Optimization

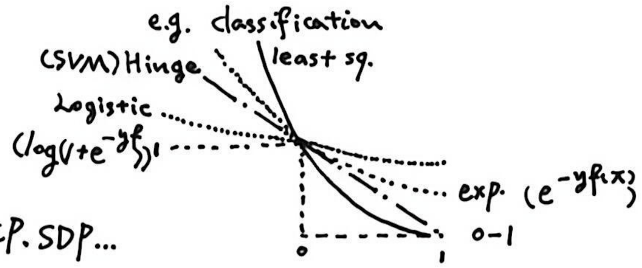
- general problem $\min f_0(x)$ s.t. $x \in X$

  (if $\inf_X f_0 = +\infty$: infeasible, $=-\infty$: unbd. from below)

  $X = \{f_i(x) \le 0,\ i=1,\dots,m;\ h_j(x)=0,\ j=1,\dots,p.\}$

  $f_0: \mathbb{R}^n \to \mathbb{R}$ LP. QP. SOCP. SDP...

  (structured $f(x)+\lambda r(x)$.) IntP. StoP..



o |

- affine set $\alpha x + (1-\alpha) y \in S, \forall x,y\in S, \alpha\in\mathbb{R}$. <=> $S = V + x_0$, V is a lin.sp.

  convex set $\dots, \alpha\in[0,1]$. (Pf of ⇒: show $S-S$ is lin.sp.)

  conv. comb. of $\{x_i\}_{i=1}^k$ is $\{y=\sum_{i=1}^k \alpha_i x_i \mid \sum_{i=1}^k \alpha_i = 1, \alpha_i \ge 0\}$. conv. hull $\mathrm{Conv}(S) = \bigcup_{\{x_i\}\subset S}$ conv.comb. of $\{x_i\}$ $= \bigcap_{S\subset C}$ conv. set C.

  cone $\alpha x\in K, \forall x\in K, \alpha\in[0,+\infty)$. conic comb. $\{\cdots \mid \alpha_i \ge 0\}$.

  polyhedron $\{x\mid Ax\le b, Cx = d\}$ (halfsp. s, hyperplanes) e.g. (prob) simplex.

  Prop. semi- pos.def. $\mathbb{S}_+^n$ is convex cone.

  (Euclidean) ellipsoid $\{x \mid (x-\bar x)' Q (x-\bar x) \le 1, Q \succeq 0\}$ ($Q = I$ for ball)

  norm cone $\{(x,t)\mid \|x\|\le t\}$ is second-~~ord.~~ Q cone.

- operations preserving conv.: $\cap$; affine func. $f(x)=Ax+b$.; perspective func., lin.frac.func.

  e.g. hyperbolic cone $\{x\mid x'Px \le (c'x)^2, c'x\ge 0, P\succeq 0\}$; $\{x \mid \sum_{i=1}^n x_i A_i \preceq B\}$. $t f(\frac{x}{t})$. $f(x) = \frac{Qx+u}{c'x+d}$, (dom $c'x+d>0$)

  Sep. hyperplane Thm.

  C and D are disjoint conv sets., $\exists a\ne 0, b$ s.t. $C\subset\{a'x\le b\}, D\subset\{a'x\ge b\}$. Supporting... C is conv, $\forall x_0\in\partial C$, $\exists a\ne 0$ s.t. $C\subset\{a'x\le a'x_0\}$.

- convex func $f: X\to\mathbb{R}$, dom f is conv. and $f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$, $\forall x_1,x_2\in\mathrm{dom}f$, $\alpha\in[0,1]$.

  (concave if $-f$ convex) strictly conv. ..<..

  (extended value func $\tilde f: X\to\bar{\mathbb{R}}$ with $\tilde f(X\setminus \mathrm{dom} f) = +\infty$)

  epigraph $\mathrm{epi}(f) = \{(x,t)\mid f(x)\le t\}$ Prop. epi(f) is conv. <=> f is conv.

  (t-sublevel set of f $L_t(f) := \{x\mid f(x)\le t\}$ f conv. => $L_t(f)$ conv. ⇍.)

  Jessen ineq. f (ext.) is conv. <=> $f(\sum_{i=1}^k \alpha_i x_i) \le \sum_{i=1}^k\alpha_i f(x_i)$, $\forall \alpha\ge0, \sum_{i=1}^k\alpha_i=1, x\in\mathbb{R}^n$.

  (Jacobian) deriv. $\lim_{y\to x\in \mathrm{int}X} \frac{\|f(y)-f(x)-Df(x)(y-x)\|}{\|y-x\|} = 0$, $Dh(x) = Dg(f(x))Df(x)$ by chain rule. diffl $h=g\circ f$.

  gradient $\nabla f(x) = Df(x)'$. (see for matrix form) $= (\frac{\partial f(x)}{\partial x_1},\dots,\frac{\partial f(x)}{\partial x_n})'$.

  Prop. diffl f is conv. <=> $f(y) \ge f(x) + \nabla f(x)'(y-x)$, $\forall x,y$. (first-ord) (strictly ..>..)

  Hessian $\nabla^2 f(x) = (\frac{\partial^2 f(x)}{\partial x_i\partial x_j})_{i,j} = D\nabla f(x) \in \mathbb{S}^n$.

  Prop. 2 cont. diffl f is conv. <=> $\nabla^2 f(x)\succeq 0, \forall x$. ($\partial^2 f$ cont.) (second-ord) (strictly ..>..)

  (Pf: Taylor's. and $\nabla^2 f$'s cont.)

  e.g. conv func: $x'Qx$ with $Q\succeq0$; e^x, x^a ($x>0$, $a\in\mathbb{R}\setminus[0,1]$); $x\log x$; $\|x\|_p$ ($p\ge1$); $\frac{x^2}{y}$.

