IOE 610/Math 660: A note on central paths in LP and a basic primal-dual feasible interior point method[1]

The logarithmic barrier approach to solving a linear program dates back to the work of Fiacco and McCormick in 1967 in their book *Sequential Unconstrained Minimization Techniques*, also known simply as SUMT. The method was not believed then to be either practically or theoretically interesting, when in fact today it is considered to be both! Interest in these methods has reignighted as a consequence of Karmarkar's interior-point method, and has been the subject of an enormous amount of research and computation, even to this day. In these notes we present the basic algorithm and a basic analysis of its performance.

1 Some nonlinear optimization preliminaries

Since barrier, or interior point, methods for solving a linear program essentially transform it into a sequence of nonlinear programming problems, we begin by reviewing some relevant facts of nonlinear programming.

1.1 A few facts from analysis

Recall that we defined a ball centered at point $z \in \mathbb{R}^n$ with radius $r > 0$ to be the set

$$B(z, r) = \{x : \|z - x\| \leq r\}.$$

We call a set $X \subseteq \mathbb{R}^n$ *open* if any point $x \in X$ is an *interior* point of the set, i.e., there exists $r > 0$ such that $B(x, r) \subset X$.

Suppose $f : X \to \mathbb{R}$ is a (nonlinear) continuous function of n variables defined over an open set X. We say that f is *differentiable* at $\bar{x} \in X$ if there exists a vector $\nabla f(\bar{x})$ (the *gradient* of f at $\bar{x}$) such that

$$\lim_{x \to \bar{x}} \frac{f(x) - (f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}))}{\|x - \bar{x}\|} = 0.$$

f is *differentiable on* X if f is differentiable $\forall \bar{x} \in X$. The gradient vector is a vector of partial derivatives:

$$\nabla f(\bar{x}) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right) \Bigg|_{x=\bar{x}}^T .$$

The *directional derivative* of f at $\bar{x}$ in the direction d is

$$\lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T d.$$

1.2 Newton's method for solving systems of nonlinear equations

Suppose we are given a differentiable vector-valued function $G = (g_1, \ldots, g_n)^T : \mathbb{R}^n \to \mathbb{R}^n$, and we need to solve the system of equations

$$G(x) = 0.$$

[1] Based in part on Robert M. Freund notes for the course *15.094: Systems Optimization: Models and Computation* at MIT and the book "Primal-Dual Interior-Point Methods" by S. J. Wright (SIAM, 1997).

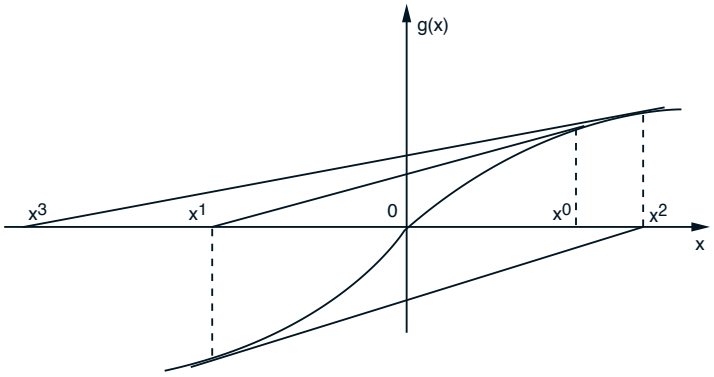
We can apply an iterative method based on the following idea. Given a point $x \in \mathbb{R}^n$, each function g_i at points nearby x can be approximated by $g_i(x+d) \approx g_i(x) + \nabla g_i(x)^T d$. We denote by $\nabla G(x) = [\nabla g_1(x), \ldots, \nabla g_1(x)]$ the Jacobian of G at x, and then we can write

$$G(x+d) \approx G(x) + \nabla G(x)^T d.$$

The iterative algorithm is as follows: starting at a point x_0, we approximate $G(x_0+d) \approx G(x_0) + \nabla G(x_0)^T d$, and provided that $\nabla g(x_0)$ is non-singular, solve the system of linear equations

$$\nabla G(x_0)^T d = -g(x_0)$$

to obtain d (it is referred to as the *Newton step*). Set the next iterate $x_1 = x_0 + d$, and continue. This method is well-studied, and is well-known for its good performance when the starting point x_0 is chosen appropriately. However, for other choices of x_0 the algorithm may not converge, as demonstrated in the following well-known picture:



The above is referred to as the Newton's method (for equation solving).[2]

1.3 Convex functions and gradient inequality

A function $f : S \to \mathbb{R}$, where $S \subseteq \mathbb{R}^n$ is a nonempty convex set is a *convex function* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \ \forall x, y \in S, \ \forall \lambda \in [0,1].$$

If f is continuously differentiable, the following serves as a necessary and sufficient condition of its convexity:

Theorem 1 (Gradient inequality) *Suppose $X \subseteq \mathbb{R}^n$ is a non-empty open convex set, and $f : X \to \mathbb{R}$ is differentiable. Then f is convex iff ("if and only if") it satisfies the* gradient inequality*:*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in X.$$

Proof: Suppose f is convex. Then, for any $\lambda \in (0,1]$,

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) \Rightarrow \frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

[2]There is a closely related Newton's method for optimizing nonlinear functions. It is essentially the Newton's method for equation solving, which is applied to the function $G(x) = \nabla f(x)$, where $f(x)$ is the function being minimized.

Letting $\lambda \to 0$, we obtain: $\nabla f(x)^T(y-x) \leq f(y) - f(x)$, establishing the "only if" part.

Now, suppose that the gradient inequality holds $\forall x, y \in X$. Let w and z be any two points in X. Let $\lambda \in [0,1]$, and set $x = \lambda w + (1-\lambda)z$. Then

$$f(w) \geq f(x) + \nabla f(x)^T(w-x) \text{ and } f(z) \geq f(x) + \nabla f(x)^T(z-x).$$

Taking a convex combination of the above inequalities,

$$\lambda f(w) + (1-\lambda) f(z) \geq f(x) + \nabla f(x)^T(\lambda(w-x) + (1-\lambda)(z-x))$$

$$= f(x) + \nabla f(x)^T 0 = f(\lambda w + (1-\lambda)z),$$

so that $f(x)$ is convex. ∎

In one dimension, the gradient inequality has the form $f(y) \geq f(x) + f'(x)(y-x)\ \forall x, y \in X$.

1.4 Some optimality conditions

Consider the following optimization problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax = b \\ & x \in X, \end{array} \tag{1}$$

where X is a convex open set and $f : X \to \mathbb{R}$ is a convex continuously differentiable function. Suppose the problem is feasible, and thus we will assume when necessary without loss of generality that matrix $A \in \mathbb{R}^{m \times n}$ has a full row rank. The following is a special case of the so-called Karush-Kuhn-Tucker optimality conditions as they apply to problem (1):

Proposition 2 (KKT optimality conditions) *Suppose problem* (1) *is as above. Then a point* $\bar{x} \in X$ *is an optimal solution of* (1) *if an only if there exists a vector* $u \in \mathbb{R}^m$ *that, together with* $\bar{x}$*, satisfies:*

$$A\bar{x} = b,\ \nabla f(\bar{x}) + A^T u = 0. \tag{2}$$

Proof: Suppose KKT conditions are satisfied for some $\bar{x} \in X$ and some u. Suppose x is another feasible point. Then we have, by gradient inequality:

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) = f(\bar{x}) + u^T A(x-\bar{x}) = f(\bar{x}),$$

thus proving that $\bar{x}$ solves the problem.

Conversely, suppose that $\bar{x}$ is an optimal solution. Then we have $\nabla f(\bar{x})^T(x-\bar{x}) \geq 0$. Why? Well, suppose this is not the case and there exists a feasible x such that $\nabla f(\bar{x})^T(x-\bar{x}) < 0$. Since the feasible region of the problem is convex, $\bar{x} + \lambda(x-\bar{x})$ is also feasible for any $\lambda \in (0,1]$. In addition, by definition of the gradient,

$$f(\bar{x} + \lambda(x-\bar{x})) = f(\bar{x}) + \lambda \nabla f(\bar{x})^T(x-\bar{x}) + \lambda \|x-\bar{x}\| \alpha_{\bar{x}}(\lambda(x-\bar{x})).$$

Rearranging the above equality and keeping in mind that $\lambda > 0$, we get:

$$\frac{f(\bar{x} + \lambda(x-\bar{x})) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T(x-\bar{x}) + \|x-\bar{x}\| \alpha_{\bar{x}}(\lambda(x-\bar{x})),$$

where, by the definition of the gradient, function $\alpha_{\bar{x}}(y)$ has limit 0 as $y \to 0$. Letting λ tend to zero (while remaining positive), we observe that for sufficiently small values of $\lambda > 0$, $f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x}) < 0$, which contradicts optimality of $\bar{x}$.

