GRADIENTS

$x, a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$

$\langle a, x \rangle = a^T x = x^T a \to a$

$\langle x, x \rangle = x^T x = \|x\|^2 \to 2x$

$x^T A x \to (A + A^T)x$

$\|Ax - b\|^2 \to 2A^T(Ax - b) \to A^T A$

$\|Ax - b\| \to A^T(Ax - b)/\|Ax - b\|$

$Ax \to A$ if $Ax \neq b$

NORMS

$\|x + y\| \le \|x\| + \|y\|$ $(\Delta \neq)$

$\|x\|_\infty \le \|x\| \le \|x\|_1$, $\le \sqrt{n}\|x\|_\infty \le n\|x\|_\infty$

$|x \cdot y| \le \|x\|\|y\|$ (CSB) $x \parallel y$: bound attained

$|x \cdot y| \le \frac{1}{2t}\|x\|^2 + \frac{t}{2}\|y\|^2 \quad \forall t > 0$ (Young)

$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$ (Parallelogram)

$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|\cos\theta$ (L. of cosines)

$= \|x\|^2 + \|y\|^2 + 2x \cdot y$ (0 if $x \perp y$)

$2(a-b)\cdot(a-c) = \|a-b\|^2 + \|a-c\|^2 - \|b-c\|^2$

$\|x-y\|^2 = \|x-z+z-y\|^2 = \|x-z\|^2 + \|z-y\|^2 + 2\langle x-z, z-y\rangle$

FACTS ABOUT FUNCTIONS

$\frac{1}{2}\|Ax-b\|^2 = \frac{1}{2}\cdot(A^TA)x - (A^Tb)\cdot x$ — $\|Ax\|$ is coercive

- convex; strong if A nons: $\mu = \min\lambda$ of A^TA — $\frac{\mu}{2}\|x\|^2$ λ-strongly convex
- Hessian $A^TA \ge 0$
- $\ker(A) = \{0\} \leftrightarrow$ argmin singleton
- strictly convex if A full column rank
- satisfies PL — quadratic

EXISTENCE OF MINIMIZERS

cont + coercive → argmin $\neq \emptyset$

D closed, f lsc; D bounded or f coercive

f lsc + D compact → argmin $\neq \emptyset$

μstrong + closed

strong/strict convexity → at most 1 minim.

f closed + μ strong → at most 1 minim.

FUNCTION PROPERTIES

nonexpansive $\|f(x) - f(y)\| \le \|x - y\|$

superadditive $f(x) + f(y) \le f(x+y)$

coercive: $\lim_{\|x\|\to\infty} f(x) = \infty$: $\forall R > 0\ \exists \rho > 0$ s.t. $f(x) > R \quad \forall x \notin B(0,\rho)$

$\lim_{\|x\|\to\infty} \|x\| = \infty \leftrightarrow \lim_{k\to\infty} f(x_k) = \infty$

lsc: $\forall \varepsilon > 0\ \exists \delta > 0$ s.t. $f(y) \ge f(x) - \varepsilon \quad \forall y \in B(x,\delta) \cap D$

$f(x) \le \liminf_{k\to\infty} f(x_k) \quad \forall x_k \to x$

strongly convex → strictly convex → coercive → at most 1 minim.

Lipschitz cont.: $|f(x) - f(y)| \le L\|x - y\|$ — closed: epif closed

(strongly convex cannot be Lipschitz continuous)

f μstrong $\leftrightarrow f(x) - \frac{\mu}{2}\|x\|^2$ convex — cont → lsc → closed

μstrong + convex = μ strong — closed + continuous = closed

closed + closed = closed

midpoint convex + continuous → convex

midpoint convex: $f(\frac{x+y}{2}) \le \frac{fx+fy}{2}$

PL: $2\mu(f(x) - \min f) \le \|\nabla f(x)\|^2$

int $D \neq \emptyset$ if at least one pnt of cont → int(epif) $\neq \emptyset$

μstrong → μ-PL; and if $A \in \mathbb{R}^{m\times n}$ → $f \circ A$ PL

proper + lsc + μstrong → exactly 1 minimizer

FERMAT'S RULES

I) $f: D \subset \mathbb{R}^n \to \mathbb{R}$; $\bar{x} \in \operatorname{argmin} f \cap \operatorname{int} D \to \nabla f(\bar{x}) = 0$

