On First-Order Potential Reduction Algorithms for Linear Programmming*

Yinyu Ye
Stanford University

July 30, 2015; updated May 20, 2016

Abstract

We describe primal, dual and primal-dual steepest-descent potential reduction methods for linear and convex minimization on the simplex in R^n and present their precise iteration complexity analyses depending on the problem dimension and the Lipschitz parameter of the constraint matrix or the convex objective function. In these methods, no matrix needs to be ever inversed so that they are pure first-order methods. We also propose a reformulation of linear programming where a Gram matrix needs to be only inversed or factorized once, and present its iteration complexity bound that is independent of any Lipschitz constant.

1 Convex Optimization with the Simplex Constraint

We consider the following optimization problem on 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 a firs-order (primal) potential reduction algorithm was proposed. In this note we develop a primal-dual first potential reduction method and prove its convergence speed is improved by a factor $\sqrt{n}$, which method and its analysis resemble those in [9].

We assume that $f(x)$ is a convex function in $x \in R^n$. Then, there is a (Wolf) dual problem

$$\begin{array}{ll} \text{Maximize} & \lambda + f(z) - \nabla f(z)^T z \\ \text{Subject to} & \nabla f(z) - e\lambda \geq 0, \\ & e^T z = 1;\ z \geq 0. \end{array} \tag{2}$$

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

The weak duality theorem is, for any feasible primal x and any feasible dual (λ, z),

$$\begin{aligned} & f(x) - (\lambda + f(z) - \nabla f(z)^T z) \\ = & f(x) - f(z) + (\nabla f(z) - e\lambda)^T z \\ = & f(x) - f(z) - (\nabla f(z) - e\lambda)^T (x - z) + (\nabla f(z) - e\lambda)^T x \\ = & f(x) - f(z) - \nabla f(z)^T (x - z) + (\nabla f(z) - e\lambda)^T x \\ \geq & (\nabla f(z) - e\lambda)^T x \\ \geq & 0, \end{aligned} \tag{3}$$

where the first inequality is from the convexity of f, and the second inequality is due to $(\nabla f(z) - e\lambda) \geq 0$ and $x \geq 0$.

Note that there is always an interior dual feasible point, that is, a (z, λ) such that $s = \nabla f(z) - e\lambda > 0$, where s is called dual slack vector. In particular, when $x = z$, then

$$f(x) - (\lambda + f(x) - \nabla f(x)^T x) = (\nabla f(x) - e\lambda)^T x. \tag{4}$$

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 [3]:

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

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

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

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

can be represented as a homogeneous linear feasibility problem (Ye et al. [6]):

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

2 Primal Potential Reduction and Complexity Analysis

We first present a first-order primal potential reduction algorithm, assume that the minimal value $f(x^*) = 0$ where x^* is a minimizer of the problem. Our complexity analysis resembles those in [8], 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.

2.1 Primal Potential Function

Concider the potential function (e.g., see [3, 5, 2, 9])

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

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

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

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 ([3, 2, 9]):

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

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

We now prove the second lemma:

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 $\rho \geq n + \sqrt{n}$. $\square$

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

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

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

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

2.2 The Gradient Projection for Potential Reduction

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

$$\begin{array}{cc} \text{Minimize} & \nabla\phi(x)^T d \\ \text{Subject to} & e^T d = 0, \; \|X^{-1}d\| \leq \beta; \end{array} \tag{5}$$

or

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

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 (7) would be

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

and

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

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

Thus,

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

For $\beta \leq 1/2$, 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

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

so that

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

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

2.3 Initial Points and Running Time Analysis

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*

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

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

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

iterations.

3 Primal-Dual Potential Reduction and Complexity Analysis

We first present a first-order primal-dual potential reduction algorithm, where our complexity analysis resembles those in [9].

3.1 Primal-Dual Potential Function

We consider the primal-dual potential function (e.g., see [5, 2, 9, 1])

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

where parameter $\rho = n + \sqrt{n}$, and

$$\delta = \lambda + f(z) - \nabla f(z)^T z = f(z) - s^T z$$

for some

$$s = \nabla f(z) - e\lambda > 0, \quad e^T z = 1,\ z > 0,$$

that is, an interior feasible (λ, z) to the dual problem of (2). From duality relation (3), δ is a lower bound on the minimal value of problem (12).

