convex set: $0\le\theta\le1$, $\theta x_1+(1-\theta)x_2\in C$

convex function: dom(f), $f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y)$

convex hull: $\sum\theta_i=1$, always convex: conv(C)

polar cone: $C^\circ=\{x: x^Ty\le0, \text{for all } y\in C\}$ always cvx

normal cone: $N_C(x)=\{g: g^T(y-x)\le0, \text{for all } y\in C\}$

tangent cone: $T_C(x)=N_C(x)^\circ$, closed, need not cvx

$(C^\circ)^\circ=C$, only close, cvx set

$N_C(x)$, for cvx set, is closed convex cone.

Three: zeroth: $f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y)$

differentiable, first order: $f(y)\ge f(x)+\nabla f(x)^T(y-x)$

twice-diff: second order: $\nabla^2 f(x)\succeq0$

Non-smooth: $f(y)\ge f(x)+\langle g_x, y-x\rangle$

$\|X\|_{op}=\sigma_1(X)$, $\|X\|_{tr}=\sum\sigma_i(X)$

cvx satisfies Jensen's $f(\mathbb{E}[X])\le\mathbb{E}f(X)$

Op that preserve: 1. nonnegative combination 2. pointwise max $g(x)=\sup_{s\in S}f_s(x)$ 3. partial min: $f(x)=\min_{y\in C}g(x,y)$

4. Affine composition 5. $f=h\circ g$, convex+↑, convex

Smoothness: $\|\nabla f(x)-\nabla f(y)\|\le\beta\|x-y\|$, $\frac{\beta}{2}\|x\|^2-f(x)$ → cvx

$f(y)\le f(x)+\nabla f(x)^T(y-x)+\frac{\beta}{2}\|y-x\|^2$

$\nabla^2 f(x)\preceq\beta I_d$

strong cvx: $f(x)-\frac{\alpha}{2}\|x\|^2$ → cvx

$f(y)\ge f(x)+\nabla f(x)^T(y-x)+\frac{\alpha}{2}\|y-x\|^2$

$\nabla^2 f(x)\succeq\alpha I_d$

strict: $f(\theta x+(1-\theta)y)<\theta f(x)+(1-\theta)f(y)$, $\cancel{\nabla^2 f(x)}\succ0$

$f(y)>f(x)+\nabla f(x)^T(y-x)$, for $x\ne y$.

optimality condition: unconstrained: $0\in\partial f(x^*)$

diff, constrained: $\nabla f(x^*)^T(y-x^*)\ge0$

$\min f(x)+\mathbb{I}_C(x)$ general, constrained: $0\in\partial f(x^*)+N_C(x^*)$

projection: $0\in x^*-y+N_C(x^*)$, $(y-x^*)^T(a-x^*)\le0$

$\|P_K(a)-P_K(b)\|\le\|a-b\|$, $(a-P_K(a))^T(P_K(a)-P_K(b))\ge0$ ···

GD: $x^{t+1}=x^t-\eta\nabla f(x^t)$

1. Fixed. 2. Exact Line: $\eta^t=\arg\min_{\eta\ge0}f(x-\eta\nabla f(x))$

3. Backtracking: $\alpha\in(0,0.5)$, $\beta\in(0,1)$, initialize $\eta=1$

if $f(x-\eta\nabla f(x))\le f(x)-\alpha\eta\|\nabla f(x)\|_2^2$, else $\eta=\beta\eta$.

Least square: $\min\frac12\|Ax-b\|_2^2$, $S=A^TA$

$\eta=\frac{2}{\lambda_{max}+\lambda_{min}}$ $k(S)=\frac{\lambda_{max}(S)}{\lambda_{min}(S)}<\infty$. $\hat{x}=(A^TA)^{-1}A^Tb$.

$x^k-\hat{x}=[I-\eta A^TA]^k(x^0-\hat{x})$

$\|x^k-\hat{x}\|_2\le\|I-\eta A^TA\|_{op}^k\|x^0-\hat{x}\|_2\to\left(\frac{k-1}{k+1}\right)^k$..

