Assignment 1

APM462 – Nonlinear Optimization – Summer 2016

Christopher J. Adkins

Solutions

Question 1 Let $f(x, y) = 2x^2 + y^2 + xy - y$

(a) Find a point satisfying the first order conditions for f

(b) Prove that the point you found in a) is a global minimum for f

Solution Notice we may rewrite f as

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x}, Q\mathbf{x}) - (b, \mathbf{x})$$

where (notation def: $(\mathbf{x}_1, \mathbf{x}_2) := \mathbf{x}_1^T \mathbf{x}_2$)

$$Q = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \quad \& \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

(a) Now it's easy to see the first order condition ($\nabla f = 0$) gives

$$\nabla f(\mathbf{x}_*) = Q\mathbf{x}_* - b = 0 \implies \mathbf{x}_* = Q^{-1}b$$

One may now easily compute $\mathbf{x}_*$:

$$Q^{-1} = \frac{1}{7}\begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} \implies \mathbf{x}_* = \begin{pmatrix} -1/7 \\ 4/7 \end{pmatrix}$$

(b) We compute the eigenvalues of Q using $P(\lambda) = \det(Q - 1\lambda) = \lambda^2 - 6\lambda + 7$. Finding the roots gives the two eigenvalues as

$$\lambda_\pm = 3 \pm \sqrt{2}$$

i.e. $\lambda_+ > \lambda_- > 0$ which means Q is positive definite. Now if we "complete the square" on f, we see

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_*, Q(\mathbf{x} - \mathbf{x}_*)) - \frac{1}{2}(\mathbf{x}_*, Q\mathbf{x}_*) \geqslant -\frac{1}{2}(\mathbf{x}_*, Q\mathbf{x}_*)$$

Thus $\mathbf{x}_*$ is a global minimum for f. □

Question 2 Find all local minimum points for the function

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 8y - 8z + 9$$

Prove that your solution really is a global minimum.

Solution Notice we may rewrite f as

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x}, Q\mathbf{x}) - (b, \mathbf{x}) + 9$$

where

$$Q = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \quad \& \quad b = \begin{pmatrix} 6 \\ 8 \\ 8 \end{pmatrix}$$

One may easily compute $\mathbf{x}_*$ using $\mathbf{x}_* = Q^{-1}b$ as we saw in the previous question. Plug and Chug

$$Q^{-1} = \frac{1}{10}\begin{pmatrix} 3 & -2 & 1 \\ -2 & 8 & -4 \\ 1 & -4 & 7 \end{pmatrix} \implies \mathbf{x}_* = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Now we show Q is positive definite by checking the eigenvalues. We see the characteristic equation is given by

$$P(\lambda) = \det(Q - 1\lambda) = -\lambda^3 + 8\lambda^2 - 18\lambda + 10$$

and the obvious bound of

$$P(\lambda) \geqslant 10 \quad \text{when} \quad \lambda \leqslant 0$$

shows that all eigenvalues are positive(since Q is symmetric, the eigenvalues must be real, and we've shown there are no non-positive eigenvalues), hence $\mathbf{x}_*$ is a global minimum by completing the square as we saw in the previous question. Another method of of checking if Q is positive definite (could use this for question 1 as well) is through Sylvester's Criterion. This states that a symmetric matrix M is positive definite if and only if the following matrices have positive determinate: the upper left 1-by-1 corner of M, the upper left 2-by-2 corner of M, ..., M itself. In this case its easy to see that

$$\det(4) \implies 4 \quad \& \quad \det\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} = 7 \quad \& \quad \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} = 10$$

Thus Sylvester's Criterion gives us Q is positive definite. □

Question 3 To approximate a function $g : [0, 1] \to \mathbb{R}$ by an n-th order polynomial, one can minimize the function f defined by

$$f(a) = \int_0^1 (g(x) - p_a(x))^2 dx$$

where, for $a = (a_0, \ldots, a_n) \in \mathbb{R}^{n+1}$, we use the notation

$$p_a(x) = a_0 + a_1 x + \ldots + a_n x^n = (\mathbf{x}, a) \quad \& \quad \mathbf{x} = (1, x, \ldots, x^n)$$

In this question we will investigate approximating the parabola $g(x) = x^2$ by the linear polynomials $p_a(x) = a_0 + a_1 x$.

(a) Show that $f(a)$ can be written in the form

$$f(a) = a^T Q a - 2b^T a + c$$

for a 2×2 matrix Q, a vector $b \in \mathbb{R}^2$ and a number c. Find formulas for Q, b and c. It should be clear from your formula that Q is symmetric.

