Line Search in Potential Reduction Algorithms for Linear Programming

Yinyu Ye †

September 1989

Abstract

We describe several line search strategies in recent potential reduction algorithms for linear programming. We clarify some concerns about the step size of the original algorithm. In particular, we illustrate that the dual step of the algorithm can be sufficiently "long". We also discuss some other implementation issues for the algorithm.

Keywords: Linear programming, line search, potential reduction algorithms.

Abbreviated title: Line Search in Potential Reduction Algorithms

† Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242

Since Karmarkar proposed his projective algorithm [5], various primal-dual potential reduction algorithms for linear programming have been developed by Anstreicher and Bosch [1], Freund [2], Gonzaga and Todd [4], Kojima, Mizuno and Yoshise [6], Liu and Goldfarb [7], McShane, Monma and Shanno [8], and Ye [10][11] among others. All of these algorithms are based on reducing a primal-dual potential function that is first appeared in Todd and Ye [9]. They showed that a Newton-type step can reduce the function by a constant from an interior point close to the central-path. Soon after, Ye [10] showed that it can be reduced by a constant from any interior point.

Since it is a function-driven algorithm, the potential reduction algorithm is quite generic. Many different strategies have been used in various forms. For examples, Ye developed the primal scaling and dual scaling forms of the algorithm in both affine space and projective space. Anstreicher and Bosch proposed the first step-size-free method. Freund also studied the primal scaling form and its step sizes. Kojima et al. extended this potential function to their and Monteiro and Adler's symmetric primal-dual scaling algorithm and it works. McShane et al. achieved impressive implementation results and compared practical performance of several algorithm parameters in the primal-dual scaling form. Liu and Goldfarb first worked out the primal scaling method to convex quadratic programming. Finally, Gonzaga and Todd developed either-primal-or-dual scaling form with very interesting discussions.

The aim of this paper is to provide a unified structure in these various approaches and describe several line search strategies. In particular, we illustrate that the dual step of the primal scaling form can be sufficiently "long" with the help of line searches, in contrast to a statement that it should be always "short". We also discuss some other implementation issues, including simultaneous movement of the primal and dual, their separate step sizes, and the role of ρ in the potential reduction algorithm.

1. A Unified Structure

Consider linear programs in the following standard form:

$$\text{LP} \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \ x \geq 0, \end{array}$$

where $c \in R^n$, $A \in R^{m \times n}$ and $b \in R^m$ are given, $x \in R^n$, and ^{T} denotes the transpose. The dual to LP can be written as

$$\text{LD} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & s = c - A^T y \geq 0, \end{array}$$

where vector $y \in R^m$ and $s \in R^n$. The components of s are called dual slacks. In this note, the upper-case letter (X) is the diagonal matrix of the vector (x) in lower-case, and e is the vector of all ones.

We consider the primal-dual potential function

$$\phi(x, s) = \rho \ln(x^T s) - \sum_{j=1}^{n} \ln(x_j s_j).$$

If ϕ is reduced below $-(\rho - n)L$, then

$$c^T x - b^T y = x^T s \leq 2^{-L}.$$

Given an interior primal solution x^k and dual solution y^k such that

$$Ax^k = b, \quad \text{and} \quad x^k > 0$$

and

$$s^k = c - A^T y^k > 0,$$

all potential reduction algorithms generate search directions via solving the system of linear equations

$$D_x \delta x + D_s \delta s = \frac{\Delta^k}{\rho} e - X^k s^k$$

$$(0)$$

$$A\delta x = 0 \quad \text{and} \quad \delta s = -A^T \delta y,$$

where $\Delta^k = (x^k)^T s^k$. The last two equations indicate that δx is in the null space of A and δs in the row space of A. These two linear spaces are orthogonal. The algorithms differ from choosing D_x and D_s.

Primal scaling form:

$$D_x = \frac{\Delta^k}{\rho}(X^k)^{-1}, \quad \text{and} \quad D_s = X^k. \tag{1}$$

Dual scaling form:

$$D_x = S^k, \quad \text{and} \quad D_s = \frac{\Delta^k}{\rho}(S^k)^{-1}. \tag{2}$$

Primal-dual scaling form:

$$D_x = S^k, \quad \text{and} \quad D_s = X^k. \tag{3}$$

Either-primal-or-dual scaling form:

Use (1) and (2) alternatively.

