TMA 4180 Optimeringsteori

Penalty and Barrier Methods - A Summary

Harald E. Krogstad, 2008/rev. 2010

The philosophy of *penalty* and *barrier* methods is simple. In the penalty methods you give a "fine" for violating the constraints, and obtain approximate solutions to your original problem by balancing the objective function and a penalty term involving the constraints. By increasing the penalty, the approximate solution is forced to approach the feasible domain, and hopefully, the solution of the original constrained problem.

A typical way of formulating a *penalty method* is to augment the objective function f with certain penalty terms, e.g.,

$$Q(x,\mu) = f(x) + \frac{1}{\mu}\sum_{i\in\mathcal{E}} |c_i(x)| + \frac{1}{\mu}\sum_{i\in\mathcal{I}} \max[0, -c_i(x)]. \tag{1}$$

Here, $|c_i(x)| = 0$ only when $c_i(x) = 0$, whereas $\max[0, -c_i(x)] = 0$ whenever $c_i(x) \geq 0$. In practice, one could consider using different μ-s for the two sets, and everything should be formulated in a dimensionally consistent way. Minimizing $Q(x,\mu)$ is now essentially an unconstrained problem, and the idea is that the corresponding solutions, $x^*(\mu)$, approach the solution of the original problem in the limit when μ tends to 0,

$$x^* = \lim_{\mu\to 0} x^*(\mu).$$

For the *barrier methods*, you move around in the interior of the feasible domain, and every time you try to approach the boundary, you feel a repulsive force. The force makes you stop close to the exact solution, if this happens to be at the boundary. As we know, this is typically the case for constrained problems. By weakening the barrier gradually, we obtain approximate solutions which hopefully converge to the exact solution. A logarithmic barrier is quite popular for inequality constraints,

$$Q(x,\mu) = f(x) - \mu\sum_{i\in\mathcal{I}} \log c_i(x). \tag{2}$$

In the barrier case, it is of course necessary that Ω has a non-empty interior, and that it is at all possible to reach a boundary point solution from the interior of Ω (can you imagine a set where this will not be possible?).

Below we shall assume that both f and the c_i-s are as smooth as needed in the proofs.

All this sounds reasonable and simple, but unfortunately, the numerical problems we encounter are not quite straightforward. This is all experienced in the special case of *quadratic penalty methods*, which we therefore consider next. Most of the theory is found in N&W, Sec. 17.1.

This note covers our curriculum for Penalty and Barrier Methods.

1 The Quadratic Penalty Method

In the *Quadratic Penalty Method* the constraint penalties are all entered in terms of quadratic functions. This may in many cases result in Least Square problems, for which there are very efficient numerical algorithms. We shall for simplicity assume only *equality* constraints, so we consider

$$\begin{aligned} &\min_x f(x), \\ &c_i(x) = 0, \ i \in \mathcal{E}. \end{aligned} \tag{3}$$

The penalty now has the form

$$\frac{1}{\mu} \sum_{i \in \mathcal{E}} c_i^2(x), \ \mu > 0, \tag{4}$$

and the unconstrained penalized problem is therefore

$$\min_x Q(x, \mu) = \min_x \left\{ f(x) + \frac{1}{\mu} \sum_{i \in \mathcal{E}} c_i^2(x) \right\}. \tag{5}$$

Let us introduce the vector of constraints,

$$c(x) = (c_1(x), c_2(x), \cdots, c_r(x))', \tag{6}$$

and the *Jacobian matrix* for the constraints,

$$A = A(x) = \begin{bmatrix} \nabla c_1(x) \\ \nabla c_2(x) \\ \vdots \\ \nabla c_r(x) \end{bmatrix}. \tag{7}$$

We recall that the LICQ condition holds at x if $A(x)$ has full row rank.

The general idea is to solve

$$x_n = \arg\min_x Q(x, \mu_n) \tag{8}$$

for a sequence of μ_n-s where $\mu_n \to 0$, and hope that at least a subsequence, $\{x_{n_k}\}$, converges to a solution $x^*$of the original problem.

The first theorem considers the case where the unconstrained penalized problems in (5) are solved exactly for each μ.

