Sparse Optimization
Lecture: Operator Splitting, Prox-Linear, ADMM

Instructor: Wotao Yin
Department of Mathematics, UCLA

July 2013

online discussions on piazza.com

Those who complete this lecture will know

- the motivation of operator splitting
- basic operator splitting approaches: forward-backward, Peaceman-Rachford, Douglas-Rachford
- the prox-linear, ADMM methods

Recall: dual explicit/implicit update

Primal problem:

$$\min f(\mathbf{x}) \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

Definitions:

- Lagrangian $\mathcal{L}(\mathbf{x}; \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$
- Augmented Lagrangian $\mathcal{L}_A(\mathbf{x}; \mathbf{y}, c) = \mathcal{L}(\mathbf{x}; \mathbf{y}) + \frac{c}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$

Dual objective:

$$d(\mathbf{y}) = -\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \mathbf{y}).$$

Dual problem:

$$\min_{\mathbf{y}} d(\mathbf{y}).$$

Recall: dual explicit/implicit update

Dual explicit (sub)gradient iteration:

$$\mathbf{y}^{k+1} = \mathbf{y}^k - c\nabla d(\mathbf{y}^k) \quad \text{or} \quad \mathbf{y}^k - c\mathbf{g}, \text{ where } \mathbf{g} \in \partial d(\mathbf{y}^k).$$

Implementation:

1. $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \mathbf{y}^k)$;
2. $\mathbf{y}^{k+1} = \mathbf{y}^k - c(\mathbf{b} - \mathbf{A}\mathbf{x}^{k+1})$.

Implicit (sub)gradient iteration:

$$\mathbf{y}^{k+1} = \mathbf{prox}_{cd}\mathbf{y}^k$$

Implementation:

1. $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}_{\mathbf{A}}(\mathbf{x}; \mathbf{y}^k, c)$;
2. $\mathbf{y}^{k+1} = \mathbf{y}^k - c(\mathbf{b} - \mathbf{A}\mathbf{x}^{k+1})$.

The implicit iteration is more stable; “step size” c does not need to diminish. However, it is also more expensive.

Separable objective function and Lagrange dual

Consider a convex program with a *separable objective* and *coupling constraints*

$$\min_{\mathbf{x},\mathbf{z}} f(\mathbf{x}) + g(\mathbf{z}) \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{b}.$$

Examples:

- $\min f(\mathbf{x}) + g(\mathbf{x}) \implies \min_{\mathbf{x},\mathbf{y}}\{f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{x} - \mathbf{z} = 0\}$
- $\min f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) \implies \min_{\mathbf{x},\mathbf{y}}\{f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{A}\mathbf{x} - \mathbf{z} = 0\}$
- $\min\{f(\mathbf{x}) : \mathbf{A}\mathbf{x} \in \mathcal{C}\} \implies \min_{\mathbf{x},\mathbf{y}}\{f(\mathbf{x}) + \iota_{\mathcal{C}}(\mathbf{z}) : \mathbf{A}\mathbf{x} - \mathbf{z} = 0\}$
- $\min \sum_{i=1}^N f_i(\mathbf{x}) \implies \min_{\{\mathbf{x}_i\},\mathbf{z}}\{\sum_{i=1}^N f_i(\mathbf{x}_i) : \mathbf{x}_i - \mathbf{z} = 0, \forall i\}$

  note: $\mathbf{x}_i$ is a copy of $\mathbf{x}$ for f_i; it is not a subvector of $\mathbf{x}$.

Separable objective function and Lagrange dual

Consider

$$\min_{\mathbf{x},\mathbf{z}} f(\mathbf{x}) + g(\mathbf{z}) \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = b.$$

Lagrangian relaxation $\quad \mathcal{L}(\mathbf{x}, \mathbf{z}; \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{b})$

Dual problem $\quad d(\mathbf{y}) := -\min_{\mathbf{x},\mathbf{z}} \mathcal{L}(\mathbf{x}, \mathbf{z}; \mathbf{y})$

Subproblem in the explicit (sub)gradient iteration ($\mathbf{x}$ and $\mathbf{z}$ are *decoupled*):

$$(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \xleftarrow{\text{solve}} \min_{\mathbf{x},\mathbf{z}} \mathcal{L}(\mathbf{x}, \mathbf{z}; \mathbf{y}^k)$$

Subproblem in the implicit (sub)gradient iteration:

$$(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \xleftarrow{\text{solve}} \min_{\mathbf{x},\mathbf{z}} \mathcal{L}_{\mathbf{A}}(\mathbf{x}, \mathbf{z}; \mathbf{y}^k)$$

Issue: $\mathcal{L}_{\mathbf{A}}(\mathbf{x}, \mathbf{z}; \mathbf{y}^k)$ contains term $\frac{c}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{b}\|^2$, so $\mathbf{x}$ and $\mathbf{z}$ are *coupled*

Question: can we decouple $\mathbf{x}$ and $\mathbf{z}$ in the implicit iteration?