There are several options to choose step sizes. Let β_x and β_s designate the step sizes for the primal and dual, respectively,

$$x^{k+1} = x^k + \beta_x \delta x \quad \text{and} \quad s^{k+1} = s^k + \beta_s \delta s.$$

For the primal and dual scaling forms, Ye [10] essentially proved that there exist $\beta_x > 0$ and $\beta_s = 0$, or $\beta_x = 0$ and $\beta_s > 0$ such that

$$\phi(x^{k+1}, s^{k+1}) \leq \phi(x^k, s^k) - 0.05.$$

For the primal-dual scaling form, Kojima et al. showed that there exist $\beta_x = \beta_s > 0$ such that

$$\phi(x^{k+1}, s^{k+1}) \leq \phi(x^k, s^k) - 0.2.$$

These results can be put into a more general theorem:

Theorem

Let $\rho \geq n + \Omega(\sqrt{n})$, and let δx and δs be resulted from the linear system (0) with D_x and D_s given by (1), (2) or (3). Then, there exist $\beta_x \geq 0$ and $\beta_s \geq 0$ such that

$$\phi(x^{k+1}, s^{k+1}) \leq \phi(x^k, s^k) - 0.05.$$

The theorem is trivially true, but it reflects many practical considerations. For examples, McShane et al. found that taking different step sizes for the primal and dual steps outperforms taking equal step sizes in the primal-dual scaling form, and Gonzaga and Todd proved that taking either primal or dual (not both) will not result in R-order convergence in the primal and dual scaling forms. Overall, a more flexible algorithm can be described as

Potential Reduction Algorithm:

Given $Ax^0 = b, x^0 > 0$ and $s^0 = c - A^T y^0 > 0$;

let $\Delta^0 = (x^0)^T s^0$ and set $k = 0$;

while $\Delta^k \geq 2^{-L}$ **do**

begin

compute δx and δs from the system (0) of (1), (2) or (3);

$x^{k+1} = x^k + \beta_x^* \delta x$ and $s^{k+1} = s^k + \beta_s^* \delta s$

where

$$(\beta_x^*, \beta_s^*) = \arg \min_{\beta_x, \beta_s \geq 0} \phi(x^k + \beta_x \delta x, s^k + \beta_s \delta s);$$

$\Delta^{k+1} = (x^{k+1})^T s^{k+1}$ and $k = k + 1$;

end.

Note again that the directions δx and δs are orthogonal and the potential function are almost separable, the bi-directional search in the above algorithm can be computed quite efficiently. In practice, the strategy of McShane et al., 99 percent to the boundary, may suffice.

2. Dual Step in the Primal Scaling Form

Now let us take a look how the primal scaling form works for the following simple example:

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{n-1} x_j \\ \text{subject to} & e^T x = n, \; x \geq 0, \end{array}$$

Let $x^0 = e$ and $y^0 = -n$, i.e., $\Delta^0 = n^2 + n - 1$. Then, one can verity that

$$\delta y = \frac{(\rho - n)\Delta^0}{n\rho} > 0$$

and

$$\|(X^0)^{-1}\delta x\| = \frac{\rho\sqrt{n(n-1)}}{n\Delta^0}.$$

For $\rho = n + \sqrt{n}$, we have

$$\|(X^0)^{-1}\delta x\| = \frac{(n+\sqrt{n})\sqrt{n(n-1)}}{n(n^2+n-1)} < 1/n.$$

This inequality indicates that $y^1 = y^0 + \delta y$ is an improving dual solution and the new primal-dual gap is reduced by a ratio $\sqrt{n}/(1+\sqrt{n})$. However, the function-driven feature of the algorithm enables us to do more—to use a line search along the direction δs to minimize $\phi(x^0, s^0 + \beta_s \delta s)$, where $\delta s = -\delta y e$. This essentially minimizes

