6. Unconstrained optimization

- optimality
- sufficient conditions

Optimality

Optimality

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x) \quad \text{where} \quad f : \mathbf{R}^n \to \mathbf{R},\ \mathcal{S} \subseteq \mathbf{R}^n$$

A point $x^* \in \mathcal{S}$ is a

1. **global minimizer** of f if $f(x^*) \leq f(x) \quad \forall\, x \in \mathcal{S}$
2. **global maximizer** of f if $f(x^*) \geq f(x) \quad \forall\, x \in \mathcal{S}$
3. **local minimizer** of f if $f(x^*) \leq f(x) \quad \forall\, x \in \mathcal{S}$ where $\|x - x^*\|_2 \leq \epsilon$
4. **local maximizer** of f if $f(x^*) \geq f(x) \quad \forall\, x \in \mathcal{S}$ where $\|x - x^*\|_2 \leq \epsilon$

for some $\epsilon > 0$.

Optimality

x^* is a **strict** global min or max if for all $x \in \mathcal{S}$, $f(x^*) = f(x) \iff x = x^*$.

The maximizer of $f(x)$ is the minimizer of $-f(x)$. (So we can only consider mins.)

Optimal attainment

- Optimal value may not always be **attained** or even **exist**

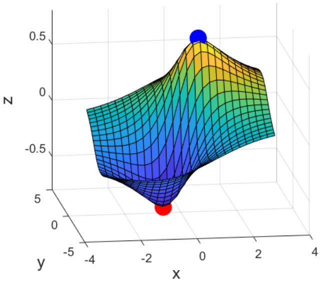
- If exists, optimal values are **always unique** even if optimal point is not

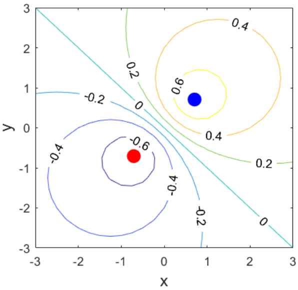
Example 1

$$\underset{x,y}{\text{minimize}} \;\; \{f(x,y) = x + y : x^2 + y^2 \leq 1\}$$

Example 2

$$\underset{x,y}{\text{minimize}} \left\{ f(x,y) = \frac{x+y}{x^2+y^2+1} : x, y \in \mathbf{R} \right\}$$

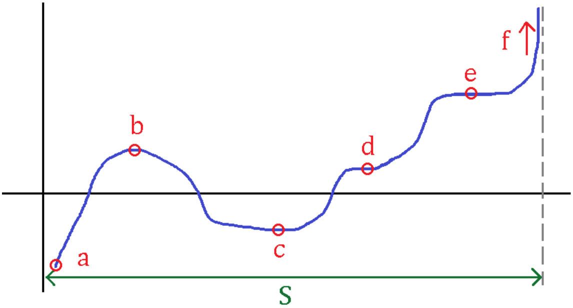




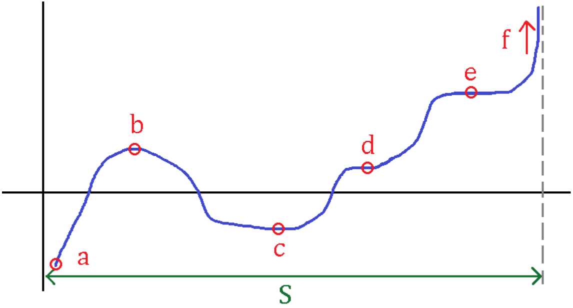
- Global maximizer $(1/\sqrt{2}, 1/\sqrt{2})$
- Global minimizer $(-1/\sqrt{2}, -1/\sqrt{2})$

Local optimality: 1-D

How can we tell if $f(x^*) < f(x)$ for all $x \in \mathcal{S}$ "close to" x^*?



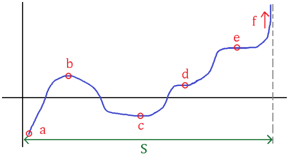
Local optimality: 1-D



Let's only consider points in the interior of $\mathcal{S}$ (b,c,d,e)

Sufficient conditions

Local optimality: 1-D



- Assume that x is in the interior of $\mathcal{S}$ (ignore boundaries for now).
- Then $x = x^*$ is a **local minimizer** (c) if

$$\underbrace{f'(x) = 0}_{f(x) \text{ is flat at } x} \qquad \text{and} \qquad \underbrace{f''(x) > 0}_{f(x) \text{ is \textbf{convex} at } x}$$

- Similarly, $x = x^*$ is a **local maximizer** (b) if

$$\underbrace{f'(x) = 0}_{f(x) \text{ is flat at } x} \qquad \text{and} \qquad \underbrace{f''(x) < 0}_{f(x) \text{ is \textbf{concave} at } x}$$

