ORIE 6300 Mathematical Programming I November 1, 2016

Lecture 19

Lecturer: Damek Davis *Scribe: Jianqiu Wang/BingZhen Zhang*

1 Review

Last time we talked about the Chambolk-Pock (CP) algorithm. The CP operator is given below:

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

We have following proposition:

Proposition 1 *If $\gamma\tau < \frac{1}{\|A\|^2}$, then T_{CP} is firmly nonexpansive:*

$$\left(\forall z_1, z_2 \in \mathbb{R}^{n+m}\right) \quad \|T_{CP}z_1 - T_{CP}z_2\| \leq \|z_1 - z_2\|_Q^2 - \|(I - T_{CP})(z_1) - (I - T_{CP})(z_2)\|_Q^2$$

in the Mahalanbois norm $\|\cdot\|_Q$, where $Q = \begin{bmatrix} \frac{1}{\gamma}I & -A \\ -A^T & \frac{1}{\tau}I \end{bmatrix}$

Thus, $2T_{\mathrm{CP}} - I$ is nonexpansive in $\|\cdot\|_Q$ and the KM iteration, initialized at $\begin{bmatrix} y^0 \\ x^0 \end{bmatrix} \in \mathbb{R}^{n+m}$ and recursively defined by

$$\begin{bmatrix} y^{k+1} \\ x^{k+1} \end{bmatrix} = \begin{bmatrix} y^k - \gamma(Ax^k - b) \\ \max\{x^k + \tau(A(2y^{k+1} - y^k) - c), 0\} \end{bmatrix},$$

converges to a primal-dual optimal pair.

2 Simplex method and nonlinear operators

So far we have developed three fixed-point algorithms: DRS, MAP, CP. All are special cases of KM iteration The KM iteration is just one of many ways to solve nonlinear, non-differentiable equations. We will discuss ways of solving differentiable equations in the coming weeks.

Today, we will view the the simplex method as just another way of solving nonlinear equations.

The following is a description of one iteration of the simplex method.

- Input: BFS x with basis B
- Compute verifying y: $y = A_B^{-T}c_B$ and reduced cost: $\bar{c} = c - A^Ty$
- $\bar{c}_B = 0, x_N = 0, x_B = A_B^{-1}b$

- If $\bar{c} \geq 0$, then x and y are primal-dual optimal.
- If $\bar{c}_j < 0$, then $j \in N$. Let $\bar{b} = A_B^{-1}A, \bar{A} = A_B^{-1}A_N$
- Check for unboundedness: If $\bar{A}_{iN(j)} \leq 0\ \forall i$, then $x_N = \varepsilon e_{N(j)}$ if feasible $\forall \varepsilon > 0$ and $\bar{c}_N^T(\varepsilon e_{N(j)}) \to \infty$ as $\varepsilon \to \infty$ So LP is unbounded
- Otherwise, we perform the ratio test. If $\exists\ i$ s.t. $\bar{A}_{iN(j)} > 0$, define

$$\varepsilon = \min_{i:\bar{A}_{iN(j)>0}} \frac{\bar{b}_i}{\bar{A}_{iN(j)}}$$

  Let i^* achieve min.
- Construct new BFS $\hat{x}$: $\hat{x}_j = \varepsilon, \hat{x}_{N\setminus\{j\}} = 0, \hat{x}_B = \bar{b} - \bar{A}\hat{x}_N$
- Set $\hat{B} = B \cup \{i^*\} \setminus \{j\}$(updated basis)
- The new point satisfies $c^T\hat{x} \leq c^T x$ and $\hat{B}$ is a basis.

How do we view the simplex method as an algorithm for solving nonlinear equations? Recall the mapping $T : \mathbb{R}^{m+n} \to \mathbb{R}^{m+n}$

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

Lemma 2 *Suppose x is a BFS and B is a basis. Let $y = A_B^{-T} c_B$ be a verifying y. Let $\bar{c} = c - A^T y$ be the reduced cost. Then*

$$T\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} y \\ x_B \\ \max\{-\bar{c}_N, 0\} \end{bmatrix}$$

Proof: We look at three components of RHS separately:

- x is feasible implies

$$y - (Ax - b) = y$$

- $\bar{c}_B = 0$ implies

$$\max\{x_B + (A^T y - c)_B, 0\} = \max\{x_B - \bar{c}_B, 0\} = \max\{x_B, 0\} = x_B$$

- $x_N = 0$ implies

$$\max\{x_N - \bar{c}_N, 0\} = \max\{-\bar{c}_N, 0\}$$

□

Thus, BFS correspond to *partial fixed-points* of T. The fixed-point residual satisfies:

$$\left\| \begin{bmatrix} y \\ x \end{bmatrix} - T\begin{bmatrix} y \\ x \end{bmatrix} \right\| = \| \max\{-\bar{c}_N, 0\}\|.$$

This shows that the reduced cost measures fixed-point violation. Clearly, $\begin{bmatrix} y \\ x \end{bmatrix}$ is not a fixed point whenever $c_N \ngeq 0$. Now suppose that all vertices are nondegenerate.