$$(n+\sqrt{n})\ln(n-1-ny) - (n-1)\ln(1-y) - \ln(-y)$$

where the minimum is

$$y^1 = \frac{1 - \sqrt{1 + 4(n-1)/(n\sqrt{n})}}{2} > -1/\sqrt{n}.$$

Thus, with the help of the line search we can bring the dual objective value at least to $-\sqrt{n}$. Then, the new primal-dual gap is $n+\sqrt{n}-1$, comparing to the old gap n^2+n-1. The gap variation is

$$\frac{n+\sqrt{n}-1}{n^2+n-1} \leq O(1/n).$$

This is much better than the theoretical bound $\sqrt{n}/(1+\sqrt{n})$.

In the large-steps path-following potential reduction method [3] the primal-dual gap is updated according to a pre-decided ratio, no more and no less, given by

$$\frac{\Delta^{k+1}}{\Delta^k} = \frac{r}{q}$$

where $q = r + O(\sqrt{n})$. Although $r = 2n$ there, the gap variation is exact $\sqrt{n}/(\sqrt{n} + O(1/2))$. The larger of r, the smaller of the gap variation. Therefore, this updating strategy still limits the permissible stepsize on any iteration.

After we update y^0 from $-n$ to $-1/\sqrt{n}$,

$$\|(X^0)^{-1}\delta x\| \geq \Omega(1),$$

and we can immediately take a primal step to reduce the potential function. Even though the norm of the primal projection vector, $\|(X^0)^{-1}\delta x\|$, is about 1, the line search will move x to

$$x_n^1 = 1 + (n-1)n^{-1/4} \quad x_j^1 = 1 - n^{-1/4} \quad \text{for} \quad j = 1, ..., n-1,$$

reducing the potential function by $O(n^{1/4})$. This gives an additional gap variation $1-n^{-1/4}$. Without going to the next iteration, the dual objective value can immediately be improved, and then the primal objective can again be improved, ... Indeed, if the bi-directional search is implemented, the optimal solutions for this LP problem can be obtained in one iteration. In this case, simultaneous movements achieve more than R-order convergence, as predicted by Gonzaga and Todd [4].

If you are not comfortable with the bi-directional search, you can implement the primal scaling form as suggested by Anstreicher and Bosch [1].

Primal Scaling Form:

Given $Ax^0 = b, x^0 > 0$ and $s^0 = c - A^T y^0 > 0$;

let $\Delta^0 = (x^0)^T s^0$ and set $k = 0$;

while $\Delta^k \geq 2^{-L}$ **do**

begin

compute δx and δs from the system of (1);

$s^{k+1} = s^k + \beta_s^* \delta s$;

where

$$\beta_s^* = \arg \min_{\beta_s \geq 0} \phi(x^k, s^k + \beta_s \delta s);$$

$x^{k+1} = x^k + \beta_x^* \delta x$;

where

$$\beta_x^* = \arg \min_{\beta_x \geq 0} \phi(x^k + \beta_x \delta x, s^{k+1});$$

$\Delta^{k+1} = (x^{k+1})^T s^{k+1}$ and $k = k + 1$;

end.

Here, we always try to update the dual (may or may not actually updated) no matter whether the norm of the primal projection vector is big or small. This will guarantee that the primal projection vector of $\|(X^k)^{-1} \delta x\|$ is always greater than $\Omega(1)$ after the dual update step.

3. The Role of ρ.

ρ plays an important role in potential reduction algorithms. While the theoretical (worst case) result indicates that ρ should be around $n + \sqrt{n}$ (Freund [2]), practical experiments constantly indicate that it is much better for ρ to be around $n + n^2$ (McShane et al [8])—another typical conflict between theoretical and practical best. In fact, recent discussions on "large steps" are actually on large ρ's.

First, let us distinct ρ in the potential function and ρ in the linear system, say, the former is ρ_p and the latter is ρ_s. Second, an interesting result of Ye [11] indicates that ρ_p and ρ_s may not necessarily be the same. In other words, we can still use $\rho_p = n + \sqrt{n}$ in the potential function to measure the iterative process, but use $\rho_s <> n + \sqrt{n}$ in the linear system (0) to obtain δx and δs.