- If $f'(x) = 0$ and $f''(x) = 0$ (d,e), not enough information

  Consider $f(x) = x^3$ and $f(x) = x^4$ at $x = 0$

Motivating proof (1-D)

Gradients

For a function $f : \mathbf{R}^n \to \mathbf{R}$, the **gradient** of f at x is a vector in $\mathbf{R}^n$

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbf{R}^n$$

Example:

$$f(x) = x_1^2 + 8x_1x_2 - 2x_3^3, \qquad \nabla f(x) =$$

Gradients

For a function $f : \mathbf{R}^n \to \mathbf{R}$, the **gradient** of f at x is a vector in $\mathbf{R}^n$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbf{R}^n$$

Example:

$$f(x) = x_1^2 + 8x_1x_2 - 2x_3^3, \qquad \nabla f(x) = \begin{bmatrix} 2x_1 + 8x_2 \\ 8x_1 \\ -6x_3^2 \end{bmatrix}$$

Hessian

For a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, the **Hessian** of f at x is a symmetric matrix in $\mathbf{R}^{n\times n}$:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial f^2(x)}{\partial x_1 \partial x_1} & \frac{\partial f^2(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_1 \partial x_n} \\ \frac{\partial f^2(x)}{\partial x_2 \partial x_1} & \frac{\partial f^2(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f^2(x)}{\partial x_n \partial x_1} & \frac{\partial f^2(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbf{R}^{n\times n}$$

Example:

$$f(x) = x_1^2 + 8x_1x_2 - 2x_3^3, \quad \nabla f(x) = \begin{bmatrix} 2x_1 + 8x_2 \\ 8x_1 \\ -6x_3^2 \end{bmatrix}, \quad \nabla^2 f(x) =$$

Hessian

For a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, the **Hessian** of f at x is a symmetric matrix in $\mathbf{R}^{n\times n}$:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial f^2(x)}{\partial x_1 \partial x_1} & \frac{\partial f^2(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_1 \partial x_n} \\ \frac{\partial f^2(x)}{\partial x_2 \partial x_1} & \frac{\partial f^2(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f^2(x)}{\partial x_n \partial x_1} & \frac{\partial f^2(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial f^2(x)}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbf{R}^{n\times n}$$

Example:

$$f(x) = x_1^2 + 8x_1x_2 - 2x_3^3, \quad \nabla f(x) = \begin{bmatrix} 2x_1 + 8x_2 \\ 8x_1 \\ -6x_3^2 \end{bmatrix}, \quad \nabla^2 f(x) = \begin{bmatrix} 2 & 8 & 0 \\ 8 & 0 & 0 \\ 0 & 0 & -12x_3 \end{bmatrix}$$

Directional derivatives

- For a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, the **directional derivative** is

$$f'(x;d) = \lim_{\alpha \to 0^+} \frac{f(x+\alpha d) - f(x)}{\alpha} = \nabla f(x)^T d$$

- A function is **flat** at x^* if its directional derivative is 0 for all $d \in \mathbf{R}^n$

$$\forall d \in \mathbf{R}^n,\ f'(x;d) = 0 \iff \nabla f(x) = 0$$

- Such a point is also called a **stationary point** of f

Directional 2nd derivatives

- For a twice-differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, the **directional second derivative** is

$$f''(x;d) = \lim_{\alpha \to 0^+} \frac{f'(x+\alpha d;d) - f'(x;d)}{\alpha} = d^T \nabla^2 f(x) d$$

- A function is **convex** if for all x, y in its domain, and all $0 \le \theta \le 1$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

- If a function is convex, its directional second derivative is positive for all $d \in \mathbf{R}^n$

$$\forall x \in \mathsf{dom} f,\ \forall d \in \mathbf{R}^n, \quad f''(x;d) \ge 0$$

Sufficient conditions for optimality

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x), \qquad f : \mathbf{R}^n \to \mathbf{R}$$

- $x^* \in \mathcal{S}$ is a **minimizer** of $f(x)$ if

$$\nabla f(x^*) = 0, \qquad \underbrace{z^T \nabla^2 f(x^*) z > 0, \quad \forall z \in \mathbf{R}^n}_{\text{positive definite}}$$

- $x^* \in \mathcal{S}$ is a **maximizer** of $f(x)$ if

$$\nabla f(x^*) = 0, \qquad \underbrace{z^T \nabla^2 f(x^*) z < 0, \quad \forall z \in \mathbf{R}^n}_{\text{negative definite}}$$

- If $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ is neither positive nor negative definite, it is **indefinite**
  - x^* is a saddle point (not a minimizer or a maximizer)