- operations preserving conv.: (for func.) $\sum$ or $\int_A w(y) f(x,y)dy = g(x)$ conv.; (if $w(y)\ge0$, $f(x,y)$ conv. in x for $\forall y\in A$.) $Af+b$, $tf(\frac{x}{t})$,

  $f(x) := \max_i\{f_i(x)\}$, (f_i conv.) $g(x) := \sup_{y\in A} f(x,y)$; ($f(x,y)$ conv. in x, $\forall y\in A$.)

  ; $f(x) = h(g_i(x))$ if g_i conv., h conv $\nearrow_i$ or if g_i concave, h conv $\searrow_i$ $\forall i$.;

  e.g.: $\sum_{i=1}^r x_{(i)}$, $\sup_{y\in A}\|x-y\|$, $\lambda_{max}(X)$; $t\log\frac{t}{x}$, $\log\sum_{i=1}^m e^{g_i(x)}$, Schur complement, $f(x,y)$ conv. in (x,y), A conv. $g(x) := \inf_{y\in A} f(x,y)$. $\inf_{y\in C}\|x-y\|$; $\|X\|_2$ ($\rho(X)$); $(\ \cdot\ C) \succeq 0$, $\succ 0$; $-\ln\det X$ (by ∇^2)

  $\tilde f(h) := f(x_0+th)$, $\forall x_0\in\mathrm{dom}f$, $t\in\mathbb{R}$. iff f is conv.

(synthesis: sum. affine. persp.; max & min; mono composition; restriction)

(conjugate $f^*(y) = \sup_x (y'x - f(x))$ is conv. e.g. $\frac{1}{2} y Q^{-1} y$ to $\frac{1}{2} x Q x$, $Q \in \mathcal{S}^n_{++}$)

∘ 2

- convex optimization problem f_0, f_i are conv. and h_j is affine.
  $\Rightarrow$ $\forall$ locally optim. solution is global.

(LP) $\min\ c'x + d$ s.t. $Gx \le h$, $Ax = b$.

Prop. lin-frac program is equiv. to LP.

$\min_x \frac{c'x+d}{e'x+f}$ s.t. $Gx \le h$, $Ax = b$. $\Leftrightarrow$ $\min_{y,z}\ c'y + dz$ s.t. $Gy \le hz$, $Ay = bz$, $e'y + fz = 1$, $z \ge 0$.

(QP) $\min \frac{1}{2} x'Px + q'x + r$ s.t. $Gx \le h$, $Ax = b$. $(P \in \mathcal{S}^n_+)$ $\in \{e'x + f > 0\}$

e.g. least-sq., $x^* = A_L^\dagger b$. Markowitz, SVM $\min \frac{1}{2}\|w\|^2 + \mu \sum_i \max\{1 - y_i(w'x_i + b), 0\}$.

(QCQP) $\min\ \frac{1}{2} x'P_0 x + q_0'x + r_0$
s.t. $\frac{1}{2} x'P_i x + q_i'x + r_i \le 0$, $i = 1, \cdots, m$
$\frac{1}{2} x'A_j x + b_j'x + c_j = 0$, $j = 1, \cdots, p$
($A_j \equiv 0$, $P_0, P_i \succeq 0$ for convex)

(SOCP) $\min f'x$
s.t. $\|A_i'x + b_i\| \le c_i'x + d_i$, $i = 1, \cdots, m$
$Fx = g$. (general than LP and QCQP)

- robust LP
  deterministic $\min c'x$ s.t. $a_i'x \le b_i$, $\forall a_i \in \mathcal{A}_i$.
  (e.g. $\mathcal{A}_i = \{\bar{a}_i + P_i u \mid \|u\| \le 1\}$, $\Rightarrow$ SOCP $\bar{a}_i'x + \|P_i'x\| \le b_i$)
  or stochastic $\min c'x$ s.t. $P(a_i'x \le b_i) \ge \eta_i$.
  (e.g. $a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$, $\Rightarrow$ SOCP $\bar{a}_i'x + \Phi^{-1}(\eta_i)\|\Sigma_i^{\frac{1}{2}} x\| \le b_i$)
- CVX. Disciplined convex programming ruleset.

∘ 3

- uncons. optimality condition:
  Prop. $f: X \subset \mathbb{R}^n \to \mathbb{R}$ is cont. diffl., (Pf by 1-Taylor $O(\alpha_0)$)
  if $\exists d \in \mathbb{R}^n$ s.t. $\nabla f(x)'d < 0$, then $\exists \alpha_0 > 0$ s.t. $f(x + \alpha d) < f(x)$, $\forall \alpha \in (0, \alpha_0)$ (descent direction)

  Prop. $f: X \subset \mathbb{R}^n \to \mathbb{R}$ is 2cont. diffl.,
  $\bar{x}$ is local min $\Rightarrow \nabla f(\bar{x}) = 0$, $\nabla^2 f(\bar{x}) \succeq 0$.
  $\nabla f(\bar{x}) = 0$, $\nabla^2 f(\bar{x}) \succ 0 \Rightarrow \bar{x}$ is local min.

  Col. local min $\bar{x}$ has $\nabla f(\bar{x}) = 0$.
  Col. $+ f$ is conv. then $\bar{x}$ is global min iff. $\nabla f(\bar{x}) = 0$.