(b) Find the first-order necessary condition for a point $a_* \in \mathbb{R}^2$ to be a minimum point for f. □

Solution

(a) Expanding out $f(a)$, we see that

$$\begin{aligned} f(a) = \int_0^1 (g(x) - p_a(x))^2 dx = \int_0^1 (g(x) - (\mathbf{x}, a))^2 dx \\ &= \int_0^1 \left[(\mathbf{x}, a)(\mathbf{x}, a) - 2g(x)(\mathbf{x}, a) + g(x)^2\right] dx \\ &= \int_0^1 \left[(a, \mathbf{x}\mathbf{x}^T a) - 2(g(x)\mathbf{x}, a) + g(x)^2\right] dx \\ &= (a, Qa) - 2(b, a) + c \end{aligned}$$

where

$$Q = \int_0^1 \mathbf{x}\mathbf{x}^T dx \implies Q_{ij} = \int_0^1 x^{i-1}x^{j-1} dx = \frac{1}{i+j-1}$$

$$b = \int_0^1 g(x)\mathbf{x} dx \implies b_i = \int_0^1 g(x)x^{i-1} dx$$

$$c = \int_0^1 g(x)^2 dx$$

Note that the particular form of Q implies that it is positive definite (since $(v, xx^T v) = (v, x)^2 \geqslant 0$ and $\det Q \neq 0$). If $g(x) = x^2$ and $\mathbf{x} = (1, x)$, we see

$$Q = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}$$

$$b = \int_0^1 \begin{pmatrix} x^2 \\ x^3 \end{pmatrix} dx = \begin{pmatrix} 1/3 \\ 1/4 \end{pmatrix}$$

$$c = \int_0^1 x^4 dx = 1/5$$

(b) As usual with functions of this form, we know that critical point a_* must satisfy $Qa_* = b$, and since Q is positive definite we have that $a_* = Q^{-1}b$. □

Question 4 Assume that g is a convex function on $\mathbb{R}^n$, that f is a convex function of a single variable, and in addition that f is a nondecreasing function (which means that $f(r) \geqslant f(s)$ whenever $r \geqslant s$).

(a) Show that $F(x) := f \circ g(x) = f(g(x))$ is convex by directly verifying the convexity inequality

$$F(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \leqslant tF(\mathbf{x}_1) + (1-t)F(\mathbf{x}_2)$$

explain where each hypothesis (convexity of g, convexity of f, and the fact that f is nondecreasing) is used in your reasoning.

(b) Now assume that f and g are both C^2. Express the matrix of second derivatives $\nabla^2 F(x)$ in terms of f and g. Prove directly (without using part a)) that $\nabla^2 F(x)$ is positive semidefinite at every x.

Solution

(a) By direct computation, we see

$$\begin{aligned} F(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) =& f(g(t\mathbf{x}_1 + (1-t)\mathbf{x}_2)) && \text{definition of } F \\ \leqslant& f(tg(\mathbf{x}_1) + (1-t)g(\mathbf{x}_2)) && g \text{ is convex and } f \text{ nondecreasing} \\ \leqslant& tf(g(\mathbf{x}_1)) + (1-t)f(g(\mathbf{x}_2)) && f \text{ is convex} \\ =& tF(\mathbf{x}_1) + (1-t)F(\mathbf{x}_2) && \text{defintion of } F \end{aligned}$$

(b) First compute the gradient, we see that

$$\nabla F(\mathbf{x}) = f'(g(\mathbf{x}))\nabla g$$

Computing the matrix of second derivatives now shows we have (by product rule)

$$\nabla^2 F(\mathbf{x}) = f''(g(\mathbf{x}))\nabla g \nabla g^T + f'(g(\mathbf{x}))\nabla^2 g$$

Since f is nondecreasing and convex, we have that $f' \geqslant 0$ and $f'' \geqslant 0$ at every x. Since g is convex, we have that $\nabla^2 g$ is positive semidefinite at every x. As we've mentioned before, matrices of the form $\mathbf{x}\mathbf{x}^T$ are positive semidefinite(in this case we have $\nabla g \nabla g^T$). Thus $\nabla^2 F$ is positive semidefinite at every x, i.e.

$$(y, \nabla^2 F(\mathbf{x})\, y) = f''(g(\mathbf{x}))(y, \nabla g \nabla g^T(\mathbf{x})\, y) + f'(g(\mathbf{x}))(y, \nabla^2 g(\mathbf{x})\, y) \geqslant 0 \quad \forall y \in \mathbb{R}^n$$

□

Question 5 Prove that if f_1 and f_2 are two convex functions on $\mathbb{R}^n$, then

$$g(x) := \max\{f_1(x), f_2(x)\}$$

is also convex.

Solution This is easy to verify directly:

$$\begin{aligned} g(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) =& \max\{f_1(t\mathbf{x}_1 + (1-t)\mathbf{x}_2), f_2(t\mathbf{x}_1 + (1-t)\mathbf{x}_2)\} \\ \leqslant& \max\{tf_1(\mathbf{x}_1) + (1-t)f_1(\mathbf{x}_2), tf_2(\mathbf{x}_1) + (1-t)f_2(\mathbf{x}_2)\} && f_1 \text{ and } f_2 \text{ are convex.} \\ \leqslant& t\max\{f_1(\mathbf{x}_1), f_2(\mathbf{x}_1)\} + (1-t)\max\{f_1(\mathbf{x}_2), f_2(\mathbf{x}_2)\} && \text{bound by the bigger function at } \mathbf{x}_1 \text{ and } \mathbf{x}_2 \\ =& tg(\mathbf{x}_1) + (1-t)g(\mathbf{x}_2) \end{aligned}$$

□