Math 661 | Homework
Optimization | August 28, 2020

Gradients of quadratic functions

In the course we will repeatedly use formulas for gradients of quadratic functions of many variables, expressed using vector and matrix notation. The goal of these problems is to derive them.

1. Consider a scalar valued function $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a vector variable in $\mathbb{R}^n$ and A is an $n \times n$ matrix.

   (a) Explain why any function in n variables that has only quadratic terms (terms of the form $b_{ij}x_ix_j$) can be written in this form. In particular give *two different* choices of A to do this for the function

   $$f(x, y, z) = 3x^2 + 2xy - 4y^2 + 5xz + 6yz + 2z^2.$$

   (b) If we impose an additional requirement that A be symmetric (i.e., $A = A^T$), then any function with only quadratic terms arises from only a single A. Explain why, and give a symmetric A for the function in part (a).

   (c) Explain why replacing a non-symmetric A by the symmetric $\frac{1}{2}(A + A^T)$ does not change the function q.

   (d) While there are several approaches to determining the gradient of q, we'll follow one that takes advantage of the vector/matrix notation as much as possible. Recall from calculus that the directional derivative in the direction $\mathbf{u}$ of a differentiable function f of several variables is

   $$\frac{d}{dt} f(\mathbf{x} + t\mathbf{u})\bigg|_{t=0} = \nabla f(\mathbf{x})^T \mathbf{u},$$

   for any unit vector $\mathbf{u}$.

      i. Expand $q(\mathbf{x} + t\mathbf{u})$ as a sum of 4 products of vectors and matrices.

      ii. Differentiate the expanded form as a function of the scalar t, and then set t=0, to get an expression with only 2 terms.

      iii. Write this last expression as some vector transposed times $\mathbf{u}$.

   (e) Explain why this shows $\nabla q = (A^T + A)\mathbf{x}$, and in the symmetric case $\nabla q = 2A\mathbf{x}$

2. Suppose f is an arbitrary quadratic function (which may have constant or linear terms).

   (a) Explain why f can always be expressed as

   $$f(\mathbf{x}) = c + \mathbf{b}^T\mathbf{x} + \mathbf{x}^T A\mathbf{x}$$

   for a scalar c, a vector $\mathbf{b}$, and a symmetric matrix A.

   (b) Show, assuming A is symmetric, the gradient of f is

   $$\nabla f = \mathbf{b} + 2A\mathbf{x}.$$