$f(x^k)-\hat{f}\le\frac{\lambda_{max}}{2}\left(\frac{k-1}{k+1}\right)^{2k}\|x^0-\hat{x}\|_2^2$

non-convex, β-smooth: $\boxed{f(x^{t+1})\le f(x^t)-\frac{\eta}{2}\|\nabla f(x^t)\|_2^2}$, $\eta\le\frac{1}{\beta}$

$\boxed{\|\nabla f(x)\|_2\le\sqrt{\frac{2\beta}{k}(f(x^0)-f(x^*))}}$

proof: $f(x^{t+1})-f(x^t)\le-\frac{f(x^0)-f(x^*)}{k}\Rightarrow f(x^k)-f(x^0)\le f(x^*)-f(x^0)$

Convex-β-smooth: $\boxed{f(x^k)-f(x^*)\le\frac{\beta\|x^0-x^*\|^2}{2k}}$

proof: $\|x^{t+1}-x^*\|_2^2=\|x^t-x^*\|_2^2-2\eta\nabla f(x^t)^T(x^t-x^*)+\eta^2\|\nabla f(x^t)\|_2^2$

by cvx $\nabla f(x^t)^T(x^t-x^*)\ge f(x^t)-f(x^*)$

by β-smooth. $\eta^2\|\nabla f(x^t)\|^2\le2\eta(f(x^t)-f(x^{t+1}))$

$2\eta(f(x^{t+1})-f(x^*))\le\|x^t-x^*\|_2^2-\|x^{t+1}-x^*\|_2^2$

α-strong cvx, β-smooth: $\boxed{\|x^k-x^*\|_2^2\le(1-\frac{\alpha}{\beta})^k\|x^0-x^*\|_2^2}$

proof: by α-strong: $\nabla f(x^t)^T(x^*-x^t)\le f(x^*)-f(x^t)-\frac{\alpha}{2}\|x^t-x^*\|^2$

$\|x^{t+1}-x^*\|_2^2\le(1-\frac{\alpha}{\beta})\|x^t-x^*\|_2^2$, $\log(\frac{1}{\epsilon})$ linear.

Subgradients: $f(y)\ge f(x)+g_x^T(y-x)$.

$f(x)=\|x\|_1$, $\partial f(x)=\text{sign}(x)$

$f(x)=\|x\|_2$, $\partial f(x)=\frac{x}{\|x\|_2}$, when $x=0$, $\|y\|\ge g_0^Ty$ any $\|g_0\|_2\le1$. C-S

$f(x)=\mathbb{I}_C(x)$, $\partial f(x)=N_C(x)$

LASSO: $x^*=\arg\min_{x\in\mathbb{R}^d}\frac12\|b-Ax\|_2^2+\lambda\|x\|_1$

$-A^T(b-Ax^*)+\lambda\,\text{sign}(x^*)=0$

subgradient method: $x^{t+1}=x^t-\eta g_{x^t}$

Fixed: $\eta=\frac{1}{\sqrt{k}}$, step-schedule: $\sum_t^\infty\eta_t=\infty$, $\sum\eta_t^2<\infty$

Assume objective G-Lipschitz, $|f(x)-f(y)|\le G\|x-y\|_2$

$\|g_x\|_2^2\le f(x)-f(x-g_x)\le|f-f|\le G\|g\|_2\Rightarrow\|g\|_2\le G$

$\boxed{f(x^{best})-f(x^*)\le\frac{\|x^0-x^*\|_2^2+G^2\sum\eta_t^2}{2\sum_0^{k-1}\eta_t}}$

$\eta_t=\frac{R}{G\sqrt{k}}$ $\le\frac{GR}{\sqrt{k}}$ $f(x^{best})-f(x^*)\to0$

$\eta_t=\frac{1}{\sqrt{t}}$, $f(x^{best})-f(x^*)\lesssim\frac{G^2\log(k)+R^2}{\sqrt{k}}$

Orade: $t\le(d-1)/2$

convex, L-Lip: $\min_{1\le s\le k}f(x^s)-\min_{\|x\|_2\le R}f(x)\ge\frac{RL}{2\sqrt{k}}$,

subg-optimal.

convex, β-smooth: $\min_{1\le s\le k}f(x^s)-f(x^*)\gtrsim\frac{\beta\|x^0-x^*\|_2^2}{k^2}$

α-strong, β-smooth. $\min_{1\le s\le k}f(x^s)-f(x^*)\gtrsim\alpha\left(\frac{\sqrt{k}-1}{\sqrt{k}+1}\right)^{2(k-1)}\|x^0-x^*\|_2^2$ ✓