Clearly, due to (3), $f(x) - \delta \geq s^T x$ so that we always have

$$n \ln(f(x) - \delta) - \sum_j \ln(x_j s_j) \geq n \ln(s^T x) - \sum_j \ln(x_j s_j) \geq n \ln(n).$$

Thus, if we start from an interior solution pair (x^0, δ^0, s^0) and generate a sequence of points (x^k, δ^k, s^k) with

$$\delta^k = \lambda^k + f(z^k) - \nabla f(z^k)^T z^k = f(z^k) - (s^k)^T z^k,\ s^k = \nabla f(z^k) - e\lambda^k > 0, \quad k = 1, ...,$$

whose potential value is strictly decreased, then, when

$$\phi(x^k, \delta^k) \leq (\rho - n)\ln(f(x^0) - \delta^0) + (\rho - n)\ln(\epsilon) + n\ln(n),$$

we must have

$$(\rho - n)\ln(f(x^k) - \delta^k) \leq (\rho - n)\ln(f(x^0) - \delta^0) + (\rho - n)\ln(\epsilon)$$

or

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

Note that the gradient vector at $x > 0$ of the potential function with respect to x is

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

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

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

where the last inequality is due to $\|X\| \leq 1$. Let $\|X^{-1}d\| = \beta < 1$ and

$$x^+ = x + d = X(e + X^{-1}d) > 0.$$

Then, from Lemma 1

$$\begin{aligned} \phi(x^+, \delta) - \phi(x, \delta) &\leq \rho \ln\left(1 + \tfrac{\nabla f(x)^T d + \frac{\gamma}{2}\|X^{-1}d\|^2}{f(x) - \delta}\right) - e^T X^{-1} d + \tfrac{\beta^2}{2(1-\beta)} \\ &\leq \rho \tfrac{\nabla f(x)^T d + \frac{\gamma}{2}\|X^{-1}d\|^2}{f(x) - \delta} - e^T X^{-1} d + \tfrac{\beta^2}{2(1-\beta)} \\ &= \nabla\phi(x)^T d + \tfrac{\rho\gamma}{2(f(x)-\delta)}\beta^2 + \tfrac{\beta^2}{2(1-\beta)}. \end{aligned} \tag{6}$$

3.2 The Gradient Projection for Potential Reduction

The first-order primal-dual potential reduction algorithm would update x by solving

$$\begin{array}{cc} \text{Minimize} & \nabla\phi(x)^T d \\ \text{Subject to} & e^T d = 0,\ \|X^{-1}d\| \le \beta; \end{array} \tag{7}$$

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.

The minimizer of (7) is the gradient projection

$$X^{-1}d = \frac{-\beta}{\|p(x,\delta)\|} p(x,\delta)$$

where

$$p(x,\delta) = X\left(\nabla\phi(x,\delta) - e\frac{e^T X^2 \nabla\phi(x,\delta)}{\|x\|^2}\right).$$

Expanding the expression of $p(x,\delta)$, we have

$$\begin{array}{rl} p(x,\delta) & = X\left(\frac{\rho}{f(x)-\delta}\nabla f(x) - X^{-1}e - e\frac{e^T X^2\nabla\phi(x,\delta)}{\|x\|^2}\right) \\ & = \frac{\rho}{f(x)-\delta} X\left(\nabla f(x) - e\lambda(x,\delta)\right) - e, \end{array} \tag{8}$$

where

$$\lambda(x,\delta) = \frac{f(x)-\delta}{\rho}\cdot\frac{e^T X^2 \nabla\phi(x,\delta)}{\|x\|^2}. \tag{9}$$

Lemma 4. *Let $\rho = n + \sqrt{n}$ and $\mu = \frac{f(x)-\delta}{n}$. Then, if, for some constant $\alpha \in (0, 1/2]$,*

$$\|p(x,\delta)\| \le \alpha\sqrt{n/(n+\alpha^2)},$$

then the following three inequalities hold:

$$\begin{array}{rl} s(x,\delta) := \nabla f(x) - e\lambda(x,\delta) & > 0, \text{ where } \lambda(x,\delta) \text{ is given by (9)} \\ \|Xs(x,\delta) - \mu^+ e\| & \le \alpha\mu^+ \text{ where } \mu^+ = \frac{x^T s(x,\delta)}{n} \\ \frac{\mu^+}{\mu} & \le \frac{n}{\rho}\cdot\left(1 + \frac{\alpha}{\sqrt{n+\alpha^2}}\right) \le (1 - \frac{\alpha}{2\sqrt{n}}). \end{array}$$

Proof. The proof is by contradiction. If the first inequality of the lemma is not true, then there is at least one index j such that $x_j s_j(x,\delta)) \le 0$ and

$$\|p(x,\delta)\| \ge 1 - \frac{\rho}{f(x)-\delta} x_j s_j(x,\delta) \ge 1,$$

which is a contradiction.

