Discussion 11B

The following note is useful for this discussion: Note 18.

1. Linear Approximation

A common way to approximate a nonlinear function is to perform linearization near a point. In the case of a one-dimensional function $f(x)$, the linear approximation of $f(x)$ at a point $x_\star$ is given by

$$\widehat{f}(x; x_\star) = f(x_\star) + f'(x_\star) \cdot (x - x_\star), \tag{1}$$

where $f'(x_\star) := \frac{\mathrm{d}f}{\mathrm{d}x}(x_\star)$ is the derivative of $f(x)$ at $x = x_\star$.

Keep in mind that wherever we see $x_\star$, this denotes a *constant value* or operating point.

We can evaluate the accuracy of our approximation by calculating the approximation error, namely $\left| f(x) - \widehat{f}(x; x_\star) \right|$.

Suppose we have the single-variable function $f(x) = x^3 - 3x^2$. We can plot the function $f(x)$ as follows:

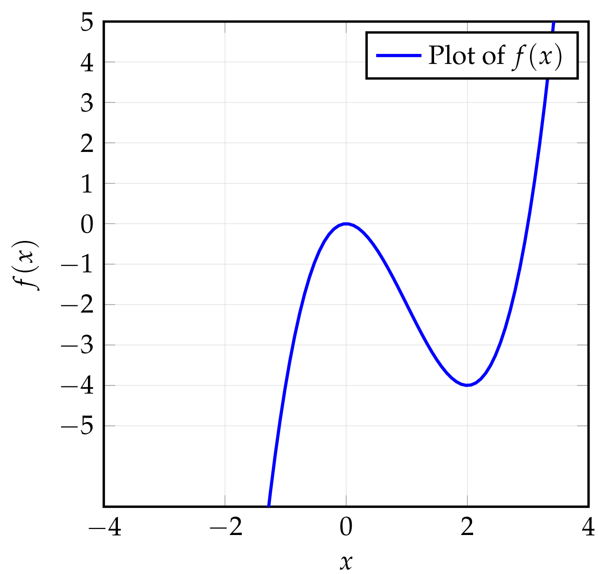


Figure 1: Plot of $f(x) = x^3 - 3x^2$

(a) **Write the linear approximation of the function around an arbitrary point $x_\star$.** **Solution:**

$$\begin{aligned}\widehat{f}(x; x_\star) &= f(x_\star) + f'(x_\star) \cdot (x - x_\star) &(2)\\ &= f(x_\star) + (3x_\star^2 - 6x_\star) \cdot (x - x_\star) &(3)\end{aligned}$$

(b) **Using the expression above, linearize the function around the point $x_\star = 1.5$. Draw the linearization into the plot in fig. 1. Then evaluate the accuracy of the linear approximation at $x = 1.7$ and $x = 2.5$.** Does the difference in accuracy make sense, based on the plot?

Solution:

$$\widehat{f}(x; x_\star) = f(1.5) + \left(3 \cdot 1.5^2 - 6 \cdot 1.5\right) \cdot (x - 1.5) \tag{4}$$
$$= -3.375 + (-2.25) \cdot (x - 1.5) \tag{5}$$

The plot is shown below:

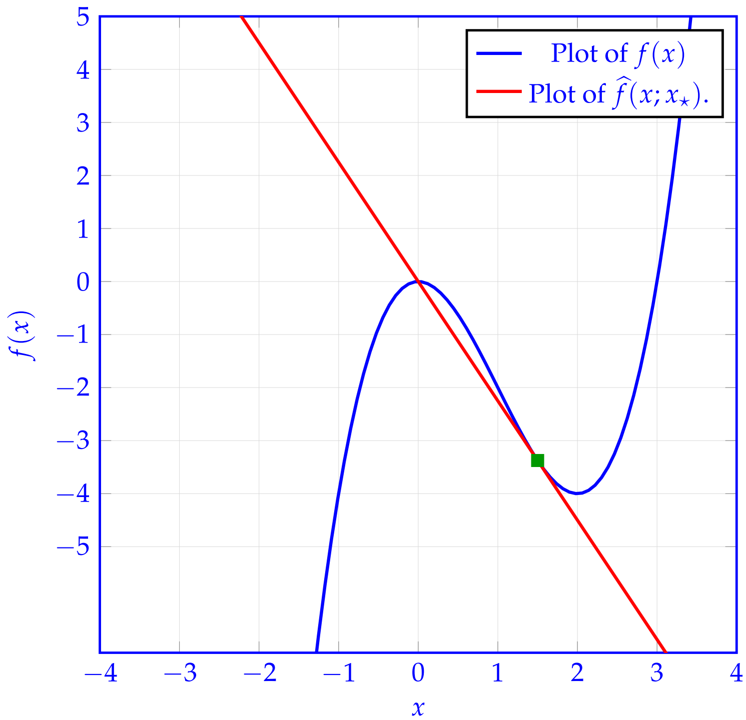


Figure 2: Plot of $\widehat{f}(x; x_\star)$ and $f(x)$

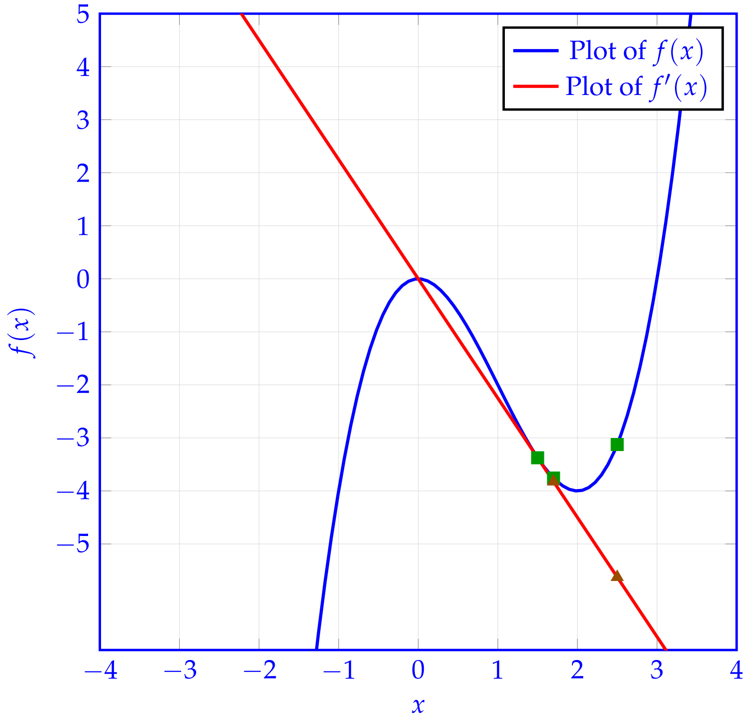
To evaluate the accuracy of $\widehat{f}(x; x_\star)$, we can compute $\left|\widehat{f}(x; x_\star) - f(x)\right|$. At $x = 1.7$:

$$\widehat{f}(1.7; x_\star) = -3.375 + (-2.25) \cdot (1.7 - 1.5) \tag{6}$$
$$= -3.375 - 0.45 \tag{7}$$
$$= -3.825 \tag{8}$$

and $f(1.7) = 1.7^3 - 3 \cdot 1.7^2 = -3.757$. Hence, $\left|\widehat{f}(1.7; x_\star) - f(1.7)\right| = 0.068$. Now, at $x = 2.5$:

$$\widehat{f}(2.5; x_\star) = -3.375 + (-2.25) \cdot (2.5 - 1.5) \tag{9}$$
$$= -3.375 - 2.25 \tag{10}$$
$$= -5.625 \tag{11}$$

and $f(2.5) = 2.5^3 - 3 \cdot 2.5^2 = -3.125$. Hence, $\left|\widehat{f}(2.5; x_\star) - f(2.5)\right| = 2.5$. We see that the error at $x = 2.5$ is about 3 times higher than the error at $x = 1.7$. We can plot the points $x = 1.7$ and $x = 2.5$ on fig. 2 to explicitly see this difference in errors:



Now, we can extend this to higher dimensional functions. In the case of a two-dimensional function $f(x,y)$, the linear approximation of $f(x,y)$ at a point $(x_\star, y_\star)$ is given by

$$\widehat{f}(x,y;x_\star,y_\star) = f(x_\star,y_\star) + \frac{\partial f}{\partial x}(x_\star,y_\star)\cdot(x-x_\star) + \frac{\partial f}{\partial y}(x_\star,y_\star)\cdot(y-y_\star). \tag{12}$$

where $\frac{\partial f}{\partial x}(x_\star,y_\star)$ is the partial derivative of $f(x,y)$ with respect to x at the point $(x_\star,y_\star)$, and similarly for $\frac{\partial f}{\partial y}(x_\star,y_\star)$