PGD: $y^{t+1}=x^t-\eta\nabla f(x^t)$, $x^{t+1}=P_C(y^{t+1})=\arg\min_{x\in C}\frac12\|x-y^{t+1}\|^2$

proj

convex, G-Lip $\boxed{f(x^{best})-f(x^*)\le\frac{\|x^0-x^*\|_2^2+G^2\sum\eta_t^2}{2\sum_0^{k-1}\eta_t}}$

proof: $\|y^{t+1}-x^*\|^2=\eta^2\|g(x^t)\|_2^2+\|x^t-x^*\|^2+2\eta g(x^t)^T(x^*-x^t)$

$\le\eta^2G^2+\|x^t-x^*\|^2+2\eta(f(x^*)-f(x^t))$

$2\eta(f(x^t)-f(x^*))\le\eta^2G^2+\|x^t-x^*\|^2-\|x^{t+1}-x^*\|^2$

proximal: $\min_{x\in\mathbb{R}^d}g(x)+h(x)$, $y^{t+1}=x^t-\eta\nabla g(x^t)$,

$x^{t+1}=\arg\min_{x\in\mathbb{R}^d}[h(x)+\frac{1}{2\eta}\|x-y^{t+1}\|_2^2]=\text{prox}_{\eta,h}(y^{t+1})$

$\min_{x\in C}f(x)\Leftrightarrow\min_{x\in\mathbb{R}^d}f(x)+\mathbb{I}_C(x)$

$\boxed{\|\text{Prox}_h(x)-\text{prox}_h(y)\|\le\|x-y\|}$

$\|\text{prox}_h(x)-\text{prox}_h(y)\|_2^2\le\langle x-y, \text{prox}_h(x)-\text{prox}_h(y)\rangle$

$\langle x-\text{prox}_h(x)-y+\text{prox}_h(y), \text{prox}_h(x)-\text{prox}_h(y)\rangle\ge0$

$0\in-x+\text{prox}_h(x)+\partial h(\text{prox}_h(x))\Rightarrow x-\text{prox}_h(x)\in\partial h(\text{prox}_h(x))$

by monocity of h

$\boxed{G_\eta(x)}=\frac{1}{\eta}[x-\text{prox}_{\eta h}(x-\eta\nabla g(x))]\Rightarrow x^{t+1}=x^t-\eta G_{\eta}(x^t)$

$G_\eta(x^*)=0\Leftrightarrow0\in\nabla g(x^*)+\partial h(x^*)$

proof: $G_\eta(\hat{x})\in\nabla g(\hat{x})+\partial h(\hat{x}-\eta G_\eta(\hat{x}))$

$\hat{x}-\eta\nabla g(\hat{x})-\hat{x}+\eta G_\eta(\hat{x})\in\eta\partial h(\hat{x}-\eta G_\eta(\hat{x}))$

if $G_\eta(x^*)=0$. $0\in\nabla g(x^*)+\partial h(x^*)$

if $0\in\nabla g(x^*)+\partial h(x^*)$, $x^*-\eta\nabla g(x^*)-x^*\in\eta\partial h(x^*)$

$x^*=\text{prox}_{\eta h}(x^*-\eta\nabla g(x^*))$, $G_\eta(x^*)=0$

Descent Lemma.

$f(x-\eta G_\eta(x))\le f(z)+G_\eta(x)^T(x-z)-\frac{\eta}{2}\|G_\eta(x)\|_2^2$

$\boxed{f(x-\eta G_\eta(x))\le f(x)-\frac{\eta}{2}\|G_\eta(x)\|_2^2}$

proxGD for β-smooth g and convex h.

$\boxed{f(x^k)-f(x^*)\le\frac{\beta\|x^0-x^*\|_2^2}{2k}}$

proof: $\|x^{t+1}-x^*\|^2=\|x^t-x^*\|^2+2\eta(\frac{\eta}{2}\|G_\eta(x^t)\|_2^2-$

$=\cdots+2\eta(f(x^*)-f(x^{t+1}))$ $(x^t-x^*)^TG_\eta(x^t))$

$g=0$: only h: $G_\eta(x)=\frac{1}{\eta}(x-\text{prox}_{\eta h}(x))$

$h(x^{t+1})\le h(x^t)-\eta\|G_\eta(x^t)\|_2^2$

$h(x^{t+1})\le h(x^*)-\eta\|G_\eta(x^t)\|_2^2+G_\eta(x^t)^T(x^t-x^*)$

proof: for $u\in\partial h(x^{t+1})$

$h(x^{t+1})\le h(z)-\eta u^T(z-x^{t+1})$, $u=G_\eta(x^t)$

$z=x^t$, $x^{t+1}=\text{prox}(x^t)$, → first claim

$z=x^*$, → second

$\boxed{h(x^k)-h(x^*)\le\frac{\|x^0-x^*\|^2}{2\eta k}}$

proof: $\|x^{t+1}-x^*\|^2=\|x^t-x^*\|^2+\eta^2\|G_\eta(x^t)\|_2^2-2\eta G_\eta(x^t)^T(x^t-x^*)$