If the second inequality does not hold, then from (8)

$$\begin{array}{rl} \|p(x,\delta)\|^2 & = \|\frac{\rho}{n\mu} Xs(x,\delta) - e\|^2 \\ & = \|\frac{\rho}{n\mu} Xs(x,\delta) - \frac{\rho\mu^+}{n\mu}e + \frac{\rho\mu^+}{n\mu}e - e\|^2 \\ & = (\frac{\rho}{n\mu})^2\|Xs(x,\delta) - \mu^+ e\|^2 + \|\frac{\rho\mu^+}{n\mu}e - e\|^2 \\ & > (\frac{\rho\mu^+}{n\mu})^2\alpha^2 + (\frac{\rho\mu^+}{n\mu} - 1)^2 n \\ & \ge \alpha^2\frac{n}{n+\alpha^2}, \end{array} \tag{10}$$

where the last relation prevails since the quadratic term yields the minimum value when

$$\frac{\rho\mu^+}{n\mu} = \frac{n}{n+\alpha^2}.$$

Finally, in view of inequality (10) we have

$$\begin{array}{ll} \alpha^2 n/(n+\alpha^2) & \geq \|p(x,\delta)\|^2 \\ & \geq (\frac{\rho\mu^+}{n\mu} - 1)^2 n \end{array}$$

implies

$$\frac{\mu^+}{\mu} \leq \frac{n}{\rho} \cdot \left(1 + \frac{\alpha}{\sqrt{n+\alpha^2}}\right).$$

$\square$

Now we consider two alternative cases: $\|p(x,\delta)\| \geq \alpha\sqrt{n/(n+\alpha^2)}$ and otherwise. We have the following theorem that resembles one in [9]:

Theorem 5. *For any given $x > 0$ on the simplex and $\delta = \lambda + f(z) - \nabla f(z)^T z$ where dual variables $s = \nabla f(z) - e\lambda > 0$, let $p(x,\delta)$ be given by (8) and $\lambda(x,\delta)$ be given by (9). Then if $\|p(x,\delta)\| \geq \alpha\sqrt{n/(n+\alpha^2)}$, assigning $x^+ = x + d = X\left(e - \frac{\beta}{\|p(x,\delta)\|}p(x,\delta)\right)$ we can choose β such that*

$$\phi(x^+,\delta) - \phi(x,\delta) \leq -\frac{n\alpha^2}{n+\alpha^2} \cdot \frac{f(x)-\delta}{2(2(f(x)-\delta)+\rho\gamma)};$$

otherwise, assigning $s^+ = \nabla f(x) - e\lambda(x,\delta)$ and $\delta^+ = f(x) - x^T s^+$ we have

$$\phi(x,\delta^+) - \phi(x,\delta) \leq -\frac{\alpha}{2} + \frac{\alpha^2}{2(1-\alpha)}.$$

Proof. In the formal case, from (6)

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

Thus,

$$\begin{array}{ll} \phi(x^+,\delta) - \phi(x,\delta) & \leq \nabla\phi(x)^T d + \frac{\rho\gamma}{2(f(x)-\delta)}\beta^2 + \frac{\beta^2}{2(1-\beta)} \\ & \leq -\beta\alpha\sqrt{n/(n+\alpha^2)} + \frac{\rho\gamma}{2(f(x)-\delta)}\beta^2 + \frac{\beta^2}{2(1-\beta)}. \end{array}$$

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

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

Thus, one can choose β to minimize the quantity at

$$\beta = \frac{\alpha\sqrt{n/(n+\alpha^2)}}{2 + \frac{\rho\gamma}{f(x)-\delta}} \leq 1/2$$

so that

$$\phi(x^+,\delta)-\phi(x,\delta)\le -\frac{n\alpha^2}{n+\alpha^2}\cdot\frac{f(x)-\delta}{2(2(f(x)-\delta)+\rho\gamma)}.$$

