1 Interior Point Methods

The interior point method for linear programming was introduced by Karmakar in 1984. It runs in polynomial time and is a practical method. For many problems it is competitive or superior to the simplex method. Many LP packages, e.g., CPLEX, offer a simplex as well as an interior point method. Links to implementations can be found on Michael Trick's web page. We give a brief account and refer the readers to [GT89] for more details.

We describe the approach for problems in the form

$$\text{maximize } c^T x \quad \text{subject to} \quad Ax \le b.$$

Let $f_0(x) = c^T x$ and let $f_i(x) = b_i - a_i^T x$ where a_i^T is the i-th row of A, $1 \le i \le m$. Let $P = \{x; Ax \le b\}$ be the feasible set, let x^* be an optimal solution, and let $p^* = c^T x^*$ be the optimal objective value.

The simplex method walks along the boundary of the feasible set, the interior point method walks through the interior. It replaces the constraint $f_i(x) \ge 0$ by a penalty term in the objective function. For real parameter $t \ge 1$, consider the function

$$g_t(x) = f_0(x) + \frac{1}{t} \sum_{1 \le i \le m} \ln f_i(x).$$

The second term

$$\Phi(x) = \sum_{1 \le i \le m} \ln f_i(x) = \ln \prod_{1 \le i \le m} f_i(x)$$

is called a *logarithmic barrier function*; $\Phi(x)$ is undefined for points outside and on the boundary of P. It assumes large negative values for points inside but close to the boundary of P; large t dampen these large negative values. Up to a scale factor (what is the factor?), $f_i(x)$ is the distance of x from the hyperplane $H_i = \{x; a_i^T x = b_i\}$. Thus $\Phi(x)$ is the (weighted) sum of the distances to the defining hyperplanes of P.

The interior point methods determines the minimizer $x^*(t)$ of g_t for larger and larger values of t. This minimizer converges against x^*.

We now give more details.

1. For any real C, consider the level set

   $$L_C = \{x; \Phi(x) = C\}.$$

   Then L_C is the boundary of the convex set $L_{\ge C} = \{x; \Phi(x) \ge C\}$. For $C \to -\infty$, L_C approaches the boundary of P, see Figure 1.

   Let us see a concrete example. Let $P = \{(x,y) \in \mathbb{R}^2; x \ge -1,\ x \le 1,\ y \ge -1,\ y \le +1\}$ be a square and let us maximize $x + y$. Then

   $$g_t(x,y) = x + y + \frac{1}{t} \ln\left((1-x^2)(1-y^2)\right)$$

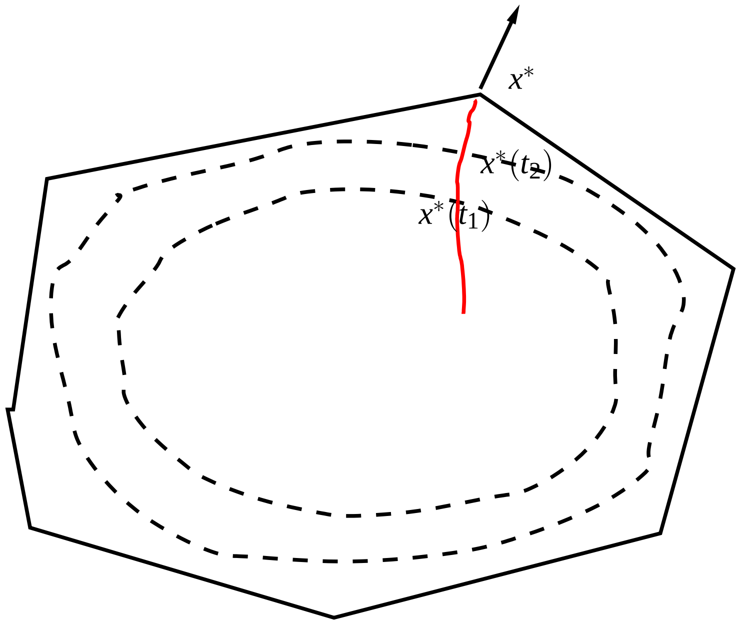

Figure 1: The solid line indicates the boundary of the feasible set, the dashed lines indicate level sets of the potential function (my appologies that the level sets are not convex). The normal of the objective function is indicated by an arrow and x^* is an optimal solution. The red path indicates the set of approximate solutions $x^*(t)$. If L_C is the level set of the barrier function containing $x^*(t)$, the tangent to L_C at $x^*(t)$ is orthogonal to the normal of the objective function.

and hence

$$L_C = \left\{ (x,y);\, (1-x^2)(1-y^2) = e^C \right\}.$$

The following points belong to L_C as a simple calculation shows:

$$(\sqrt{1-e^C}, 0) \quad \text{and} \quad (\sqrt{1-e^{C/2}}, \sqrt{1-e^{C/2}}).$$

Observe that the former point converges to $(1,0)$ and the latter point converges to $(1,1)$ for $C \to -\infty$.

