C&O367: Nonlinear Optimization (Winter 2013) Assignment 4 H. Wolkowicz

Posted Mon, Feb. 8

Due: Thursday, Feb. 28 10:00AM (before class),

1 Matrices

1.1 Positive Definite Matrices

1. Let $A \in \mathcal{S}^n$, i.e., let A be a symmetric, real $n \times n$ matrix. Show that $A \succ 0$ if, and only if, $A^{-1} \succ 0$. (Recall that $A \succ 0$ denotes that A is positive definite.)

   Solution: $A \succ 0$ if and only if all the eigenvalues of A are positive. If $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of A, the eigenvalues of A^{-1} are $\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_n}$. Hence if all the eigenvalues of A are positive, all the eigenvalues of A^{-1} are also positive and $A^{-1} \succ 0$.

2. Let $A \in \mathcal{S}^n_+$. Show that A has a square root $A = P^2$ with $P \in \mathcal{S}^n_+$. And, show that this square root is unique if $A \succ 0$.

   Solution: By SVD decomposition, we can write any positive semidefinite matirx A as $A = Q^T D Q$ where $Q^T Q = QQ^T = I$ and $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix that the diagonal entries are the eigenvalues of A. Now define $P := Q^T D^{\frac{1}{2}} Q$ where $D^{\frac{1}{2}}$ is a diagonal matrix where the diagonal entries are the square root of the eigenvalues of A. Now P is positive semidefinite and we have $P^2 = Q^T D^{\frac{1}{2}} Q Q^T D^{\frac{1}{2}} Q = Q^T D^{\frac{1}{2}} D^{\frac{1}{2}} Q = Q^T D Q = A$.

   I claim that if we put the restriction $P \in \mathcal{S}^n_+$, P is unique for all positive semidefinite matrices. Assume that $C^2 = P^2 = A$. As the eigenvalues of P and C are the same, we have $C = U^T D^{\frac{1}{2}} U$, $UU^T = U^T U = I$. So we have $U^T D U = Q^T D Q$ or $TD = DT$, where $T := QU^T$. Note that $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$, so if we look at $TD = DT$ element wise, we have $t_{ij}\lambda_j = \lambda_i t_{ij}$. This is equivalent to $t_{ij}\lambda_j^{\frac{1}{2}} = \lambda_i^{\frac{1}{2}} t_{ij}$, which means $TD^{\frac{1}{2}} = D^{\frac{1}{2}} T$. $T = QU^T$, so $QU^T D^{\frac{1}{2}} = D^{\frac{1}{2}} QU^T$ or $C = U^T D^{\frac{1}{2}} U = Q^T D^{\frac{1}{2}} Q = P$.

2 Steepest Descent

2.1 Positive Definiteness and Steepest Descent

1. Suppose that $f(x) = \frac{1}{2}(x_1^2 + \alpha x_2^2)$, where $x = (x_1, x_2)^T$ and $\alpha \geq 1$. Suppose we use the method of Steepest Descent with exact line search starting from $x^0 = (\alpha, 1)^T$. Find the sequence that the method generates. Hence, show that

$$\frac{f(x^{k+1})}{f(x^k)} = \left(\frac{\alpha - 1}{\alpha + 1}\right)^2.$$

(Note that this is a worst case error in terms of the condition number of the Hessian.)

Solution: I want to prove by induction that $x^k = [\alpha(\frac{\alpha-1}{\alpha+1})^k \quad (\frac{1-\alpha}{\alpha+1})^k]^T$. We start from $x^0 = (\alpha, 1)^T$, so the base of the induction is true. For the body of induction, note that by steepest descent we have $x^{k+1} = x^k - t\nabla f(x^k) = ((1-t)x_1^k, (1-\alpha t)x_2^k)^T$. By exact line search, we want to find t that minimizes $f(x^{k+1})$. By taking derivative with respect to t and putting equal to zero we have $\bar{t} = \frac{(x_1^k)^2 + \alpha^2 (x_2^k)^2}{(x_1^k)^2 + \alpha^3 (x_2^k)^2}$. By substituting x^k we get $t = \frac{2}{1+\alpha}$. Then we have:

$$x^{k+1} = \begin{bmatrix} (1-t)x_1^k \\ (1-\alpha t)x_2^k \end{bmatrix} = \begin{bmatrix} (1 - \frac{2}{1+\alpha})\alpha(\frac{\alpha-1}{\alpha+1})^k \\ (1 - \frac{2\alpha}{1+\alpha})(\frac{1-\alpha}{\alpha+1})^k \end{bmatrix} = \begin{bmatrix} \alpha(\frac{\alpha-1}{\alpha+1})^{k+1} \\ (\frac{1-\alpha}{\alpha+1})^{k+1} \end{bmatrix},$$

and we are done. $\frac{f(x^{k+1})}{f(x^k)} = \left(\frac{\alpha-1}{\alpha+1}\right)^2$ is just a simple substitution.

