Optimization for Machine Learning

Lecture 1a: Introduction & Convexity

Martin Jaggi
EPFL

PKU Summer School
`github.com/epfml/optml-pku`
July 31th, 2023

Outline

- Monday
  Introduction, Convexity, Gradient Descent
- Tuesday
  Projected, Proximal and Subgradient Descent,
  Stochastic Gradient Descent, Non-Convex Optimization
- Wednesday
  Newton's Method & Quasi-Newton,
  Coordinate Descent
- Thursday
  Frank-Wolfe,
  Accelerated Gradient, Gradient-free, Adaptive Methods for Deep learning
- Friday
  Optimization for ML in Practice,
  Parallel and Distributed Optimization, Federated & Decentralized, Collaborative Learning

Course Materials

- Lecture Slides
- Lecture Notes PDF
- Exercises: Theoretical and Practical Python notebooks

See details on the course webpage:

github.com/epfml/optml-pku

Optimization

- General optimization problem (**unconstrained minimization**)

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{with} \quad & \mathbf{x} \in \mathbb{R}^d \end{aligned}$$

  - candidate solutions, variables, parameters $\mathbf{x} \in \mathbb{R}^d$
  - objective function $f : \mathbb{R}^d \to \mathbb{R}$
  - typically: technical assumption: f is continuous and differentiable

Why? And How?

Optimization is everywhere

machine learning, big data, statistics, data analysis of all kinds, finance, logistics, planning, control theory, mathematics, search engines, simulations, and many other applications ...

- **Mathematical Modeling**:
  - *defining & modeling the optimization problem*
- **Computational Optimization**:
  - *running an (appropriate) optimization algorithm*

Optimization for Machine Learning

- **Mathematical Modeling**:
  - defining & and measuring the machine learning model
- **Computational Optimization**:
  - learning the model parameters
- Theory vs. practice:
  - libraries are available, algorithms treated as "black box" by most practitioners
  - **Not here:** we look inside the algorithms and try to understand why and how fast they work!

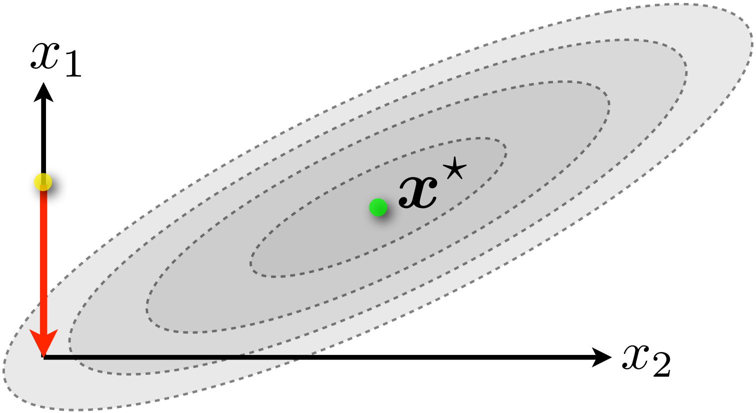
Optimization Algorithms

- Optimization at large scale: **simplicity** rules!
- Main approaches:
  - **Gradient Descent**
  - **Stochastic Gradient Descent** (SGD)
  - **Coordinate Descent**
- History:
  - 1847: Cauchy proposes gradient descent
  - 1950s: Linear Programs, soon followed by non-linear, SGD
  - 1980s: General optimization, convergence theory
  - 2005-2015: Large scale optimization (mostly convex), convergence of SGD
  - 2015-today: Improved understanding of SGD for deep learning

Example: Coordinate Descent

Goal: Find $\mathbf{x}^\star \in \mathbb{R}^d$ minimizing $f(\mathbf{x})$.