(c) Now, let's see how we can find partial derivatives. When we are given a function $f(x,y)$, we calculate the partial derivative of f with respect to x by fixing y and taking the derivative with respect to x. **Given the function $f(x,y) = x^2y$, find the partial derivatives $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial y}$.**

Solution: We have

$$\frac{\partial f(x,y)}{\partial x} = 2xy \tag{13}$$

$$\frac{\partial f(x,y)}{\partial y} = x^2. \tag{14}$$

(d) **Write out the linear approximation of f near $(x_\star,y_\star)$.**

Solution: Based on the formula in eq. (12), we can write that:

$$\widehat{f}(x,y;x_\star,y_\star) = f(x_\star,y_\star) + 2x_\star y_\star\cdot(x-x_\star) + x_\star^2\cdot(y-y_\star) \tag{15}$$

(e) We want to see if the approximation arising from linearization of this function is reasonable for a point close to our point of evaluation. Suppose we want to evaluate the accuracy of our approximation at some point $(x_\star+\delta, y_\star+\delta)$, where $x_\star = 2$ and $y_\star = 3$. **Find the accuracy of this approximation in terms of δ. What if $\delta = 0.01$?**

Solution: The true value of $f(2+\delta, 3+\delta)$ is

$$f(2+\delta,3+\delta) = (2+\delta)^2(3+\delta) = (4+4\delta+\delta^2)(3+\delta) = 12+16\delta+7\delta^2+\delta^3 \tag{16}$$

On the other hand, our approximation is

$$\widehat{f}(2+\delta,3+\delta;x_\star,y_\star) = f(2,3) + 2\cdot 2\cdot 3\cdot\delta + 2^2\cdot\delta = 12+16\delta \tag{17}$$

So the approximation error is

$$\left|f(2+\delta,3+\delta) - \widehat{f}(2+\delta,3+\delta;x_\star,y_\star)\right| = \left|7\delta^2+\delta^3\right| \tag{18}$$

When δ is sufficiently small (i.e. close to 0), the δ^2 and δ^3 terms become very small, and hence our approximation is reasonable. For $\delta = 0.01$, the approximation error is $\left|7\delta^2+\delta^3\right| = 0.000701$.

(f) Suppose we have now a scalar-valued function $f(\vec{x},\vec{y})$, which takes in vector-valued arguments $\vec{x}\in\mathbb{R}^n$, $\vec{y}\in\mathbb{R}^k$ and outputs a scalar $\in\mathbb{R}$. That is, $f(\vec{x},\vec{y})$ is $\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$.

One way to linearize the function f is to do it for every single element in $\vec{x} = \begin{bmatrix}x_1 & x_2 & \ldots & x_n\end{bmatrix}^\top$ and $\vec{y} = \begin{bmatrix}y_1 & y_2 & \ldots & y_k\end{bmatrix}^\top$. Then, when we are looking at x_i or y_j, we fix everything else as constant. This would give us the linear approximation

$$f(\vec{x},\vec{y}) \approx f(\vec{x}_\star,\vec{y}_\star) + \sum_{i=1}^{n} \frac{\partial f(\vec{x},\vec{y})}{\partial x_i}\Big|_{(\vec{x}_\star,\vec{y}_\star)}(x_i - x_{i,\star}) + \sum_{j=1}^{k} \frac{\partial f(\vec{x},\vec{y})}{\partial y_j}\Big|_{(\vec{x}_\star,\vec{y}_\star)}(y_j - y_{j,\star}). \tag{19}$$

In order to simplify this equation, we can define the following two vector quantities:

$$J_{\vec{x}}f = \begin{bmatrix}\frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n}\end{bmatrix} \tag{20}$$

$$J_{\vec{y}}f = \begin{bmatrix}\frac{\partial f}{\partial y_1} & \cdots & \frac{\partial f}{\partial y_k}\end{bmatrix} \tag{21}$$

First, how can we "vectorize" eq. (19) using $J_{\vec{x}}f$ and $J_{\vec{y}}f$? Next, assume that $n = k$ and we define the function $f(\vec{x},\vec{y}) = \vec{x}^\top\vec{y} = \sum_{i=1}^{k} x_i y_i$. Find $J_{\vec{x}}f$ and $J_{\vec{y}}f$ for this specific f.
(HINT: For vectorizing, think about replacing the summations as the multiplication of a row and column vector. What would these vectors be?)