Suppose we want to generate a new point $\bar{x}$ s.t. $A\bar{x} = b, \bar{x} \geq 0, c^T\bar{x} < c^T x$ and

$$T \begin{bmatrix} y \\ \bar{x} \end{bmatrix} = \begin{bmatrix} y \\ \bar{x}_B \\ ?? \end{bmatrix},$$

where "??" signifies that we do not constrain those components. Then, to ensure that the objective function decreases, we have

$$c_B^T \bar{x}_B + c_N^T \bar{x}_N < c_B^T x_B + c_B^T x_N = c_B^T x_B$$

This is equivalent to

$$c_B^T(\bar{x}_B - x_B) + c_N^T \bar{x}_N < 0 \tag{1}$$

Use

$$A\bar{x} = b \Leftrightarrow \bar{x}_B = A_B^{-1} b - A_B^{-1} A_N \bar{x}_N$$

and $x_B = A_B^{-1} b$, to get

$$\begin{aligned} (1) &\Leftrightarrow -c_B^T(A_B^{-1} A_N \bar{x}_N) + c_N^T \bar{x}_N < 0 \\ &\Leftrightarrow (c_N - A_N y)^T \bar{x}_N < 0 \\ &\Leftrightarrow \bar{c}^T \bar{x}_N < 0 \end{aligned}$$

So we can choose j s.t. $\bar{c}_j < 0$, choose $\bar{x}_{N \setminus \{j\}} = 0$, and $\bar{x}_j = \varepsilon$ where $\varepsilon > 0$ is chosen so that

$$0 \leq \bar{x}_B = A_B^{-1} b - A_B^{-1} A_N \bar{x}_N = A_B^{-1} b - A_B^{-1} A_N \bar{x}_N \varepsilon e_{N(j)}$$

i.e.,

$$\varepsilon = \min_{i:(A_B^{-1} A_N)_{ij} > 0} \frac{(A_B^{-1} b)_i}{(A_B^{-1} A_N)_{ij}}$$

Then by working backwards we can check that $c^T\bar{x} \leq c^T x + \varepsilon \bar{c}_j < c^T x$. Of course this $\bar{x}$ is just $\hat{x}$ generated by simplex.

Moreover

$$\begin{aligned} \left(T \begin{bmatrix} y \\ \bar{x} \end{bmatrix}\right)_y &= y - (A\bar{x} - b) = y \\ \left(T \begin{bmatrix} y \\ \bar{x} \end{bmatrix}\right)_{\bar{x}_B} &= \max\{\bar{x}_B - \bar{c}_B, 0\} = \max\{\bar{x}_B, 0\} = \bar{x}_B \\ \left(T \begin{bmatrix} y \\ \bar{x} \end{bmatrix}\right)_{\bar{x}_j} &= \max\{\bar{x}_j - \bar{c}_j, 0\} > \bar{x}_j \\ \left(T \begin{bmatrix} y \\ \bar{x} \end{bmatrix}\right)_{\bar{x}_{N \setminus \{j\}}} &= \max\{\bar{x}_{N \setminus \{j\}} - \bar{c}_{N \setminus \{j\}}, 0\} = \max\{-\bar{c}_{N \setminus \{j\}}, 0\} \end{aligned}$$

Let $\hat{B}$ be the updated basis. How can we choose $\bar{y}$ such that

$$T \begin{bmatrix} \bar{y} \\ \bar{x} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \hat{x}_{\hat{B}} \\ ?? \end{bmatrix}$$

By the lemma, choose $\bar{y} = A_{\hat{B}}^{-T} c_{\hat{B}}$. Thus,

the simplex method tries to maintain a partial fixed point while simultaneously decreasing the objective.

This view of the simplex method seems to be new.

We list three open questions:

1. Given that $\left\| \begin{bmatrix} y \\ x \end{bmatrix} - T \begin{bmatrix} y \\ x \end{bmatrix} \right\| = \|\max\{-\bar{c}_N, 0\}\|$, is it better to choose $\hat{x}$ so that

$$\left\| \begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix} - T \begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix} \right\| < \left\| \begin{bmatrix} y \\ x \end{bmatrix} - T \begin{bmatrix} y \\ x \end{bmatrix} \right\|$$

   And how would we do this?

2. The arguments above generalize to the CP operator. We know that

$$\left\| T_{\text{CP}} \left(T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix} \right) - T_{\text{CP}} \left(\begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix} \right) \right\| < \left\| T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix} - \begin{bmatrix} y \\ x \end{bmatrix} \right\|$$

   Does the closest BFS $\begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix}$ to $T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix}$ satisfy

$$\left\| \begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix} - T_{\text{CP}} \begin{bmatrix} \hat{y} \\ \hat{x} \end{bmatrix} \right\| < \left\| \begin{bmatrix} y \\ x \end{bmatrix} - T_{\text{CP}} \begin{bmatrix} y \\ x \end{bmatrix} \right\|?$$

3. Can we interpret the simplex method as a fixed-point method on the DRS and MAP operators?