To summarize, for any feasible solution x, $\nabla f(\bar{x})^T(x - \bar{x}) \geq 0$. In other words, the following linear system does not have a solution:

$$\nabla f(\bar{x})^T d < 0, \;\; Ad = 0.$$

Using an appropriate theorem of the alternative (or a duality result), we can show that this implies that (2) has a solution. ∎

2 The barrier problem and approximately optimal solutions

Consider the linear programming problem in standard form:

$$\begin{array}{lll} \text{(P)} & \min & c^T x \\ & \text{s.t.} & Ax = b \\ & & x \geq 0, \end{array}$$

where x is a vector of n variables, whose standard linear programming dual problem is:

$$\begin{array}{lll} \text{(D)} & \max & b^T p \\ & \text{s.t.} & A^T p + s = c \\ & & s \geq 0. \end{array}$$

Given a feasible solution x of (P) and a feasible solution (p, s) of (D), the duality gap is simply

$$c^T x - b^T p = x^T s \geq 0.$$

Recall that we can write the optimality conditions for both (P) and (D) as follows:

$$\begin{array}{rrrl} Ax & & & = b \\ & A^T p & +s & = c \\ c^T x & -b^T p & & = 0 \\ x \geq 0 & & s \geq 0. & \end{array} \tag{3}$$

So, if it were not for the nonnegativity constraints, (P) and (D) could have been solved via a simple system of linear equations. Thus, it is the nonnegativity constraints that make LP difficult! One way to view many interior point method approaches to linear programming is as ways to ignore (at least temporarily) nonnegativity constraints, or to prevent them from becoming active.

We introduce the following notation which will be very convenient for manipulating equations, etc. Suppose that $x > 0$. Define the matrix X to be the $n \times n$ diagonal matrix whose diagonal entries are precisely the components of x. Then X looks like:

$$\begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_n \end{pmatrix}.$$

We can also define X^2, which looks like:

$$\begin{pmatrix} x_1^2 & 0 & \dots & 0 \\ 0 & x_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_n^2 \end{pmatrix}.$$

Similarly, the matrices X^{-1} and X^{-2} look like:

$$\begin{pmatrix} 1/x_1 & 0 & \dots & 0 \\ 0 & 1/x_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/x_n \end{pmatrix}$$

and

$$\begin{pmatrix} 1/x_1^2 & 0 & \dots & 0 \\ 0 & 1/x_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/x_n^2 \end{pmatrix}.$$

Some tricks of using this notation: if e is the vector of ones, i.e., $e = (1, 1, 1, \dots, 1)^T$, $Xe = x$, and for any $s \in \mathbb{R}^n$, $Xs = XSe = (x_1 s_1, \dots, x_n s_n)^T$.

To "avoid" the nonnegativity constraints in (P), let us introduce a *logarithmic barrier* term for (P). The resulting (nonlinear) optimization problem is

$$\begin{array}{lll} \mathrm{P}(\mu) & \min & c^T x - \mu \sum_{j=1}^{n} \ln(x_j) \\ & \text{s.t.} & Ax = b \\ & & x \in X = \{x > 0\}. \end{array}$$

It is not hard to show that, for any $\mu > 0$, the objective function of P(μ) is convex, and so this problem fits into the framework of (1).

Notice that the feasible region of P(μ) is the relative interior of the feasible region of (P), i.e., it contains all feasible points with $x > 0$. The objective function of P(μ) can be seen as balancing two goals: on the one hand, minimizing the original objective function $c^T x$, and on the other hand keeping components of x as far from 0 as possible. The value of μ dictates how these two goals are "balanced" by the optimal solution of the problem. (For $\mu = +\infty$, the corresponding problem becomes $\min \left\{ -\sum_{j=1}^{n} \ln(x_j) : Ax = b,\ x > 0 \right\}$, and its optimal solution is referred to as the *analytic center* of (P).)

Because the gradient of the objective function of P(μ) is simply $c - \mu X^{-1} e$, the Karush-Kuhn-Tucker conditions for P(μ) are:

$$\begin{aligned} Ax = b,\ x > 0 \\ c - \mu X^{-1} e = A^T p. \end{aligned} \tag{4}$$

If we define $s = \mu X^{-1} e$, then

$$\frac{1}{\mu} Xs = e,$$

equivalently

$$\frac{1}{\mu}XSe = e,$$

and we can rewrite the Karush-Kuhn-Tucker conditions as:

$$\begin{aligned} Ax = b,\ x > 0 \\ A^T p + s = c \\ \frac{1}{\mu}XSe - e = 0. \end{aligned} \tag{5}$$

From the equations of (5) it follows that if $(x(\mu), p(\mu), s(\mu))$ is a solution of (5), then $x(\mu)$ is feasible for P, $(p(\mu), s(\mu))$ is feasible for D, and the resulting duality gap is:

$$x(\mu)^T s(\mu) = e^T X(\mu) S(\mu) e = \mu e^T e = \mu n.$$

The set of points

$$\{(x(\mu), p(\mu), s(\mu)) : \mu > 0\} \in \mathbb{R}^{n+m+n}$$

is referred to as the *(primal-dual) central path*, and is illustrated in Figure 1. The above duality gap result for points on the central path suggests that we try solving P(μ) for a variety of values of μ as $\mu \to 0$.

However, we cannot usually solve (5) exactly, because the third equation group is not linear in the variables. We will instead define a "β-approximate solution" of the Karush-Kuhn-Tucker conditions (5). A β-approximate solution of P(μ) is defined as any solution (x, p, s) of

$$\begin{aligned} Ax = b,\ x > 0 \\ A^T p + s = c \\ \left\| \frac{1}{\mu}Xs - e \right\| \leq \beta. \end{aligned} \tag{6}$$

Here the norm $\|\cdot\|$ is the Euclidean norm, or the 2-norm.

Lemma 3 *If $(\bar{x}, \bar{p}, \bar{s})$ is a β-approximate solution of P(μ) and $\beta < 1$, then $\bar{x}$ is feasible for P, $(\bar{p}, \bar{s})$ is feasible for D, and the duality gap satisfies:*

$$n\mu(1-\beta) \leq c^T\bar{x} - b^T\bar{p} = \bar{x}^T\bar{s} \leq n\mu(1+\beta). \tag{7}$$

Proof: Primal feasibility is obvious. To prove dual feasibility, we need to show that $\bar{s} \geq 0$. To see this, note that the third equation system of (6) implies that

$$-\beta \leq \frac{\bar{x}_j\bar{s}_j}{\mu} - 1 \leq \beta$$

which we can rearrange as:

$$\mu(1-\beta) \leq \bar{x}_j\bar{s}_j \leq \mu(1+\beta). \tag{8}$$

Therefore $\bar{x}_j\bar{s}_j \geq (1-\beta)\mu > 0$, which implies $\bar{x}_j\bar{s}_j > 0$, and so $\bar{s}_j > 0$.

From (8) we have

$$n\mu(1-\beta) = \sum_{j=1}^{n} \mu(1-\beta) \leq \sum_{j=1}^{n} \bar{x}_j\bar{s}_j = \bar{x}^T\bar{s} \leq \sum_{j=1}^{n} \mu(1+\beta) = n\mu(1+\beta).$$

■

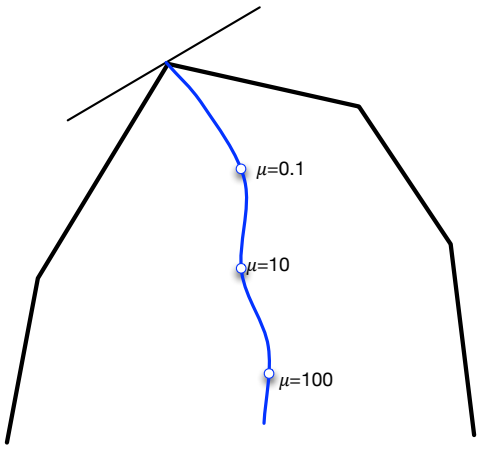


Figure 1: A conceptual picture of the central path.

3 The primal-dual Newton step for the barrier problem

Consider solving P(μ) by applying Newton's method to (5). Let $(\bar{x}, \bar{p}, \bar{s})$ be our current iterate, which we assume is primal and dual feasible, namely:

$$A\bar{x} = b,\ \bar{x} > 0, \quad A^T\bar{p} + \bar{s} = c, \bar{s} > 0. \tag{9}$$

Introducing a direction $(\Delta x, \Delta p, \Delta s)$, the next iterate will be $(\bar{x}, \bar{p}, \bar{s}) + (\Delta x, \Delta p, \Delta s)$, and we want to solve:

$$\begin{aligned} A(\bar{x} + \Delta x) &= b,\ \bar{x} + \Delta x > 0 \\ A^T(\bar{p} + \Delta p) + (\bar{s} + \Delta s) &= c \\ \frac{1}{\mu}(\bar{X} + \Delta X)(\bar{S} + \Delta S)e - e &= 0. \end{aligned}$$

Keeping in mind that $(\bar{x}, \bar{p}, \bar{s})$ is primal-dual feasible and so satisfies (9), we can rearrange the above to be:

$$\begin{aligned} A\Delta x = &\quad 0 \\ A^T\Delta p + \Delta s = &\quad 0 \\ \bar{S}\Delta x + \bar{X}\Delta s = &\quad \mu e - \bar{X}\bar{S}e - \Delta X \Delta S e. \end{aligned}$$