On the other hand, when $\|p(x,\delta)\|\le \alpha\sqrt{n/(n+\alpha^2)}$, the three inequalities hold in Lemma 4: The first indicates that $(\lambda^+=\lambda(x,\delta), z^+=x)$ is an *interior* dual feasible solutions with positive slacks $s^+=s(x,\delta)$. Using the second inequality and applying Lemma 1 to vector Xs^+/μ^+ with $\mu^+=x^Ts^+/n$, we have

$$\begin{aligned}
n\ln(x^Ts^+)-\textstyle\sum_{j=1}^n\ln(x_js_j^+) &= n\ln(x^Ts^+/\mu^+)-\textstyle\sum_{j=1}^n\ln(x_js_j^+/\mu^+)\\
&= n\ln n-\textstyle\sum_{j=1}^n\ln(x_js_j^+/\mu^+)\\
&\le n\ln n+\tfrac{\|Xs^+/\mu^+-e\|^2}{2(1-\|Xs^+/\mu^+-e\|_\infty)}\\
&\le n\ln n+\tfrac{\alpha^2}{2(1-\alpha)}\\
&\le n\ln x^Ts-\textstyle\sum_{j=1}^n\ln(x_js_j)+\tfrac{\alpha^2}{2(1-\alpha)}\\
&\le n\ln(f(x)-\delta)-\textstyle\sum_{j=1}^n\ln(x_js_j)+\tfrac{\alpha^2}{2(1-\alpha)}.
\end{aligned}$$

Then, according to the third inequality and $\delta^+=f(x)-x^Ts^+$, we have

$$\sqrt{n}\ln(f(x)-\delta^+)-\sqrt{n}\ln(f(x)-\delta)=\sqrt{n}\ln(n\mu^+)-\sqrt{n}\ln(n\mu)=\sqrt{n}\ln\frac{\mu^+}{\mu}\le-\frac{\alpha}{2}.$$

Adding the two inequalities, we have

$$\phi(x,\delta^+)-\phi(x,\delta)\le-\frac{\alpha}{2}+\frac{\alpha^2}{2(1-\alpha)}.$$

This completes the proof. $\square$

Using the gradient projection, our proposed primal-dual potential reduction algorithm would generate a sequence of (x^k,δ^k,s^k):

$$\begin{array}{ll}
\text{if } \|p(x,\delta)\|\ge\alpha\frac{n}{n+\alpha^2} & : x^{k+1}=X^k\left(e-\frac{\beta}{\|p(x^k,\delta^k)\|}p(x^k,\delta^k)\right),\ s^{k+1}=s^k,\ \delta^{k+1}=\delta^k;\\
\quad\text{otherwise} & : s^{k+1}=\nabla f(x^k)-e\lambda(x^k,\delta^k), \delta^{k+1}=f(x^k)-(x^k)^Ts^{k+1}, x^{k+1}=x^k.
\end{array}$$

That is, we update primal and dual alternatively, and the potential function is reduced in either way. One can see that the dual-update creates a constant reduction, but the primal-update creates a much smaller reduction depending on duality gap $f(x)-\delta$.

3.3 Initial Points and Running Time Analysis

Starting from $x^0=\frac{1}{n}e$ and $z^0=x^0$, we select λ^0 sufficiently lower such that $s^0=\nabla f(x^0)-e\lambda^0>0$ and

$$\|X^0s^0-\mu^0e\|\le\alpha\mu^0,\text{ where }\mu^0=(x^0)^Ts^0/n=(f(x^0)-\delta^0)/n,\ \delta^0=f(x^0)-(x^0)^Ts^0. \tag{11}$$

Our analysis would consider the least reduction by the primal

$$\begin{aligned}
\phi(x^{k+1},\delta^{k+1})-\phi(x^k,\delta^k) &\le -\tfrac{n\alpha^2}{n+\alpha^2}\cdot\tfrac{f(x^k)-\delta^k}{2(2(f(x^k)-\delta^k)+\rho\gamma)}\\
&= -\tfrac{n\alpha^2}{n+\alpha^2}\cdot\tfrac{n\mu^k}{2(2n\mu^k+\rho\gamma)}\\
&= -\tfrac{n\alpha^2}{n+\alpha^2}\cdot\tfrac{\mu^k}{2(2\mu^k+\rho\gamma/n)}.
\end{aligned}$$

As n sufficiently large, the first term $\frac{n\alpha^2}{n+\alpha^2}$ becomes constant α^2 and $\rho\gamma/n = (n+\sqrt{n})\gamma/n$ becomes just γ. Also, we expect $2\mu^k \leq 2\mu^0 << \gamma$. Thus, we have

$$\phi(x^{k+1}, \delta^k) - \phi(x^k, \delta^k) \leq -\alpha^2 \frac{\mu^k}{3\gamma}, \;\; k = 1, 2, ...$$

where we recall that the average duality gap $\mu^k = (f(x^k) - \delta^k)/n$.