2. Let $x^*(t)$ be the unique minimizer of $g_t(x)$. For $t \to \infty$, $x^*(t)$ converges to an optimal solution of the LP. In particular

$$f_0(x^*(t)) \geq p^* - m/t.$$

We continue with our example. In order to find the optimizer of g_t, it suffices to find the zero of the partial derivatives. The partial derivative with respect to x is

$$1 + \frac{1}{t} \cdot \frac{-2x}{1-x^2}.$$

This has zero

$$x = -\frac{1}{t} + \sqrt{1 + \frac{1}{t^2}} \approx 1 - \frac{1}{t} + \frac{1}{2t^2},$$

where the approximation is valid for large t. Thus the optimizer of g_t has coordinates

$$(-\frac{1}{t} + \sqrt{1 + \frac{1}{t^2}}, -\frac{1}{t} + \sqrt{1 + \frac{1}{t^2}}) \approx (1 - \frac{1}{t} + \frac{1}{2t^2}, 1 - \frac{1}{t} + \frac{1}{2t^2})$$

This point converges to $(1,1)$ for $t \to \infty$. Also the objective value at $(x^*(t), y^*(t)$ is approximately $2 - 2/t + 1/t^2$. This is slightly better than claimed.

3. The function $g_t(x)$ is a strictly concave function on P. Given a reasonably good approximation for $x^*(t)$, an excellent approximation for $x^*(t)$ is easy to find, e.g., by Newton iteration.

4. $x^*(t)$ is a reasonably good approximation for $x^*(\mu t)$ and μ not too large, say $\mu \le 20$.

5. The full algorithm is then as follows:

   Input: The LP plus a tolerance parameter ε.

   Output: a feasible point x such that $p^* - c^T x \le \varepsilon$.

   solve the auxiliary LP $\min\{s; Ax \le b + s1\}$ where 1 is the all-ones vector and s is a scalar. If the minimum is larger than 0, the LP is infeasible. Otherwise we have a feasible point x_0 for our LP. If the minimum is smaller than zero, x_0 lies in the interior of the feasible region. If $s = 0$, the feasible region is lower dimensional and special steps have to be taken. We ignore this complication.
   set $t = 1$ and find $x^*(1)$ by Newton iteration from x_0.
   while $t/m > \varepsilon$ **do**
   $t = \mu t$; — μ is a small constant larger than one
   find $x^*(t)$ from $x^*(t/\mu)$; — Newton iteration
   end while
   output $x^*(t)$

We prove some of the items.

At (1): Let x_0 and x_1 be two points in $L_{\ge C}$ and let $x = \alpha x_0 + (1-\alpha)x_1$ with $0 \le \alpha \le 1$ be any point on the line segment with endpoints x_0 and x_1. We show $\Phi(x) \ge \min(\Phi(x_0), \Phi(x_1))$; the inequality is strict for $0 < \alpha < 1$. Indeed

$$\begin{aligned}
\Phi(x) &= \sum_{1 \le i \le m} \ln\left(b_i - a_i^T(\alpha x_0 + (1-\alpha)x_1\right) \\
&= \sum_{1 \le i \le m} \ln\left(\alpha(b_i - a_i^T x_0) + (1-\alpha)(b_i - a_i^T x_1)\right) \\
&\ge \alpha \sum_{1 \le m} \ln\left(b_i - a_i^T x_0\right) + (1-\alpha) \sum_{1 \le i \le m} \ln\left(b_i - a_i^T x_1\right) \\
&= \alpha\Phi(x_0) + (1-\alpha)\Phi(x_1) \\
&\ge \min(\Phi(x_0), \Phi(x_1)).
\end{aligned}$$

The non-trivial step is the inequality. It holds since for any r and s

$$\ln(\alpha r + (1-\alpha)s) \geq \alpha \ln r + (1-\alpha)\ln s.$$

This can be seen as follows. Consider the line segment with endpoints $(r, \ln r)$ and $(s, \ln s)$. The LHS is the value of the ln-function at $\alpha r + (1-\alpha)s$, the RHS is the value of the line segment. The ln-function is above the line segment and strictly above except at the endpoints. We have now shown the convexity of the set $L_{\geq C}$. The argument above also shows that points x with $\Phi(x) > C$ lie in the interior of $L_{\geq C}$. Thus any point in the boundary belongs to L_C. In fact, L_C is equal to the boundary (Please complete the argument).

