Analysis of the logarithmic barrier method

Mark S. Gockenbach

1 Introduction

I will now analyze the performance of the logarithmic barrier method for the inequality-constrained nonlinear program

$$\begin{aligned} \min \quad & f(x) & (1)\\ s.t. \quad & h(x) \geq 0. & (2) \end{aligned}$$

My approach will be the same as for the quadratic penalty method or the augmented Lagrangian method: I apply the implicit function theorem to a system of equations defining the first-order necessary conditions.

2 Existence of minimizers of the logarithmic barrier function

The logarithmic barrier function is

$$B(x;\mu) = f(x) - \mu \sum_{i=1}^{p} \log\left(h_i(x)\right),$$

and so

$$\begin{aligned} \nabla B(x;\mu) &= \nabla f(x) - \mu \sum_{i=1}^{p} \frac{1}{h_i(x)} \nabla h_i(x) \\ &= \nabla f(x) - \nabla h(x)(\mu h(x)^{-1}). \end{aligned}$$

Therefore, a stationary point of $B(\cdot;\mu)$ must satisfy

$$\nabla f(x) - \nabla h(x)\lambda,$$

where

$$\lambda = \mu h(x)^{-1}.$$

This last equation can be written

$$\lambda h(x) = \mu e,$$

where e is the vector of all ones and $\lambda h(x)$ is defined by

$$\lambda h(x) \in \mathrm{R}^p,\ (\lambda h(x))_i = \lambda_i h_i(x),\ i = 1, 2, \ldots, p.$$

The result is the system

$$\begin{aligned} \nabla f(x) - \nabla h(x)\lambda &= 0, & (3)\\ \lambda h(x) &= \mu e & (4) \end{aligned}$$

of $n+p$ equations in $n+p$ unknowns (the components of x and λ).

The reader will recall that the first-order conditions for a minimizer of (1–2) are

$$\begin{aligned} \nabla f(x^*) - \nabla h(x^*)\lambda^* &= 0, \\ h(x^*) &\geq 0, \\ \lambda^* &\geq 0, \\ \lambda^* h(x^*) &= 0. \end{aligned}$$

Equations (3–4) are thus the Lagrange multiplier equation together with a perturbation of the complementarity condition. I will define $F : \mathrm{R}^n \times \mathrm{R}^p \times \mathrm{R} \rightarrow \mathrm{R}^n \times \mathrm{R}^p$ by

$$F(x, \lambda; \mu) = \begin{bmatrix} \nabla f(x) - \nabla h(x)\lambda = 0 \\ \lambda h(x) - \mu e \end{bmatrix}.$$

If $x^* \in \mathrm{R}^n$ and $\lambda^* \in \mathrm{R}^p$ satisfy the first-order necessary conditions, then

$$F(x^*, \lambda^*; 0) = 0.$$

In order to write the Jacobian of F conveniently, I will adopt the following convention: For any vector denoted by a lower-case letter, the corresponding capital letter will indicate the diagonal matrix with the components of the vector as the diagonal entries. In particular, I write

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}, \ H(x) = \begin{bmatrix} h_1(x) & & & \\ & h_2(x) & & \\ & & \ddots & \\ & & & h_p(x) \end{bmatrix}.$$

The Jacobian of F (with respect to (x, λ)) is

$$J(x, \lambda; \mu) = \begin{bmatrix} \nabla^2 \ell(x; \lambda) & -\nabla h(x) \\ \Lambda \nabla h(x)^T & H(x) \end{bmatrix}.$$

In order that the implicit function theorem be applicable, it is necessary that the Jacobian $J^* = J(x^*, \lambda^*; 0)$ be nonsingular. As I will show below, this is true only if x^*, λ^* satisfy *strict complementarity*:

$$\text{Exactly one of } \lambda_i^* \text{ and } h_i(x^*) \text{ is zero.} \tag{5}$$

I will therefore define x^* to be a nonsingular point of the NLP (1–2) if

1. x^* is a regular point;
2. there exists a Lagrange multiplier λ^* such that
   (a) x^* and λ^* satisfy the strict complementarity condition;
   (b) x^* and λ^* satisfy the second-order sufficient condition.