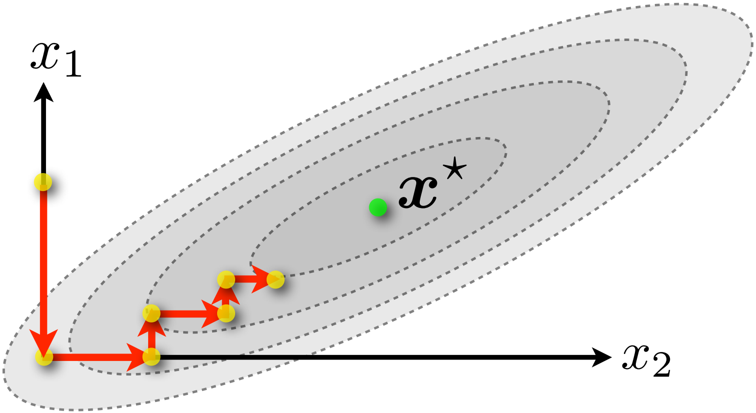
(Example: $d = 2$)



Idea: Update one coordinate at a time, while keeping others fixed.

Example: Coordinate Descent

Goal: Find $\mathbf{x}^\star \in \mathbb{R}^d$ minimizing $f(\mathbf{x})$.



Idea: Update one coordinate at a time, while keeping others fixed.

Chapter 1

Theory of Convex Functions

Warmup: The Cauchy-Schwarz inequality

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Cauchy-Schwarz inequality (Proof in Section 1.1.2):

$$|\mathbf{u}^\top \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \, .$$

Notation:

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{pmatrix}$$

$$\mathbf{u}^\top = \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix}$$

- $\mathbf{u} = (u_1, \ldots, u_d), \mathbf{v} = (v_1, \ldots, v_d)$, d-dimensional column vectors with real entries
- $\mathbf{u}^\top$, transpose of $\mathbf{u}$, a d-dimensional row vector
- $\mathbf{u}^\top \mathbf{v} = \sum_{i=1}^d u_i v_i$, scalar (or inner) product of $\mathbf{u}$ and $\mathbf{v}$
- $|\mathbf{u}^\top \mathbf{v}|$, absolute value of $\mathbf{u}^\top \mathbf{v}$
- $\|\mathbf{u}\| = \sqrt{\mathbf{u}^\top \mathbf{u}} = \sqrt{\sum_{i=1}^d u_i^2}$, Euclidean norm of $\mathbf{u}$

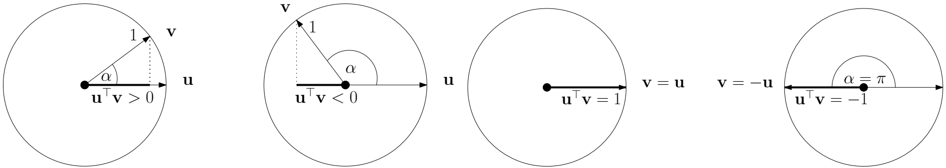
The Cauchy-Schwarz inequality: Interpretation

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Cauchy-Schwarz inequality: $|\mathbf{u}^\top \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$.

For nonzero vectors, this is equivalent to

$$-1 \leq \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|} \leq 1.$$

Fraction can be used to define the angle α between $\mathbf{u}$ and $\mathbf{v}$: $\cos(\alpha) = \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$



Examples for unit vectors ($\|\mathbf{u}\| = \|\mathbf{v}\| = 1$)

Equality in Cauchy-Schwarz if and only if $\mathbf{u} = \mathbf{v}$ or $\mathbf{u} = -\mathbf{v}$.

Convex Sets

A set C is **convex** if the line segment between any two points of C lies in C, i.e., if for any $\mathbf{x}, \mathbf{y} \in C$ and any λ with $0 \leq \lambda \leq 1$, we have

$$\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in C.$$

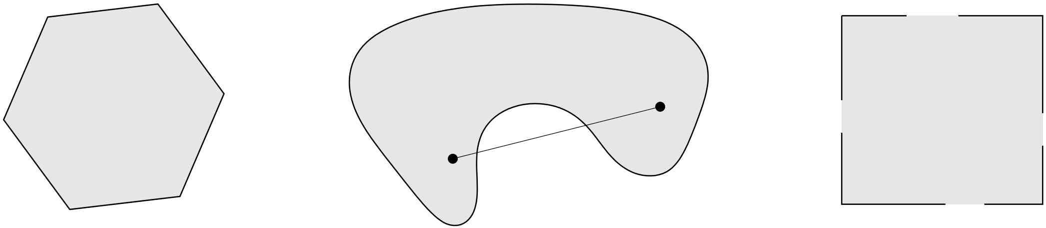


*Figure 2.2 from S. Boyd, L. Vandenberghe

- **Left** Convex.
- **Middle** Not convex, since line segment not in set.
- **Right** Not convex, since some, but not all boundary points are contained in the set.

