Optimization for Data Science Lecture 03: Convexity and First Order Optimality Conditions

Kimon Fountoulakis

School of Computer Science
University of Waterloo

17/09/2019

Outline

- Convex functions
- First-order optimality conditions
- Application to denoising

Parameter tuning

$$\text{minimize } \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2$$

- User defined parameter that controls the balance between regularization and fitting to the noisy image.

$\lambda = 0.1$

$\lambda = 3$

$\lambda = 10$

How to detect solutions?

$$\text{minimize } \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2$$

- How can we know if a given “x” is a minimizer for the above problem?

- Is it possible find criteria to check if a given “x” is a minimizer?

Some comments

- For the sake of generalization let's consider the following problem

$$\text{minimize } f(x)$$

- For simplicity, let's assume that the domain of f is $\mathbb{R}^n$, i.e., there are no constraints.

- Our analysis will be done for this general problem, and it will hold for our denoising problem by simply setting

$$f(x) = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2$$

Local minimizer

- We say that x is a local minimizer of $f(x)$ if
- $\exists\ \epsilon > 0$ such that for any y in the ϵ-neighborhood of x

$$\|y - x\|_2 \leq \epsilon$$

- Then

$$f(x) \leq f(y)$$

Strict local minimizer

- We say that x is a strict local minimizer of $f(x)$ if

- $\exists\ \epsilon > 0$ such that for any $y \neq x$ in the ϵ-neighborhood of x