If $\Phi(x) = C$, then there must be an i such that $\ln f_i(x) \leq C/n$ or $f_i(x) \leq e^{C/n}$. Thus if $C \to -\infty$, the level sets converge to the boundary of P.

At (2): Consider $g_t(x)$ restricted to a segment with endpoints x_0 and x_1. The function f_0 is linear and hence concave. We showed in the previous item that $\Phi(x)$ is strictly concave. Thus $g_t(x)$ is the sum of a concave function and a strictly concave function and hence strictly concave. Strictly concave functions have unique maxima.

The gradient vector (= vector of partial derivatives) of g_t is given by

$$\begin{aligned}\nabla g_t(x) &= \nabla f_0(x) + \frac{1}{t}\sum_{1\leq i\leq m}\frac{1}{f_i(x)}\nabla f_i(x)\\ &= c - \frac{1}{t}\sum_{1\leq i\leq m}\frac{1}{b_i - a_i x}\cdot a_i.\end{aligned}$$

Observe that $\frac{1}{b_i - a_i x} a_i$ is a scaled version of the vector a_i. It is instructive to compute the 2-norm (= length) of this vector. We have

$$\begin{aligned}\left\|\frac{1}{b_i - a_i x}\cdot a_i\right\|_2 &= \left\|\frac{1}{\frac{b_i}{\|a_i\|_2} - \frac{a_i}{\|a_i\|_2}x}\cdot\frac{a_i}{\|a_i\|_2}\right\|_2\\ &= \left|\frac{1}{\frac{b_i}{\|a_i\|_2} - \frac{a_i}{\|a_i\|_2}x}\right|\cdot\left\|\frac{a_i}{\|a_i\|_2}\right\|_2\\ &= \frac{1}{\frac{b_i}{\|a_i\|_2} - \frac{a_i}{\|a_i\|_2}x}\\ &= \frac{1}{dist(x, H_i)},\end{aligned}$$

where $dist(x, H_i)$ is the distance of x from the hyperplane $a_i x = b_i$. We still need to argue the sequence of equalities. The first equality is a simple rewrite. The second equality holds since the norm of a scaled vector is the absolute value of the scaling factor time the norm of the vector. The third equality holds since $a_i/\|a_i\|_2$ is a unit vector and hence has norm 1. The last equality holds since the value of $b - ax$ is the distance from the hyperplane $b = ax$ if a is a unit vector.

At $x^*(t)$ the function $g_t(x)$ must have a gradient vector equal to zero, i.e.,

$$0 = \nabla g_t(x^*(t)) = c - \frac{1}{t}\sum_{1\leq i\leq m}\frac{a_i}{b_i - a_i x^*(t)}.$$

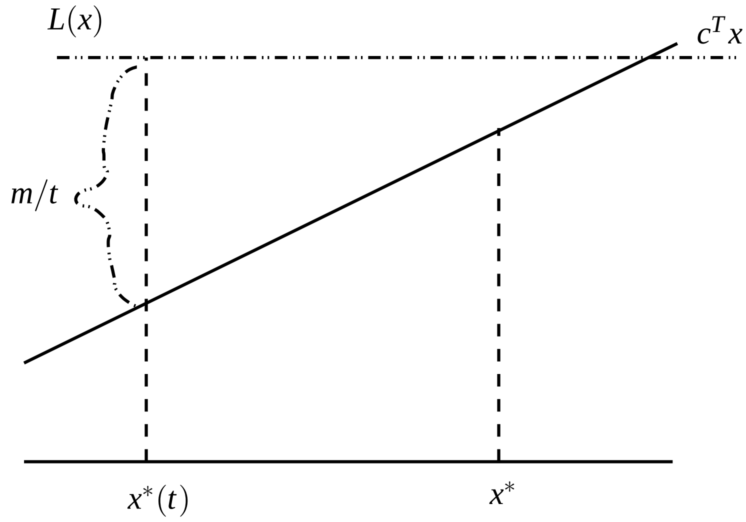

Figure 2: $c^T x^*(t) \geq c^T x^* - m/t$.

So at $x^*(t)$, the vector tc is a linear combination of the normal vectors of the constraints. The scalar for a_i is $1/(b_i - a_i x^*(t))$ which by the above is the inverse of the distance of $x^*(t)$ from the hyperplane H_i.

