A General Framework for a Class of First Order Primal-Dual Algorithms for TV Minimization

Ernie Esser (UCI)

Joint work with Xiaoqun Zhang (UCLA) and Tony Chan (HKUST)

4-14-2010

Outline

- Motivation to Study Primal Dual Methods
- A Model Convex Minimization Problem
- The Primal Dual Hybrid Gradient (PDHG) Method
- Primal, Dual and Saddle Point Formulations
- PDHG-Based Framework for a Class of Primal Dual Methods
  - Proximal Forward Backward Splitting (PFBS) $\leftrightarrow$ Alternating Minimization Algorithm (AMA)
  - Alternating Direction Method of Multipliers (ADMM) $\leftrightarrow$ Split Bregman $\leftrightarrow$ Douglas Rachford Splitting
  - Split Inexact Uzawa (SIU) $\leftrightarrow$ Modified PDHG (PDHGM)
- Comparison of Algorithms
- Conclusions

Motivation to Study Primal Dual Algorithms

- Split Bregman and PDHG, both primal-dual methods, demonstrated clear potential to be significantly more efficient than previous methods used to solve convex models in image processing, but convergence properties were initially unclear
- Overwhelming number of seemingly related methods, but with often unclear connections
- Need for methods with simple, explicit iterations capable of solving large scale, often non-differentiable convex models

A Model Convex Minimization Problem

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

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)

The PDHG Method

$$\min_u J(Au) + H(u) \qquad (P)$$

from definition of convex conjugate

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

Saddle Point Formulation:

biconjugate

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

Interpret PDHG as primal-dual proximal point method:

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

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

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.

Dual Problem and Strong Duality

The dual problem is

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

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

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

- Having assumed u^* is an optimal solution to (P), it follows that 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}

Ref: R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

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

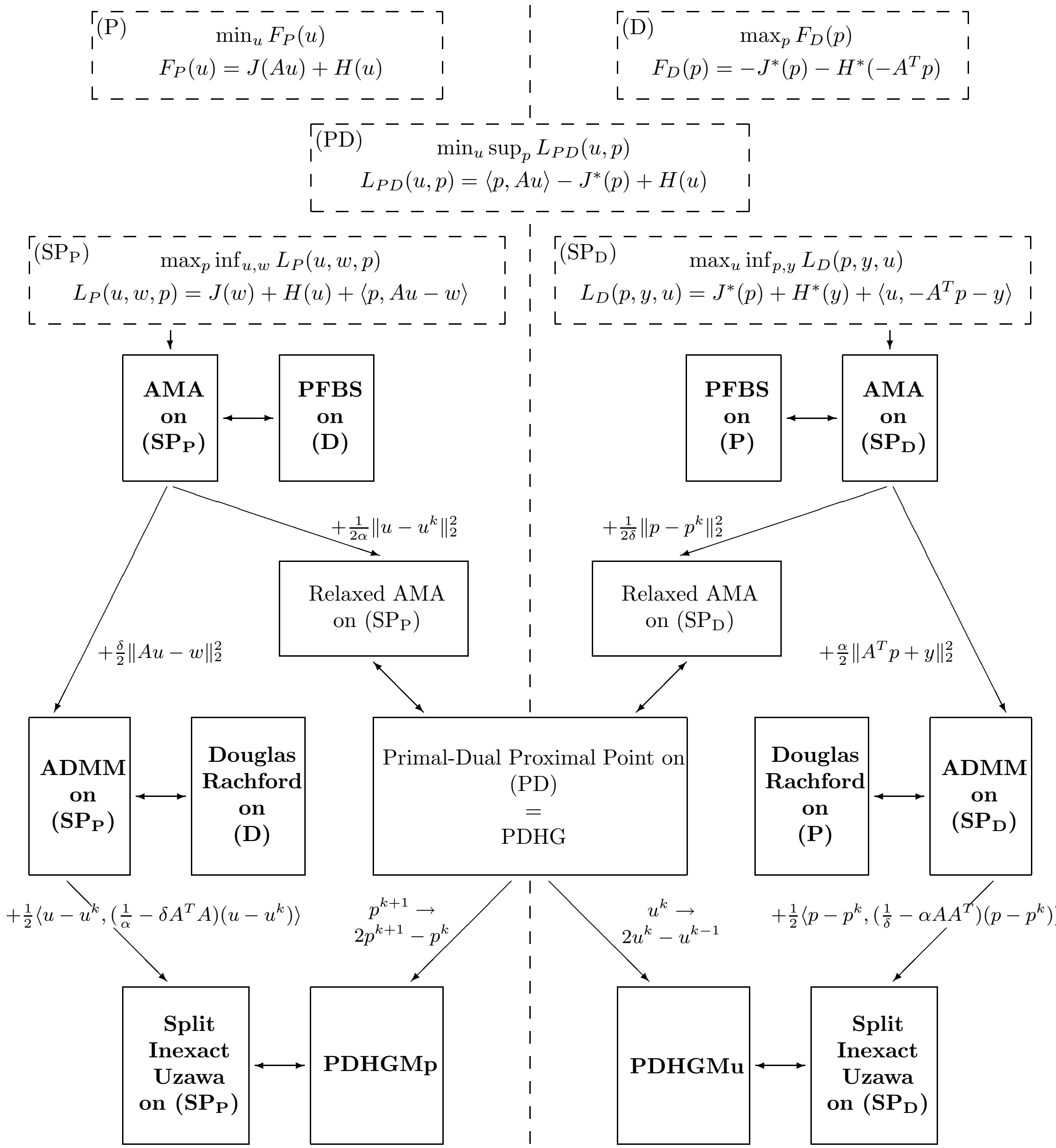
$$\max_{p \in \mathbb{R}^n} \inf_{u \in \mathbb{R}^m, w \in \mathbb{R}^n} L_P(u, w, p) \qquad (SPP)$$