Theorem (N&W 17.1): *Let $\{x_n\}$ be global minima of Eqn. 5 for a sequence of μ_n-s converging to* 0, *and such that $x_n \longrightarrow \bar{x}$. Then $\bar{x}$ is a global minimum of the problem in (3).*

Proof: Assume that x^* is a global minimum of problem (3). Since x_n are global minima of the penalized problems we must have

$$f(x_n) + \frac{1}{\mu_n} \sum_{i \in \mathcal{E}} c_i^2(x_n) \le f(x^*) + \frac{1}{\mu_n} \sum_{i \in \mathcal{E}} c_i^2(x^*) = f(x^*) \tag{9}$$

($x^* \in \Omega$, so that $c_i(x^*) = 0$!). Hence,

$$0 \leq \sum_{i\in\mathcal{E}} c_i^2(x_n) \leq \mu_n \left[f(x^*) - f(x_n)\right]. \tag{10}$$

Since $f(x^*) - f(x_n) \to f(x^*) - f(\bar{x})$ and $\mu_n \to 0$,

$$\sum_{i\in\mathcal{E}} c_i^2(\bar{x}) = \lim_{n\to\infty} \sum_{i\in\mathcal{E}} c_i^2(x_n) = 0. \tag{11}$$

Thus, for all $i \in \mathcal{E}$, $c_i(\bar{x}) = 0$, and $\bar{x}$ *is* feasible. Moreover, Eqn. 9 also shows that

$$0 \leq \frac{1}{\mu_n} \sum_{i\in\mathcal{E}} c_i^2(x_n) \leq f(x^*) - f(x_n), \tag{12}$$

and in the limit,

$$0 \leq f(x^*) - f(\bar{x}). \tag{13}$$

But x^* is already the global minimum of $f(x)$ on Ω, and hence $f(x^*) = f(\bar{x})$. Thus, $\bar{x}$ is feasible and a global minimum to problem (3) as well.

The next theorem deals with the more practical situation where the penalized problems are solved only approximately. In this case, the approximate solutions will not necessary converge to a minimum, but under favorable conditions, to a KKT-point:

Theorem (N&W 17.2): *Assume that $\{x_n\}$ is a sequence of approximate solutions of problem (5) where*

$$\begin{aligned} \left|\nabla_x Q(x_n, \mu_n)\right| &\leq \tau_n, \\ \tau_n, \mu_n &\to 0. \end{aligned} \tag{14}$$

Assume that $x_n \to x^*$ and that the LICQ holds at x^*. Then x^* will be a KKT-point and

$$\lambda_i^* = \lim_{n\to\infty} \frac{-2c_i(x_n)}{\mu_n},\ i \in \mathcal{E}. \tag{15}$$

Proof: It is not really necessary that $\{x_n\}$ converges to x^*. As long as $\{x_n\}$ is bounded, there exists convergent subsequences which could be used instead.

We first of all observe that

$$\nabla\left(\sum_{i\in\mathcal{E}} c_i^2(x)\right) = 2\sum_{i\in\mathcal{E}} c_i(x)\nabla c_i(x). \tag{16}$$

According to the assumptions,

$$\left|\nabla_x Q(x_n, \mu_n)\right| = \left|\nabla f(x_n) + \frac{2}{\mu_n}\sum_{i\in\mathcal{E}} c_i(x_n)\nabla c_i(x_n)\right| \leq \tau_n. \tag{17}$$

Since $\nabla f(x_n) \to \nabla f(x^*)$ and thus stays finite and $\mu_n \to 0$, we *must* have that

$$\lim_{n\to\infty} \sum_{i\in\mathcal{E}} c_i(x_n)\nabla c_i(x_n) = 0. \tag{18}$$

But this implies that

$$\sum_{i\in\mathcal{E}} c_i(x^*)\nabla c_i(x^*) = c(x^*)' A(x^*) = 0, \tag{19}$$

and, since the LICQ was supposed to holds at x^*, that $c(x^*) = 0$ (clever argument!). Thus, $x^* \in \Omega$. It is now reasonable to introduce the vector

$$\lambda_n \triangleq \frac{-2c_i(x_n)}{\mu_n} \tag{20}$$

for each n. Then

$$\nabla_x Q(x_n, \mu_n) = \nabla f(x_n) - \lambda_n' A(x_n). \tag{21}$$

Since the LICQ condition holds at x^*, it also holds at neighboring points (we assume that $A(x)$ is a continuous matrix of x). For such x_n-s we have

$$A(x_n) A(x_n)' \lambda_n = A(x_n)(\nabla f(x_n) - \nabla_x Q(x_n, \mu_n))', \tag{22}$$

and

$$\begin{aligned} \lambda^* &= \lim_{n\to\infty} \lambda_n \\ &= \lim_{n\to\infty} \left[A(x_n) A(x_n)'\right]^{-1} \left(A(x_n)(\nabla f(x_n) - \nabla_x Q(x_n, \mu_n))'\right) \\ &= \left[A(x^*) A(x^*)'\right]^{-1} A(x^*) \nabla f(x^*)'. \end{aligned} \tag{23}$$

In conclusion,

$$\nabla f(x^*) = \lambda^{*\prime} A(x^*), \tag{24}$$

which, together with $x^* \in \Omega$ and $\lambda^{*\prime} c(x^*) = 0$, shows that x^* is a KKT-point.