- Fritz-John necessary cond. $\bar{x}$ is a local min of (P) $\min f(x)$ s.t. $g_i(x) \le 0$, $h_j(x) = 0$, $\exists \lambda_i, \nu_j \in \mathbb{R}$ s.t.
  (FJ) $u \nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^p \nu_j \nabla h_j(\bar{x}) = 0$, (Lagrange multiplier)
  $u, \lambda_i \ge 0$, $\lambda_i g_i(\bar{x}) = 0$, $(u, \lambda_i, \nu_j) \ne 0$.

  Karush-Kuhn-Tucker cond. Suppose $\bar{x}$ is regular $\{\nabla g_i(\bar{x})\}_{i \in I} \cup \{\nabla h_j(\bar{x})\}_{j=1}^p$ is lin. ind. (LICQ) where $I = \{i \in \{1, \cdots, m\} \mid g_i(\bar{x}) = 0\}$ (active)
  (KKT) $\nabla_x L := \nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^p \nu_j \nabla h_j(\bar{x}) = 0$,
  (u can be set to 1) $(\bar{x}, \lambda, \nu)$
  $\lambda \ge 0$, $\lambda_i g_i(\bar{x}) = 0$. (complementary slackness) + primal cons. $g_i(\bar{x}) \le 0$, $h_j(\bar{x}) = 0$.

  Global minimum existence Thm. cont. f over $S \ne \phi$,
  (Weierstrass) compact S, $\exists$ global min & max.
  (coerciveness) $\lim_{\|x\| \to \infty} f(x) = +\infty$, closed S, $\exists$ global min.

  KKT for convex.: Slater cond. $\exists x' \in S$ s.t. $g_i(x') < 0$, $\forall i \in I$.
  (g_i convex and h_j affine) (also a CQ) (allow $g_i(x') = 0$ for affine g_i) $\Rightarrow \bar{x}$ satisfies KKT. (see for below)
  (if g_i concave, $\bar{x}$ satisfies KKT)
  ($+ f$ conv. to be VP.) sufficient cond. $+ f$ is conv., if $(\bar{x}, \bar{\lambda}, \bar{\nu})$ is a solution to KKT then it's an optimal to (P).
  (i.e. (VP) has KKT $\Leftrightarrow$ optim.)

  e.g. water-filling
  $\min \sum_{i=1}^n -\log(x_i + \alpha_i)$ s.t. $x \ge 0$, $1'x = 1$.

- $L(x,\lambda,\nu) = f(x) + \lambda' g(x) + \nu' h(x)$ dual func $\theta(\lambda,\nu) = \inf_{x\in X} L(x,\lambda,\nu)$

(P) $\inf_{x\in X} \sup_{\lambda\in\mathbb{R}^m_+, \nu\in\mathbb{R}^p} L(x,\lambda,\nu)$. since $\forall\varphi$, $\sup_y \inf_x \varphi(x,y) \leq \inf_x \sup_y \varphi(x,y)$, $\theta: \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R}$.

(Lagrangian) dual problem (D) $\sup_{\lambda\in\mathbb{R}^m_+, \nu\in\mathbb{R}^p} \theta(\lambda,\nu)$. (many duals since cons. can be treated as X or g, h)

Weak duality Thm. $\bar{x}$ to (P), $(\bar\lambda,\bar\nu)$ to (D), $\theta(\bar\lambda,\bar\nu) \leq f(\bar x)$.

Prop. θ is concave. (Pf: $-\theta$ is p.w. sup.)

Prop. (Fenchel) $\inf_x f_1(x) + f_2(x)$ has dual $\inf_w f_1^*(w) + f_2^*(-w)$.

(if it's $f_1(Ax)+f_2(x)$ then dual is $f_1^*(w) + f_2^*(-A'w)$.)

(Pf: let $f_1(y)+f_2(z)$ s.t. $y=z$ or $Az=y$.) (ADMM)

e.g. (LP') $\max -b'y$ s.t. $A'y+c \geq 0$ (note for feasible) or truncation

(SDP) $\min tr(CX)$ s.t. $tr(A_i X) = b_i$, $i=1,\cdots,m$, $X\in\mathcal{S}^n_+$. ($A_i, C\in\mathcal{S}^n$)

(SDP') $\max b'y$ s.t. $C - \sum_{i=1}^m y_i A_i \in \mathcal{S}^n_+$.

e.g. LASSO $\min \frac{1}{2}\|Ax-b\|^2 + \|x\|_1$ dual $\max -y'b - \frac{1}{2}\|y\|^2$ s.t. $\|A'y\|_\infty \leq 1$.

(note conj. of norm $\|\cdot\|_p$ is $\mathcal{I}_{\{\|\cdot\|_q\leq 1\}} := \begin{cases} 0 & \|x\|_q \leq 1 \\ \infty & \cdot\cdot > 1 \end{cases}$.)

($X\succeq 0$ in SDP can also be $-tr(ZX)$ in Lagrange) with $Z\succeq 0$.

Prop. LP, SOCP, SDP can be formulated as (CLP) $\min c\cdot x$ s.t. $a_i\cdot x = b_i$, $i=1,\cdots,m$, $x \succeq_{\mathcal{K}} 0$ (or $x\in\mathcal{K}$ closed convex cone)

(Pf by slackness e.g. $Gx+h\leq 0 \Rightarrow s\leq 0$, $s=Gx+h$)

- saddle point $(\bar x,\bar\lambda,\bar\nu) \in X\times\mathbb{R}^m_+\times\mathbb{R}^p$ s.t. $L(\bar x,\lambda,\nu) \leq L(\bar x,\bar\lambda,\bar\nu) \leq L(x,\bar\lambda,\bar\nu)$, $\forall x,\lambda,\nu \in ..$, $\Leftrightarrow$ strong duality. ($\bar x$ solves (P), $(\bar\lambda,\bar\nu)$ solves (D), while gap is 0.)

iff. (primal feas.) $\bar x\in X$, $g(\bar x)\leq 0$, $h(\bar x)=0$; (La. optim.) $\bar\lambda\geq 0$, $\bar x = \arg\min_x L(x,\bar\lambda,\bar\nu)$; (complementary) $\bar\lambda' g(\bar x) = 0$.

