Mathematical Optimization
YTU 2019

Chathuranga Weeraddana

August 2019

Lecture 8: Math Review: Part 7

Derivative

- suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and x is an element in $\mathbb{R}^n$ and in the interior of domain f
- the derivative of f at x is a row vector denoted $Df(x)$ and its jth element is given by

$$Df(x)_{1,j} = \frac{\partial f(x)}{\partial x_j},$$

  provided that the partial derivative exists

- note that $j \in \{1, 2, \ldots, n\}$

Derivative

- some notations !
- recall that the derivative of f at x is $Df(x)$
- similarly, the derivative of f at x_0 is $Df(x_0)$
- when we are given the derivative at a point x, [i.e., $Df(x)$], we can easily compute the derivative at another point x_0 as follows:

$$Df(x_0) = Df(x) \Big|_{x=x_0}$$

Derivative

- if the partial derivatives exists for all $j \in \{1, 2, \ldots, n\}$, we say f is differentiable at x
- if the domain of f is open and f is differentiable at every point in its domain, we say f is differentiable

First-Order Approximation of f at x

- consider the function f_1 of z given by

$$f_1(z) = f(x) + Df(x)(z - x)$$

- f_1 is called the first-order approximation of f at (or near) x
- when z is close to x, f_1 is very close to f
- graphically interpret this phenomenon

Gradient

- gradient of a real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ at x is

$$\nabla f(x) = Df(x)^T$$

- i.e., the transpose of the derivative
- note that $\nabla f(x) \in \mathbb{R}^n$, a n-vector
- moreover, it is easy to see that

$$\nabla f(x)_j = \frac{\partial f(x)}{\partial x_j}, \quad j = 1, \ldots, n$$

- the first-order approximation f_1 of f at x is simply

$$f_1(z) = f(x) + \nabla f(x)^T (z - x)$$

Gradient

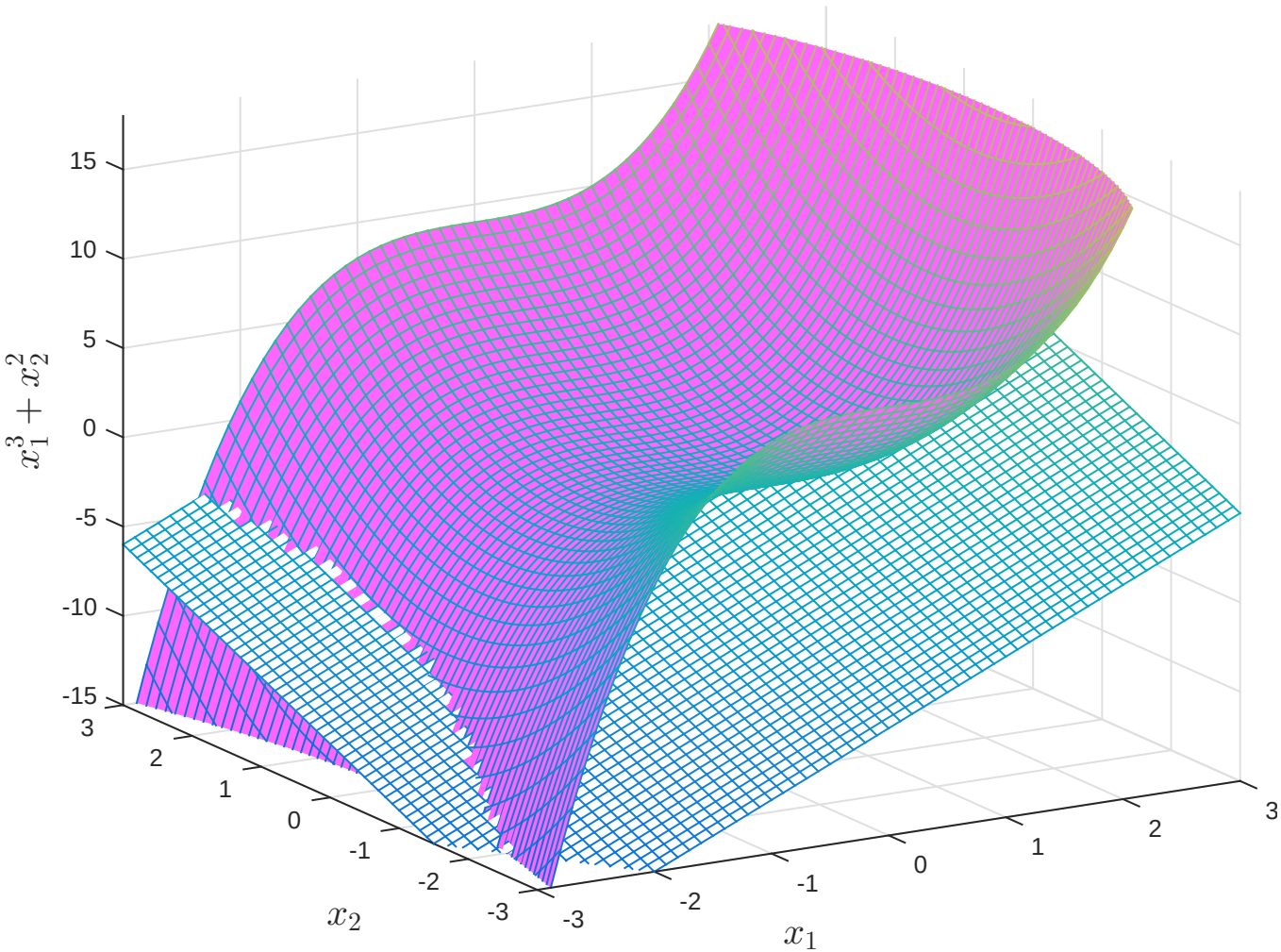
- determine the derivatives, gradients, and first-order approximations of
  - $f(x) = ||x||^2$
  - $f(x_1, x_2) = x_1^2 + 2x_2$
  - $f(x_1, x_2) = \sin(x_1) + x_2$
  - $f(x_1, x_2) = 2x_1 + 3x_2$

