L. Vandenberghe June 6, 2018

ECE236B homework #9 solutions

1. *Exercise T10.14.* Eliminating x_3 using the equality constraint $x_1 + x_2 + x_3 = 1$ gives the equivalent problem

$$\begin{array}{ll} \text{maximize} & (1/3)\log(1 + x_1 + 0.3x_2) + (1/6)\log(1 + x_1 - 0.5x_2) \\ & + (1/3)\log(1 - 0.5x_1 + 0.3x_2) + (1/6)\log(1 - 0.5x_1 - 0.5x_2), \end{array}$$

with two variables x_1 and x_2. We use Newton's method with backtracking parameters $\alpha = 0.01$, $\beta = 0.5$, stopping criterion $\lambda < 10^{-8}$, and initial point $x = (0, 0, 1)$. The algorithm converges in five steps to the solution

$$x_1 = 0.4972, \qquad x_2 = 0.1994, \qquad x_3 = 0.3034.$$

No backtracking steps were needed.

```
pi = [1/3; 1/6; 1/3; 1/6];
b = ones(4,1);
A = [1, .3;  1, -0.5;  -0.5, 0.3;  -0.5, -0.5];
x = zeros(2,1);
for iters = 1:50
    y = A*x+b;
    val = -pi'*log(y);
    grad = -A'*(pi./y);
    hess = A'*diag(pi./y.^2)*A;
    v = -hess\grad;
    fprime = grad'*v;
    if (sqrt(-fprime) < 1e-8), break; end;
    t = 1;
    while (min(b+A*(x+t*v)) <= 0.0), t = t/2; end;
    while (-pi'*log(b+A*(x+t*v)) > val + 0.01*t*fprime), t = t/2; end;
    x = x+t*v;
end;
```

2. *Exercise A8.2.* Strong duality holds ($y = 0$ is strictly feasible). The optimality conditions for y are:

   (a) $Ay = 0$ and $y^T \nabla^2 f(\hat{x}) y \le 1$.

   (b) $\mu \ge 0$.

   (c) $\mu(1 - y^T \nabla^2 f(\hat{x}) y) = 0$.

   (d) $\nabla f(\hat{x}) + A^T \nu + 2\mu \nabla^2 f(\hat{x}) y = 0$.

Here μ is the multiplier for the inequality $y^T \nabla^2 f(\hat{x}) y \leq 1$ and ν is the multiplier for $Ay = 0$.

If the multiplier μ is zero, condition (d) reduces to $\nabla f(\hat{x}) + A^T \nu = 0$. This means that $\hat{x}$ is optimal, and contradicts the assumption that $\lambda(\hat{x}) \neq 0$. We can therefore assume that $\mu > 0$. Condition (b), together with $Ay = 0$ from (a), can be written as

$$\begin{bmatrix} \nabla^2 f(\hat{x}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} 2\mu y \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f(\hat{x}) \\ 0 \end{bmatrix}.$$

Since the coefficient matrix is nonsingular, this shows that $2\mu y = \Delta x$.

The value of the multiplier μ follows from

$$\mu^2 = \mu^2 (y^T \nabla^2 f(\hat{x}) y) = \frac{1}{4} (\Delta x^T \nabla^2 f(\hat{x}) \Delta x) = \frac{1}{4} \lambda(\hat{x})^2.$$

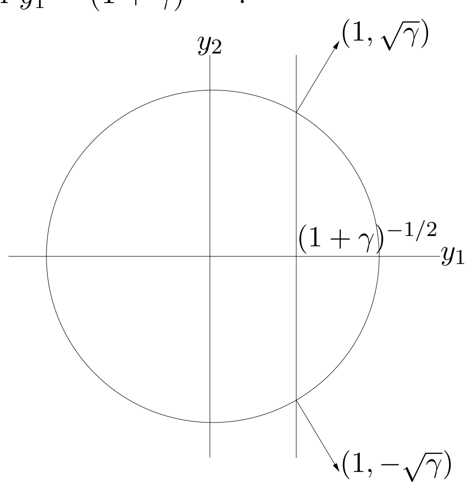
Hence $\mu = \lambda(\hat{x})/2$ and $y = (1/\lambda(\hat{x}))\Delta x$.