Separable objective function and Lagrange dual

$\mathcal{L}(\mathbf{x}, \mathbf{z}; \mathbf{y})$ is separable

$$\mathcal{L}(\mathbf{x}, \mathbf{z}; \mathbf{y}) = \underbrace{f(\mathbf{x}) + \mathbf{y}^T \mathbf{A}\mathbf{x}}_{\mathcal{L}_1(\mathbf{x};\mathbf{y})} + \underbrace{g(\mathbf{z}) + \mathbf{y}^T B\mathbf{z} - \mathbf{y}^T \mathbf{b}}_{\mathcal{L}_2(\mathbf{z};\mathbf{y})}$$

Define

$$d_1(\mathbf{y}) := -\min_{\mathbf{x}} \mathcal{L}_1(\mathbf{x}; \mathbf{y}),$$

$$d_2(\mathbf{y}) := -\min_{\mathbf{z}} \mathcal{L}_2(\mathbf{z}; \mathbf{y}),$$

Therefore

$$d(\mathbf{y}) = d_1(\mathbf{y}) + d_2(\mathbf{y}).$$

We shall obtain $\mathbf{y}^*$ such that

$$0 \in \partial d_1(\mathbf{y}^*) + \partial d_2(\mathbf{y}^*)$$

then, recover $\mathbf{x}^*$ and $\mathbf{y}^*$ (conditions?)

$$\mathbf{x}^* \xleftarrow{\text{solve}} \min_{\mathbf{x}} \mathcal{L}_1(\mathbf{x}; \mathbf{y}^*),$$

$$\mathbf{z}^* \xleftarrow{\text{solve}} \min_{\mathbf{z}} \mathcal{L}_2(\mathbf{z}; \mathbf{y}^*).$$

Operator splitting

Assume d_1 is differentiable and d_2 is sub-differentiable

- **forward-backward splitting** (FBS)

$$
\begin{aligned}
0 \in \nabla d_1(\mathbf{y}) + \partial d_2(\mathbf{y}) &\iff y - c\,\nabla d_1(\mathbf{y}) \in \mathbf{y} + c\,\partial d_2(\mathbf{y}) \\
&\iff (I + c\,\partial d_2)^{-1}(I - c\nabla d_1)\mathbf{y} = \mathbf{y} \\
&\iff \mathbf{prox}_{cd_2}(I - c\nabla d_1)\mathbf{y} = \mathbf{y}
\end{aligned}
$$

- y^* minimizes $d_1(\mathbf{y}) + d_2(\mathbf{y})$ if and only if

$$
\mathbf{prox}_{cd_2}(I - c\nabla d_1)\mathbf{y}^* = \mathbf{y}^*.
$$

- $(I - c\nabla d_1)$: forward (explicit gradient) operator w.r.t. d_1,

- $\mathbf{prox}_{cd_2}$: backward (implicit gradient) operator w.r.t. d_2.

Forward-backward splitting iteration

$$\mathbf{y}^{k+1} = \mathbf{prox}_{cd_2}(I - c\nabla d_1)\mathbf{y}^k$$

At iteration k:

step 1: compute $\nabla d_1(\mathbf{y}^k)$ and apply $\mathbf{y}^{k+1/2} = (I - c\nabla d_1)\mathbf{y}^k$

step 2: compute $\mathbf{y}^{k+1} = \mathbf{prox}_{cd_2}\mathbf{y}^{k+1/2} = (I + \partial d_2(\mathbf{y}^{k+1}))^{-1}\mathbf{y}^{k+1/2}$

Primal forward-backward splitting iteration

Assume f is differentiable. Consider convex problem

$$\min_{\mathbf{x}} r(\mathbf{x}) + f(\mathbf{x})$$

Optimality condition:

$$0 \in \partial r(\mathbf{x}^*) + \nabla f(\mathbf{x}^*).$$

FB iteration:

$$\mathbf{x}^{k+1} = \mathbf{prox}_{cr}(I - c\nabla f)\mathbf{x}^k.$$

Equivalent to:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}}\ r(\mathbf{x}) + \langle \nabla f, \mathbf{x} - \mathbf{x}^k \rangle + \frac{1}{2c}\|\mathbf{x} - \mathbf{x}^k\|_2^2.$$

Other names: **prox-linear** iteration, **majorization** iteration, etc.

