is equivalent to the second step of PFBS by direct application of Moreau's decomposition.

$$p^{k+1} = \arg\min_p J^*(p) + \frac{1}{2\delta_k} \|p - (p^k + \delta_k A u^{k+1})\|_2^2$$

AMA/PFBS Connection to PDHG

PFBS on (D) plus additional proximal penalty is PDHG

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle + \frac{1}{2\alpha_k} \|u - u^k\|_2^2$$

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) + \langle p, -Au^{k+1} \rangle + \frac{1}{2\delta_k} \|p - p^k\|_2^2$$

AMA on (SPP) with first step relaxed by same proximal penalty is PDHG

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle + \frac{1}{2\alpha_k} \|u - u^k\|_2^2$$

$$w^{k+1} = \arg\min_{w \in \mathbb{R}^n} J(w) - \langle p^k, w \rangle + \frac{\delta_k}{2} \|Au^{k+1} - w\|_2^2$$

$$p^{k+1} = p^k + \delta_k (Au^{k+1} - w^{k+1})$$

- PFBS on (P) and AMA on (SPD) are connected to PDHG analogously
- Can think of PDHG as a relaxed version of AMA

AMA Connection to ADMM

AMA on (SPP) with $\frac{\delta}{2}\|Au - w^k\|_2^2$ added to first step is ADMM applied to (SPP):

$$
\begin{aligned}
u^{k+1} &= \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u\rangle + \frac{\delta}{2}\|Au - w^k\|_2^2 \\
w^{k+1} &= \arg\min_{w \in \mathbb{R}^n} J(w) - \langle p^k, w\rangle + \frac{\delta}{2}\|Au^{k+1} - w\|_2^2 \\
p^{k+1} &= p^k + \delta(Au^{k+1} - w^{k+1})
\end{aligned}
$$

- ADMM alternately minimizes the augmented Lagrangian $L_P + \frac{\delta}{2}\|Au - w\|_2^2$ with respect to u and w before updating the Lagrange multiplier p
- Equivalent to Split Bregman, a method which combines Bregman iteration and operator splitting to solve constrained convex optimization problems

D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*, 1989.

T. Goldstein and S. Osher, *The Split Bregman Algorithm for L1 Regularized Problems*, SIIMS, Vol. 2, No. 2, 2008.

Equivalence to Douglas Rachford Splitting

Can apply Moreau decomposition twice along with an appropriate change of variables to show ADMM on (SPP) or (SPD) is equivalent to Douglas Rachford Splitting on (D) and (P) resp.

Douglas Rachford splitting on (D):

$$z^{k+1} = \arg\min_q H^*(-A^T q) + \frac{1}{2\delta}\|q - (2p^k - z^k)\|_2^2 + z^k - p^k$$

$$p^{k+1} = \arg\min_p J^*(p) + \frac{1}{2\delta}\|p - z^{k+1}\|_2^2$$

Note: $z^k = p^k + \delta w^k$ with p^k and w^k the same as in ADMM on (SPP)

S. Setzer, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, LNCS, 2008.

J. Eckstein, and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program. 55, 1992.

P.L. Combettes and J-C. Pesquet, *A Douglas-Rachford Splitting Approach to Nonsmooth Convex Variational Signal Recovery*, IEEE, 2007.

Split Inexact Uzawa Method

Special case: only linearize the u^{k+1} step of ADMM applied to (SPP) by adding $\frac{1}{2}\langle u - u^k, (\frac{1}{\alpha} - \delta A^T A)(u - u^k)\rangle$ to the objective function, with $0 < \alpha < \frac{1}{\delta \|A\|^2}$.

Split Inexact Uzawa applied to (SPP):

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u\rangle + \frac{1}{2\alpha}\|u - u^k + \delta\alpha A^T(Au^k - w^k)\|_2^2$$

$$w^{k+1} = \arg\min_{w \in \mathbb{R}^n} J(w) - \langle p^k, w\rangle + \frac{\delta}{2}\|Au^{k+1} - w\|_2^2$$

$$p^{k+1} = p^k + \delta(Au^{k+1} - w^{k+1})$$

By only modifying the first step of ADMM, we obtain an interesting PDHG-like interpretation.

Modified PDHG (PDHGMp)

Replace p^k in first step of PDHG with $2p^k - p^{k-1}$ to get PDHGMp:

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \left\langle A^T \left(2p^k - p^{k-1}\right), u \right\rangle + \frac{1}{2\alpha}\|u - u^k\|_2^2$$

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) - \langle p, Au^{k+1} \rangle + \frac{1}{2\delta}\|p - p^k\|_2^2$$

- Can show equivalence to SIU on (SPP) using Moreau's decomposition

Related Works:

- G. Chen and M. Teboulle, *A Proximal-Based Decomposition Method for Convex Minimization Problems*, Mathematical Programming, Vol. 64, 1994.
- T. Pock, D. Cremers, H. Bischof, and A. Chambolle, *An Algorithm for Minimizing the Mumford-Shah Functional*, ICCV, 2009.
- A. Chambolle, V. Caselles, M. Novaga, D. Cremers and T. Pock, *An introduction to Total Variation for Image Analysis*, http://hal.archives-ouvertes.fr/docs/00/43/75/81/PDF/preprint.pdf, 2009.
- A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, 2010.

Equivalence of PDHGMp and SIU on (SPP)

SIU on (SPP): (the only change from PDHG is addition of blue term)

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^m} H(u) + \langle A^T p^k, u\rangle + \frac{1}{2\alpha}\|u - u^k + \delta\alpha A^T(Au^k - w^k)\|_2^2$$

$$w^{k+1} = \arg\min_{w\in\mathbb{R}^n} J(w) - \langle p^k, w\rangle + \frac{\delta}{2}\|Au^{k+1} - w\|_2^2$$

$$p^{k+1} = p^k + \delta(Au^{k+1} - w^{k+1})$$

Replace $\delta(Au^k - w^k)$ in the u^{k+1} update with $p^k - p^{k-1}$.
Combine p^{k+1} and w^{k+1} to get

$$p^{k+1} = (p^k + \delta Au^{k+1}) - \delta \arg\min_w J(w) + \frac{\delta}{2}\|w - \frac{(p^k + \delta Au^{k+1})}{\delta}\|^2$$

and apply Moreau's decomposition.