3. *Exercise A8.1.*

(a) We follow the hint and examine the optimization problem

$$\begin{array}{ll} \text{minimize} & x_1 y_1 + \sqrt{\gamma} x_2 y_2 \\ \text{subject to} & y_1^2 + y_2^2 \leq 1 \\ & y_1 \geq 1/\sqrt{1+\gamma} \end{array}$$

with variables y_1, y_2. The feasible set is the part of the unit disk to the right of the vertical line through $y_1 = (1+\gamma)^{-1/2}$.



We maximize the inner product of y with the coefficient vector $(x_1, \sqrt{\gamma} x_2)$. There are three cases to distinguish, depending on the orientation of $(x_1, \sqrt{\gamma} x_2)$.

- If $x_1 > 0$ and $|x_2| \leq x_1$, the coefficient vector lies in the cone between the vectors $(1, -\sqrt{\gamma})$ and $(1, \sqrt{\gamma})$, and the optimum is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{x_1^2 + \gamma x_2^2} \begin{bmatrix} x_1 \\ \sqrt{\gamma} x_2 \end{bmatrix}.$$

The optimal value is $(x_1^2+\gamma x_2^2)^{1/2}$.

- If $x_2 \geq 0$, and $x_1 < x_2$, the point

$$\left[\begin{array}{c} y_1 \\ y_2 \end{array}\right] = \frac{1}{\sqrt{1+\gamma}} \left[\begin{array}{c} 1 \\ \sqrt{\gamma} \end{array}\right]$$

is optimal, and the optimal value is $(x_1+\gamma x_2)/(1+\gamma)^{1/2}$.

- If $x_2 \leq 0$, and $x_1 < -x_2$, the point

$$\left[\begin{array}{c} y_1 \\ y_2 \end{array}\right] = \frac{1}{\sqrt{1+\gamma}} \left[\begin{array}{c} 1 \\ -\sqrt{\gamma} \end{array}\right]$$

is optimal, and the optimal value is $(x_1-\gamma x_2)/(1+\gamma)^{1/2}$.

(b) We first note that the iterates given in the problem satisfy $|x_2^{(k)}| < x_1^{(k)}$, so they are in the interior of the region where $f(x_1,x_2) = (x_1^2+\gamma x_2^2)^{1/2}$. In this region the function is differentiable with gradient

$$\nabla f(x) = \frac{1}{\sqrt{x_1^2+\gamma x_2^2}} \left[\begin{array}{c} x_1 \\ \gamma x_2 \end{array}\right].$$

Since we use an exact line search, only the direction of $\nabla f(x)$ matters.

We now verify the expressions

$$x_1^{(k)} = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma-1}{\gamma+1}\right)^k.$$

For $k=0$, this is the starting point $x^{(0)} = (\gamma, 1)$. The gradient at $x^{(k)}$ is proportional to $(x_1^{(k)}, \gamma x_2^{(k)})$, and therefore the exact line search minimizes f along the line

$$\left[\begin{array}{c} (1-t)x_1^{(k)} \\ (1-\gamma t)x_2^{(k)} \end{array}\right] = \left(\frac{\gamma-1}{\gamma+1}\right)^k \left[\begin{array}{c} (1-t)\gamma \\ (1-\gamma t)(-1)^k \end{array}\right].$$

Along this line f is given by

$$f\left((1-t)x_1^{(k)}, (1-\gamma t)x_2^k\right) = \left(\gamma^2(1-t)^2 + \gamma(1-\gamma t)^2\right)^{1/2} \left(\frac{\gamma-1}{\gamma+1}\right)^k.$$

This is minimized by $t = 2/(1+\gamma)$, so we have

$$x^{(k+1)} = \left(\frac{\gamma-1}{\gamma+1}\right)^k \left[\begin{array}{c} (1-t)\gamma \\ (1-\gamma t)(-1)^k \end{array}\right] = \left(\frac{\gamma-1}{\gamma+1}\right)^{k+1} \left[\begin{array}{c} \gamma \\ (-1)^{k+1} \end{array}\right].$$