Solution: To vectorize eq. (19), we can try to replace the summations with a dot product. That is, if we were to multiply the row vector $\begin{bmatrix}\frac{\partial f}{\partial x_1}\Big|_{(\vec{x}_\star,\vec{y}_\star)} & \cdots & \frac{\partial f}{\partial x_n}\Big|_{(\vec{x}_\star,\vec{y}_\star)}\end{bmatrix} = J_{\vec{x}}f\Big|_{(\vec{x}_\star,\vec{y}_\star)}$ with the column vector $\begin{bmatrix}x_1 - x_{1,\star}\\ \vdots \\ x_n - x_{n,\star}\end{bmatrix} = \vec{x}-\vec{x}_\star$, then we would get the same summation (and similarly for y_j). Writing this more compactly,

$$\widehat{f}(\vec{x},\vec{y};\vec{x}_\star,\vec{y}_\star) = f(\vec{x}_\star,\vec{y}_\star) + J_{\vec{x}}f\Big|_{(\vec{x}_\star,\vec{y}_\star)}(\vec{x}-\vec{x}_\star) + J_{\vec{y}}f\Big|_{(\vec{x}_\star,\vec{y}_\star)}(\vec{y}-\vec{y}_\star) \tag{22}$$

Now, for the specific $f(\vec{x},\vec{y})$ in this problem, we apply the definition (and write out the given function explicitly as $x_1y_1 + x_2y_2 + \ldots + x_ky_k$) to obtain:

$$J_{\vec{x}}f = \begin{bmatrix}y_1 & y_2 & \cdots & y_k\end{bmatrix} = \vec{y}^\top \tag{23}$$

and

$$J_{\vec{y}}f = \begin{bmatrix}x_1 & x_2 & \cdots & x_k\end{bmatrix} = \vec{x}^\top \tag{24}$$

(g) Following the above part, **find the linear approximation of $f(\vec{x},\vec{y})$ near $\vec{x}_\star = \begin{bmatrix}1\\2\end{bmatrix}$ and $\vec{y}_\star = \begin{bmatrix}-1\\2\end{bmatrix}$. Recall that $f(\vec{x},\vec{y}) = \vec{x}^\top\vec{y} = \sum_{i=1}^{k} x_i y_i$.**

Solution: From the solution in the previous part, we can write

$$\widehat{f}(\vec{x},\vec{y};\vec{x}_\star,\vec{y}_\star) = f(\vec{x}_\star,\vec{y}_\star) + J_{\vec{x}}f\Big|_{(\vec{x}_\star,\vec{y}_\star)}(\vec{x}-\vec{x}_\star) + J_{\vec{y}}f\Big|_{(\vec{x}_\star,\vec{y}_\star)}(\vec{y}-\vec{y}_\star) \tag{25}$$

$$= \vec{x}_\star^\top\vec{y}_\star + \vec{y}_\star^\top(\vec{x}-\vec{x}_\star) + \vec{x}_\star^\top(\vec{y}-\vec{y}_\star) \tag{26}$$

Putting in $\vec{x}_\star = \begin{bmatrix}1\\2\end{bmatrix}$ and $\vec{y}_\star = \begin{bmatrix}-1\\2\end{bmatrix}$,

$$\widehat{f}(\vec{x},\vec{y};\vec{x}_\star,\vec{y}_\star) = 3 + \begin{bmatrix}-1\\2\end{bmatrix}^\top\vec{x} - 3 + \begin{bmatrix}1\\2\end{bmatrix}^\top\vec{y} - 3 \tag{27}$$

$$= \begin{bmatrix}-1\\2\end{bmatrix}^\top\vec{x} + \begin{bmatrix}1\\2\end{bmatrix}^\top\vec{y} - 3 \tag{28}$$

These linearizations are important for us because we can do many easy computations using linear functions.

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.