PDHGMp: (the only change from PDHG is that p^k became $2p^k - p^{k-1}$)

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^m} H(u) + \left\langle A^T \left(2p^k - p^{k-1}\right), u\right\rangle + \frac{1}{2\alpha}\|u - u^k\|_2^2$$

$$p^{k+1} = \arg\min_{p\in\mathbb{R}^n} J^*(p) - \langle p, Au^{k+1}\rangle + \frac{1}{2\delta}\|p - p^k\|_2^2$$

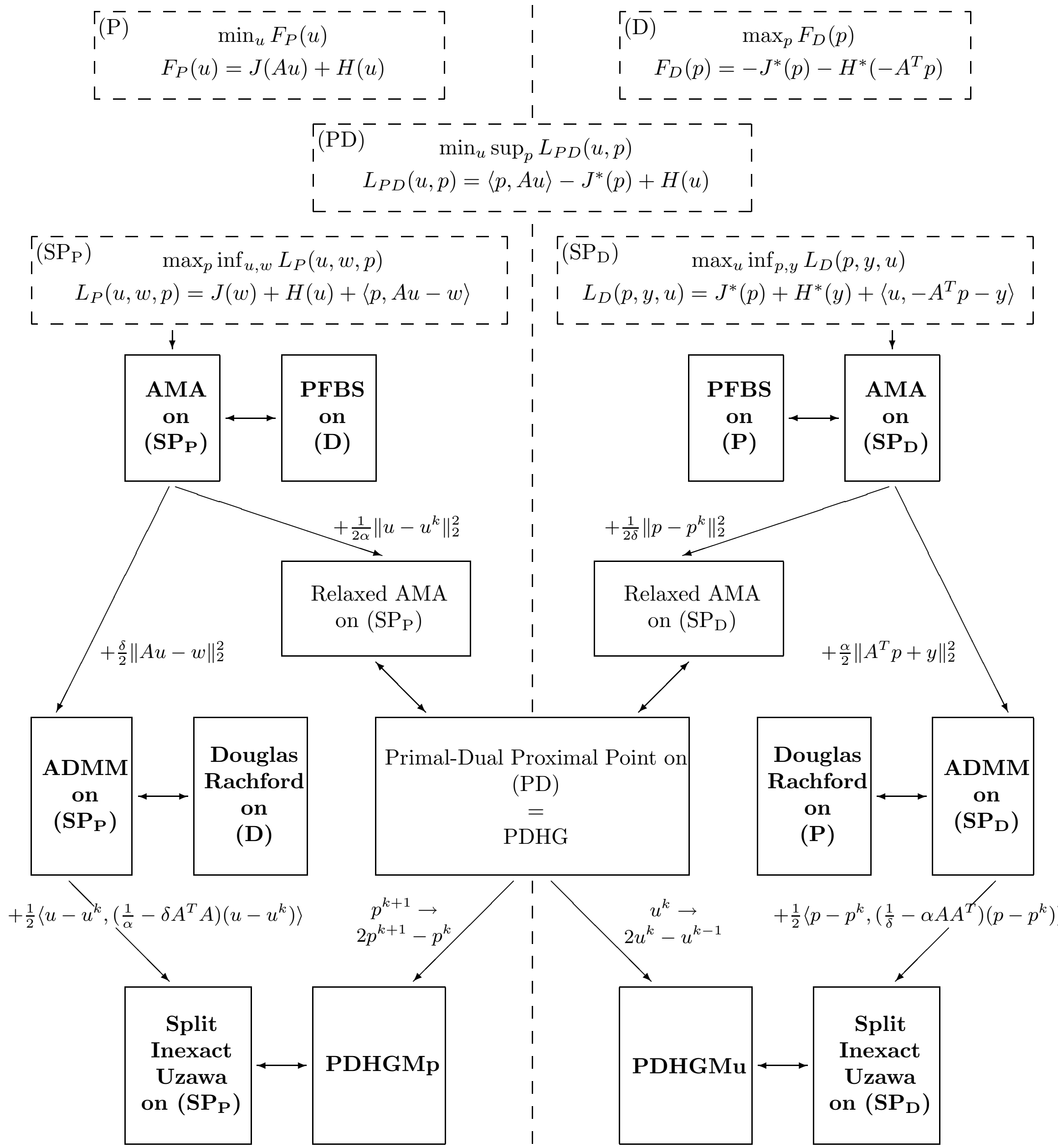
Modified PDHG (PDHGMu)

PDHGMu: (the only change from PDHG is that u^k became $2u^k - u^{k-1}$)

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) - \left\langle p, A\left(2u^k - u^{k-1}\right)\right\rangle + \frac{1}{2\delta}\|p - p^k\|_2^2$$

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \left\langle A^T p^{k+1}, u\right\rangle + \frac{1}{2\alpha}\|u - u^k\|_2^2$$

PDHGMu is analogously equivalent to the split inexact Uzawa (SIU) method applied to (SPD)



Legend: (P): Primal
(D): Dual
(PD): Primal-Dual
($\mathrm{SP_P}$): Split Primal
($\mathrm{SP_D}$): Split Dual

AMA: Alternating Minimization Algorithm (4.2.1)
PFBS: Proximal Forward Backward Splitting (4.2.1)
ADMM: Alternating Direction Method of Multipliers (4.2.2)
PDHG: Primal Dual Hybrid Gradient (4.2)
PDHGM: Modified PDHG (4.2.3)
Bold: Well Understood Convergence Properties

Comparison of Algorithms

	Step size restrictions (α and δ parameters)	Additional smooth or strong convexity assumptions	Can decouple variables coupled by linear constraints	Assumes objective is written as sum of n terms, where n is
MM/Bregman	no	no	no	1
BOS	yes	no	yes	arbitrary
ADMM/DR	no	no	no	2
AMA/PFBS	yes	yes	yes	2
PDHG	yes	no	yes	2
PDHGM/SIU	yes	no	yes	2

More About Algorithm Connections

- S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, 2010.
- P.L. Combettes and J-C. Pesquet, *Proximal Splitting Methods in Signal Processing*, 2009.
- J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph. D. Thesis, MIT, Dept. of Civil Engineering, 1989.
- E. Esser, *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*, UCLA CAM Report [09-31], 2009.
- E. Esser, X. Zhang, and T. F. Chan, *A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science*, SIAM J. Imaging Sci. Volume 3, Issue 4, pp. 1015-1046, 2010.
- R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM, 1989.
- C. Wu and X.C Tai, *Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and High Order Models*, 2009.

Accelerated Variant of PFBS

FISTA algorithm modifies PFBS on (D):

$$p^{k+1} = \arg\min_p J^*(p) + \frac{1}{\delta_k}\|p - (p^k + \delta_k A\nabla H^*(-A^T p^k))\|^2$$

by replacing p^k with $p^k + \frac{t_k - 1}{t_{k+1}}(p^k - p^{k-1})$,

where $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$. (initialize at $t_1 = 1$)

Complexity of FISTA is $O(\frac{1}{k^2})$:

$$J(Au^k) + H(u^k) - J(Au^*) - H(u^*) \leq \frac{C}{k^2}$$

C is independent of k, but does depend on the Lipschitz constant and step size.

A. BECK, AND M. TEBOULLE, *Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems*, 2009.

Accelerated Variant of PDHGM

$$p^{k+1} = \arg\min_{p \in \mathbb{R}^n} J^*(p) - \left\langle p, Aw^k \right\rangle + \frac{1}{2\delta_k}\|p - p^k\|_2^2$$

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) + \left\langle A^T p^{k+1}, u \right\rangle + \frac{1}{2\alpha_k}\|u - u^k\|_2^2$$

Instead of $w^k = 2u^k - u^{k-1}$, let $w^k = u^k + \theta_k(u^k - u^{k-1})$
Well chosen θ_k accelerates convergence rate: $O(\frac{1}{N})$ in general case, $O(\frac{1}{N^2})$ if J^* or H is strongly convex, and linear if both are strongly convex

In particular, in the H strongly convex case, let

$$\begin{aligned}
\theta_k &= \frac{1}{\sqrt{1 + 2\gamma\alpha_k}} \qquad (\nabla H^* \text{ is } \frac{1}{\gamma} \text{ Lipschitz}) \\
\alpha_{k+1} &= \theta_k \alpha_k \\
\delta_{k+1} &= \frac{\delta_k}{\theta_k} \\
w^{k+1} &= u^{k+1} + \theta_k(u^{k+1} - u^k)
\end{aligned}$$

A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, 2010.