Strong duality for convex.: suppose (P) has a solution satisfying Slater's cond. $\Rightarrow$ (D) also has. and (D) = (P).

(X open & conv., f and g_i conv., h_j affine) & cont. diff'l. (by argmin becomes KKT)

Col. g, h affine then (D) = (P).

(synthesis: $\begin{cases} \text{feasibility} \to \text{truncate.} \\ \text{slacken (substitute) (pick off)} \\ \text{append (diagl.)} \end{cases}$) (e.g. $\begin{pmatrix} A & b \\ b' & c\end{pmatrix} \succeq 0 \Leftrightarrow x'Ax + 2b'x + c \geq 0$ $\forall x$. $A\succeq 0$)

∘ 4

- proper cone conv., closed, (solid) int $\neq\phi$, (pointed) $x, -x\in K \Rightarrow x=0$.

e.g. self-dual cones: orthant $\mathbb{R}^n_+$; Lorentz/SOC cone $\{(t,x) \mid t\geq\|x\|_2\}$; psd. cone $\mathcal{S}^n_+$.

dual cone $K^* = \{y \mid x'y\geq 0, \forall x\in K\}$ Prop. proper cone's dual is proper.

K-convex func. $\cdot\cdot \leq_K \cdot\cdot$ e.g. $f(X) = X^2 \in \mathcal{S}^n$ is $\mathcal{S}^n_+$-conv.

SDP relxation for QCQP: $\min f_0(x) = x'A_0x + b_0'x\ (+c_0)$ s.t. $f_i(x) = x'A_ix + b_i'x + c_i \leq 0$, $A_0, A_i \in\mathcal{S}^n$. $i=1,\cdots,m$

$\min A_0\cdot X + b_0'x$ s.t. $A_i\cdot X + b_i'x + c_i\leq 0$ $\begin{pmatrix} X & x \\ x' & 1\end{pmatrix}\succeq 0$ (drop the rank cons. ≤ 1)

have the same La. dual.

$\max \sum_{i=1}^m \lambda_i c_i - \tau$ s.t. $\lambda\geq 0$, $\begin{pmatrix} \sum_{i=0}^m \lambda_i A_i & \sum_{i=0}^m \lambda_i b_i \\ \cdot\cdot' & \tau \end{pmatrix} \succeq 0$. ($\lambda_0 = 1$)

tight cond.:

homogenous $b_i = 0$, only $X\succeq 0$.

$m\leq 2$ yields $r(X^*)\leq 1$, $\exists X^*\succeq 0$.; (S-Lem.) $m\leq 1$ and Slater's holds.

if $A_0 \succeq \lambda A_1$, $\exists\lambda>0$, then $\forall x$, $f_1(x)\geq 0 \Rightarrow f_0(x)\geq 0$;

conversely need + $\exists x_1$ s.t. $f_1(x_1) > 0$ (Slater)

- SOCP represetation:

$\min \|Fx+g\| \Rightarrow \min t$ s.t. $\|Fx+g\| \leq t$.

$\min \dfrac{\|Fx+g\|^2}{a'x+b}$ s.t. $a'x+b>0$ $\Rightarrow \min t$ s.t. $a'x+b>0$, $(Fx+g)'(Fx+g) \leq (a'x+b)t$.

Prop. SOCP can be formulated as (SDP) (by Schur's comp)

$\min f'x$ s.t. $\begin{pmatrix} c_i'x+d_i & (A_ix+b_i)' \\ A_i'x+b_i & (c_i'x+d_i)I \end{pmatrix} \succeq 0$, $Fx = g$.

- (RLP) robust $\forall a_i\in\mathcal{U}_i, b_i\in\mathcal{V}_i$. transform: reformulated as LP. ; safe tractable approx. of chance cons. (e.g. Hoeffding)

- LP Strong duality if (P) has an optim. solution x^*, then (D) has a y^*, s.t. $c'x^* = b'y^*$.

(std.) (P) $\min c'x$ s.t. $Ax=b$, $x \geq 0$. (D) $\max b'y$ s.t. $A'y \leq c$. (bd. from below & feas.)

CLP duality (weak) $b'\bar{y} \leq c \cdot \bar{x}$;

(P) $\min c'x$ s.t. $Ax=b$, $x \in \mathcal{K}$. (D) $\max b'y$ s.t. $c - A'y \in \mathcal{K}^*$. (or inn. prod.)

(Strong) if (P) bd. from below & strictly feas., then $p^* = d^*$, and (D) $\exists \bar{y}$ s.t. $b'\bar{y} = d^*$. ($\bar{x} \succ_{\mathcal{K}} 0$ or $\bar{x} \in \operatorname{int} \mathcal{K}$)

if (D) bd. from above & strictly feas., .. $\exists \bar{x}$ s.t. $c \cdot \bar{x} = p^*$.

Col. when (P) and (D) are both not strictly feas. there exists duality gap. (note inf and sup in P and D may not be attainable.)

(KKT can be generalized to CLP.) $Ax=b$, e.g. $x \in \mathcal{K}$, $\lambda \in \mathcal{K}^*$, $\lambda' x = 0$. of (P). (CVP. thus equiv. to optim.)

