Annika Hirsch - 08.11.2024

Optimization Cheat Sheet

Preliminaries

Def: f is coercive iff for every seq. $(x_k) \subseteq \mathbb{R}^n$ s.t. $\lim_{k\to\infty} \|x_k\| = \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$
(not proven)

Def: The sublevel set of $f: \mathbb{R}^N \to \mathbb{R}$ is defined as $[f \le \gamma] = \{x \in \mathbb{R}^N : f(x) \le \gamma\}$

Prop: The sublevel sets of a coercive function must be bounded

Prop: If $f: \mathbb{R}^N \to \mathbb{R}$ continuous and coercive, then $\operatorname{argmin}(f) \neq \emptyset$

Def: $f: D \subset \mathbb{R}^N \to \mathbb{R}$ is lower-semicontinuous at $x \in D$ if, $\forall \varepsilon > 0 \ \exists \delta > 0$ s.t. $f(y) > f(x) - \varepsilon$ for every $y \in B(x, \delta) \cap D$

Note: Every continuous function satisfies this.

Prop: $f: D \subset \mathbb{R}^N \to \mathbb{R}$ is lower-semicon. at $x \in D$ iff $f(x) \le \lim_{k\to\infty}(\inf(f(x_k)))$ $\forall (x_k) \to x$ in D

Cauchy-Schwartz Ineq: $\forall x, y \in \mathbb{R}^N$, $|x \cdot y| \le \|x\| \|y\|$

Def: $f: D \subset \mathbb{R}^N \to \mathbb{R}$ is L-smooth if it is differentiable and its gradient is L-Lipschitz continuous: $\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$

L-smoothness

Prop: Let $f: D \subset \mathbb{R}^N \to \mathbb{R}$ be L-smooth and $\operatorname{seg}(x,y) \subseteq D$. Then, $|f(y) - f(x) - \nabla f(x) \cdot (y - x)| \le \frac{L}{2}\|x - y\|^2$

First Optimality Condition:

Let $f: D \subset \mathbb{R}^N \to \mathbb{R}$ differentiable at $\hat{x}$. Then, if $\hat{x} \in \operatorname{argmin}_D(f)$, $\langle \nabla f(\hat{x}), y - \hat{x} \rangle \ge 0$ $\forall y \in C$ convex + closed where $C \subseteq D$
And if $\hat{x} \in \operatorname{int}(D)$, then $\nabla f(\hat{x}) = 0$

Second order Optimality Conditions:

Let $D \subset \mathbb{R}^N$ be open, let $f: D \to \mathbb{R}$ be of class C^2 and let $\hat{x} \in D$.

i) If $\hat{x} \in \operatorname{argmin}(f)$, then $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x})$ is positive semidefinite.

ii) If $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x})$ is positive definite there is $\varepsilon > 0$ such that $f(\hat{x}) < f(y)$ for every $y \in B(\hat{x}, \varepsilon)$

A is positive definite iff $x^T A x \ge 0 \ \ \forall x \in \mathbb{R}^N$

Second order Taylor approximation:

Let $f: D \subset \mathbb{R}^N \to \mathbb{R} \in C^2$ and $x \in A$. For each $v \in \mathbb{R}^N$, $\lim_{t\to 0} \frac{1}{t^2} \left| f(x + tv) - f(x) - t\nabla f(x) \cdot v - \frac{t^2}{2} \nabla^2 f(x) v \cdot v \right| = 0$

An affine subspace is $C = \{x_0\} + V = \{x_0 + v : v \in V\}$ & Closed halfspaces are $[v \le \gamma] = \{x \in \mathbb{R}^N : v \cdot x \le \gamma\}$
A polyhedron is the intersection of a $<\infty$ number of halfspaces $\Rightarrow$ polyhedra are convex

Projections

Let $C \subset \mathbb{R}^N$ s.t. $C \neq \emptyset$, closed and convex. For each $x \in \mathbb{R}^N$ $\exists P_C(x) \in C$ unique such that $\|x - P_C(x)\| = \min\{\|x - z\| : z \in C\} = \operatorname{dist}(x, C)$

The point $P_C(x)$ is the projection of x onto C.

