MSE2122 - Nonlinear Optimization
Lecture Notes XI

Fernando Dias (based on previous version by Fabricio Oliveira)

November 8, 2023

Abstract

In this lecture, we consider a class of methods known as the barrier methods. We describe the behaviour of such methods in terms of finding optimal solutions for constrained optimisation problems, and provide a general convergence result. Then, we describe the most widespread barrier method, generally known as the interior point method for solving LPs. We develop the main settings of the algorithm and discuss some implementation aspects.

Contents

1 Barrier functions

In essence, barrier methods also use proxies for the constraints in the objective function so that an unconstrained optimisation problem can be solved instead. However, the concept of barrier functions differs from penalty functions in that they are defined to *prevent* the solution search method from leaving the feasible region, which is why some of these methods are also called *interior point methods.*

Consider the primal problem P being defined as:

$$\begin{aligned} (P): \text{ min. } & f(x) \\ \text{subject to: } & g(x) \leq 0 \\ & x \in X. \end{aligned}$$

We define the *barrier problem* BP as:

$$\begin{aligned} (BP): \ & \inf_{\mu} \theta(\mu) \\ \text{subject to: } & \mu > 0 \end{aligned}$$

where $\theta(\mu) = \inf_x \{f(x) + \mu B(x) : g(x) < 0, x \in X\}$ and $B(x)$ is a *barrier function*. The barrier function is such that its value approaches $+\infty$ as the boundary of the region $\{x : g(x) \leq 0\}$ is approached from its interior. In practice, the constraint $g(x) < 0$ can be dropped, as the barrier function automatically enforces them.

The barrier function $B : \mathbb{R}^m \to \mathbb{R}$ is such that:

$$B(x) = \sum_{i=1}^{m} \phi(g_i(x)), \text{ where } \begin{cases} \phi(y) \geq 0, & \text{if } y < 0; \\ \phi(y) = +\infty, & \text{when } y \to 0^-. \end{cases} \tag{1}$$

Examples of barrier functions that impose this behaviour are:

- $B(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)}$
- $B(x) = -\sum_{i=1}^{m} \ln(\min\{1, -g_i(x)\})$.

Perhaps the most important barrier function is the *Frisch's log barrier function*, used in the highly successful primal-dual interior point methods. We will describe its use later. The log barrier is defined as:

$$B(x) = -\sum_{i=1}^{m} \ln(-g_i(x)).$$

Figure 1 illustrates the behaviour of the barrier function. Ideally, the barrier function $B(x)$ has the role of an *indicator function*, which is zero for all feasible solutions $x \in \{x : g(x) < 0\}$ but assume infinite value if a solution is at the boundary $g(x) = 0$ or outside the feasible region. This is illustrated in the dashed line in Figure 1. The barrier functions for different values of barrier term μ illustrate how the log barrier mimics this behaviour, becoming more and more pronounced as μ decreases.

2 The barrier method

Similar to what was developed for penalty methods, we can devise a solution method that successively solves the barrier problem $\theta(\mu)$ for decreasing the barrier term μ.

We start by stating the result that guarantees convergence of the barrier method.

