EECS 16B Designing Information Devices and Systems II
Fall 2021 Discussion Worksheet Discussion 12B

The following notes are useful for this discussion:

1. Quadratic Approximation and Vector Differentiation

As shown in the previous discussion, a common way to approximate a non-linear high-dimensional functions is to perform linearization near a point. In the case of a two-dimensional function $f(x,y)$ with scalar output, the linear approximation of $f(x,y)$ at a point $(x_\star, y_\star)$ is given by

$$f(x,y) \approx f(x_\star, y_\star) + f_x(x_\star, y_\star)(x - x_\star) + f_y(x_\star, y_\star)(y - y_\star) \tag{1}$$

where as in the previous section,

$$f_x(x_\star, y_\star) = \left.\frac{\partial f(x,y)}{\partial x}\right|_{(x_\star, y_\star)} \quad \text{and} \quad f_y(x_\star, y_\star) = \left.\frac{\partial f(x,y)}{\partial y}\right|_{(x_\star, y_\star)}. \tag{2}$$

In vector form, this can be written as:

$$f(\vec{x}) \approx f(\vec{x}_\star) + \left[D_{\vec{x}} f|_{\vec{x}_\star} \right] (\vec{x} - \vec{x}_\star). \tag{3}$$

Recall from the previous discussion that $D_{\vec{x}} f$ is a row-vector filled with the partial derivatives $\frac{\partial f(\vec{x})}{\partial x_i}$:

$$D_{\vec{x}} f = \begin{bmatrix} \frac{\partial f(\vec{x})}{\partial x_1} & \cdots & \frac{\partial f(\vec{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} f_{x_1}(\vec{x}) & \cdots & f_{x_n}(\vec{x}) \end{bmatrix}. \tag{4}$$

Our goal is to extend this idea to a quadratic approximation. To do this, we need some notion of a second derivative. For this discussion, we will only be considering these types of functions from $\mathbb{R}^n \to \mathbb{R}$, since that is the typical form for a cost function used during optimization.

(a) **Given the function $f(x) = \mathrm{e}^{-2x}$, find the first and second derivatives, and write out its quadratic approximation at $x = x_\star$.** *Hint: Use Taylor's theorem.*

(b) To write second partial derivatives compactly, we will introduce a new notation that builds off the notation f_x and f_y introduced previously. To compute f_{xy}, we first take the derivative in x, then in y:

$$f_{xy}(x_\star, y_\star) = \left.\frac{\partial f_x(x,y)}{\partial y}\right|_{(x_\star, y_\star)} = \left.\frac{\partial^2 f(x,y)}{\partial y \partial x}\right|_{(x_\star, y_\star)}. \tag{5}$$

Given the function $f(x,y) = x^2y^2$, find all of the first and second partial derivatives.

(c) To find the quadratic approximation of $f(x,y)$ near $(x_\star, y_\star)$, we plug in $f(x_\star + \Delta x, y_\star + \Delta y)$ and drop the terms that are higher order than quadratic:

$$\begin{aligned}
f(x_\star + \Delta x, y_\star + \Delta y) &= (x_\star + \Delta x)^2 (y_\star + \Delta y)^2 && (6)\\
&= (x_\star^2 + 2x_\star \Delta x + (\Delta x)^2)(y_\star^2 + 2y_\star \Delta y + (\Delta y)^2) && (7)\\
&\approx x_\star^2 y_\star^2 + 2x_\star y_\star^2 \Delta x + 2x_\star^2 y_\star \Delta y && (8)\\
&\quad + y_\star^2 (\Delta x)^2 + 4x_\star y_\star (\Delta x)(\Delta y) + x_\star^2 (\Delta y)^2 && (9)\\
&= f(x_\star, y_\star) + f_x(x_\star, y_\star)\Delta x + f_y(x_\star, y_\star)\Delta y && (10)\\
&\quad + \frac{1}{2} f_{xx}(x_\star, y_\star)(\Delta x)^2 + \frac{1}{2} f_{yy}(x_\star, y_\star)(\Delta y)^2 && (11)\\
&\quad + f_{xy}(x_\star, y_\star)(\Delta x)(\Delta y). && (12)
\end{aligned}$$