Moreover, $P_C(x)$ is the only point in C that satisfies $(x - P_C(x)) \cdot (z - P_C(x)) \le 0$ $\forall z \in C$

Also note $\|P_C(x) - P_C(y)\| \le \|x - y\|$ $\forall x, y \in \mathbb{R}^N$ $\Rightarrow P_C$ is non-expansive.

Epigraph

The epigraph is $\operatorname{epi}(f) = \{(x, t) \in \mathbb{R}^N \times \mathbb{R} : t \ge f(x)\}$

We have shown: f convex $\Leftrightarrow \operatorname{epi}(f)$ convex

Prop: If $D \subset \mathbb{R}^N$ has a nonempty interior, and f has at least one point of continuity, then $\operatorname{epi}(f)$ has a non-empty interior

Convexity

Prop: Let C be convex, let $x \in \operatorname{int}(C)$, and let $y \in \bar{C}$. Then, $\lambda x + (1-\lambda) y \in \operatorname{int}(C)$ for every $\lambda \in (0, 1]$

Def: A function $f: D \subset \mathbb{R}^N \to \mathbb{R}$ is convex if D is a convex set and $f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$ $\forall x, y \in D$ and $\lambda \in (0, 1)$

Note: Any norm is convex

Prop: Every sublevel set of a convex function is convex. Also, if f is convex, then $\operatorname{argmin}(f)$ is convex.

Prop: If $D \subset \mathbb{R}^N$ is convex, for each $x \in \operatorname{int}(D)$, there exists $L_x, r_x > 0$ such that for all $z, y \in B(x, r_x)$. In particular, f is continuous in $\operatorname{int}(D)$.

Prop: Let $D \subset \mathbb{R}^N$ be convex, and let $f: D \to \mathbb{R}$ be differentiable. The following are equivalent:

i) f is convex

ii) $\forall x, y \in D$, $f(y) \ge f(x) + \nabla f(x) \cdot (y - x)$

iii) $\forall x, y \in D$, $(\nabla f(y) - \nabla f(x)) \cdot (y - x) \ge 0$

If f is twice differentiable,

iv) $\forall x \in D$, $\nabla^2 f(x)$ is positive semidefinite

Baillon-Haddad Thm: Let $f: \mathbb{R}^N \to \mathbb{R}$ be convex and differentiable, and let $L > 0$. The following are equivalent:

i) f is L-smooth

ii) $f(z) \le f(x) + \nabla f(x) \cdot (z - x) + \frac{L}{2}\|z - x\|^2$ $\forall x, z \in \mathbb{R}^N$

iii) $f(x) \ge f(y) + \nabla f(y) \cdot (x - y) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$ $\forall x, y \in \mathbb{R}^N$

iv) $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$ $\forall x, y \in \mathbb{R}^N$

Def: f is strictly convex if D is convex and $f(\lambda x + (1-\lambda)) < \lambda f(x) + (1-\lambda) f(y)$ $\forall \lambda \in (0,1)$ and $\forall x, y \in D$ s.t. $x \neq y$

Prop: A strictly convex function can have, at most, one minimizer.

Def: Given $\mu > 0$, $f: D \subset \mathbb{R}^N \to \mathbb{R}$ is μ-strongly convex if D is convex and $f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x - y\|^2$ $\forall x, y \in D$ and $\lambda \in (0,1)$

Note: strongly convex functions cannot be Lipschitz continuous

Also: strongly convex functions have unique minimizers

Ex. (not proven): f is μ-strongly convex iff $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ is convex

Note: Proof relies on $(1-\lambda)\|x\|^2 + \lambda\|y\|^2 - \|(1-\lambda)x + \lambda y\|^2 = \lambda(1-\lambda)\|x - y\|^2$

Prop: Let $f: D \subset \mathbb{R}^N \to \mathbb{R}$ be differentiable. The following are equivalent:

i) f is μ-strongly convex

ii) $\forall x, y \in D$, $f(y) \ge f(x) + \nabla f(x) \cdot (y - x) + \frac{\mu}{2}\|x - y\|^2$

iii) $\forall x, y \in D$, $(\nabla f(y) - \nabla f(x)) \cdot (y - x) \ge \mu\|x - y\|^2$