Notice that the only nonlinear term in the above system of equations in $(\Delta x, \Delta p, \Delta s)$ is the term "$\Delta X \Delta S e$" in the last line. If we erase this term, i.e., linearize the equations, we obtain the following primal-dual Newton equation system:

$$\begin{aligned} A\Delta x &= 0 \\ A^T\Delta p + \Delta s &= 0 \\ \bar{S}\Delta x + \bar{X}\Delta s &= \mu e - \bar{X}\bar{S}e. \end{aligned} \tag{10}$$

The solution $(\Delta x, \Delta p, \Delta s)$ of the system (10) is called the primal-dual Newton step. We can manipulate these equations to yield the following formulas for the solution:

$$\begin{aligned} \Delta x &\leftarrow \left[I - \bar{X}\bar{S}^{-1}A^T\left(A\bar{X}\bar{S}^{-1}A^T\right)^{-1}A\right]\left(-\bar{x} + \mu\bar{S}^{-1}e\right), \\ \Delta p &\leftarrow \left(A\bar{X}\bar{S}^{-1}A^T\right)^{-1}A\left(\bar{x} - \mu\bar{S}^{-1}e\right), \\ \Delta s &\leftarrow A^T\left(A\bar{X}\bar{S}^{-1}A^T\right)^{-1}A\left(-\bar{x} + \mu\bar{S}^{-1}e\right). \end{aligned} \tag{11}$$

Notice, by the way, that the computational effort in these equations lies primarily in solving a single equation:

$$\left(A\bar{X}\bar{S}^{-1}A^T\right)\Delta p = A\left(\bar{x} - \mu\bar{S}^{-1}e\right).$$

Once this system is solved, we can easily substitute:

$$\begin{array}{rcl} \Delta s & \leftarrow & -A^T\Delta p \\ \Delta x & \leftarrow & -\bar{x} + \mu\bar{S}^{-1}e - \bar{S}^{-1}\bar{X}\Delta s. \end{array} \tag{12}$$

In what follows, however, we will mostly work with the equations of the primal-dual Newton system (10) rather than formulae (11).

Suppose that $(\Delta x, \Delta p, \Delta s)$ is the (unique) solution of the primal-dual Newton system (10). We obtain the new value of the variables (x, p, s) by taking the Newton step:

$$(x', p', s') = (\bar{x}, \bar{p}, \bar{s}) + (\Delta x, \Delta p, \Delta s).$$

We have the following very powerful convergence theorem which demonstrates the *quadratic convergence* of Newton's method for this problem, with an explicit guarantee of the range in which quadratic convergence takes place.

Theorem 4 (Explicit Quadratic Convergence of Newton's Method) *Suppose that $(\bar{x}, \bar{p}, \bar{s})$ is a β-approximate solution of P(μ) and $\beta < \frac{1}{3}$. Let $(\Delta x, \Delta p, \Delta s)$ be the solution to the primal-dual Newton equations (10), and let:*

$$(x', p', s') = (\bar{x}, \bar{p}, \bar{s}) + (\Delta x, \Delta p, \Delta s).$$

Then (x', p', s') is a $\left(\frac{1+\beta}{(1-\beta)^2}\right)\beta^2$-approximate solution of P(μ).

Proof: Our current point $(\bar{x}, \bar{p}, \bar{s})$ satisfies:

$$\begin{array}{r} A\bar{x} = b, \bar{x} > 0 \\ A^T\bar{p} + \bar{s} = c \\ \left\| \frac{1}{\mu}\bar{X}\bar{S}e - e \right\| \leq \beta. \end{array} \tag{13}$$

Furthermore the primal-dual Newton step $(\Delta x, \Delta p, \Delta s)$ satisfies:

$$\begin{array}{r} A\Delta x = 0 \\ A^T\Delta p + \Delta s = 0 \\ \bar{S}\Delta x + \bar{X}\Delta s = \mu e - \bar{X}\bar{S}e. \end{array} \tag{14}$$

Note from the first two equations of (14) that $\Delta x^T\Delta s = 0$. From the third equation of (13) we have

$$1 - \beta \leq \frac{\bar{s}_j\bar{x}_j}{\mu} \leq 1 + \beta, \ j = 1, \ldots, n, \tag{15}$$

which implies:

$$\bar{x}_j \geq \frac{(1-\beta)\mu}{\bar{s}_j} \text{ and } \bar{s}_j \geq \frac{(1-\beta)\mu}{\bar{x}_j}, \ j = 1, \ldots, n. \tag{16}$$

As a result of this we obtain:

$$\begin{aligned}\mu(1-\beta)\|\bar{X}^{-1}\Delta x\|^2 &= \sum_{j=1}^{n} \frac{\mu(1-\beta)}{\bar{x}_j} \cdot \frac{\Delta x_j^2}{\bar{x}_j} \\ &\leq \sum_{j=1}^{n} \bar{s}_j \cdot \frac{\Delta x_j^2}{\bar{x}_j} = \Delta x^T \bar{X}^{-1} \bar{S} \Delta x \\ &= \Delta x^T \bar{X}^{-1} \left(\mu e - \bar{X}\bar{S}e - \bar{X}\Delta s\right) \\ &= \Delta x^T \bar{X}^{-1} \left(\mu e - \bar{X}\bar{S}e\right) \\ &\leq \|\bar{X}^{-1}\Delta x\| \cdot \|\mu e - \bar{X}\bar{S}e\| \\ &\leq \|\bar{X}^{-1}\Delta x\| \beta\mu.\end{aligned}$$

From this it follows that

$$\|\bar{X}^{-1}\Delta x\| \leq \frac{\beta}{1-\beta} < 1.$$

Therefore

$$x' = \bar{x} + \Delta x = \bar{X}(e + \bar{X}^{-1}\Delta x) > 0.$$

We replicate the exact same chain of inequalities for the dual variables to get:

$$\mu(1-\beta)\|\bar{S}^{-1}\Delta s\|^2 \leq \|\bar{S}^{-1}\Delta s\|\beta\mu.$$

From this it follows that

$$\|\bar{S}^{-1}\Delta s\| \leq \frac{\beta}{1-\beta} < 1.$$

Therefore

$$s' = \bar{s} + \Delta s = \bar{S}(e + \bar{S}^{-1}\Delta s) > 0.$$

Next note from (14) that for $j = 1, \ldots, n$ we have:

$$x'_j s'_j = (\bar{x}_j + \Delta x_j)(\bar{s}_j + \Delta s_j) = \bar{x}_j\bar{s}_j + \bar{x}_j\Delta s_j + \Delta x_j\bar{s}_j + \Delta x_j\Delta s_j = \mu + \Delta x_j\Delta s_j.$$

Therefore

$$\left(e - \frac{1}{\mu}X'S'e\right)_j = -\frac{\Delta x_j \Delta s_j}{\mu}.$$

From this we obtain:

$$\begin{aligned}\left\|e - \frac{1}{\mu}X'S'e\right\| &\leq \left\|e - \frac{1}{\mu}X'S'e\right\|_1 \\ &= \sum_{j=1}^{n} \frac{|\Delta x_j \Delta s_j|}{\mu} \\ &= \sum_{j=1}^{n} \frac{|\Delta x_j|}{\bar{x}_j} \frac{|\Delta s_j|}{\bar{s}_j} \frac{\bar{x}_j\bar{s}_j}{\mu} \\ &\leq \sum_{j=1}^{n} \frac{|\Delta x_j|}{\bar{x}_j} \frac{|\Delta s_j|}{\bar{s}_j}(1+\beta) \\ &\leq \|\bar{X}^{-1}\Delta x\|\|\bar{S}^{-1}\Delta s\|(1+\beta) \\ &\leq \left(\frac{\beta}{1-\beta}\right)^2 (1+\beta).\end{aligned}$$

■

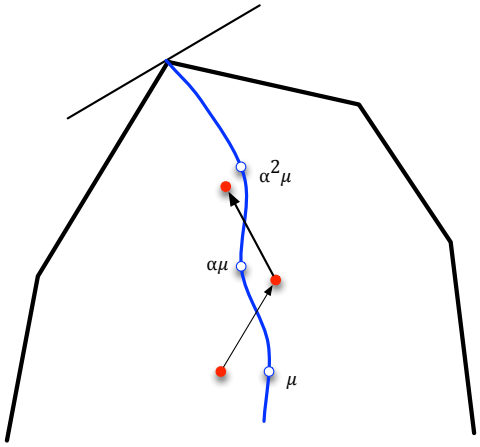


Figure 2: A conceptual picture of the interior-point algorithm.

4 The primal-dual interior point (barrier) algorithm

Based on the analysis just presented, we are motivated to develop the following algorithm:

Step 0: Initialization Data is (x^0, p^0, s^0, μ_0). $k = 0$. Assume that (x^0, p^0, s^0) is a β-approximate solution of $P(\mu_0)$ for some known value of β.

Step 1: Set current values $(\bar{x}, \bar{p}, \bar{s}) = (x^k, p^k, s^k)$, $\mu = \mu_k$.

Step 2: Shrink μ. Set $\mu' = \alpha\mu$ for some $\alpha \in (0, 1)$.