∘ 5

- iterative algo. for $\min_{x \in \mathbb{R}^n} f(x)$: (f diffl.) $x_{k+1} = x_k + t_k p_k$. descent direction $p_k' \nabla f(x_k) < 0$.

gradient descent method $p_k = -\nabla f(x_k)$. (Interpretation: $x_{k+1} = \arg\min_x f(x_k) + \nabla f(x_k)'(x - x_k) + \frac{1}{2t_k}\|x - x_k\|^2$.)

termination criteria e.g. $\|\nabla f_k\| \leq \varepsilon$. $\frac{\|x_{k+1} - x_k\|}{\max\{1, \|x_k\|\}} \leq \varepsilon$.

step size rules $t_k \equiv \alpha > 0$; $t_k = \arg\min_{t \geq 0} f(x_k + t p_k)$; (backtracking line search) Armijo.

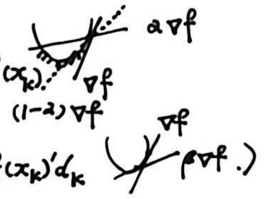
(exact line search)

$f(x_k + t_k p_k) \leq f(x_k) + \alpha t_k p_k' \nabla f(x_k)$

$t_k = \hat{t}\beta^s$, $\beta \in (0,1)$; $\alpha \in (0,1)$

(Armijo-Goldstein. $+ .. \geq f(x_k) + (1-\alpha) t_k p_k' \nabla f(x_k)$.

- Wolfe. $+ \nabla f(x_k + t_k p_k)' d_k \geq \beta \nabla f(x_k)' d_k$.)



- convergence analysis f is diffl., ∇f is L-Lipschitz cont., optim. value is bd. $C_L^{1,1}(\mathbb{R}^n)$

e.g. $x'Ax + b'x + c$ is Lip. on bd. set and its ∇ is Lip.

Prop. (quadratic upper bd.)

if $\nabla f \in C_L^{1,1}$, equiv.: i) $f(y) \leq f(x) + \nabla f(x)'(y-x) + \frac{L}{2}\|y-x\|^2$.
ii) $\frac{L}{2}x'x - f(x)$ is conv.

and when f is 2cont.diffl., ii) can be iii) $\nabla^2 f \preceq LI$, $\forall x$.

(Pf i) ⇒ ii) $f(y) - f(x) - \nabla f(x)'(y-x) = \int_0^1 (\nabla f(x + t(y-x)) - \nabla f(x))'(y-x)\,dt \leq \int_0^1 Lt\|y-x\|^2 dt$.)

Prop. if f is 2cont.diffl., $\nabla f \in C_L^{1,1} \iff \|\nabla^2 f\| \leq L$, $\forall x$. (stronger than $\nabla^2 f \preceq LI$)

(Pf by $\forall h$, $\frac{\|\nabla^2 f \cdot h\|}{\|h\|} \leq L$; ⇒

⇐ by $\nabla f(y) - \nabla f(x) = \int_0^1 \nabla^2 f(x + t(y-x))\,dt\,(y-x)$.)

Sufficient decrease Lem.

$\nabla f \in C_L^{1,1}$, then $f(x) - f(x - t\nabla f(x)) \geq t(1 - \frac{Lt}{2})\|\nabla f(x)\|^2$.

choice $\bar{t} = \frac{1}{L}$ to get $.. \geq \frac{1}{2L}\|\nabla f(x)\|^2$. (let $y = x - t\nabla f(x)$ in Prop.)

Convergence of grad. method: (of constant t) $f(x_{k+1}) < f(x_k)$ unless $\nabla f(x_k) = 0$, and $\nabla f(x_k) \to 0$, $k \to +\infty$.

Col. $\min_{0 \leq i \leq k} \|\nabla f(x_i)\| \leq \sqrt{\frac{f(x_0) - f^*}{(k+1)M}}$ where $f^* = \lim_i f(x_i)$, $M = t(1 - \frac{Lt}{2})$.

- ... for convex.: Prop. conv. f with min f attained by some x^*, $f(x_k) - f(x^*) \leq \frac{1}{2kt}\|x_0 - x^*\|^2$. (assume $tL \leq 1$)

(Pf: $f(x_{k+1}) \leq f(x^*) + \nabla f(x_k)'(x_k - x^*) - \frac{t}{2}\|\nabla f(x_k)\|^2 = f(x^*) + \frac{1}{2t}(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2)$ and note $f(x_{k+1}) < f(x_k)$.)

Prop. (control the grad) Col. iter # of $O(\frac{1}{\varepsilon})$.

$\|\nabla f(x_k)\|^2 \leq \frac{L}{kt}\|x_0 - x^*\|^2$ (Pf by Col. $f(x^*) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$.)
(let $y = x_k - \frac{1}{L}\nabla f(x_k)$ in upper bd.)

for backtracking ($\alpha = \frac{1}{2}$) $f(x_k) - f(x_k + t_k p_k) \geq t_k \alpha \|\nabla f(x_k)\|^2$, t_k is bd. by $t_{\min} = \min\{\hat{t}, \frac{\beta}{L}\} \leq t_k$, $\forall k$.

(modify the above Props. to get Armijo.)

Prop. (quadratic grad lower bd.)

if $\nabla f \in C_L^{1,1}$, conv. f, then $f(y) \geq f(x) + \nabla f(x)'(y-x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2$.

(Pf by using Col. for $g_x(y) = f(y) - \nabla f(x)'y$ which is conv. and $\nabla g_x \in C_L^{1,1}$)

- ... for strongly convex.: Prop. μ-sg.conv.f has (quadratic lower bd.)

