Lorenzo Zambelli

web: lorenzozambelli.it

Intro to Optimization Cheat Sheet

First Optimality Condition

Th: (Fermat's Rule) Let $f: \mathrm{Dom}(f) \subset \mathbb{R}^n \to \mathbb{R}$ and let $\emptyset \neq C \subset \mathrm{Dom}(f)$ be convex.
If $\hat{x} \in C$ is such that $f(\hat{x}) \le f(y)$ $\forall y \in C$ and if f is diff. at $\hat{x}$, then
$$\langle \nabla f(\hat{x}), y - \hat{x} \rangle \ge 0 \quad \forall y \in C$$
If, moreover, $\hat{x} \in \mathrm{int}(C)$, then $\nabla f(\hat{x}) = 0$

Second Optimality Condition

Let $f: \mathrm{Dom}(f) \to \mathbb{R}$ be twice diff. at $\hat{x}$
i) If $\hat{x}$ is a local minimizer of f, then $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x}) \ge 0$
ii) If $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x}) > 0$, then $\hat{x}$ is a strict local minimizer of f

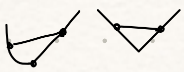
Convex Functions

Def: $f: \mathrm{Dom}(f) \to \mathbb{R}$ is convex if $\mathrm{Dom}(f)$ is convex and
$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$$
$\forall x, y \in \mathrm{Dom}(f)$ and $\forall \lambda \in (0,1)$

Prop: Local minimizer of convex functions are global minimizer

Prop: A diff. f is convex if, and only if $\mathrm{Dom}(f)$ is convex and
$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x)$$
$\forall x, y \in \mathrm{Dom}(f)$

Prop: $f: \mathbb{R}^n \to \mathbb{R}$ convex, L-smooth
$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x)$$
$$f(y) \le f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$
$$(\nabla f(y) - \nabla f(x)) \cdot (y - x) \ge \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|^2$$
$\forall x, y \in \mathbb{R}^n$

Ex:  convex 

Descent Lemma

If f is L-smooth and $\mathrm{Dom}(f)$ is convex, then
$$f(y) \le f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$
for all $x, y \in \mathrm{Dom}(f)$

Existence of Solution + L.S.C.

Def: A function is lower semi continuous (L.S.C) at $x \in \mathrm{Dom}(f)$ if
$$f(x) \le \liminf_{n \to \infty} f(x_n)$$
for every sequence x_n converging to x

Prop.: The following are equivalent:
i) f is L.S.C (at every $x \in \mathrm{Dom}(f)$)
ii) $\forall \gamma \in \mathbb{R}$, the γ-sublevel set $[f \le \gamma] := \{x \in \mathrm{Dom}(f) : f(x) \le \gamma\}$, is closed

Thm: Let $f: \mathrm{Dom}(f) \subset \mathbb{R}^n \to \mathbb{R}$ be L.S.C. If $[f \le \gamma]$ is bounded for some $\gamma > \inf f(t)$, the f attains its min.

Cor: i) If f is L.S.C and $\mathrm{Dom}(f)$ is bounded, then f attains its minimum
ii) If f is L.S.C. and C is compact, then f attains its minimum over C.

Remark: If $\lim_{\|x\| \to \infty} f(x) = \infty \Rightarrow [f \le \gamma]$ bounded (or empty) $\forall \gamma \in \mathbb{R}$

Strongly Convex Functions

Def: $\mu > 0$, $f: \mathrm{Dom}(f) \to \mathbb{R}$ is μ-strongly convex if $\mathrm{Dom}(f)$ is convex and
$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x - y\|^2$$
$\forall x, y \in \mathrm{Dom}(f)$, $\lambda \in (0,1)$ or equivalently if
$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x) + \mu\|x - y\|^2$$
$\forall x, y \in \mathrm{Dom}(f)$