(necessary: $\bar{x} \in \operatorname{int} D$, f diff at $\bar{x}$)

II) D open; $C \subset D$ nonempty, closed, convex, f diff

$\bar{x} \in \operatorname{argmin}_C f \to \nabla f(\bar{x})\cdot(y - \bar{x}) \ge 0 \quad \forall y \in C$

for affine $C = \{x_0\} + V \subseteq D$: $\bar{x} \in \operatorname{argmin} f \to \nabla f(\bar{x}) \perp V$

III) $f: \mathbb{R}^n \to \mathbb{R}$; $\bar{x} \in \operatorname{argmin} f \leftrightarrow 0 \in \partial f(\bar{x})$

IV) $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$; $\bar{x} \in \operatorname{argmin} f \leftrightarrow 0 \in \partial f(\bar{x})$

2ND ORDER OPTIMALITY: $f \in C^2$, $\bar{x} \in D$

$\bar{x} \in \operatorname{argmin} f \leftrightarrow \nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x}) \ge 0$

$\to \exists \varepsilon > 0$ s.t. $f(\bar{x}) < f(y)\ \forall y \in B(\bar{x}, \varepsilon)$

CONVEX RESULTS

bounded above → f constant

$\exists L_x, r_x > 0$ s.t. $|f(z) - f(y)| \le L_x|z - y| \quad \forall z, y \in B(x, r_x)$

epif and argmin convex

MATRICES AS FUNCTIONS $A \in \mathbb{R}^{m\times n}$

injectivity equivalences:
- $\ker A = \{0\}$
- lin indep columns
- (necessary: $N \le M$)

surjectivity equivalences:
- $\ker(A^T) = \{0\}$
- lin indep rows
- (necessary $N \ge M$)

bijectivity equivalences
- A invert. (hons.)
- $\ker A = \{0\}$
- $0 \neq$ eigval of A
- rows lin indep
- cols lin indep $(*)$
- $\det A \neq 0$
- (necessary: $N = M$)
- $(*) \leftrightarrow \mu > 0$; if $A = Q^TQ$, $\ker Q = \{0\}$

$\ker(A^TA) = \ker A = \operatorname{ran}(A^T)^\perp$

$\ker(A)^\perp = \operatorname{ran}(A^T)$

$\operatorname{ran}(AA^T) = \operatorname{ran}(A) =$ span of cols

adjoint: $\phi(u)\cdot v = u \cdot \phi^*(v)$

$(Au)\cdot v = u\cdot(A^Tv)$

Spectral Thm: for Q symm.

$\exists (v_i, \lambda_i)$ s.t. $\{v_i\}$ orthonormal basis of $\mathbb{R}^N$, and $Q = \sum \lambda_i v_i v_i^T$

$Q \ge 0$: $x^TQx \ge 0$

$Q \ge \mu I$: $Q - \mu I \ge 0$

$\ge$: λ nonneg; $>$: $\lambda > 0$

$L\|x\|^2 \ge \|Ax\|^2 = x^TA^TAx \ge \mu\|x\|^2$

condition no: $\kappa = L/\mu$ $(\mu > 0)$

A^TA symm and ≥ 0

$Q = A^TA \to x\cdot Qx = \|A\|^2$

CONVEX SETS: $C_1, C_2 \in \mathbb{R}^N$, $C_3 \in \mathbb{R}^M$ convex

$\cap C_i$ also convex (infinity also works)

$\alpha C_1 + C_2$, $C_1 \times C_3$, $A(C_1)$, $A^{-1}(C_3)$, $\operatorname{int} C$, $\bar{C}$ } also convex

$x \in \operatorname{int} C$, $y \in \bar{C} \to \operatorname{seg}[x,y) \subset \operatorname{int} C$