Given that strict complementarity holds, the second-order sufficient condition reduces to

$$z \neq 0, \ \nabla h_i(x^*) \cdot z = 0 \text{ for all } i \in \mathcal{A}(x^*) \ \Rightarrow \ z \cdot \nabla^2 \ell(x^*; \lambda^*) z > 0.$$

I can now show that

$$J^* = \begin{bmatrix} \nabla^2 \ell(x^*; \lambda^*) & -\nabla h(x^*) \\ \Lambda^* \nabla h(x^*)^T & H(x^*) \end{bmatrix}$$

is nonsingular. Suppose $(z, w) \in \mathrm{R}^n \times \mathrm{R}^p$ satisfies $J^*(z, w) = 0$. Then

$$\begin{aligned} \nabla^2 \ell(x^*; \lambda^*) z - \nabla h(x^*) w &= 0, && (6) \\ \Lambda^* \nabla h(x^*)^T z + H(x^*) w &= 0. && (7) \end{aligned}$$

The second equation yields

$$\lambda_i^* \nabla h_i(x^*) \cdot z + h_i(x^*) w_i = 0, \ i = 1, 2, \ldots, p. \tag{8}$$

If $h_i(x^*) = 0$ (that is, if $i \in \mathcal{A}(x^*)$), then (8) shows that

$$\lambda_i^* \nabla h_i(x^*) \cdot z = 0.$$

By strict complementarity, $h_i(x^*) = 0$ implies that $\lambda_i^* > 0$, and so

$$i \in \mathcal{A}(x^*) \ \Rightarrow \ \nabla h_i(x^*) \cdot z = 0. \tag{9}$$

On the other hand, $h_i(x^*) > 0$ implies that $\lambda_i^* = 0$, and so (8) yields

$$h_i(x^*) w_i = 0$$

and hence

$$i \notin \mathcal{A}(x^*) \ \Rightarrow w_i = 0.$$

Taking the dot product of (6) with z, I obtain

$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z - z \cdot \nabla h(x^*) w = 0. \tag{10}$$

But

$$z \cdot \nabla h(x^*) w = (\nabla h(x^*)^T z) \cdot w = \sum_{i=1}^{p} w_i \nabla h_i(x^*) \cdot z = 0,$$

since, for each i, either $w_i = 0$ or $\nabla h_i(x^*) \cdot z = 0$. Therefore, (10) becomes

$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z = 0,$$

which, since z satisfies (9), shows that $z = 0$. Finally, since $w_i = 0$ for $i \neq \mathcal{A}(x^*)$ and $z = 0$, Equation (6) yields

$$\nabla h(x^*) w = \sum_{i \in \mathcal{A}(x^*)} w_i \nabla h_i(x^*) = 0. \tag{11}$$

Since x^* is a regular point of the NLP,

$$\{\nabla h_i(x^*) \, : \, i \in \mathcal{A}(x^*)\}$$

is linearly independent, and so (11) shows that $w = 0$.

I have therefore showed that $J^*(z, w) = 0$ holds only for $(z, w) = 0$, which proves that J^* is nonsingular. The implicit function theorem then shows that there exist $\hat{\mu} > 0$ and continuously differentiable functions $x : [0, \hat{\mu}] \to \mathrm{R}^n$ and $\lambda : [0, \hat{\mu}] \to \mathrm{R}^p$ such that

$$x(0) = x^*, \ \lambda(0) = \lambda^*$$

and

$$\begin{aligned} \nabla f(x(\mu)) - \nabla h(x(\mu)) \lambda(\mu) &= 0, \\ \lambda(\mu) h(x(\mu)) &= \mu e \end{aligned}$$

for all $\mu \in [0, \hat{\mu}]$. Moreover, $x(\mu)$ and $\lambda(\mu)$ form the unique solution to this system in a neighborhood of (x^*, λ^*).

I still need to show that $\lambda(\mu) > 0$ and $h(x(\mu)) > 0$ for $\mu \in [0, \hat{\mu}]$ (so that $x(\mu)$ is a strictly feasible point and hence a candidate for a minimizer of the logarithmic barrier function). But this follows from the continuity of x and λ and from the strict complementarity of x^*, λ^*. The equation

$$\lambda_i(\mu) h_i(x(\mu)) = \mu > 0$$

shows that $\lambda_i(\mu)$ and $h_i(x(\mu))$ are either both positive or both negative. For $i \in \mathcal{A}(x^*)$, λ_i^* is positive and hence, by continuity, so is $\lambda(\mu)$ for all μ sufficiently small. On the other hand, for $i \notin \mathcal{A}(x^*)$, $h_i(x^*)$ is positive and therefore so is $h_i(x(\mu))$ for all μ sufficiently small. Thus, in either case, $\lambda_i(\mu)$ and $h_i(x(\mu))$ are both positive for all μ sufficiently small.