  at x_0

- note that

$$\nabla f(x_0) = \nabla f(x) \Big|_{x=x_0}$$

First-Order Approximation of $f(x) = x_1^3 + x_2^2$ at $x = (1, 1)$



Derivative, Gradient

- some useful results
  - $D(q^T x) = q^T, \quad q \in \mathbb{R}^m$
  - $D(q^T A x) = q^T A, \quad q \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$
  - $D(x^T A x) = x^T (A + A^T), \quad A \in \mathbb{R}^{n \times n}$
  - $D(x^T A x) = x^T (A + A^T) = 2x^T A, \quad A \in \mathbb{S}^{n \times n}$
  - $D(x^T x) = x^T (I + I^T) = 2x^T I = 2x^T$
- recall: the gradient is simply the transpose of the derivative

Second Derivative

- suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and x is an element in $\mathbb{R}^n$ and in the interior of domain f
- the second derivative of Hessian matrix of f at x, denoted $\nabla^2 f(x)$, is given by

$$\nabla^2 f(x)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n$$

provided f is twice differentiable at x

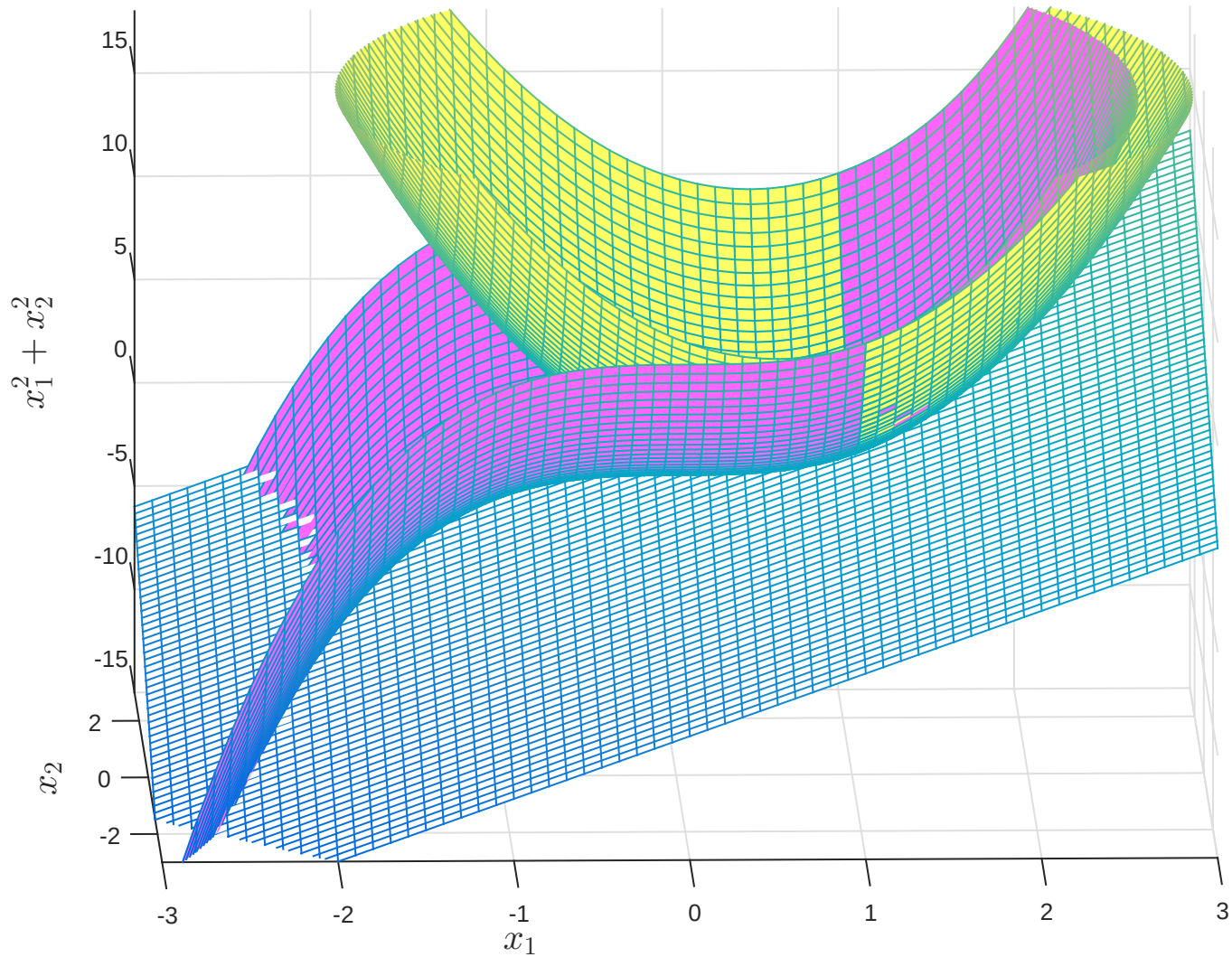
Second-Order Approximation of f at x