$\overline{\operatorname{int} C} = \bar{C}$, $\operatorname{int} C = \operatorname{int}(\bar{C})$ (nonempty int nec)

$C = \operatorname{conv}(C)$

if C not cont $\{0\} \to \exists v \neq 0$ s.t. $v \cdot x \le 0\ \forall x \in C$

C open, $0 \notin C$, $N \ge 2 \to \exists$ nontrivial subspace of $\mathbb{R}^N$ that doesn't intersect C

$\operatorname{conv}(C)$ = convex hull = all convex combos of elements of C

convex sets:

$\{x_0\} + V$, open/closed halfspace $[v \le \gamma] = \{x \in \mathbb{R}^N : v \cdot x \le \gamma\}$

$[v = \gamma]$ hyperplane

$\{x \mid Ax = b\} = \cap[v_i \le \gamma_i]$ polyhedron

MERIT FUNCTION

$\Phi(\bar{x}) = 0 \leftrightarrow \bar{x} \in \operatorname{argmin} f$

$\Phi(z) < \Phi(w) \leftrightarrow z$ better approx sol than w

Examples: $f(z) - \inf f$

$\operatorname{dist}(z, \operatorname{argmin} f)^2$

$\|z - \bar{x}\|^2$, where $\{\bar{x}\} = \operatorname{argmin} f$

$\|\nabla f(z)\|^2$, where f convex + differentiable

CONVERGENCE

Let $(p_k) \to 0$ positive

$\Phi_k = O(p_k)$: $\sup_{k\ge 0} \frac{\Phi_k}{p_k} < +\infty$

$= o(p_k)$: $\lim_{k\to\infty} \Phi_k/p_k = 0$

linear convergence: $\exists C > 0$, $Q \in [0,1)$ s.t. $\Phi(x_k) \le CQ^k$; $\Phi_k = O(q^k)$, $q \in (0,1)$

sublinear: $\Phi_k \neq O(q^k)\ \forall q \in (0,1)$, often still $O(k^{-p})$ $p > 0$

superlinear: $\Phi_k = o(q^k)\ \forall q \in (0,1)$

ratio test: order $p > 0$ rate $R \in (0,1)$

$\lim_{k\to\infty} \sup\left[\frac{\Phi_{k+1}}{\Phi_k^p}\right] = R$; order 1 rate $p \in (0,1)$: linear

polynomial convergence: $\Phi(x_k) \le C/k^p$, $p > 0$

DIFFERENTIABILITY

Gateaux: $\nabla f(x)\cdot h = f'(x;h) = \lim_{t\to 0} \frac{f(x+th) - f(x)}{t}$

gradient orthogonal to the level sets of the function

Fréchet: $\exists$ a vector $G(x) \in \mathbb{R}^n$ s.t. → Gateaux

$\lim_{h\to 0} \frac{|f(x+h) - f(x) - G(x)\cdot h|}{\|h\|} = 0$ → continuity

CONJUGATE DIRECTIONS

$\{d_0, \dots, d_{N-1}\}$ are conjugate wrt Q if $d_i \cdot Qd_j = 0\ \forall i \neq j$ → implies lin. indep.

We use conj. in. prod $\langle x, y\rangle_Q := x \cdot Qy$

orth. comp. $W^\perp = \{z : \langle z, w\rangle = 0\ \forall w \in W\}$

for kernel: $v \in \ker A \leftrightarrow -v \in \ker A$ so $z \cdot v \le 0$ and $z\cdot(-v) \le 0 \to z \cdot v \ge 0 \to z \cdot v = 0\ \forall v \in \ker A$

PROJECTIONS

$\hat{z} = \operatorname{Proj}_y(x) = \min_{z \parallel y} \|z - x\| = \frac{x\cdot y}{\|y\|^2} y$ $\quad \hat{z} \parallel y$ $\quad x - \hat{z} \perp y$