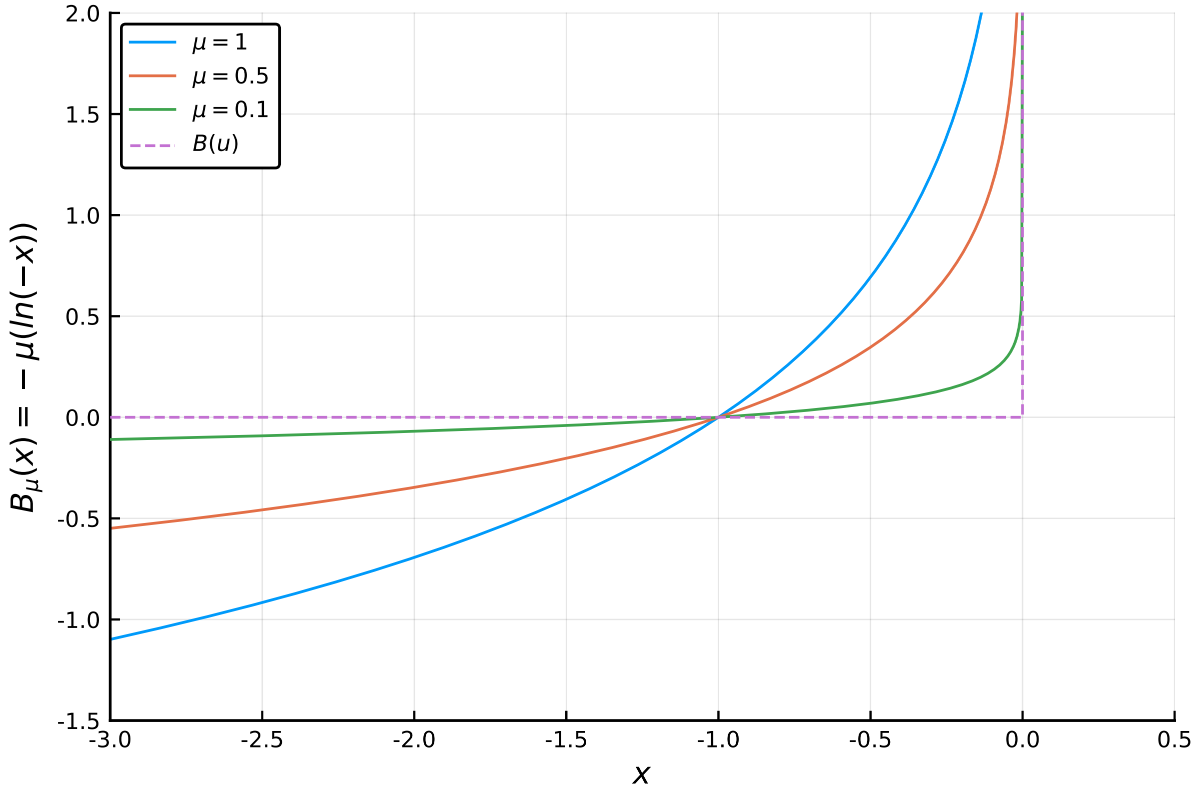

Figure 1: The barrier function for different values of μ

Theorem 2.1. Convergence of barrier methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be continuous functions and $X \in \mathbb{R}^n$ a nonempty closed set in problem P. Suppose $\{x \in \mathbb{R}^n : g(x) < 0, x \in X\}$ is not empty. Let $\overline{x}$ be the optimal solution of P such that, for any neighbourhood $N_\epsilon(\overline{x}) = \ x : ||x - \overline{x}|| \leq \epsilon$, there exists $x \in X \cap N_\epsilon$ for which $g(x) < 0$. Then:

$$\min\{f(x) : g(x) \leq 0, x \in X\} = \lim_{\mu \to 0^+} \theta(\mu) = \inf_{\mu > 0} \theta(\mu).$$

Letting $\theta(\mu) = f(x_\mu) + \mu B(x_\mu)$, where $B(x)$ is a barrier function as described in (1), $x_\mu \in X$ and $g(x_\mu) < 0$, the limit of $\{x_\mu\}$ is optimal to P and $\mu B(x_\mu) \to 0$ as $\mu \to 0^+$.

Proof. First, we show that $\theta(\mu)$ is a nondecreasing function in μ. For $\mu > \lambda > 0$ and x such that $g(x) < 0$ and $x \in X$, we have:

$$\begin{aligned}
&f(x) + \mu B(x) \geq f(x) + \lambda B(x) \\
&\inf_x \{f(x) + \mu B(x)\} \geq \inf_x \{f(x) + \lambda B(x)\} \\
&\theta(\mu) \geq \theta(\lambda).
\end{aligned}$$

From these, we conclude that $\lim_{\mu \to 0^+} \theta(\mu) = \inf\{\theta(\mu) : \mu > 0\}$. Now, let $\epsilon > 0$. As $\overline{x}$ is optimal, by assumption there exists some $\hat{x} \in X$ with $g(\hat{x}) < 0$ such that $f(\overline{x}) + \epsilon > f(\hat{x})$. Then, for $\mu > 0$ we have:

$$f(\overline{x}) + \epsilon + \mu B(\hat{x}) > f(\hat{x}) + \mu B(\hat{x}) \geq \theta(\mu).$$