- consider the function f_2 of z given by

$$f_2(z) = f(x) + \nabla f(x)^T(z-x) + (1/2)(z-x)^T\nabla^2 f(x)(z-x)$$

- f_2 is called the second-order approximation of f at (or near) x
- when z is close to x, f_2 is very very close to f
- graphically interpret this phenomenon

Second-Order Approximation of $f(x) = x_1^3 + x_1^2$ at $x = (1, 1)$



Second Derivative

- determine the Hessian and second-order approximations of
  - $f(x) = ||x||^2$
  - $f(x_1, x_2) = x_1^2 + 2x_2$
  - $f(x_1, x_2) = \sin(x_1) + x_2$
  - $f(x_1, x_2) = 2x_1 + 3x_2$

  at x_0

- note that

$$\nabla^2 f(x_0) = \nabla^2 f(x) \Big|_{x=x_0}$$

Second Derivative

- some useful results
  - $\nabla^2(q^T x) = 0 \in \mathbb{R}^{m\times m}, \quad q \in \mathbb{R}^m$
  - $\nabla^2(q^T A x) = 0 \in \mathbb{R}^{n\times n}, \quad q \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$
  - $\nabla^2(x^T A x) = (A + A^T), \quad A \in \mathbb{R}^{n\times n}$
  - $\nabla^2(x^T A x) = (A + A^T) = 2A, \quad A \in \mathbb{S}^{n\times n}$
  - $\nabla^2(x^T x) = (I + I^T) = 2I$