Motivating proof

- If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, then for any x and $z = x - x^*$:

$$f(x) = f(x^*) + \underbrace{\nabla f(x^*)^T z}_{=0} + \underbrace{\frac{1}{2} z^T \nabla^2 f(x^*) z}_{\text{strictly positive}} + \underbrace{O(\|z\|^3)}_{\text{really small}} > f(x^*)$$

  when x is close enough to x^* (local minimum)

- If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is negative definite, then for any x and $z = x - x^*$:

$$f(x) = f(x^*) + \underbrace{\nabla f(x^*)^T z}_{=0} + \underbrace{\frac{1}{2} z^T \nabla^2 f(x^*) z}_{\text{strictly negative}} + \underbrace{O(\|z\|^3)}_{\text{really small}} < f(x^*)$$

  when x is close enough to x^* (local maximum)

Example

$$\underset{x,y}{\text{minimize}} \quad f(x,y) := \frac{x+y}{x^2+y^2+1}$$

Gradient of f:

$$\nabla f(x,y) = \frac{1}{(x^2+y^2+1)^2} \begin{bmatrix} y^2 - 2xy - x^2 + 1 \\ x^2 - 2xy - y^2 + 1 \end{bmatrix}$$

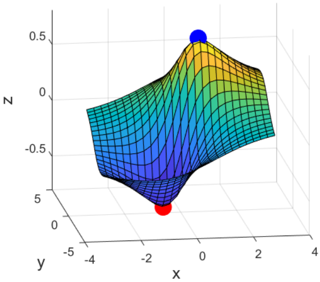
Where is $\nabla f(x,y) = 0$?

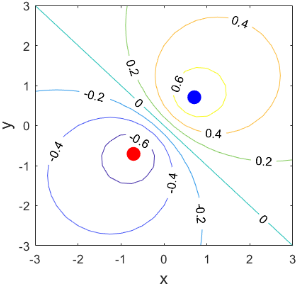
$$(x^*, y^*) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \quad \text{and} \quad (x^*, y^*) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

Hessian of f at these points:

$$\nabla^2 f(x,y) = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \text{and} \quad \nabla^2 f(x,y) = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Example





$$\underbrace{(x^*, y^*) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)}_{\text{minimum}} \quad \text{and} \quad \underbrace{(x^*, y^*) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)}_{\text{maximum}}$$

$$\underbrace{\nabla^2 f(x, y) = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{positive definite}} \quad \text{and} \quad \underbrace{\nabla^2 f(x, y) = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{negative definite}}$$

Convexity and optimality: 1-D

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x), \qquad f(x) \quad \text{is differentiable everywhere.}$$

- Suppose for some point x^* in the interior of $\mathcal{S}$, $f'(x^*) = 0$.
- Then,
  - if $f''(x^*) > 0$, x^* is a local minimum
  - if $f''(x^*) < 0$, x^* is a local maximum
  - if $f''(x^*) = 0$, x^* could be a local minimum, maximum, or saddle point.
- Example of third case:
  - $f(x) = x^4$, $x^* = 0$ is a local minimum
  - $f(x) = -x^4$, $x^* = 0$ is a local maximum
  - $f(x) = x^3$, $x^* = 0$ is a saddle point

Convexity and optimality: 1-D

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x), \qquad f(x) \quad \text{is differentiable everywhere.}$$

Claim: If $f''(x) \geq 0$ for all $x \in \mathcal{S}$, then

$$f'(x^*) = 0 \iff x^* \quad \text{is a global minimum (maybe not unique)}$$

Convexity and optimality: 1-D

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x), \qquad f(x) \quad \text{is differentiable everywhere.}$$

Claim: If $f''(x) \geq 0$ for all $x \in \mathcal{S}$, then ← We don't need strict inequality!

$$f'(x^*) = 0 \iff x^* \quad \text{is a global minimum (maybe not unique)}$$

Convexity and optimality: 1-D

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x), \qquad f(x) \quad \text{is differentiable everywhere.}$$

Claim: If $f''(x) \geq 0$ for all $x \in \mathcal{S}$, then $\leftarrow$ We don't need strict inequality!

$$f'(x^*) = 0 \iff x^* \quad \text{is a global minimum (maybe not unique)}$$

Proof: use mean value theorem from calculus.

- Suppose there exists some $\bar{x}$ where $f(\bar{x}) < f(x^*)$
- Without loss of generality, assume $\bar{x} < x^*$. (Just reverse the proof otherwise.)
- By MVT, there exists $\tilde{x} \in (\bar{x}, x^*)$ where $f'(\tilde{x}) > 0$
- By MVT, there exists $\hat{x} \in (\tilde{x}, x^*)$ where $f''(\hat{x}) < 0$
- Contradiction!