Dual forward-backward splitting iteration

$$\mathbf{y}^{k+1} = \mathbf{prox}_{cd_2}(I - c\nabla d_1)\mathbf{y}^k$$

Step by step (right to left):

- forward step $\mathbf{y}^{k+1/2} := (I - c\nabla d_1)\mathbf{y}^k$; details:
  - obtain $\nabla d_1(\mathbf{y}^k) = -\mathbf{A}\mathbf{x}^k$, where $\mathbf{x}^k = \arg\min_{\mathbf{x}} \mathcal{L}_1(\mathbf{x}; \mathbf{y}^k)$,
  - set $\mathbf{y}^{k+1/2} := \mathbf{y}^k + c\mathbf{A}\mathbf{x}^k$,
- backward step $\mathbf{y}^{k+1} = \mathbf{prox}_{cd_2}\mathbf{y}^{k+1/2}$; details:
  - compute $\mathbf{b} - \mathbf{B}\mathbf{z}^{k+1} \in \partial d_2(\mathbf{y}^{k+1})$, where $\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \mathcal{L}_{2\mathbf{A}}(\mathbf{z}; \mathbf{y}^{k+1/2})$
  - set $\mathbf{y}^{k+1} = \mathbf{y}^{k+1/2} + c(\mathbf{B}\mathbf{z}^{k+1} - \mathbf{b})$,
- combine the two steps

$$\mathbf{y}^{k+1} = \mathbf{y}^{k+1/2} + c(\mathbf{B}\mathbf{z}^{k+1} - \mathbf{b}) = \mathbf{y}^k + c(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{b}).$$

Dual forward-backward splitting iteration

Simplified iteration

1. $\mathbf{x}^k \xleftarrow{\text{solve}} \min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{y}^k, \mathbf{A}\mathbf{x} \rangle$,
2. $\mathbf{z}^{k+1} \xleftarrow{\text{solve}} \min_{\mathbf{z}} g(\mathbf{z}) + \langle \mathbf{y}^k, \mathbf{B}\mathbf{z} \rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z} - \mathbf{b}\|^2$,
3. $\mathbf{y}^{k+1} = \mathbf{y}^k + c(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{b})$.

f and g appear in different subproblems, sometimes significantly simplifies the computation.

note: strictly convex f is required for unique $\mathbf{x}^k$ and differentiable d_1

question: how to deal with non-differentiable d_1

Peaceman-Rachford splitting

For now, still assume d_1 is differentiable.

Peaceman-Rachford splitting (PRS) extends FBS

$$
\begin{aligned}
0 \in \nabla d_1(\mathbf{y}) + \partial d_2(\mathbf{y}) &\Leftrightarrow (I + c\,\partial d_2)^{-1}(I - c\nabla d_1)\mathbf{y} = \mathbf{y} \\
&\Leftrightarrow (I + c\,\partial d_2)^{-1}(I - c\nabla d_1)\mathbf{y} + c\,\nabla d_1(\mathbf{y}) = \mathbf{y} + c\,\nabla d_1(\mathbf{y}) \\
&\Leftrightarrow (I + c\,\nabla d_1)^{-1}\left((I + c\,\partial d_2)^{-1}(I - c\nabla d_1) + c\,\nabla d_1\right)\mathbf{y} = \mathbf{y} \\
&\Leftrightarrow \mathbf{prox}_{cd_1}\left(\mathbf{prox}_{cd_2}(I - c\nabla d_1) + c\,\nabla d_1\right)\mathbf{y} = \mathbf{y}
\end{aligned}
$$

PRS uses implicit d_1 and d_2.

FBS uses only implicit d_2.

In fact, d_1 can be sub-differential (once we eliminate ∇d_1 from the splitting)

Eliminating ∇d_1

Introduce $\mathbf{w}$ such that $\mathbf{y} = \mathbf{prox}_{cd_1}\mathbf{w}$. Then,

$$
\begin{aligned}
& \mathbf{prox}_{cd_1}\left(\mathbf{prox}_{cd_2}(I - c\nabla d_1) + c\,\nabla d_1\right)\mathbf{y} = \mathbf{y} \\
\Leftrightarrow\ & \left(\mathbf{prox}_{cd_2}(I - c\nabla d_1) + c\,\nabla d_1\right)\mathbf{y} = \mathbf{w} \\
& \quad \text{(plug in the formulas of } (I - c\nabla d_1)\mathbf{y} \text{ and } \nabla d_1 \mathbf{y}\text{)} \\
\Rightarrow\ & \left(\mathbf{prox}_{cd_2}\mathbf{refl}_{cd_1} - \frac{1}{2}\mathbf{refl}_{cd_1} + \frac{1}{2}I\right)\mathbf{w} = \mathbf{w} \\
\Leftrightarrow\ & \frac{1}{2}\left((2\mathbf{prox}_{cd_2} - I)\mathbf{refl}_{cd_1} + I\right)\mathbf{w} = \mathbf{w} \\
\Leftrightarrow\ & \frac{1}{2}\left(\mathbf{refl}_{cd_2}\mathbf{refl}_{cd_1} + I\right)\mathbf{w} = \mathbf{w} \\
\Leftrightarrow\ & \mathbf{refl}_{cd_2}\mathbf{refl}_{cd_1}\mathbf{w} = \mathbf{w}
\end{aligned}
$$

Reflection operator

Fixed-point optimality: $\mathbf{x}^*$ minimizes f if and only if

$$0 \in \partial f(\mathbf{x}^*) \iff \mathbf{prox}_f \mathbf{x}^* = \mathbf{x}^* \iff \mathbf{refl}_f \mathbf{x}^* = \mathbf{x}^*.$$

However, iteration $\mathbf{x}^{k+1} = \mathbf{refl}_f(\mathbf{x}^k)$ may "orbit" near a fixed distance to $\mathbf{x}^*$.