$$\|y - x\|_2 \leq \epsilon$$

- Then

$$f(x) < f(y)$$

Global minimizer

- We say that x is a global minimizer of $f(x)$ if

- for any y we have

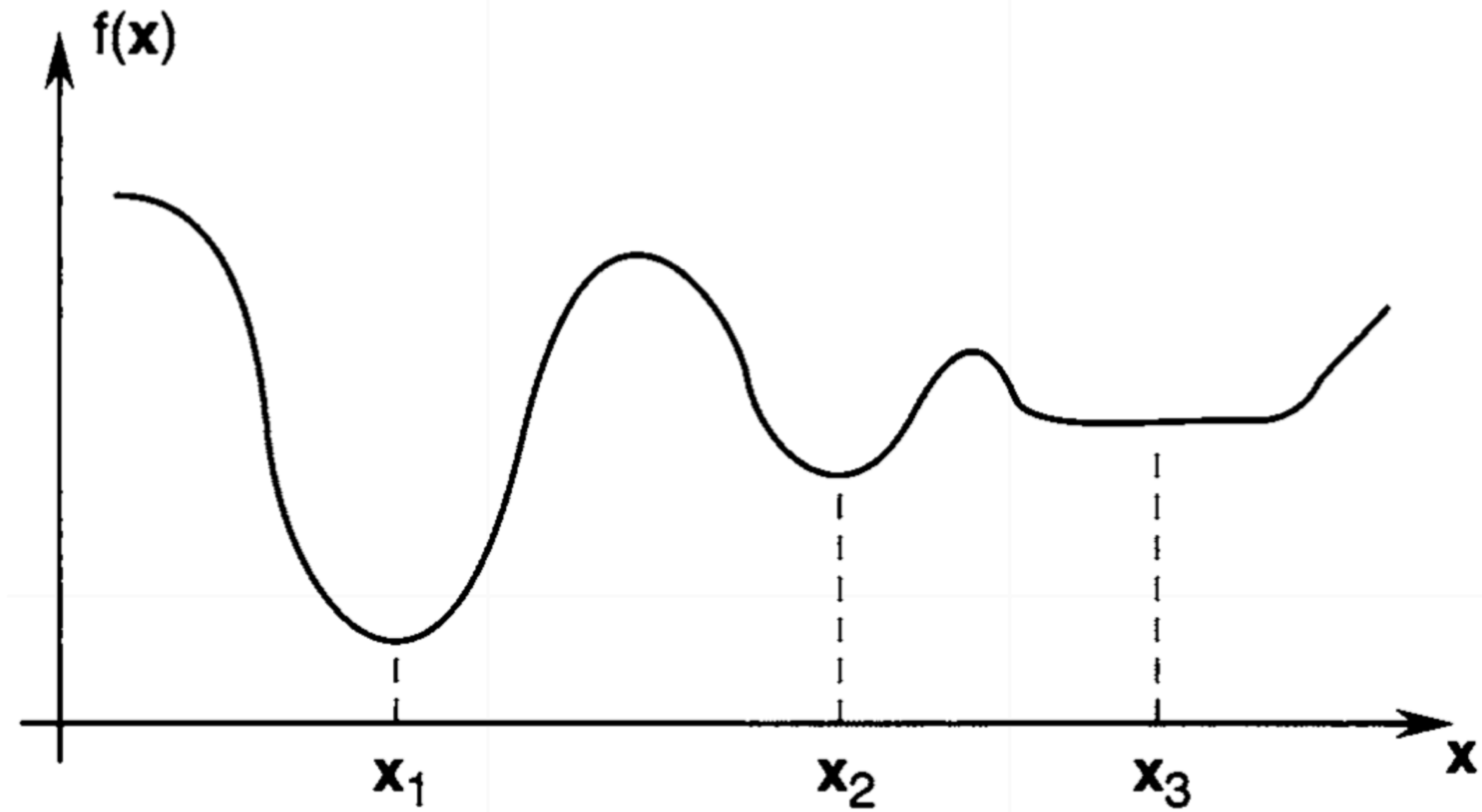
$$f(x) \leq f(y)$$

Strict global minimizer

- We say that x is a strict global minimizer of $f(x)$ if
- for any $y \neq x$ we have

$$f(x) < f(y)$$

Minimizers



x_1: strict global minimizer; x_2: strict local minimizer; x_3: local minimizer

Convex problems

- Consider the problem

$$\text{minimize} \;\; f(x)$$

- We will assume that function $f(x)$ is **convex**.

- Convexity is usually a good indicator for tractability. Minimizing convex functions is considered to be an easy task, at least easier than minimizing non-convex functions.

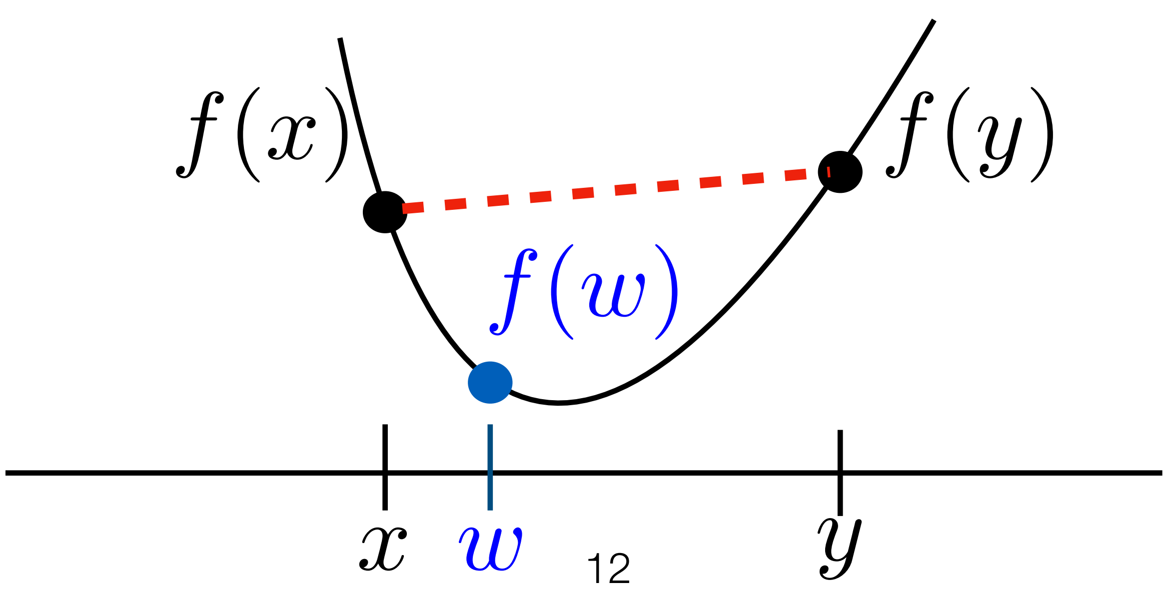
Convexity: definition

- A function $f(x)$ is convex if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$