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

Moreau Decomposition

Moreau's decomposition will be a main tool in demonstrating connections between the algorithms that follow.

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

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

J. J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93, 1965, pp. 273-299.

PFBS on (D)

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

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

Ref: P. Combettes and W. Wajs, *Signal Recovery by Proximal Forward-Backward Splitting*, Multiscale Modeling and Simulation, 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.

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

Ref: 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):

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

PFBS on (D) and AMA on (SPP) already 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) with additional proximal penalty,

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

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

AMA applied to (SPP) with first step relaxed by same additional penalty,

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

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

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

- PFBS on (P) and AMA on (SPD) are connected to PDHG in the analogous way.

AMA Connection to ADMM

ADMM applied to (SPP):

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

Ref: J. ECKSTEIN, AND D. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program. 55, 1992.

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.

Example: Douglas Rachford splitting on (D) with $z^k = p^k + \delta w^k$:

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

Ref: S. SETZER, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, http://kiwi.math.uni-mannheim.de/˜ssetzer/pub/setzer_fba_fbs_frames08.pdf, 2008.

Ref: J. ECKSTEIN, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph. D. Thesis, Massachusetts Institute of Technology, Dept. of Civil Engineering, http://hdl.handle.net/1721.1/14356, 1989.

Split Inexact Uzawa Method

Consider adding $\frac{1}{2}\langle u - u^k, (\frac{1}{\alpha} - \delta A^T A)(u - u^k)\rangle$ to the first step of the Alternating Direction Method of Multipliers (ADMM) applied to (SPP), with $0 < \alpha < \frac{1}{\delta\|A\|^2}$.

Split Inexact Uzawa applied to (SPP):

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

Note: In general we could similarly modify both minimization steps in ADMM, but by only modifying the first step we can obtain an interesting PDHG-like interpretation.