If f is twice differentiable

iv) $\forall x \in D$, $\nabla^2 f(x)$ is positive definite, with $v \cdot \nabla^2 f(x) v \ge \mu\|v\|^2$ for every $v \in \mathbb{R}^N$

Prop: If $f: \mathbb{R}^N \to \mathbb{R}$ is L-smooth and μ-strongly convex, then $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \ge \frac{\mu L}{L + \mu}\|x - y\|^2 + \frac{1}{L + \mu}\|\nabla f(x) - \nabla f(y)\|^2$ $\forall x, y \in \mathbb{R}^N$

Subgradient

Def: $v \in \mathbb{R}^N$ is a subgradient of f at x if $f(y) \ge f(x) + v \cdot (y - x)$ $\forall y \in \mathbb{R}^N$. The subdifferential of f at x, $\partial f(x)$, is the set of all subgradients.
closed and convex

Prop: $\forall x, y \in \mathbb{R}$ and $u \in \partial f(x)$, $v \in \partial f(y)$, it holds that $(x - y) \cdot (u - v) \ge 0$

Prop: $x \in \operatorname{argmin}(f)$ iff $0 \in \partial f(x)$ ← Fermat's Rule II

Prop: If f is convex, then $\partial f(x)$ is nonempty and bounded $\forall x \in \mathbb{R}^N$ & **Prop:** f convex and diff. at $x \Rightarrow \partial f(x) = \{\nabla f(x)\}$

Duality

Def: The fenchel conjugate of f is $f^*(y)=\sup_{x\in\mathbb{R}^N}\{y\cdot x-f(x)\}$
→ If $f\neq\infty$ then f^* is closed and convex.

Prop:
- If $f\leq g$ then $f^*\geq g^*$
- Fenchel-Young: $f(x)+f^*(x^*)\geq x^*\cdot x$ (There is equality iff $x^*\in\partial f(x)$)
- Legendre-Fenchel Reciprocity Formula: $x^*\in\partial f(x)$ iff $x\in\partial f^*(x^*)$
- Let $\mu\ell=1$. Then, f is μ-strongly convex iff f^* is ℓ-smooth

Fenchel-Rockafellar Duality

Let $f:\mathbb{R}^N\to\mathbb{R}\cup\{\infty\}$ and $g:\mathbb{R}^M\to\mathbb{R}\cup\{\infty\}$ be closed & convex

Primal Problem: $\min_{x\in\mathbb{R}^N}\{f(x)+g(Ax)\}$, with opti value $v\in\mathbb{R}$

Dual Problem: $\min_{y\in\mathbb{R}^M}\{f^*(-A^Ty)+g^*(y)\}$, with opti value $v^*\in\mathbb{R}$

The duality gap $v+v^*$ is nonnegative

Primal-Dual Method

Consider $\min\{f(x)+g(Ax)+h(x)\}$

where $A\in\mathbb{R}^{M\times N}$, $f:\mathbb{R}^N\to\mathbb{R}\cup\{+\infty\}$ and $g:\mathbb{R}^M\to\mathbb{R}\cup\{+\infty\}$ are closed and convex, and $h:\mathbb{R}^N\to\mathbb{R}$ is ℓ-smooth and convex.

The Primal-Dual Algorithm is given by: $\begin{cases} x_{k+1}=\operatorname{prox}_{\tau f}(x_k-\tau\nabla h(x_k)-\tau A^Ty_k) \\ y_{k+1}=\operatorname{prox}_{\sigma g^*}(y_k+\sigma A(2x_{k+1}-x)) \end{cases}$ with $\tau\sigma\|A\|^2+\frac{2\ell}{2}\leq 1$

Moreau's Identity: Let $g:\mathbb{R}^M\to\mathbb{R}\cup\{+\infty\}$ be closed and convex, and let $\sigma>0$. Then,
$\operatorname{prox}_{\sigma g^*}(y)=y-\sigma\operatorname{prox}_{\sigma^{-1}g}(\sigma^{-1}y)\quad\forall y\in\mathbb{R}^M$