($f(x) - \frac{\mu}{2}x'x$ is conv.) ($\mu > 0$)

$f(y) \geq f(x) + \nabla f(x)'(y-x) + \frac{\mu}{2}\|y-x\|^2$,

and when f is 2diffl., it has $\nabla^2 f \succeq \mu I$.

Prop. (lin. convergence) $f(x_k) - f(x^*) \leq (1 - \frac{\mu}{L})^k (f(x_0) - f(x^*))$.

(Pf by $f(x^*) \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$ by lower bd.) (also for Armijo. with $\alpha = \frac{1}{2}$, $\hat{t}L \leq 1$)

(synthesis: condition number $\frac{L}{\mu}$. iters to $f_k - f^* \le \varepsilon$ is at most $\approx \frac{L}{\mu}\log\frac{f_0 - f^*}{\varepsilon}$ for constant step size. ,)
(slow and may depend on scaling)

(convergence rate $e_k = \|x_k - x^*\|$ or $\|f_k - f^*\|$, $\mu := \lim_{k\to\infty} \frac{e_{k+1}}{e_k}$ $\begin{cases} =1 & \text{sub-lin.} \\ \in (0,1) & \text{lin.} \\ 0 & \text{super-lin.} \end{cases}$)

e.g. quadratic in super-lin. $\lim_k \frac{e_{k+1}}{e_k^2} = \mu_2 > 0$.

(supplements: double sufficient decrease Lem. $(tL \le 1)$
if $\nabla f \in C_L^{1,1}$, conv. f, then $f(x_k) - f(x_{k+1}) \ge \frac{t}{2}\|\nabla f(x_k)\|^2 + \frac{t}{2}\|\nabla f(x_{k+1})\|^2$.

low bd. $\xleftrightarrow{\text{Lip.}}$ upp. bd.
$\Delta_* f$
conv.

monotonicity. $(tL \le 2)$ $\nabla f(x_k)'\nabla f(x_{k+1}) \ge 0$ and $\|\nabla f(x_{k+1})\| \le \|\nabla f(x_k)\|$.)

• 6

(affine invariance)

• Newton step $x_{k+1} = x_k - \nabla^2 f(x_k)^{-1}\nabla f(x_k)$ (quadratic approx.) solve $\nabla f_k + \nabla^2 f_k p_k = 0$. (Cholesky $\frac{1}{3}n^3$, $\nabla^2 f = LL'$. Newton-CG ..)
($\nabla^2 f \succ 0$, decrease)

pure Newton may diverge. damped ~ $f(x_k) - f(x_k + t_k p_k) \ge \alpha t_k (\nabla f_k' \nabla^2 f_k^{-1} \nabla f_k)$. (Armijo.) (otherwise $t_k \leftarrow \beta t_k$, $t_0 = 1$)

($\eta_k \in (0,1)$ lin. inexact $\|r_k\| \le \eta_k \|\nabla f_k\|$
$\eta_k \to 0$ superlin. $O(\|\nabla f\|)$ is quad.)

convergence analysis. f is 2 cont. diff'l. (∇f is M-Lip. and sg. conv. with m) $MI \succcurlyeq \nabla^2 f \succcurlyeq mI$.
$\nabla^2 f$ is L-Lip. cont.

Prop. f is 2 cont. diff'l., $\nabla^2 f$ is L-Lip.
then $f(y) \le f(x) + \nabla f(x)'(y-x) + \frac{1}{2}(y-x)'\nabla^2 f(x)(y-x) + \frac{L}{6}\|y-x\|^3$.

$\exists\, 0 < \eta \le \frac{m^2}{L}$, $\gamma > 0$ s.t. $\begin{cases} \text{phase 1 (sublin.)} & f_{k+1} - f_k \le -\gamma \quad \|\nabla f_k\| \ge \eta. \text{ (damped)} \\ \text{phase 2 (quad)} & t_k \equiv 1,\ \frac{L}{2m^2}\|\nabla f_{k+1}\| \le \left(\frac{L}{2m^2}\|\nabla f_k\|\right)^2 \ .. < \eta. \text{ (pure)} \end{cases}$

Adv. fast conv., affine invar. (not conditioning)
or high dim.
DisAdv. $\nabla^2 f$ and solve $O(n^3)$, memory $O(n^2)$.

$\min m_k(p) \triangleq f_k + \nabla f_k' p + \frac{1}{2}p' B_k p$. $p_k = -B_k^{-1}\nabla f_k$ let B satisfies

$s_k \triangleq x_{k+1} - x_k = t_k p_k$,
$y_k \triangleq \nabla f(x_{k+1}) - \nabla f(x_k)$.
i.e. secant equation $B_{k+1}s_k = y_k$.
(or $s_k = H_{k+1}y_k$)

approx $B_k \to \nabla^2 f(x_k)$. $H_k \to \nabla^2 f(x_k)^{-1}$.
(auto) $\nabla m_{k+1}(0) = \nabla f(x_{k+1})$, $\nabla m_{k+1}(-t_k p_k) = \nabla f(x_k)$.

• quasi-Newton method

curvature condition (for pos. def. of B) $s_k' y_k > 0$.

DFP update $\min_{B_k} \|B - B_k\|_{F,W}$ s.t. $B = B'$, $Bs_k = y_k$.

Prop. Wolfe condition meets curv. condition in quasi-N.
(Pf: $y_k' s_k \ge (c_2 - 1) t_k \nabla f_k' p_k > 0$.)

$\Rightarrow B_{k+1} = (I - \rho_k y_k s_k')B_k(I - \rho_k s_k y_k') + \rho_k y_k y_k'$.