$\|x - P_C(x)\| = \min\{\|x - z\| : z \in C\} = \operatorname{dist}(x, C)$

only point s.t. $(x - P_C(x))\cdot(z - P_C(x)) \le 0\ \forall z \in C$

$\|P_C(x) - P_C(y)\| \le \|x - y\|$

$f \in C^2$

$\to \forall v \in \mathbb{R}^N$ $\lim_{t\to 0} \frac{1}{t^2}|f(x+tv) - f(x) - t\nabla f(x)\cdot v - \frac{t^2}{2}\nabla^2 f(x)v\cdot v| = 0$

Descent direction: $\exists \bar{\alpha} > 0$ s.t. $f(x + \alpha d) < f(x)$ $\forall \alpha \in (0, \bar{\alpha})$

$\theta = \cos^{-1}\left(\frac{x\cdot y}{\|x\|\|y\|}\right)$

SUBGRADIENTS (f convex)

$f(y) \geq f(x) + v \cdot (y-x) + \frac{\mu}{2}\|x-y\|^2$

$f \to \mathbb{R}$

$\partial f(x)$ $\partial f(y)$

$(x-y)\cdot(u-v) \geq 0 + \mu\|x-y\|^2$

$\partial f(x)$ always closed and convex; nonempty and bounded if domain $\mathbb{R}^n$

$\partial(f+g) = \partial f + \partial g$;

Piecewise: $f(x) = \max_i f_i(x) \to \partial f(x) = \prod \partial f_i(x)$

$f \to \mathbb{R} \cup \{+\infty\}$

$\partial f(x)$ always closed and bounded, nonempty and bounded if $x \in \text{int } D$ (empty if $x \notin \text{dom}$)

$\partial(f+g) \supseteq \partial f + \partial g$

equality holds if at least one of the fns is cont. at a point in the domain of the other

If f μ-strong, $\forall y^* \in \mathbb{R}^n$ $\exists x^* \in \mathbb{R}^N$ s.t. $y^* \in \partial f(x^*)$

Some examples:

$|x| \to \begin{cases} \{1\} & x>0 \\ [-1,1] & x=0 \\ \{-1\} & x<0 \end{cases}$

$\|x\| \to \begin{cases} \{\frac{x}{\|x\|}\} & x \neq 0 \\ \bar{B}(0,1) & x=0 \end{cases}$

$\max\{f_1(x),\dots,f_m(x)\}$, f_i convex $\to \mathbb{R}$

Let $a(x) = \{i : f(x) = f_i(x)\}$

$\partial f(x) = \text{conv}(\cup_{i \in a(x)} \partial f_i(x))$

$\|x\|_1$: piecewise (note this is the sum of $|x_1| + |x_2| + \dots$

Reciprocity Formula:

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

EXTENDED REAL-VALUED FUNCTIONS

f convex $\iff$ epi f convex $\to$ dom f, argmin f, $[f \leq \gamma]$ convex $\forall \gamma \in \mathbb{R}$

f strictly convex $\to$ argmin empty or singleton

convex + (strictly) convex = (strictly) convex

Closed equivalences:

f closed ; $\forall \gamma > \inf f$, $[f \leq \gamma]$ closed; $f|_{\text{dom } f}$ is lsc;

$\forall x \in \mathbb{R}^n$ and all $(x_k) \to x$, $f(x) \leq \liminf_{k\to\infty} f(x_k)$

affine minorant: f closed + convex $\to$ $\exists c \in \mathbb{R}^n, \alpha \in \mathbb{R}$ s.t. $f(x) \geq c \cdot x + \alpha$ $\forall x \in \mathbb{R}^n$

Indicator function: $\iota_C = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}$

dom $(\iota_C) = C$ epi $(\iota_C) = C \times [0, +\infty)$

ι_C closed $\iff$ C closed ι_C convex $\iff$ C convex

$\min\{f(x) : x \in C\} = \min\{f(x) + \iota_C(x) : x \in \mathbb{R}^n\}$

Normal cone: $\partial \iota_C(x) = \{N_C(x)$ if $x \in C$, $\emptyset$ if $x \notin C\}$

$N_C(x) = \{g \in \mathbb{R}^n : \langle g, y - x \rangle \leq 0 \ \forall y \in C\}$

