Large-Scale Convex Optimization

Stadelmann Silvan silvasta@ethz.ch

13. Juni 2025

1 Convex sets and convex functions

Definition 1 (Convex Set). A set $\mathcal{C}$ is convex if and only if $\forall x, y \in \mathcal{C}$ and $\forall \theta \in [0,1]$: $\theta x + (1-\theta) y \in \mathcal{C}$.
Examples of convex sets:

- hyperplane $\{x \in \mathbb{R}^n \mid a^\mathsf{T} x = b\}$
- half-space $\{x \in \mathbb{R}^n \mid a^\mathsf{T} x \leq b\}$
- polyhedron $\{x \in \mathbb{R}^n \mid Ax \leq b,\ Cx = d\}$
- ...more...

Operations that preserve convexity (sets)

- **Intersection** $\mathcal{C}_1, \mathcal{C}_2$ convex $\Rightarrow \mathcal{C}_1 \cap \mathcal{C}_2$ convex
- **Image under affine map** $\mathcal{C} \subseteq \mathbb{R}^n$ convex $\Rightarrow \{Ax + b \mid x \in \mathcal{C}\}$ convex
- inverse image of an affine map: ...

Separating Hyperplane Theorem

Theorem 1. Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $y \notin \mathcal{C}$. Then $\exists\ a \neq 0, b \in \mathbb{R}$, s. t. $a^\mathsf{T} x + b < a^\mathsf{T} y + b,\ \forall x \in \mathcal{C}$
Proof
Step 1: Claim $\exists \hat{x} \in C$ s.t. $|\hat{x} - y| \leq |x - y|$
Step 2: hyperplane $a = y - \hat{x}, b = -a^T \hat{x} = -(y - \hat{x})^T \hat{x}$
$\rightarrow$ we note $a^\mathsf{T} y + b = |y - xh|^2 > 0$
$\rightarrow$ we need to show $a^\mathsf{T} x + b \leq 0\ \forall x \in \mathcal{C}$
$(y - \hat{x})^T (x - \hat{x}) < 0$
(Details in Lecture notes)
Corollary
A closed convex set $\mathcal{C} = \mathbb{R}^n$ is the intersection of the closed half-spaces that contain $\mathcal{C}$
Proof $\mathcal{S}$ is the intersection of all half-spaces containing $\mathcal{C}$
1) $\mathcal{C} contains \mathcal{S}$: $x \in \mathcal{C}$ => x in all half-spaces...
2) Assume not, ...

Support function

Idea represent any closed convex set by its supporting hyperplanes
support function $\sigma_\mathcal{C}(a) = sup_{x \in \mathcal{C}} a^T x$
CALCULATION EXAMPLE
If we know the support function, we arrive at at

$$\mathcal{C} = \bigcap_{a \in \mathbb{R}^n} \{x \in \mathbb{R}^n \mid a^\mathsf{T} x - \sigma_c(a) \leq 0\}$$

$$= \{x \in \mathbb{R}^n \mid \sup_{x \in \mathcal{C}} a^T x - \sigma_\mathcal{C}(a) \leq 0\}$$

Definition 2. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if its epigraph is a convex set, where $epi(f) := \{(x,t) \in \mathbb{R}^{n+1} | f(x) \leq t\}$
$\rightarrow$ this provides a link between sets and functions

Operations that preserve convexity (functions)

- the pointwise maximum of convex functions is convex
- the sum of convex functions is convex
- $f(Ax + b)$ is convex if f is convex

How to check f convex?

- if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice differentiable d2/dx2>=0
- if $g : \mathbb{R} \rightarrow \mathbb{R}$ with $g(t) = f(x + tv)$ is convex in t for all $x, v \in \mathbb{R}^n$, then f is convex

$\rightarrow$ It will be convenient to introduce extended real numbers

$$\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$$

$\rightarrow$ Indicator function $INF x \notin C\ 0 x \in C$
$\rightarrow$ we can then write minf as minf+y NOTES
Definition 3 (3). f : Rn □ R is called proper if f is bounded below and if x □ Rn s. t. f (x) < ∞.
Definition 4 (name of the definition). Let f : Rn □ R. The conjugate function of f is defined as

$$f^\star(y) = sup_{CC} y^T x - f(x)$$

2 KKT and Lagrange Duality

Example

$$\min_{x \in \mathbb{R}^\epsilon} f(x) \text{ s.t. } h(x) = 0$$

Generalization

Generalization to n $\leq$ 2 and presence of inequality constraints

$$f^\star = \inf_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h(x) = 0, g(x) \leq 0$$

$\rightarrow$ the corresponding Lagrange function is then:

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \lambda^\mathsf{T} g(x) + \nu^\mathsf{T} h(x)$$

cond...
Proposition 1 (Weak Duality). The dual function

$$d(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \nu)$$

satisfies $d(\lambda, \nu) \leq f^\star,\ \forall \lambda \geq 0,\ \nu \in \mathbb{R}^n$
proof short
Definition 5 (Constraint qualification). Let $\mathcal{C}$ be Convex, Slaters Condition is satisfied if $\exists \lambda \geq 0, v Rn$ s.t. $d(l, v) = f*$
Proposition 2 (Strong Duality). If Slater's condition holds and (1 TODO) is convex then $\exists \lambda \geq 0,\ \nu \in \mathbb{R}^n$ such that $d(\lambda, \nu) = f^\star$
proof extended, important graphic

KKT

Theorem 2 (KKT Conditions). Let (1,TODO) be convex and Slaters condition hold. Then $x^\star \in \mathbb{R}^n$ is a minimizer of the primal (1t) and $\lambda^\star \geq 0,\ \nu^\star$ maximizer of the dual if and only if:

$$KKT - 1\ (Stationary\ Lagrangian)$$

$$\nabla_x \mathcal{L}(x^\star, \lambda^\star, \nu^\star) = 0$$

$$KKT - 2\ (primal\ feasibility)$$

$$g(x^\star) \leq 0, h(x^\star) = 0$$

$$KKT - 3\ (dual\ feasibility)$$

$$\lambda^\star \leq 0, \nu^\star \in \mathbb{R}^{n_h}$$

$$KKT - 4\ (complementary\ slackness)$$

$$\lambda^{\star\mathsf{T}} g(x^\star) = 0, \nu^{\star\mathsf{T}} h(x^\star) = 0$$

INF=SUP
Remark Without Slater, KKT 1 to 4 still implies $x^\star$ minimizes (1t) and (λ,ν) maximizes the dual, but the convergence is no longer true!
FORCE BALLANCE

What if f, g not differentiable?

Example $\inf_{x \in \mathbb{R}^n} |Ax - b|^2 + |x|_1$
where (l_1)-norm not differentiable at 0

Subdifferential

for convex f...
Definition 6 (name of the definition). $f \mathbb{R}^n \rightarrow \mathbb{R}$ convex. The subdifferential of f at $\bar{x}$ is: $\delta f(\bar{x}) := \{\lambda \in \mathbb{R}^n \mid f...\}$
Proposition 3. $f \mathbb{R}^n \rightarrow \mathbb{R}$ convex. $x^\star \in argmin...$
Proposition 4. f convex, $epi(f)$ closed $y \in df(x) arrow x \in \delta f^\star(y)$

EXAMPLE

3 Convex Optimization Problem

minimize $f(x), g \leq 0, h(x) = 0$
1) Feasibility Problem
minimmize s
s.t. $g_i(x) \leq s \quad \forall i, \ldots, n_g, h(x) = 0$
2) Linear Programming
minimmize $c^\mathsf{T} x$ s.t. $Ax - b \geq 0$ and $x \geq 0$
$\rightarrow$ derive dual:
Step 1: $\mathcal{L}(x, \lambda_1, \lambda_1) = c^\mathsf{T} x - \lambda_1^\mathsf{T}(Ax - b) - \lambda_2^\mathsf{T} x$
Step 2: $\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda_1, \lambda_1) = \{-\infty else, \lambda_1^\mathsf{T} b, if\ A^\mathsf{T} \lambda_1 + \lambda_2 = c$
Step 3: Dual Problem
$\sup_{\lambda_1 \geq 0} \lambda_1 b$ s.t. $0 \leq c - A^\mathsf{T} \lambda_2$

- dual as a linear program
- Skech, polyhedron, c-vector normal gives 'Level-sets' and optimal solution in or trough a corner (if exists)

Proposition 5. The optimal solution of a linear program (if it exists) lies always on the boundary of the feasible set and there exists an optimal solution that is a vertex of the feasible set.
Example Shortest Path
Analogie with Fluid
Soltuion greater 0, not optimal edges = 0
3) Quadratic Programming
minimmize $\frac{1}{2} x^\mathsf{T} P x + q^T x$
s.t. $Gx \leq h,\ Ax = b$
$\rightarrow$ if $P = P^\mathsf{T}$ is positive semi-definite then the problem is convex.