Step 3: Compute the primal-dual Newton direction. Compute the Newton step $(\Delta x, \Delta p, \Delta s)$ for the equation system (5) at $(x, p, s) = (\bar{x}, \bar{p}, \bar{s})$ for μ', by solving the following system of equations in the variables $(\Delta x, \Delta p, \Delta s)$:

$$\begin{aligned} A\Delta x &= 0 \\ A^T\Delta p + \Delta s &= 0 \\ \bar{S}\Delta x + \bar{X}\Delta s &= \bar{X}\bar{S}e - \mu' e. \end{aligned} \tag{17}$$

Denote the solution to this system by $(\Delta x, \Delta p, \Delta s)$.

Step 4: Update All Values.

$$(x', p', s') = (\bar{x}, \bar{p}, \bar{s}) + (\Delta x, \Delta p, \Delta s)$$

Step 5: Reset Counter and Continue. $(x^{k+1}, p^{k+1}, s^{k+1}) = (x', p', s')$. $\mu_{k+1} = \mu'$. $k \leftarrow k + 1$. Go to Step 1.

Figure 2 shows a picture of the algorithm.

Some of the issues regarding this algorithm include:

- how to set the approximation constant β and the fractional decrease parameter α (we will see that it will be convenient to set $\beta = \frac{3}{40}$ and $\alpha = 1 - \frac{\frac{1}{8}}{\frac{1}{5}+\sqrt{n}}$)

- whether or not successive iterative values (x^k, p^k, s^k) are β-approximate solutions to $P(\mu_k)$
- how long does the algorithm take to obtain a satisfactory solution of (P) and (D).

5 Analysis of the algorithm

Theorem 5 (Relaxation Theorem) *Suppose that $(\bar{x}, \bar{p}, \bar{s})$ is a $\frac{3}{40}$-approximate solution of P(μ). Let*

$$\alpha = 1 - \frac{\frac{1}{8}}{\frac{1}{5} + \sqrt{n}}$$

and let $\mu' = \alpha\mu$. Then $(\bar{x}, \bar{p}, \bar{s})$ is a $\frac{1}{5}$-approximate solution of $P(\mu')$.

Proof: The triplet $(\bar{x}, \bar{p}, \bar{s})$ satisfies $A\bar{x} = b, \bar{x} > 0$, and $A^T\bar{p} + \bar{s} = c$, and so it remains to show that

$$\left\| \frac{1}{\mu'} \bar{X}\bar{s} - e \right\| \leq \frac{1}{5}.$$

We have

$$\begin{aligned} \left\| \frac{1}{\mu'} \bar{X}\bar{s} - e \right\| = \left\| \frac{1}{\alpha\mu} \bar{X}\bar{s} - e \right\| &= \left\| \frac{1}{\alpha} \left(\frac{1}{\mu} \bar{X}\bar{s} - e \right) - \left(1 - \frac{1}{\alpha} \right) e \right\| \\ &\leq \left(\frac{1}{\alpha} \right) \left\| \frac{1}{\mu} \bar{X}\bar{s} - e \right\| + \left| \frac{1-\alpha}{\alpha} \right| \|e\| \\ \leq \frac{\frac{3}{40}}{\alpha} + \left(\frac{1-\alpha}{\alpha} \right) \sqrt{n} \quad &= \quad \frac{\frac{3}{40} + \sqrt{n}}{\alpha} - \sqrt{n} = \frac{1}{5}. \end{aligned}$$

■

Theorem 6 (Convergence Theorem). *Suppose that (x^0, p^0, s^0) is a $\beta = \frac{3}{40}$-approximate solution of $P(\mu_0)$. Then for all $k = 1, 2, 3, \ldots$, (x^k, p^k, s^k) is a $\beta = \frac{3}{40}$-approximate solution of $P(\mu_k)$.*

Proof: By induction, suppose that the theorem is true for iterates $0, 1, 2, ..., k$.

Then (x^k, p^k, s^k) is a $\beta = \frac{3}{40}$-approximate solution of $P(\mu_k)$. From the Relaxation Theorem, (x^k, p^k, s^k) is a $\frac{1}{5}$-approximate solution of $P(\mu_{k+1})$ where $\mu_{k+1} = \alpha\mu_k$.

From the Quadratic Convergence Theorem, $(x^{k+1}, p^{k+1}, s^{k+1})$ is a β-approximate solution of $P(\mu_{k+1})$ for

$$\beta = \frac{1 + \frac{1}{5}}{\left(1 - \frac{1}{5}\right)^2} \left(\frac{1}{5} \right)^2 = \frac{3}{40}.$$

Therefore, by induction, the theorem is true for all values of k. ■

Theorem 7 (Complexity Theorem) *Suppose that (x^0, p^0, s^0) is a $\beta = \frac{3}{40}$-approximate solution of $P(\mu_0)$. In order to obtain primal and dual feasible solutions (x^k, p^k, s^k) with a duality gap of at most ϵ, one needs to run the algorithm for at most*

$$k = \left\lceil 10\sqrt{n} \ln \left(\frac{43}{37} \frac{(x^0)^T s^0}{\epsilon} \right) \right\rceil$$

iterations.

Proof: Let k be as defined above. Note that

$$\alpha = 1 - \frac{\frac{1}{8}}{\frac{1}{5} + \sqrt{n}} = 1 - \frac{1}{\left(\frac{8}{5} + 8\sqrt{n}\right)} \leq 1 - \frac{1}{10\sqrt{n}}.$$

Therefore

$$\mu_k \leq \left(1 - \frac{1}{10\sqrt{n}}\right)^k \mu_0.$$

This implies that

$$c^T x^k - b^T p^k = (x^k)^T s^k \leq \mu_k n(1+\beta) \leq \left(1 - \frac{1}{10\sqrt{n}}\right)^k \left(\frac{43}{40}n\right) \mu_0$$

$$\leq \left(1 - \frac{1}{10\sqrt{n}}\right)^k \left(\frac{43}{40}n\right) \left(\frac{(x^0)^T s^0}{\frac{37}{40}n}\right),$$

from (7). Taking logarithms, we obtain

$$\ln(c^T x^k - b^T p^k) \leq k \ln\left(1 - \frac{1}{10\sqrt{n}}\right) + \ln\left(\frac{43}{37}(x^0)^T s^0\right)$$

$$\leq \frac{-k}{10\sqrt{n}} + \ln\left(\frac{43}{37}(x^0)^T s^0\right)$$

$$\leq -\ln\left(\frac{43}{37}\frac{(x^0)^T s^0}{\epsilon}\right) + \ln\left(\frac{43}{37}(x^0)^T s^0\right) = \ln(\epsilon).$$

The second inequality uses the fact that $\ln(1-t) \leq -t$ for all $t < 1$. Therefore $c^T x^k - b^T p^k \leq \epsilon$. ∎

6 A broader view of path-following IPMs

In this section we generalize the algorithm we considered before to provide a more general "path-following framework" for interior point methods.

Let us review some of the basic ideas of the primal-dual feasible interior point method we've developed so far. We are considering a primal-dual pair of problems

$$\begin{array}{llll} \text{(P)} \ \min & c^T x & \text{(D)} \ \max & b^T p \\ & Ax = b & & A^T p + s = c \\ & x \geq 0 & & s \geq 0. \end{array}$$

Let us assume that both problems have interior feasible solutions: $(x > 0, p, s > 0)$. (Then we can assume without loss of generality that A has full row rank, and $c \notin \text{Range}(A^T)$.)

We defined the primal-dual central path for this pair to be the trajectory of points $(x(\mu), p(\mu), s(\mu))$ such that

$$Ax(\mu) = b, \ A^T p(\mu) + s(\mu) = c, \ X(\mu)S(\mu)e = \mu e, \ (x(\mu), s(\mu)) > 0, \tag{18}$$

for all $\mu > 0$, which can be viewed both as the relaxed optimality conditions of the original primal and dual LPs, and as the KKT conditions for the primal and dual barrier problems with μ being

the "weight" of the barrier function. Note that the parameter μ above can be viewed as a measure of the duality gap between the solutions on the central path, since $x(\mu)^T s(\mu) = n\mu$. We haven't studied the issue of existence of solutions to (18), but it can be shown that a unique solution does indeed exist for any $\mu > 0$ under our assumption of existence of interior feasible solutions.[3]

The general "philosophy" of the path-following IPMs can be described as follows: We would like to generate a sequence of primal-dual (strictly feasible) iterates (x^k, p^k, s^k) such that:

- The duality gap $(x^k)^T s^k$ decreases with k (ideally approaching 0 in the limit), and
- Each of the iterates is "close" to some point $(x(\mu_k), p(\mu_k), s(\mu_k))$ on the central path — we call this property "centrality".

The main reason we would like to maintain centrality is that centrally-located points are guaranteed to (strictly) satisfy nonnegativity constraints of the LP. Moreover, if an iterate is located close to the point on the central path corresponding to a known value of μ_k, we can provide an estimate of the duality gap at that iterate as well.

To facilitate the discussion below, let us give an alternative interpretation of centrality. Note that at points on the central path (i.e., perfectly central points), the products $x_i(\mu)s_i(\mu)$ are the same for all $i = 1, \ldots, n$ — namely, they are equal to μ. We can think of a strictly feasible point as having good centrality if the products $x_i s_i$ are approximately the same for all i.