$\le\|x^t-x^*\|^2+2\eta(h(x^{t+1})-h(x^*))$

SGD: $x^{t+1}=x^t-\eta_t g(x^t,\xi^t)$, $\mathbb{E}_\xi g(x,\xi)=\nabla f(x)$

1. noisy gradient: $g(x,\xi)=\nabla f(x)+\xi$, $\mathbb{E}\xi=0$

2. Incremental: $f(x)=\frac{1}{n}\sum f_i(x)$

$x^{t+1}=x^t-\eta_t\nabla f_{i^t}(x^t)$

3. ERM: $R(f)=\mathbb{E}_{x,y}[L(f(x),y)]=\frac{1}{n}\sum L(f(x_i),y_i)$

4. minibatch.

5. Randomized: $x^{t+1}=x^t-\eta_t\nabla_{i^t}f(x^t)$ diff ✓ diff+con

One-pass SGD: $\min\frac12\mathbb{E}_{x\sim p}\|X-x\|^2$

$g(x^t,X_i)=X_i-x^t$, $\eta_t=\frac{1}{t+1}$

$\hat{x}=\frac{1}{n}\sum X_i$,

$f(\hat{x})-f(x^*)=\frac12\mathbb{E}_{x\sim p}\|X-\hat{x}\|^2-\frac12\mathbb{E}_{x\sim p}\|X-\mu\|^2=\frac{\sigma^2}{2n}$

$\frac{1}{k}$ type convergence.

SGD for non-smooth. $\mathbb{E}_\xi[g(x,\xi)]\in\partial f(x)$

$\mathbb{E}\|g(x,\xi)\|_2^2\le G^2$. $\eta=\frac{\sqrt{R}}{\sqrt{G}k}$

$\mathbb{E}f(\frac{1}{k}\sum^k x^t)-f(x^*)\le\frac{RG}{\sqrt{k}}$

proof: $\mathbb{E}[\|x^{t+1}-x^*\|_2^2|x^t]=\|x^t-x^*\|_2^2+\eta^2\mathbb{E}[\|g(x^t,\xi)\|_2^2|x^t]$

$-2\eta(x^t-x^*)^T\mathbb{E}[g(x^t,\xi)|x^t]\le\|x^t-x^*\|_2^2+\eta^2G^2-2\eta(x^t-x^*)^Tg_{x^t}$

$\le\|x^t-x^*\|_2^2+\eta^2G^2+2\eta(f(x^*)-f(x^t))$ by cvx

tower property.

$\mathbb{E}[\|x^{t+1}-x^*\|_2^2]\le\mathbb{E}_{x^t}\|x^t-x^*\|_2^2+\eta^2G^2+2\eta(f(x^*)-\mathbb{E}f(x^t))$

$\frac{1}{k}\sum^k\mathbb{E}f(x^t)-f(x^*)\le\boxed{\frac{G^2\eta}{2}+\frac{\|x^0-x^*\|_2^2}{2k\eta}}$

$\boxed{\mathbb{E}f(\frac{1}{k}\sum x^t)-f(x^*)}\le\frac{1}{k}\sum^k\mathbb{E}f(x^t)$

SGD for α-strong: not linear, bc variance

for $\eta<1/\alpha$, $\mathbb{E}\|x^k-x^*\|_2^2\le(1-\alpha\eta)^k\|x^0-x^*\|_2^2+\frac{\eta G^2}{\alpha}$

$\eta_t=\frac{1}{\alpha(t+1)}$, $\mathbb{E}f(\frac{1}{k}\sum^k x^t)-f(x^*)\le\frac{G^2(1+\log k)}{2\alpha k}$

