EECS 127 Crib Sheet - DIVIT RAWAL

Standard form:

$$\min_{\vec x\in\mathbb R^n} f_0(\vec x) \text{ s.t. } f_i(\vec x)\le 0\ \forall i\in[m],\quad h_j(\vec x)=0\ \forall j\in[p]$$

Norms: $f: V\to\mathbb R$
- PD; $f(\vec x)\ge 0\ \forall \vec x\in V$ & $f(\vec x)=0$ iff $\vec x=\vec 0$
- $f(\alpha\vec x)=|\alpha|f(\vec x)\ \forall \vec x\in V$ & $\alpha\in\mathbb R$
- $f(\vec x+\vec y)\le f(\vec x)+f(\vec y)\ \forall \vec x,\vec y\in V$

$\|\vec x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}\ \forall p\le 1$

$\|\vec x\|_\infty = \max_i |x_i|$

Hölder's Ineq. Let $1\le p,q\le\infty$ s.t. $\frac1p+\frac1q=1$
$\forall \vec x,\vec y\in\mathbb R^n$ $|\vec x^T\vec y|\le \sum_{i=1}^n |x_iy_i| \le \|\vec x\|_p\|\vec y\|_q$

G-S Alg: in: LI set $\{a_i\}_{i=1}^k$
$\vec q_1 = \vec a_1/\|a_1\|_2$
for $i\in\{2,\dots,k\}$
$\vec p_i = \sum_{j=1}^{i-1}\vec q_j(\vec q_j^T\vec a_i)$
$\vec s_i = \vec a_i - \vec p_i$
$\vec q_i = \vec s_i/\|\vec s_i\|_2$
return $\{q_i\}_{i=1}^k$

$\vec a_i = r_{1i}\vec q_1 + r_{2i}\vec q_2+\dots+r_{ii}\vec q_i$

$[\vec a_1\cdots\vec a_k]=[\vec q_1\cdots\vec q_k]\begin{bmatrix} r_{11}&r_{12}&\cdots&r_{1k}\\ \vdots&\ddots&&\vdots\\ r_{k1}&\cdots&&r_{kk}\end{bmatrix}$

$A\in\mathbb R^{n\times k}$ for $k\le n$ & A is full rank
$\Rightarrow \exists Q\in\mathbb R^{n\times k}$, $R\in\mathbb R^{k\times k}$ upper tri & orthonormal s.t. $A=QR$

FTLA: $N(A)\oplus R(A^T)=\mathbb R^n = N(A^T)\oplus R(A)$

Sym. has $\lambda\in\mathbb R$ & guaranteed diag.

Spectral Thm. $A\in \mathrm{Sym}(n)$; λ_i w/ alg. mult μ_i
$\Phi_i = N(\lambda_i I - A)$ w/ geom. mult. $\phi_i = \dim(\Phi_i)$
- $\lambda_i\in\mathbb R\ \forall i$; $\vec p_i\in\Phi_i$, $\vec p_j\in\Phi_j \Rightarrow p_i^Tp_j = \vec 0$;

$\mu_i=\phi_i\ \forall i$; $A=U\Lambda U^T$
$U=[\vec u_1\cdots\vec u_n]\in\mathbb R^{n\times n}$ eig. are (λ_i,μ_i)

PD: $x^TAx>0\ \forall\vec x\ne 0$ iff $\lambda_i>0\ \forall i$
PSD: $x^TAx\ge0\ \forall \vec x$ iff $\lambda_i\ge0\ \forall i$

SVD: A has rank r. $A=U\Sigma V^T = U_r\Sigma_r V_r^T = \sum_{i=1}^r \sigma_i\vec u_i\vec v_i^T$

$\|A\|_F^2=\sum_{i=1}^m\sum_{j=1}^n A_{ij}^2 = \mathrm{tr}(A^TA)=\sum_{i=1}^r\sigma_i^2$
& invariant to orthonormal transf.