Properties of Convex Sets

- Intersections of convex sets are convex

  Observation 1.2. Let $C_i, i \in I$ be convex sets, where I is a (possibly infinite) index set. Then $C = \bigcap_{i \in I} C_i$ is a convex set.

- (later) Projections onto convex sets are *unique*, and *often* efficient to compute

$$P_C(\mathbf{x}') := \operatorname{argmin}_{\mathbf{y} \in C} \|\mathbf{y} - \mathbf{x}'\|$$

Convex Functions

Definition

A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if (i) $\mathbf{dom}(f)$ is a convex set and (ii) for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$, and λ with $0 \leq \lambda \leq 1$, we have

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

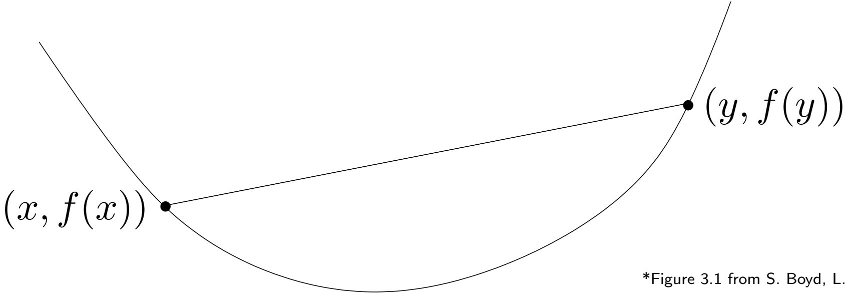


*Figure 3.1 from S. Boyd, L. Vandenberghe

Geometrically: The line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ lies above the graph of f.

Motivation: Convex Optimization

Convex Optimization Problems are of the form

$$\min \; f(\mathbf{x}) \qquad \text{s.t.} \qquad \mathbf{x} \in X$$

where both

- f is a convex function
- $X \subseteq \mathbf{dom}(f)$ is a convex set (note: $\mathbb{R}^d$ is convex)

Crucial Property of Convex Optimization Problems

- Every local minimum is a **global minimum**, see later...

Motivation: Solving Convex Optimization - Provably

For convex optimization problems, all algorithms

- Coordinate Descent, Gradient Descent, Stochastic Gradient Descent, Projected and Proximal Gradient Descent

do **converge** to the global optimum! (assuming f differentiable)

Example Theorem: The **convergence rate** is proportional to $\frac{1}{t}$, i.e.

$$f(\mathbf{x}_t) - f(\mathbf{x}^\star) \leq \frac{c}{t}$$

(where $\mathbf{x}^\star$ is some optimal solution to the problem.)

Meaning: **Approximation error** converges to 0 over time.