Mirror descent. cvx f, L_∞ bounded $\{x\ge0, \sum^d x_i=1\}$

subgradient: $\frac{\sqrt{d}}{\sqrt{k}}$, $\frac{\sqrt{\log d}}{}$

mirrordescent: $\frac{}{\sqrt{k}}$

mirror map: Φ diff, convex, open domain D. $\bar{D}\supseteq C$

Φ: α-strong cvx on C. with r.t $\|\cdot\|$

$\Phi(y)\ge\Phi(x)+\nabla\Phi(x)^T(y-x)+\frac{\alpha}{2}\|x-y\|^2$

Bregman Divergences:

$D_\Phi(x,y)=\Phi(x)-\Phi(y)-\nabla\Phi(y)^T(x-y)$

Bregman projection $\Pi_C(y)=\arg\min_{x\in C}D_\Phi(x,y)$

- $\Phi(x)=\frac12\|x\|_2^2$ Eucl norm.

$D_\Phi(x,y)=\frac12\|x\|_2^2-\frac12\|y\|^2-y^T(x-y)=\frac12\|x-y\|_2^2$

- $\Phi(x)=\sum x_i\log x_i$, strict cvx, L_1-norm.

$D_\Phi(x,y)=\sum x_i\log(x_i/y_i)-\sum_{i=1}^d(x_i-y_i)=KL(x,y)$ $\log d$. $\ge\frac12\|x-y\|_1^2$

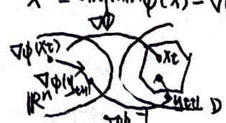
$x^{t+1}=\arg\min_{x\in C}f(x^t)+\nabla f(x^t)^T(x-x^t)+\frac{1}{\eta}D_\Phi(x,x^t)$

$x^{t+1}=\arg\min_{x\in C}\Phi(x)-(\nabla\Phi(x^t)-\eta\nabla f(x^t))^Tx$

$\nabla\Phi(y_{t+1})=\nabla\Phi(x^t)-\eta\nabla f(x^t)$

$x^{t+1}=\arg\min\Phi(x)-\nabla\Phi(y^{t+1})^Tx=\arg\min_{x\in C}D_\Phi(x,y^{t+1})$

$=\Pi_C(y^{t+1})$



f is G-Lip. $D_\Phi(x^0,x^*)\le R^2$

$f(\frac{1}{k}\sum^{k-1}x^i)-f(x^*)\le RG\sqrt{\frac{2}{\alpha k}}$ $R^2\le\log d$ G by L_∞ norm

$y^{t+1}=x^t\exp(-\eta\nabla f(x^t))$

$x^{t+1}=y^{t+1}/\|y^{t+1}\|_1$ minimize f over simplex

LP $\min_x c^Tx$

subject to $Ax=b$, $Gx\le h$

where $c\in\mathbb{R}^d, A\in\mathbb{R}^{m\times n}, b\in\mathbb{R}^m$, $G\in\mathbb{R}^{r\times d}, h\in\mathbb{R}^r$

dual: $\max_{u,v} -u^Tb-v^Th$

$-A^Tu-G^Tv=c, v\ge0$

$u^T(Ax-b)+v^T(Gx-h)\le0$.

$(-A^Tu-G^Tv)^Tx\ge -u^Tb-v^Th$

$p^*=\infty$, infeasible ⇐ dual is unbounded $d^*=\infty$

$p^*=-\infty$, unbounded → dual is infea $d^*=-\infty$

$c^Tx\ge c^Tx+u^T(Ax-b)+v^T(Gx-h)$

$g(u,v)=\min_x L(x,u,v)\le p^*$

$=\begin{cases}-b^Tu-h^Tv, & \text{if } c=-A^Tu-G^Tv\\ -\infty, & \text{otherwise}\end{cases}$

$\max_{u,v} g(u,v)$, $v\ge0$

OT: $\min_M \sum_{ij}C_{ij}M_{ij}$, $M\in\mathbb{R}^{m\times n}$

$\sum_{j=1}^m M_{ij}=p_i$, 1-n, $\sum_{i=1}^n M_{ij}=q_j$, 1-m

$M_{ij}\ge0$

$L(M,u,v,w)=\sum_{ij}C_{ij}M_{ij}+\sum_{i=1}^n u_i(p_i-\sum_j^m M_{ij})+\sum_{j=1}^m v_j(q_j-\sum_i^n M_{ij})-\sum_{ij}w_{ij}M_{ij}$

$w_{ij}\ge0$

$g(u,v,w)=\begin{cases}\sum_{i=1}^n u_ip_i+\sum_{j=1}^m v_jq_j, & \text{if } C_{ij}-u_i-v_j-w_{ij}=0\\ -\infty, & \end{cases}$

$\max_{u,v,w}\sum_{i=1}^n u_ip_i+\sum_j^m v_jq_j$ | $\max_{u,v}$ ....

