UC Berkeley
Department of Electrical Engineering and Computer Science

EECS 227A
NONLINEAR AND CONVEX OPTIMIZATION

Solutions 7
Fall 2009

Problem 7.1

(a) The derivatives are

$$f'(x) = -1/x^2, f''(x) = 2/x^3, f'''(x) = -6/x^4$$

so the self-concordance condition is

$$\frac{6}{x^4} \leq 2\left(\frac{2}{x^3}\right)^{3/2} = \frac{4\sqrt{2}}{x^4\sqrt{x}}$$

which holds if $\sqrt{x} \leq 4\sqrt{2}/6 = \sqrt{8/9}$

(b) If we make an affine change of variables $y_i = 8(b_i - a_i^T x)/(9\alpha)$, then $y_i < 8/9$ for all $x \in \mathbf{dom} f$. The function f reduces to $\sum_{i=1}^m 1/y_i$, which is self concordant by the result in (a).

Problem 7.2

(a) We have

$$\frac{d}{dx} f''(x)^{-1/2} = (-1/2)\frac{f'''(x)}{f''(x)^{3/2}}$$

Integrating

$$\frac{d}{dx} f''(x)^{-1/2} = 1$$

gives $f(x) = -\log(x + c_0) + c_1 x + c_2$. Integrating

$$\frac{d}{dx} f''(x)^{-1/2} = -1$$

gives $f(x) = -\log(-x + c_0) + c_1 x + c_2$

(b) Suppose $f''(0) > 0, f''(\bar{x}) = 0$ for $\bar{x} > 0$, and $f''(x) > 0$ on the interval between 0 and $\bar{x}$. The inequality

$$-1 \leq \frac{d}{dx} f''(x)^{-1/2} \leq 1$$

holds for x between 0 and $\bar{x}$. Integrating gives

$$f''(\bar{x})^{-1/2} - f''(0)^{-1/2} \leq \bar{x}$$

which contradicts $f''(\bar{x}) = 0$

Problem 7.3

(a) The composition rules show that $tf_0(x) + \phi_h(x)$ is convex in x, since h is increasing and convex, and f_i are convex.

(b) The minimizer of $tf_0(x) + \phi_h(x), z = x^*(t)$, satisfies $t\nabla f_0(z) + \nabla\phi(z) = 0$. Expanding this we get

$$t\nabla f_0(z) + \sum_{i=1}^{m} h'(f_i(z))\nabla f_i(z) = 0$$

This shows that z minimizes the Lagrangian $f_0(z) + \sum_{i=1}^{m} \lambda_i f_i(z)$, for

$$\lambda_i = h'(f_i(z))/t, i = 1, \dots, m$$

The associated dual function value is

$$g(\lambda) = f_0(z) + \sum_{i=1}^{m} \lambda_i f_i(z) = f_0(z) + \sum_{i=1}^{m} h'(f_i(z)) f_i(z)/t,$$

so the duality gap is

$$(1/t)\sum_{i=1}^{m} h'(f_i(z))(-f_i(z))$$

(d) The only way the expression above does not depend on the problem data (except t and m is for $h'(u)(-u)$ to be constant. This means $h'(u) = a/(-u)$ for some constant a, so $h(u) = -a\log(-u) + b$, for some constant b. Since h must be convex and increasing, we need $a > 0$. Thus, h gives rise to a scaled, offset log barrier. In particular, the central path associated with h is the same as for the standard log barrier.

Problem 7.4

(a) Follows from the convex-concave property of f_0, convexity of $-\log(-f_i)$, and concavity of $\log(-\tilde{f}_i)$

(b) Since $(w^*(t), z^*(t))$ is a saddle-point of the function

$$tf_0(w,z) - \sum_{i=1}^{m} \log(-f_i(w)) + \sum_{i=1}^{\tilde{m}} \log(-\tilde{f}_i(z))$$

its gradient with respect to w, and also with respect to z, vanishes there:

$$\begin{aligned} t\nabla_w f_0(w^*(t), z^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(w^*(t))}\nabla f_i(w^*(t)) &= 0 \\ t\nabla_z f_0(w^*(t), z^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(z^*(t))}\nabla f_i(z^*(t)) &= 0 \end{aligned}$$

It follows that $w^*(t)$ minimizes

$$f_0(w, z^*(t)) + \sum_{i=1}^{m} \lambda_i f_i(w)$$

over w where $\lambda_i = 1/(-tf_i(w^*(t)))$, i.e. for all w, we have

$$f_0(w^*(t), z^*(t)) + \sum_{i=1}^{m} \lambda_i f_i(w^*(t)) \leq f_0(w, z^*(t)) + \sum_{i=1}^{m} \lambda_i f_i(w)$$

The LHS is equal to $f_0(w^*(t), z^*(t)) - m/t$, and for all $w \in W$, the second term on the RHS is nonpositive, so we have

$$f_0(w^*(t), z^*(t)) \leq \inf_{w \in W} f_0(w, z^*(t)) + m/t$$

A similar argument shows that

$$f_0(w^*(t), z^*(t)) \geq \sup_{z \in Z} f_0(w^*(t), z) - m/t$$

Problem 7.5

(a) We implement a barrier method to solve the SDP (11.66). The only constraint in the problem is the linear matrix inequality (LMI) $W + \mathbf{diag}(\nu) \succeq 0$, for which we will use the log barrier $-\log\det(W + \mathbf{diag}(\nu))$. To start the barrier method, we need a strictly feasible point, but this is easily found. If $\lambda_{\min}(W)$ is the smallest eigenvalue of the matrix W then $W + (-\lambda_{\min}(W) + 1)I$ has smallest eigenvalues 1, and so is positive define. Thus $\nu = (-\lambda_{\min}(W) + 1)\mathbf{1}$ is a strictly feasible starting point.