Letting $\mu \to 0^+$, it follows that $f(\overline{x}) + \epsilon \geq \lim_{\mu \to 0^+} \theta(\mu)$, which implies $f(\overline{x}) \geq \lim_{\mu \to 0^+} \theta(\mu) = \inf\{\theta(\mu) : \mu > 0\}$. Conversely, since $B(x) \geq 0$ and $g(x) < 0$ for some $\mu > 0$, we have:

$$\begin{aligned}
\theta(\mu) &= \inf\{f(x) + \mu B(x) : g(x) < 0, x \in X\} \\
&\geq \inf\{f(x) : g(x) < 0, x \in X\} \\
&\geq \inf\{f(x) : g(x) \leq 0, x \in X\} = f(\overline{x}).
\end{aligned}$$

Thus $f(\overline{x}) \leq \lim_{\mu \to 0^+} \theta(\mu) = \inf\{\theta(\mu) : \mu > 0\}$. Therefore, $f(\overline{x}) = \lim_{\mu \to 0^+} \theta(\mu) = \inf\{\theta(\mu) : \mu > 0 \geq\}$.

The proof has three main steps. First, we show that $\theta(\mu)$ is a nondecreasing function in μ, which implies that $\lim_{\mu \to 0^+} \theta(\mu) = \inf\{\theta(\mu) : \mu > 0\}$. This can be trivially shown as only feasible solutions x must be considered.

Next, we show the convergence of the barrier method by showing that $\inf_{\mu>0} \theta(\mu) = f(\overline{x})$, where $\overline{x} = \text{argmin}\{f(x) : g(x) \leq 0, x \in X\} = \lim_{\mu \to 0^+} \theta(\mu) = \inf_{\mu>0} \theta(\mu)$. The optimality of $\overline{x}$ implies that $f(\hat{x}) - f(\overline{x}) < \epsilon$ for feasible $\hat{x}$ and $\epsilon > 0$. Moreover, $B(\hat{x}) \geq 0$ by definition. In the last part, we argue that including the boundary can only improve the objective function value, leading to the last inequality. It is worth highlighting that, to simplify the proof, we have assumed that the barrier function has the form described in (1). However, a proof in the veins of Theorem 2.1 can still be developed for the Frisch log barrier (for which $B(x)$ is not necessarily nonnegative) since, essentially, (1) only needs to be observed in a neighbourhood of $g(x) = 0$.

The result in Theorem 2.1 allows the design of an optimisation method that, starting from a strictly feasible (interior) solution, is based on successively reducing the barrier term until a solution with an arbitrarily small barrier term is obtained. Algorithm 1 present a pseudo code for such a method.

Algorithm 1 Barrier method

```
1: initialise. ε > 0, x^0 ∈ X with g(x^k) < 0, μ^k, β ∈ (0,1), k = 0.
2: while μ^k B(x^k) > ε do
3:     x̄^{k+1} = argmin{f(x) + μ^k B(x) : x ∈ X}
4:     μ^{k+1} = βμ^k, k = k + 1
5: end while
6: return x^k.
```

One important aspect to notice is that starting the algorithm requires a strictly feasible point, which, in some applications, might be challenging to obtain. This characteristic renders the name *interior point methods* for this class of algorithms.

Consider the following example. Let $P = \{(x+1)^2 : x \geq 0\}$. Let us assume we use the barrier function $B(x) = -\ln(x)$. Then, the unconstrained barrier problem becomes:

$$(BP) : \min_{x}. \ (x+1)^2 - \mu \ln(x). \tag{2}$$

Since this is a convex function, the first-order condition $f'(x) = 0$ is necessary and sufficient for optimality. Thus, solving $2(x+1) - \frac{\mu}{x} = 0$ we obtain the positive root and unique solution $x_\mu = \frac{-1}{2} + \frac{\sqrt{4+8\mu}}{4}$. Figure X shows the behaviour of the problem as μ converges to zero. As can be seen, as $\mu \to 0$, the optimal solution x_μ converges to the constrained optimum $\overline{x} = 0$.