$\|A\|_2 = \max_{\vec x\in\mathbb R^n, \|\vec x\|_2=1}\|A\vec x\|_2=\sigma_1$

Rayleigh Quotient, $A\in\mathrm{Sym}(n)$
$\lambda_{max}(A)=\max_{\vec x\in\mathbb R^n,\vec x\ne\vec 0}\frac{\vec x^TA\vec x}{\vec x^T\vec x} = \max_{\vec x\in\mathbb R^n,\|\vec x\|_2=1}\vec x^TA\vec x$ & analog for min

Eckhart-Young
$\|A-A_k\|\le\|A-B\|_2$ $\forall B\in\mathbb R^{m\times n}\mid \mathrm{rank}(B)\le k$
& same for $\|\cdot\|_F$

Block Matrices
$\begin{bmatrix}A&B\\C&D\end{bmatrix}^T=\begin{bmatrix}A^T&C^T\\B^T&D^T\end{bmatrix}$

Vector Calc $h(x)=f(g(x))$
$\frac{dh}{dx}(x)=\sum_{i=1}^n\frac{\partial f}{\partial g_i}g(x)\cdot\frac{dg_i}{dx}(x)$

Jacobian $f:\mathbb R^n\to\mathbb R^m$
$D\vec f(\vec x)=\begin{bmatrix}\nabla f_1(\vec x)^T\\ \vdots\\ \nabla f_m(\vec x)^T\end{bmatrix}$
$\to D\vec h(\vec x)=[D\vec f(\vec g(\vec x))][D\vec g(\vec x)]$
$\nabla h(\vec x)=[D\vec g(\vec x)]^T\nabla f(\vec g(\vec x))$

$\nabla^2 f(\vec x)=D(\nabla f)(\vec x)=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2}(\vec x)&\cdots&\frac{\partial^2 f(\vec x)}{\partial x_n\partial x_1}\\ \vdots&&\vdots\\ \frac{\partial^2 f}{\partial x_1\partial x_n}(\vec x)&\cdots&\frac{\partial^2 f(\vec x)}{\partial x_n^2}\end{bmatrix}$

Taylor's Thm. (order 2)
$\hat f_2(\vec x;\vec x_0)=f(\vec x_0)+[\nabla f(\vec x_0)]^T(\vec x-\vec x_0)+\frac12(\vec x-\vec x_0)^T[\nabla^2f(\vec x_0)](\vec x-\vec x_0)$
$\vec f(\vec x+\vec\delta)=\vec f(\vec x)+[D\vec f(\vec x)]\vec\delta+o(\|\vec\delta\|_2)$
@ solns. to opt probl. $\nabla f(\vec x^*)=0$

LinReg. & Ridge Reg.
$\kappa(A)=\frac{\sigma_1(A)}{\sigma_n(A)}$
$\min_{\vec x\in\mathbb R^n}\|A\vec x-\vec y\|_2^2+\lambda\|\vec x\|_2^2$
$\to \vec x^* = (A^TA+\lambda I)^{-1}A^Ty$

$\boxed{\min\max \ge \max\min}$

Convexity
C is convex if $\forall\theta\in[0,1]$ & $x_1,x_2\in C$ $\theta x_1+(1-\theta)x_2\in C$
convex hull: set of all convex combs of points in S
f convex if
$f(\theta\vec x_1+(1-\theta)\vec x_2)\le\theta f(\vec x_1)+(1-\theta)f(\vec x_2)$

Jensen's Ineq. for $\theta_i\in[0,1]$ $\sum\theta_i=1$
$f\left(\sum_{i=1}^k\theta_i x_i\right)\le\sum_{i=1}^k\theta_i f(x_i)$ & f convex

f is convex iff $\nabla^2 f(\vec x)\succeq 0\ \forall x\in\Omega$
$\vec x^TQ\vec x+\vec b^T\vec x+c$ convex iff $Q\succeq0$.
Strict convexity: $<$ not $\le$, $\succ$ not $\succeq$.
affine is concave & convex
Optimization problem in standard form is convex if $f_0\dots f_m$ convex & $h_1\dots h_p$ are affine.
for convex, local extrema are global & unique