Pt: $2\mu(f(x) - \min f) \leq \min\{\|v\|^2 : v \in \partial f(x)\}$ $\forall x \in \text{dom}(\partial f)$

μ-strong satisfies Pt

PROX AND f_z (f closed + convex)

$f_z(x) = f(x) + \frac{1}{2}\|x - z\|^2$

$z - \text{prox}_f(z) \in \partial f(\text{prox}_f(z))$ $\text{prox}_f(z) = \text{argmin}_x \{f_z(x)\}$

$\partial f_z(x) = \partial f(x) + x - z$ $\forall x, z$ f_z has unique minim. $\hat{x}$ s.t. $0 \in \partial f_z(\hat{x})$

Resolvent: $\text{prox}_f(z) = \underbrace{(I + \partial f)^{-1}}_{\text{resolvent of } \partial f}(z)$ $\|\text{prox}_f(x) - \text{prox}_f(y)\| \leq \|x - y\|$ so prox$_f$ continuous

$C \neq \emptyset$, closed and convex: $\text{prox}_{\iota_C}(x) = \text{Proj}_C(x)$

DUALITY

primal $\min_x \{f(x) + g(Ax)\}$ dual $\min_y \{f^*(-A^T y) + g^*(y)\}$

f, g closed + convex

Duality gap: $v + v^*$ strong duality: gap $= 0$, weak duality: gap ≥ 0

Strong duality equivalences:

- $\hat{x} \in S$, $\hat{y} \in S^*$
- $-A^T\hat{y} \in \partial f(\hat{x})$ and $\hat{y} \in \partial g(A\hat{x})$
- $f(\hat{x}) + g(A\hat{x}) + f^*(-A^T\hat{y}) + g^*(\hat{y}) = 0$
- $f(\hat{x}) + f^*(-A^T\hat{y}) = -A^T\hat{y} \cdot \hat{x}$ and $g(A\hat{x}) + g^*(\hat{y}) = \hat{y} \cdot A\hat{x}$

Moreover, if f cont and $\{g \circ A$ poly or g continuous$\}$: $\hat{x} \in S \to \exists \hat{y} \in \mathbb{R}^m$ s.t. all four statements above hold

Moreau's Identity: let g closed + convex, $\delta > 0$

$\forall y \in \mathbb{R}^m$ $\text{prox}_{\delta g^*}(y) = y - \delta\,\text{prox}_{\delta^{-1} g}(\delta^{-1} y)$

FENCHEL CONJUGATE

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

Biconjugate: $f^{**} = (f^*)^*$; $f^{**} \leq f$ with eq. if f closed, convex and dom $f \neq \emptyset$

$f \leq g \to f^* \geq g^*$ f closed + convex $\to f^*(y) > -\infty$ and not id. $+\infty$

f not id $+\infty \to f^*$ closed + convex

f closed, convex, not id $+\infty$: Fenchel-Young Ineq: $f(x) + f^*(x^*) \geq x^* \cdot x$

- $f^{**} = f$
- Reciprocity: $x^* \in \partial f(x) \iff x \in \partial f^*(x^*)$
- if $\mu L = 1$ then f μ-strong iff f^* L-smooth

(eq. iff $x \in \partial f(x)$) $f(x) = c \cdot x + \alpha$ $\to f^*(y) = \iota_C(y) - \alpha$

Support functional: $\emptyset = C$, closed, convex: $\delta_C(y) := \iota_C^*(y) = \sup\{y \cdot x : x \in C\}$

LAGRANGIANS

Let f diff, take $\min\{f(x) : Ax = b\}$

associated Lagrangian: $L(x, z) = f(x) + z^T(Ax - b)$ x primal variable, z dual or Lagrange multiplier

$\hat{x}$ solution and $\hat{z}$ corresp. mulhplier

$\hookrightarrow 0 = \nabla L(\hat{x}, \hat{z}) = \begin{pmatrix} \nabla f(\hat{x}) + A^T\hat{z} \\ A\hat{x} - b \end{pmatrix}$ primal feasability: $A\hat{x} = b$; dual $\dots$: $-A^T\hat{z} = \nabla f(\hat{x})$

augmented Lagrangian: $L_\alpha(x, \lambda) = f(x) + \lambda^T(Ax - b) + \frac{\alpha}{2}\|Ax - b\|^2$