We now consider a more involved example. Let us consider the problem:

$$P = \min. \ \{(x_1 - 2)^4 + (x_1 - 2x_2)^2 : x_1^2 - x_2 \leq 0\}$$

with $B(x) = -\frac{1}{x_1^2 - x_2}$. We implemented Algorithm 1 and solved it with two distinct values for the penalty term μ and reduction term β. Figure 2 illustrates the trajectory of the algorithm with each parameterisation, exemplifying how these can affect the convergence of the method.

3 Interior point method for LP/QP problems

Perhaps ironically, the most successful applications of barrier methods in terms of efficient implementations are devoted to solving linear and quadratic programming (LP/QP) problems. In the last decade, the primal-dual interior point method has become the algorithm of choice for many applications involving large-scale LP/QP problems.

To see how barrier methods can be applied to LP problems, consider the following primal/dual pair

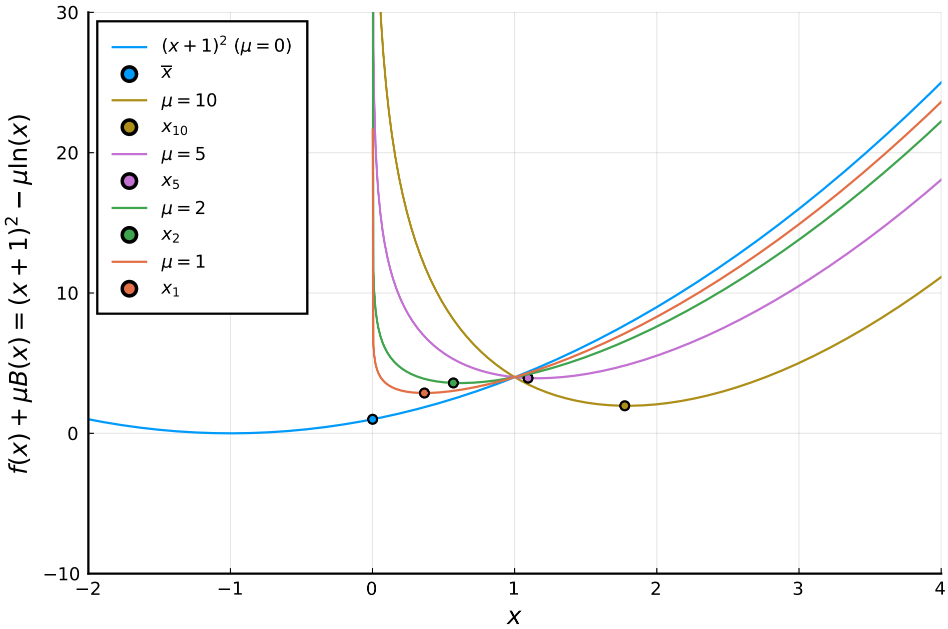

Figure 2: Example 1: solving a one-dimensional problem with the barrier method

formed by an LP primal P:

$$\begin{aligned}(P) \ : \ \text{min.} \ & c^\top x \\ \text{subject to:} \ & Ax = b \ : v \\ & x \geq 0 \ : u\end{aligned}$$

and its respective dual formulation D:

$$\begin{aligned}(D) \ : \ \text{max.} \ & b^\top v \\ \text{subject to:} \ & A^\top v + u = c \\ & u \geq 0, v \in \mathbb{R}^m.\end{aligned}$$

The optimal solution $(\overline{x}, \overline{v}, \overline{u}) = \overline{w}$ is such that it satisfies KKT conditions of P, given by:

$$\begin{aligned}& Ax = b, \ x \geq 0 \\ & A^\top v + u = c, \ u \geq 0, \ v \in \mathbb{R}^m \\ & u^\top x = 0.\end{aligned}$$