$f_k(\vec x)\le 0$ active if $f_k(x_0)=0$, inactive otherwise

μ-strongly convex iff $\nabla^2 f(\vec x)-\mu I\succeq0$
↳ exactly one global minimizer

L-smooth ⇒ $f(\vec y)\le f(\vec x)+[\nabla f(\vec x)]^T(\vec y-\vec x)+\frac L2\|\vec y-\vec x\|_2^2$
↳ $\|\nabla f(\vec x)\|_2^2\le 2L\left(f(\vec x)-\min_{\vec x'\in\mathbb R^n}f(\vec x')\right)$

Thm. Let f be μ-strongly convex, L-smooth.
$p^*=\min_{\vec x\in\mathbb R^n}f(\vec x)$. $\eta=\frac1L$ s.t. $\vec x_{t+1}=\vec x_t-\eta\nabla f(\vec x_t)$
$\to\forall x_0$ $\|\vec x_{t+1}-\vec x^*\|_2^2\le\left(1-\frac{\mu}{L}\right)\|\vec x_t-\vec x^*\|_2^2$.

Duality (Ω is feasible set)
$\mathcal L(\vec x,\vec\lambda,\vec\nu)=f_0(\vec x)+\sum_{i=1}^m\lambda_if_i(\vec x)+\sum_{j=1}^p\nu_jh_j(\vec x)$
$(\vec\lambda,\vec\nu)\mapsto\mathcal L(\vec x,\vec\lambda,\vec\nu)$ is affine & concave
Dual: $d^*=\max_{\lambda\in\mathbb R^m,\nu\in\mathbb R^p} g(\vec\lambda,\vec\nu)$ s.t. $\lambda_i\ge0\ \forall i$ for $g(\vec\lambda,\vec\nu)=\min_{\vec x\in\mathbb R^n}\mathcal L(\vec x,\vec\lambda,\vec\nu)$

D is always convex.
$f_0(\vec x)\ge p^*$ & $g(\vec\lambda,\vec\nu)\le d^*$
$f_0(\vec x)\ge\mathcal L(\vec x,\vec\lambda,\vec\nu)\ge g(\vec\lambda,\vec\nu)$
$f_0(\vec x)\ge d^*$ & $g(\vec\lambda,\vec\nu)\le p^*$
p^*-d^* is duality gap
gap $=0$ ⇒ strong duality, else weak

Slater's Condition (str. duality) & always holds! via minimax
$\exists\vec x\in\mathrm{relint}(\Omega)$ s.t.
$\forall i$ s.t. f_i is affine $f_i(\vec x)\le0$
$\forall i$ s.t. f_i is not affine $f_i(\vec x)<0$
$\forall j$ $h_j(\vec x)=0$

KKT Conditions (ass. $f_0,\dots f_m,h_1,\dots,h_p$ differentiable)
$(\tilde x,\tilde\lambda,\tilde\nu)$ fulfills KKT conditions if
$\tilde x$ feasible for P: $f_i(\tilde x)\le0$, $h_j(\tilde x)=0$
$(\tilde\lambda,\tilde\nu)$ feasible for D: $\tilde\lambda\ge0$
$\tilde\lambda_if_i(\tilde x)=0$
$\vec 0=\nabla_{\vec x}\mathcal L(\tilde x,\tilde\lambda,\tilde\nu) = \nabla f(\tilde x)+\sum_{i=1}^m\tilde\lambda_i\nabla f_i(\tilde x)+\sum_{j=1}^p\tilde\nu_j\nabla h_j(\tilde x)$

Thm. If strong duality holds, KKT conditions are necessary for optimality
Thm. If convexity holds, KKT cond. are sufficient for optimality

Fenchel Conj. $f^*(y)=\sup_x\{y^Tx-f(x)\}$
f^* always convex. $f^{**}=f$ if f closed, proper, convex
$y\in\partial f(x)\Leftrightarrow x\in\partial f^*(y) \Leftrightarrow f(x)+f^*(y)=y^Tx$

P: $\min_x f(x)+g(Ax)$
$=\max_y -f^*(-A^Ty)-g^*(y)$
compute via $\nabla f(x)=y$
$f(x)+f^*(x)\ge y^Tx$

$f^{**}\le f$

Types of Problems

Linear Program

$p^* = \min_{x \in \mathbb{R}^n} \vec{c}^T x$ s.t. $A\vec{x} = \vec{y}$, $\vec{x} \geq \vec{0}$

Always convex.