MATHEMATICAL PROGRAMMING

$f, h_i \in C^1(\mathbb{R}^n; \mathbb{R})$ $(i \in \{1, \dots, M\})$

$\min_C\{f(x)\}$ $C = \{x : h_m(x) = 0\}$

- no ineq constraints: regular point $\hat{x} \in C$ s.t. $V = \{\nabla h_i(x)\}$ lin indep.

Lag Mulhplier Thm

$\hat{x}$ local minimizer $\to \exists \hat{\lambda} \in \mathbb{R}^M$ unique s.t.

$\nabla f(\hat{x}) + \sum_1^M \hat{\lambda} \nabla h_m(\hat{x}) = 0$ (if also $f, h_i \in C^2 \to$ $y(\nabla^2 f(\hat{x}) + \sum \hat{\lambda} \nabla^2 h_m(\hat{x}))y \geq 0$

and order: if $f, h_i \in C^2$, $\hat{x}, \hat{\lambda}$ as $\uparrow$ w strict inequality $\uparrow$ $\forall y \in V^\perp \setminus \{0\}$

Then $\hat{x}$ strict local minimizer

- General case: now $f, g_j, h_m \in C^1$ and we want $\{h_m(x) = 0, g_j(x) \leq 0 \ \forall m, j\}$

Active constraints: $A(\hat{x}) = \{j : g_j(\hat{x}) = 0\}$

regular point: $V = \{\nabla g_j(x), \nabla h_m(\hat{x}) ; j \in A(\hat{x})\}$ lin. indep.

Karush-Kuhn-Tucker: $\hat{x}$ local min. $\to$ $\exists$ unique $\hat{\lambda} \in \mathbb{R}^m$ and $\hat{\mu} \in \mathbb{R}^J_+$ s.t.

$\nabla f(\hat{x}) + \sum_1^N \hat{\mu}_j \nabla g_j(\hat{x}) + \sum_1^M \hat{\lambda}_m \nabla h_m(\hat{x}) = 0$ and $\hat{\mu}_j = 0 \ \forall j \notin A(\hat{x})$ so $\hat{\mu}_j \nabla g_j(\hat{x}) = 0 \ \forall j$

If also $f, g_j, h_m \in C^2 \to y \cdot (\nabla^2 f(\hat{x}) + \sum_1^M \hat{\lambda}_m \nabla^2 h_m(\hat{x}))y \geq 0$ $\forall y \in V^\perp$

GM/GD: $x_{k+1} = x_k - \gamma \nabla f(x_k)$; $\gamma \in (0, \frac{2}{L})$

Line search, greedy: $\gamma_k := \text{argmin}_{\gamma \geq 0 \text{ or } \gamma \in (0,\bar{\gamma})} f(x_k + \gamma d_k)$

Line search, backtracking: $\gamma_k = \bar{\gamma}\beta^{m_k}$ ($\bar{\gamma}$ initial guess, $m_k \geq 0$ smallest possible integer

- Armijo: $f(x_k + \gamma d_k) \leq f(x_k) + \delta\gamma \nabla f(x_k) \cdot d_k$, $\delta \in (0,1)$, typ. $\sim 10^{-4}$
- Goldstein: $f(x_k + \gamma d_k) \geq f(x_k) + (1-\delta)\gamma \nabla f(x_k) \cdot d_k$
- Weak Wolfe: $\nabla f(x_k + \gamma d_k) \cdot d_k \geq \tau \nabla f(x_k) \cdot d_k$, $\tau \in (\delta, 1)$, typ in $(0.1; 0.9)$
- Strong Wolfe: $|\nabla f(x_k + \gamma d_k) \cdot d_k| \leq \tau |\nabla f(x_k) \cdot d_k|$