$$\forall x \in \mathbb{R}^n, y \in \mathbb{R}^n \;\; \alpha \in [0,1]$$

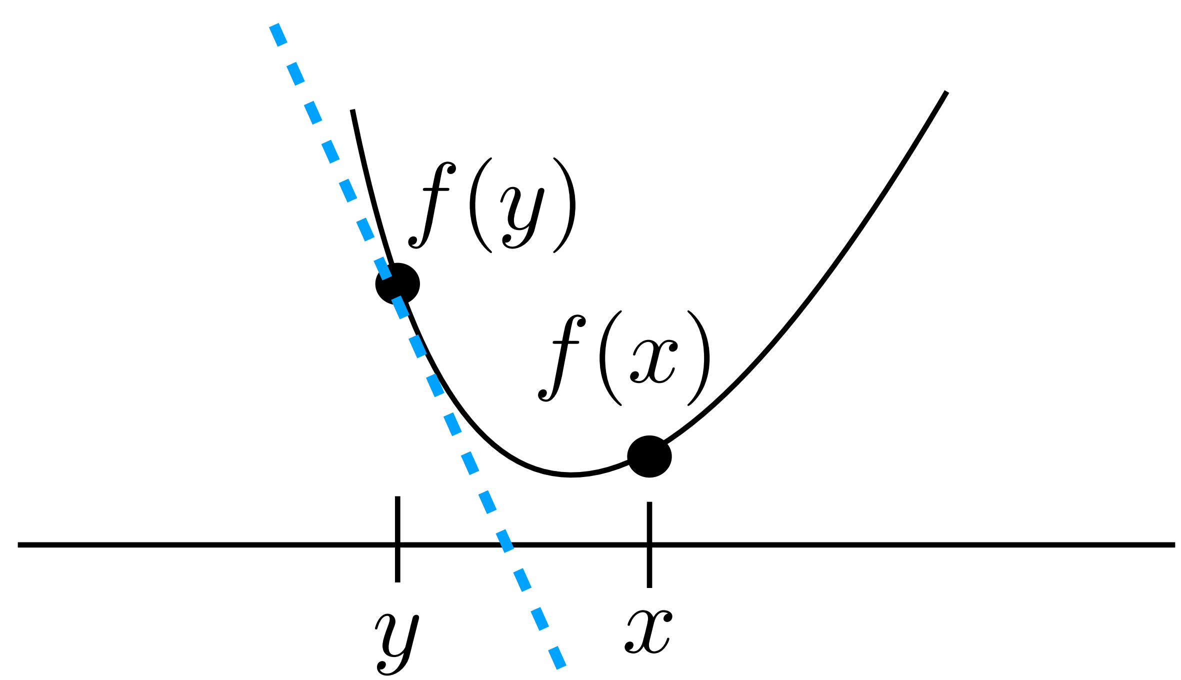
- This means that the function is always below the line between two points



Convexity: definition

- If function $f(x)$ is differentiable, then an equivalent definition for convexity is

$$f(x) \geq f(y) + \nabla f(y)^T (x - y) \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^n$$



$$f(x) \geq f(y) + \nabla f(y)^T (x - y)$$

Convexity: definition

- If function $f(x)$ is **twice** differentiable, then an equivalent definition for convexity is

$$\nabla^2 f(x) \succeq 0 \ \forall x \in \mathbb{R}^n$$

- which means that the second order derivative of $f(x)$ is positive semi-definite.

- This also means that

$$y^T \nabla^2 f(x) y \geq 0 \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^n$$

Convexity: definition

- A twice differentiable function "f" is convex if

$$y^T \nabla^2 f(x) y \geq 0 \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^n$$

- This also means that all eigenvalues of $\nabla^2 f(x)$ are non-negative for any "x".

- All definitions of convexity are equivalent.

Convexity implies nice properties

- If a function is convex, then this implies that **any local minimizer is a global minimizer**.

- If “x” is a minimizer of $f(x)$ then any other point “y” satisfies

$$f(x) \leq f(y)$$

- This makes the problem easy, since if we find one such minimizer then we can stop searching for more.

Proof

First order optimality conditions and convexity

- If a function $f(x)$ is convex and differentiable then all minimizers satisfy

$$\nabla f(x) = 0$$

- This is a necessary and sufficient condition for convex functions.