$\rho_k = \frac{1}{y_k' s_k}$. SMW formula yields $H_{k+1} = H_k - \frac{H_k y_k y_k' H_k}{y_k' H_k y_k} + \frac{s_k s_k'}{y_k' s_k}$. BFGS update H_k.
exchange B with H, y with s.

($B = A + UCV'$,
$B^{-1} = A^{-1} - A^{-1}U(C^{-1} + V'A^{-1}U)^{-1}V'A^{-1}$.) (bad condition if weaker than BFGS)

Starts $t_k = 1$, $c_1 = 10^{-4}$, $c_2 = 0.9$,
$H_0 = \frac{y_k' s_k}{y_k' y_k} I$.

... for BFGS: (Pf by Zoutendijk's Thm.)
f is 2 cont. diff'l. and $\{x : f(x) \le f(x_0)\}$ convex, $mI \le \nabla^2 f \le MI$, $\forall x \in \{\cdot\}$.

L-BFGS. $H_{k+1} = (I - \rho_k s_k y_k')H_k(I - \rho_k y_k s_k') + \rho_k s_k s_k'$ Store (up to m steps) s_i, y_i. $O(mn)$.
(two-loop recursion algo.) compact representation. ~.

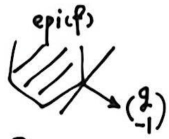
BB gradient method. $-t_k \nabla f_k$ approx. $-\nabla^2 f_k^{-1}\nabla f_k$. $\min \|s_{k-1} - D_k y_{k-1}\|$ yields $t_k = \frac{s_{k-1}' y_{k-1}}{y_{k-1}' y_{k-1}}$.
(approx. secant eq.) $\min \|D_k^{-1} s_{k-1} - y_{k-1}\|$ $t_k = \frac{s_{k-1}' s_{k-1}}{s_{k-1}' y_{k-1}}$.
($D_k = t_k I$ for H_k)

trust region. $\min m_k(p_k)$ s.t. $\|p_k\| \le \Delta_k$. (bd. from line search) (adjusting Δ in steps)

Cubically reg. Newton. (like grad. method) use $.. + \frac{L}{6}\|x - x_k\|^3$ in Prop. to min. $O(\varepsilon^{-\frac{3}{2}})$.

Gauss-Newton method. $\min_x \frac{1}{2}\|g(x)\|^2$. cont diff'l. $g: \mathbb{R}^n \to \mathbb{R}^m$.
linear approx. $\min \frac{1}{2}\|g(x_k) + J_k(x - x_k)\|^2$ + stepsize to get $x_{k+1} = x_k - t_k(J_k'J_k)^{-1}J_k' g_k$.
(i.e. Newton's. drop term in $\nabla^2 f$.)
(esp. $g(x) \approx 0$)
LM's: ($J_k' J_k + \mu_k I$ if not full-rank)

· 7



· subgradient. $f(y) \geq f(x) + g'(y-x), \forall y$. subdiff. $\partial f(x) = \{g \mid g'(y-x) \leq f(y) - f(x), \forall y\}$.

Calculus rules: $\alpha>0$, $\partial(\alpha f) = \alpha \partial f$.
$\partial(f_1+f_2) = \partial f_1 + \partial f_2$.
affine. $g(x) := f(Ax+b)$, $\partial g = A' \partial f(Ax+b)$.
(fin.) p.w. max. $f = \max_{i=1,\cdots,m} f_i(x)$, $\partial f = \mathrm{Conv} \cup \{\partial f_i(x) \mid f_i(x) = f(x)\}$.

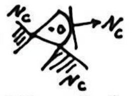
Prop. i) $0 \in \partial f(x_{min})$. ii) if f is conv. and dom(f) open, $\partial f(x) \neq \phi$ bd.
iii) + if f diffl. $\partial f(x) = \{\nabla f(x)\}$.

e.g. $f(x) = \|x\|_2$, $\partial f = \begin{cases} \frac{x}{\|x\|} & \neq 0 \\ \{g \mid \|g\|_2 \leq 1\} & = 0 \end{cases}$.

$f(x) = \frac{1}{2}\|x-y\|_2^2 + \lambda \|x\|_1$,
$\partial f = \begin{cases} x-\lambda & x<0 \\ [x-\lambda, x+\lambda] & x=0 \\ x+\lambda & x>0 \end{cases}$ soft-thres.

nondiffl. KKT (similar) $\min f_0$ s.t. $f_i \leq 0$.
if strong duality, $f_i^* \leq 0$; $\lambda^* \geq 0$; $0 \in \partial f_0(x^*) + \sum_i \lambda_i^* \partial f_i(x^*)$; $\lambda_i^* f_i(x^*) = 0$.

esp. $\min_{x \in C} f(x)$, let $I_C = \begin{cases} 0 & x \in C \\ \infty & \notin \end{cases}$, then $0 \in \partial f(x) + N_C(x)$.



e.g. proj. $\min_y \frac{1}{2}\|y-x\|^2 + I_C(y)$,
i.e. $(x-y)'(z-y) \leq 0, \forall z \in C$.

· subgradient method. proj. conv. f on closed conv. X. $y_{k+1} = x_k - \alpha_k g_k$, $g_k \in \partial f(x_k)$.
$x_{k+1} = \Pi_X(y_{k+1})$.
Lem. $\forall x \in X, y$, $\|\Pi_X(y) - x\|^2 \leq \|y-x\|^2$.

convergence analysis: conv. f on $\mathbb{R}^n$.
not descent. $f(x^*) = \inf f > -\infty$.
$f_{bs} \triangleq \min_i f_i(x)$, then $\|g\| \leq G$, $\forall g \in \partial f$.

