CNAc: Continuous Optimization
Solutions to problem set 6 — SQP methods

Honour School of Mathematics, Oxford University
Hilary Term 2006, Dr Nick Gould

Problem 1.

(a) Since $\nabla_x F(x_*)$ is non singular, let $\mathcal{B}$ be the set of points for which

$$\|(\nabla_x F(x))^{-1}\|_2 \leq 2\|(\nabla_x F(x_*))^{-1}\|_2. \tag{1}$$

Let γ^L be the Lipschitz constant for $\nabla_x F(x)$ over $\mathcal{B}\bigcup\{x \mid \|x - x_*\|_2 \leq 1\}$, and let

$$0 < \kappa < \min(1, 1/(\gamma^L\|(\nabla_x F(x_*))^{-1}\|))$$

be chosen sufficiently small that $\mathcal{X} = \{x \mid \|x - x_*\|_2 \leq \kappa\} \subseteq \mathcal{B}$.

Suppose that $x_k \in \mathcal{X}$. Then the next Newton iterate x_{k+1} satisfies

$$\begin{aligned} x_{k+1} - x_* = &\ x_k - x_* - (\nabla_x F(x_k))^{-1}F(x_k) = x_k - x_* - (\nabla_x F(x_k))^{-1}\left(F(x_k) - F(x_*)\right) \\ = &\ (\nabla_x F(x_k))^{-1}\left(F(x_*) - F(x_k) - (\nabla_x F(x_k)(x_* - x_k)\right). \end{aligned} \tag{2}$$

But Theorem 1.3 gives that

$$\|F(x_*) - F(x_k) - \nabla_x F(x_k)(x_* - x_k)\|_2 \leq \tfrac{1}{2}\gamma^L\|x_* - x_k\|_2^2. \tag{3}$$

Hence (1)–(3) and the Cauchy-Schwartz inequality gives

$$\|x_{k+1} - x_*\|_2 \leq \gamma^L\|(\nabla_x F(x_*))^{-1}\|_2\|x_k - x_*\|_2^2 \tag{4}$$

It then follows from the definition of $\mathcal{X}$ and (4) that $x_{k+1} \in \mathcal{X}$. Hence if $x_0 \in \mathcal{X}$, all $x_k \in \mathcal{X}$, and (4) implies that $\{x_k\}$ converges to x_* Q-quadratically.

(b) The equation has a single (repeated) root $x_* = 0$. The Newton iteration is

$$x_{k+1} = x_k - \frac{x_k^2}{2x_k} = \tfrac{1}{2}x_k,$$

and thus $\|x_{k+1} - x_*\| = \frac{1}{2}\|x_k - x_*\|$. The convergence rate is Q-linear. The Jacobian at x_* is singular since $\nabla_x = 2x_* = 0$.

Problem 2.

(a) The first-order necessary optimality conditions are that

$$\begin{pmatrix} 4[x_*]_1 - 1 \\ 4[x_*]_2 \end{pmatrix} - y_* \begin{pmatrix} 2[x_*]_1 \\ 2[x_*]_2 \end{pmatrix} = 0 \ \text{ and } \ [x_*]_1^2 + [x_*]_2^2 - 1 = 0.$$

This has two solutions $x_* = (1, 0)^T$, with $y_* = 3/2$ and $x_* = (-1, 0)^T$, with $y_* = 5/2$. For the former the Hessian of the Lagrangian is I, while for the latter it is $-I$. Thus the former is an isolated local (and actually global) minimizer, while the latter is an isolated local (and actually global) maximizer.

(b) The SQP step satisfies the equations

$$\begin{pmatrix} 1 & 0 & -2\cos\theta \\ 0 & 1 & -2\sin\theta \\ 2\cos\theta & 2\sin\theta & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ y^+ \end{pmatrix} = - \begin{pmatrix} 4\cos\theta \\ 4\sin\theta \\ 0 \end{pmatrix},$$

which has the solution $s = (\sin^2\theta, -\cos\theta\sin\theta)^T$ and $y^+ = 2 - \frac{1}{2}\cos\theta$. But then $\cos(x+s) = \sin^2\theta$ and $f(x+s) - f(x) = \sin^2\theta$ which are both positive unless $\theta = 0$.

(c) The second-order correction satisfies the equations

$$\begin{pmatrix} 1 & 0 & -2\cos\theta \\ 0 & 1 & -2\sin\theta \\ 2\cos\theta & 2\sin\theta & 0 \end{pmatrix} \begin{pmatrix} s_1^{\mathrm{C}} \\ s_2^{\mathrm{C}} \\ y^{\mathrm{C}} \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ \sin^2\theta \end{pmatrix},$$

which has the solution $s^{\mathrm{C}} = (-\frac{1}{2}\cos\theta\sin^2\theta, -\frac{1}{2}\sin^3\theta)^T$ and $y^{\mathrm{C}} = -\sin^2\theta$. In particular $\|s\|_2 = \sin\theta$ but $\|s^{\mathrm{C}}\|_2 = \frac{1}{2}\sin^2\theta$, and thus the second-order correction is small relative to the SQP step.

Problem 3.

The problem we must solve is to minimize $\|s\|_2$ subject to $As = c$. As $\|\cdot\|_2$ is not differentiable, we solve instead the equivalent differentiable problem of minimizing $f(s) = \frac{1}{2}\|s\|_2^2$ subject to the same constraints.

First-order necessary optimality conditions are that

$$\nabla_s f(s) = s = A^T y, \;\; \text{where} \;\; As = -c.$$

These are the required equations. Since the Hessian of the Lagrangian is I, second-order sufficiency conditions hold, and thus our equations provide the required solution.

Problem 4.

The problem may be rewritten as

$$\underset{s\in\mathbb{R}^n, t\in\mathbb{R}}{\text{minimize}} \;\; g_k^T s + \tfrac{1}{2}s^T B_k s + \rho t \;\; \text{subject to} \;\; \|c_k + A_k s\|_\infty \le t \;\; \text{and} \;\; \|s\|_1 \le \Delta_k$$

But $\|c_k + A_k s\|_\infty \le t$ is the same as $|[c_k + A_k s]_i| \le t$ for all i, or equivalently $-t \le [c_k + A_k s]_i \le t$ and $t \ge 0$. The trust-region constraint $\|s\|_1 \le \Delta_k$ is equivalent to the 2^n linear constraints $\sum_{i=1}^n \sigma_i s_i \le \Delta$ where $\sigma_i = \pm 1$. Thus the ℓ_∞ QP problem with an ℓ_1-norm trust region is equivalent to the quadratic program

$$\begin{array}{rl} \underset{s\in\mathbb{R}^n, t\in\mathbb{R}}{\text{minimize}} & g_k^T s + \frac{1}{2}s^T B_k s + \rho t \\ \text{subject to} & -t \le [c_k + A_k s]_i \le t, \\ & t \ge 0, \\ \text{and} & \sum_{i=1}^n \sigma_i s_i \le \Delta \;\; \text{for all combinations of} \;\; \sigma_i = \pm 1. \end{array}$$

Problem 5.

The proof is essentially the same as for Theorem 7.1. The only significant difference is that now

$$\nabla_x \Phi(x_k, \mu_k) = g(x_k) + \|c(x_k)\|_2^2 \sum_{i=1}^m a_i(x_k) c_i(x_k)/\mu_k.$$

Now simply replace every mention of $\|c(x_k)\|_2^2$ by $\|c(x_k)\|_2^4$ in the original proof.