Motivation: Convergence Theory

f	Algorithm	Rate	# Iter	Cost/iter
non-smooth	center of gravity	$\exp\left(-\frac{t}{n}\right)$	$n\log\left(\frac{1}{\varepsilon}\right)$	1 ∇, 1 n-dim $\int$
non-smooth	ellipsoid method	$\frac{R}{r}\exp\left(-\frac{t}{n^2}\right)$	$n^2\log\left(\frac{R}{r\varepsilon}\right)$	1 ∇, mat-vec $\times$
non-smooth	Vaidya	$\frac{Rn}{r}\exp\left(-\frac{t}{n}\right)$	$n\log\left(\frac{Rn}{r\varepsilon}\right)$	1 ∇, mat-mat $\times$
quadratic	CG	exact $\exp\left(-\frac{t}{\kappa}\right)$	n $\kappa\log\left(\frac{1}{\varepsilon}\right)$	1 ∇
non-smooth, Lipschitz	PGD	$RL/\sqrt{t}$	R^2L^2/ε^2	1 ∇, 1 proj.
smooth	PGD	$\beta R^2/t$	$\beta R^2/\varepsilon$	1 ∇, 1 proj.
smooth	AGD	$\beta R^2/t^2$	$R\sqrt{\beta/\varepsilon}$	1 ∇
smooth (any norm)	FW	$\beta R^2/t$	$\beta R^2/\varepsilon$	1 ∇, 1 LP
strong. conv., Lipschitz	PGD	$L^2/(\alpha t)$	$L^2/(\alpha\varepsilon)$	1 ∇ , 1 proj.
strong. conv., smooth	PGD	$R^2\exp\left(-\frac{t}{\kappa}\right)$	$\kappa\log\left(\frac{R^2}{\varepsilon}\right)$	1 ∇ , 1 proj.
strong. conv., smooth	AGD	$R^2\exp\left(-\frac{t}{\sqrt{\kappa}}\right)$	$\sqrt{\kappa}\log\left(\frac{R^2}{\varepsilon}\right)$	1 ∇
$f+g$, f smooth, g simple	FISTA	$\beta R^2/t^2$	$R\sqrt{\beta/\varepsilon}$	1 ∇ of f Prox of g
$\max_{y\in\mathcal{Y}}\varphi(x,y)$, φ smooth	SP-MP	$\beta R^2/t$	$\beta R^2/\varepsilon$	MD on $\mathcal{X}$ MD on $\mathcal{Y}$
linear, $\mathcal{X}$ with F ν-self-conc.	IPM	$\nu\exp\left(-\frac{t}{\sqrt{\nu}}\right)$	$\sqrt{\nu}\log\left(\frac{\nu}{\varepsilon}\right)$	Newton step on F
non-smooth	SGD	$BL/\sqrt{t}$	B^2L^2/ε^2	1 stoch. ∇, 1 proj.
non-smooth, strong. conv.	SGD	$B^2/(\alpha t)$	$B^2/(\alpha\varepsilon)$	1 stoch. ∇, 1 proj.
$f=\frac{1}{m}\sum f_i$ f_i smooth strong. conv.	SVRG	–	$(m+\kappa)\log\left(\frac{1}{\varepsilon}\right)$	1 stoch. ∇

(Bubeck [Bub15])

Convex Functions & Sets

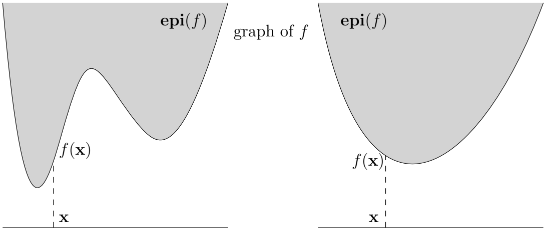
The **graph** of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\{(\mathbf{x}, f(\mathbf{x})) \,|\, \mathbf{x} \in \mathbf{dom}(f)\},$$

The **epigraph** of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathbf{epi}(f) := \{(\mathbf{x}, \alpha) \in \mathbb{R}^{d+1} \,|\, \mathbf{x} \in \mathbf{dom}(f), \alpha \geq f(\mathbf{x})\},$$

Observation 1.4. A function is convex *iff* its epigraph is a convex set.



Convex Functions & Sets

Proof: recall $\mathbf{epi}(f) := \{(\mathbf{x}, \alpha) \in \mathbb{R}^{d+1} \,|\, \mathbf{x} \in \mathbf{dom}(f), \alpha \geq f(\mathbf{x})\}$

Convex Functions

Examples of convex functions

- Linear functions: $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$
- Affine functions: $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$
- Exponential: $f(x) = e^{\alpha x}$
- Norms. Every norm on $\mathbb{R}^d$ is convex.

Convexity of a norm $\|\mathbf{x}\|$

By the triangle inequality $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ and homogeneity of a norm $\|a\mathbf{x}\| = |a| \, \|\mathbf{x}\|$, a scalar:

$$\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\| \leq \|\lambda\mathbf{x}\| + \|(1-\lambda)\mathbf{y}\| = \lambda \|\mathbf{x}\| + (1-\lambda) \|\mathbf{y}\| .$$

We used the triangle inequality for the inequality and homogeneity for the equality.

Jensen's Inequality

Lemma (Jensen's inequality)

Let f *be convex,* $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbf{dom}(f)$, $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$ *such that* $\sum_{i=1}^m \lambda_i = 1$. *Then*

$$f\left(\sum_{i=1}^m \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^m \lambda_i f(\mathbf{x}_i).$$

For $m = 2$, this is convexity. The proof of the general case is Exercise 2.