This is slightly different from the expression we get via the Taylor series expansion. **How would we rewrite this expression, so that *all* second derivatives are involved, each with a coefficient of $\frac{1}{2}$?**

(d) Just as we created the derivative row vector to hold all the first partial derivatives to help in writing linearization in matrix/vector form:

$$D_{\vec{x}} f = \begin{bmatrix} \frac{\partial f(\vec{x})}{\partial x_1} & \cdots & \frac{\partial f(\vec{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} f_{x_1}(\vec{x}) & \cdots & f_{x_n}(\vec{x}) \end{bmatrix} \tag{13}$$

we would like to create a matrix to hold all the second partial derivatives to help in writing quadratic approximation in matrix/vector form:

$$H_{\vec{x}} f = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\vec{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_n^2} \end{bmatrix} = \begin{bmatrix} f_{x_1 x_1}(\vec{x}) & \cdots & f_{x_n x_1}(\vec{x}) \\ \vdots & \ddots & \vdots \\ f_{x_1 x_n}(\vec{x}) & \cdots & f_{x_n x_n}(\vec{x}) \end{bmatrix} \tag{14}$$

This matrix is the *Hessian* of f. Note that this quantity is different from the *Jacobian* matrix that was covered in the previous discussion. In contrast to the Hessian, which is the matrix of second partial derivatives of a *scalar-valued vector-input* function $f\colon \mathbb{R}^n \to \mathbb{R}$, the Jacobian is the matrix of first partial derivatives of a *vector-valued vector-input* function $\vec{f}\colon \mathbb{R}^n \to \mathbb{R}^k$.

In fact, the Hessian is the (Jacobian) derivative of the derivative; if we let $\vec{g}(\vec{x}) = (D_{\vec{x}} f)^\top$ (so that it's a column vector and the dimensions work out), then $H_{\vec{x}} f = D_{\vec{x}} \vec{g}$. **To get a feel for the Hessian of f, find $H_{(x,y)} f$ for the f above, that is, $f(x, y) = x^2 y^2$.**

(e) **Using the Hessian, write out the general formula for the quadratic approximation of a scalar-valued function f of a vector $\vec{x}$ in vector/matrix form.**

(f) **[Practice]: Show that the quadratic approximation for the scalar-valued function $f(\vec{w}) = \mathrm{e}^{\vec{x}^\top \vec{w}}$ around $\vec{w} = \vec{w}_\star$ is**

$$f(\vec{w}_\star + \Delta\vec{w}) \approx \mathrm{e}^{\vec{x}^\top \vec{w}_\star}\left(1 + \vec{x}^\top(\Delta\vec{w}) + \frac{1}{2}\left(\vec{x}^\top(\Delta\vec{w})\right)^2\right). \tag{15}$$

assuming that $\vec{x}$ is just some constant, given vector.

Hint: You can compute the following partial derivatives:

$$f_{w_i}(\vec{w}) = x_i f(\vec{w}) \tag{16}$$

$$f_{w_i w_j}(\vec{w}) = x_i x_j f(\vec{w}). \tag{17}$$

Now compute $D_{\vec{w}} f$ and $H_{\vec{w}} f$, and plug it into the quadratic approximation formula.

(g) Using the result in the previous subpart, **use linearity to give the quadratic approximation for the function $\sum_{i=1}^{m} \mathrm{e}^{\vec{x}_i^\top \vec{w}}$ around $\vec{w} = \vec{w}_\star$. Here, assume that the $\vec{x}_i$ are just some given vectors.**

(h) **[Practice]:** The second derivative also has an interpretation as the derivative of the derivative. However, we saw that the derivative of a scalar-valued function with respect to a vector is naturally a row. **If you wanted to approximate how much the derivative changed by moving a small amount $\Delta\vec{w}$, how would you get such an estimate using your expression for the second derivative?**

Contributors:

- Neelesh Ramachandran.
- Druv Pai.
- Pavan Bhargava.
- Anant Sahai.

