EECS 16B DIS 13B Gaoyue

Quadratic Approximation

Review:

Linear approximation:

$$f(x,y) \approx f(x_*, y_*) + f_x(x_*, y_*)(x-x_*) + f_y(x_*, y_*)(y-y_*)$$

For a general $f(\vec{x})$, $\vec{x} = [x_1 \ x_2 \ \cdots \ x_n]^T$

$$f(\vec{x}) \approx f(\vec{x}_*) + \left[D_{\vec{x}} f \big|_{\vec{x}_*} \right] (\vec{x} - \vec{x}_*)$$

$$D_{\vec{x}} f = \underbrace{\left[\frac{\partial f(\vec{x})}{\partial x_1} \ \cdots \cdots \ \frac{\partial f(\vec{x})}{\partial x_n} \right]}_{n} \Big\} 1$$

Taylor Series for scalar valued $f(x)$:

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(x_*) \frac{(x-x_*)^k}{k!}$$

Jacobian $D_{\vec{x}} \vec{f}$ for $\vec{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$ $f_i(x_1, \ldots, x_n)$

$$D_{\vec{x}} \vec{f} = \begin{bmatrix} D_{\vec{x}} f_1 \\ \vdots \\ D_{\vec{x}} f_m \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}}_{n} \Bigg\} m$$

v'

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

$f(x) = e^{-2x}$
$f'(x) = -2e^{-2x}$
$f''(x) = 4e^{-2x}$

Quadratic Approx: keep 2nd order term in Taylor expansion

Taylor expand $f(x)$ around x_*:

$f(x) = \sum_{k=0}^{\infty} f^{(k)}(x_*) \frac{(x-x_*)^k}{k!}$

$= f(x_*) + f'(x_*)(x-x_*) + \frac{1}{2} f''(x_*)(x-x_*) + \cdots$

$$f(x) \approx f(x_*) + f'(x_*)(x-x_*) + \frac{1}{2} f''(x_*)(x-x_*)^2$$
$$= e^{-2x_*} - 2e^{-2x_*}(x-x_*) + 2e^{-2x_*}(x-x_*)^2$$

(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.

1^{st} order derivatives: f_x, f_y

2^{nd} order derivatives: f_{xx}, f_{xy}, f_{yx}, f_{yy}

$$f_x(x,y) = \frac{\partial f(x,y)}{\partial x} = 2xy^2$$
$$f_y(x,y) = \frac{\partial f(x,y)}{\partial y} = 2x^2y$$

$$f_{xx}(x,y) = \frac{\partial f_x(x,y)}{\partial x} = \frac{\partial}{\partial x}(2xy^2) = 2y^2$$
$$f_{xy}(x,y) = \frac{\partial f_x(x,y)}{\partial y} = \frac{\partial}{\partial y}(2xy^2) = 4xy$$
$$f_{yx}(x,y) = \frac{\partial f_y(x,y)}{\partial x} = \frac{\partial}{\partial x}(2x^2y) = 4xy$$
$$f_{yy}(x,y) = \frac{\partial f_y(x,y)}{\partial y} = \frac{\partial}{\partial y}(2x^2y) = 2x^2$$

(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:

$f(x,y) = x^2y^2$

$\cancel{\Delta x^2 \Delta y^2}$

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

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 $\frac{1}{2}$?

note: $f_{xy} = f_{yx}$

$$f_{xy}(x,y) = \frac{1}{2} f_{xy}(x,y) + \frac{1}{2} f_{yx}(x,y)$$

Plug back in:

$$
\begin{aligned}
f(x+\Delta x, y+\Delta y) &= f(x_*, y_*)\\
&\quad + f_x(x_*, y_*)\Delta x + f_y(x_*, y_*)\Delta y\\
&\quad + \frac{1}{2} f_{xx}(x_*, y_*)(\Delta x^2) + \frac{1}{2} f_{yy}(x_*, y_*)(\Delta y)^2\\
&\quad + \frac{1}{2} f_{yx}(x_*, y_*)(\Delta y)(\Delta x) + \frac{1}{2} f_{xy}(x_*, y_*)(\Delta x)(\Delta y)
\end{aligned}
$$

(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:

$f_1 \quad f_2 \quad f_n$

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

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

$$H_{\vec{x}}f = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\vec{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_n^2} \end{bmatrix} = \left.\underbrace{\begin{bmatrix} f_{x_1x_1}(\vec{x}) & \cdots & f_{x_nx_1}(\vec{x}) \\ \vdots & \ddots & \vdots \\ f_{x_1x_n}(\vec{x}) & \cdots & f_{x_nx_n}(\vec{x}) \end{bmatrix}}_{n}\right\} n \tag{14}$$

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, $f(x,y) = x^2y^2$.

$$H_{(x,y)}f = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix}$$

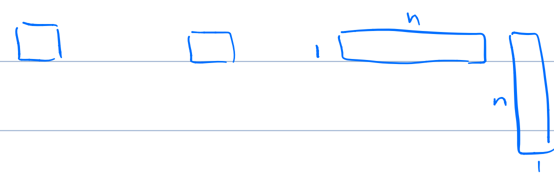
$$= \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

$$= \begin{bmatrix} 2y^2 & 4xy \\ 4xy & 2x^2 \end{bmatrix}$$

(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.