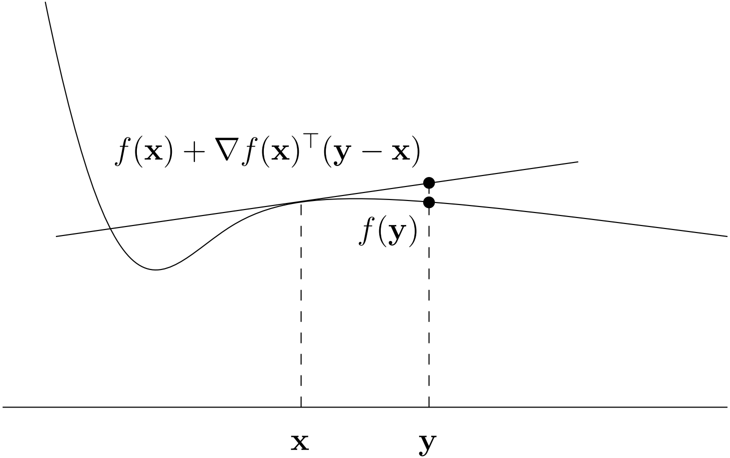
Convex Functions are Continuous

Lemma 1.6.: Let f be convex and suppose that $\mathbf{dom}(f)$ is open. Then f is continuous.

Not entirely obvious (Exercise 3).

Differentiable Functions

Graph of the affine function $f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x})$ is a tangent hyperplane to the graph of f at $(\mathbf{x}, f(\mathbf{x}))$.



First-order Characterization of Convexity

Lemma ([BV04, 3.1.3])
Suppose that $\mathbf{dom}(f)$ *is open and that* f *is differentiable; in particular, the* **gradient** *(vector of partial derivatives)*

$$\nabla f(\mathbf{x}) := \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_d}(\mathbf{x}) \right)$$

exists at every point $\mathbf{x} \in \mathbf{dom}(f)$*. Then* f *is convex if and only if* $\mathbf{dom}(f)$ *is convex and*

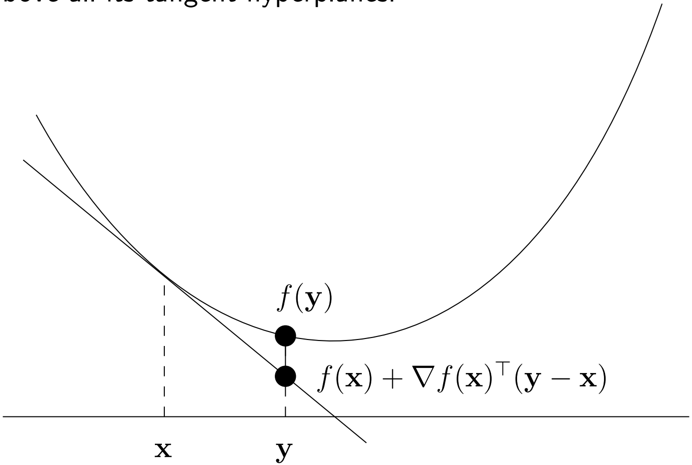
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \qquad (1)$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$*.*

First-order Characterization of Convexity

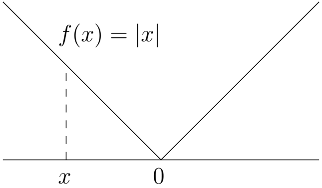
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}), \quad \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f).$$

Graph of f is above all its tangent hyperplanes.

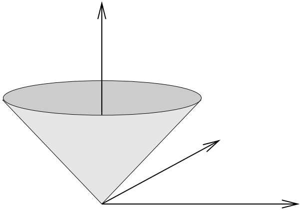


Nondifferentiable Functions. . .

are also relevant in practice.