$d^* = \max_{\lambda \in \mathbb{R}^n, \nu \in \mathbb{R}^p} -\vec{y}^T\vec{\nu}$ s.t. $\vec{c} - \vec{\lambda} + A^T\vec{\nu} = \vec{0}$, $\vec{\lambda} \geq \vec{0}$

Quadratic Program

$p^* = \min_{x \in \mathbb{R}^n} \frac{1}{2}\vec{x}^T H \vec{x} + \vec{c}^T\vec{x}$ s.t. $A\vec{x} \leq \vec{y}$, $C\vec{x} = \vec{z}$. for $H \in \mathrm{Sym}(n)$

not always convex.

convex $\iff H \succeq 0$

Second-Order Cone Program

$p^* = \min_{\vec{x} \in \mathbb{R}^n} \vec{c}^T\vec{x}$ s.t. $\|A_i\vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^T\vec{x} + z_i$

are convex.

dual of SOCP is SOCP

$LP \subseteq QP \subseteq SOCP \subseteq SDP$

Regularization & Sparsity

Regularized Pz:

$p^*_\lambda = \min_{\vec{x} \in \Omega} \{f_0(\vec{x}) + \lambda R(\vec{x})\}$

LASSO Reg.

$\min_{\vec{x} \in \mathbb{R}^n} \underbrace{\|A\vec{x} - \vec{y}\|_2^2}_{f_0} + \lambda\|\vec{x}\|_1$

f_0 is convex. If A has full rank, f_0 is μ-str-convex w/ $\mu = 2\sigma_n(A)^2$, a soln. always exists, if A has full rank, soln is unique.

Optimality Cond.:

$0 \in A^T(Ax^* - b) + \lambda\partial\|x^*\|_1$

$-A^T(Ax^* - b) \in \lambda\partial\|x^*\|_1$

componentwise a_i is col. i of A

$a_i^T(b - Ax^*) = \lambda\,\mathrm{sgn}(x_i^*)$ if $x_i^* \neq 0$

$|a_i^T(b - Ax^*)| \leq \lambda$ if $x_i^* = 0$

Common Fenchel Conj

$f(x) = \frac{1}{2}x^TQx \quad Q \succeq 0$

$f^*(y) = \frac{1}{2}y^TQ^{-1}y$

$f(x) = \mathbb{1}_{\{x \in C\}} = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$

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

Norms. $f(x) = \|x\|$

$\|x\|_1^* = \mathbb{1}_{\{\|y\|_\infty \leq 1\}}$

$\|x\|_2^* = \mathbb{1}_{\{\|y\|_2 \leq 1\}}$

$\|x\|_\infty^* = \mathbb{1}_{\{\|y\|_1 \leq 1\}}$

LP → simplex alg → vertex is opt.

Linear $f(x) = a^Tx$

$f^*(y) = \mathbb{1}_{\{y = a\}}$

If $g(x) = f(x) + a^Tx + b$

$g^*(y) = f^*(y - a) - b$

$f(x) = g(Ax - b)$

$f^*(y) = \inf g^*(z) + z^Tb$ s.t. $A^Tz = y$

$\min_x f(Ax) + g(x)$

$f^*(y) = \sup_z y^Tz - f(z)$

$g^*(s) = \sup_x s^Tx - g(x)$

Fenchel Dual:

$\max_y -f^*(y) - g^*(-A^Ty)$

Vector Derivatives & Stuff

$\nabla_x a^Tx = a = \nabla_x x^Ta$

$\nabla_x c = 0$, $\nabla_x x^Tx = \nabla_x\|x\|_2^2 = 2x$

$\nabla_x(x^TQx) = (Q + Q^T)x \quad \nabla^2(\frac{1}{2}x^TQx) = Q$ if $Q = Q^T$

$\nabla_x\|Ax - b\|_2^2 = 2A^T(Ax - b)$

$\nabla_x^2\|Ax - b\|_2^2 = 2A^TA \quad \nabla_x\|Ax\|_2^2 = 2A^TAx$

$\nabla_x^2\left(\frac{1}{2}(Ax - b)^TQ(Ax - b)\right) = A^TQ(Ax - b)$

$\nabla_y(A^Tf(Ay)) = A^T\mathrm{diag}(f'(Ay))A$ if f is elementwise

For $\min_x \frac{1}{2}\|x - z\|_2^2 + \lambda\|x\|_1$

$x_i^* = S_\lambda(z_i)$ for $S_\lambda(z_i) = \mathrm{sgn}(z_i)\max\{|z_i| - \lambda, 0\}$

if $A^TA = I$, $x^* = S_\lambda(A^Tb)$

Grad. Desc.

convex $f \Rightarrow f(y) \geq f(x) + \nabla f(x)^T(y - x)$

$\Rightarrow (\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$

convex + L-smooth: $\alpha \in [0, \frac{1}{L}]$.

if μ-str-convex & L-smooth: $\alpha \in [0, \frac{2}{L}]$

for QP, $\alpha \in (0, \frac{2}{\lambda_{max}(Q)}]$

Newton's Method

$\Delta x_{nt} = -(\nabla^2 f(x))^{-1}\nabla f(x)$

$x^{k+1} = x^k + (\Delta x_{nt})t$ for $t \in (0, 1]$ ← full newton step

$\lambda(x)^2 = \nabla f(x)^T(\nabla^2 f(x))^{-1}\nabla f(x)$

$\equiv \lambda(x)^2 = -\nabla f(x)^T\Delta x_{nt}$

Stopping criterion $\frac{\lambda(x)^2}{2} \leq \varepsilon$

Support Vector Machines (x_i, y_i) $y_i \in \{\pm 1\}$

$\hat{y} = \mathrm{sgn}(w^Tx + b)$. functional margin: $y_i(w^Tx_i + b)$

dec. boundary: $w^Tx + b = 0$ margin width: $\frac{2}{\|w\|_2}$

If data linearly separable $\min_{w,b} \frac{1}{2}\|w\|_2^2$ s.t. $y_i(w^Tx_i + b) \geq 1 \;\forall i$

support vectors satisfy $y_i(w^Tx_i + b) = 1$

Data not separable: $\min_{w,b,\xi} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \xi_i$ s.t. $y_i(w^Tx_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

↳ class A pt is a convex comb of class B pts.

Slack Vars & Formulation Stuff

$|u| \leq t \rightarrow \pm(u) \leq t$

$\min_x \max_i f_i(x) \iff \min_{x,t} t$ s.t. $f_i(x) \leq t$

$\|u\|_1 \leq t \iff \exists s \,|\, \pm u \leq s \quad 1^Ts \leq t$

$\|u\|_\infty \leq t \iff \pm u \leq t\mathbb{1}$ $\quad \|u\|_2 \leq t \Rightarrow$ SOCP

Leftover Slack $a^Tx \leq b \iff a^Tx + s = b$ for $s \geq 0$

Epigraph Reformulation: standard form

$\rightarrow \min_{t \in \mathbb{R}, \vec{x} \in \mathbb{R}^n} t$ s.t. $t \geq f_0(\vec{x})$

$f_i(\vec{x}) \leq 0 \;\forall i \in [m]$

$\min_x \sum_i \max(0, a_i^Tx + b_i)$

↓

$\min_{x,t} \sum_i t_i$ s.t. $t_i \geq 0$, $t_i \geq a^Tx + b_i$

Farkas' Lem. for A, b

(i) $Ax = b$, $x \geq 0$

(ii) $A^Ty \leq 0$, $y^Tb > 0$

exactly one holds

$\frac{d}{dx}\frac{f}{g} = \frac{gf' - fg'}{g^2}$

for $f = Ax$, $Df = A$