For example, if the ratio of $(\min(X^k s^k)/(x^k)^T s^k)$ is bounded by $\Omega(1/n)$, ρ_s may even be set to ∞ and the $n + \sqrt{n}$ potential function is still reduced by a constant (in this case, x^k and s^k are close to the central path). Thus, an adaptive selection of ρ_s can be implemented according to this ratio. This adds another dimensional search for ρ_s (tri-directional search overall) into the potential reduction algorithm, while ρ_p remains $n + \Omega(\sqrt{n})$ to maintain the best complexity result.

In a coming paper, we show that ρ_s can be adaptively chosen from a wide range in $[n + \sqrt{n}, \infty)$ to satisfy the theoretical and practical needs. More specifically, let

$$h^k = \sum_j \left(\frac{(x^k)^T s^k}{x_j^k s_j^k}\right) - n^2,$$

where by the well-known Harmonic inequality,

$$h^k \geq 0$$

and the equality holds if an only if $X^k s^k$ is a multiple of e.

Let

$$Z = [n + \sqrt{n}, n + \Omega(\sqrt{n} + \sqrt{h^k})] \quad \text{if} \quad h^k > \Omega(n)$$

and

$$Z = [n, n + \infty) \quad \text{if} \quad h^k \leq O(n).$$

Then, for any ρ_s in Z the directions resulted from the system (3) of linear equations are theoretically acceptable, meaning there exits $\beta_x > 0$ and $\beta_s > 0$ such that

$$\phi(x^k + \beta_x \delta x, s^k + \beta_s \delta s) \leq \phi(x^k, s^k) - 0.06$$

while $\rho_p = n + \sqrt{n}$ in ϕ.

An interesting point here is that the larger of the h^k, the larger of ρ_s. Typically, h^k is in the range of $\Omega(n^2)$. Then, ρ_s can be set up to $n + \Omega(n)$. In the case $((x^k)^T s^k / \min(X^k s^k))$ is very large like $\Omega(n^4)$, then h^k is at least $\Omega(n^4)$ and ρ_s can be set up to $n + \Omega(n^2)$, still achieving the best theoretical complexity. If a very large step-size strategy is implemented as it is done by McShane et al., h^k could easily reach that bound.

References

[1] K. M. Anstreicher and R. A. Bosch, "Long steps in a $O(n^3L)$ algorithm for linear programming," manuscript, Yale School of Management (New Haven, CT, 1988).

[2] R. M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function, " manuscript, Sloan School, M.I.T. (Cambridge, MA, 1988).

[3] C. C. Gonzaga, "Large-steps path-following methods for linear programming: Potential reduction method," Technical Report ES-211/89, Department of Systems Engineering and Computer Sciences, COPPE-Federal University of Rio de Janeiro (Rio de Janeiro, Brazil, 1989).

[4] C. C. Gonzaga and M. J. Todd, "An $O(\sqrt{n}L)$-iteration large-step primal-dual affine algorithm for linear programming," Technical Report No. 862, School of ORIE, Cornell University (Ithaca, NY, 1989).

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

[6] M. Kojima, S. Mizuno and A. Yoshise, "An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems," Research Report, Department of Information Sciences, Tokyo Institute of Technology (Tokyo, Japan, 1988).

[7] S. Liu and D. Goldfarb, "Interior point potential function reduction algorithms for solving convex quadratic programs," manuscript, Department of IEOR, Columbia University (New York, NY, 1989).

[8] K. A. McShane, C. L. Monma and D. F. Shanno, "An implementation of a primal-dual interior point method for linear programming," *ORSA Journal on Computing* **1** (1989) 70-83.

[9] M. J. Todd and Y. Ye, "A centered projective algorithm for linear programming," Technical Report 763, School of ORIE, Cornell University (Ithaca, NY, 1987), to appear in *Mathematics of Operations Research*.

[10] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," manuscript (1988), to appear in *Mathematical Programming*.

[11] Y. Ye, "A class of projective transformations for linear programming," manus-cript (1988), to appear in *SIAM J. on Computing*.