2 Numerical Aspects

In the methods above, we solve unconstrained problems for each μ, and the convergence behavior is therefore connected to the behavior of the Hessian of the objective function near to the solution. Recalling how we computed the Hessian for Least Square problems, it is easy to derive that

$$\begin{aligned} \nabla^2_{xx} Q(x, \mu) &= \nabla^2_{xx}\left(f(x) + \frac{1}{\mu}\sum_{i\in\mathcal{E}} c_i^2(x)\right) \\ &= \nabla^2 f(x) + \frac{2}{\mu}\sum_{i\in\mathcal{E}} c_i(x)\nabla^2 c_i(x) + \frac{2}{\mu} A'(x) A(x). \end{aligned} \tag{25}$$

Close to the solution we therefore have

$$\begin{aligned} \nabla^2_{xx} Q(x, \mu) &\approx \nabla^2 f(x) - \sum_{i\in\mathcal{E}} \lambda_i^* \nabla^2 c_i(x) + \frac{2}{\mu} A'(x) A(x) \\ &= \nabla^2_{xx} \mathcal{L}(x, \lambda^*) + \frac{2}{\mu} A'(x) A(x). \end{aligned} \tag{26}$$

The Hessian for Q around the solution is therefore a sum of a slowly varying matrix, $\nabla^2_{xx}\mathcal{L}(x,\lambda^*)$, and a positive semi-definite matrix, $\frac{2}{\mu}A'(x)A(x)$, of order μ^{-1}. If $\nabla^2_{xx}\mathcal{L}(x,\lambda^*) > 0$, also $\nabla^2_{xx}Q(x,\mu) > 0$, and the problem has a unique minimum,– if it has solutions at all. For algorithms like the Steepest Descent or the Conjugate Gradient (CG) method, the speed of convergence is governed by the eigenvalue distribution of the Hessian, and there is a small lemma which describes this in a very neat way:

Key Lemma: *Consider a matrix of the form*

$$M = G + \frac{1}{\mu}A'A \tag{27}$$

where $G > 0$, A has full row rank r, and Z is a matrix of basisvectors for $\mathcal{N}(A)$. Then the eigenvalues of M split into two groups when $\mu \to 0$. The first group behaves asymptotically as

$$\left\{\frac{\lambda_1}{\mu}, \frac{\lambda_2}{\mu}, \cdots, \frac{\lambda_r}{\mu}\right\}, \tag{28}$$

where $\{\lambda_1, \lambda_2, \cdots, \lambda_r\}$ are the non-zero eigenvalues of $A'A$. The remaining $n-r$ eigenvalues converge to the eigenvalues of $Z'GZ$.

Proof: The proof for the first group is simple: Consider the matrix $\mu M = A'A + \mu G$, which asymptotically has $\{\lambda_1, \lambda_2, \cdots, \lambda_r\}$ as non-zero eigenvalues, since $\mu G \to 0$.

For the other group, extend Z to an orthogonal matrix $P = [Z\ V]$ such that

$$P'A'AP = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda \end{bmatrix}. \tag{29}$$

Then

$$P'MP = P'GP + \begin{bmatrix} 0 & 0 \\ 0 & \Lambda/\mu \end{bmatrix} = \begin{bmatrix} \tilde{G}_{11} & \tilde{G}_{12} \\ \tilde{G}_{21} & \tilde{G}_{22} + \Lambda/\mu \end{bmatrix}. \tag{30}$$

Sublemma: *Let*

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}(\mu) \end{bmatrix} > 0 \tag{31}$$

and assume that $B_{22}(\mu) = \mathcal{O}(1/\mu)$ when $\mu \to 0$. Then

$$\lim_{\mu\to 0} B^{-1}(\mu) = \begin{bmatrix} B_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{32}$$

Proof: Left to you! (First recall that both B_{11} and $B_{22}(\mu)$ are positive definite, since B is positive definite. Then consider the equations for $C = B^{-1}$ in a partitioned form, either by solving the matrix equations,

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & I_2 \end{bmatrix}, \tag{33}$$

or looking up the result in a linear algebra textbook).