Finally, I have shown that $x(\mu)$ is a stationary point of $B(\cdot;\mu)$ for all μ sufficiently small. It remains to prove that $\nabla^2 B(x(\mu);\mu)$ is poisitive definite and hence that $x(\mu)$ is a local minimizer of $B(\cdot;\mu)$ for all μ sufficiently small. I can write the Hessian in the following way:

$$\begin{aligned}\nabla^2 B(x(\mu);\mu) \quad &= \quad \nabla^2 \ell(x(\mu); \lambda(\mu) + \sum_{i \notin \mathcal{A}(x^*)} \frac{\lambda_i(\mu)}{h_i(x(\mu))} \nabla h(x(\mu)) \nabla h(x(\mu))^T + \\ &\sum_{i \in \mathcal{A}(x^*)} \frac{\lambda_i(\mu)}{h_i(x(\mu))} \nabla h(x(\mu)) \nabla h(x(\mu))^T.\end{aligned}$$

For all μ sufficiently small, the first two terms are positive definite on all z satisfying $\nabla h_i(x^*) \cdot z = 0$ for all $i \in \mathcal{A}(x^*)$, while the third term vanishes on all such z. On the other hand, the first term is bounded, the second term nonnegative, and the third term large and positive when acting on any z for which $\nabla h_i(x^*) \cdot z \neq 0$ for some $i \in \mathcal{A}(x^*)$ (provided μ is small). Using these ideas, one can construct a rigorous proof that $\nabla^2 B(x(\mu);\mu)$ is positive definite for all μ sufficiently small. The details are left to the reader.

This proves the following theorem:

Theorem 2.1 *Suppose* $f : \mathrm{R}^n \to \mathrm{R}$ *and* $h : \mathrm{R}^n \to \mathrm{R}^p$ *are twice continuously differentiable and* x^* *is a local minimizer and nonsingular point of*

$$\begin{array}{ll} \min & f(x) \\ s.t. & h(x) \geq 0. \end{array}$$

Then there exist $\hat{\mu} > 0$*, a neighborhood* N *of* x^* *and a continuously differentiable function* $x : [0, \hat{\mu}] \to N$ *such that*

1. *for all* $\mu \in [0, \hat{\mu}]$, $x(\mu)$ *is a local minimizer of the logarithmic barrier function* $B(\cdot;\mu)$*; and*
2. $x(\mu) \to x^*$ *as* $\mu \to 0$*;*
3. $\lambda(\mu) = \mu h(x(\mu))^{-1} \to \lambda^*$ *as* $\mu \to 0$*, where* λ^* *is the Lagrange multiplier corresponding to* x^*.

3 Rate of convergence

Earlier I presented an example in which I applied the logarithmic barrier method to solve a simple inequality-constrained nonlinear program. In that example, $\|x(\mu) - x^*\| = O(\mu)$ was observed to hold. The following theorem shows that this is typical.

Theorem 3.1 *Suppose* $f, h, x^*, \lambda^*, x : [0, \hat{\mu}] \to \mathrm{R}^n, \lambda : [0, \hat{\mu}] \to \mathrm{R}^p$ *are all as in the previous theorem. Then*

$$\|x(\mu) - x^*\| = O(\mu) \text{ and } \|\lambda(\mu) - \lambda^*\| = O(\mu) \text{ as } \mu \to 0.$$

Proof: Since both x and λ are continuously differentiable and $x(0) = x^*$, $\lambda(0) = \lambda^*$, I can write

$$\begin{aligned} x(\mu) &= x^* + \int_0^\mu \dot{x}(\eta)\, d\eta, \\ \lambda(\mu) &= \lambda^* + \int_0^\mu \dot{\lambda}(\eta)\, d\eta. \end{aligned}$$

Assuming that, for some constant M,

$$\|\dot{x}(\mu)\| \leq M, \ \|\dot{\lambda}(\mu)\| \leq M \tag{12}$$

for all μ sufficiently small, then

$$\|x(\mu) - x^*\| \leq \int_0^\mu \|\dot{x}(\eta)\|\, d\eta \leq \int_0^\mu M\, d\eta = \mu M,$$

and similarly for λ. So it suffices to prove (12).