1st order approximation:

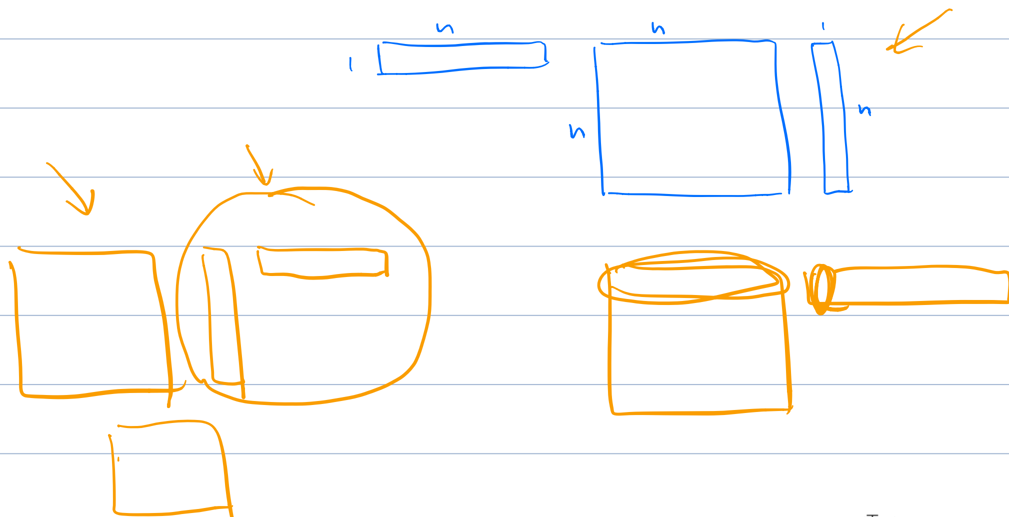
$$f(\vec{x}_* + \Delta\vec{x}) \approx f(\vec{x}_*) + \left[D_{\vec{x}} f|_{\vec{x}_*}\right](\Delta\vec{x})$$



2nd order approximation:

$$f(\vec{x}_* + \Delta x) \approx f(\vec{x}_*) + \left[D_{\vec{x}} f|_{\vec{x}_*}\right](\Delta\vec{x})$$

$$+ \frac{1}{2}(\Delta\vec{x})^T \left[H_{\vec{x}} f|_{\vec{x}_*}\right](\Delta\vec{x})$$



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

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

Here, assume that $\vec{x}$ is just some given vector — a constant 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.

$$f_{w_i}(\vec{w}) = \frac{\partial}{\partial w_i} e^{x \cdot \vec{w}}$$
$$= x_i \, e^{\vec{x}^T \vec{w}}$$
$$= x_i \, f(\vec{w})$$

$$f_{w_i w_j} = \frac{\partial f_{w_i}(\vec{w})}{\partial w_j}$$

$$= \frac{\partial}{\partial w_j}\left(x_i \, f(\vec{w})\right)$$

$$= x_i x_j f(\vec{w})$$

$$D_{\vec{w}} f = \left[f_{w_1}(\vec{w}) \quad \cdots \quad f_{w_n}(\vec{w}) \right]$$
$$= f(\vec{w}) \, \vec{x}^T$$

$$H_{\vec{w}} f = \begin{bmatrix} f_{w_1 w_1}(\vec{w}) & \cdots & f_{w_1 w_n}(\vec{w}) \\ \vdots & & \vdots \\ f_{w_n w_1}(\vec{w}) & & f_{w_n w_n}(\vec{w}) \end{bmatrix}$$

$$= f(\vec{w}) \, \vec{x} \, \vec{x}^T$$

$$f(\vec{w}_* + \Delta w) \approx f(\vec{w}_*) + D_{\vec{w}} f \big|_{\vec{w}_*} (\Delta \vec{w})$$
$$+ \frac{1}{2} (\Delta w)^T \left[H_{\vec{w}} f \big|_{w_*} \right] (\Delta \vec{w})$$
$$= \cdots$$
$$= e^{\vec{x}^T \vec{w}_*} \left(1 + \vec{x}^T (\Delta w) + \frac{1}{2} \left(\vec{x}^T (\Delta \vec{w}) \right)^2 \right)$$

$\vec{x}_i^T \vec{w}$

$(f_i = e \quad)$

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

$f = \sum_i f_i$

$$f(\vec{w}_* + \Delta\vec{w}) \approx \sum_{i=1}^{m} e^{\vec{x}_i^\top \vec{w}} \left(1 + \vec{x}_i(\Delta\vec{w}) + \frac{1}{2}(\vec{x}_i^\top(\Delta w))^2\right)$$

(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?

linear approximate scalar-valued f:

$$f(\vec{w}_* + \Delta w) \approx f(\vec{w}_*) + f'(\vec{w}_*)\Delta w$$

let $f'(\vec{w}_* + \Delta w)$ to be "f":

$$f'(\vec{w}_* + \Delta\vec{w}) \approx f'(w_*) + f''(w_*)\Delta w$$

but now "f" is $D_{\vec{w}} f$ $\boxed{}$

$$D_{\vec{w}} f \approx D_{\vec{w}} f|_{\vec{w}_*} + ?$$

$$\boxed{} \approx \boxed{} + \boxed{}$$

$$D_{\vec{w}} f \approx D_{\vec{w}} f \big|_{\vec{w}_*} + (\Delta w)^T \left[H_{\vec{w}} f \big|_{\vec{w}_*} \right]$$