Once the sublemma is proved, the main lemma follows by observing that $\tilde{G}_{11} = Z'GZ$.

Corollary: *The condition number of* $\nabla^2_{xx} Q(x,\mu)$ *is* $\mathcal{O}(1/\mu)$ *when* $\mu \to 0$.

Example: The *Rosenbrock Banana function* is often used as an example of a notorious difficult case for unconstrained algorithms. The function may be considered as solving the following trivial problem by means of a Quadratic Penalty Method:

$$\begin{aligned} &\min_{x\in\mathbb{R}^2} (1-x_1)^2, \\ &x_2 - x_1^2 = 0. \end{aligned} \tag{34}$$

Thus,

$$Q(x,\mu) = (1-x_1)^2 + \frac{1}{\mu}\left(x_2 - x_1^2\right)^2, \tag{35}$$

Recall that we experienced the minimum of $Q(x,\mu)$ to be virtually unreachable with the Steepest Descent method already for $\mu = 0.01$.

It is worth noting that the CG method will be excellent for the above problem if the number of constraints (r) is small. This is actually discussed (without any reference to the later penalty methods) in N&W following Theorem 5.5 on p. 115–117, and is related to the optimal interpolation property of the CG. According to Luenberger, the way to operate the CG method is then to make a *restart*, with a Steepest Descent step, every r-th iteration (However, according to some limited numerical experiments carried out using Matlab, the restart does not seem to be really worth the effort).

3 The Barrier Methods

The theory for *Barrier Methods* resembles in many ways the theory above. As noted in N&W, the most popular barrier is the logarithm:

$$B(x) = -\sum_{i\in\mathcal{I}} \log\left(c_i(x)\right),\ x \in \Omega^i, \tag{36}$$

where Ω^i denotes the *interior of* Ω. The term interior may be understood in a *relative* sense, e.g. when the search is limited to a simple set such as a hyperplane.

The logarithmic barrier function was, according to Fletcher, first introduced by Ragnar Frisch in 1955 (First Norwegian Nobel Prize winner in Economics).

As mentioned above, barrier methods require that Ω has a non-empty interior and that *all points on the boundary can be reached from the interior* (as limits of sequences consisting of interior points). This last condition is not explicitly mentioned in N&W, but is of course essential. Sets with this property are (reasonably enough) called *robust* in Luenberger.

In many practical cases, $-c_i(x)$ will be *smooth* convex functions, and then Ω will be convex. Also $B(x)$ will be convex for $x \in \Omega^i$ in this case. This follows easily by computing the

Hessian of B:

$$\nabla^2 B(x) = \nabla^2\left(-\sum_{i\in\mathcal{I}} \log\left(c_i(x)\right)\right)$$
$$= -\sum_{i\in\mathcal{I}} \frac{1}{c_i(x)}\nabla^2 c_i(x) + \sum_{i\in\mathcal{I}} \frac{1}{c_i(x)^2}\nabla c_i(x)' \nabla c_i(x),\ \ x\in\Omega^i. \tag{37}$$

Since $c_i(x) > 0$, and both $\nabla^2(-c_i(x))$ and $\nabla c_i(x)\nabla c_i(x)'$ are positive semidefinite, so is $\nabla^2 B(x)$.

If, in addition, f is convex, then

$$Q(x,\mu) = f(x) + \mu B(x),\ \ \mu > 0, \tag{38}$$

will be convex as well. If Ω is bounded, the problem

$$\min_x Q(x,\mu) \tag{39}$$

will then have a convex set of global minimizers.

At a minimum x_μ,

$$0 = \nabla Q(x,\mu) = \nabla\left(f(x) + \mu B(x)\right)$$
$$= \nabla f(x_\mu) + \sum_{i\in\mathcal{I}} \frac{-\mu}{c_i(x_\mu)}\nabla c_i(x_\mu), \tag{40}$$

and it is also here reasonable to introduce

$$\lambda_i(\mu) \triangleq \frac{\mu}{c_i(x_\mu)}. \tag{41}$$

If now $x_{\mu_n} \underset{n\to\infty}{\longrightarrow} x^*$ and the LICQ holds for the active constraints at x^*, we have, exactly as above, that for $i\in\mathcal{A}(x^*)$,

$$\lim_{n\to\infty} \lambda_i(\mu_n) = \lambda_i^* = \left[A(x^*)A(x^*)'\right]^{-1} A(x^*)\nabla f(x^*)' \tag{42}$$

(Note that $A(x^*)$ *only* contains the gradients of the active constraints). Being at the same time a limit of the non-negative numbers $\mu_n/c_i(x_{\mu_n})$, the Lagrange multipliers will be non-negative, and also equal to 0 for the non-active constraints at x^*(since $c(x^*) > 0$).

