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) ↔ Alternating Minimization Algorithm (AMA)
  • Alternating Direction Method of Multipliers (ADMM) ↔ Split Bregman ↔ Douglas Rachford Splitting
  • Split Inexact Uzawa (SIU) ↔ 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

\[
\min_{u \in \mathbb{R}^m} J(Au) + H(u) \quad (P)
\]

\(J, H\) closed proper convex

\(H : \mathbb{R}^m \to (\mathbb{R}, \infty]\)

\(J : \mathbb{R}^n \to (\mathbb{R}, \infty]\)

\(A \in \mathbb{R}^{n \times m}\)

Assume there exists an optimal solution \(u^*\) to \((P)\)
The PDHG Method

\[
\begin{align*}
\min_u J(Au) + H(u) & \quad (P) \\
J(Au) = J^{**}(Au) = \sup_p \langle p, Au \rangle - J^*(p)
\end{align*}
\]

Saddle Point Formulation:

\[
\begin{align*}
\min_u \sup_p -J^*(p) + \langle p, Au \rangle + H(u) & \quad (PD)
\end{align*}
\]

Interpret PDHG as primal-dual proximal point method:

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

Dual Problem and Strong Duality

The dual problem is

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

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 $L_{PD}$

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

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

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,

$$\max_{u \in \mathbb{R}^m} \inf_{p \in \mathbb{R}^n, y \in \mathbb{R}^m} L_D(p, y, u) \quad (SPD)$$
\[
\begin{align*}
\{ \text{(P)} \} & \quad \min_u F_P(u) \\
& \quad F_P(u) = J(Au) + H(u) \\
\{ \text{(D)} \} & \quad \max_p F_D(p) \\
& \quad F_D(p) = -J^*(p) - H^*(-A^T p) \\
\{ \text{(PD)} \} & \quad \min_u \sup_p L_{PD}(u, p) \\
& \quad L_{PD}(u, p) = \langle p, Au \rangle - J^*(p) + H(u) \\
\{ \text{(SP)} \} & \quad \max_u \inf_w L_{P}(w, u) \\
& \quad L_P(w, u) = J(w) + H(u) + \langle p, w \rangle \\
\end{align*}
\]

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

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

PFBS on (D)

PFBS alternates a gradient descent step with a proximal step:

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

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

Since \( u^{k+1} = \nabla H^*(-A^T p^k) \iff -A^T p^k \in \partial H(u^{k+1}) \), 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

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

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

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 $p$.

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

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

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

Equivalence by Moreau Decomposition

AMA applied to (SPP):

\[ u^{k+1} = \text{arg min}_{u \in \mathbb{R}^m} H(u) + \langle A^T p^k, u \rangle \]

\[ w^{k+1} = \text{arg min}_{w \in \mathbb{R}^n} J(w) - \langle p^k, w \rangle + \frac{\delta_k}{2} \| A u^{k+1} - w \|^2_2 \]

\[ p^{k+1} = p^k + \delta_k (A u^{k+1} - w^{k+1}) \]

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

Combining the last two steps of AMA yields

\[ p^{k+1} = (p^k + \delta_k A u^{k+1}) - \delta_k \text{arg min}_{w} J(w) + \frac{\delta_k}{2} \left\| w - \frac{(p^k + \delta_k A u^{k+1})}{\delta_k} \right\|^2_2, \]

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

\[ p^{k+1} = \text{arg min}_{p} J^*(p) + \frac{1}{2\delta_k} \| p - (p^k + \delta_k A u^{k+1}) \|^2_2 \]
AMA/PFBS Connection to PDHG

PFBS on (D) with additional proximal penalty,

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

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

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

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

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

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

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

ADMM applied to (SPP):

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

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

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

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 $z^k = p^k + \delta w^k$:

\[
q^{k+1} = \arg \min_q H^*(-A^Tq) + \frac{1}{2\delta} \|q + z^k - 2p^k\|^2
\]

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

\[
z^{k+1} = z^k + q^{k+1} - p^k
\]

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):

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

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

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

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.

Equivalence to Modified PDHG (PDHGMp)

Split Inexact Uzawa applied to (SPP):

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

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

\[
p^{k+1} = (p^k + \delta Au^{k+1}) - \delta \arg \min_w J(w) + \frac{\delta}{2} \| w - \left( \frac{p^k + \delta Au^{k+1}}{\delta} \right) \|^2_2
\]

and apply Moreau’s decomposition.

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

\[
\begin{align*}
u^{k+1} &= \arg \min_{u \in \mathbb{R}^m} H(u) + \langle A^T (2p^k - p^{k-1}), u \rangle + \frac{1}{2\alpha} \| u - u^k \|^2_2 \\
p^{k+1} &= p^k + \delta (Au^{k+1})
\end{align*}
\]

\[
p^{k+1} = \arg \min_{p \in \mathbb{R}^n} J^*(p) - \langle p, Au^{k+1} \rangle + \frac{1}{2\delta} \| p - p^k \|^2_2
\]
\[
\begin{align*}
(P) & \quad \min_u F_P(u) \\
& \quad F_P(u) = J(Au) + H(u)
\end{align*}
\]
\[
\begin{align*}
(D) & \quad \max_p F_D(p) \\
& \quad F_D(p) = -J^*(p) - H^*(-A^T p)
\end{align*}
\]
\[
\begin{align*}
(PD) & \quad \min_u \sup_p L_{PD}(u, p) \\
& \quad L_{PD}(u, p) = \langle p, Au \rangle - J(u) - H(u)
\end{align*}
\]
\[
\begin{align*}
(SP_P) & \quad \max_y \inf_u \inf_w L_P(u, w, p) \\
& \quad L_P(u, w, p) = J(u) + H(w) + \langle p, Au - w \rangle
\end{align*}
\]
\[
\begin{align*}
(SP_D) & \quad \max_u \inf_p \inf_y L_D(p, y, u) \\
& \quad L_D(p, y, u) = J^*(p) + H^*(y) + \langle u, A^T p - y \rangle
\end{align*}
\]

Legend: (P): Primal
(D): Dual
(PD): Primal-Dual
(SP_P): Split Primal
(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


Comparison of Algorithms

Assume $J$ and $H$ are closed proper convex functions

• PFBS on (D) convergence requires $H^*$ differentiable, $\nabla(H^*(-A^T p))$ Lipschitz continuous with Lipschitz constant equal to $\frac{1}{\beta}$, and

$0 < \inf \delta_k \leq \sup \delta_k < \frac{2}{\beta}$

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

$$\min_{u \in \mathbb{R}^m} J(Au) + H(u) \quad (P)$$
Original, Noisy and Benchmark Images

Use $256 \times 256$ cameraman image.

Add white Gaussian noise having standard deviation 20.

Let $\lambda = .053$. 
## Iterations Required for TV Denoising

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