Other Algorithm Improvements

With a suitable correction step, new alternating direction methods based on the augmented Lagrangian no longer require the objective functional to be split into two parts.

Being able to apply more than two alternating steps is advantageous when linear constraints couple many variables.

B. HE, Z. PENG AND X. WANG, *Proximal alternating direction-based contraction methods for separable linearly constrained convex optimization*, 2010.

B. HE, M. TAO AND X. YUAN, *A splitting method for separate convex programming with linking linear constraints*, 2010.

Smart application of Newton-based methods to primal-dual optimality conditions can achieve superlinear convergence.

R. CHAN, Y. DONG AND M. HINTERMULLER, *An Efficient Two-Phase L_1-TV Method for Restoring Blurred Images with Impulse Noise*, 2010.

T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal dual method for total variation based image restoration*, 1999.

Outline for Implementation Details

- Operator splitting
- Convex constraints
- TV discretization
- Easy to handle functions
- Examples
  - PFBS for TVL2 denoising (ROF)
  - ADMM and BOS for TVL1 minimization
  - ADMM for sparse/low rank decomposition
  - PDHGM for constrained deblurring
  - PDHGM for multiphase segmentation
  - ADMM for nonnegative matrix factorization

Operator Splitting for PDHGM

Applying PDHGMp to $\min_u \sum_{i=1}^N J_i(A_i u) + H(u)$ yields:

$$u^{k+1} = \arg\min_u H(u) + \frac{1}{2\alpha}\left\| u - \left(u^k - \alpha \sum_{i=1}^N A_i^T (2p_i^k - p_i^{k-1}) \right) \right\|_2^2$$

$$p_i^{k+1} = \arg\min_{p_i} J_i^*(p_i) + \frac{1}{2\delta}\left\| p_i - \left(p_i^k + \delta A_i u^{k+1} \right) \right\|_2^2 \qquad i = 1, ..., N$$

where $J(Au) = \sum_{i=1}^N J_i(A_i u)$.

Letting $p = \begin{bmatrix} p_1 \\ \vdots \\ p_N \end{bmatrix}$, the decoupling follows from $J^*(p) = \sum_{i=1}^N J_i^*(p_i)$

- The p_i subproblems are decoupled
- Need $0 < \alpha < \frac{1}{\delta \|A\|^2}$ for stability with $A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}$
- Preconditioning is possible

Convex Constraints

We want the algorithms to have simple, explicit solutions to their minimization subproblems.

A convex constraint $u \in T$ can be handled by adding the convex indicator function

$$g_T(u) = \begin{cases} 0 & \text{if } u \in T \\ \infty & \text{otherwise.} \end{cases}$$

This leads to a simple update when the orthogonal projection

$$\Pi_T(z) = \arg\min_u g_T(u) + \|u - z\|^2$$

is easy to compute. For example,

$$T = \{z : \|z - f\|_2 \leq \epsilon\} \Rightarrow \Pi_T(z) = f + \frac{z - f}{\max\left(\frac{\|z-f\|}{\epsilon}, 1\right)}$$

TV Discretization (1)

Temporarily thinking of u as a $M_r \times M_c$ matrix, discretize $\|u\|_{TV}$ using forward differences and assuming Neumann BC by

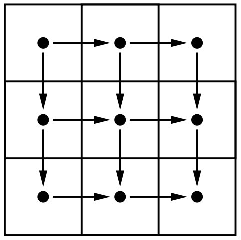
$$\|u\|_{TV} = \sum_{r=1}^{M_r} \sum_{c=1}^{M_c} \sqrt{(D_c^+ u_{r,c})^2 + (D_r^+ u_{r,c})^2}$$

Vectorize $M_r \times M_c$ matrix u by stacking columns

Define a discrete gradient matrix D and a norm $\| \cdot \|_E$ such that $\|Du\|_E = \|u\|_{TV}$.

Define a directed grid-shaped graph with $m = M_r M_c$ nodes corresponding to matrix elements (r, c).

3×3 example:



For each edge η with endpoint indices (i, j), $i < j$, define:

$$D_{\eta,k} = \begin{cases} -1 & \text{for } k = i, \\ 1 & \text{for } k = j, \\ 0 & \text{for } k \neq i, j. \end{cases} \qquad E_{\eta,k} = \begin{cases} 1 & \text{if } D_{\eta,k} = -1, \\ 0 & \text{otherwise.} \end{cases}$$

TV Discretization (2)

Can use E to define norm $\|\cdot\|_E$ on $\mathbb{R}^e$ by

$$\|w\|_E = \left\| \sqrt{E^T(w^2)} \right\|_1 = \sum_{i=1}^{m} \left(\sqrt{E^T(w^2)} \right)_i = \sum_{i=1}^{m} \|w_i\|_2$$

where w_i is the vector of edge values for directed edges coming out of node i.

$$\|p\|_{E^*} = \left\| \sqrt{E^T(p^2)} \right\|_\infty = \max_i \left(\sqrt{E^T(p^2)} \right)_i = \max_i \|w_i\|_2$$

Again, p_i is the vector of edge values for directed edges coming out of node i.

For TV regularization, $J(Au) = \|Du\|_E = \|u\|_{TV}$

Some Easy Functions to Deal With

The methods discussed are most efficient when the decoupled minimization subproblems can be easily computed. A few examples (there are many more) of "nice" functions and their Legendre transforms include

$$J(z) = \frac{1}{2\alpha}\|z\|_2^2 \qquad J^*(p) = \frac{\alpha}{2}\|p\|_2^2$$

$$J(z) = \|z\|_2 \qquad J^*(p) = g_{\{p:\|p\|_2 \leq 1\}}$$

$$J(z) = \|z\|_1 \qquad J^*(p) = g_{\{p:\|p\|_\infty \leq 1\}}$$

$$J(z) = \|z\|_E \text{ where } \|Du\|_E = \|u\|_{TV} \qquad J^*(p) = g_{\{p:\|p\|_{E^*} \leq 1\}}$$

$$J(z) = \|z\|_\infty \qquad J^*(p) = g_{\{p:\|p\|_1 \leq 1\}}$$

$$J(z) = \max(z) \qquad J^*(p) = g_{\{p:p\geq 0 \text{ and } \|p\|_1 = 1\}}$$

$$J(z) = \inf_w F(w) + H(z-w) \qquad J^*(p) = F^*(p) + H^*(p)$$

Note: All indicator functions g are for convex sets that are easy to project onto.

Although there's no simple formula for projecting a vector onto the l_1 unit ball (or its positive face) in $\mathbb{R}^n$, this can be computed with $O(n \log n)$ complexity.

Preconditioning

When A is poorly conditioned, the rate of convergence can be poor. Preconditioning is sometimes a practical necessity.

Example for BOS: Recall that the standard BOS iteration computes (u^{k+1}, p^{k+1}) as a saddle point of

$$\min_u \max_p J(u) + \langle p, b - Au \rangle - \frac{1}{2\delta} \|p - p^k\|^2 + \frac{1}{2} \|u - u^k\|_D^2$$

with $D = \frac{1}{\alpha} - \delta A^T A$ and $\alpha, \delta > 0$ chosen to ensure D is positive definite.