$C_{ij}-u_i-v_j-w_{ij}=0$, $w_{ij}\ge0$ | $C_{ij}-u_i-v_j\ge0$

$\begin{cases}\min_x f(x)\\ h_i(x)\le0, i\in\{1,m\}\\ l_j(x)=0, j\in\{1,\dots r\}\end{cases}$ f, h – convex, l – affine.

$f(x)\ge L(x,u,v)=f(x)+\sum^r u_jl_j(x)+\sum^m v_ih_i(x)$

$p^*\ge\min_x L(x,u,v)=g(u,v)$.

dual: $\max_{u,v} g(u,v)$, $v\ge0$

QP: $\min_x \frac12x^TQx+c^Tx$

$Ax=b$

$Gx\le h$

$L(x,u,v)=\frac12x^TQx+c^Tx+u^T(Ax-b)+v^T(Gx-h)$

deriv: $Qx+c+A^Tu+G^Tv=0$

$x^*=-Q^{-1}(c+A^Tu+G^Tv)$

dual: $\max_{u,v} -\frac12(c+A^Tu+G^Tv)^TQ^{-1}(c+A^Tu+G^Tv)-u^Tb-v^Th$, $v\ge0$

$g(u,v)=\begin{cases}--- & \text{if } (c+A^Tu+G^Tv)\perp \text{null } Q\\ -\infty, & \text{otherwise}\end{cases}$

Relaxed Slater's Condition

strong, most convex, rarely non-cvx

Exist x_0:

$l_j(x_0)=0, j\in\{1,r\}$

$h_i(x_0)\le0, i\in\{1,\dots k\}$

$h_i(x_0)<0, i\in\{k+1,m\}$

strictly feasible for non-affine inequality

LP-QP. feasible ⇔ strong duality ⇐ all affine

Derive primal:

$p^*=\inf_x\sup_{u,v\ge0}L(x,u,v)=\begin{cases}f(x), & \text{if } x \text{ fea}\\ \infty, & \text{otherwise}\end{cases}$

$d^*=\sup_{u,v\ge0}\inf_x L(x,u,v)$

$L(x,u^*,v^*)\ge L(x^*,u^*,v^*)\ge L(x^*,u,v)$

primal dual: $f(x)-p^*\le f(x)-g(u,v)$.

KKT $(\hat x,\hat u,\hat v)$

$h_i(\hat x)\le0, i\in\{1,\dots,m\}$

$l_j(\hat x)=0, i\in\{1,\dots r\}$

$\hat v\ge0$

$\hat v_ih_i(\hat x)=0, i\in\{1,\dots,m\}$ complementary slackness

$\nabla f(\hat x)+\sum^m\hat v_i\nabla h_i(\hat x)+\sum^r\hat u_i\nabla l_j(\hat x)=0$ Stationarity

Sufficiency:

KKT $(\hat x,\hat u,\hat v)$: $f(\hat x)=g(\hat u,\hat v)\le f(x)$

Strong duality: $g(\hat u,\hat v)=f(\hat x)\ge g(u,v)$

Necessity:

strong duality holds, given opt sol (x^*,u^*,v^*)

$f(x^*)=g(u^*,v^*)=\inf_x[f(x)+\sum\ ]$

$\le f(x^*)+\dots$

is KKT point.

Fenchel Conjugate:

$f^*(y)=\sup_{x\in dom(f)}(y^Tx-f(x))$

$f(x)\ge y^Tx-b$, $b\ge y^Tx-f(x)$

$b=\sup_{x\in dom f}[y^Tx-f(x)]=f^*(y)$

• $f(x)=\frac12x^TQx$, $Q\succ0$

$f^*(y)=\sup_x[x^Ty-\frac12x^TQx]=\frac12y^TQ^{-1}y$

$\nabla f(x)=Qx$ the grad of f^* is the inverse of the grad of f

$\nabla f^*(y)=Q^{-1}y$

$\nabla f^*(\nabla f(x))=x$

$\nabla f(\nabla f^*(y))=y$

• $f(x)=\|x\|$

$f^*(y)=\sup_x[x^Ty-\|x\|]=\begin{cases}0, & \text{if } \|y\|_*\le1, \text{ s-c}\\ \infty, & \text{otherwise}\end{cases}$

convex indicator

• $f(x)=I_C(x)$

$f^*(y)=\sup_{x\in C}x^Ty$ → support function

$x^Ty-f^*(y)\le0$ C on one side of $x^Ty-f^*(y)=0$

if C closed

Fenchel Inequality

$f^*(y)\ge x^Ty-f(x)$.