2. Suppose that $f(x)$ is a quadratic function on $\mathbb{R}^n$, $f(x) := a + b^T x + \frac{1}{2}x^T A x$, where $a \in \mathbb{R}, b \in \mathbb{R}^n, A \in \mathcal{S}^n$ is positive definite.

   (a) Show that $f(x)$ has a unique global minimizer.

   Solution: We had this question before. Because A is positive definite, $f(x)$ is strictly convex and had a unique minimizer x^* that is the solution of $Ax^* + b = 0$.

   (b) Show that if the initial point x^0 for Steepest Descent is selected so that $x^0 - x^*$ is an eigenvector of A, then the Steepest Descent sequence $\{x^k\}$ with initial point x^0 reaches the optimum x^* in one step, i.e., $x^1 = x^*$.

   Solution: If we start with x^0, the next point by the steepest descent is $x^1 = x^0 - t(Ax^0 + b)$. But $b = -Ax^*$, so $x^1 = x^0 - t(Ax^0 - Ax^*) = x^0 - t(A(x^0 - x^*))$. By hypothesis, $x^0 - x^*$ is an eigenvector of A, so there exists $\lambda > 0$ that $A(x^0 - x^*) = \lambda(x^0 - x^*)$. Hence $x^1 = x^0 - t\lambda(x^0 - x^*)$. By chooseing $t = \frac{1}{\lambda}$ we get $x^1 = x^*$ as we wanted.

2.2 Steepest Descent for Different Norms

1. Let $P \succ 0$. What is the Steepest Descent direction for the quadratic norm defined by $\|z\|_P := \sqrt{\langle z^T P z \rangle}$.

   Solution: From the course notes, steepest descent direction is

   $$d_{sd} := \text{argmin}\{\nabla f(x)^T z : \|z\| \leq 1\},$$

   where the norm in the definition is the desired norm. To solve this problem, we consider the Lagrangian $L(z, \lambda) := \nabla f(x)^T z + \lambda(\|z\| - 1)$, $\lambda \geq 0$, and minimize that. To do that we put $\nabla L(z, \lambda) = 0$.
   For quadratic norm $\|z\|_P := \sqrt{\langle z^T P z \rangle}$, we have $\nabla \|z\|_P = \frac{1}{2\sqrt{\langle z^T P z \rangle}} P z$, so

   $$\nabla L(z, \lambda) = \nabla f(x) + \frac{\lambda}{2\sqrt{\langle z^T P z \rangle}} P z = 0 \quad \Rightarrow \quad z = -\left(\frac{2\sqrt{\langle z^T P z \rangle}}{\lambda}\right) P^{-1} \nabla f(x).$$

   As $\left(\frac{2\sqrt{\langle z^T P z \rangle}}{\lambda}\right) \geq 0$, the Steepest descent direction is $d_{sd} = -P^{-1} \nabla f(x)$.

2. What is the Steepest Descent direction for the ℓ_1 norm.

   Solution: ℓ_1 norm is not differentiable at all the points, so we can not use the previous method. Our constraint is $\|v\|_1 = \sum_{i=1}^n |v_i| =\leq 1$. To minimize $\nabla f(x)^T v$, it is clear that we can pick the entry of $\nabla f(x)$ that has the largest absolute value, say $(\nabla f(x))_i = \frac{\partial f(x)}{\partial x_i}$, and then put $v_i = \pm 1$, and $v_j = 0$, $j \in \{1, \cdots, n\} \backslash \{i\}$. The sign of v_i is such that $\frac{\partial f(x)}{\partial x_i} v_i \leq 0$. Hence the Steepest descent direction is $d_{sd} = \text{sign}\left(\frac{\partial f(x)}{\partial x_i}\right) e_i$, where e_i is the ith standard basis of $\mathbb{R}^n$.

3 Linear Least Squares

1. Consider the problem $\min_x \|Ax - b\|^2$, where A is an $m \times n$ matrix, $b \in \mathbb{R}^m$, and the solution $x \in \mathbb{R}^n$.

   (a) What is the first order necessary condition for optimality? Is it also a sufficient condition? If so, why?

   Solution: We can write

   $$f(x) := \|Ax - b\|^2 = (Ax - b)^T (Ax - b) = x^T A^T A x - 2b^T A x + b^T b.$$

   Hence the first order necessary condition is $\nabla f(x) = 2(A^T A)x - 2A^T b = 0$. Note that for any A, $A^T A$ is always a positive semidefinite matrix. Hence, $f(x)$ is a convex quadratic function and so the $\nabla f(x) = 0$ is also a sufficient condition.

