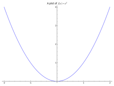
Nonlinear optimization I: Extrema of functions

Discrete Mathematics and Optimization
Bioinformatics

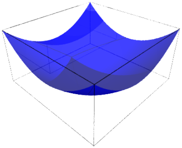
Functions of several variables

How to deal with functions of several variables?

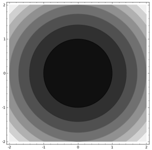
$$f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$$



- $f : \mathbb{R} \longrightarrow \mathbb{R}$: visualize with plot of f,



- $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$: visualize with plot of f,



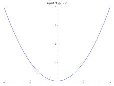
visualize level curves of f

- $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, mmmh!

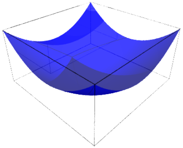
Functions of several variables

How to deal with functions of several variables?

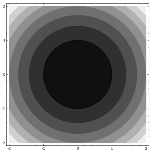
$$f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$$



- $f : \mathbb{R} \longrightarrow \mathbb{R}$: visualize with plot of f,



- $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$: visualize with plot of f,



visualize level curves of f

- $f : \mathbb{R}^{300} \longrightarrow \mathbb{R}$, mmmmmmmmmmmmh!

1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

- Reduction to one dimensional case: If $g(h) = f(\vec{x} + h\vec{e_i})$ then

$$\frac{\partial f}{\partial x_i}(\vec{x}) = g'(0)$$

1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

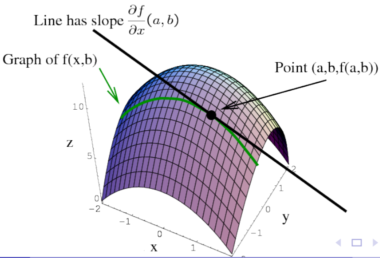
Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

- Geometric interpretation: $\frac{\partial f}{\partial x_i}(\vec{x})$ is the slope of the tangent line to the graph of $g(h) = f(\vec{x} + h\vec{e_i})$ at $h = 0$.



1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

Definition

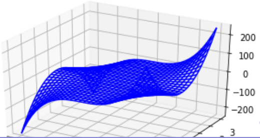
Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

- How to compute: Think of all variables but the one we consider and use differentiation rules of one variable.

$$f(x,y) = x^3y^2 : \frac{\partial f}{\partial x}(x,y) = 3x^2y^2$$
$$\frac{\partial f}{\partial y}(x,y) = 2x^3y$$



1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

- How to compute: Think of all variables but the one we consider and use differentiation rules of one variable.

$$f(x, y, z) = e^{xy} \cos z$$

$$\frac{\partial f}{\partial x}(x, y, z) = ye^{xy} \cos z$$
$$\frac{\partial f}{\partial y}(x, y, z) = xe^{xy} \cos z$$
$$\frac{\partial f}{\partial z}(x, y, z) = -e^{xy} \sin z$$

1. Partial derivatives

How to extend the tools of calculus to functions of several variables?

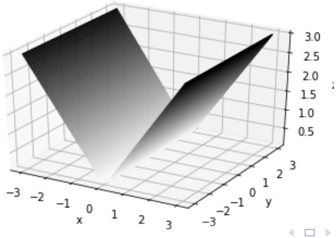
Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The partial derivative of f at $\vec{x} \in D$ with respect to the i–th coordinate is

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e_i}) - f(\vec{x})}{h}$$

if the limit exists.

- Why should it fail to exist? The function $g(h) = f(\vec{x} + h\vec{e_i})$ may be not differentiable at 0: not continuous, not smooth,...



2. The gradient

Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = (\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}))$$

if all the partial derivatives exist.

2. The gradient

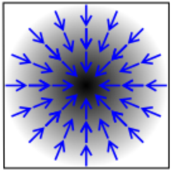
Definition

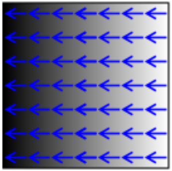
Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = \left(\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x})\right)$$

if all the partial derivatives exist.

- If the gradient exists and is nonzero, then it is perpendicular to the level sets of f, pointing at the direction of maximum (local) increasing rate of f.