$\mathbf{refl}_f := 2\mathbf{prox}_f - I$ is nonexpansive, but *not firmly* nonexpansive.

Solution: damp the iteration as

$$\mathbf{x}^{k+1} = ((1-\alpha)I + \alpha \mathbf{refl}_f)\, \mathbf{x}^k$$

for $\alpha \in (0,1)$; then $\mathbf{x}^k$ converges to a fixed point of $\mathbf{refl}_f$.

(This trick works for any nonexpansive operator.)

Peaceman-Rachford splitting iteration

PRS iteration

$$\mathbf{w}^{k+1} = \mathbf{refl}_{cd_2}\mathbf{refl}_{cd_1}\mathbf{w}^{k}$$

Convergence generally needs

- at least one of $\mathbf{refl}_{cd_1}$ and $\mathbf{refl}_{cd_2}$ to be contractive
- the other can be nonexpansive.

Upon convergence, **recover**

$$\mathbf{y}^{k+1} = \mathbf{prox}_{cd_1}\mathbf{w}^{k+1}$$

Douglas-Rachford splitting iteration

DRS is *damped* PRS:

$$\begin{aligned} \textbf{iteration:} \quad & \mathbf{w}^{k+1} = (1-\alpha)\mathbf{w}^k + \alpha\, \mathbf{refl}_{cd_2} \mathbf{refl}_{cd_1} \mathbf{w}^k, \\ \textbf{return:} \quad & \mathbf{y}^{k+1} = \mathbf{prox}_{cd_1} \mathbf{w}^{k+1}. \end{aligned}$$

where $\alpha \in (0,1)$, often set as $1/2$.

It always converge (as long as d_1 and d_2 are proper closed convex functions)

Peaceman-Rachford splitting applied to the dual

Task 1: break PRS into two steps

1. $\mathbf{w}^{k+\frac{1}{2}} = \mathbf{refl}_{cd_1}\mathbf{w}^k$
2. $\mathbf{w}^{k+1} = \mathbf{refl}_{cd_2}\mathbf{w}^{k+\frac{1}{2}}$

and implement each step

Recall: $\mathbf{refl}_{cd_1}\mathbf{w}^k = \mathbf{prox}_{cd_1}\mathbf{w}^k + (\mathbf{prox}_{cd_1} - I)\mathbf{w}^k = \mathbf{y}^k + (\mathbf{y}^k - \mathbf{w}^k)$

Also recall: $\mathbf{y}^k = \mathbf{prox}_{cd_1}\mathbf{w}^k$, so $\mathbf{y}^k - \mathbf{w}^k \in c\partial d_1(\mathbf{y}^k)$.

Hence, $\mathbf{y}^k - \mathbf{w}^k = c\mathbf{A}\mathbf{x}^k$, where

$$\mathbf{x}^k = \arg\min_{\mathbf{x}} \mathcal{L}_1(\mathbf{x}; \mathbf{y}^k) = \arg\min_{\mathbf{x}} \mathcal{L}_{1\mathbf{A}}(\mathbf{x}; \mathbf{w}^k).$$

Step 1 is equivalent to $\mathbf{w}^{k+\frac{1}{2}} \leftarrow \mathbf{y}^k + (\mathbf{y}^k - \mathbf{w}^k) = \mathbf{y}^k + c\mathbf{A}\mathbf{x}^k$

Peaceman-Rachford splitting applied to the dual

Following the similar arguments:

$$\mathbf{refl}_{cd_2}\mathbf{w}^{k+\frac{1}{2}} = \mathbf{prox}_{cd_2}\mathbf{w}^{k+\frac{1}{2}} + (\mathbf{prox}_{cd_1} - I)\mathbf{w}^{k+\frac{1}{2}} = \mathbf{y}^{k+\frac{1}{2}} + (\mathbf{y}^{k+\frac{1}{2}} - \mathbf{w}^{k+\frac{1}{2}})$$

Recall $\mathbf{y}^{k+1/2} = \mathbf{prox}_{cd_2}\mathbf{w}^{k+1/2}$, so $\mathbf{y}^{k+1/2} - \mathbf{w}^{k+1/2} \in c\partial d_2(\mathbf{y}^{k+1/2})$.
Hence, $\mathbf{y}^{k+\frac{1}{2}} - \mathbf{w}^{k+\frac{1}{2}} = c(\mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$, where

$$\mathbf{z}^{k+\frac{1}{2}} = \arg\min_{\mathbf{z}} \mathcal{L}_2(\mathbf{z}; \mathbf{y}^{k+\frac{1}{2}}) = \arg\min_{\mathbf{z}} \mathcal{L}_{2A}(\mathbf{z}; \mathbf{w}^{k+\frac{1}{2}})$$

Step 2 is equivalent to $\mathbf{w}^{k+1} \leftarrow \mathbf{y}^{k+\frac{1}{2}} + c(\mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$

Peaceman-Rachford splitting applied to the dual

Task 2: express the iterations in $\mathbf{y}$, eliminating $\mathbf{w}$.