We next show that the objective value at $x^*(t)$ is at most m/t away from the optimal objective value. Consider the function

$$\begin{aligned}
L(x) &= c^T x + \frac{1}{t} \sum_{1 \leq i \leq m} \frac{b_i - a_i^T x}{b_i - a_i^T x^*(t)} \\
&= \frac{1}{t} \sum_{1 \leq i \leq m} \frac{b_i}{b_i - a_i^T x^*(t)} + \left(c^T + \frac{1}{t} \sum_{1 \leq i \leq m} \frac{-a_i^T}{b_i - a_i^T x^*(t)} \right) x \\
&= \frac{1}{t} \sum_{1 \leq i \leq m} \frac{b_i}{b_i - a_i^T x^*(t)} + \nabla g_t(x^*(t))^T x \\
&= \frac{1}{t} \sum_{1 \leq i \leq m} \frac{b_i}{b_i - a_i^T x^*(t)}.
\end{aligned}$$

Observe that the denumerator of the terms in the sum does not depend on x and hence $L(x)$ is linear in x. Its gradient at $x^*(t)$ is equal to $\nabla g_t(x^*(t))$ and hence is zero. Since $L(x)$ is linear, this implies that $L(x)$ is a constant function. Thus

$$f_0(x^*(t)) \leq f_0(x^*) \leq L(x^*) = L(x^*(t)) = f_0(x^*(t)) + m/t,$$

where the first inequality uses the fact that $x^*(t)$ is feasible and x^* is an optimal feasible point and the second inequality uses the fact that x^* and $x^*(t)$ satisfy all constraints and hence $(b_i - a_i^T x^*)/(b_i - a_i^T x^*(t))$ is positive for all i. Figure 2 captures this reasoning.

At (3): In (2), we have already shown that g_t is strictly concave. For strictly concave functions, a Newton iteration will find the maximum. We refer our students to their course in numerical methods.

At (4): Consult your course on numerical methods.

1.1 A Word on Newton Iteration for Unconstrained Maximization

Newton's method is an iterative method for finding the maximum (or minimum) of a function $g_t(x)$. Assume we have a current approximation x_0 for the maximizer. We model g_t by a quadratic function whose value, derivative and second derivative are those of g_t at x_0. We compute the maximizer of the quadratic function and take it as the next approximation.

Let us first study the one-dimensioncal case, say we want to maximize the function $f(x)$ of one variable. The quadratic functions is

$$q(h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)}{2}h^2.$$

If $f''(x_0) < 0$, q is a parabola with a unique maximum at

$$h = \frac{-f'(x_0)}{f''(x_0)}.$$

Thus the next iterate is

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

The generalization to a function is several variables is straightforward. Assume that f is a function of n variables. The quadratic function is

$$q(h) = f(x_0) + (\nabla f(x_0))^T h + h^T(\nabla^2 f(x_0))h$$

where $\nabla^2 f$ is the matrix of partial second derivatives. If the matrix is negative definite, i.e. $h^T(\nabla^2 f(x_0))h < 0$ for all $h \neq 0$, the quadratic function is a paraboloid with a unique maximum at

$$h = -(\nabla^2 f(x_0))^{-1}\nabla f(x_0).$$

Thus the next iterate is

$$x_1 = x_0 - (\nabla^2 f(x_0))^{-1}\nabla f(x_0).$$

Observe that the step h is computed by solving a linear system.

Let us work out the details for the function g_t. The gradient of g_t at x is

$$\nabla g_t(x) = c - \frac{1}{t}\sum_{1 \le i \le m} \frac{1}{b_i - a_i x} \cdot a_i.$$

The matrix of second derivatives is

$$\nabla^2 g_t(x) = -\frac{1}{t}\sum_{1 \le i \le m} \frac{1}{(b_i - a_i x)^2} \cdot a_i a_i^T.$$

This matrix is negative definite, i.e., for any vector $h \neq 0$ we have $h^T \nabla^2 g_t(x) h \leq 0$. Indeed,

$$-\frac{1}{t} \sum_{1 \leq i \leq m} \frac{1}{(b_i - a_i x)^2} \cdot h^T a_i a_i^T h = -\frac{1}{t} \sum_{1 \leq i \leq m} \frac{1}{(b_i - a_i x)^2} \cdot (a_i^T h)(a_i^T h) \leq 0.$$

The value zero is attained if $a_i^T h = 0$ for all i. This implies $h = 0$ (assuming that the rows of A span the full space). Thus the quadratic function is a paraboloid with a unique maximum. This maximum can be obtained by solving a linear system.

References

[GT89] D. Goldfarb and M.J. Todd. Linear programming. In G.L. Nemhauser et. al., editor, *Handbooks in Combinatorial Optimization and Management Science*. Elsevier, 1989.