COL 726

Homework V

Due on April 16, 2018

1. (**Heath**) Let A be an $n \times n$ real symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. We know that $\lambda_1 = \min_x \frac{x^T A x}{x^T x}$ and $\lambda_n = \max_x \frac{x^T A x}{x^T x}$, with the minimum and the maximum occurring at the corresponding eigenvectors.

   (a) Use an unconstrained optimization routine to compute the extreme eigenvalues and the corresponding eigenvectors of the matrix $A = \begin{pmatrix} 6 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. Is the solution unique in each case? Why?

   (b) The foregoing characterization of λ_1 and λ_n remain valid if we restrict the vector x to be normalized by taking $x^T x = 1$. Repeat the above part, but use a constrained optimization routine to impose this normalization constraint. What is the significance of the Lagrange multiplier in this context?

2. **(Boyd, Vandenberghe)** Newton's method with fixed step size ($\eta = 1$) can diverge if the initial point is not close to $x^\star$. In this problem we consider two examples (plot the function values at each of the iterates):

   (a) $f(x) = \log(e^x + e^{-x})$ has a unique minimizer $x^\star = 0$. Run Newton's method with fixed step size $\eta = 1$, starting at $x(0) = 1$ and at $x(0) = 1.1$.

   (b) (b) $f(x) = -\log x + x$ has a unique minimizer at $x^\star = 1$. Run Newton's method with fixed step size $\eta = 1$, starting at $x(0) = 3$.

3. **(Boyd, Vandenberghe)** Consider the optimization problem:

$$\min f(x) = -\sum_{i=1}^{n} x_i \log x_i, \text{ subject to } Ax = b,$$

   where $x \in \Re^n$ with all coordinates being positive, and A is $p \times n$ matrix, where $p < n$. Generate a problem instance with $n = 100$ and $p = 30$ by choosing A randomly (checking that it has full rank), choosing $\hat{x}$ as a random positive vector (e.g., with entries uniformly distributed on $[0, 1]$) and then setting $b = A\hat{x}$ (Thus, $\hat{x}$ is feasible). Compute a solution using Newton's method. Plot a graph to show the progress of the method.

4. **(Boyd, Vandenberghe)** Let $\gamma > 1$ and consider the function

$$f(x_1, x_2) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & \text{if } |x_2| < x_1 \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{otherwise} \end{cases}$$

Prove that f is convex. Consider the gradient descent algorithm applied to f, with starting point $x(0) = (\gamma, 1)$ and exact line search. Show that the iterates are

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

Therefore x^k converges to $(0,0)$. However, this is not the optimum, since f is unbounded below.

5. **(Boyd, Vandenberghe)** Consider the unconstrained problem

$$\min f(x) = -\sum_{i=1}^{m} \log(1 - a_i^T x) - \sum_{i=1}^{n} \log(1 - x_i^2),$$

where $x \in \Re^n$ and $a_1, \ldots, a_m$ are vectors in $\Re^n$. Note that we can choose $x(0) = 0$ as our initial point. You can generate instances of this problem by choosing a_i from some distribution on $\Re^n$.

- Use the gradient descent method to solve the problem, using reasonable choices for the back-tracking parameters, and a stopping criterion of the form $||\nabla f(x)|| \leq \eta$. Plot the objective function and step length versus iteration number. (Once you have determined $f(x^\star)$ to high accuracy, you can also plot $f(x) - f(x^\star)$ versus iteration.) Experiment with the backtracking parameters α and β to see their effect on the total number of iterations required. Carry these experiments out for several instances of the problem, of different sizes.
- Repeat using Newton's method, with stopping criterion based on the Newton decrement. Look for quadratic convergence. You do not have to use an efficient method to compute the Newton step; you can use a general purpose solver for system of equations.