More generally, $f(\mathbf{x}) = \|\mathbf{x}\|$ (Euclidean norm). For $d = 2$, graph is the ice cream cone:



Second-order Characterization of Convexity

Lemma ([BV04, 3.1.4])

Suppose that $\mathbf{dom}(f)$ *is open and that* f *is twice differentiable; in particular, the* **Hessian** *(matrix of second partial derivatives)*

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_d \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d}(\mathbf{x}) \end{pmatrix}$$

exists at every point $\mathbf{x} \in \mathbf{dom}(f)$ *and is symmetric. Then* f *is convex if and only if* $\mathbf{dom}(f)$ *is convex, and for all* $\mathbf{x} \in \mathbf{dom}(f)$*, we have*

$$\nabla^2 f(\mathbf{x}) \succeq 0 \quad \textit{(i.e. } \nabla^2 f(\mathbf{x}) \textit{ is positive semidefinite).}$$

(A symmetric matrix M *is positive semidefinite if* $\mathbf{x}^\top M \mathbf{x} \geq 0$ *for all* $\mathbf{x}$*, and positive definite if* $\mathbf{x}^\top M \mathbf{x} > 0$ *for all* $\mathbf{x} \neq \mathbf{0}$*.)*

Second-order Characterization of Convexity

Example: $f(x_1, x_2) = x_1^2 + x_2^2$.

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succeq 0.$$

Operations that Preserve Convexity

Lemma (Exercise 5)

(i) *Let* $f_1, f_2, \ldots, f_m$ *be convex functions,* $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}_+$. *Then* $f := \sum_{i=1}^m \lambda_i f_i$ *is convex on* $\mathbf{dom}(f) := \bigcap_{i=1}^m \mathbf{dom}(f_i)$.

(ii) *Let* f *be a convex function with* $\mathbf{dom}(f) \subseteq \mathbb{R}^d$, $g : \mathbb{R}^m \to \mathbb{R}^d$ *an affine function, meaning that* $g(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, *for some matrix* $A \in \mathbb{R}^{d \times m}$ *and some vector* $\mathbf{b} \in \mathbb{R}^d$. *Then the function* $f \circ g$ *(that maps* $\mathbf{x}$ *to* $f(A\mathbf{x} + \mathbf{b})$*) is convex on* $\mathbf{dom}(f \circ g) := \{\mathbf{x} \in \mathbb{R}^m : g(\mathbf{x}) \in \mathbf{dom}(f)\}$.

Local Minima are Global Minima

Definition

A **local minimum** of $f : \mathbf{dom}(f) \to \mathbb{R}$ is a point $\mathbf{x}$ such that there exists $\varepsilon > 0$ with

$$f(\mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{dom}(f) \text{ satisfying } \|\mathbf{y} - \mathbf{x}\| < \varepsilon.$$

Lemma

Let $\mathbf{x}^\star$ *be a* **local minimum** *of a convex function* $f : \mathbf{dom}(f) \to \mathbb{R}$*. Then* $\mathbf{x}^\star$ **is a global minimum**, *meaning that* $f(\mathbf{x}^\star) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{dom}(f)$.

Proof.

Suppose there exists $\mathbf{y} \in \mathbf{dom}(f)$ such that $f(\mathbf{y}) < f(\mathbf{x}^\star)$.
Define $\mathbf{y}' := \lambda \mathbf{x}^\star + (1 - \lambda)\mathbf{y}$ for $\lambda \in (0, 1)$.
From convexity, we get that that $f(\mathbf{y}') < f(\mathbf{x}^\star)$. Choosing λ so close to 1 that $\|\mathbf{y}' - \mathbf{x}^\star\| < \varepsilon$ yields a contradiction to $\mathbf{x}^\star$ being a local minimum. $\square$

Critical Points are Global Minima

Lemma

*Suppose that f is convex and differentiable over an open domain $\mathbf{dom}(f)$. Let $\mathbf{x} \in \mathbf{dom}(f)$. If $\nabla f(\mathbf{x}) = \mathbf{0}$ (**critical point**), then $\mathbf{x}$ is a **global minimum**.*

Proof.

Suppose that $\nabla f(\mathbf{x}) = \mathbf{0}$. According to our Lemma on the first-order characterization of convexity, we have

$\square$

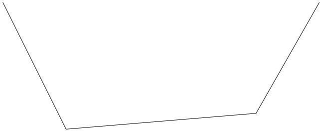
Geometrically, tangent hyperplane is horizontal at $\mathbf{x}$.