We can precondition by working in a different metric defined by a positive definite matrix M, computing instead

$$\min_u \max_p J(u) + \langle p, M(b - Au) \rangle - \frac{1}{2\delta} \|p - p^k\|_M^2 + \frac{1}{2} \|u - u^k\|_{D_M}^2$$

now with $D_M = \frac{1}{\alpha} - \delta A^T M A$.

We can also precondition using a change of variables $u = Sv$, but must be careful not to overly complicate subproblems by the choice of M or S.

PFBS for TVL2 denoising

TVL2 denoising (ROF model): $\min_u \|u\|_{TV} + \frac{\lambda}{2}\|u - f\|_2^2$

Let $A = D$, $J(Au) = \|Du\|_E$ and $H(u) = \frac{\lambda}{2}\|u - f\|_2^2$ to write the model in the form of $\min_u J(Au) + H(u)$.
The dual form of PFBS yields the following iterations:

$$u^{k+1} = \arg\min_u H(u) + \langle D^T p^k, u \rangle$$

$$p^{k+1} = \arg \min_{\|p\|_{E^*} \leq 1} \langle p, -Du^{k+1} \rangle + \frac{1}{2\delta_k}\|p - p^k\|^2$$

These can be explicitly computed:

$$u^{k+1} = f - \frac{1}{\lambda} D^T p^k$$

$$p^{k+1} = \Pi_X (p^k + \delta_k D u^{k+1})$$

where to simplify notation, $X = \{p : \|p\|_{E^*} \leq 1\}$.

Gradient Projection Interpretation

From the optimality condition for $u^{k+1} = \arg\min_u H(u) + \langle D^T p^k, u\rangle$

$$-D^T p^k \in \partial H(u^{k+1}) \qquad \Leftrightarrow \qquad u^{k+1} = \nabla H^*(-D^T p^k)$$

Therefore

$$p^{k+1} = \Pi_X \left(p^k - \delta_k \nabla(H^*(-D^T p^k))\right)$$

$$\Pi_X(p) = \arg\min_{q \in X} \|q - p\|_2^2 = \frac{p}{E \max\left(\sqrt{E^T(p^2)}, 1\right)} \quad \text{(componentwise)}$$

$$\Pi_X(p)_i = \frac{p_i}{\max(\|p_i\|_2, 1)}$$

where p_i is the vector of edge values for directed edges out of node i.

Original, Noisy and Denoised Images

Use 256×256 cameraman image.

Add white Gaussian noise having standard deviation 20.

Let $\lambda = .053$.

ADMM for TVL1

$$\min_u \|u\|_{TV} + \beta\|Ku - f\|_1$$

Rewrite as

$$\min_u \|Du\|_E + \beta\|Ku - f\|_1$$

Let $z = \begin{bmatrix} w \\ v \end{bmatrix} = \begin{bmatrix} Du \\ Ku - f \end{bmatrix}$, $\qquad B = -I \qquad A = \begin{bmatrix} D \\ K \end{bmatrix}$, $\qquad b = \begin{bmatrix} 0 \\ f \end{bmatrix}$

to put in form $\qquad \min_{z,u} F(z) + H(u) \qquad$ s.t. $Bz + Au = b$

where $F(z) = \|w\|_E + \beta\|v\|_1$ and $H(u) = 0$.

Introduce dual variable $\lambda = \begin{bmatrix} p \\ q \end{bmatrix}$.

Assume $\ker(D) \bigcap \ker(K) = \{0\}$.

Augmented Lagrangian and ADMM Iterations

$$L_\delta(z,u,\lambda) = \|w\|_E + \beta\|v\|_1 + \langle p, Du - w\rangle + \langle q, Ku - f - v\rangle +$$
$$\frac{\delta}{2}\|w - Du\|^2 + \frac{\delta}{2}\|v - Ku + f\|^2$$

The ADMM iterations are given by

$$\begin{aligned}
w^{k+1} &= \arg\min_w \|w\|_E + \frac{\delta}{2}\|w - Du^k - \frac{p^k}{\delta}\|^2 \\
v^{k+1} &= \arg\min_v \beta\|v\|_1 + \frac{\delta}{2}\|v - Ku^k + f - \frac{q^k}{\delta}\|^2 \\
u^{k+1} &= \arg\min_u \frac{\delta}{2}\|Du - w^{k+1} + \frac{p^k}{\delta}\|^2 + \frac{\delta}{2}\|Ku - v^{k+1} - f + \frac{q^k}{\delta}\|^2 \\
p^{k+1} &= p^k + \delta(Du^{k+1} - w^{k+1}) \\
q^{k+1} &= q^k + \delta(Ku^{k+1} - f - v^{k+1}),
\end{aligned}$$

where $p^0 = q^0 = 0$, u^0 is arbitrary and $\delta > 0$.

Explicit Iterations

The explicit formulas for w^{k+1}, v^{k+1} and u^{k+1} are given by

$$
\begin{aligned}
w^{k+1} &= \tilde{S}_{\frac{1}{\delta}}(Du^k + \frac{p^k}{\delta}) \\
v^{k+1} &= S_{\frac{\beta}{\delta}}(Ku^k - f + \frac{q^k}{\delta}) \\
u^{k+1} &= (D^T D + K^T K)^{-1} \left(D^T w^{k+1} - \frac{D^T p^k}{\delta} + K^T (v^{k+1} + f) - \frac{K^T q^k}{\delta} \right) \\
&= (D^T D + K^T K)^{-1} \left(D^T w^{k+1} + K^T (v^{k+1} + f) \right).
\end{aligned}
$$

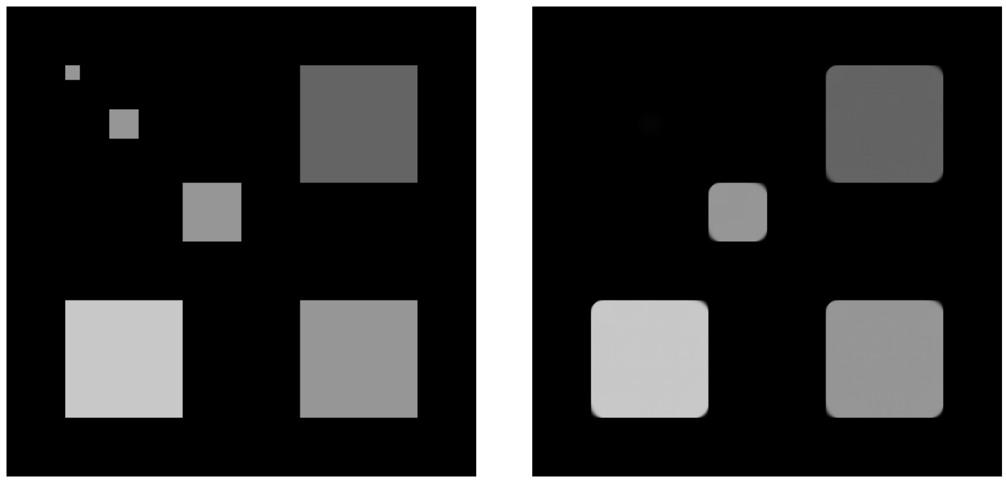
where

$$
\tilde{S}_c(f) = f - \Pi_{cX}(f) \ \text{ and } \ S_c(f) = f - \Pi_{\{q:\|q\|_\infty \le c\}}(f)
$$

are both soft thresholding formulas in vector and scalar cases, respectively.