Let the current (strictly feasible) iterate of our algorithm be $(\bar{x}, \bar{p}, \bar{s}) = (x^k, p^k, s^k)$, and we would like to generate the next iterate to be close to the central path, for the above reasons. First, we need to decide which point on the central path we want to approach, i.e., what should be the value of μ in (18). There are several options, motivated by our goals:

- Set $\mu = 0$. In this case, we are trying to decrease the duality gap between the primal and the dual solutions.
- Set $\mu = \bar{x}^T \bar{s}/n$. In this case, we are not trying to change the duality gap, but rather trying to improve centrality by having all individual products $x_i s_i$ revert to the average, while maintaing the same value of the duality gap.
- Set $\mu = \sigma \cdot (\bar{x}^T \bar{s}/n)$ with $\sigma \in (0, 1)$ — this is a compromise between the two options above.

Once we have settled on the value of μ, we have identified the nonlinear system of equation of the form (18) we would like our next iterate to (approximately) solve:

$$F(x, p, s) \stackrel{\triangle}{=} \begin{bmatrix} Ax - b \\ A^T p + s - c \\ XSe \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mu e \end{bmatrix}.$$

We will use ideas from the Newton's method to solve that system, starting at the current point $(\bar{x}, \bar{p}, \bar{s})$. To find the Newton direction for the above system of equations at the point $(\bar{x}, \bar{p}, \bar{s})$, we set up the following approximation:

$$F((\bar{x}, \bar{p}, \bar{s}) + (\Delta x, \Delta p, \Delta s)) \approx F(\bar{x}, \bar{p}, \bar{s}) + J(\bar{x}, \bar{p}, \bar{s})^T \begin{bmatrix} \Delta x \\ \Delta p \\ \Delta s \end{bmatrix},$$

[3]Proven, for instance, in "Linear Programming: Foundations and Extensions" by Robert J. Vanderbei, Kluwer. An electronic version of the latest edition of this book is available through the UofM library website.

where $J(\cdot)^T$ is the Jacobian of the function $F(\cdot)$ — in our setting,

$$J(\bar{x},\bar{p},\bar{s})^T = \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \bar{S} & 0 & \bar{X} \end{bmatrix}.$$

Motivated by the above approximation, we obtain the direction $(\Delta x, \Delta p, \Delta s)$ by solving the linear system

$$F(\bar{x},\bar{p},\bar{s}) + J(\bar{x},\bar{p},\bar{s})^T \begin{bmatrix} \Delta x \\ \Delta p \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mu e \end{bmatrix}, \text{ or}$$

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \bar{S} & 0 & \bar{X} \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta p \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\bar{X}\bar{S}e + \mu e \end{bmatrix}. \tag{19}$$

It can be shown that as long as A has full row rank and $(\bar{x},\bar{s}) > 0$, the matrix in the system (19) is nonsingular. Finally, our next iterate is obtained by setting

$$(x^{k+1}, p^{k+1}, s^{k+1}) = (\bar{x},\bar{p},\bar{s}) + \lambda(\Delta x, \Delta p, \Delta s), \tag{20}$$

where $\lambda > 0$ is an appropriate stepsize. In particular, λ should be sufficiently small, to ensure that $(\bar{x} + \lambda\Delta x, \bar{s} + \lambda\Delta s) > 0$.

What are "desirable" values of σ and λ? The following result provides us with some intuition:

Lemma 8 *Let* $(x(\lambda), p(\lambda), s(\lambda)) = (\bar{x},\bar{p},\bar{s}) + \lambda(\Delta x, \Delta p, \Delta s)$. *Then*

$$\Delta x^T \Delta s = 0, \text{ and } x(\lambda)^T s(\lambda) = (1 - \lambda(1-\sigma))\bar{x}^T\bar{s}.$$

Proof: $\Delta x^T \Delta s = 0$ follows from the first two equations of (19). Using this fact and the third equation,

$$\begin{aligned} x(\lambda)^T s(\lambda) &= \bar{x}^T\bar{s} + \lambda(\bar{x}^T\Delta s + \bar{s}^T\Delta x) \\ &= \bar{x}^T\bar{s} + \lambda e^T(\bar{X}\Delta s + \bar{S}\Delta x) \\ &= \bar{x}^T\bar{s} + \lambda e^T(-\bar{X}\bar{S}e + \mu e) \\ &= (1 - \lambda(1-\sigma))\bar{x}^T\bar{s}, \end{aligned}$$

since $\mu = \sigma\bar{\mu} = \sigma \cdot \bar{x}^T\bar{s}/n$. ∎

Considering that the reduction in the duality gap is given by the above lemma, we want to choose smaller values of σ on the one hand, and be able to use large values of λ and still maintain feasibility on the other.

General framework for a path-following primal-dual feasible IPM

Given (x^0, p^0, s^0) – strictly feasible

for $k = 0, 1, 2, \ldots$

solve

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S^k & 0 & X^k \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta p \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -X^kS^ke + \sigma_k\mu_k e \end{bmatrix},$$

where $\sigma_k \in [0,1]$ and $\mu_k = (x^k)^T s^k/n$.

set $(x^{k+1}, p^{k+1}, s^{k+1}) = (x^k, p^k, s^k) + \lambda_k(\Delta x, \Delta p, \Delta s)$ where $\lambda_k > 0$ is chosen, in part, so that $(x^{k+1}, s^{k+1}) > 0$

end

Depending on the starting conditions (i.e., properties of (x^0, p^0, s^0)) and rules for choosing σ_k and λ_k, we specify this framework to include a variety of algorithms.

6.1 Short-step algorithm

Let $\beta = \frac{3}{40}$, $\sigma = 1 - \frac{\frac{1}{8}}{\frac{1}{5}+\sqrt{n}}$. In the algorithm, choose $\sigma_k = \sigma$, $\lambda_k = 1$ for all k. In this algorithm,

$$(x^k, p^k, s^k) \in \mathcal{N}_2(\beta) \stackrel{\triangle}{=} \{(x,p,s) \text{ feasible} : \|XSe - e\mu\|_2 \leq \beta\mu, \text{ where } \mu = x^T s/n\}$$

for all $k > 0$, provided that the condition is satisfied for $k = 0$.

This is exactly the algorithm we analyzed in the previous section. It is referred to as a *short-step* algorithm, since in every iteration it makes a short, safe move in a narrow neighborhood of the central path. Recall that we showed that, if we initialize the algorithm with a feasible point $(x^0, p^0, s^0) \in \mathcal{N}_2(\beta)$ for a sufficiently small β, the algorithm will produce a feasible solution with duality gap of at most ϵ in at most

$$O(\sqrt{n}\ln((x^0)^T s^0/\epsilon))$$

iterations. This is the best known theoretical complexity result for an interior point algorithm. However, in practice this algorithm is considered too conservative — precisely because it is, by design, a short-step algorithm.

6.2 Predictor-corrector algorithm

This is the so-called Mizuno-Todd-Ye predictor-corrector algorithm. Another well known algorithm with a similar name is the so-called Mehrotra predictor-corrector algorithm. (The latter is actually the algorithm which most closely resembles IPMs implemented in most commercial linear programming solvers. It uses a lot of the ideas we are developing here, plus adds a few clever twists, to come up with a very practical IPM.)

Start with $(x^0, p^0, s^0) \in \mathcal{N}_2(0.25)$. This predictor-corrector algorithm uses different rules to select values of σ_k and λ_k at alternating iterations:

- For k even, set $\sigma_k = 0$. Select the largest λ_k such that $(x^k, p^k, s^k) + \lambda_k(\Delta x, \Delta p, \Delta s) \in \mathcal{N}_2(0.5)$. This is a *predictor*, or affine scaling, step. Its goal is to take a large step and significantly improve the duality gap.
- For k odd, set $\sigma_k = 1$ and $\lambda_k = 1$. This is a *corrector* step, the goal of which is to return to $\mathcal{N}_2(0.25)$ without changing μ.

To analyze the performance of this algorithm, we need to consider what happens after one predictor and one corrector step. Using an argument similar to that of Theorem 4, we can show that if the

current iterate is in $\mathcal{N}(0.5)$, after applying one corrector step, the next iterate will be in $\mathcal{N}_2(0.25)$, so the corrector step indeed restores better centrality. Also, since $\sigma_k = 1$, Lemma 8 indicates that the duality gap does not change.

It only remains to see how much progress a predictor step makes towards reducing the duality gap, i.e., we need to provide a lower bound on the stepsize α_k (recall Lemma 8). This can be done via the following two lemmas:

Lemma 9 *If $(x^k, p^k, s^k) \in \mathcal{N}_2(\beta)$ for some $\beta \in (0,1)$, then*

$$\|\Delta X \Delta S e\| \leq \frac{\beta^2 + n(1-\sigma)^2}{2^{3/2}(1-\beta)}\mu_k.$$

Lemma 10 *If $(x^k, p^k, s^k) \in \mathcal{N}_2(0.25)$ and $(\Delta x, \Delta p, \Delta s)$ is calculated using $\sigma_k = 0$, then for all $\lambda \in [0, \bar{\lambda}]$, where*

$$\bar{\lambda} = \min\left(\frac{1}{2}, \left(\frac{\mu_k}{8\|\Delta X \Delta S e\|}\right)^{1/2}\right),$$

we have $(x(\lambda), p(\lambda), s(\lambda)) \in \mathcal{N}_2(0.5)$.