Lagrange Duality:

Given the primal problem $\min_{x\in\mathbb{R}^N}\{f(x)$ s.t. $h_i(x)\leq 0\ \forall i\in[n]$ we define the Langrangian
$L(x,\lambda)=f(x)+\sum_{i=1}^{n}\lambda_i h_i(x)$

Then the dual problem becomes $\max_{\lambda\geq 0}(\min_x L(x,\lambda))$

Multiplier Methods:

Consider the problem $\min\{f(x): Ax=b\}$

Then $\mathcal{L}_\alpha(x,z)=\underbrace{f(x)+z^T(Ax-b)}_{\mathcal{L}(x,z)}+\frac{\alpha}{2}\|Ax-b\|^2$

Such that the Method of Multipliers uses
$\begin{cases} x_{k+1}\in\operatorname{argmin}\{\mathcal{L}_\alpha(x,z_k): x\in\mathbb{R}^N\} \\ z_{k+1}=z_k+\alpha(Ax_{k+1}-b \end{cases}$

Proximal:

Def: $\operatorname{prox}_f(z)=\operatorname{argmin}\{f(x)+\frac{1}{2}\|x-z\|^2\}$ **Note:** $\hat{x}\in\operatorname{argmin}(f)\iff\hat{x}=\operatorname{prox}_f(\hat{x})$
Note: prox is continuous

Prop: If $C\subset\mathbb{R}^N$ nonempty, closed and convex, then $\operatorname{prox}_{i_C}(x)=P_C(x)$

Prop: For every closed convex function $f:\mathbb{R}^N\to\mathbb{R}\cup\{+\infty\}$, $\|\operatorname{prox}_f(x)-\operatorname{prox}_f(y)\|\leq\|x-y\|\quad\forall x,y\in\mathbb{R}^N$

Optimization Problems with Eq. Constraints

Consider problem (P): $\min_{x\in C} f(x)$ where the feasable set is $C=\{x\in\mathbb{R}^n: h_m(x)=0,\ \forall m\}$

Note $\hat{x}\in C$ is regular if $V=\{\nabla h_1(\hat{x}),\dots,\nabla h_M(\hat{x})\}$ is lin. independent

Lagrange Multiplier Thm: Let $\hat{x}$ be a regular, local minimizer of (P). There is a unique $\hat{\lambda}\in\mathbb{R}^M$, such that
$\nabla f(\hat{x})+\sum_{m=1}^{M}\hat{\lambda}_m\nabla h_m(\hat{x})=0$. In addition $y\cdot(\nabla^2 f(\hat{x})+\sum_{m=1}^{M}\hat{\lambda}_m\nabla^2 h_m(\hat{x}))y\geq 0\quad\forall y\in V=\{\nabla_1(\hat{x}),\dots,\nabla h_m(\hat{x})\}$

Theorem: Let $f,h_1,\dots,h_M\in C^2(\mathbb{R}^N,\mathbb{R})$ and $\hat{x}\in C,\hat{\lambda}\in\mathbb{R}^M$ satisfy the Lagrange Thm $\forall y\in V^\perp\setminus\{0\}$. Then $\hat{x}$ is a strict local minimizer of (P)
$f(x)>f(\hat{x})\ \forall x\in C$

Optimization Problems with Ineq. Constraints

Consider $\min_{x\in C} f(x)$ and the feasable set $C=\{x\in\mathbb{R}^N: h_m(x)=0,\ g_j(x)\leq 0,\ \forall m,j\}$

Def: The set of active ineq. constraints at a feasable point $\hat{x}$ is $A(\hat{x})=\{j: g_j(\hat{x})=0\}$

Note: $\hat{x}\in C$ is regular if $V=\{\nabla g_j(\hat{x}),\nabla h_m(\hat{x}): j\in A(\hat{x}), m=1,\dots,M\}$ is lin. independent