TVL1 Results ($K = I$ case)

f $\qquad\qquad$ u



TV-l_1 Minimization of 512×512 Synthetic Image

Image Size	Iterations	Time
64×64	40	1s
128×128	51	5s
256×256	136	78s
512×512	359	836s

Iterations until $\|u^k - u^{k-1}\|_\infty \leq .5$, $\|Du^k - w^k\|_\infty \leq .5$ and $\|v^k - u^k + f\|_\infty \leq .5$

$\beta = .6, .3, .15$ and $.075$, $\qquad \delta = .02, .01, .005$ and $.0025$

BOS for TVL1

$$\min_{w,v,u} \|w\|_E + \beta\|v\|_1 \text{ such that } \begin{bmatrix} I & 0 & -D \\ 0 & I & -K \end{bmatrix} \begin{bmatrix} w \\ v \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ -f \end{bmatrix}$$

$$\begin{aligned}
w^{k+1} &= \arg\min_w \|w\|_E + \frac{1}{2\alpha}\|w - w^k + \alpha\delta(w^k - Du^k - \frac{p^k}{\delta})\|^2 \\
v^{k+1} &= \arg\min_v \beta\|v\|_1 + \frac{1}{2\alpha}\|v - v^k + \alpha\delta(v^k - Ku^k + f - \frac{q^k}{\delta})\|^2 \\
u^{k+1} &= \arg\min_u \frac{1}{2\alpha}\|u - u^k + \alpha\delta(-D^T w^k + D^T D u^k - K^T v^k + K^T K u^k \\
&\quad - K^T f + \frac{D^T p^k}{\delta} + \frac{K^T q^k}{\delta})\|^2 \\
p^{k+1} &= p^k + \delta(w^{k+1} - Du^{k+1}) \\
q^{k+1} &= q^k + \delta(v^{k+1} - Ku^{k+1} + f)
\end{aligned}$$

ADMM for Sparse/Low Rank Decomposition

$$\min \|u\|_* + \lambda \|e\|_1 \qquad \text{such that} \qquad f = u + e$$

$$L_\delta(u,e,p) = \|u\|_* + \lambda\|e\|_1 + \langle p, u+e-f\rangle + \frac{\delta}{2}\|u + e^k - f + \frac{p^k}{\delta}\|^2$$

$$\begin{aligned}
u^{k+1} &= \arg\min_u \|u\|_* + \frac{\delta}{2}\|u + e^k - f + \frac{p^k}{\delta}\|^2 \\
e^{k+1} &= \arg\min_e \lambda\|e\|_1 + \frac{\delta}{2}\|e + u^{k+1} - f + \frac{p^k}{\delta}\|^2 \\
p^{k+1} &= p^k + \delta(u^{k+1} + e^{k+1} - f)
\end{aligned}$$

Explicit Iterations

$$
\begin{aligned}
u^{k+1} &= f - e^k - \frac{p^k}{\delta} - \Pi_{\|\cdot\|_2 \le \frac{1}{\delta}}(f - e^k - \frac{p^k}{\delta}) \\
e^{k+1} &= f - u^{k+1} - \frac{p^k}{\delta} - \Pi_{\|\cdot\|_\infty \le \frac{\lambda}{\delta}}(f - u^{k+1} - \frac{p^k}{\delta}) \\
p^{k+1} &= p^k + \delta(u^{k+1} + e^{k+1} - f)
\end{aligned}
$$

In this context, $\|\cdot\|_2$ denotes the spectral norm, which is the largest singular value.

The u^{k+1} update can be computed by using the singular value decomposition to soft threshold the singular values of $f - e^k - \frac{p^k}{\delta}$.

Constrained TV Deblurring Example

$$\min_{\|Ku-f\|_2 \leq \epsilon} \|u\|_{TV}$$

can be rewritten as

$$\min_u \|Du\|_E + g_T(Ku),$$

where g_T is the indicator function for $T = \{z : \|z - f\|_2 \leq \epsilon\}$

In order to treat both D and K explicitly, let

$$H(u) = 0 \qquad \text{and} \qquad J(Au) = J_1(Du) + J_2(Ku),$$

where $A = \begin{bmatrix} D \\ K \end{bmatrix}$.

Write the dual variable as $p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$ and apply PDHGMp.

PDHGMp for Constrained TV Deblurring

$$
\begin{aligned}
u^{k+1} &= \arg\min_u H(u) + \left\langle A^T \left(2p^k - p^{k-1}\right), u\right\rangle + \frac{1}{2\alpha}\|u - u^k\|_2^2 \\
p^{k+1} &= \arg\min_p J^*(p) - \langle p, Au^{k+1}\rangle + \frac{1}{2\delta}\|p - p^k\|_2^2
\end{aligned}
$$

$$
\begin{aligned}
u^{k+1} &= u^k - \alpha \left(D^T(2p_1^k - p_1^{k-1}) + K^T(2p_2^k - p_2^{k-1})\right) \\
p_1^{k+1} &= \Pi_X \left(p_1^k + \delta D u^{k+1}\right) \\
p_2^{k+1} &= p_2^k + \delta K u^{k+1} - \delta \Pi_T \left(\frac{p_2^k}{\delta} + K u^{k+1}\right),
\end{aligned}
$$

where Π_X and Π_T are defined as before.

Both projections are simple to compute:

- Π_X is analogous to orthogonal projection onto an l_∞ ball
- Π_T is orthogonal projection onto the l_2 ϵ-ball centered at f

Deblurring Parameters

$$\min_u \|u\|_{TV} \qquad \text{such that} \qquad \|Ku - f\|_2 \leq \epsilon$$

K is a convolution operator corresponding to a normalized Gaussian blur with a standard deviation of 3 in a 17×17 window.

Letting h denote the clean image, the given data f is $f = Kh + \eta$, where η is zero mean Gaussian noise with standard deviation 1.

Let $\epsilon = 256$, and choose algorithm parameters $\alpha = .33$ and $\delta = .33$.

Original, blurry/noisy and image recovered from 300 iterations

Multiphase Segmentation Example

Recall the convex approximation we considered for multiphase segmentation.

Goal: Segment a given image, $h \in \mathbb{R}^M$, into W regions where the intensities in the w^{th} region are close to given intensities $z_w \in \mathbb{R}$ and the lengths of the boundaries between regions are not too long.

$$\min_c g_C(c) + \sum_{w=1}^{W} \left(\|c_w\|_{TV} + \frac{\lambda}{2} \langle c_w, (h - z_w)^2 \rangle \right)$$

$$C = \{c = (c_1, ..., c_W) : c_w \in \mathbb{R}^M, \sum_{w=1}^{W} c_w = 1, c_w \geq 0\}$$

This is a convex approximation of the related nonconvex functional which additionally requires the labels, c, to only take on the values zero and one.