2. The gradient

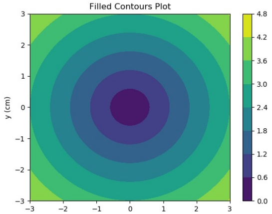
Definition

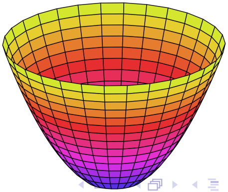
Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = (\frac{\partial f}{\partial x_1}(\vec{x}), \dots, \frac{\partial f}{\partial x_n}(\vec{x}))$$

if all the partial derivatives exist.

- If the gradient exists and it is nonzero, then it is perpendicular to the level sets of f, pointing at the direction of maximum (local) increasing rate of f.
  $f(x, y) = x^2 + y^2$
  $\nabla f(x, y) = (2x, 2y),\ \nabla f(1, 1) = (2, 2)$





2. The gradient

Definition

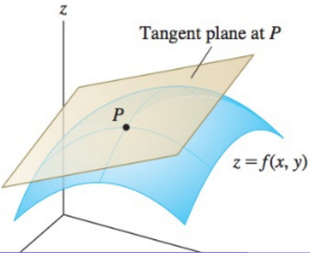
Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = \left(\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x})\right)$$

if all the partial derivatives exist.

- If the gradient exists at $\vec{x}$ one gets the linear approximation of f at $\vec{x}$

$$f(\vec{x} + \vec{h}) \approx f(\vec{x}) + \nabla f(\vec{x}) \cdot \vec{h}$$



2. The gradient

Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = (\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}))$$

if all the partial derivatives exist.

- If the gradient exists at $\vec{x}$ one gets the linear approximation of f at $\vec{x}$

$$f(\vec{x} + \vec{h}) \approx f(\vec{x}) + \nabla f(\vec{x}) \cdot \vec{h}$$

$$f(x,y) = x^2 + y^2, \nabla f(x,y) = (2x, 2y)$$

$$\nabla f(1,2) = (2,4)$$

$$f(1+h, 2+k) \approx f(1,2) + \nabla f(1,2) \cdot (h,k) = 5 + 2h + 4k$$

2. The gradient

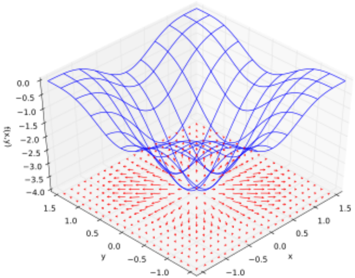
Definition

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The gradient of f at $\vec{x} \in D$ is the vector in $\mathbb{R}^n$

$$\nabla f(\vec{x}) = (\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}))$$

if all the partial derivatives exist.

- If the gradient exists and is zero at a point $\vec{x}$ then x is a critical point of f: all curves on the graph of f through $\vec{x}$ have a critical point in $\vec{x}$.



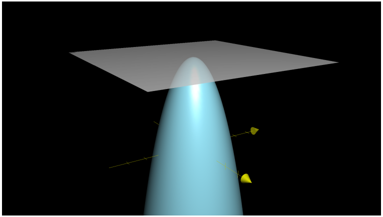
3. Extrema of functions

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. If f has a local extremum at $\vec{x} \in D$ and ∇f exists at $\vec{x}$ then

$$\nabla f(\vec{x}) = 0.$$

- The point $\vec{x}$ is a local maximum of f if $f(\vec{x}) \geq f(\vec{x} + \vec{h})$ for every ' small' $\vec{h} \in \mathbb{R}^n$.
  Then all curves on the graph of f through $\vec{x}$ have a local maximum at $\vec{x}$.



3. Extrema of functions

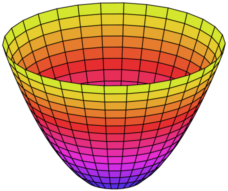
Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. If f has a local extremum at $\vec{x} \in D$ and ∇f exists at $\vec{x}$ then

$$\nabla f(\vec{x}) = 0.$$

- The point $\vec{x}$ is a local minimum of f if $f(\vec{x}) \leq f(\vec{x} + \vec{h})$ for every ' small' $\vec{h} \in \mathbb{R}^n$.
  Then all curves on the graph of f through $\vec{x}$ have a local minimum at $\vec{x}$.

$$f(x, y) = x^2 + y^2,\ \nabla f(x, y) = 0 \ \Rightarrow (x, y) = (0, 0)$$



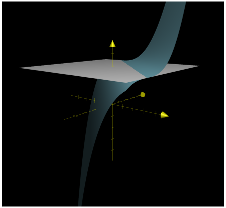
3. Extrema of functions

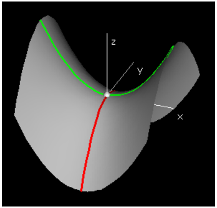
Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. If f has a local extremum at $\vec{x} \in D$ and ∇f exists at $\vec{x}$ then

$$\nabla f(\vec{x}) = 0.$$

- We may have $\nabla f(\vec{x}) = 0$ in points which are not local extrema.