Since prior to the prediction step conditions of Lemma 9 are satisfied for $\beta = 0.25$, we can use Lemma 10 to show that the stepsize λ_k in the predictor step will be at least $\frac{0.4}{\sqrt{n}}$, and thus after the predictor step,

$$\mu_{k+1} \leq \left(1 - \frac{0.4}{\sqrt{n}}\right)\mu_k.$$

Thus, the duality gap is decreased by a factor of at least $\left(1 - \frac{0.4}{\sqrt{n}}\right)$ every two iterations, which means that if we initialize the algorithm with a feasible point $(x^0, p^0, s^0) \in \mathcal{N}_2(0.25)$, the algorithm will produce a feasible solution with duality gap of at most ϵ in at most

$$O(\sqrt{n}\ln((x^0)^T s^0/\epsilon))$$

iterations. Note that, in theory, this is the same worst-case complexity as in the short step algorithm (although the constants "hidden" by the big-O notation are different). However, in practice this algorithm has more freedom to select the stepsize, and thus to make more progress in the predictor step if a good predictor direction is found (and thus the actual λ_k is often larger than the worst-case bound given by Lemma 10).

6.3 Long-step algorithm

This algorithm is designed to allow us to make more rapid progress in reducing the duality gap by relaxing the centrality demands on our iterates. Instead of requiring that all our iterates lie in a neighborhood $\mathcal{N}_2(\beta)$ of the central path, for some $\beta \in (0,1)$, this algorithm allows the iterates to be in a larger neighborhood, namely,

$$\mathcal{N}_{-\infty}(\gamma) \stackrel{\triangle}{=} \{(x,p,s) \text{ feasible} : x_i s_i \geq \gamma\mu \ \forall i, \text{ where } \mu = x^T s/n\},$$

for some $\gamma \in (0,1)$. Note that this is a "one-sided" neighborhood, i.e., whereas the neighborhood $\mathcal{N}_2(\beta)$ requires that the products $x_i s_i$ do not deviate much from the average $x^T s/n$ in either direction, the neighborhood $\mathcal{N}_{-\infty}(\gamma)$ only requires that none of these products becomes too small

compared to the average. (The larger γ is, the "tighter" the neighborhood.) Note, incidentally, that given a strictly feasible solution (x, p, s) we can, for instance, take

$$\gamma = \min\left(0.5,\ n \min_i \frac{x_i s_i}{x^T s}\right),$$

and have $(x, p, s) \in \mathcal{N}_{-\infty}(\gamma)$.

The long step algorithm at initialization is given some values $0 < \sigma_{\min} < \sigma_{\max} < 1$ and $\gamma \in (0, 1)$, and its initial iterate should satisfy

$$(x^0, p^0, s^0) \in \mathcal{N}_{-\infty}(\gamma).$$

At the kth iteration of the algorithm, we choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ and pick the largest λ_k such that

$$(x^k, p^k, s^k) + \lambda_k(\Delta x, \Delta p, \Delta s) \in \mathcal{N}_{-\infty}(\gamma).$$

Since this is a wider neighborhood of the central path than $\mathcal{N}_2(\beta)$,[4] we are allowed to take longer steps, leading to more significant improvements in μ, while some proximity to the central path is still maintained.

Note that in this algorithm the iterates stay in the neighborhood $\mathcal{N}_{-\infty}(\gamma)$ by design, and so to analyze the progress the algorithm makes, we just need to provide a lower bound on the stepsize λ_k at a typical iteration.

Lemma 11 *If $(x^k, p^k, s^k) \in \mathcal{N}_\infty(\gamma)$ for some $\gamma \in (0, 1)$, then*

$$\|\Delta X \Delta S e\| \leq 2^{-3/2}(1 + 1/\gamma) n \mu_k.$$

Theorem 12 *Given the parameters γ, $\sigma_{\min}$ and $\sigma_{\max}$ in the long-step algorithm, if $(x^k, p^k, s^k) \in \mathcal{N}_{-\infty}(\gamma)$, there exists a constant δ, independent of n (but tied to the values of the above parameters) such that*

$$\mu_{k+1} \leq \left(1 - \frac{\delta}{n}\right) \mu_k.$$

Thus, the duality gap is decreased by a factor of at least $\left(1 - \frac{\delta}{n}\right)$ at every iteration, which means that if we initialize the algorithm with a feasible point $(x^0, p^0, s^0) \in \mathcal{N}_\infty(\gamma)$, the algorithm will produce a feasible solution with duality gap of at most ϵ in at most

$$O(n \ln((x^0)^T s^0/\epsilon))$$

iterations. The constant "hidden" by the big-O notation depends on the parameters used in the implementation of the algorithm.

Note that, in theory, this algorithm has a worse complexity than the previous two algorithms with considered. However, in practice it often performs better, again, because it can capitalize on good directions by being flexible with the stepsize selection.

[4] This is an imprecise statement. Of course, the sizes of these neighborhood are determined by the actual values of β and γ.

7 Finite termination of IPMs

In the homework assignment, you have shown (or you will show) that the sequence (x^k, s^k) generated by many IPMs, and in particular, path-following IPMs we considered herein, has limit points, and its limit points correspond to strictly complementary primal and dual optimal solutions of (P) and (D).

We define:

$$\mathcal{B} = \{i : x_i^\star > 0 \text{ for some optimal } x^\star\}$$

and

$$\mathcal{N} = \{i : s_i^\star > 0 \text{ for some optimal } (p^\star, s^\star)\}.$$

Therefore, the sets of optimal solutions for the primal and the dual can be characterized, respectively, as

$$P^\star = \{x : Ax = b,\ x \geq 0,\ x_i = 0 \text{ for } i \in \mathcal{N}\}$$

and

$$D^\star = \{(p, s) : A^T p + s = c,\ s \geq 0,\ s_i = 0 \text{ for } i \in \mathcal{B}\}.$$

From the homework, we know that in every strictly complementary primal-dual pair of solutions (x, p, s), $x_i > 0$ iff $i \in \mathcal{B}$ and $s_i > 0$ iff $i \in \mathcal{N}$.

Furthermore, from the same homework, we know that there exists a constant $\bar{C} > 0$ such that, for a sequence of iterates (x^k, s^k) generated by any (feasible) interior point algorithm that keeps its iterates in a neighborhood $\mathcal{N}_{-\infty}(\gamma)$ of the central path for some $\gamma > 0$, we have

$$0 < x_i^k \leq \frac{\mu^k}{\bar{C}}\ \forall i \in \mathcal{N},\ 0 < s_i^k \leq \frac{\mu^k}{\bar{C}}\ \forall i \in \mathcal{B}, \tag{21}$$

$$s_i^k \geq \bar{C}\gamma\ \forall i \in \mathcal{N},\ x_i^k \geq \bar{C}\gamma\ \forall i \in \mathcal{B}, \tag{22}$$

where $\mu^k = (x^k)^T s^k / n$. Since in any of the algorithms we discussed μ^k tends to zero, for k sufficiently large, the components $x_i^k,\ i \in \mathcal{N}$ and $s_i^k,\ i \in \mathcal{B}$ will become small, while $x_i^k,\ i \in \mathcal{B}$ and $s_i^k,\ i \in \mathcal{N}$ will be converging to positive values. Thus, in the algorithm we can keep track of the relative sizes of x_i^k and s_i^k, and once the contents of $\mathcal{B}$ and $\mathcal{N}$ become clear to us, we can:

- Set $x_\mathcal{N}$ and $s_\mathcal{B}$ to zero;
- Find $x_\mathcal{B}$ and p to satisfy $A_\mathcal{B} x_\mathcal{B} = b$ and $A_\mathcal{B}^T p = c_\mathcal{B}$;
- Confirm that $x_\mathcal{B} > 0$ and $s_\mathcal{N} = c_\mathcal{N} - A_\mathcal{N}^T p > 0$.

Once we successfully perform the above steps, we can terminate our algorithm, since an optimal solution has been found.

Here is one possible approach to executing this finite termination strategy: given $(x, p, s) \in \mathcal{N}_{-\infty}(\gamma)$, let

$$\mathcal{B}(x, s) = \{i : x_i \geq s_i\} \text{ and } \mathcal{N}(x, s) = \{i : s_i > x_i\} = \{1, \ldots, n\} \setminus \mathcal{B}(x, s). \tag{23}$$

Note that strictly feasible points (x, p, s) that are close to a given strictly complementary solution satisfy $\mathcal{B}(x, s) = \mathcal{B}$ and $\mathcal{N}(x, s) = \mathcal{N}$, and so the "guess" in (23) is correct.

