The First-Order Homogeneluos and Self-Dual Potential Reduction Method for Linear Optimization

***

March 25, 2016

Abstract

We propose a primal-dual first-order potential reduction method for linear optimization, and show that the method generates an ϵ solution in $\log(1/\epsilon)$ Lagrange update steps.

1 The Homogeneous and Self-Dual Linear Programming

Consider linear program and its dual

$$\begin{array}{rc} \min & c^T x \\ s.t. & Ax \geq b; \\ & x \geq 0. \end{array} \qquad \begin{array}{rc} \max & b^T y \\ s.t. & A^T y + s = c,\ (y, s) \geq 0. \end{array}$$

The problem can be solved by finding a nontrivial feasible solution to a homogeneous and self-dual system (Ye et al. [5]):

$$\begin{array}{c} Ax - b\tau - z = 0; \\ -A^T y + c\tau - s = 0; \\ b^T y - c^T x - \kappa = 0; \\ e^T x + e^T z + e^T y + e^T s + \tau + \kappa = 1; \\ (x, y, s, z, \tau, \kappa) \geq 0; \end{array}$$

where e is the vector of all ones.

The problem can be further written and generalized as finding a pair of $(x \in R^n, s \in R^n)$ such that:

$$\begin{array}{c} Mx - s = 0; \\ e^T x + e^T s = 2n; \\ (x, s) \geq 0; \end{array} \tag{1}$$

where M is a skew-symmetric matrix, that is, $M = -M^T$. More precisely, for linear programming above,

$$M = \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix}.$$

For any pair $(x, s) \geq 0$ of system (1), we define

$$f(x, s, \theta) = x^T s + \theta^2 + \frac{1}{2}\|Mx - s - \theta(Me - e)\|^2.$$

One can see $f(x, s, \theta) \geq 0$ since it is the sum of two nonnegative terms, and it is a homogeneous quadratic function. From the linear programming theorems, there is always a nontrivial pair (x, s) feasible to (1) or $f(x, s, \theta) = 0$.

2 A Primal-Dual Potential Function

Now we consider the potential function

$$\phi(x, s, \theta) = \rho \ln(f(x, s, \theta)) - \sum_j \ln(x_j s_j),$$

where $(x, s) > 0$ and $\rho = n + \sqrt{n}$. One can see

$$\begin{aligned} \phi(x, s, \theta) &= (\rho - n)\ln(f(x, s, \theta)) + n\ln(f(x, s, \theta)) - \textstyle\sum_j \ln(x_j s_j) \\ &\geq (\rho - n)\ln(f(x, s, \theta)) + n\ln(x^T s) - \textstyle\sum_j \ln(x_j s_j) \\ &\geq (\rho - n)\ln(f(x, s, \theta)) + n\ln(n) \\ &= \sqrt{n}\ln(f(x, s, \theta)) + n\ln(n). \end{aligned}$$

Also note that the gradient vector of the function at $(x, s) > 0$ and θ is

$$\begin{aligned} \nabla\phi_x(x, s, \theta) &= \tfrac{\rho}{f(x,s,\theta)}\nabla f_x(x, s, \theta) - X^{-1}e, \\ \nabla\phi_s(x, s, \theta) &= \tfrac{\rho}{f(x,s,\theta)}\nabla f_x(x, s, \theta) - S^{-1}e, \\ \nabla\phi_\theta(x, s, \theta) &= \tfrac{\rho}{f(x,s,\theta)}\nabla f_\theta(x, s, \theta). \end{aligned}$$

where X and S denote the diagonal matriices whose diagonal entries are elements of vectors x and s, respectively.

The following lemma is well known in the literature of interior-point algorithms ([2, 1, 6]):

Lemma 1. *Let $x > 0$ and $\|X^{-1}d\| \leq \beta < 1$. Then*

$$-\sum_j \ln(x_j + d_j) + \sum_j \ln(x_j) \leq -e^T X^{-1} d + \frac{\beta^2}{2(1-\beta)}.$$

Let γ be the largest eigenvalue of matrix $[M - I(Me - e)]^T[M - I(Me - e)]$. Then, for any given $d = [d_x \in R^n; d_s \in R^n, d_\theta \in R]$,

$$f(x + d_x, s + d_x, \theta + d_\theta) - f(x, s, \theta) \leq \nabla f(x, s, \theta)^T d + \frac{\gamma}{2}\|d\|^2.$$

Denote by $x^+ = x + d_x,\ s^+ = s + d_s,\ \theta^+ = \theta + d_\theta$.