Application of PDHGMp

$$\text{Let } \ H(c) = g_C(c) + \frac{\lambda}{2}\langle c, \sum_{w=1}^{W} \mathcal{X}_w^T (h - z_w)^2\rangle,$$

$$J(Ac) = \sum_{w=1}^{W} J_w(D\mathcal{X}_w c),$$

where $A = \begin{bmatrix} D\mathcal{X}_1 \\ \vdots \\ D\mathcal{X}_W \end{bmatrix}$, $\qquad \mathcal{X}_w c = c_w \qquad$ and

$$J_w(D\mathcal{X}_w c) = \|D\mathcal{X}_w c\|_E = \|Dc_w\|_E = \|c_w\|_{TV}.$$

PDHGMp iterations:

$$c^{k+1} = \Pi_C\left(c^k - \alpha \sum_{w=1}^{W} \mathcal{X}_w^T (D^T(2p_w^k - p_w^{k-1}) + \frac{\lambda}{2}(h - z_w)^2)\right)$$

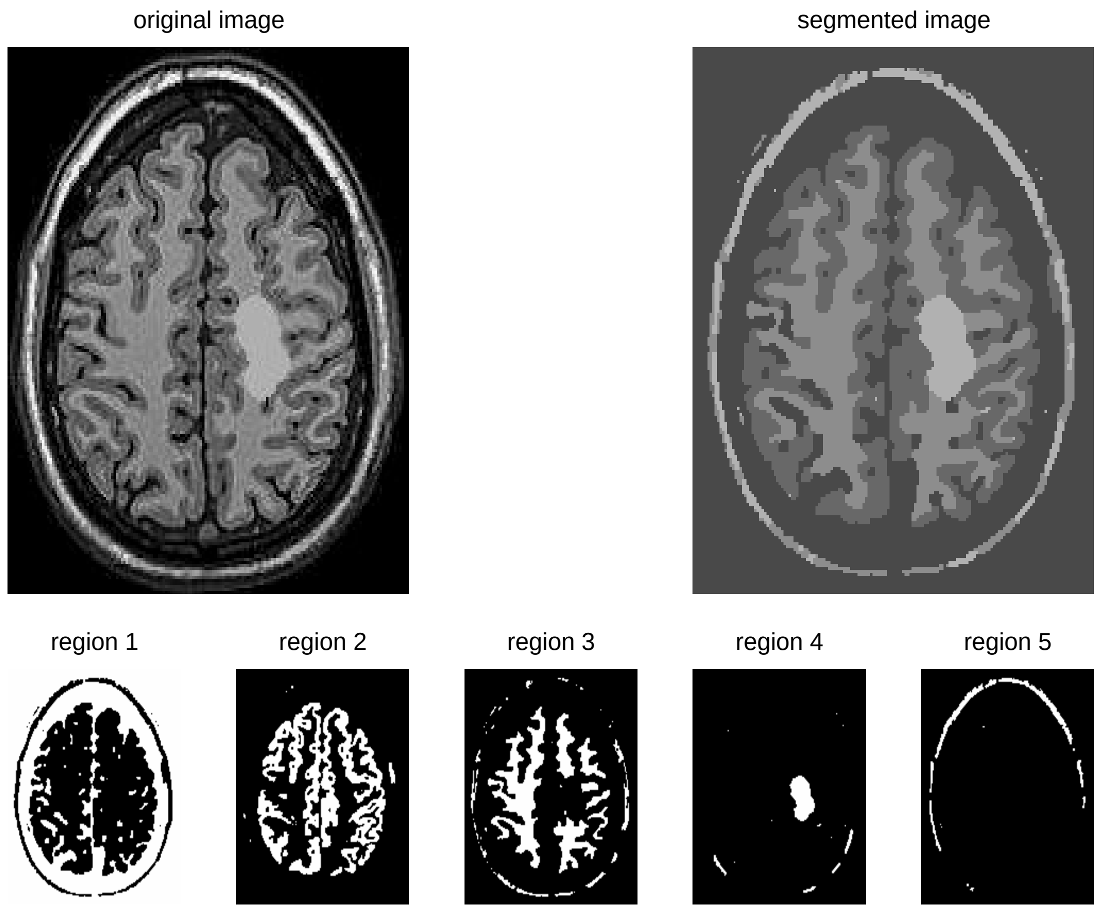
$$p_w^{k+1} = \Pi_X\left(p_w^k + \delta D\mathcal{X}_w c^{k+1}\right) \qquad \text{for } w = 1, \ldots, W.$$

Segmentation Numerical Result

$$\lambda = .0025 \qquad z = \begin{bmatrix} 75 & 105 & 142 & 178 & 180 \end{bmatrix}$$

$$\alpha = \delta = \frac{.995}{\sqrt{40}}$$

Threshold c when each $\|c_w^{k+1} - c_w^k\|_\infty < .01$ (150 iterations)



Segmentation of Brain Image Into 5 Regions

ADMM for Special Case of NMF

Nonnegative matrix factorization (NMF): Given nonnegative $X \in \mathbb{R}^{m \times d}$
Find nonnegative matrices $A \in \mathbb{R}^{m \times n}$ and $S \in \mathbb{R}^{n \times d}$ such that $X \approx AS$

- NMF is a very ill-posed problem

Additional assumptions:

- Assume columns of dictionary A come from data X
- Possibly additional assumptions about S (ie: sparsity)

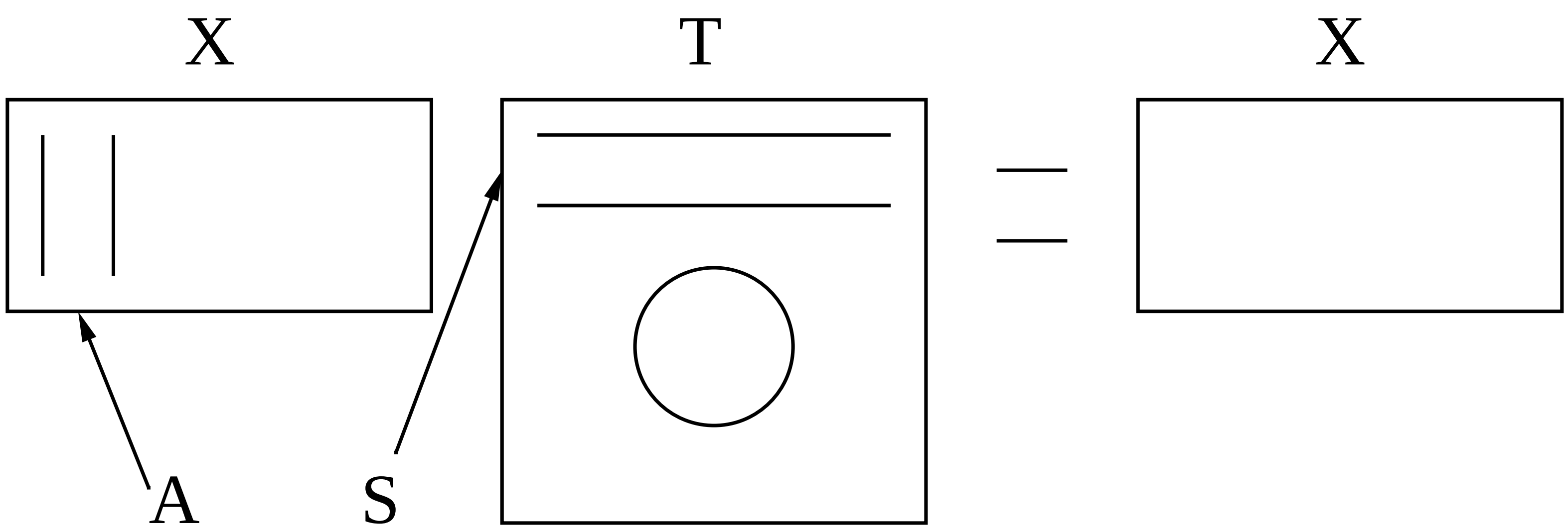
Geometric interpretation: Find a small number of columns of X that span a cone containing most of the data

Our General Strategy

Let I index the columns of X that cannot be written as nonnegative linear combinations of the other columns. Any column X_j in X can be written as

$$X_j = \sum_{i \in I} X_i T_{i,j} \qquad \text{for } T_{i,j} \geq 0$$

Our Strategy: Find a nonnegative matrix T such that $XT = X$ and as many rows of T as possible are zero.