Example [optimal control] (basis for mpc)

Second-order cone program (SOCP)

minimmize $f^T x$
s.t. $|A_i x + b| \leq c_i^T x + d_i,\ Fx = g$
Cone: Cn+1=
Example [Markovitz portfolio optimization:]
- n number of assets/stocks
- x_i relative value of asset i
- p_i price change of stock i
- $p^T x$ overall return

Constraints
- $x^T \mathbf{1} = B$, total amount
- $x \geq 0$, no short position

CALCULATIONS

Semidefinite programming (SDP)

minimmize $c^\mathsf{T} x$
s.t. $x_1 F_1, \cdots + x_n F_n <= 0$ and $Ax - b = b$
$\rightarrow$ the 'standard' form

$$\min_{x \in \mathbb{R}^{n*n}} tr(CX)$$

Diagramm

4 Gradient methods - Part I

Definition 7 (smoothness)**.** The function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is L-smooth if $\nabla f(x)$ satisfies

$$|\nabla f(x) - \nabla f(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R}$$

This result (with Taylors'Theorem) in:

$$f(y) \leq f(x) + \nabla f(x)^\mathsf{T}(y - x) + \frac{L}{2}|x - y|^2 \quad \forall x, y \in \mathbb{R}$$

Definition 8 (strong convexity)**.** The function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is μ strongly convex it it satisfies

$$f(y) \geq f(x) + \nabla f(x)^\mathsf{T}(y - x) + \frac{\mu}{2}|x - y|^2 \quad \forall x, y \in \mathbb{R}$$

Gradient Descent

PSEUDOCODE
$x = x_0 \in \mathbb{R}^n$
for x in range N ...
HERLEITUNG
Optimal Step Size
$\mu \leq h \leq L$

$$T^\star = \frac{2}{L + M}$$

GRAFIK

$$\rho(T^\star) = |XXX|$$

therefore with stepsize $T^\star$
$|x_N - x^\star| \leq \epsilon$

Momentum-based methods

$$q_{k+1} = q_k + T_{p_{k+1}}$$

$$p_{k+1} = (1 - 2dT)p_k - T\nabla f(q_k + \beta p_k)/L$$

Discretization of $\dot{q} = p, \dot{p} = -2d...$
Spring damper analogy
- for $T = 1, d = \frac{1}{\sqrt{k}+1}, \beta = \frac{\sqrt{k}-1}{\sqrt{k}+1}$

Nesterovs accelerated gradient methods
- for $T = \frac{2\sqrt{k}}{\sqrt{k}+1}, d = \frac{1}{\sqrt{k}+1}, \beta = 0$

Heavy Ball (tuned quadratics)
What is the convergence rate?
EXAMPLE DIAGONALIZATION
EIGENVALUE analysis
ROOT Locus
- Nesterov on circle $c = (r/0), r = \lambda_i/L = \mu/L$
- Heavy ball circle $c = ((\lambda - L)/2, 0), r = \lambda + L$

Theorem 3 (NOT Nesterovs)**.** $f\mu$ strongly convex, L smooth Nesterovs Method satisfies

$$|x_N - x^\star| \leq (1 - \frac{2}{\sqrt{k} + 1})|x_0 - x^\star| \forall k \geq 0$$

proof with H Function

Gradient Descent - Part II

Projected gradient descent(smooth,strongly convex f)
Definition 9. $prox_{\mathcal{C}}(x) = argmin 1/2|x - y|^2$
C closed convex
CAUCHY SCHWARZ
This implies: $|prox_{\mathcal{C}}(x) - prox_{\mathcal{C}}(y)| \leq |x - y|$

Proof Other Information. TODO □

Algorithm
Proposition 6. satisfaction of GD

Proof. Restricted on quadtratic functions: $\frac{1}{2}x^T H x + b^\mathsf{T} x + c$ □

– norm ball
- probability simplex

when are projetions computationally cheap?
What if f is not strongly convex? ($\mu = 0$)
$\rightarrow$idea: apply small amount of regulairzation
$f : \mathbb{R}^n \rightarrow \mathbb{R}$ L-smooth, convex

$$\hat{f}(x) = f(x) + \frac{\mu}{2}|x - x_0|^2$$

and
XXX (IEQ 1 2)
are satisfied $\forall x_0\ in \mathbb{R}^n, \mu > 0$, where...x(hat)star argmin f(hat)

Proof. XXX □

hence we can apply GD or Nesterov
calc
For Nesterov: ... $e^{sqrt \frac{\mu}{L+\mu}}$...
sqrt ESSENCE of morning
chose .. $\frac{2\ln(N)}{N}$
BOX Hence if f smooth and (not strongly) convex we need aproximately N tilde $L|x^\star - x_0|^2/epsilon$ iterations to reach $f(x_N) - f(x_0) \leq \epsilon$
What if f is non-smooth?
i.e. L_f Lipschitz but nor neccessairly differentiable
Example $f(x) = |x|$
Leads to osciliations with $\nabla f = \{+1 \mid -1\}$
Proposition 7 (Subgradient Method)**.** Closed, convex set $\mathcal{C}$ contained in ball of $r = R$
Consider update rule: $x_{k+1} = prox_{\mathcal{C}}(x_k - Tg_k), \ldots$ then $x_0, \ldots$

Proof. NOT SHOWED □

TABLE
GRAPH with rates, IMPORTANTe