A short tutorial on interior point methods

log-barrier method

Consider the following constrained minimzation problem

$$\begin{aligned} &\min_{\boldsymbol{x}} f_0(\boldsymbol{x}) \\ &\text{subject to} \\ &f_1(\boldsymbol{x}) \leq 0 \\ &\vdots \\ &f_m(\boldsymbol{x}) \leq 0, \\ &\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^p$, $f_0 : \mathbb{R}^p \to \mathbb{R}$ is called objective function $f_k : \mathbb{R}^p \to \mathbb{R}$, $k = 1, \ldots, m$, are called constraint functions. We assume that $\boldsymbol{A}$ is a $l \times p$ matrix of full rank. Both the objective function and the constraint functions are assumed to be convex functions of $\boldsymbol{x}$, i.e. we consider the minimization of a convex function over a convex set. If $\boldsymbol{x} \in \mathbb{R}^p$ satisfies all constraints induced by the functions $\{f_k\}_{k=1}^m$ and the matrix $\boldsymbol{A}$, it is said to be feasible. If all constraints are satisfied with strict inequality, $\boldsymbol{x}$ is said to be strictly feasible. The idea of the log-barrier method is to solve above minimization problem by solving a sequence of unconstrained convex problems. For $t > 0$, these minimization problems are of the form

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}) - \underbrace{\frac{1}{t}\sum_{k=1}^{m} \log(-f_k(\boldsymbol{x}))}_{\mathcal{B}(\boldsymbol{x})} + \boldsymbol{\nu}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}). \tag{2}$$

where t is increased when switching from one problem to the next. The term $\mathcal{B}(\boldsymbol{x})$ is called log-barrier. Let $\{\boldsymbol{x}^\star(t),\ t > 0\}$ the set of all minimizers of all problem instances indexed by the parameter t. Then the $\boldsymbol{x}^\star(t)$ are strictly feasible, and each of the $\boldsymbol{x}^\star(t)$ yields a dual feasible point, $\boldsymbol{\lambda}^\star(t) = (\lambda_k^\star(t))_{k=1}^m$, say, and hence a lower bound for the optimal function value $p^\star$ of the optimization problem (1). To see this, the optimality condition of the log-barrier of the problem (2) is obtained by setting the gradient w.r.t. $\boldsymbol{x}$ equal to zero read

$$\begin{aligned} &\nabla f_0(\boldsymbol{x}^\star(t)) + \frac{1}{t}\nabla\mathcal{B}(\boldsymbol{x}^\star(t)) + \boldsymbol{A}^\top\boldsymbol{\nu}^\star(t) = \boldsymbol{0} \\ \Leftrightarrow\ &\nabla f_0(\boldsymbol{x}^\star(t)) - \frac{1}{t}\sum_{k=1}^{m}\frac{\nabla f_k(\boldsymbol{x}^\star(t))}{f_k(\boldsymbol{x}^\star(t))} + \boldsymbol{A}^\top\boldsymbol{\nu}^\star(t) = \boldsymbol{0} \\ \Leftrightarrow\ &\nabla f_0(\boldsymbol{x}^\star(t)) + \sum_{k=1}^{m}\lambda_k^\star(t)\nabla f_k(\boldsymbol{x}^\star(t)) + \boldsymbol{A}^\top\boldsymbol{\nu}^\star(t) = \boldsymbol{0}, \end{aligned}$$

by setting $\lambda_k^\star(t) = -\frac{1}{tf_k(\boldsymbol{x}^\star(t))}$, $k = 1, \ldots, m$, which are all strictly positive since $\boldsymbol{x}^\star(t)$ is strictly feasible. The last equation has the typical form of the KKT optimality condition. To be more precise, consider the primal Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{k=1}^{m}\lambda_k f_k(\boldsymbol{x}) + \boldsymbol{\nu}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

Then $\boldsymbol{x}^\star(t)$ minimizes the Lagrangian for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star(t)$ and $\boldsymbol{\nu} = \boldsymbol{\nu}^\star(t)$. Since the pair $(\boldsymbol{\lambda}^\star(t), \boldsymbol{\nu}^\star(t))$ is dual feasible, the dual Lagrangian is finite, and

$$\begin{aligned}\mathcal{L}^*(\boldsymbol{\lambda}^\star(t), \boldsymbol{\nu}^\star(t)) &= f_0(\boldsymbol{x}^\star(t)) + \sum_{k=1}^{m} \lambda_k^\star(t) f_k(\boldsymbol{x}^\star(t)) + \boldsymbol{\nu}^\star(t)(\boldsymbol{A}\boldsymbol{x}^\star(t) - \boldsymbol{b}) \\ &= f_0(\boldsymbol{x}^\star(t)) - m/t,\end{aligned}$$

concluding that the duality gap is no more than m/t, i.e. $\boldsymbol{x}^\star(t)$ is at most m/t suboptimal. The suboptimality also becomes obvious when considering the complementarity slackness condition

$$-\sum_{k=1}^{m} \lambda_k^\star(t) f_k(\boldsymbol{x}^\star(t)) = 1/t.$$

All in all, these considerations prompt the following algorithm.

1. Initialize t. Find a strictly primal feasible solution.
2. Solve the log-barrier problem for the parameter t using e.g. the Newton method. This yields an intermediate solution $\boldsymbol{x}^\star(t)$.
3. Increase the parameter t by a factor $\mu > 0$ and go back to 2., using the current $\boldsymbol{x}^\star(t)$ as starting value.
4. Stop when t exceeds a pre-specified tolerance.