4. *Exercise A8.7.* The Hessian is

$$H = I + A^T \left(\mathbf{diag}(z) - zz^T \right) A.$$

where

$$z_i = \frac{\exp(a_i^T x + b_i)}{\sum_i \exp(a_i^T x + b_i)},$$

so H is diagonal plus a low rank term, and we can more or less follow the method of page 10-30 of the lecture notes. However $\mathbf{diag}(z) - zz^T$ is singular, since $(\mathbf{diag}(z) - zz^T)\mathbf{1} = 0$, so we cannot directly factor it using the Cholesky factorization. Note that

$$\mathbf{diag}(z) - zz^T = L\, \mathbf{diag}(z)^{-1} L^T$$

where

$$L = \mathbf{diag}(z) - zz^T.$$

The Newton system

$$\left(I + A^T \left(\mathbf{diag}(z) - zz^T \right) A \right) \Delta x = g$$

is therefore equivalent to

$$\left[\begin{array}{cc} I & A^T L \\ L^T A & -\mathbf{diag}(z) \end{array} \right] \left[\begin{array}{c} \Delta x \\ \Delta u \end{array} \right] = \left[\begin{array}{c} g \\ 0 \end{array} \right].$$

Eliminating Δx gives an equation

$$(\mathbf{diag}(z) + L^T A A^T L)\Delta u = L^T A g.$$

with $m + 1$ variables.

The cost is roughly $(1/3)m^3 + m^2 n$ flops.

5. *Exercise A8.9.*

   (a) The problem is

$$\text{minimize} \;\; -\sum_{\bar{y}_i = 1}^{m} \log \Phi(-b_i + a_i^T x) - \sum_{\bar{y}_i = -1}^{m} \log \Phi(b_i - a_i^T x).$$

   where

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt.$$

   $\Phi(u)$ is log-concave (it is the cumulative distribution function of a log-concave density; see page 27 of lecture 3 and exercise T3.55). Therefore $\Phi(a_i^T x - b_i)$ and $\Phi(b_i - a_i^T x)$ are log-concave.

(b) To simplify notation we redefine A and b as

$$A := \mathbf{diag}(\bar{y})A, \qquad b := \mathbf{diag}(\bar{y})b.$$

This allows us to express the problem as

$$\text{minimize } h(Ax - b),$$

where $h : \mathbf{R}^m \to \mathbf{R}$ is defined as

$$h(w) = -\sum_{i=1}^{m} \log \Phi(w_i).$$

The gradient and Hessian of $f(x) = h(Ax - b)$ are given by

$$\nabla f(x) = A^T \nabla h(Ax - b), \qquad \nabla^2 f(x) = A^T \nabla^2 h(Ax - b) A.$$

The first derivatives of h are

$$\frac{\partial h(w)}{\partial w_i} = -\frac{\Phi'(w_i)}{\Phi(w_i)} = \frac{-1/\sqrt{2\pi}}{\exp(w_i^2/2)\Phi(w_i)}.$$

The Hessian $\nabla^2 h(w)$ is diagonal with diagonal elements

$$\begin{aligned}
\frac{\partial^2 h(w)}{\partial w_i^2} &= -\frac{\Phi''(w_i)}{\Phi(w_i)} + \frac{\Phi'(w_i)^2}{\Phi(w_i)^2} \\
&= \frac{w_i/\sqrt{2\pi}}{\exp(w_i^2/2)\Phi(w_i)} + \left(\frac{1/\sqrt{2\pi}}{\exp(w_i^2/2)\Phi(w_i)}\right)^2.
\end{aligned}$$

In the following MATLAB code we take the least squares solution as starting point.

```
one_bit_meas_data;
[m, n] = size(A);
A = diag(y)*A;
b = y.*b;
x = A\b;
for k=1:50
    w = A*x-b;
    Phi = 0.5*erfc(-w/sqrt(2));
    Phix = 0.5*sqrt(2*pi)*erfcx(-w/sqrt(2));
    val = -sum(log(Phi));
    grad = -A'*(1./Phix);
    hess = A'*diag((w + 1./Phix)./Phix)*A;
    v = -hess\grad;
    fprime = grad'*v
```

```
        if (-fprime/2 < 1e-8), break;  end;
        t = 1;
        while ( -sum(log(0.5*erfc(-(A*(x+t*v)-b)/sqrt(2)))) > ...
            val + 0.01*t*fprime )
            t = t/2;
        end;
        x = x + t*v;
    end;
```

This converges in a few iterations to

$$x = (-0.27,\ 9.15,\ 7.98,\ 6.70,\ 6.02,\ 5.0,\ 4.30,\ 2.68,\ 2.02,\ 0.68).$$