After K iterations, by the choice of the initial point (11) we have

$$\begin{array}{ll} \frac{\alpha^2}{3\gamma}\sum_{k=0}^{K}\mu^k & \leq \phi(x^0, \delta^0) - \phi(x^{K+1}, \delta^{K+1}) \\ & = \sqrt{n}\ln(\mu^0/\mu^{K+1}) + (n\ln(n\mu^0) - \sum_j \ln(x_j^0 s_j^0)) \\ & \quad -(n\ln(n\mu^{K+1}) - \sum_j \ln(x_j^{K+1}s_j^{K+1})) \\ & \leq \sqrt{n}\ln(\mu^0/\mu^{K+1}) + n\ln(n) + \frac{\alpha^2}{2(1-\alpha)} - n\ln(n) \\ & = \sqrt{n}\ln(\mu^0/\mu^{K+1}) + \frac{\alpha^2}{2(1-\alpha)}. \end{array}$$

If still $\mu^{K+1}/\mu^0 \geq \epsilon$, we must have

$$\frac{\alpha^2}{3\gamma}\sum_{k=0}^{K}\mu^k \leq \sqrt{n}\ln(1/\epsilon) + \frac{\alpha^2}{2(1-\alpha)}.$$

Thus, there must be one $k \in (0, K]$ such that $\mu^k/\mu^0 \leq \epsilon$ after $K = O(\gamma\sqrt{n}\ln(1/\epsilon)/(\epsilon\mu^0))$ iterations.

Theorem 6. *Let $\mu^k = (f(x^k) - \delta^k)/n$. Then, the primal-dual first-order potential reduction algorithm terminates in $O(\gamma\sqrt{n}\ln(1/\epsilon)/(\epsilon\mu^0))$ iterations such that the duality gap ratio*

$$\frac{\mu^k}{\mu^0} \leq \epsilon;$$

or in $O(\gamma\sqrt{n}\ln(\mu^0/\epsilon)/\epsilon)$ iterations such that the average duality gap

$$\mu^k \leq \epsilon.$$

In practice, one may use the Mehrotra's predictor and corrector algorithm [4] to improve the practical efficiency. In particular, the high-order or conjugate gradient correction in the primal-update may further reduce the dependency on γ for the complexity bound.

3.4 Convex Optimization with the Box Constraint

We can consider the following optimization problem over the box constraint:

$$\begin{array}{cc} \text{Minimize} & f(x) \\ \text{Subject to} & 0 \leq x \leq e. \end{array} \tag{12}$$

Then, we consider the primal potential function (assuming $f(x^*) = 0$)

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

where $\rho = 2n + \sqrt{2n}$ over the simplex. Clearly, $x^0 = \frac{1}{2}e$, is the analytic center of the box.

Furthermore, we can consider the primal-dual potential function:

$$\phi(x, \delta) = \rho \ln(f(x) - \delta) - \sum_j \ln(x_j(1 - x_j)) - \sum_j \ln(-s_j y_j),$$

where parameter $\rho = 2n + \sqrt{2n}$, and

$$\delta = f(z) - s^T z - y^T(1 - z)$$

and some variables

$$s = \nabla f(z) + y > 0, \quad y > 0, \quad 0 < z < 1.$$

Again, from duality relation, δ is a lower bound on the minimal value of the box constrained problem.