Strictly Convex Functions

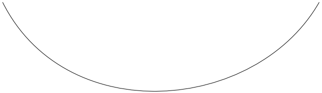
Definition ([BV04, 3.1.1])

A function $f : \mathbf{dom}(f) \to \mathbb{R}$ is **strictly convex** if (i) $\mathbf{dom}(f)$ is convex and (ii) for all $\mathbf{x} \neq \mathbf{y} \in \mathbf{dom}(f)$ and all $\lambda \in (0, 1)$, we have

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}). \tag{2}$$



convex, but not strictly convex



strictly convex

Lemma

Let $f : \mathbf{dom}(f) \to \mathbb{R}$ be strictly convex. Then f has at most one global minimum.

Constrained Minimization

Definition

Let $f : \mathbf{dom}(f) \to \mathbb{R}$ be convex and let $X \subseteq \mathbf{dom}(f)$ be a convex set. A point $\mathbf{x} \in X$ is a minimizer of f over X if

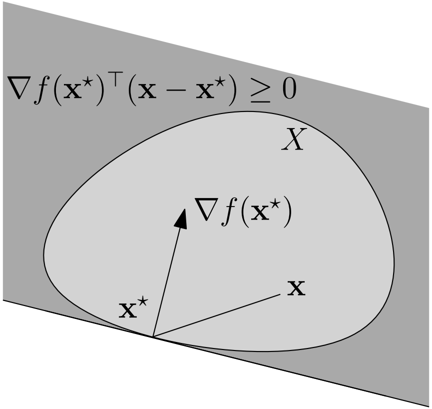
$$f(\mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in X.$$

Lemma

Suppose that $f : \mathbf{dom}(f) \to \mathbb{R}$ is convex and differentiable over an open domain $\mathbf{dom}(f) \subseteq \mathbb{R}^d$, and let $X \subseteq \mathbf{dom}(f)$ be a convex set. Point $\mathbf{x}^\star \in X$ is a minimizer of f over X if and only if

$$\nabla f(\mathbf{x}^\star)^\top (\mathbf{x} - \mathbf{x}^\star) \geq 0 \quad \forall \mathbf{x} \in X.$$

Constrained Minimization



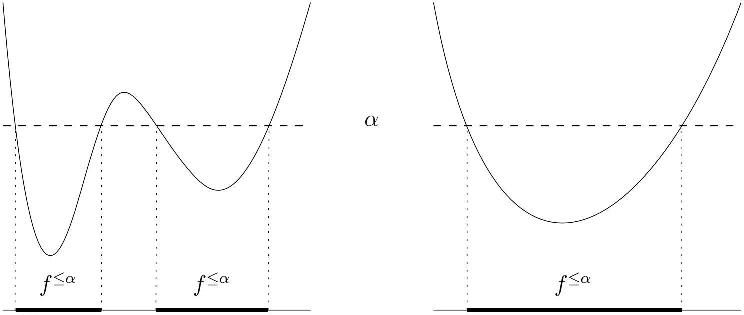
Existence of a minimizer

How do we know that a global minimum exists?

Not necessarily the case, even if f bounded from below ($f(x) = e^x$)

Definition

$f : \mathbb{R}^d \to \mathbb{R}$, $\alpha \in \mathbb{R}$. The set $f^{\leq \alpha} := \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \leq \alpha\}$ is the α-sublevel set of f



The Weierstrass Theorem

Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function, and suppose there is a nonempty and bounded sublevel set $f^{\leq \alpha}$. Then f has a global minimum.

Proof:

We know that f—as a continuous function—attains a minimum over the closed and bounded (= compact) set $f^{\leq \alpha}$ at some $\mathbf{x}^\star$. This $\mathbf{x}^\star$ is also a global minimum as it has value $f(\mathbf{x}^\star) \leq \alpha$, while any $\mathbf{x} \notin f^{\leq \alpha}$ has value $f(\mathbf{x}) > \alpha \geq f(\mathbf{x}^\star)$.

Generalizes to suitable domains $\mathbf{dom}(f) \neq \mathbb{R}^d$.

Bibliography

Sébastien Bubeck.
Convex Optimization: Algorithms and Complexity.
Foundations and Trends in Machine Learning, 8(3-4):231–357, 2015.

Stephen Boyd and Lieven Vandenberghe.
Convex Optimization.
Cambridge University Press, New York, NY, USA, 2004.
https://web.stanford.edu/~boyd/cvxbook/.