These are going to be useful as a reference for the next developments. We start by considering the *barrier problem* for P by using the logarithmic barrier function to represent the condition $x \geq 0$. Thus, the barrier problem BP can be defined as:

$$\begin{aligned}(BP) \ : \ \text{min.} \ & c^\top x - \mu \sum_{i=1}^{n} \ln(x_j) \\ \text{subject to:} \ & Ax = b.\end{aligned}$$

Notice that this problem is a strictly equality-constrained problem that can be solved using the constrained variant of Newton's method (which simply consists of employing Newton's method to solve the KKT conditions of equality-constrained problems). Moreover, observe that the KKT conditions of BP are:

$$\begin{aligned}& Ax = b, \ x > 0 \\ & A^\top v = c - \mu \left(\frac{1}{x_1}, \dots, \frac{1}{x_n}\right)\end{aligned}$$

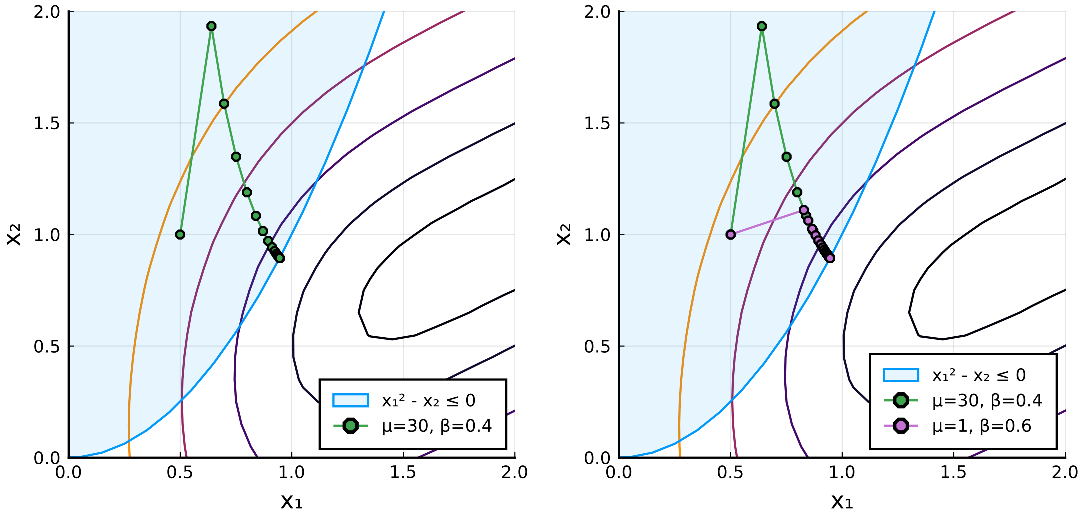

Figure 3: The trajectory of the barrier method for problem P. Notice how the parameters influence the trajectory and number of iterations. The parameters on the left require 27 iterations, while those on the right require 40 for convergence.

Notice that, since $\mu > 0$ and $x > 0$, $u = \mu\left(\frac{1}{x_1}, \ldots, \frac{1}{x_n}\right)$ serves as an estimate for the Lagrangian dual variables. To further understand the relationship between the optimality conditions of BP and P, let us define $X \in \mathbb{R}^{n\times n}$ and $U \in \mathbb{R}^{n\times n}$ as:

$$X = \mathbf{diag}(x) = \begin{bmatrix} \ddots & & \\ & x_i & \\ & & \ddots \end{bmatrix} \text{ and } U = \mathbf{diag}(u) = \begin{bmatrix} \ddots & & \\ & u_i & \\ & & \ddots \end{bmatrix}$$

and let $e = [1, \ldots, 1]^\top$ be a vector of ones of suitable dimension. We can rewrite the KKT conditions of BP as:

$$Ax = b,\ x > 0 \tag{3}$$

$$A^\top v + u = c \tag{4}$$

$$u = \mu X^{-1} e \ \Rightarrow\ XUe = \mu e. \tag{5}$$

Notice how the condition (5) resembles the complementary slackness from P, but *relaxed* to be μ instead of zero. This system is often called the *perturbed KKT system.*