$XT = X$ with the index set I appearing first

ADMM for Solving Convex Model

Apply ADMM to solve

$$\min_{T \geq 0} \zeta \sum_i \max_j (T_{i,j}) + \langle \sigma, T \rangle + \frac{\beta}{2} \|(XT - X)\|_F^2$$

by finding a saddle point of

$$\begin{aligned} L_\delta(T, Z, P) = g_{\geq 0}(T) + \zeta \sum_i \max_j (T_{i,j}) + \langle \sigma, T \rangle \\ + \frac{\beta}{2} \|(XZ - X)\|_F^2 + \langle P, Z - T \rangle + \frac{\delta}{2} \|Z - T\|_F^2, \end{aligned}$$

where $g_{\geq 0}$ is an indicator function for the $T \geq 0$ constraint and $\delta > 0$.

E. ESSER, M. MOLLER, S. OSHER, G. SAPIRO, J. XIN, *A convex model for non-negative matrix factorization and dimensionality reduction on physical space*, UCLA CAM Report [11-06], 2011.

Application of ADMM

Initialize T^0 and P^0 and then iterate

$$Z^{k+1} = \arg\min_Z \langle P^k, Z\rangle + \frac{\beta}{2}\|(XZ - X)\|_F^2 + \frac{\delta}{2}\|Z - T^k\|_F^2$$

$$T^{k+1} = \arg\min_T g_{\geq 0}(T) + \zeta \sum_i \|T_i\|_\infty + \langle \sigma, T\rangle - \langle P^k, T\rangle + \frac{\delta}{2}\|T - Z^{k+1}\|_F^2$$

$$P^{k+1} = P^k + \delta(Z^{k+1} - T^{k+1})$$

- This converges for any $\delta > 0$ if a saddle point exists.
- Each minimization step is straightforward to compute.

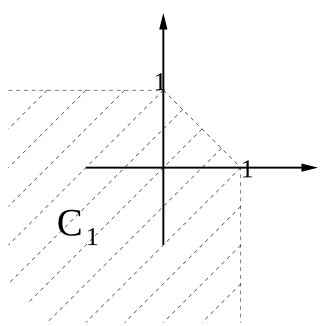
Solving for T Update

Note that the T update can be computed one row at a time.

Let $J(T_i) = g_{\geq 0}(T_i) + \zeta \max_j(T_{i,j})$.

$$\begin{aligned}
J^*(Q) &= \sup_{T_i} \langle Q, T_i \rangle - g_{\geq 0}(T_i) - \zeta \max_j (T_{i,j}) \\
&= \sup_{T_i} (\max_j (T_{i,j}))(\|\max(Q,0)\|_1 - \zeta) \\
&= \begin{cases} 0 & \text{if} \quad \|\max(Q,0)\|_1 \leq \zeta \\ \infty & \text{otherwise} \end{cases}
\end{aligned}$$

Let C_ζ denote the convex set $C_\zeta = \{Q \in \mathbb{R}^{d_s} : \|\max(Q,0)\|_1 \leq \zeta\}$



The Update Formulas

Using the Moreau decomposition, we can write the T update in terms of an orthogonal projection onto the convex set C.

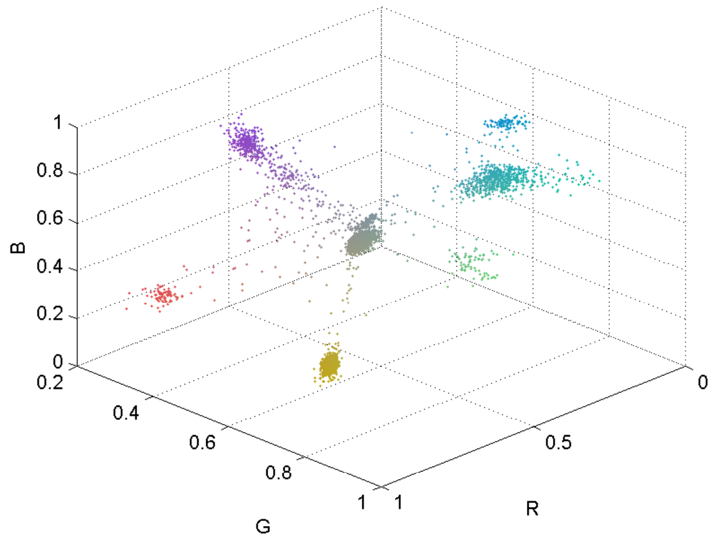
$$Z_j^{k+1} = (\beta X^T X + \delta I)^{-1}(\beta X^T X_j + \delta T_j^k - P_j^k)$$

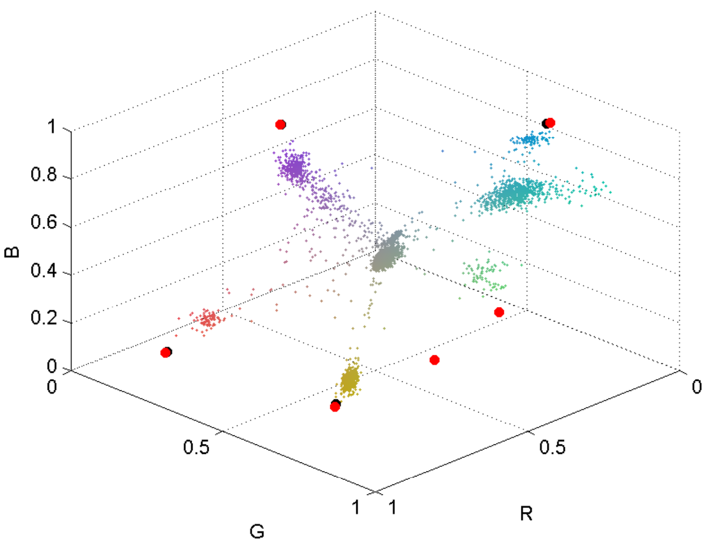
$$T^{k+1} = Z^{k+1} + \frac{P^k}{\delta} - \frac{\sigma}{\delta} - \Pi_{C_{\frac{\zeta}{\delta}}}(Z^{k+1} + \frac{P^k}{\delta} - \frac{\sigma}{\delta})$$

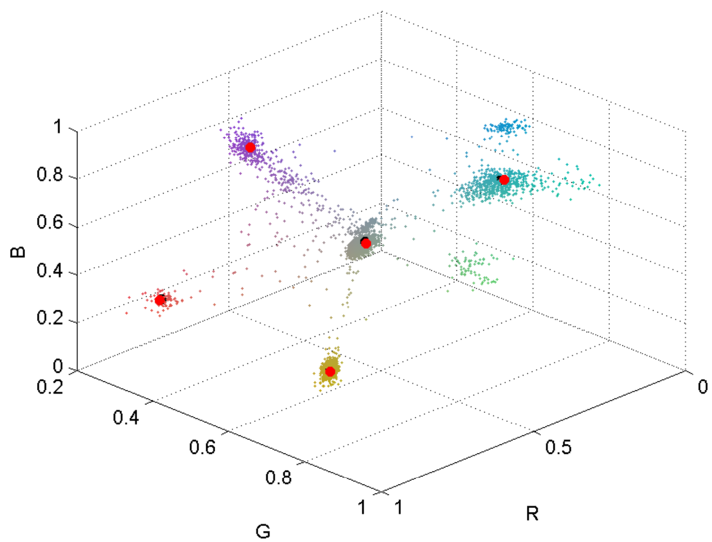
$$P^{k+1} = P^k + \delta(Z^{k+1} - T^{k+1})$$