Following the definition $\mathbf{y} = \mathbf{prox}(\mathbf{w})$, we have

$$\mathbf{y}^{k+\frac{1}{2}} = \mathbf{prox}_{cd_2}\mathbf{w}^{k+\frac{1}{2}} = \mathbf{w}^{k+\frac{1}{2}} + c(\mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b}) = \mathbf{y}^k + c(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$$

also,

$$\mathbf{y}^{k+1} = \mathbf{prox}_{cd_1}\mathbf{w}^{k+1} = \mathbf{w}^{k+1} + c\mathbf{A}\mathbf{x}^{k+1} = \mathbf{y}^{k+\frac{1}{2}} + c(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$$

Task 3: express $\mathbf{x}, \mathbf{z}$ in terms of $\mathbf{y}$ instead of $\mathbf{w}$

Recall $\mathbf{w}^{k+1} = \mathbf{y}^{k+\frac{1}{2}} + c(\mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$. So,

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}_{1\mathbf{A}}(\mathbf{x}; \mathbf{w}^{k+1}) = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{y}^{k+\frac{1}{2}}, \mathbf{A}\mathbf{x}\rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b}\|^2$$

Recall $\mathbf{w}^{k+\frac{1}{2}} = \mathbf{y}^k + c\mathbf{A}\mathbf{x}^k$. So,

$$\mathbf{z}^{k+\frac{1}{2}} = \arg\min_{\mathbf{x}} \mathcal{L}_{2\mathbf{A}}(\mathbf{x}; \mathbf{w}^{k+\frac{1}{2}}) =$$
$$\arg\min_{\mathbf{z}} g(\mathbf{z}) + \langle \mathbf{y}^k, \mathbf{B}\mathbf{z}\rangle + \tfrac{c}{2}\|\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z} - \mathbf{b}\|^2$$

Peaceman-Rachford splitting applied to the dual

Final Task: simplify the iteration

1. $\mathbf{z}^{k+\frac{1}{2}} = \arg\min_{\mathbf{z}} g(\mathbf{z}) + \langle \mathbf{y}^k, B\mathbf{z}\rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x}^k + B\mathbf{z} - \mathbf{b}\|^2$
2. $\mathbf{y}^{k+\frac{1}{2}} = \mathbf{y}^k + c(\mathbf{A}\mathbf{x}^k + B\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$
3. $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{y}^{k+\frac{1}{2}}, \mathbf{A}\mathbf{x}\rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x} + B\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b}\|^2$
4. $\mathbf{y}^{k+1} = \mathbf{y}^{k+\frac{1}{2}} + c(\mathbf{A}\mathbf{x}^{k+1} + B\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b})$

- Every iteration has two decoupled primal subproblems, interleaved with two dual updates.
- Unlike dual FBS iteration, both subproblems have the augmented term.
- Lagrange multipliers $\mathbf{y}$ is immediately updated when either $\mathbf{z}$ or $\mathbf{x}$ is updated

Peaceman-Rachford splitting applied to the dual

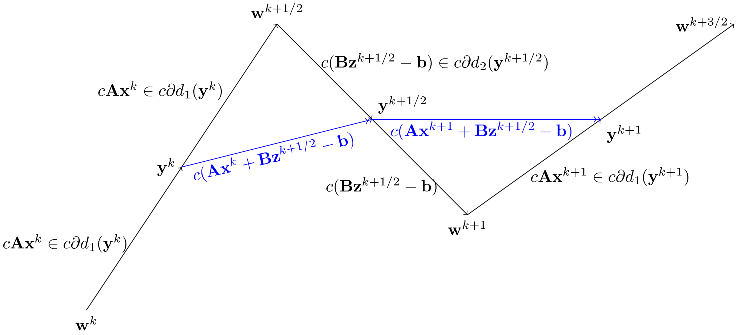

Figure: The illustration of PRS

Douglas-Rachford splitting applied to the dual

Recall DRS:

$$\begin{aligned}\textbf{iteration:}\quad \mathbf{w}^{k+1} &= (1-\alpha)\mathbf{w}^k + \alpha\,\mathbf{refl}_{cd_2}\mathbf{refl}_{cd_1}\mathbf{w}^k,\\ \textbf{return:}\quad \mathbf{y}^{k+1} &= \mathbf{prox}_{cd_1}\mathbf{w}^{k+1}.\end{aligned}$$

where $\alpha \in (0,1]$.