Theorem 2.1[1] guarantees that $w_\mu = (x_\mu, v_\mu, u_\mu)$ approaches the optimal primal-dual solution of P as $\mu \to 0^+$. The trajectory formed by successive solutions $\{w_\mu\}$ is called the *central path* due to the interiority enforced by the barrier function. When the barrier term μ is large enough, the solution of the barrier problem is close to the analytic centre of the feasibility set. The analytic centre of a polyhedral set $S = \{x \in \mathbb{R}^n : Ax \le b\}$ is given by:

$$\begin{aligned} &\max_x. \prod_{i=1}^{m} (b_i - a_i^\top x) \\ &\text{subject to: } x \in X, \end{aligned}$$

Which corresponds to finding the point of maximum distance to each of the hyperplanes forming the polyhedral set. This is equivalent to the convex problem:

[1] In fact, we require a slight variant of Theorem 1 that allows for $B(x) \ge 0$ only being required in a neighbourhood of $g(x) = 0$.

$$\min_x. \sum_{i=1}^{m} -\ln\left(b_i - a_i^\top x\right)$$
$$\text{subject to: } x \in X,$$

And thus justifying the nomenclature.

One aspect should be observed for defining the stopping criterion. Notice that the term $u^\top x$ is such that it measures the duality gap at a given solution. That is, notice that:

$$\begin{aligned} c^\top x &= (A^\top v + u)^\top x \\ &= (A^\top v)^\top x + u^\top x \\ &= v^\top (Ax) + u^\top x \\ c^\top x - b^\top v = u^\top x &= \sum_{i=1}^{n} u_i x_i = \sum_{i=1}^{n} \left(\frac{\mu}{x_i}\right) x_i = n\mu. \end{aligned}$$

This gives the *total slack violation* that can be used to determine the algorithm's convergence.

3.1 Primal/dual path-following interior point method

The primal/dual path-following interior point method (IPM) is the specialised version of the setting described earlier for solving LP/QP problems.

It consists of building upon employing logarithmic barriers to LP/QP problems and solving the system (3) - (5) using Newton's method. However, instead of solving the problem to optimality for each μ, only a *single* Newton step is taken before the barrier term μ is reduced.

Suppose we start with a $\overline{\mu} > 0$ and a $w^k = (x^k, v^k, u^k)$ sufficiently close to $w_{\overline{\mu}}$. Then, for a sufficiently small $\beta \in (0, 1)$, $\beta\overline{\mu}$ will lead to a w^{k+1} sufficiently close to $w_{\beta\overline{\mu}}$. Figure 4 illustrates this effect, showing how a suboptimal solution x^k does not necessarily need to be in the central path (denoted by the dashed line) to guarantee convergence, as long as they are guaranteed to remain within the same neighbourhood $N_\mu(\theta)$ of the central path.

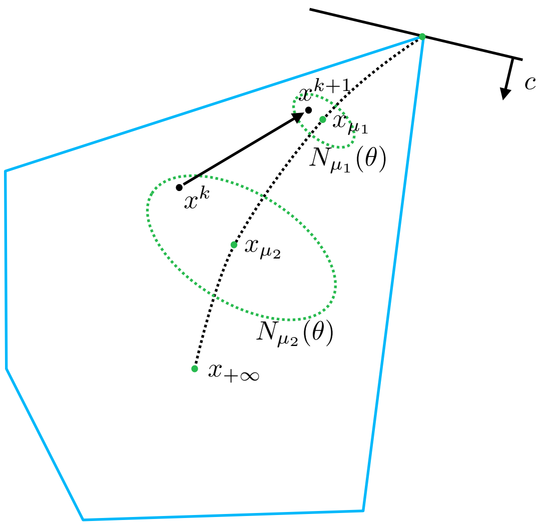

Figure 4: an illustrative representation of the central path and approximately how the IPM follows it.

For example, let $N_\mu(\theta) = ||X_\mu U_\mu e - \mu e|| \leq \theta\mu$. Then, by selecting $\beta = 1 - \frac{\sigma}{\sqrt{n}}$, $\sigma = \theta = 0.1$, and $\mu^0 = (x^\top u)/n$, successive Newton steps are guaranteed to remain within $N_\mu(\theta)$.