The functions x and λ satisfy the equations

$$\begin{aligned} \nabla f(x(\mu)) - \nabla h(x(\mu))\lambda(\mu) &= 0, \\ \lambda(\mu) h(x(\mu)) &= \mu e. \end{aligned}$$

Differentiating with respect to μ yields the following system of equations for $\dot{x}(\mu)$, $\dot{\lambda}(\mu)$:

$$\begin{aligned} \nabla^2 \ell(x(\mu); \lambda(\mu))\dot{x}(\mu) - \nabla h(x(\mu))\dot{\lambda}(\mu) &= 0, \\ \Lambda(\mu)\nabla h(x(\mu))^T \dot{x}(\mu) + H(x(\mu))\dot{\lambda}(\mu) &= e. \end{aligned}$$

In matrix-vector form, the system is

$$\begin{bmatrix} \nabla^2 \ell(x(\mu); \lambda(\mu)) & -\nabla h(x(\mu)) \\ \Lambda(\mu)\nabla h(x(\mu))^T & H(x(\mu)) \end{bmatrix} \begin{bmatrix} \dot{x}(\mu) \\ \dot{\lambda}(\mu) \end{bmatrix} = \begin{bmatrix} 0 \\ e \end{bmatrix},$$

and thus

$$\begin{bmatrix} \dot{x}(\mu) \\ \dot{\lambda}(\mu) \end{bmatrix} = \begin{bmatrix} \nabla^2 \ell(x(\mu); \lambda(\mu)) & -\nabla h(x(\mu)) \\ \Lambda(\mu)\nabla h(x(\mu))^T & H(x(\mu)) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ e \end{bmatrix}.$$

Since

$$\begin{bmatrix} \nabla^2 \ell(x(\mu); \lambda(\mu)) & -\nabla h(x(\mu)) \\ \Lambda(\mu)\nabla h(x(\mu))^T & H(x(\mu)) \end{bmatrix}$$

converges to the nonsingular matrix

$$\begin{bmatrix} \nabla^2 \ell(x^*; \lambda^*) & -\nabla h(x^*) \\ \Lambda^*\nabla h(x^*)^T & H(x^*) \end{bmatrix}$$

as $\mu \to 0$, it follows that

$$\begin{bmatrix} \nabla^2 \ell(x(\mu); \lambda(\mu)) & -\nabla h(x(\mu)) \\ \Lambda(\mu)\nabla h(x(\mu))^T & H(x(\mu)) \end{bmatrix}^{-1}$$

is uniformly bounded in norm, say by $\tilde{M}$. This gives

$$\left\| \begin{bmatrix} \dot{x}(\mu) \\ \dot{\lambda}(\mu) \end{bmatrix} \right\| = \left\| \begin{bmatrix} \nabla^2 \ell(x(\mu); \lambda(\mu)) & -\nabla h(x(\mu)) \\ \Lambda(\mu)\nabla h(x(\mu))^T & H(x(\mu)) \end{bmatrix}^{-1} \right\| \left\| \begin{bmatrix} 0 \\ e \end{bmatrix} \right\| \leq \tilde{M} \left\| \begin{bmatrix} 0 \\ e \end{bmatrix} \right\| = M.$$

This gives the desired bounds on $\|\dot{x}(\mu)\|$ and $\|\dot{\lambda}(\mu)\|$. QED

4 Advantages and disadvantages of the logarithmic barrier method

The advantages and disadvantages of the logarithmic barrier method are very similar to those of the quadratic penalty method. On the one hand, the method is fairly simple to implement, assuming that a good code for unconstrained minimization is available. On the other hand, assuming the barrier parameters are chosen as $\mu_k = \beta^k \mu_0$, the convergence of $\{x^{(k)}\}$ to x^* is only linear. Moreover, it is easy to show that the problem of minimizing $B(\cdot;\mu)$ becomes arbitrarily ill-conditioned as $\mu \to 0$.