ORIE 6300 Mathematical Programming I	October 27, 2016

Lecture 18

Lecturer: Damek Davis *Scribe: Chao Ning*

1 Last Time

Last time, we learned the Douglas-Rachford splitting (DRS) method.

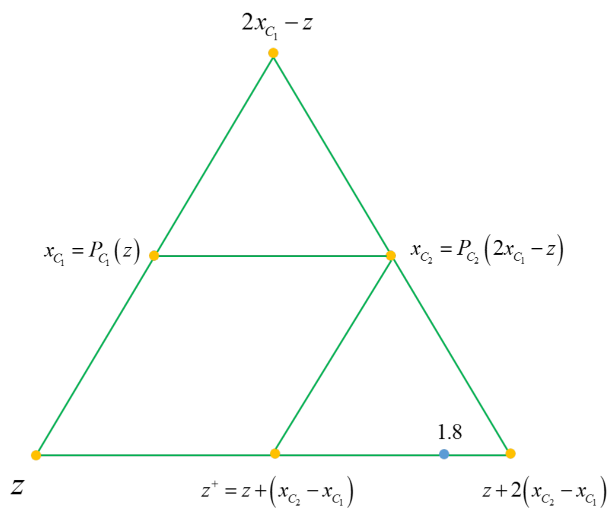

Figure 1: The Douglas-Rachford splitting method

2 Chambolle-Pork Algorithm

- Simplex method, MAP, DRS all require matrix inversions and linear equation solves.

	Per iteration cost	Accuracy after T iterations
Simplex method	$O(m^3 + mn)$	??
MAP/DRS with precomputation of Moore-Penrose inverse	$O((2m+n)(m+n+1))$	$O(\delta^T)$
MAP/DRS with no precomputation	$O((2m+n)^3)$	$O(\delta^T)$
Chambolle-Pork (Today's lecture)	$O(mn)$	$O(\frac{1}{T})$

In general, DRS and MAP allow inexact equation solves, say three iterations of conjugate gradient method. We can decrease per iteration complexity of MAP/DRS by exploiting the structure of

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ c^T & -b^T & 0 \end{bmatrix}$$

The Chambolle-Pork (CP) algorithm is unlike DRS, MAP and Simplex method in that it requires no equation solves. It is similar to MAP and DRS in that it is an instance of the KM algorithm. Therefore, we need to introduce the CP operator.

Definition 1 (Chambolle-Pork operator) *Consider a primal-dual pair of linear programs*

$$\min\{c^T x \mid Ax = b, x \geq 0\} \qquad \text{and} \qquad \max\{b^T y \mid A^T y \leq c\},$$

Let $\gamma, \tau > 0$. The CP operator, $T_{\text{CP}} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$
is defined by

$$(\forall y \in \mathbb{R}^m)\,(\forall x \in \mathbb{R}^n)$$

$$T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix} := \begin{bmatrix} y - \gamma\,(Ax - b) \\ \max\{x + \tau\left(A^T(y - 2\gamma(Ax - b)) - c\right), 0\} \end{bmatrix}$$

Definition 2 *The operator T (from Homework 6, Problem 3) is defined by*

$$T \begin{bmatrix} y \\ x \end{bmatrix} := \begin{bmatrix} y - \gamma\,(Ax - b) \\ \max\{x + \tau\left(A^T y - c\right), 0\} \end{bmatrix}$$

Lemma 1 *$Fix(T_{\text{CP}}) = Fix(T)$.*
Thus, by Homework 6, Problem 3

$$Fix(T_{\text{CP}}) = \{(y, x) \mid x \text{ is primal optimal, y is dual optimal}\}.$$

Proof: "$\subseteq$": Suppose that $\begin{bmatrix} y \\ x \end{bmatrix} \in Fix(T_{\text{CP}})$. Then

$$y = y - \tau(Ax - b) \;\Rightarrow\; Ax = b$$

$$\begin{aligned} x &= \max\{x + \tau\left(A^T(y - 2\gamma(Ax - b)) - c\right), 0\} \\ &= \max\{x + \tau\left(A^T y - c\right), 0\} \end{aligned}$$

Therefore, we have $T \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$, which implies that $\begin{bmatrix} y \\ x \end{bmatrix} \in Fix(T)$.
"$\supseteq$" is an exercise. □

Question: why introduce T_{CP} when T is much simpler? Answer: Because T_{CP} is nonexpansive, while T is, in general, expansive. T_{CP} is not nonexpansive under the usual norm $\left\| \begin{bmatrix} y \\ x \end{bmatrix} \right\|^2 = \|y\|^2 + \|x\|^2$, though. Instead we introduce a new norm.

Definition 3 *Let $Q \in \mathbb{R}^{d\times d}$ be a symmetric strongly positive definite (SSPD) matrix.*

$$(\exists \mu_Q > 0) : (\forall x \in \mathbb{R}^d)\ \langle Qx, x\rangle\ \geq \mu_Q \|x\|^2$$

Then we call $(\forall x \in \mathbb{R}^d)$ $\|x\|_Q = \sqrt{\langle Qx, x\rangle}$ the Mahalanobis norm induced by Q.