Now, let $(x, s) > 0$ and $\|[X^{-1}d; S^{-1}d_s; d_\theta]\| = \beta < 1$. Then $x^+ = x + d_x = X(e + X^{-1}d_x) > 0$ and $s^+ = s + d_s = S(e + S^{-1}d_s) > 0$. Morevoer, from Lemma 1

$$\begin{aligned} \phi(x^+, s^+, \theta^+) - \phi(x, s, \theta) &\leq \rho\ln\left(1 + \tfrac{\nabla f(x,s,\theta)^T d + \frac{\gamma}{2}\|d\|^2}{f(x,s,\theta)}\right) - e^T X^{-1} d + \tfrac{\beta^2}{2(1-\beta)} \\ &\leq \rho\tfrac{\nabla f(x,s,\theta)^T d + \frac{\gamma}{2}\|d\|^2}{f(x,s,\theta)} - e^T X^{-1} d + \tfrac{\beta^2}{2(1-\beta)} \\ &= \nabla\phi(x, s, \theta)^T d + \tfrac{\rho\gamma}{2f(x,s,\theta)}\beta^2 + \tfrac{\beta^2}{2(1-\beta)}. \end{aligned}$$

Lemma 2. *For any* $(x, s) > 0$ *and* $(x, s, \theta) \neq (x^*, s^*, 0)$*, a matrix* $A \in R^{m\times(2n+1)}$ *with* $A[x; s; \theta] = A[x^*; s^*; 0]$*, and a vector* $\bar{\lambda} \in R^m$*, consider vector*

$$p(x, s, \theta) = \Delta\left(\nabla\phi(x, s, \theta) - A^T\bar{\lambda}\right),$$

where

$$\Delta = \begin{pmatrix} X & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then,

$$\|p(x, s, \theta)\| \geq 1.$$

Proof. First,

$$\begin{aligned} p(x, s, \theta) &= \Delta\left(\tfrac{\rho}{f(x,s,\theta)}\nabla f(x, s, \theta) - [X^{-1}e; S^{-1}e; 0] - A^T\bar{\lambda}\right) \\ &= \tfrac{\rho}{f(x,s,\theta)}\Delta\left(\nabla f(x, s, \theta) - \tfrac{f(x,s,\theta)}{\rho}A^T\bar{\lambda}\right) - [e; e; 0]. \end{aligned}$$

If any entry of $(\nabla f(x, s, \theta) - \frac{f(x,s,\theta)}{\rho}A^T\bar{\lambda})_{1:2n}$ (i.e., the first $2n$ entries) is equal or less than 0, then

$$\|p(x, s, \theta)\| \geq \|p(x, s, \theta)\|_\infty \geq 1.$$

On the other hand, if $\left(\nabla f(x, s, \theta) - \frac{f(x,s,\theta)}{\rho}A^T\bar{\lambda}\right)_{1:2n} > 0$, we have

$$\left(\nabla f(x, s, \theta) - \frac{f(x, s, \theta)}{\rho}A^T\bar{\lambda}\right)_{1:n}^T [x^*; s^*] \geq 0.$$

Then, from convexity and $A[x; s; \theta] = A[x^*; s^*; 0]$,

$$f(x^*) - f(x) \geq \nabla f(x)^T(x^* - x) = \left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T (x^* - x).$$

Thus, from $f(x^*) = 0$

$$f(x) \leq \left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x.$$

Furthermore,

$$\begin{aligned} \|p(x)\|^2 &= \tfrac{\rho^2}{f(x)^2}\|X\left(\nabla f(x) - \tfrac{f(x)}{\rho}A^T\bar{\lambda}\right)\|^2 - 2\tfrac{\rho}{f(x)}\left(\nabla f(x) - \tfrac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x + n \\ &\geq \tfrac{\rho^2}{n\cdot f(x)^2}\|X\left(\nabla f(x) - \tfrac{f(x)}{\rho}A^T\bar{\lambda}\right)\|_1^2 - 2\tfrac{\rho}{f(x)}\left(\nabla f(x) - \tfrac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x + n \\ &\geq \tfrac{\rho^2}{n}\left(\tfrac{\left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x}{f(x)}\right)^2 - 2\rho\left(\tfrac{\left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x}{f(x)}\right) + n \\ &= \tfrac{(\rho z)^2}{n} - 2\rho z + n = \tfrac{1}{n}(\rho z - n)^2, \end{aligned}$$

where

$$z = \frac{\left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x}{f(x)} \geq 1.$$

The above quadratic function of z has the minimizer at $z = 1$ if $\rho \geq n$, so that

$$\frac{1}{n}(\rho z - n)^2 \geq \frac{1}{n}(\rho - n)^2 \geq 1$$

for $\rho \geq n + \sqrt{n}$. □

3 A Gradient Projection for Potential Reduction

The first order gradient-projection potential reduction algorithm would update (x, s, θ) by solving d from

$$\begin{array}{cc} \text{Minimize} & \nabla\phi(x,s,\theta)^T d \\ \text{Subject to} & e^T d_x + e^T d_s = 0,\ \|[X^{-1}d_x; S^{-1}d_s; d_\theta]\| = \beta; \end{array} \tag{2}$$

or

$$\begin{array}{cc} \text{Minimize} & \nabla\phi(x)^T X d' \\ \text{Subject to} & e^T X d' = 0,\ \|d'\| \le \beta; \end{array}$$

where parameter $\beta < 1$ is yet to be determined.

