

Math 417
Mathematical Programming

Homework X

Frédéric Boileau

Prof. Tim Hoheisel

2 janvier 2017

Contents

1	Gradient of quadratic function	2
2	Linear least squares problem	3
3	KKT conditions	5

1 Gradient of quadratic function

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2}x^\top Mx + c^\top x \quad (1)$$

where $M \in \mathbb{R}^{n \times n}$ is a matrix not necessarily symmetric and $c \in \mathbb{R}^n$. Compute the derivative of $f(x)$. We will first prove the following lemma. It makes the computation direct and allows us to forbade partial derivatives in those cases in the (near) future.

Lemma: Let $g(x) := x^\top Ax$ for some real square matrix A . Then the derivative of $g(x)$ is $x^\top(A + A^\top)$

Proof:

First of all we notice that $f(x)$ can be rewritten as follows

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j \quad (2)$$

And thus:

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^n a_{ki}x_i + \sum_{j=1}^n a_{jk}x_j = \langle x, a_k + a'_k \rangle \quad (3)$$

With a_i and a'_j denoting the i th row and the j th column of A respectively. From 3 we have that

$$\nabla f(x) = (\langle x, a_1 + a'_1 \rangle, \dots, \langle x, a_n + a'_n \rangle) \quad (4)$$

$$= \langle x, (a_1 + a'_1, \dots, a_n + a'_n) \rangle \quad (5)$$

$$= \langle x, (a_1, \dots, a_n) \rangle + \langle x, (a'_1, \dots, a'_n) \rangle \quad (6)$$

$$= x^\top A^\top + x^\top A = x^\top(A + A^\top) \quad (7)$$

■

This lemma with the obvious fact that $\nabla c^\top x = c^\top$ gives us:

$$\nabla f(x) = \frac{1}{2}x^\top(A + A^\top) + c^\top \quad (8)$$

2 Linear least squares problem

For $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ consider

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \|Ax - b\|_2^2 \quad (9)$$

a) Claim: $\text{im}(A^\top A) = \text{im} A^\top$

Proof: We know from basic linear algebra that the images of any two matrices that live in the same space are the same if and only if their rank are equal. We prove this for A and AA^\top . Let $x \in \mathbb{R}^n$ such that

$$(AA^\top)x = 0 \iff x^\top(AA^\top)x = 0 \quad x \neq 0 \quad (10)$$

But by construction we know that

$$x^\top(AA^\top)x = \|A^\top x\|_2^2 \quad (11)$$

Hence 10 implies that a vector is in the kernel of AA^\top if and only if it is in the kernel of A^\top . Hence they have the same rank and the proof is complete. ■

b) Claim: f is convex

Proof: First we note that the mapping $x \mapsto Ax - b$ is affine and therefore convex which means that its image is a convex set. Moreover we know scaling a set does not affect convexity. So it remains to show the operation $\|\cdot\|_2^2$ is a convex function. We have shown in the previous assignment "a continuous, twice differentiable function of several variables is convex on a convex set if and only if its Hessian matrix is positive semidefinite on the interior of the convex set."¹. Let us find the Hessian of the squared euclidean norm.

$$f(x) := \sum_{i=1}^n x_i^2 \quad (12)$$

$$\frac{\partial}{\partial x_i} f(x) = \frac{\partial}{\partial x_i} \sum_{k=1}^n x_k^2 = 2x_i \quad (13)$$

$$\nabla f(x) = \sum_{i=1}^n 2x_i e_i \Rightarrow H(f(x)) = 2I \quad (14)$$

Hence the Hessian is clearly positive definite. And the proof is complete. To sum up: we disregard the scaling, we consider the composition of the squared euclidean norm with an affine mapping. The image of any affine mapping is convex and the euclidean norm with a convex set as support is a convex function. ■

¹wikipedia

- c) The last claim is that the problem in 9 always has a solution and that it is unique if the matrix A has full rank.

Claim: There is always a solution to the problem.

Proof: We find the gradient of the function:

$$\begin{aligned} f(x) &:= \|Ax - b\|_2^2 \\ &= \|Ax\|^2 + \|b\|^2 - 2\langle Ax, b \rangle \\ &= x^T A^T Ax - 2(Ax)^T b + b^T b \end{aligned}$$

We apply the basic rules of matrix calculus to get

$$\nabla f(x) = 2A^T Ax - 2A^T b = 0 \quad (15)$$

Hence the set of solutions is defined by the next set of equations:

$$A^T Ax = A^T b \quad (16)$$

Moreover as we have shown that the image of A^T is the same as the image of AA^T it follows directly that there is always an $x \in \mathbb{R}^n$ that fulfills the above equations. In other words $A^T b$ is in the image of AA^T so by definition of the image we know there exists an x such that $AA^T x = A^T b$

■

Claim: The solution is unique whenever A has full column rank.

Proof: A has full column rank so AA^T is positive definite so it is invertible. Thus we can solve for x in the following way:

$$x = (AA^T)^{-1} A^T b \quad (17)$$

and the solution is unique because we are solving a matrix equation of the form $\hat{A}x = y$ where $\hat{A} = AA^T$ and $y = A^T b$ by inverting \hat{A} .

■

3 KKT conditions

a)

b) Claim: If \bar{x} solves the optimality conditions described above then it solves the general quadratic problem.

Proof: First let us say that if we can solve the conditions described above then clearly the feasible set is not empty as some of the conditions are exactly the feasibility conditions $Bx = b \ Ax \leq a$. Moreover we have that the gradient is $Qx + c$ and hence the Hessian is just Q , positive semi-definite by assumption. The objective function is continuous over its feasible set which is compact (a polyhedron is generally) so since the Hessian (Q) is positive semi-definite the problem is convex. Finally

- i. A point is a local minimizer if and only if it respects the optimality conditions thus \bar{x} is a local minimizer
- ii. Every local minimizer is a global minimizer since the problem is convex

Hence \bar{x} is a global minimizer if it respects the optimality conditions described above, which is another way to say it solves the problem.

■