Using a similar analysis, we have

Corollary 7. *Let $\mu^k = (f(x^k) - \delta^k)/n$. Then, the primal-dual first-order potential reduction algorithm terminates in $O(\gamma\sqrt{n}\ln(1/\epsilon)/(\epsilon\mu^0))$ iterations such that the duality gap ratio*

$$\frac{\mu^k}{\mu^0} \leq \epsilon;$$

or in $O(\gamma\sqrt{n}\ln(\mu^0/\epsilon)/\epsilon)$ iterations such that the average duality gap

$$\mu^k \leq \epsilon.$$

4 Complexity of the Algorithm for Linear Programming

As we mentioned earlier, linear programming problems can be casted as finding a (primal) feasible point x to meet

$$\begin{aligned} Ax &= 0; \\ e^T x &= 1; \\ x &\geq 0; \end{aligned}$$

and it can be formulated as our model (12) with $f(x) = \frac{1}{2}\|Ax\|^2$, where A has a full row rank. Then, the Lipschitz parameter γ of $f(x)$ in our complexity bounds would be the largest eigenvalue of matrix $A^T A$.

One way to remove the dependency on γ is to consider

$$\begin{array}{ll} \text{Minimize} & f(x) := \frac{1}{2}\|P_A x\|^2 \\ \text{Subject to} & e^T x = 1, \ x \geq 0; \end{array} \tag{13}$$

where the (positive semidefinite) projection matrix

$$P_A = A^T(AA^T)^{-1}A$$

and it largest eigenvalue is 1. Note that $P_A x = 0$ if and only if $Ax = 0$.

Since $P_A^T P_A = P_A$, the Hessian matrix of $f(x)$ is P_A so that its Lipschitz parameter is $\gamma = 1$. Thus, we have the following corollary

Corollary 8. *Let $\mu^k = (f(x^k) - \delta^k)/n$. Then, the primal-dual first-order potential reduction algorithm, for solving the reformulated linear feasibility problem (13), terminates in $O(\sqrt{n}\ln(1/\epsilon)/(\epsilon\mu^0))$ iterations such that the duality gap ratio*

$$\frac{\mu^k}{\mu^0} \leq \epsilon;$$

or in $O(\sqrt{n}\ln(\mu^0/\epsilon)/\epsilon)$ iterations such that the average duality gap

$$\mu^k \leq \epsilon.$$

Note that AA^T need only be inversed or factorized once for computing the gradient vector of the new objective function.

One may also consider find a (dual) feasible point (y, s) to meet

$$\begin{array}{c} A^T y + s = 0; \\ e^T s = 1; \\ s \geq 0; \end{array}$$

One can cast the problem as (12) with objective $f(s) = \frac{1}{2}\|(I - P_A)s\|^2$ and solving

$$\begin{array}{cl} \text{Minimize} & \frac{1}{2}\|(I - P_A)s\|^2 \\ \text{Subject to} & e^T s = 1,\ s \geq 0. \end{array} \tag{14}$$

Again, the Hessian matrix of $f(s)$ is projection matrix $I - P_A$ so that its Lipschitz parameter is $\gamma = 1$. Solving the dual seems make more sense since the objective function is exactly equivalent to the original one without changing the metric.

Note further that, when $\|X^{-1}d\|_\infty \leq 1/2$, the potential reduction is upper-bounded by a quadratic function

$$\begin{array}{rl} \phi(x^+) - \phi(x) & \leq \rho \ln\left(1 + \frac{\nabla f(x)^T d + \frac{\gamma}{2}\|d\|^2}{f(x)}\right) - e^T X^{-1} d + \|X^{-1}d\|^2 \\ & \leq \frac{\rho}{f(x)}\left(\nabla f(x)^T X d' + \frac{\gamma}{2} d'^T X^2 d'\right) - e^T d' + \|d'\|^2 \\ & = \left(\frac{\rho}{f(x)} X \nabla f(x) - e\right)^T d' + \frac{1}{2} d'^T \left(\frac{\gamma\rho}{f(x)} X^2 + 2I\right) d', \end{array}$$

with direction vector $d' = X^{-1}d$. This is a separable quadratic function so that each iteration can be carried out as separable quadratic minimization over the simplex, rather than the gradient projection onto the simplex as it is discussed earlier.

One can replace the simplex constraint with the box-constraint. Or remove the simplex constraint and treat the problem as a completely unconstrained problem. At the end you can scale the final solution x by $\sum_j x_j$ to obtain the solution on the simplex.

References

[1] R. M. Freund. Polynomial–time algorithms for linear programming based only on primal scaling and projected gradients of a potential function. *Math. Programming*, 51:203–222, 1991.

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

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

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

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

[6] 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.

[7] Y. Ye, On a First-Order Potential Reduction Algorithm for Linear Programming, Teaching Note, http://www.stanford.edu/˜yyye/FO-potential-reduction.pdf, 2015.

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

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