Consider $(\bar{x}, \bar{p}, \bar{s}) \in \mathcal{N}_{-\infty}(\gamma)$ for some $\gamma \in (0, 1)$ (this, for instance, could be an iterate of an IPM). We can attempt to construct an optimal solution from this point as follows:

Finite termination (attempt) procedure

- Find $\mathcal{B}(\bar{x}, \bar{s})$ and $\mathcal{N}(\bar{x}, \bar{s})$ from (23)
- Solve the following problem:

$$\begin{array}{ll} \min_{x,p,s} & \frac{1}{2}\|x - \bar{x}\|^2 + \frac{1}{2}\|s - \bar{s}\|^2 \\ \text{s.t.} & Ax = b \\ & A^T p + s = c \\ & x_i = 0 \ \forall i \in \mathcal{N}(\bar{x}, \bar{s}) \\ & s_i = 0 \ \forall i \in \mathcal{B}(\bar{x}, \bar{s}). \end{array} \tag{24}$$

- If for the solutions found $x_{\mathcal{B}(\bar{x},\bar{s})} > 0$ and $s_{\mathcal{N}(\bar{x},\bar{s})} > 0$, declare success: (x, p, s) is a strictly complementary solution. Otherwise, declare failure.

(24) is a convex QP (i.e., optimization problem with a quadratic objective function and linear constraints) with equality constraints only, and thus can be solved by solving its KKT conditions, which are a system of linear equations. Solving these linear equations requires approximately the same amount of effort as solving the Newton system in each iteration of an IPM. However, it should be noted that the constraints of this optimization problem, once the zero values of the chosen variables are substituted, may not have full rank. I.e., they might be infeasible or overdefined. Thus, careful use of Gaussian elimination must be employed to detects linear dependencies.

The following theorem establishes sufficient conditions under which the finite termination procedure will eventually succeed.

Theorem 13 *Let $\gamma \in (0, 1)$. Then there is a threshold value $\bar{\mu}$ such that for all $(x, p, s) \in \mathcal{N}_{-\infty}(\gamma)$ such that $0 < \mu = \frac{x^T s}{n} \leq \bar{\mu}$, we have*

(i) $\mathcal{B}(x, s) = \mathcal{B}$ and $\mathcal{N}(x, s) = \mathcal{N}$, i.e., actual index sets $\mathcal{B}$ and $\mathcal{N}$ are correctly identified by (23)*.*

(ii) The optimization problem (24) *yields a strictly complementary solution (x, p, s).*

The tools for proving part *(i)* of this theorem are, once again, provided in your homework: this part follows from equations (21) and (22). Part *(ii)* follows from the analysis of the KKT conditions for the system (24).

In an IPM, once the value of $\mu_k = (x^k)^T s^k / n$ falls below the threshold $\bar{\mu}$ in Theorem 13, the algorithm will be terminated. Thus,

- Short-step and predictor-corrector IPMs can be finitely terminated in at most

  $$O(\sqrt{n} \ln((x^0)^T s^0 / \bar{\mu}))$$

  iterations, and
- Long-step IPM can be finitely terminated in at most

  $$O(n \ln((x^0)^T s^0 / \bar{\mu}))$$

  iterations.

It is natural to conclude that the threshold value $\bar{\mu}$ is linked to the value of $\bar{C}$ in (21) and (22). In the homework, you establish an estimate of $\bar{C}$. Namely, we define the value $\epsilon(A, b, c)$ to be the

value of the "smallest large variable:"

$$\epsilon(A,b,c) = \min\left(\min_{i\in\mathcal{B}} \sup_{x^\star \text{ optimal}} x_i^\star,\ \min_{i\in\mathcal{N}} \sup_{s^\star \text{ optimal}} s_i^\star\right). \tag{25}$$

The definition of $\epsilon(A,b,c)$ is somewhat contrived, but the crux of it is as follows: pick an arbitrary $i \in \{1,\ldots,n\}$. If $i \in \mathcal{B}$, there exists *some* primal optimal solution $x^\star$ with $x_i^\star \geq \epsilon(A,b,c)$. If $i \in \mathcal{N}$, there exists *some* dual optimal solution $(p^\star, s^\star)$ with $s_i^\star \geq \epsilon(A,b,c)$. In fact, $\epsilon(A,b,c)$ is the largest positive number with this property. One of your homework problems will lead you to show that $\epsilon(A,b,c) > 0$, and another one — that $\bar{C}$ in (21) and (22) can be taken to be equal to $\epsilon(A,b,c)/n$.

A value of $\bar{\mu}$ that makes Theorem 13 work is proportional to $\epsilon(A,b,c)^2$, and also depends on n and a certain "measure of conditioning" of the matrix A. For the case when all the data of the LP are integers, and the size of the LP instance is bounded by L, all of those quantities (and thus $\bar{\mu}$) can be bounded below by 2^{-L}, and thus the complexity of our algorithms can be expressed as

$$O(\sqrt{n}L\ln((x^0)^T s^0)) \text{ or } O(nL\ln((x^0)^T s^0)).$$

8 Initialization of IPMs

Recall that up until this point we assumed that we have available a strictly feasible point (x^0, p^0, s^0) that is close to the central path, and thus can be used to successfully initialize our algorithms. How can such a point be obtained? There are many possible approaches, including:

- A big-M type method, which involves adding artificial columns and/or rows to the LPs, and a large ($M = O(2^L)$) penalty parameter to force solutions to become feasible,
- A two-phase approach, in which the Phase I is dedicated to finding a feasible point — a task just as hard as solving the original LP,
- etc.

Another approach, which allowed for easy initialization of a slightly re-designed algorithm, was presented by Ye, Todd and Mizuno in their 1994 paper in the journal *Mathematics of Operations Research*. In addition to simplifying initialization, their algorithm also removed the assumption that the LP in fact has a solution, since their approach is also capable of detecting primal or dual infeasibility.

This method, which is summarized below, utilizes a re-formulation of the primal and dual LPs into one LP problem, which is closely related to the homogeneous self-dual reformulation we considered in an earlier homework.

The problem, which the authors term (HLP) for *homogeneous LP* is as follows: suppose we are given (or we selected) some $(x^0 > 0, p^0, s^0 > 0)$, which is not necessarily primal-dual feasible. Define:

$$\bar{b} = b - Ax^0,\ \bar{c} = c - A^T p^0 - s^0,\ \bar{z} = c^T x^0 + 1 - b^T p^0.$$

Our problem is then

$$\begin{array}{llccccccccc}
(\text{HLP}) & \min & & & & & & ((x^0)^T s^0 + 1)\theta & & & \\
& \text{s.t.} & & Ax & & -b\tau & & +\bar{b}\theta & & = & 0 \\
& & -A^T p & & -s & +c\tau & & -\bar{c}\theta & & = & 0 \\
& & b^T p & -c^T x & & & & \bar{z}\theta & -\kappa & = & 0 \\
& & -\bar{b}^T p & +\bar{c}^T x & & -\bar{z}\tau & & & & = & -(x^0)^T s^0 - 1 \\
& & p \text{ free}, & x \geq 0, & s \geq 0 & \tau \geq 0, & & \theta \text{ free}, & \kappa \geq 0 & &
\end{array}$$

Notice that a solution to (HLP) with $\tau = 1$ and $\theta = 0$ satisfies primal and dual feasibility and a "reverse" weak duality, i.e., is a primal and dual optimal solution to the original LPs.

(HLP) has the following properties:

1. It is a self-dual system;
2. It has a strictly feasible point:
$$(x, p, s) = (x^0 > 0, p^0, s^0 > 0),\ \tau = 1,\ \theta = 1,\ \kappa = 1;$$
3. (HLP) has an optimal solution and its optimal solution set is bounded;
4. The optimal value of (HLP) is zero, and every feasible point satisfies
$$((x^0)^T s^0 + 1)\theta = x^T s + \tau\kappa;$$
5. (HLP) has a strictly self-complementary optimal solution with $\theta^\star = 0$ and
$$x^\star + s^\star > 0 \text{ and } \tau^\star + \kappa^\star > 0.$$

It will be convenient for us to take $p^0 = 0$ and $x^0 = s^0 = e$. Then (HLP) becomes

$$\begin{array}{llccccccccc}
(\text{HLP}) & \min & & & & & & ((n+1)\theta & & & \\
& \text{s.t.} & & Ax & & -b\tau & & +\bar{b}\theta & & = & 0 \\
& & -A^T p & & -s & +c\tau & & -\bar{c}\theta & & = & 0 \\
& & b^T p & -c^T x & & & & \bar{z}\theta & -\kappa & = & 0 \\
& & -\bar{b}^T p & +\bar{c}^T x & & -\bar{z}\tau & & & & = & -(n+1) \\
& & p \text{ free}, & x \geq 0, & s \geq 0 & \tau \geq 0, & & \theta \text{ free}, & \kappa \geq 0 & &
\end{array}$$

We can prove the following theorem, using techniques similar to what you used in the homework:

Theorem 14 *Let $(p^\star, x^\star, s^\star, \tau^\star, \theta^\star = 0, \kappa^\star)$ be a strictly self-complementary solution to (HLP). Then*

1. *(P) has an optimal solution iff $\tau^\star > 0$, in which case $x^\star/\tau^\star$ and $(p^\star/\tau^\star, s^\star/\tau^\star)$ are primal and dual optimal solutions.*
2. *If $\tau^\star = 0$, then $\kappa^\star > 0$, which implies that at least one of $c^T x^\star$ and $-b^T y^\star$ is negative. If the former is true, then (D) is infeasible, and if the latter is true, (P) is infeasible.*

Below is an interior point method designed for solving (HLP). Notice that since (HLP) is self-dual, the definitions of the central path, its neighborhood, and the Newton system do not require "additional" dual information.