Set $\alpha = \frac{1}{2}$ and obtain simplified iteration:

1. $\mathbf{z}^{k+\frac{1}{2}} = \arg\min_{\mathbf{z}} g(\mathbf{z}) + \langle \mathbf{y}^k, B\mathbf{z}\rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x}^k + B\mathbf{z} - \mathbf{b}\|^2$,
2. $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{y}^k, \mathbf{A}\mathbf{x}\rangle + \frac{c}{2}\|\mathbf{A}\mathbf{x} + B\mathbf{z}^{k+\frac{1}{2}} - \mathbf{b}\|^2$,
3. $\mathbf{y}^{k+1} = \mathbf{y}^k + c(A\mathbf{x}^{k+1} + B\mathbf{z}^{k+\frac{1}{2}} - b)$.

A.k.a. the alternating direction method of multipliers (**ADM** or **ADMM**)

Douglas-Rachford splitting applied to the dual

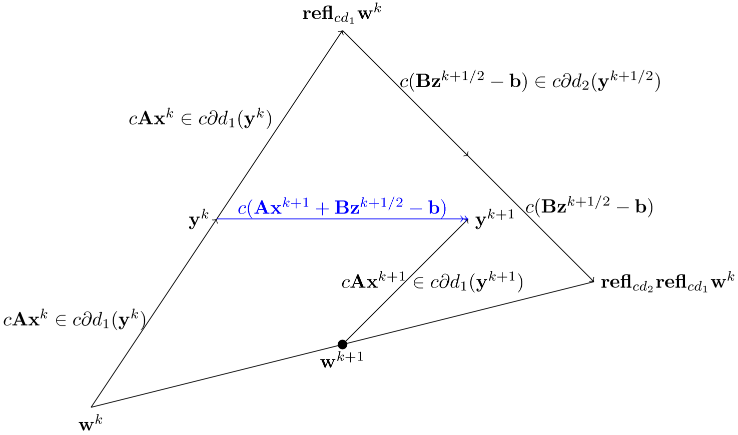

Figure: The illustration of DRS

Example: LASSO I

Model

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

Split $\|\cdot\|_1$ and $\|\cdot\|_2^2$

$$\min_{\mathbf{x},\mathbf{z}} \ \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2^2, \quad \text{s.t. } \mathbf{x} - \mathbf{z} = \mathbf{0}$$

Objective $\|\cdot\|_1$ and $\|\cdot\|_2^2$ end up in different subproblems

ADMM:

- $\mathbf{x}$-subproblem (minimizing $\|\cdot\|_1$) is soft-thresholding
- $\mathbf{z}$-subproblem is convex quadratic

Example: LASSO II

Same model

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

Alternative forms:

$$\min_{\mathbf{x},\mathbf{z}} \ \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{b}\|_2^2, \quad \text{s.t. } \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{0}$$

or

$$\min_{\mathbf{x},\mathbf{z}} \ \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{z}\|_2^2, \quad \text{s.t. } \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{b}$$

ADMM:

- $\mathbf{x}$-subproblem involve both ℓ_1 and $\mathbf{A}$, generally not simple! Exception: $\mathbf{A}$ is orthogonal. Or, solve an inexact $\mathbf{x}$-subproblem (a variant of ADMM).
- $\mathbf{z}$-subproblem is trivial

Example: ℓ_2-constrained basis pursuit

Model

$$\min \|\mathbf{x}\|_1, \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \le \sigma.$$

Decouple ℓ_2 from $\mathbf{A}\mathbf{x} - \mathbf{b}$ and rewrite the problem as

$$\min \|\mathbf{x}\|_1, \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b},\ \|\mathbf{z}\|_2 \le \sigma.$$

View $\|\mathbf{z}\|_2 \le \sigma$ as a part of the objective:

$$\min \|\mathbf{x}\|_1 + \iota_{\|\cdot\|_2 \le \sigma}(\mathbf{z}), \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}.$$

ADMM

- $\mathbf{x}$-subproblem involve both ℓ_1 and $\mathbf{A}$, generally not simple. Exception: $\mathbf{A}$ is orthogonal. Or, solve an inexact $\mathbf{x}$-subproblem.
- $\mathbf{z}$-subproblem is the projection to $B_\sigma = \{\mathbf{z} : \|\mathbf{z}\|_2 \le \sigma\}$.

Example: find a point in the intersection

Problem: given convex sets $\mathcal{A}$ and $\mathcal{B}$, find $\mathbf{x} \in \mathcal{A} \cap \mathcal{B}$.