To see how the setting works, let the perturbed KKT system (3) – (5) for each $\hat{\mu}$ be denoted as

$H(w) = 0$. Let $J(\overline{w})$ be the Jacobian of $H(w)$ at $\overline{w}$.

Applying Newton's method to solve $H(w) = 0$ for $\overline{w}$, we obtain:

$$J(\overline{w})d_w = -H(\overline{w}) \tag{6}$$

where $d_w = (w - \overline{w})$. By rewriting $d_w = (d_x, d_v, d_u)$, (6) can be equivalently stated as:

$$\begin{bmatrix} A & 0^\top & 0 \\ 0 & A^\top & I \\ \overline{U} & 0^\top & \overline{X} \end{bmatrix} \begin{bmatrix} d_x \\ d_v \\ d_u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hat{\mu}e - \overline{X}\,\overline{U}e \end{bmatrix}. \tag{7}$$

The system (7) is often called the Newton's system.

In practice, the updates incorporate primal and dual infeasibility, which precludes the need for additional mechanisms to guarantee primal and dual feasibility throughout the algorithm. This can be achieved with a simple modification in the Newton system, rendering the direction update step:

$$\begin{bmatrix} A & 0^\top & 0 \\ 0 & A^\top & I \\ U^k & 0^\top & X^k \end{bmatrix} \begin{bmatrix} d_x^{k+1} \\ d_v^{k+1} \\ d_u^{k+1} \end{bmatrix} = - \begin{bmatrix} Ax^k - b \\ A^\top v^k + u^k - c \\ X^k U^k e - \mu^{k+1} e \end{bmatrix}, \tag{8}$$

To see how this still leads to primal and dual feasible solutions, consider the primal residuals (i.e., the amount of infeasibility) as $r_p(x, u, v) = Ax - b$ and the dual residuals $r_d(x, u, v) = A^\top v + u - c$. Now, let $r(w) = r(x, u, v) = (r_p(x, u, v), r_d(x, u, v))$, recalling that $w^k = (x, v, u)$. The optimality conditions can be expressed as requiring that the residuals vanish, that is $r(\overline{w}) = 0$.

Now, consider the first-order Taylor approximation for r at w for a step d_w:

$$r(w + d_w) \approx r(w) + Dr(w)d_w,$$

Where $Dr(w)$ is the derivative of r evaluated at w, given by the two first rows of the Newton system (7). The step d_w for which the residue vanishes is:

$$Dr(w)d_w = -r(w), \tag{9}$$

Which is the same as (6) without the bottom equation. Now, if we consider the directional derivative of the square of the norm of r in the direction d_w, we obtain:

$$\frac{d}{dt}||r(w + td_w)||_2^2 \Big|_{t \to 0^+} = 2r(w)^\top Dr(w)d_w = -2r(w)^\top r(w), \tag{10}$$

Which is strictly decreasing. That is, the step d_w is such that it will make the residual decrease and eventually become zero. From that point onwards, the Newton system will take the form of (7).

The algorithm proceeds by iteratively solving the system (8) with $\mu^{k+1} = \beta\mu^k$ with $\beta \in (0, 1)$ until $n\mu^k$ is less than a specified tolerance. Algorithm 2 summarises a simplified form of the IPM.

Algorithm 2 Interior point method (IPM) for LP

1: **initialise.** primal-dual feasible w^k, $\epsilon > 0$, μ^k, $\beta \in (0, 1)$, $k = 0$.
2: **while** $n\mu = c^\top x^k - b^\top v^k > \epsilon$ **do**
3: compute $d_{w^{k+1}} = (d_{x^{k+1}}, d_{v^{k+1}}, d_{u^{k+1}})$ using (8) and w^k.
4: $w^{k+1} = w^k + d_{w^{k+1}}$
5: $\mu^{k+1} = \beta\mu^k$, $k = k + 1$
6: **end while**
7: **return** w^k.