At each outer iteration, we use the Newton's method to minimize

$$f(\nu) = t\mathbf{1}^T \nu - \log\det(W + \mathbf{diag}(\nu))$$

We can start with $t = 1$, and at the end of each outer iteration increase t by a factor of μ, say $\mu = 10$ until the desired accuracy is reached. At the end of each iteration, the duality gap is exactly n/t, with the dual feasible point

$$Z = (n/t)(W + \mathbf{diag}(\nu))^{-1}$$

We will return ν and Z, at the end of the first outer iteration to satisfy $n/t \leq \epsilon$ where ϵ is the required tolerance.

Now we turn to the question of how to compute the gradient and Hessian of f. We know that for $X \in \mathbf{S}^n_{++}$, the gradient of the function $g(X) = \log\det(X)$ at X is given by $\nabla g(X) = X^{-1}$. We use chain rule, with

$$X = W + \mathrm{diag}(\nu) = W + \sum_{i=1}^{n} \nu_i E_{ii}$$

where E_{ii} is the matrix with a one in the i, i entry and zeros elsewhere, to obtain

$$\begin{aligned}\nabla f(\nu)_i &= t - \operatorname{tr}((W + \mathbf{diag}(\nu))^{-1} E_{ii}) \\ &= t - ((W + \mathbf{diag}(\nu)))^{-1}_{ii}\end{aligned}$$

for $i = 1, \ldots, n$. Thus we have the simple formula

$$\nabla f(\nu) = t\mathbf{1} - \mathbf{diag}((W + \mathbf{diag}(\nu))^{-1})$$

The second derivative of $\log\det X$, at $X \in \mathbf{S}^n_{++}$, is given by the bilinear form

$$\nabla^2 g(X)[Y, Z] = -\operatorname{tr}(X^{-1} Y X^{-1} Z)$$

Applying this to our function f yields, with $X = W + \mathbf{diag}(\nu)$,

$$\nabla^2 f(\nu)_{ij} = \operatorname{tr}(X^{-1} E_{ii} X^{-1} E_{jj}) = (X^{-1})^2_{ij}$$

for $i = 1, j = 1, \ldots, n$. Thus we have the very simple formula for the Hessian

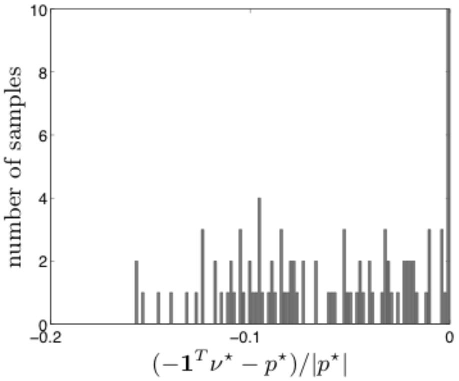
$$\nabla^2 f(\nu) = ((W + \mathbf{diag}(\nu))) \circ ((W + \mathbf{diag}(\nu)))$$

where for $U, V \in \mathbf{S}^n$, the Hadamard product of U and V, denoted by $T = U \circ V$ is defined by $T_{ij} = U_{ij} V_{ij}$

We first test the method on some small problems. We generate random symmetric matrices $W \in \mathbf{S}^{10}$, with off-diagonal elements generated from independent $\mathcal{N}(0, 1)$ distributions, and zero diagonal elements. The figure shows the distribution of the relative error

$$\frac{-\mathbf{1}^T \nu^* - p^*}{|p^*|}$$

for 100 randomly generated matrices.



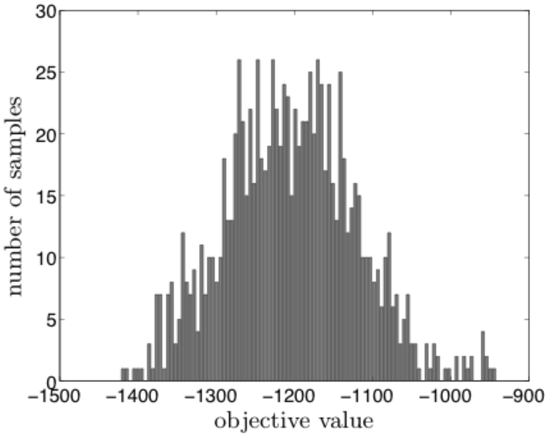
We notice that the lower bound is equal (or very close) to p^* in 10 cases, and never less than about 15% below p^*.

We also generate a larger problem instance, with $n = 100$. The optimal value of the relaxation is -1687.5. The lowerbound from the eigenvalue decomposition of W is $n\lambda_{\min}(W) = -1898.4$.

(b) We first try the heuristic on the family of 100 problems with $n = 10$. The heuristic gave the correct solution in 70 instances. For the larger problem, the heuristic gives the upper bound -1336.5. At this point, we can stay that the optimal value of the larger problem lies between -1336.5 and -1687.5.

(c) We first try this heuristic, with $K = 10$, on the family of 100 problems with $n = 10$. The heuristic gave the correct solution in 88 instances.

We plot below a histogram of the objective obtained by the randomized heuristic, over 1000 samples.



Many of these samples have an objective value larger than the one found in part (b) above, but some have a lower cost. The minimum value is -1421.7, so p^* lies between -1421.7 and -1687.5.