Model

$$\min_{\mathbf{x},\mathbf{z}} \iota_{\mathcal{A}}(\mathbf{x}) + \iota_{\mathcal{B}}(\mathbf{z}), \quad \text{s.t. } \mathbf{x} - \mathbf{z} = \mathbf{0}.$$

ADMM

- $\mathbf{x}$-update is ℓ_2 projection to $\mathcal{A}$
- $\mathbf{z}$-update is ℓ_2 projection to $\mathcal{B}$
- Faster than alternatively projecting one point to $\mathcal{A}$ and $\mathcal{B}$ iteratively
- Has a long history, including cases where the sets are non-convex

Example: ℓ_1-ℓ_1 Model

Model

$$\min \|\mathbf{x}\|_1 + \mu\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1$$

Split two $\|\cdot\|_1$.

$$\min \|\mathbf{x}\|_1 + \mu\|\mathbf{z}\|_1, \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}.$$

ADMM

- $\mathbf{x}$-subproblem involves ℓ_1 and $\mathbf{A}$
- $\mathbf{y}$-subproblem is soft-thresholding

Example: group LASSO

Recall

$$\|\mathbf{x}\|_{\mathcal{G},2,1} = \sum_{s=1}^{S} w_s \|\mathbf{x}_{\mathcal{G}_s}\|_2$$

Assume that the groups $\mathcal{G}_s$ may *overlap*.

Model

$$\min \|\mathbf{x}\|_{\mathcal{G},2,1} + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

Rewrite

$$\min \ \frac{\beta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \sum_{s=1}^{S} w_s \|\mathbf{z}_s\|_2, \quad \text{s.t. } \mathbf{x}_{\mathcal{G}_s} - \mathbf{z}_s = \mathbf{0}, \ \forall s.$$

Different $\mathbf{z}_s$ *do not overlap*.

ADMM: quadratic $\mathbf{x}$-update and separable $\mathbf{z}_s$-updates.

ADMM applied to dual

The dual of BP is

$$\max\ \mathbf{b}^T\mathbf{y}, \quad \text{s.t. } \|\mathbf{A}^T\mathbf{y}\|_\infty \le 1$$

equivalent to

$$\max_{\mathbf{y},\mathbf{z}}\ \mathbf{b}^T\mathbf{y} + \iota_{\|\cdot\|_\infty \le 1}(\mathbf{z}), \quad \text{s.t. } \mathbf{A}^T\mathbf{y} + \mathbf{z} = \mathbf{0}.$$

ADMM

- $\mathbf{y}$-update is quadratic
- $\mathbf{z}$-update is projection to ℓ_∞-ball
- $\mathbf{x}$ becomes the Lagrange multipliers

ADMM applied to dual

The dual of

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \le \sigma$$

is

$$\max \ \mathbf{b}^T\mathbf{y} - \sigma\|\mathbf{y}\|_2, \quad \text{s.t. } \|\mathbf{A}^T\mathbf{y}\|_\infty \le 1$$

equivalent to

$$\max_{\mathbf{y},\mathbf{z}} \ (\mathbf{b}^T\mathbf{y} - \sigma\|\mathbf{y}\|_2) + \iota_{\|\cdot\|_\infty \le 1}(\mathbf{z}), \quad \text{s.t. } \mathbf{A}^T\mathbf{y} + \mathbf{z} = \mathbf{0}.$$

ADMM

- $\mathbf{y}$-update is nonsmooth
- $\mathbf{z}$-update is projection to ℓ_∞-ball

ADMM applied to dual

The dual of

$$\min \ \|\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

is

$$\min \ \mathbf{b}^T\mathbf{y} - \frac{1}{2\mu}\|\mathbf{y}\|_2^2, \quad \text{s.t. } \|\mathbf{A}^T\mathbf{y}\|_\infty \leq 1$$

equivalent to

$$\min_{\mathbf{y},\mathbf{z}} \ (\mathbf{b}^T\mathbf{y} - \frac{1}{2\mu}\|\mathbf{y}\|_2^2) + +\iota_{\|\cdot\|_\infty \leq 1}(\mathbf{z}), \quad \text{s.t. } \mathbf{A}^T\mathbf{y} + \mathbf{z} = \mathbf{0}.$$

ADMM

- $\mathbf{y}$-update is quadratic
- $\mathbf{z}$-update is projection to ℓ_∞-ball