Let the scaled gradient projection vector

$$p(x) = \left(I - \frac{1}{\|x\|^2} X e e^T X\right) X\nabla\phi(x) = X\left(\frac{\rho}{f(x)}\left(\nabla f(x) - e\cdot\lambda(x)\right)\right) - e,$$

where

$$\lambda(x) = \frac{e^T X^2 \nabla\phi(x)\cdot f(x)}{\|x\|^2\cdot\rho}.$$

Then the minimizer of problem (2) would be

$$d = -\frac{\beta}{\|p(x)\|} X p(x),$$

and

$$\nabla\phi(x)^T d = -\frac{\beta}{\|p(x)\|}\|p(x)\|^2 = -\beta\|p(x)\| \le -\beta,$$

since $\|p(x)\| \ge 1$ based on Lemma 2.

Thus,

$$\phi(x^+) - \phi(x) \le -\beta + \frac{\rho\gamma}{2f(x)}\beta^2 + \frac{\beta^2}{2(1-\beta)}$$

For $\beta \le 1/2$, the above quantity is less than

$$-\beta + \left(2 + \frac{\rho\gamma}{f(x)}\right)\beta^2/2.$$

Thus, one can choose β to minimize the quantity at

$$\beta = \frac{1}{2 + \frac{\rho\gamma}{f(x)}} \le 1/2$$

so that

$$\phi(x^+) - \phi(x) \le \frac{-f(x)}{2(f(x) + 2\rho\gamma)}.$$

One can see that the larger value of $f(x)$, the greater reduction of the potential function.

Starting from $x^0 = \frac{1}{n}e$, we iteratively generate x^k, $k = 1, ...,$ such that

$$\phi(x^{k+1}) - \phi(x^k) \le \frac{-f(x^k)}{2(f(x^k) + 2\rho\gamma)} \le \frac{-f(x^k)}{2(f(x^0) + 2\rho\gamma)} \le \frac{-f(x^k)}{4\max\{f(x^0), 2\rho\gamma\}}.$$

The second inequality is due to $f(x^k) < f(x^0)$ from $\phi(x^k) < \phi(x^0)$ for all $k \geq 1$ and x^0 is the analytic center of the simplex.

Thus, if $\frac{f(x^k)}{f(x^0)} \geq \epsilon$ for $1 \leq k \leq K$, we must have

$$\phi(x^0) - \phi(x^K) \leq \rho \ln(\frac{1}{\epsilon}),$$

so that

$$\sum_{k=1}^{K} \frac{f(x^k)}{4\max\{f(x^0), 2\rho\gamma\}} \leq \rho \ln(\frac{1}{\epsilon})$$

or

$$K\epsilon f(x^0) \leq 4\max\{f(x^0), 2\rho\gamma\}\rho \ln(\frac{1}{\epsilon}).$$

Note that $\rho = n + \sqrt{n} \leq 2n$. We conclude

Theorem 3. *The steepest descent potential reduction algorithm generates a x^k with $f(x^k)/f(x^0) \leq \epsilon$ in no more than*

$$4(n+\sqrt{n})\frac{\max\{1, 2(n+\sqrt{n})\gamma/f(x^0)\}}{\epsilon} \ln(\frac{1}{\epsilon})$$

steps.

Let $x^0 = s^0 = e$ and $\theta^0 = 1$, where e is the vector of all ones, then $e^T x^0 + e^T s^0 = 2n$ and

$$f(x^0, s^0) = n + \frac{1}{2}\|Me - e\|^2 = \frac{3}{2}n + \frac{1}{2}\|Me\|^2$$

and

$$\phi(x^0, s^0) = \rho \ln(f(x^0, s^0)).$$

Second, one could develop a primal-dual potential reduction algorithm (e.g., [4])

$$\phi(x) = \rho \ln(s(x,\lambda)^T x) - \sum_j \ln(x_j) - \sum_j \ln(s(x,\lambda)_j),$$

where $s(x,\lambda) = \nabla f(x) - e \cdot \lambda > 0$. Then, such an algorithm would save the complexity iteration bound by a factor $\sqrt{n}$.

Moreover, one may use the Mehrotra's predictor and corrector algorithm [3] to improve the practical efficiency. In particular, the high-order or conjugate gradient correction may further reduce the dependency on γ for the complexity bound.

References

[1] C. C. Gonzaga, Polynomial affine algorithms for linear programming, *Math. Programming* 49 (1990) 7–21.

[2] N. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica* 4 (1984) 373-395.

[3] S. Mehrotra. On the implementation of a primal–dual interior point method. *SIAM J. Optimization*, 2(4):575–601, 1992.

[4] M. J. Todd and Y. Ye, A centered projective algorithm for linear programming, *Math. Oper. Res.* 15 (1990) 508-529.

[5] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$ - iteration homogeneous and self-dual linear programming algorithm, *Math. Oper. Res.* 19 (1994) 53–67.

[6] Y. Ye, An $O(n^3L)$ potential reduction algorithm for linear programming, *Math. Programming* 50 (1991) 239–258.

[7] Y. Ye, On the complexity of approximating a KKT point of quadratic programming, *Math. Programming* 80 (1998) 195-211.