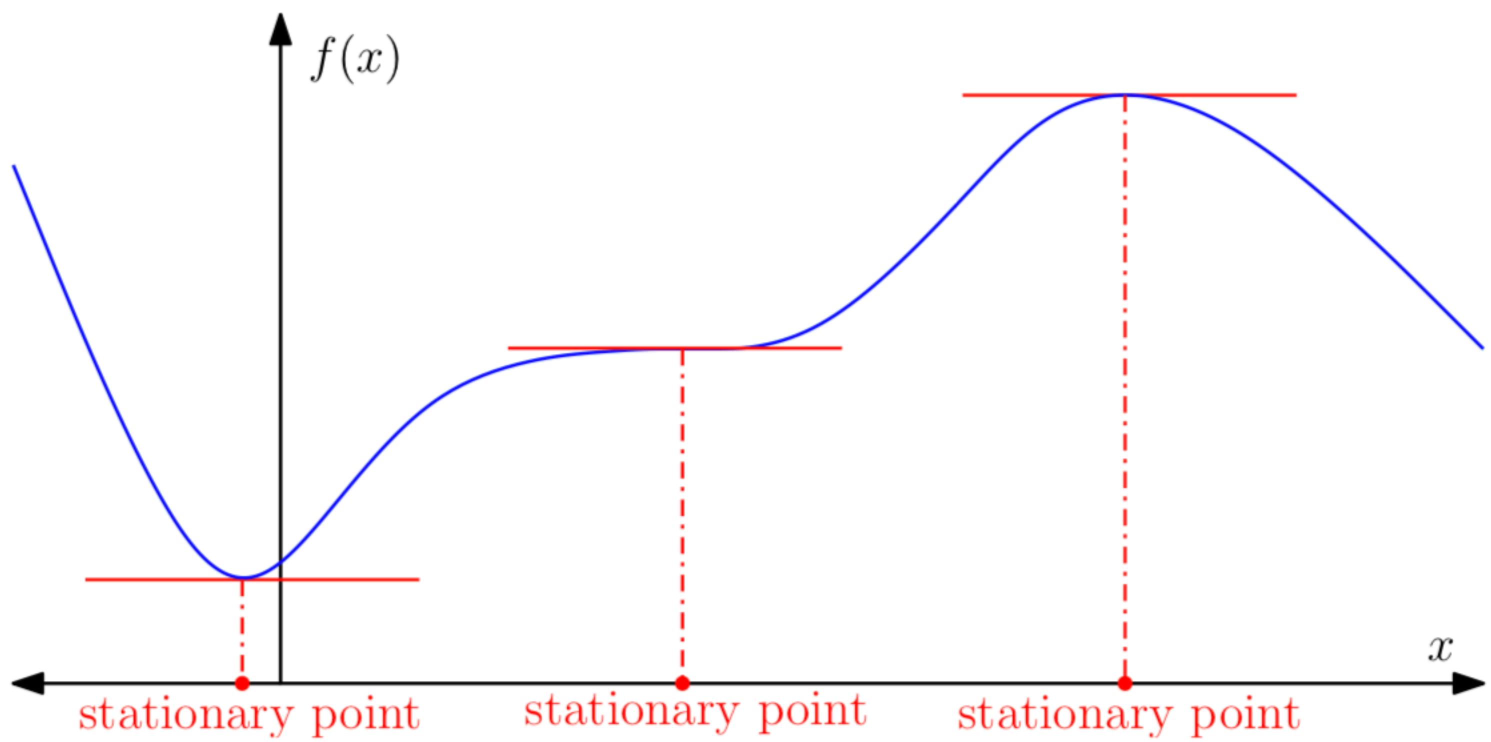
- Therefore, if we find an “x” that the gradient of the function is zero then we can stop searching for other solutions.

Proof

First order optimality conditions and convexity

- If function f is **not convex**, then $\nabla f(x) = 0$ does not imply that x is a local minimizer.

- We call points that satisfy $\nabla f(x) = 0$ stationary points.



Is the denoising problem convex?

- The objective function of the denoising problem is

$$f(x) = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2$$

- This function is twice differentiable. Let’s use the third definition of convexity, i.e., the second-order derivative must be positive semi-definite.

Is the denoising problem convex?

- The objective function of the denoising problem is

$$f(x) = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2$$

- The second order derivative is

$$\nabla^2 f(x) = \lambda \text{real}(D^* D) + I = \lambda\left(D_h^T D_h + D_v^T D_v\right) + I$$

- where $D = D_h + iD_v$

Is the denoising problem convex?

- The second order derivative is

$$\nabla^2 f(x) = \lambda \mathrm{real}(D^* D) + I = \lambda \left(D_h^T D_h + D_v^T D_v \right) + I$$

- where $D = D_h + iD_v$

- Therefore, we have that

$$\lambda y^T \left(D_h^T D_h + D_v D_v \right) y^T + y^T y = \lambda \|D_h y\|_2^2 + \lambda \|D_v y\|_2^2 + \|y\|_2^2 \geq 0 \quad \forall y \in \mathbb{R}^n$$

- which means that the denoising objective function is convex.

What are the optimality conditions of the denoising problem?

$$\nabla f(x) = \lambda \mathrm{real}(D^* D)x + x - z_{noisy} = 0$$

- This is equivalent to requiring that “x” satisfies the linear system

$$(\lambda \mathrm{real}(D^* D) + I)\, x = z_{noisy}$$