Matlab package: YALL1

Software package YALL1 implemented ADMM applied to the duals of

$$
\begin{aligned}
\text{BP}: &\quad \min_{\mathbf{x}\in\mathbb{C}^n} \quad \|\mathbf{W}\mathbf{x}\|_{\mathbf{w},1} \quad \text{s.t. } \mathbf{A}\mathbf{x}=\mathbf{b} \\
\text{L1/L1}: &\quad \min_{\mathbf{x}\in\mathbb{C}^n} \quad \|\mathbf{W}\mathbf{x}\|_{\mathbf{w},1}+\tfrac{1}{\nu}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 \\
\text{L1/L2}: &\quad \min_{\mathbf{x}\in\mathbb{C}^n} \quad \|\mathbf{W}\mathbf{x}\|_{\mathbf{w},1}+\tfrac{1}{2\rho}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 \\
\text{BP+}: &\quad \min_{\mathbf{x}\in\mathbb{R}^n} \quad \|\mathbf{x}\|_{\mathbf{w},1} \quad \text{s.t. } \mathbf{A}\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq 0 \\
\text{L1/L1+}: &\quad \min_{\mathbf{x}\in\mathbb{R}^n} \quad \|\mathbf{x}\|_{\mathbf{w},1}+\tfrac{1}{\nu}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 \quad \text{s.t. } \mathbf{x}\geq 0 \\
\text{L1/L2+}: &\quad \min_{\mathbf{x}\in\mathbb{R}^n} \quad \|\mathbf{x}\|_{\mathbf{w},1}+\tfrac{1}{2\rho}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 \quad \text{s.t. } \mathbf{x}\geq 0
\end{aligned}
$$

For details, see `yall1.blogs.rice.edu`.

Example: total variation

Model

$$\min \mathrm{TV}(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

Applications (where $\mathbf{x}$ represents the pixels of a 2D image)

- Denoising: $\mathbf{A} = I$
- Deblurring and deconvolution: $\mathbf{A}$ is a circulant matrix or convolution operator
- MRI CS: $\mathbf{A} = \mathbf{P}\mathbf{F}$ downsampled Fourier transform; $\mathbf{P}$ is a row selector, $\mathbf{F}$ is the (discrete) Fourier transform
- Circulant CS: $\mathbf{A} = \mathbf{P}\mathbf{C}$ downsampled convolution; $\mathbf{C}$ is a circulant matrix or convolution operator

Difficulty: TV is the composite of ℓ_1 and $\nabla\mathbf{x}$, defined as

$$\mathrm{TV}(\mathbf{x}) := \|\nabla\mathbf{x}\|_1.$$

Assuming the periodic boundary condition, $\nabla\cdot$ is a convolution operator.

Example: total variation

Decouple ℓ_1 from $\nabla \mathbf{x}$:

$$\min_{\mathbf{x},\mathbf{z}} \frac{\mu}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \|\mathbf{z}\|_1, \quad \text{s.t. } \nabla \mathbf{x} - \mathbf{z} = \mathbf{0}$$

ADMM

- $\mathbf{x}$-update is quadratic in the form of

  $$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathbf{x}^T(\mu \mathbf{A}^T\mathbf{A} + \beta \nabla^T \nabla)\mathbf{x} + \text{linear terms}$$

  If $\mathbf{A}$ is identity, convolution, or partial Fourier, then

  $$\mathbf{F}(\mu \mathbf{A}^T\mathbf{A} + \beta \nabla^T \nabla)\mathbf{F}^{-1}$$

  is a diagonal matrix. So, $\mathbf{x}$-update becomes *very easy*.
- $\mathbf{y}$-subproblem is soft-thresholding

This splitting approach is often faster than the splitting

$$\min \mathrm{TV}(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{z}\|_2^2, \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}$$

because the $\mathbf{x}$-update is not in closed form.

Example: transform ℓ_1 (analysis ℓ_1)

Model

$$\min \|\mathbf{L}\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

Examples of $\mathbf{L}$ are:

- anisotropic finite difference operators
- orthogonal transforms: DCT, orthogonal wavelets
- frames: curvelets, shearlets

New models for ADMM

$$\min_{\mathbf{x},\mathbf{z}} \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{z}\|_1, \quad \text{s.t. } \mathbf{L}\mathbf{x} - \mathbf{z} = \mathbf{0},$$

or

$$\min_{\mathbf{x},\mathbf{z}} \|\mathbf{L}\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2^2, \quad \text{s.t. } \mathbf{x} - \mathbf{z} = \mathbf{0}.$$

Example: ℓ_1 fitting

Model

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1$$

New model

$$\min_{\mathbf{x},\mathbf{z}} \|\mathbf{z}\|_1, \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}.$$

ADMM

- $\mathbf{x}$-update is quadratic
- $\mathbf{y}$-update is soft-thresholding

Example: robust (Huber-norm) fitting

Model

$$\min_{\mathbf{x}} H(\mathbf{A}\mathbf{x} - \mathbf{b}) = \sum_{i=1}^{m} h(\mathbf{A}_i\mathbf{x} - b_i)$$

where

$$h(y) = \begin{cases} \frac{y^2}{2\epsilon}, & 0 \le |y| \le \epsilon, \\ |y| - \frac{\epsilon}{2}, & |y| > \epsilon. \end{cases}$$

New model

$$\min_{\mathbf{x},\mathbf{z}} H(\mathbf{z}), \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}.$$

ADMM

- $\mathbf{x}$-update is quadratic, involving $\mathbf{A}\mathbf{A}^T$
- $\mathbf{z}$-update is component-wise separable