(b) Is the optimal solution unique? Give reasons for your answer. If it is not unique, is it a convex set? And, is there a unique element of minimal norm in this set?

Solution: The optimal solution is unique if and only if $2(A^TA)x - 2A^Tb = 0$ has a unique solution, which is equivalent to A^TA being a positive definite matrix. A^TA is a positive definite matrix if and only if A has full column rank. Let D be the set of optimal solutions. D is the set of solutions of the linear system $2(A^TA)x - 2A^Tb = 0$, so it is clearly closed and convex. We want to minimize $\|x\|$, which is a strictly convex function, over the set D. $\|x\|$ also has a lower bound (zero is a lower bound), so it has a unique minimizer over D.

(c) Can you give a closed form expression for the optimal solution? Specify any assumptions that you may need.

Solution: If we assume that A has full column rank, then A^TA is invertible and as we explained above, we have a unique minimizer $x^* = (A^TA)^{-1}A^Tb$.
If A has full row rank, I want to find a formulation for the answer with the minimum norm (we also need it for question 2). We want to solve the problem $\min\{\frac{1}{2}\|x\|^2 : A^TAx = A^Tb\}$. I chose $\frac{1}{2}\|x\|^2$ to simplify the calculations. There is a Lagrangian multiplier vector λ that satisfies $x^* - A^TA\lambda = 0$. Substituting x^* in $A^TAx^* = A^Tb$ we get $A^TAA^TA\lambda = A^Tb$. A has full row rank, so this equation is equivalent to $AA^TA\lambda = b$. A has full low rank also results in AA^T being invertible, so $A\lambda = (AA^T)^{-1}b$. Hitting both sides of the last equation with A^T and using the fact that $x^* = A^TA\lambda$ we have $x^* = A^T(AA^T)^{-1}b$.

(d) Show that the residual $Ax - b$ at the optimal x is orthogonal to the columns of A.

Solution: Any optimal solution is a solution of $(A^TA)x - A^Tb = A^T(Ax - b) = 0$. This is equivalent to $Ax - b$ is orthogonal to the columns of A.

(e) Use the two different factorizations QR and SVD to solve the problem for

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad b = \begin{pmatrix} 2 \\ 6 \\ 2 \\ 0 \end{pmatrix}.$$

State whether or not the system $AX = b$ is consistent and why. Also, compare the norm of the residual that you obtain with that obtained by applying Gaussian elimination to the normal equations.

Solution: The QR factorization for A is:

$$Q = \begin{bmatrix} 0.8944 & -0.0877 & -0.2224 \\ 0 & 0.8771 & 0.2965 \\ 0 & 0.4385 & -0.8154 \\ 0.4472 & 0.1754 & 0.4447 \end{bmatrix} \qquad R = \begin{bmatrix} 2.2361 & -0.8944 & -0.4472 \\ 0 & 2.2804 & 1.9295 \\ 0 & 0 & 1.0377 \end{bmatrix}$$

Then x^* is the solution of $Rx = Qb$, which is $x^* = (2, 2.8571, -.2857)^T$. The SVD decomposition is $A = UDV$ where you can get by the function svd of MATLAB. Then x^* is the solution to $DVx = U^T b$. By solving this we again get $x^* = (2, 2.8571, -.2857)^T$.
The system $Ax = b$ is not sonsistent, becuase if we solve the fisrt 3 equations, we get the answer $x = (2, 2, 1)^T$ which does not satisfy the forth equation. By using Gaussin elimination for normal equations, we get:

$$\begin{bmatrix} 5 & -2 & 1 & 4 \\ -2 & 6 & 4 & 12 \\ 1 & 4 & 5 & 12 \end{bmatrix} \quad \to R = \begin{bmatrix} 0 & 0 & 1 & \frac{-2}{7} \\ 0 & 1 & 0 & \frac{20}{7} \\ 1 & 0 & 0 & 2 \end{bmatrix}$$

From all three methods, the norm of the residual is approximately 2.2678.

2. Find the minimum norm solution of the underdetermined linear system with

$$A = \begin{bmatrix} 2 & 1 & 1 & 5 \\ -1 & -1 & 3 & 2 \end{bmatrix}, \quad b = \begin{pmatrix} 8 \\ 0 \end{pmatrix}.$$

Solution: A has full row rank, and we derived an exact formulation for the minimum norm solution in the answer for question 3-1-(c); $x^* = A^T(AA^T)^{-1}b$. Using that we have: $x^* = (0.8767, 0.5479, -0.3288, 1.2055)^T$.