8.1 Central path and its neighborhood for (HLP)

Let $\mathcal{F}$ denote the set of feasible solutions of (HLP), and let $\mathcal{F}^0$ denote the set of strictly feasible solutions (i.e., ones with $(x, s, \tau, \kappa) > 0$).

For μ, there is a unique point $(p, x, s, \tau, \theta, \kappa) \in \mathcal{F}^0$ (I suppressed dependence on μ in the notation) such that

$$\begin{pmatrix} XSe \\ \tau\kappa \end{pmatrix} = \mu e.$$

Thus the central path of the (HLP) is the set

$$\left\{(p, x, s, \tau, \theta, \kappa) \in \mathcal{F}^0 : \begin{pmatrix} XSe \\ \tau\kappa \end{pmatrix} = \frac{x^T s + \tau\kappa}{n+1} e\right\},$$

and its β-neighborhood is

$$\mathcal{N}_2(\beta) = \left\{(p, x, s, \tau, \theta, \kappa) \in \mathcal{F}^0 : \left\| \begin{pmatrix} XSe \\ \tau\kappa \end{pmatrix} - \mu e \right\| \le \beta\mu, \text{ where } \mu = \frac{x^T s + \tau\kappa}{n+1}\right\}$$

(we can also define $\mathcal{N}_{-\infty}(\gamma)$ analogously).

In particular, if we choose $p^0 = 0$ and $x^0 = s^0 = e$ (same values we used to calculate $\bar{b}$, $\bar{c}$ and $\bar{z}$), and let $\tau^0 = \theta^0 = \kappa^0 = 1$, we have a perfectly centered initial point for our algorithm, with $\mu^0 = 1$!

8.2 Newton system

Given a current iterate $(p, x, s, \tau, \theta, \kappa) \in \mathcal{F}^0$ and $\mu = \frac{x^T s + \tau\kappa}{n+1}$, we set the next target value of the barrier parameter to be $\sigma\mu$ for an appropriate choice of σ and determine the Newton step by solving the linear system of equations

$$\begin{array}{rrrrrrr}
 & A\Delta x & & -b\Delta\tau & +\bar{b}\Delta\theta & & = 0 \\
-A^T\Delta p & & -\Delta s & +c\Delta\tau & -\bar{c}\Delta\theta & & = 0 \\
b^T\Delta p & -c^T\Delta x & & & \bar{z}\Delta\theta & -\Delta\kappa & = 0 \\
-\bar{b}^T\Delta p & +\bar{c}^T\Delta x & & -\bar{z}\Delta\tau & & & = 0
\end{array}$$

and

$$\begin{pmatrix} X\Delta s + S\Delta x \\ \tau\Delta\kappa + \kappa\Delta\tau \end{pmatrix} = \sigma\mu e - \begin{pmatrix} XSe \\ \tau\kappa \end{pmatrix}.$$

Notice the size of this system is about the same as that of the Newton system we solved in the algorithms described before — there are only three more variables/equations.

With the Newton system defined, and a well-centered point at which to initialize the algorithm, we can now specify a number of interior point methods by giving rules for selecting σ and step sizes at each iterations. They will have properties similar to the algorithms described in Section 6 and can thus be finitely terminated, along the lines of our discussion in Section 7.

9 An infeasible primal-dual interior-point algorithm

In addition to the methods described in the previous section, some IPMs do not require feasibility of iterates at all. Here is one possible implementation of such an algorithm.

Our current iterate is $(\bar{x}, \bar{p}, \bar{s})$ for which $\bar{x} > 0$ and $\bar{s} > 0$, but quite possibly $A\bar{x} \neq b$ and/or $A^T\bar{p} + \bar{s} \neq c$. We are given a target value of the central path barrier parameter $\mu > 0$. We want to compute a direction $(\Delta x, \Delta p, \Delta s)$ so that $(\bar{x} + \Delta x, \bar{p} + \Delta p, \bar{s} + \Delta s)$ approximately solves the central path equations. We set up the system:

$$\begin{aligned} &A(\bar{x} + \Delta x) = b \\ &A^T(\bar{p} + \Delta p) + (\bar{s} + \Delta s) = c \\ &(\bar{X} + \Delta X)(\bar{S} + \Delta S)e = \mu e. \end{aligned}$$

We linearize this system of equations and rearrange the terms to obtain the Newton equations for the current point $(\bar{x}, \bar{p}, \bar{s})$:

$$\begin{aligned} &A\Delta x = b - A\bar{x} =: r_1 \\ &A^T\Delta p + \Delta s = c - A^T\bar{p} - \bar{s} =: r_2 \\ &\bar{S}\Delta x + \bar{X}\Delta s = \mu e - \bar{X}\bar{S}e =: r_3 \end{aligned}$$

Note that this primal-dual Newton direction differs from the one derived earlier in that earlier it was assumed that $r_1 = 0$ and $r_2 = 0$.

Given our current point $(\bar{x}, \bar{p}, \bar{s})$ and a given value of μ, we compute the Newton direction $(\Delta x, \Delta p, \Delta s)$ and we update our variables by choosing primal and dual stepsizes λ_P and λ_D to obtain new values:

$$(\tilde{x}, \tilde{p}, \tilde{s}) \leftarrow (\bar{x} + \lambda_P\Delta x, \bar{p} + \lambda_D\Delta p, \bar{s} + \lambda_D\Delta s).$$

Note that we might use different stepsizes in the primal and dual.

In order to ensure that $\tilde{x} > 0$ and $\tilde{s} > 0$, we choose a value of r satisfying $0 < r < 1$ ($r = 0.99$ is a common value in practice), and determine λ_P and λ_D as follows:

$$\begin{aligned} \lambda_P &= \min\left\{1,\ r \min_{\Delta x_j < 0}\left\{\frac{\bar{x}_j}{-\Delta x_j}\right\}\right\} \\ \lambda_D &= \min\left\{1,\ r \min_{\Delta s_j < 0}\left\{\frac{\bar{s}_j}{-\Delta s_j}\right\}\right\}. \end{aligned}$$

These stepsizes ensure that the next iterate $(\tilde{x}, \tilde{p}, \tilde{s})$ satisfies $\tilde{x} > 0$ and $\tilde{s} > 0$.

We typically declare the problem solved when it is "almost" primal feasible, "almost" dual feasible, and there is "almost" no duality gap. We set our tolerance ϵ to be a small positive number, typically $\epsilon = 10^{-8}$, for example, and we stop when:

$$\begin{aligned} &\|A\bar{x} - b\| \leq \epsilon \\ &\|A^T\bar{p} + \bar{s} - c\| \leq \epsilon \\ &\bar{s}^T\bar{x} \leq \epsilon \end{aligned}$$

An infeasible primal-dual interior point algorithm

1. Given (x^0, p^0, s^0) satisfying $x^0 > 0$, $s^0 > 0$, and $r \in (0,1)$, $\sigma \in (0,1)$, and $\epsilon > 0$. Set $k \leftarrow 0$.

2. Test stopping criterion. Check if:

$$\begin{array}{l} \|Ax^k - b\| \leq \epsilon \\ \|A^T p^k + s^k - c\| \leq \epsilon \\ (s^k)^T x^k \leq \epsilon. \end{array}$$

   If so, STOP. If not, proceed.

3. Set $\mu \leftarrow \sigma \left(\dfrac{(x^k)^T (s^k)}{n} \right)$

4. Solve the Newton equation system:

$$\begin{array}{l} A\Delta x = b - Ax^k =: r_1 \\ A^T \Delta p + \Delta s = c - A^T p^k - s^k =: r_2 \\ S^k \Delta x + X^k \Delta s = \mu e - X^k S^k e =: r_3 \end{array}$$

5. Determine the stepsizes:

$$\begin{aligned} \mu_P &= \min \left\{ 1,\ r \min_{\Delta x_j < 0} \left\{ \frac{x_j^k}{-\Delta x_j} \right\} \right\} \\ \mu_D &= \min \left\{ 1,\ r \min_{\Delta s_j < 0} \left\{ \frac{s_j^k}{-\Delta s_j} \right\} \right\}. \end{aligned}$$

6. Update values:

$$(x^{k+1}, p^{k+1}, s^{k+1}) \leftarrow (x^k + \lambda_P \Delta x, p^k + \lambda_D \Delta p, s^k + \lambda_D \Delta s).$$

   Re-set $k \leftarrow k + 1$ and return to 2.

The algorithm just described is almost exactly what is used in commercial interior-point method software. A typical interior-point code will solve a linear or quadratic optimization problem in 25-80 iterations, regardless of the dimension of the problem. These days, interior-point methods have been extended to allow for the solution of a very large class of convex nonlinear optimization problems.