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

Ex:  strong convex

Descend Methods

$$x_{n+1} = x_n - \alpha_n d_n$$

→ **Gradient:**
$$d_n = \nabla f(x_n)$$
$$0 < \alpha_n < 2/L$$

Conv: $O\left(\frac{1}{N}\right)$
Str. conv: $O(q^n)$ $q \in (0,1)$
General: $\min \|\nabla f(x_n)\| = O(1/\sqrt{n})$

Other choice of d_n
$a\|d_n\|^2 \le \|\nabla f(x_n)\|^2 \le -b \nabla f(x_n) d_n$

→ **Newton's method**
$$d_n = [\nabla^2 f(x_n)]^{-1} \nabla f(x_n)$$
Other choices of α_n:
- $\alpha_n \equiv \alpha \in (0, 2/L)$
- $\min_{\alpha > 0} f(x_n + \alpha d_n)$
- $\alpha_n \to 0$, $\sum \alpha_n = \infty$
- Backtracking: Armijo, Goldstein

→ **Conjugate**
$$d_i \cdot A d_j = 0$$
- If A has at most k eing. algorithm ends in k steps and function is quadratic

Non quadratic: unknown.

→ **Quasi-Newton**
$$d_n = D_n \nabla f(x_n)$$
$$D_n \sim \nabla^2 f(x_n)^{-1}$$

→ **Acceleration:**
→ Nesterov's - Polyak
$$\begin{cases} y_n = x_n + \beta_n (x_n - x_{n-1}) \\ x_{n+1} = y_n - \alpha_n \nabla f(x_n) \end{cases}$$
$O(1/n^2)$

→ **Stochastic**

Nonsmooth Functions & Convex Subgradients

Def: $v \in \mathbb{R}^n$ is a subgradient of f at x if
$$f(y) \ge f(x) + \langle v, y - x \rangle \quad \forall y \in \mathbb{R}^n$$

Def: The subdifferential of f at x, denoted by $\partial f(x)$ is the set of all subgrad. of f at x

Fac: $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex, L.S.C, $y \in \mathbb{R}^n$, $\gamma > 0$
$$x \mapsto f(x) + \frac{1}{2\gamma}\|x - y\|^2$$
strongly convex and L.S.C.. Its subdiff. is
$$\partial f(x) + \frac{1}{\gamma}(x - y) \quad \forall x \in \mathrm{Dom}(f)$$
Then, its unique minimizer $\hat{x}$ satisfies
$$0 \in \partial f(\hat{x}) + \frac{1}{\gamma}(\hat{x} - y)$$
Note: Fact used to define optimality condition

Subgradient Method

$x_{n+1} = x_n - \alpha v_n$, $v_n \in \partial f(x_n)$ $\quad O(1/\sqrt{n})$

→ **Proximal method** $O\left(\frac{d^2}{2\gamma_n}\right) \le O\left(\frac{d^2}{2\sum \gamma_n}\right)$

(1)
$$x_{n+1} = \mathrm{Argmin}\left\{ f(x) + \frac{1}{2\gamma_n}\|x - x_n\|^2 \right\}$$
$$\Rightarrow 0 \in \partial f(x_{n+1}) + \frac{1}{\gamma_n}(x_{n+1} - x_n) \Rightarrow x_{n+1} \in \underbrace{(I + \gamma_n \partial f)^{-1} x_n}_{\mathrm{prox}_f(x_n)}$$

(2) Let $f = g + h$, $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex & L.S.C.
$h: \mathbb{R}^n \to \mathbb{R}$ convex and L-smooth
$$x_{n+1} = \mathrm{Argmin}\left\{ g(x) + h(x_n) + \langle \nabla h(x_n), x - x_n \rangle + \frac{1}{2\gamma}\|x - x_n\|^2 \right\}$$
$$\Rightarrow x_{n+1} = (I + \gamma \partial g)^{-1}(x_n - \gamma_n \nabla h(x_n)) \quad n \ge 0 \quad \gamma \in (0, 2/L)$$

$$O(1/n) \text{ or } O\left(1/\textstyle\sum \gamma_n\right)$$

Lorenzo Zambelli
web: lorenzozambelli.it

The Fenchel conjugate

Fenchel conj. of $f:\mathbb{R}^n \to \mathbb{R}$

$$f^*(x^*) = \sup_{x\in\mathbb{R}^n}\{x^*\cdot x - f(x)\}$$

Fenchel-Young inequality:

$$f(x)+f^*(x^*) \geq x^*\cdot x$$

with equality if, and only if, $x^* \in \partial f(x)$

Useful properties:

→ Legendre's reciprocity formula:

$$x^* \in \partial f(x) \iff x \in \partial f^*(x^*)$$

→ $\partial f^*(x^*) = (\partial f)^{-1} x^*$

→ if $f \leq g \Rightarrow f^* \geq g^*$

Fenchel-Rockafellar duality

Let $L \in \mathbb{R}^{m\times n}$, and let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ and $g:\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ be lower-semi.cont. and convex.

Primal problem: $\inf_{x\in\mathbb{R}^n}\{f(x)+g(Lx)\}$

α = opt. value; S := set of primal solutions

Dual problem: $\inf_{y\in\mathbb{R}^m}\{f^*(-L^*y)+g^*(y)\}$

α^* = opt. value; S^* := set of dual solution

Prop.: duality gap $\alpha+\alpha^*$ is nonnegative

Th: The following are equivalents; let $\hat{x}\in\mathbb{R}^n$, $\hat{y}\in\mathbb{R}^m$

i) $-L^*\hat{y} \in \partial f(\hat{x})$ and $\hat{y}\in\partial g(L\hat{x})$

ii) $f(\hat{x})+f^*(-L^*\hat{y}) = -L^*\hat{y}\cdot\hat{x}$ and $g(L\hat{x})+g^*(\hat{y}) = \hat{y}\cdot L\hat{x}$

iii) $f(\hat{x})+g(L\hat{x})+f^*(-L^*\hat{y})+g^*(\hat{y})=0$

iv) $\hat{x}\in S$ and $\hat{y}\in S^*$ and $\alpha+\alpha^*=0$

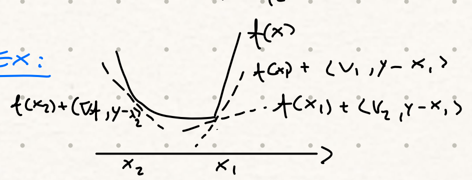
Subgradients properties

→ $0 \in \partial f(x) \iff x$ global min. of f

→ f diff $\partial f(x) = \nabla f(x)$

→ if $f = f_1+f_2+\dots+f_m$ and convex
$\Rightarrow \partial f = \partial f_1 + \partial f_2 + \dots + \partial f_m$

Ex:



Convergence rates and Complexity

Def: Let (p_k) be positive with $\lim_{k\to\infty} p_k = 0$
$\phi_k = \mathcal{O}(p_k)$ if $\phi_k \to 0$ at least as fast as p_k: $C := \sup_{k\geq 0}\left[\frac{\phi_k}{p_k}\right] < +\infty$

Def: on the other hand $\phi_k = o(p_k)$, $\phi_k \to 0$ strictly faster than p_k: $\lim_{k\to 0}\left[\frac{\phi_k}{p_k}\right] = 0$

Primal-Dual method

Prob: $\min\{f(x)+h(x)+g(Lx)\}$

where: $L\in\mathbb{R}^{m\times n}$, $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and $g:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ are l.c.s convex
$h:\mathbb{R}^n\to\mathbb{R}$ convex and L-smooth

→ **Case I: $h\equiv 0$**

op. conditions: $-L^*\hat{y}\in\partial f(\hat{x})$ and $\hat{y}\in\partial g(L\hat{x})$

Chambolle-Pock 2011: with $\tau\sigma\|L\|^2\leq 1$

$$\begin{cases} x_{k+1} = (I+\tau\partial f)^{-1}(x_k - \tau L^* y_k) \\ y_{k+1} = (I+\sigma\partial g^*)^{-1}(y_k + \sigma L(2x_{k+1}-x_k)) \end{cases}$$