Karush-Tucker Conditions: Let $\hat{x}$ be a regular local minimizer of (P). There exist unique $\hat{\lambda}\in\mathbb{R}^M$ and $\hat{\mu}\in\mathbb{R}_+^J$ such that
$\nabla f(\hat{x})+\sum_{j=1}^{J}\hat{\mu}_j\nabla g_j(\hat{x})+\sum_{m=1}^{M}\hat{\lambda}_m\nabla h_m(\hat{x})=0$ and $\hat{\mu}_j g_j(\hat{x})=0\ \forall j$.
If in addition, $f,g,h_m\in C^2(\mathbb{R}^N;\mathbb{R})$ then $y\cdot(\nabla^2 f(\hat{x})+\sum_{m=1}^{M}\hat{\lambda}_m\nabla^2 h_m(\hat{x}))y\geq 0\quad\forall y\in V^\perp$

The Convergence rate is given by ϕ_k.
Let (p_k) be positive with $\lim_{k\to\infty}p_k=0$. $\phi_k=O(p_k)$ if $C=\sup_{k\geq 0}\left[\frac{\phi_k}{p_k}\right]<\infty$

Or in other words if $\lim_{k\to\infty}\left[\frac{\phi_k}{p_k}\right]=0$

Examples: $\frac{1}{k^2+1}=o(\frac{1}{k})$, but $\frac{1}{k^2+1}\neq o(\frac{1}{k^2})$

Def: If $\phi_k=O(c^k)$ for some $c\in(0,1)$, we say ϕ_k converges linearly to zero with rate c.

Def: If $\phi_k\neq O(c^k)\ \forall c\in(0,1)$ then ϕ_k converges sublinearly.

A function $f:\mathbb{R}^N\to\mathbb{R}$ satisfies the **(PŁ) Inequality** with $\mu>0$ if $2\mu(f(x)-\min(f))\leq\|\nabla f(x)\|^2\quad\forall x\in\mathbb{R}^N$

Note: If f is μ-strongly convex, it satisfies the PŁ Inequality with constant μ.

Subgradient Method

Iterate $x_{k+1}=x_k-\gamma d_k$ with $d_k\in\partial f(x_k)$

Prop: Let $f:\mathbb{R}^N\to\mathbb{R}$ be convex and Lipschitz-continuous with constant M. Suppose $S:=\operatorname{argmin}(f)\neq\emptyset$, and let (x_k) be defined by the Subgradient method. Set $\bar{x}_k=\frac{1}{k+1}\sum_{j=0}^{k}x_j$. Then, $f(\bar{x}_k)-\min(f)\leq\frac{\gamma M^2}{2}+\frac{\operatorname{dist}(x_0,S)^2}{2\gamma(k+1)}$ **Note:** Complexity is ε^{-2}

Proximal-Gradient Method

Consider $f:\mathbb{R}^N\to\mathbb{R}$ convex and smooth, but $g:\mathbb{R}^N\to\mathbb{R}\cup\{\infty\}$ closed and convex.

We iterate $x_{k+1}=\operatorname{argmin}_z\{\gamma g(z)+\frac{1}{2}\|z-(x_k-\gamma\nabla f(x_k))\|^2: z\in\mathbb{R}^N\}$

Thm: Set $\gamma\leq\frac{1}{L}$ and define (x_k) by above. If $\operatorname{argmin}(f+g)\neq\emptyset$, then x_k converges, as $k\to\infty$ to a minimizer of $f+g$. Moreover, $\forall k\geq 1$, we have
$(f+g)(x_k)-\min(f+g)\leq\frac{\operatorname{dist}(x_0,\operatorname{argmin}(f+g))^2}{2\gamma k}$

Thm: Set $\gamma\leq\frac{2}{L}$ and define (x_k) by above. If $\Phi=f+g$ satisfies the PŁ Inequality with $\mu>0$, and $\operatorname{argmin}(\Phi)\neq\emptyset$, then
$\Phi(x_k)-\min(\Phi)\leq\frac{\Phi(x_0)-\min(\Phi)}{(1+\eta)^k}$, where $\eta=\mu(2-\gamma L)$