Prop. $\|g\| \leq G$, $\forall g \in \partial f \iff f$ is G-Lip.
(Pf of $\Rightarrow$: $-g_x'(y-x) \geq f(x) - f(y) \geq g_y'(x-y)$.)

const. α: $f_{bs}^k - f^* \leq \frac{R^2 + G^2 k \alpha^2}{2k\alpha} \to \frac{G^2\alpha}{2}$. $\alpha_0 = \frac{R}{G\sqrt{k}}$. $O(\frac{RG}{\sqrt{k}})$.

const. length $\underbrace{\alpha\|g\|}_{:=\gamma}$: $\leq \frac{R^2+\gamma^2 k}{2\gamma k} G \to \frac{G\gamma}{2}$. $\gamma_0 = \frac{R}{\sqrt{k}}$. (R is dom f's diam.)

Polyak's step size.
(when f^* is known.) $\alpha_k = \frac{f_k - f^*}{\|g_k\|^2}$. it also, $\leq \frac{RG}{\sqrt{k}}$.

(synthesis:
nonsmooth + conv. $\frac{GD}{\sqrt{k}}$;
smooth (Lip ∇) + conv. $\frac{LD}{k}$;
sm. + sg. conv. $(1-\frac{\mu}{L})^k$, nonconv. $\frac{\cdots}{\sqrt{k}}$.)
nonsm. + sg. conv. $\frac{2G^2}{\mu k}$; Nesterov. $\frac{2LD^2}{k^2}$;
(in sm. + conv.)

e.g. find point in $\cap_j C_j$ i.e. $\min_x (\max_j d_j(x))$, use subgrad
(alternating projs.) $g_k = \frac{x_k - P_{C_j}(x_k)}{\|\cdot\cdot\|_2}$ (C_j is the farest set), $\alpha_k = f(x_k)$,
i.e. $x_{k+1} = P_{C_j}(x_k)$.

e.g. $Ax=b$, $x \geq 0$.
$S_1 = \{x : Ax = b\}$, $S_2 = \mathbb{R}_+^n$.
then $x_{k+1} = (x_k - A'(AA')^{-1}(Ax_k - b))^+$.

· 8

· $f = g + h$, diffl. g ($\mathrm{dom}(g) = \mathbb{R}^n$), nondiffl. but conv. h.

when $\min_{x \in X} g(x)$, interpretation of proj. grad. $\min_u g(x_{k-1}) + \nabla g(x_{k-1})'(u - x_{k-1}) + \frac{1}{2t_k}\|u - x_{k-1}\|^2 + I_X(u)$.

here $\min_x f(x)$, similarly $\min_u \cdots + h(u)$ i.e. $\min_u \frac{1}{2t_k}\|u - (x_{k-1} - t_k \nabla g(x_{k-1}))\|^2 + h(u)$

prox. mapping (operator) (local g(u) around x_{k-1}) $= \mathrm{prox}_{t_k h}(x_{k-1} - t_k \nabla g_{k-1})$. (e.g. LAsso)

$\mathrm{prox}_h(x) = \arg\min_u \frac{1}{2}\|u-x\|^2 + h(u)$. e.g. $h = I_C$, proj. on C; $h = \|x\|_1 \cdot \lambda$, soft-thres. (shrinkage);

$h = \|x\|_2 \cdot \lambda$, $\mathrm{prox}_h(x) = \begin{cases} (1 - \frac{t}{\|x\|_2})x & \|x\|_2 \geq t \\ 0 & < \end{cases}$;

$C = \mathrm{SOC}\ \{(x,t) \in \mathbb{R}^{n+1} : \|x\|_2 \leq t\}$, $P_C(v,s) = \begin{cases} 0 & \|v\| \leq -s \\ (v,s) & \|v\| \leq s \\ \frac{1}{2}(1 + \frac{s}{\|v\|})(v, \|v\|_2) & \|v\| > |s| \end{cases}$.

$\cdots$ of prox. grad.: g is m-sg. conv. and ∇g L-Lip.
lin. conv., $O(\frac{1}{k})$ with $t_k \equiv \frac{1}{L}$.

(Pf:) grad. mapping $G_t(x) := \frac{1}{t}(x - \mathrm{prox}_{th}(x - t\nabla g(x)))$,
verify $G_t(x) \in \nabla g(x) + \partial h(x - tG_t(x))$ (note but $G_t(x)$ is not subgrad. of f.)
and $G_t(x_{min}) = 0$. use g's L-upper bd. for $x - tG_t(x)$, g's m-lower bd. and ∂h's def. to get

Prop. $f(x - tG_t(x)) \leq f(z) + G_t(x)'(x-z) - \frac{t}{2}\|G_t(x)\|^2 - \frac{m}{2}\|z-x\|^2$.

Col. let $z = x$, $f(x') \leq f(x) - \frac{t}{2}\|G_t\|^2$; let $z = x^*$, $f' - f^* \leq \frac{1}{2t}(\|x-x^*\|^2 - \|x'-x^*\|^2)$.

thus $f_k - f^* \leq \frac{1}{2kt}\|x_0 - x^*\|^2$, $\|x_k - x^*\|^2 \leq (1 - \frac{m}{L})^k \|x_0 - x^*\|^2$. (from above)
$(0 \leq)$

line search.
$t \geq t_{min} = \min\{\hat{t}, \frac{\beta}{L}\}$.
(to satisfy g's bd with $\alpha = \frac{1}{2}$)
(same for $\frac{1}{2kt_{min}}$, $(1 - mt_{min})$)
(g's bd. i.e. sufficient decrease)

f_k is decreasing and $G_{t_k}(x_k) \to 0$, $k \to \infty$. (same as grad. method $h \equiv 0$)
faster than subgrad. useful when prox_h simple.