Excercise. Let $Q \in \mathbb{R}^{d\times d}$ be SSPD, then

$$(\forall x \in \mathbb{R}^d), (\forall y \in \mathbb{R}^d), (\forall a \in \mathbb{R})$$

1. $\|x\|_Q \geq 0$
2. $\|x + y\|_Q \leq \|x\|_Q + \|y\|_Q$
3. $\|ax\|_Q \leq |a|\|x\|_Q$
4. $\|x\|_Q = 0 \Leftrightarrow x = 0$

- The Mahalanobis distance skews space.

Example.
The two examples are shown in Figure 2.

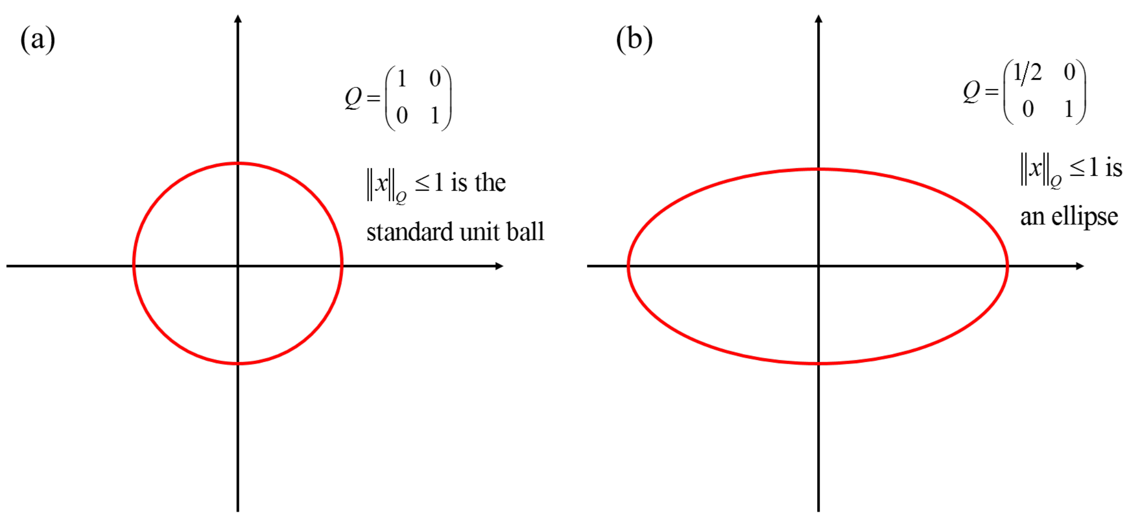

Figure 2: The two examples.

Chambolle-Pock Algorithm: Let $z^0 = \begin{bmatrix} y^0 \\ x^0 \end{bmatrix} \in \mathbb{R}^{m+n}$. Then $(\forall k \in \mathbb{N})$

$$z^{k+1} = T_{\text{CP}} z^k$$

Simplification Given $\begin{bmatrix} y \\ x \end{bmatrix} \in \mathbb{R}^{m+n}$.

$$\begin{bmatrix} y^+ \\ x^+ \end{bmatrix} = T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} y - \gamma\left(Ax - b\right) \\ \max\{x + \tau\left(A^T(2y^+ - y) - c\right), 0\} \end{bmatrix}$$

Lemma 2 *Define*

$$Q = \begin{bmatrix} \frac{1}{\gamma}I & -A \\ -A^T & \frac{1}{\tau}I \end{bmatrix}.$$

Then if $\gamma\tau < \frac{1}{\|A\|^2}$*,* Q *is SSPD (where* $\|A\| = \sup_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|}$*).*

Proof: $(\forall y \in \mathbb{R}^m), (\forall x \in \mathbb{R}^n)$

$$\begin{aligned} \left\langle Q \begin{bmatrix} y \\ x \end{bmatrix}, \begin{bmatrix} y \\ x \end{bmatrix} \right\rangle &= \left\langle \begin{bmatrix} \frac{1}{\gamma}y - Ax \\ \frac{1}{\tau}x - A^T y \end{bmatrix}, \begin{bmatrix} y \\ x \end{bmatrix} \right\rangle \\ &= \frac{1}{\gamma}\|y\|^2 + \frac{1}{\tau}\|x\|^2 - \langle Ax, y\rangle - \left\langle A^T y, x \right\rangle \\ &= \frac{1}{\gamma}\|y\|^2 + \frac{1}{\tau}\|x\|^2 - 2\langle Ax, y\rangle \qquad (*) \end{aligned}$$

Note that

$$\begin{aligned} -2\langle Ax, y\rangle &\geq -2\|Ax\|\|y\| \\ &\geq -2\|A\|\|x\|\|y\|. \end{aligned}$$