Figure 5 illustrates the behaviour of the IPM when employed to solve the linear problem:

$$\begin{aligned} \text{min. } & x_1 + x_2 \\ \text{subject to: } & 2x_1 + x_2 \geq 8 \\ & x_1 + 2x_2 \geq 10, \\ & x_1, x_2 \geq 0 \end{aligned}$$

Considering two distinct initial penalties μ. Notice how higher penalty values enforce a more central convergence of the method.

Some points are worth noticing concerning Algorithm 2. First, notice that a fixed step size is considered

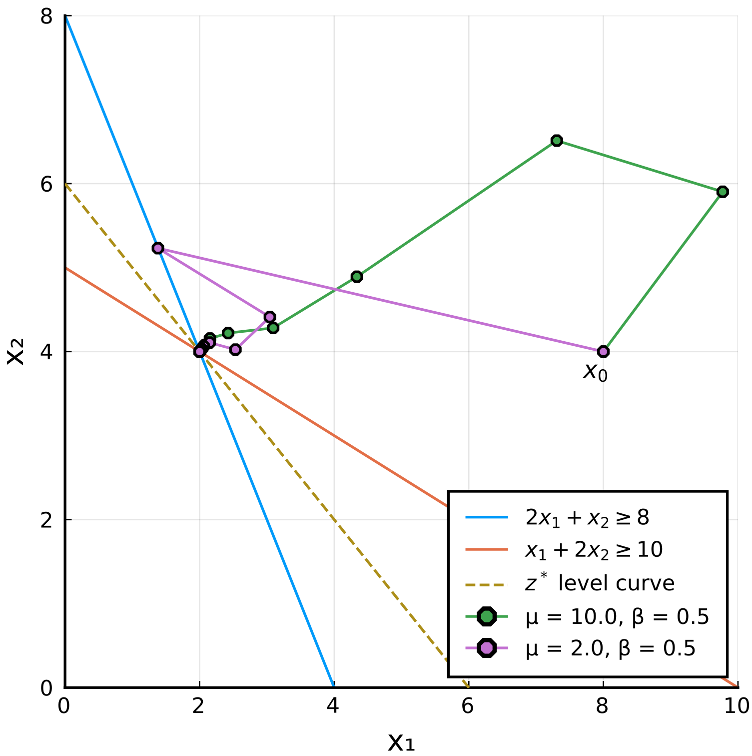

Figure 5: IPM applied to an LP problem with two different barrier terms

in Line 4. A line search can be incorporated to prevent infeasibility and improve numerical stability. Typically, it is used $\lambda_i^k = \min\{\alpha, -\frac{x_i^k}{d_i^k}\}$ with $\alpha < 1$ but close to 1.

Also, even though the algorithm is initialised with a feasible solution w^k, this might, in practice, not be necessary. Implementations of the infeasible IPM method can efficiently handle primal and dual infeasibility.

Under specific conditions, the IPM can be shown to have complexity of $O(\sqrt{n}\ln(1/\epsilon))$, which is polynomial and of much better worst-case performance than the simplex method, which makes it the algorithm of choice for solving large-scale LPs. Another important advantage is that IPM can be modified with little effort to solve a wider class of problems under the class of *conic optimisation problems*.

Predictor-corrector methods are variants of IPM that incorporate a two-phase direction calculation using a *predicted* direction d_w^{pred}, calculated by setting $\mu = 0$ and a *correcting* direction, which is computed considering the impact that d_w^{cor} would have in the term $\overline{X}\,\overline{U}e$.

Let $\Delta X = \mathbf{diag}(d_x^{\text{pred}})$ and $\Delta U = \mathbf{diag}(d_u^{\text{pred}})$.

Then:

$$\begin{aligned}(X + \Delta X)(U + \Delta U)e &= XUe + (U\Delta X + X\Delta U)e + \Delta X \Delta U e \\ &= XUe + (0 - XUe) + \Delta X \Delta U e \\ &= \Delta X \Delta U e \end{aligned} \tag{11}$$

Using the last equation (11), the corrector Newton step becomes $\overline{U}d_x + \overline{X}d_u = \hat{\mu}e - \Delta X \Delta U e$.

Finally, d_w^k is set to be a combination of d_w^{pred} and d_w^{cor}.