In conclusion, we have shown that x^* is a KKT-point for the original problem.

The Hessian, $\nabla^2_{xx}Q(x,\mu)$, has a similar structure as above. Near x^* we have in particular

$$\nabla^2_{xx}Q(x,\mu) = \nabla^2 f(x) - \sum_{i\in\mathcal{I}} \frac{\mu}{c_i(x)}\nabla^2 c_i(x) + \sum_{i\in\mathcal{I}} \frac{\mu}{c_i(x)^2}\nabla c_i(x)'\nabla c_i(x)$$
$$\approx \nabla^2_{xx}\mathcal{L}(x^*,\lambda^*) + \sum_{i\in\mathcal{A}(x^*)} \frac{(\lambda_i^*)^2}{\mu}\nabla c_i(x^*)'\nabla c_i(x^*). \tag{43}$$

Again, the Hessian has eigenvalues splitting in two groups, but now the group tending to infinity when $\mu\to 0$ only corresponds to the *active* constraints at x^*.

4 The Logarithmic Barrier Method for the LP-problem

The material in this section is not mentioned in N&W. Nevertheless, it forms the basis of what is probably among the fastest interior methods for large scale LP-problems that are currently available. It is different from the interior point methods treated in Chapter 14. My main reference is Gonzaga (1992), not mentioned in N&W.

The positive cone,

$$\mathbb{R}_+^n = \{x \ ; \ x \geq 0\} \tag{44}$$

is a simple robust set, and the main idea is trying to solve the standard LP-problem

$$\begin{aligned} &\min_x c'x, \\ &Ax = b, \\ &x \geq 0, \end{aligned} \tag{45}$$

by introducing a logarithmic barrier for the cone,

$$B(x) = -\sum_{i=1}^{n} \log(x_i), \ x > 0. \tag{46}$$

We observe that

$$\begin{aligned} \nabla B(x)' &= -\operatorname{vect}\left\{\frac{1}{x_i}\right\}, \\ \nabla^2 B(x) &= \operatorname{diag}\left\{\frac{1}{x_i^2}\right\}. \end{aligned} \tag{47}$$

The barrier problem is thus

$$\begin{aligned} \min_x Q(x,\mu) &= \min_x \{c'x + \mu B(x)\}, \\ Ax &= b \text{ (and } x > 0). \end{aligned} \tag{48}$$

(Recall that A is assumed to have full row rank r). It is clear that the problem is convex, and assuming that solutions exist, the solution is in fact unique, since Q is *strictly* convex. This set of solutions $\{x(\mu)\}_{\mu>0}$ is called the *Central Path* (The same term is also used for various other situations in N&W). The central path turns out to have a surprising property, but let us first recall the *Dual Problem* to the standard form:

$$\begin{aligned} &\max_\pi b'\pi, \\ &A'\pi + s = c, \\ &s \geq 0, \end{aligned} \tag{49}$$

and the essence of the *Duality Theorem*:

$$b'\pi \leq \max_\pi b'\pi = \min_x c'x \leq c'x. \tag{50}$$

Theorem: *Let $x(\mu)$ be the solution of problem (48). Then*

$$s(\mu) = \mu \operatorname{vect}\left\{\frac{1}{x_i(\mu)}\right\} \tag{51}$$

and

$$\pi(\mu) = (AA')^{-1} A (c - s(\mu)) \tag{52}$$

are feasible vectors for the dual problem, and

$$x(\mu)' s(\mu) = c' x(\mu) - b' \pi(\mu) = \mu n. \tag{53}$$

This theorem gives us a bound on how far we are from the optimal objective value and shows in particular that $c'x(\mu)$ will *always* approach the optimal value when $\mu \to 0$.