Then apply Young's inequality: $((\forall \epsilon > 0),\ 2ab \leq a^2\epsilon + \frac{b^2}{\epsilon})$: set $a = \|x\|, b = \|y\|, \epsilon = \sqrt{\frac{\gamma}{\tau}}$, so

$$-2\|A\|\|x\|\|y\| \geq -\sqrt{\frac{\gamma}{\tau}}\|A\|\|x\|^2 - \sqrt{\frac{\tau}{\gamma}}\|A\|\|y\|^2.$$

Thus,

$$\begin{aligned} (*) &\geq \left(\frac{1}{\gamma} - \sqrt{\frac{\tau}{\gamma}}\|A\|\right)\|y\|^2 + \left(\frac{1}{\tau} - \sqrt{\frac{\gamma}{\tau}}\|A\|\right)\|x\|^2 \\ &\geq \min\left\{\left(\frac{1}{\gamma} - \sqrt{\frac{\tau}{\gamma}}\|A\|\right), \left(\frac{1}{\tau} - \sqrt{\frac{\gamma}{\tau}}\|A\|\right)\right\}\left(\|x\|^2 + \|y\|^2\right) \end{aligned}$$

Define $\mu_Q := min\left\{\left(\frac{1}{\gamma} - \sqrt{\frac{\tau}{\gamma}}\|A\|\right), \left(\frac{1}{\tau} - \sqrt{\frac{\gamma}{\tau}}\|A\|\right)\right\}$. Clearly, $\mu_Q > 0 \Leftrightarrow \gamma\tau < \frac{1}{\|A\|^2}$ □

The following Proposition is an exercise on Homework 7.

Proposition 1 *Let $\gamma\tau < \frac{1}{\|A\|^2}$. Then*

$(\forall z_1 \in \mathbb{R}^{m+n}), (\forall z_2 \in \mathbb{R}^{m+n})$

$$\|T_{\text{CP}}z_1 - T_{\text{CP}}z_2\|_Q^2 \leq \|z_1 - z_2\|_Q^2 - \|(I - T_{\text{CP}})z_1 - (I - T_{\text{CP}})z_2\|_Q^2$$

In fact, $2T_{\text{CP}} - I$ is nonexpansive in $\|\cdot\|_Q$.

<u>Remark</u>: Given an operator $T : \mathbb{R}^d \to \mathbb{R}^d$ that satisfies $2T - I$ is nonexpansive. The operator T is called *firmly nonexpansive*.

Theorem 1 *Let $\begin{bmatrix} y^0 \\ x^0 \end{bmatrix} \in \mathbb{R}^{m+n}$ and choose $\gamma, \tau > 0$, such that $\gamma\tau < \frac{1}{\|A\|^2}$. Then the sequence*

$$y^{k+1} = y^k - \gamma(Ax^k - b)$$

$$x^{k+1} = \max\left\{x^k + \tau\left(A^T(2y^{k+1} - y^k) - c\right), 0\right\}.$$

converges to a primal-dual optimal pair.

Proof: Let $N = 2T_{\text{CP}} - I$. Apply KM iteration with $\lambda = \frac{1}{2} : N_{\frac{1}{2}} = T_{\text{CP}}$ □

2.1 Convergence Rates

The following results are known, but not explicitly recorded in existing papers:

Proposition 2

1. $\|Ax^k - b\| = o\left(\frac{1}{\sqrt{k+1}}\right)$
2. $\|x^{k+1} - \max\left\{x^k + \tau\left(A^T(2y^{k+1} - y^k) - c\right), 0\right\}\| = o\left(\frac{1}{\sqrt{k+1}}\right)$
3. *Replace x^k and y^k by $\overline{x}^k = \frac{1}{k+1}\sum_{i=0}^k x^i$, $\overline{y}^k = \frac{1}{k+1}\sum_{i=0}^k y^i$, rates improve to $O\left(\frac{1}{k+1}\right)$.*

2.2 General Case

In general, CP solves

$$\min_x F(Ax - b) + g(x),$$

with algorithm

$$y^{k+1} = (y^k - \gamma(Ax^k - b)) - \gamma\text{argmin}_y\left\{F(y) + \frac{\gamma}{2}\|y - \frac{1}{\gamma}(y^k - \gamma(Ax^k - b)\|^2\right\}$$

$$x^{k+1} = \text{argmin}_x\left\{g(x) + \frac{1}{2\tau}\|x - (x^k + \tau A^T(2y^{k+1} - y^k)\|^2\right\}$$

In our case:

$$F(y) = \iota_{\{0\}}(y) = \begin{cases} 0 & \text{if y=0} \\ \infty & \text{otherwise} \end{cases}$$

$$g(x) = \langle c, x \rangle + \iota_{\mathbb{R}_{\geq 0}}(x) = \begin{cases} \langle c, x \rangle & \text{if x}\geq\text{0} \\ \infty & \text{otherwise} \end{cases}$$