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

Equivalence to Modified PDHG (PDHGMp)

Split Inexact Uzawa applied to (SPP):

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

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

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

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

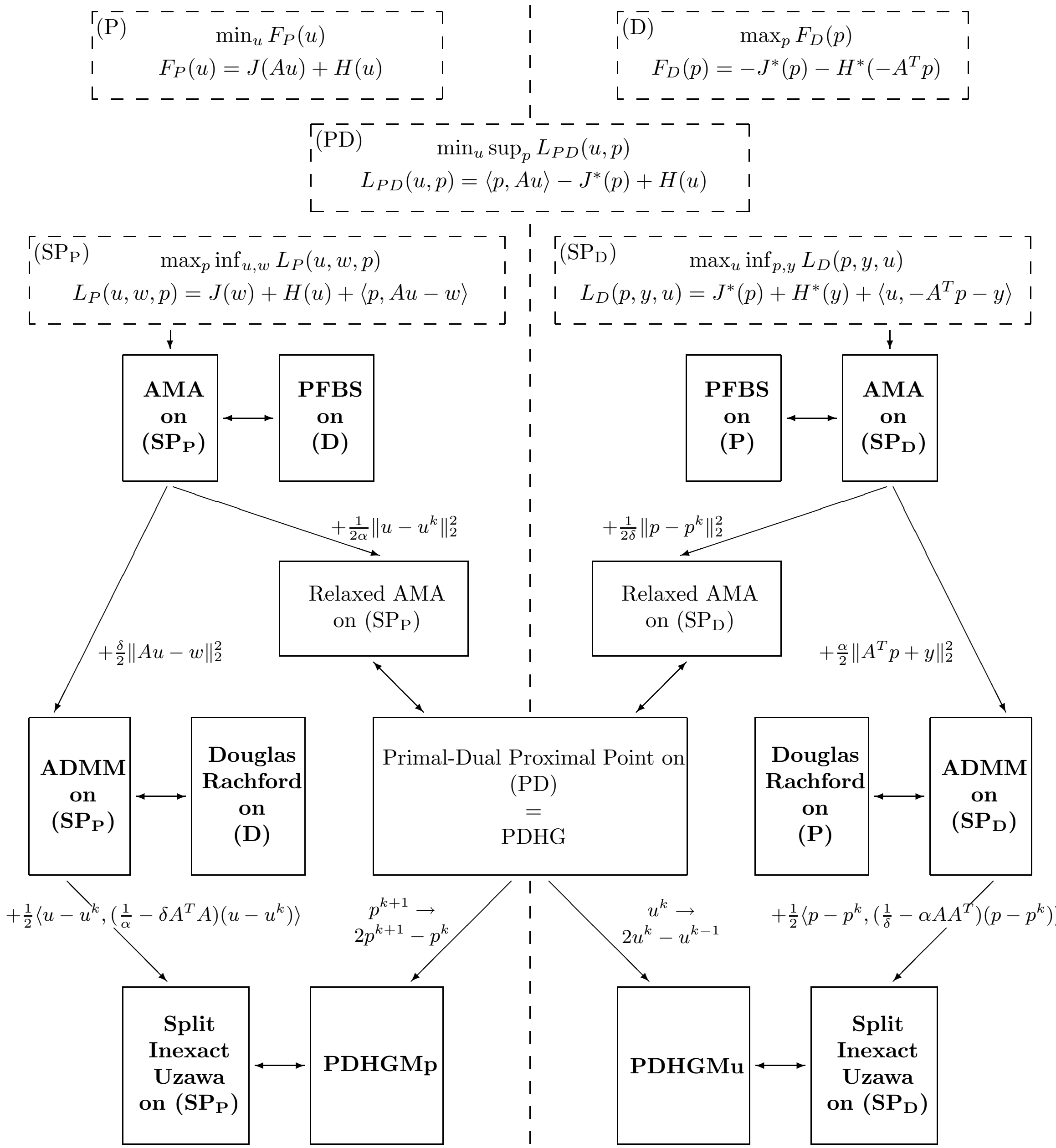
$$p^{k+1} = (p^k + \delta Au^{k+1}) - \delta \arg\min_{w} J(w) + \frac{\delta}{2}\|w - \frac{(p^k + \delta Au^{k+1})}{\delta}\|^2$$

and apply Moreau's decomposition.