$f^{**}\le f$, $f(x)\ge x^Ty-f^*(y)$

$f(x)\ge\sup_y[x^Ty-f^*(y)]=f^{**}(x)$

if f is closed and convex. $f^{**}=f$

f α-strongly → f^* $\frac1\alpha$ smooth ✓

f β-smooth → f^* $\frac1\beta$ strongly

$\min_x f(x)+g(x)\to\min_x f(x)+g(z)$, $z=x$

$L(x,z,u)=f(x)+g(z)+u^T(x-z)$

$\sup_u\inf_x[f(x)+u^Tx]+\inf_z g(z)-u^Tz$

$=\sup_u -f^*(-u)-g^*(u)$

$x_0\in dom(f)\cap cont(g)$ Fenchel duality theorem.

LASSO: $\min_x f(Ax)+g(x)\to\min_{x,z}f(z)+g(x)$, $Ax=z$

$\max_u -f^*(u)-g^*(-A^Tu)$

$\min_x\frac12\|y-Ax\|_2^2+\lambda\|x\|_1:\min_x f(Ax)+g(x)$

$\max_u[-y^Tu-\frac12\|u\|_2^2-\mathbb{I}(\|A^Tu\|_\infty\le\lambda)]$

⇔ $\min_{\hat u}\frac12\|y-\hat u\|_2^2$, $\hat u=y-Ax$

$\|A^T\hat u\|_\infty\le\lambda$

PL condition $\frac12\|\nabla f(x)\|_2^2\ge\mu(f(x)-f(x^*))$

β-smooth, unconstrained. need not cvx, assume x^* exist, not nec unique

☆ α-strongly convex satisfies α-PL

$f(y)\ge f(x)+\nabla f(x)^T(y-x)+\frac\alpha2\|x-y\|_2^2$

$f(x^*)\ge f(x)-\frac1{2\alpha}\|\nabla f(x)\|_2^2$

☆ β-smooth, μ-PL. GD with $\eta=1/\beta$.

$f(x^k)-f(x^*)\le(1-\frac\mu\beta)^k(f(x^0)-f(x^*))$

$f(x^{t+1})\le f(x^t)-\frac1{2\beta}\|\nabla f(x^t)\|_2^2\le f(x^t)-\frac\mu\beta(f(x^t)-f(x^*))$

☆ quadratic growth $\frac\mu2\|x-x^*\|_2^2\le f(x)-f(x^*)$

Inexact GD: $\|\overset{\text{oracle}}{g_x}-\nabla f(x)\|_2\le\mu_1\|x-x^*\|_2+\mu_2$, $\mu_1,\mu_2\ge0,\mu_1<\alpha$

☆ α-strongly, β-smooth. $\eta=2/(\alpha+\beta)$

$\boxed{\|x^k-x^*\|_2\le(1-k_\alpha)^k\|x^0-x^*\|+\frac{\mu_2}{1-k_\alpha}}$, $0<k_\alpha=\frac{2(\alpha-\mu_1)}{(\alpha+\beta)}<1$

Matrix completion: $\min_M\sum(Y_{ij}-M_{ij})^2+\lambda\|M\|_{tr}$

M → soft thresholding the singular of Y.

$prox_{\lambda\eta}(Y)=U\Sigma_\lambda V^T$, $Y=U\Sigma V^T$, $\Sigma_\lambda=\max(0,\Sigma-\lambda)$

$\begin{cases}Z^{t+1}=x^t-\eta(x^t-Y)\circ\Omega. \ \Omega_{ij}=1 \text{ if } (i,j)\in\Omega. \text{ 0 otherw}\\ x^{t+1}=prox_{\lambda\eta}(Z^{t+1})\end{cases}$

SVM $\min_{\beta,\beta_0,\xi}\frac12\|\beta\|_2^2+C\sum_{i=1}^n\xi_i$ QP. strong duality ✓

$\xi_i\ge0$, for $i\in\{1,\dots,n\}$.

$y_i(x_i^T\beta+\beta_0)\ge1-\xi_i$