The algorithm is subject to the following drawbacks: i) it is in general unclear how exact the solution of in 2. has to be, and ii) the number of outer iterations (= number of log-barrier problems) is fixed in advance. These two problems are circumevented in the approach described in the following.

primal-dual interior point method

The following approach is somewhat more direct than the log-barrier method, to which it is closely related (see Boyd and Vandenberghe (2004) for a discussion). In particular, both have in common that they work well in combination with the Newton method. As starting point, let us consider the following m/t-'approximate version' ($t > 0$) of the KKT optimality conditions.

$$\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0},$$

where

$$\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla f_0(\boldsymbol{x}) + \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda} + \boldsymbol{A}^\top \boldsymbol{\nu} \\ -\mathrm{diag}(\lambda_1, \ldots, \lambda_k)\boldsymbol{f}(\boldsymbol{x}) - \frac{m}{t}\boldsymbol{1} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{\text{dual}} \\ \boldsymbol{r}_{\text{cent}} \\ \boldsymbol{r}_{\text{primal}} \end{bmatrix},$$

$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))^\top$,

$$\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \nabla f_1(\boldsymbol{x})^\top \\ \vdots \\ \nabla f_k(\boldsymbol{x})^\top \end{bmatrix}.$$

In order to solve the nonlinear equation $r_t(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathbf{0}$, we apply a Taylor expansion. Denoting the current point by $\boldsymbol{y} = (\boldsymbol{x}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$ and the Newton step direction by $\boldsymbol{\Delta y} = (\boldsymbol{\Delta x}^\top, \boldsymbol{\Delta \lambda}^\top, \boldsymbol{\Delta \nu}^\top)$, we expand

$$\begin{aligned} & r_t(\boldsymbol{y} + \boldsymbol{\Delta y}) \approx \boldsymbol{r}_t(\boldsymbol{y}) + \boldsymbol{D} r_t(\boldsymbol{y}) \boldsymbol{\Delta y} = \mathbf{0} \\ \Leftrightarrow \; & \boldsymbol{\Delta y} = -\boldsymbol{D r}_t(\boldsymbol{y})^{-1}, \end{aligned}$$

where $\boldsymbol{D r}_t(\boldsymbol{y})$ is given by the block partitioned matrix

$$\boldsymbol{D r}_t(\boldsymbol{y}) = \begin{bmatrix} \nabla^2 f_0(\boldsymbol{x}) + \sum_{k=1}^m \lambda_k \nabla^2 f_k(\boldsymbol{x}) & \boldsymbol{Df}(\boldsymbol{x})^\top & \boldsymbol{A}^\top \\ -\operatorname{diag}(\lambda_1, \ldots, \lambda_k) \boldsymbol{Df}(\boldsymbol{x}) & -\operatorname{diag}(f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})) & \mathbf{0} \\ \boldsymbol{A} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Solving the Newton equation gives a step direction $(\boldsymbol{\Delta x}^\top, \boldsymbol{\Delta \lambda}^\top, \boldsymbol{\Delta \nu}^\top)$. The stepsize is determined using a line search which takes care of all inequalities (see Boyd and Vandenberghe (2004) for a detailed description).
We see that − as opposed to the log barrier method − both the primal and dual problem are dealt with simultaneously. Similiarly to the log-barrier method, the quality of approximation is improved in each step by incrementing the parameter t appropriately.
In some applications, it is possible and useful to solve for $\boldsymbol{\Delta x}, \boldsymbol{\Delta \nu}$ first and then resolve for $\boldsymbol{\Delta \lambda}$ (cf. example below). We have

$$\begin{aligned} & -\operatorname{diag}(\lambda_1, \ldots, \lambda_k) \boldsymbol{Df}(\boldsymbol{x}) \boldsymbol{\Delta x} - \operatorname{diag}(f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x})) \boldsymbol{\Delta \lambda} = -\boldsymbol{r}_{\text{cent}} \\ \Leftrightarrow \; & \boldsymbol{\Delta \lambda} = \operatorname{diag}(1/f_1(\boldsymbol{x}), \ldots, 1/f_k(\boldsymbol{x})) \left(\boldsymbol{r}_{\text{cent}} - \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \boldsymbol{Df}(\boldsymbol{x}) \boldsymbol{\Delta x} \right) \end{aligned}$$

Substituting this result back yields the block-partioned linear system

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^\top \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta x} \\ \boldsymbol{\Delta \nu} \end{bmatrix} = - \begin{bmatrix} \boldsymbol{r}_{\text{dual}} + \boldsymbol{Df}(\boldsymbol{x})^\top \operatorname{diag}(1/f_1(\boldsymbol{x}), \ldots, 1/f_m(\boldsymbol{x})) \boldsymbol{r}_{\text{cent}} \\ \boldsymbol{r}_{\text{pri}} \end{bmatrix},$$

with $\boldsymbol{H}$ given by

$$\boldsymbol{H} = \nabla^2 f_0(\boldsymbol{x}) + \sum_{k=1}^m \lambda_k \nabla^2 f_k(\boldsymbol{x}) + \sum_{k=1}^m \frac{\lambda_k}{-f_k(\boldsymbol{x})} \nabla f_k(\boldsymbol{x}) \nabla f_k(\boldsymbol{x})^\top.$$

With these preparations, we can write down the structure of the primal-dual interior point algorithm.

1. Initialize with a strictly feasible triple $(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$.
2. Beginning of each iteration: compute the so-called surrogate duality gap $\eta = -\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda}$, which is used to set $t = \mu m / \eta$, for some $\mu > 0$.
3. Perform the Newton step outlined above.
4. Stop if the duality gap and the size of the residuals $\boldsymbol{r}_{\text{pri}}$ and $\boldsymbol{r}_{\text{dual}}$ as measured by their norms drop below a prespecified tolerance.