The proof is not particularly difficult. Let P be the projection operator on $\mathcal{N}(A)$. Suppressing the dependence on μ, it is always possible to write

$$c - s = P(c - s) + A'\pi \tag{54}$$

for some $\pi \in \mathbb{R}^r$. Thus, s is feasible if and only if

$$\begin{aligned} s &\geq 0, \\ P(c - s) &= 0. \end{aligned} \tag{55}$$

Since $x(\mu)$ *is* a minimum,

$$\nabla Q(x, \mu) d \geq 0 \tag{56}$$

for all feasible directions, which in this case are simply the non-zero vectors in $\mathcal{N}(A)$. If d is feasible, so is $-d$ as well, and therefore,

$$\nabla Q(x, \mu) d = 0 \text{ for all } d \in \mathcal{N}(A). \tag{57}$$

Thus, $\nabla Q(x, \mu)$ is orthogonal to $\mathcal{N}(A)$ and

$$\begin{aligned} 0 &= P\left(\nabla Q(x, \mu)'\right) \\ &= P\left(c - \mu \operatorname{vect}\left\{\frac{1}{x_i(\mu)}\right\}\right) \\ &= P(c - s). \end{aligned} \tag{58}$$

Moreover, $s > 0$ by the definition in the theorem.

Since we have verified that $P(c - s) = 0$, we also have from Eqn. 54 that

$$A'\pi = c - s, \tag{59}$$

which may be solved for π, as given in the theorem.

Finally,

$$\begin{aligned} c'x(\mu) - b'\pi &= x(\mu)' c - x(\mu)' A'\pi \\ &= x(\mu)' s \\ &= \sum_{i=1}^{n} x_i(\mu) \mu \frac{1}{x_i(\mu)} = n\mu. \end{aligned} \tag{60}$$

By staying on the central path

- we easily find a feasible dual solution
- we have an exact expression for the duality gap, that is, how far we are from the optimal objective value.

How do we stay on the central path? Actually, there are several approaches, but first of all we need to find a feasible start vector, that is, an *interior feasible point.* The following neat trick is somewhat similar to Phase 1 for the SIMPLEX method.

Consider the standard form in Eqn. 45. Define

$$\begin{aligned} a &= b - Ae, \\ e &= (1, 1, \cdots, 1)', \end{aligned} \tag{61}$$

and introduce an additional component x_{n+1} to x. Then solve the *extended problem*

$$\min \left[c' \ 1\right] \begin{bmatrix} x \\ x_{n+1} \end{bmatrix} = \min \left\{c'x + x_{n+1}\right\} \tag{62}$$

with the constraints

$$\begin{aligned} \begin{bmatrix} A & a \end{bmatrix} \begin{bmatrix} x \\ x_{n+1} \end{bmatrix} &= b, \\ x \geq 0 \ , \ x_{n+1} &\geq 0. \end{aligned} \tag{63}$$

The extended problem has the interior feasible point

$$\begin{bmatrix} e \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{64}$$

(Check it!).

Moreover, the optimal solution will be

$$\begin{bmatrix} x^* \\ 0 \end{bmatrix}, \tag{65}$$

if the original problem has a solution x^* at all.

Since both the gradient and the Hessian of $Q(x, \mu)$ are very simple, a gradient projection approach looks reasonable, and the optimal methods now appears to be the projected Newton direction followed by a line search: From x_0 we first define

$$d = -\nabla^2 Q(x_0, \mu)^{-1} \nabla Q(x_0, \mu)' \tag{66}$$

$$= \operatorname{diag}\left\{x_{0i}^2\right\} \left(\operatorname{vect}\left\{\frac{1}{x_{0i}}\right\} - \frac{c}{\mu}\right). \tag{67}$$

and then we solve

$$x_1 = \arg\min_{\alpha} Q\left(x_0 + \alpha P d, \mu\right), \tag{68}$$

where P is the projection operator on $\mathcal{N}(A)$.

Note that the projection operator has the form

$$P = I - A'\left(AA'\right)^{-1} A \tag{69}$$

and is therefore independent of μ.

There are apparently different strategies whether one should take several projected Newton iterations between each time μ is decreased, or take small steps in μ for each Newton iteration. In any case, the theorem above tells us how far we are from the solution, and these methods solve the LP problem to L significant digits in $\mathcal{O}\left(Ln^3\right)$ operations.

References

[1] C.C. Gonzaga: Path Following Methods for Linear Programming, *SIAM Review,* **34** (1992) pp. 167–224.

[2] R. Fletcher: *Practical Methods of Optimization*, *2nd Edition*, Wiley, 1996.

[3] R. Frisch: *The Logarithmic Potential Method for Convex Programming*, Report Universitetet i Oslo, 1955.

[4] D.G. Luenberger, : *Linear and Nonlinear Programming*, 2nd ed., Addison Westley, 1984