→ **Case II: General**

Condat-Vũ 2013:

$$\begin{cases} x_{k+1} = (I+\tau\partial f)^{-1}(x_k - \tau\nabla h(x_k) - \tau L^* y_k) \\ y_{k+1} = (I+\sigma\partial g)^{-1}(y_k+\sigma L(2x_{k+1}-x_k)) \end{cases}$$

→ **Three-operator splitting** Davis-Yin 2017

$$\begin{cases} x_k = (I+\tau\partial f)^{-1} z_k \\ y_k = (I+\tau\partial g)^{-1}(2x_k - z_k - \tau\nabla h(x_k)) \\ z_{k+1} = z_k + y_k - x_k \end{cases}$$

Optimization problems with eq. constraints

Let $f, h_1, h_2, \dots, h_m : \mathbb{R}^n \to \mathbb{R}$ cont. diff, and

$$(\mathcal{P}) \quad \min_{x\in C} f(x)$$

where the feasible set $C := \{x\in\mathbb{R}^n : h_m(x) = 0,\ m = 1,\dots,M\}$

Thm: Lagrange Multiplier Let $\hat{x}$ be a regular local minimizer of $(\mathcal{P})$. There is a unique $\hat{\lambda}\in\mathbb{R}^M$ called Lagrange multiplier vector, such that

$$\nabla f(\hat{x}) + \sum_{m=1}^{M}\hat{\lambda}_m\nabla h_m(\hat{x}) = 0$$

If in addition, $f, h_1, h_2, \dots, h_m \in C^2(\mathbb{R}^n;\mathbb{R})$, then

$$y\cdot\left(\nabla^2 f(\hat{x}) + \sum_{m=1}^{M}\hat{\lambda}_m\nabla^2 h_m(\hat{x})\right)y \geq 0$$

$\forall y\in V = \{\nabla h_1(\hat{x}),\dots,\nabla h_m(\hat{x})\}$

Thm: Let $f, h_1,\dots,h_m:\mathbb{R}^n\to\mathbb{R}$ be twice diff. and $\hat{x}\in C$, $\hat{\lambda}\in\mathbb{R}^M$ satisfy the Lagrange theorem $\forall y\in V^{\perp}\setminus\{0\}$. Then, $\hat{x}$ is a strict local minimizer of $(\mathcal{P})$

Optimization problems with inequ. constraints

Let $f, g_j, h_m \in C^1(\mathbb{R}^n,\mathbb{R})$ and consider the const. prob.

$$(\mathcal{P}) \quad \min_{x\in C} f(x)$$

where $C := \{x\in\mathbb{R}^N : h_m(x)=0,\ g_j(x)\leq 0\ \forall m,j\}$

Def: Active inequality constraints at a local solution $\hat{x}$ is $A(\hat{x}) = \{j : g_j(\hat{x}) = 0\}$

Thm: Karush-Kuhn-Tucker Let $\hat{x}$ a regular local minimizer of $(\mathcal{P})$. $\exists\,\hat{\lambda}\in\mathbb{R}^M$ and $\hat{\mu}\in\mathbb{R}^J_+$ Lagrange multiplier vectors 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 $\hat{\mu}_j = 0\ \forall j\notin A(\hat{x})$. If in addition, $f, g_j, h_m \in C^2(\mathbb{R}^n;\mathbb{R})$

$$y\cdot\left(\nabla^2 f(\hat{x}) + \sum_{m=1}^{M}\hat{\lambda}_m\nabla^2 h_m(\hat{x})\right)y \geq 0$$

$\forall y\in V^{\perp}$, where $V = \{\nabla g_j(\hat{x}), \nabla h_m(\hat{x}), j\in A(\hat{x}), m=1,\dots,M\}$

Note: $\phi_k = o(p_k) \Rightarrow \phi_k = \mathcal{O}(p_k)$

Note: if $\phi_k = \mathcal{O}(q^k)$, $q\in(0,1) \Rightarrow \phi_k$ converge q-linearly
if $\phi_k \neq \mathcal{O}(q^k) \Rightarrow \phi_k$ converges sublinearly.