On a First-Order Potential Reduction Algorithm for Linear Programming*

Yinyu Ye
Stanford University

July 30, 2015

Abstract

We describe a steepest-descent potential reduction method for linear and convex minimization over a simplex in R^n and its precise complexity analysis. In this method, no matrix needs to be ever inversed so that it is a pure first-order method.

1 Convex optimization over the simplex constraint

We consider the following optimization problem over the simplex:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to} & e^T x = 1;\ x \geq 0, \end{array} \tag{1}$$

where e is the vector of all ones. Such a problem in considered in [7], where function $f(x)$ does not need to be convex and a FPTAS algorithm was developed for computing an approximate KKT point of general quadratic programming. The following algorithm and analysis resemble those in [7].

We assume that $f(x)$ is a convex function in $x \in R^n$ and $f(x^*) = 0$ where x^* is a minimizer of the problem. Furthermore, we make a standard Lipschitz assumption such that

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

where positive γ is the Lipschitz parameter. Note that any homogeneous linear feasibility problem, e.g., the canonical Karmarkar form in [2]:

$$\begin{array}{c} Ax = 0; \\ e^T x = 1; \\ x \geq 0. \end{array}$$

can be formulated as the model with $f(x) = \frac{1}{2}\|Ax\|^2$ and γ as the half of the largest eigenvalue of matrix $A^T A$.

*This was a teaching note for course MS&E310, Linear Optimization

Furthermore, any linear programming problem in the standard form and its dual

$$\begin{array}{lclc} \text{Minimize} & c^T x & \text{Maximize} & b^T y \\ \text{Subject to} & Ax = b;\; x \geq 0; & \text{Subject to} & A^T y + s = c;\; s \geq 0 \end{array}$$

can be represented as a homogeneous linear feasibility problem (Ye et al. [5]):

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

We consider the potential function (e.g., see [2, 4, 1, 6])

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

where $\rho \geq n$ over the simplex. Clearly, if we start from $x^0 = \frac{1}{n}e$, the analytic center of the simplex, and generate a sequence of points x^k, $k = 1, ...,$, whose potential value is strictly decreased, then when

$$\phi(x^k) - \phi(x^0) \leq -\rho \ln(1/\epsilon),$$

we must have

$$\rho \ln(f(x^k)) - \rho \ln(f(x^0)) \leq -\rho \ln(1/\epsilon)$$

or

$$\frac{f(x^k)}{f(x^0)} \leq \epsilon.$$

This is because on the simplex

$$\sum_j \ln(x_j^k) \leq \sum_j \ln(x_j^0), \forall k = 1, ...$$

We now describe a first order steepest descent potential reduction algorithm in the next section.

2 Steepest-Descent Potential Reduction and Complexity Analysis

Note that the gradient vector of the potential function of $x > 0$ is

$$\nabla \phi(x) = \frac{\rho}{f(x)} \nabla f(x) - X^{-1} e.$$

where in this note X denotes the diagonal matrix whose diagonal entries are elements of vector x.

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*

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

Lemma 2. *For any $x > 0$ and $x \neq x^*$, a matrix $A \in R^{m \times n}$ with $Ax = Ax^*$, and a vector $\bar{\lambda} \in R^m$, consider vector*

$$p(x) = X\left(\nabla\phi(x) - A^T\bar{\lambda}\right).$$

Then,

$$\|p(x)\| \geq 1.$$

Proof. First,

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

If any entry of $(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda})$ is equal or less than 0, then $\|p(x)\| \geq \|p(x)\|_\infty \geq 1$. On the other hand, if $\left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right) > 0$, we have $\left(\nabla f(x) - \frac{f(x)}{\rho}A^T\bar{\lambda}\right)^T x^* \geq 0$. Then, from convexity and $Ax = Ax^*$,

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

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 $\rho \geq n + \sqrt{n}$. □

For any given $x > 0$ in the simplex and any d with $e^T d = 0$,

$$f(x+d) - f(x) \le \nabla f(x)^T d + \frac{\gamma}{2}\|d\|^2 \le \nabla f(x)^T d + \frac{\gamma}{2}\|XX^{-1}d\|^2 \le \nabla f(x)^T d + \frac{\gamma}{2}\|X^{-1}d\|^2,$$

where the last inequality is due to $\|X\| \le 1$. Let $\|X^{-1}d\| = \beta < 1$ and $x^+ = x + d = X(e + X^{-1}d) > 0$. Then, from Lemma 1

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

The first order steepest descent potential reduction algorithm would update x by solving

$$\begin{array}{ll} \text{Minimize} & \nabla \phi(x)^T d \\ \text{Subject to} & e^T d = 0, \ \|X^{-1}d\| \le \beta; \end{array} \tag{2}$$

or

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

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

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

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

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

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

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

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

so that

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

or

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

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

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

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

steps; ot it generates a x^k with $f(x^k) \le \epsilon$ in no more than

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

3 Further Remarks

First, we relax the assumption that $f(x^*) = 0$ where x^* is a minimizer. As in the (primal) interior-point potential reduction algorithm, we consider

$$\phi(x) = \rho \ln(f(x) - \lambda) - \sum_j \ln(x_j),$$

where λ is any lower bound of the objective function. Then, during the potential reduction process, if the norm of the scaled gradient projection $\|p(x)\| < 1$, one can generate a new $\lambda^+(> \lambda)$ that remains a lower bound of the objective function; see, e.g., [1, 6] More precisely, consider the Karmarkar canonical linear programming form and its dual

$$\begin{array}{lll} \min & c^T x & \\ s.t. & Ax = 0; & \\ & e^T x = 1; & \\ & x \geq 0. & \end{array} \qquad \begin{array}{ll} \max & \lambda \\ s.t. & A^T y + e \cdot \lambda \leq c. \end{array}$$

For any feasible (y^0, λ^0) for the dual, λ^0 would be a lower bound for the primal objective. Then we define

$$f(x) = (c - A^T y^0 - e \cdot \lambda^0)^T x - + \frac{1}{2}\|Ax\|^2.$$

One can verify $f(x) \geq 0$ since it is the sum of two nonnegative terms for any $x \geq 0$ on the simplex. When $\|p(x)\| < 1$, one must have a new feasible (y^1, λ^1) for the dual and $\lambda^1 > \lambda^0$.

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.