Acceleration methods: add a momentum term $(x_k - x_{k+1})$

- Heavy-Ball $y_k = x_k + \beta(x_k - x_{k-1})$; $x_{k+1} = y_k - \alpha \nabla f(x_k)$
- Nesterov: $y_k = x_k + \beta(x_k - x_{k-1})$; $x_{k+1} = y_k - \alpha \nabla f(y_k)$ ($\mu > 0$) strongly convex

Conjugate Gradient: for quadr. fns. $\frac{1}{2}x^TQx - b^Tx + c$, Q sq, $\tau I \preceq \nabla f(x) = Qx - b \preceq LI$

(exp. subspace min.; line search along each direction)

$x_{k+1} = x_k + \alpha_k p_k$; $r_{k+1} = r_k + \alpha_k Q p_k$; $p_{k+1} = -r_{k+1} + \gamma_k p_k$; stops for $r_k = 0$

$\alpha_{-1} = 0$; $r_k = \nabla f(x_k)$; $\gamma_k = (r_{k+1}^T r_{k+1})/(r_k^T r_k)$; $\alpha_k = -(p_k^T r_k)/(p_k^T Q p_k)$

Q k dist λvals $\to$ termination in at most k steps

Newton's method: $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$; Quasi-N choose $D_k \sim \nabla^2 f(x_k)^{-1}$

Secant condition: $[\nabla^2 f(x_k)]^{-1}(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k$

Proj GM: $x_{k+1} = P_C(x_k - \gamma \nabla f(x_k))$ for $\min_{x \in C} f(x)$; f L-sm $\neq C$, closed, convex easy to project onto

Subgr. M $x_{k+1} = x_k - \gamma d_k$; $d_k \in \partial f(x_k)$ // $\text{prox}_{\gamma g}(x_k - \gamma \nabla f(x_k))$

PPM: $x_{k+1} = \text{prox}_{\gamma f}(x_k)$; PGM: $x_{k+1} = \text{argmin}_z \{g(z) + f(x_k) + \nabla f(x_k) \cdot (z - x_k) + \frac{1}{2\gamma}\|z - x_k\|^2\}$

PrDuM: $x_{k+1} = \text{prox}_{\tau f}(x_k - \tau \nabla h(x_k) - \tau A^T y_k)$; $y_{k+1} = \text{prox}_{\sigma g^*}(y_k + \sigma A(2x_{k+1} - x_k))$

$\tau\sigma\|A\|^2 + \frac{\tau L}{2} \leq 1$; $(x_k, y_k) \to (\hat{x}, \hat{y}) \Rightarrow \hat{x}, \hat{y}$ sols; $x_{k+1} = x_k, y_{k+1} = y_k \to (x_k, y_k)$ sols

---

MetMult: $x_{k+1} = \text{argmin}\{L_\alpha(x, \lambda_k) : x \in \mathbb{R}^n\}$; $\lambda_{k+1} = \lambda_k + \alpha(Ax_{k+1} - b)$ (dual feas. ent: $\nabla f(x_{k+1}) + A^T\lambda_{k+1} = 0$)

proximal: $x_{k+1} = \text{argmin}\{L_\alpha(x, \lambda_k) + \frac{\alpha}{2}\|x - x_k\|^2\}$, $\lambda_{k+1} = \lambda_k + \alpha(Ax_{k+1} - b)$

pred-cor: $p_{k+1} = \lambda_k + \alpha(Ax_k - b)$; $x_{k+1} = \text{argmin}\{L(x, p_{k+1}) + \frac{1}{2\alpha}\|x - x_k\|^2\}$; $\lambda_{k+1} = \lambda_k + \alpha(Ax_{k+1} - b)$ (normal L here!)

ADMM for $\min\{f(x) + g(y) : Ax + By = c\}$

$x_{k+1} = \text{argmin}\{L_\alpha(x, y_k, \lambda_k)\}$; $y_{k+1} = \text{argmin}\{L_\alpha(x_{k+1}, y, \lambda_k)\}$; $\lambda_{k+1} = \lambda_k + \alpha(Ax_{k+1} + By_{k+1} - c)$