Note: The projection for each row of the T update can be computed with complexity $O(d \log d)$

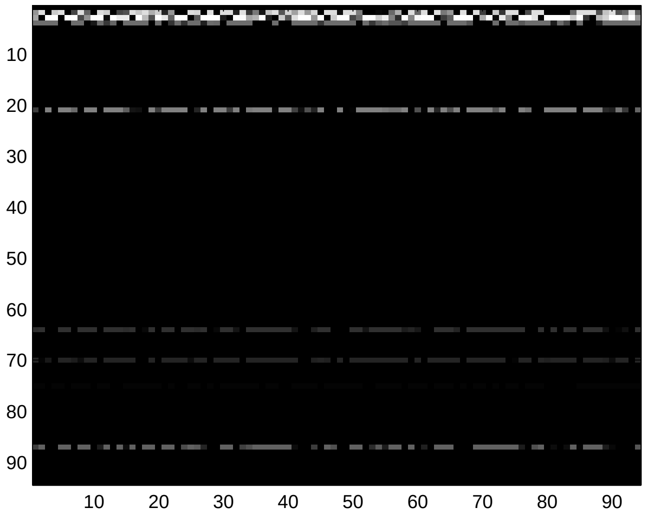
RGB Visualization (w/o and with sparsity)

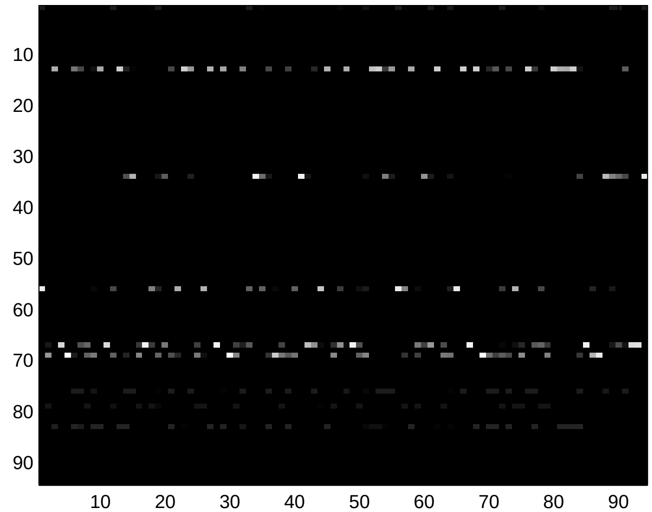




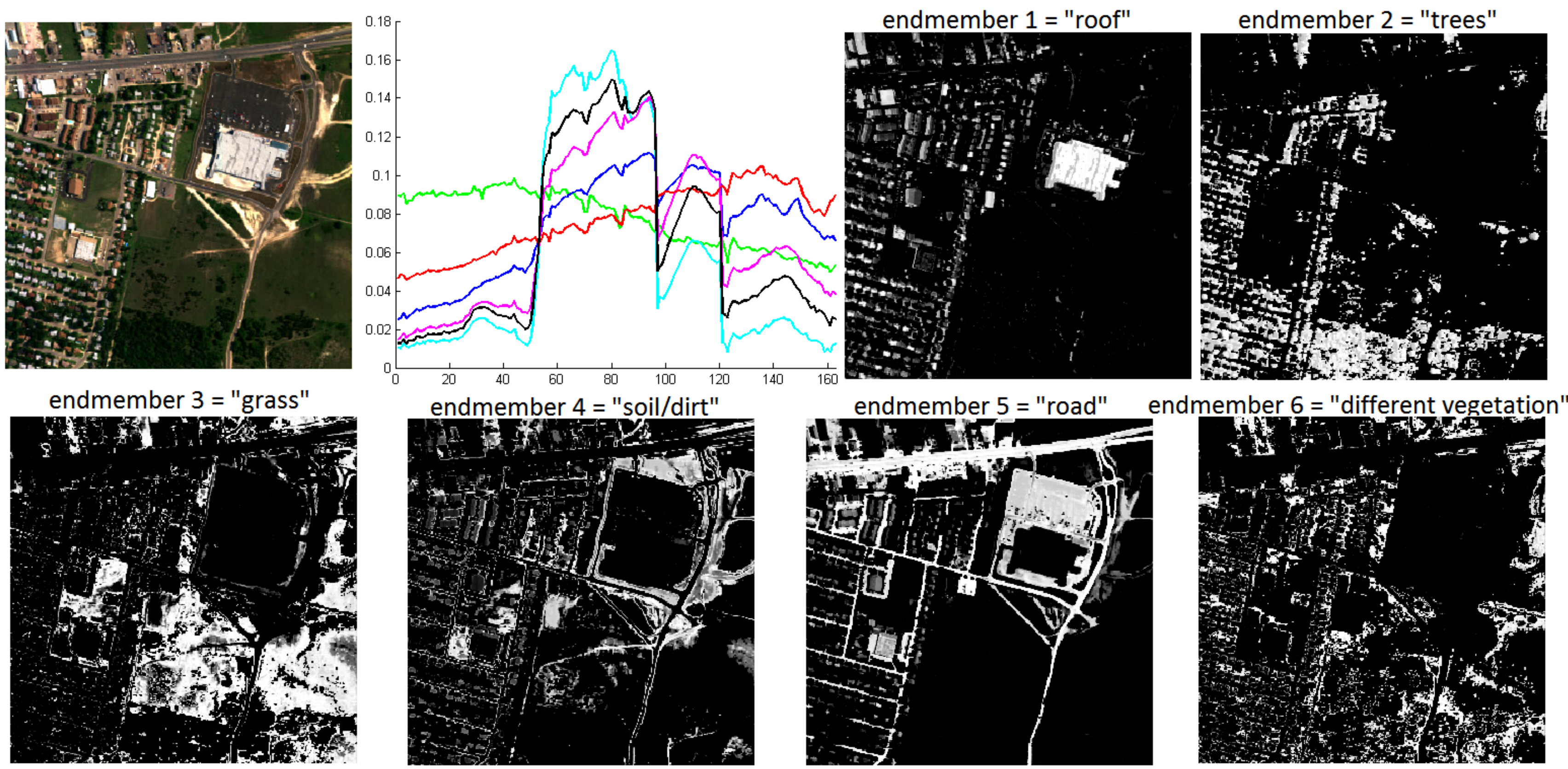


Abundance Matrix Comparison





Application to Urban Hyperspectral image



Conclusions

- The primal dual methods discussed here are practical for many of the large scale, non-differentiable convex minimization problems that arise in image processing, computer vision and elsewhere.
  - They have simple iterations
  - They converge under minimal assumptions
  - They can take advantage of separable structure

- The large amount of recent work in the imaging science literature about these and related methods demonstrates their usefulness in this area.

- Recent and ongoing work is
  - Generalizing applicability
  - Improving convergence rates
  - Even successfully applying to some nonconvex problems