How to decide the nature of a critical point?

4. The Hessian

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist in all points in a neighborhood of $\vec{x} \in D$. The second derivative of f with respect to x_i and x_j at $\vec{x}$ is

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\frac{\partial f}{\partial x_i}}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\vec{x} + h\vec{e_j}) - \frac{\partial f}{\partial x_i}(\vec{x})}{h}$$

if the limit exists.

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

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

$$\frac{\partial^2 f}{\partial x^2} = 2x, \ \frac{\partial^2 f}{\partial y^2} = 2y, \ \frac{\partial^2 f}{\partial x \partial y} = 2x + 2y$$

4. The Hessian

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist in all points in a neighborhood of $\vec{x} \in D$. The second derivative of f with respect to x_i and x_j at $\vec{x}$ is

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\frac{\partial f}{\partial x_i}}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\vec{x} + h\vec{e_j}) - \frac{\partial f}{\partial x_i}(\vec{x})}{h}$$

if the limit exists.

- If all second partial derivatives of f exist in $\vec{x} \in D$ then the Hessian matrix of f at $\vec{x}$ is

$$Hf(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{x}) & \frac{\partial^2 f}{\partial x_1 x_2}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 x_n}(\vec{x}) \\ \frac{\partial^2 f}{\partial x_2 x_1}(\vec{x}) & \frac{\partial^2 f}{\partial x_2^2}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 x_n}(\vec{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1}(\vec{x}) & \frac{\partial^2 f}{\partial x_n x_2}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\vec{x}) \end{pmatrix}$$

4. The Hessian

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist in all points in a neighborhood of $\vec{x} \in D$. The second derivative of f with respect to x_i and x_j at $\vec{x}$ is

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\frac{\partial f}{\partial x_i}}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\vec{x} + h\vec{e_j}) - \frac{\partial f}{\partial x_i}(\vec{x})}{h}$$

if the limit exists.

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

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

$$Hf(x,y) = \begin{pmatrix} 2y & 2x+2y \\ 2x+2y & 2x \end{pmatrix}, \ Hf(1,-1) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

4. The Hessian

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist in all points in a neighborhood of $\vec{x} \in D$. The second derivative of f with respect to x_i and x_j at $\vec{x}$ is

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\frac{\partial f}{\partial x_i}}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\vec{x} + h\vec{e_j}) - \frac{\partial f}{\partial x_i}(\vec{x})}{h}$$

if the limit exists.

- If the second partial derivatives are continuous functions in a neighborhood of x then

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})$$

  The Hessian matrix in this case is symmetric.
- We denote by $\mathcal{C}^2(D)$ the class of functions $f : D \subset \mathbb{R}^n \to \mathbb{R}$ which have all second derivatives in D and are continuous functions.

4. The Hessian

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist in all points in a neighborhood of $\vec{x} \in D$. The second derivative of f with respect to x_i and x_j at $\vec{x}$ is

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}) = \frac{\frac{\partial f}{\partial x_i}}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\vec{x} + h\vec{e_j}) - \frac{\partial f}{\partial x_i}(\vec{x})}{h}$$

if the limit exists.

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2(D)$. Then

$$f(\vec{x} + \vec{h}) \approx f(\vec{x}) + \nabla f(\vec{x}) \cdot \vec{h} + \frac{1}{2}\vec{h}^T \cdot Hf(\vec{x}) \cdot \vec{h}$$

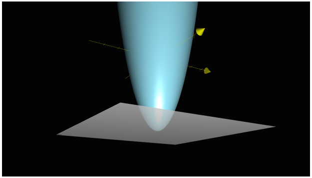
- The second order approximation in the theorem allows us to determine the nature of extrema of f at critical points.