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

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

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



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

More Related Works

- Ref: T. GOLDSTEIN AND S. OSHER, *The Split Bregman Algorithm for L1 Regularized Problems*, SIIMS, Vol. 2, No. 2, 2008.
- Ref: S. SETZER, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, LNCS, 2008.
- G. CHEN AND M. TEBOULLE, *A Proximal-Based Decomposition Method for Convex Minimization Problems*, Mathematical Programming, Vol. 64, 1994.
- X. ZHANG, M. BURGER, AND S. OSHER, *A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration*, UCLA CAM Report [09-99], 2009.
- Ref: 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.
- Ref: J. ECKSTEIN, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph. D. Thesis, MIT, Dept. of Civil Engineering, 1989.
- Ref: C. WU AND X.C TAI, *Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and High Order Models*, 2009.
- Ref: E. ESSER, X. ZHANG, AND T. F. CHAN, *A General Framework for a Class of First Order Primal-Dual Algorithms for TV Minimization*, UCLA CAM Report [09-67], 2009.

Comparison of Algorithms

Assume J and H are closed proper convex functions

- PFBS on (D) convergence requires H^* differentiable, $\nabla(H^*(-A^Tp))$ Lipschitz continuous with Lipschitz constant equal to $\frac{1}{\beta}$, and $0 < \inf \delta_k \leq \sup \delta_k < 2\beta$
  AMA on (SPP) convergence analogously requires H strongly convex
  Variables not coupled by matrix A
- ADMM on (SPP) convergence requires $\delta > 0$ and $H(u) + \|Au\|^2$ to be strictly convex (ensures convergence to saddle point, not needed for Douglas Rachford)
  Variables coupled by A
- SIU on (SPP) and PDHGMp convergence requires $0 < \delta < \frac{1}{\alpha\|A\|^2}$
  Variables not coupled by A
- PDHG convergence in general is still an open problem
  Variables not coupled by A

Application to TV Minimization Problems

Discretize $\|u\|_{TV}$ using forward differences and assuming Neumann BC

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

Vectorize $M_r \times M_c$ matrix by stacking columns

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

To compare numerical performance on TV denoising,

$$\min_u \|u\|_{TV} + \frac{\lambda}{2}\|u - f\|_2^2,$$

first let $A = D$, $J(Au) = \|Du\|_E$ and $H(u) = \frac{\lambda}{2}\|u - f\|_2^2$ to write the model in form of

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

Original, Noisy and Benchmark Images

Use 256×256 cameraman image.

Add white Gaussian noise having standard deviation 20.

Let $\lambda = .053$.

Iterations Required for TV Denoising

Algorithm	tol $= 10^{-2}$	tol $= 10^{-4}$	tol $= 10^{-6}$
PDHG (adaptive)	14	70	310
PDHGMu (adaptive)	19	92	365
PDHG $\alpha = 5, \delta = .025$	31	404	8209
PDHG $\alpha = 1, \delta = .125$	51	173	1732
PDHG $\alpha = .2, \delta = .624$	167	383	899
PDHGMu $\alpha = 5, \delta = .025$	21	394	8041
PDHGMu $\alpha = 1, \delta = .125$	38	123	1768
PDHGMu $\alpha = .2, \delta = .624$	162	355	627
PDHG $\alpha = 5, \delta = .1$	22	108	2121
PDHG $\alpha = 1, \delta = .5$	39	123	430
PDHG $\alpha = .2, \delta = 2.5$	164	363	742
PFBS $\delta = .0132$	48	750	15860
ADMM $\delta = .025$	17	388	7951
ADMM $\delta = .125$	22	100	1804
ADMM $\delta = .624$	97	270	569

Conclusions

- The PDHG-related framework for a class of primal-dual algorithms hopefully reduces the dauntingly large space of potential methods by showing some close connections between them.
- These primal dual algorithms (with the exception of PDHG) converge under few assumptions.
- The methods discussed can be efficiently applied to convex models with separable structure which can be written in the form of (P). They are practical for many large scale convex optimization problems that arise in image processing, including total variation minimization.