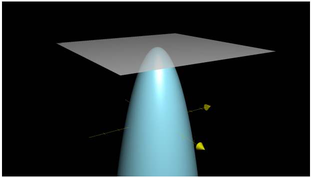
5. The nature of critical points

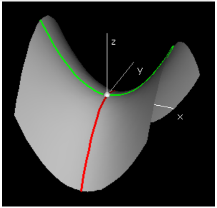
Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2(D)$. Let $\vec{x}$ be a critical point of f. Assume that $H \neq 0$.

- *If H is positive definite at $\vec{x}$ then $\vec{x}$ is a (strict) local minimum of f.*
- *If H is negative definite at $\vec{x}$ then $\vec{x}$ is a (strict) local maximum of f.*
- *if H is not definite at $\vec{x}$ then $\vec{x}$ is a saddle point of f.*







5. The nature of critical points

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2(D)$. Let $\vec{x}$ be a critical point of f. Assume that $H \neq 0$.

- *If H is positive definite at $\vec{x}$ then $\vec{x}$ is a (strict) local minimum of f.*
- *If H is negative definite at $\vec{x}$ then $\vec{x}$ is a (strict) local maximum of f.*
- *if H is not definite at $\vec{x}$ then $\vec{x}$ is a saddle point of f.*

- A simple criteria to decide if H is definite: Let Δ_k denote the determinant of the $k \times k$ principal submatrix of H, $k = 1, \ldots, n$.
  - H is positive definite if and only if $\Delta_k > 0$ for $k = 1, \ldots, n$.
  - H is negative definite if and only if $(-1)^k \Delta_k > 0$ for $k = 1, \ldots, n$.

6. Global extrema

When local extrema can be ensured to be global?

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2(D)$. Then

$$f(\vec{x}+\vec{h}) = f(\vec{x}) + \nabla f(\vec{x}) \cdot \vec{h} + \frac{1}{2} h^T \cdot H(\vec{x}+\theta\vec{h}) \cdot \vec{h},\ 0 \leq \theta \leq 1.$$

6. Global extrema

When local extrema can be ensured to be global?

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2(D)$. Then

$$f(\vec{x}+\vec{h}) = f(\vec{x}) + \nabla f(\vec{x}) \cdot \vec{h} + \frac{1}{2} h^T \cdot H(\vec{x}+\theta\vec{h}) \cdot \vec{h},\ 0 \leq \theta \leq 1.$$

- If $H(\vec{x})$ is positive definite for all $\vec{x} \in \mathbb{R}^n$ then x is a (strict) global minimum of f.
- If $H(\vec{x})$ is negative definite for all $\vec{x} \in \mathbb{R}^n$ then $\vec{x}$ is a (strict) global maximum of f.
- if $H(\vec{x})$ is not definite at $\vec{x}$ then $\vec{x}$ is a saddle point of f.

6. Global extrema

When local extrema can be ensured to be global?

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a continous function on D. If D is a closed and bounded then f has a global maximum and a global minimum.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continous function. If

$$\lim_{\|x\| \to \infty} f(x) = \infty,$$

then f has a global minimum.

6. Global extrema

When local extrema can be ensured to be global?

Theorem

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a continous function on D. If D is a closed and bounded then f has a global maximum and a global minimum.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continous function. If

$$\lim_{\|x\| \to \infty} f(x) = \infty,$$

then f has a global minimum.

- If $H(y)$ is positive definite for all $y \in D$ then x is a (strict) global minimum of f.
- If $H(y)$ is negative definite for all $y \in D$ then x is a (strict) global maximum of f.

Summary

- The tools of calculus are useful for nonlinear optimization: methods for locating extreme values of functions.
- Partial derivatives, gradient and Hessian are the several variable analogous to derivatives in one variable.
- The linear (local) approximation of a function in $\mathcal{C}^2$ is

$$f(\vec{y}) = f(\vec{x}) + \nabla f(\vec{x}) \cdot (\vec{y} - \vec{x}) + o(\|\vec{y} - \vec{x}\|)$$

- The quadratic (local) approximation of a function in $\mathcal{C}^3$ is

$$f(\vec{y}) = f(\vec{x}) + \nabla f(\vec{x}) \cdot (\vec{y} - \vec{x}) + (\vec{y} - \vec{x})^T Hf(\vec{x})(\vec{y} - \vec{x}) + o(\|\vec{y} - \vec{x}\|^2).$$

- Local extrema of a function are to be found among its critical points: $\nabla f(\vec{x}) = \mathbf{0}$.
- The nature of critical points can be elucidated by the Hessian.