

## 0.1 Symbols

$S^n \equiv$  set of all symmetric matrices

$M_+^n \equiv$  set of all positive definite matrices (PD)

$S_+^n \equiv$  set of all symmetric positive semidefinite matrices (SPSD)

$S_+ +^n \equiv$  set of all symmetric positive definite matrices (SPD)

$cl(X) \equiv$  closure of  $X$

$relint(X) \equiv$  relative interior of  $X$

$int(X) \equiv$  interior of  $X$

## 0.2 Preliminary

Consider  $f: \mathbb{R}^n \rightarrow \mathbb{R}$

Gradient of  $f$ :  $\nabla f(x) = \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \end{bmatrix}$

$$f(x) = a^T x \implies \nabla f(x) = a$$

$$f(x) = x^T P x, P = P^T \implies \nabla f(x) = 2Px$$

$$f(x) = x^T P x \implies \nabla f(x) = 2\left(\frac{P^T + P}{2}\right)x = (P^T + P)x$$

Taylor expansion approximation:

$$f(x) \approx f(x_0) + \nabla^T f(x_0)(x - x_0) + o((x - x_0)^2)$$

$$f(x + \delta x) \approx f(x_0) + \nabla^T f(x_0)\delta x + o((\delta x)^2)$$

Chain rule:

$$f: \mathbb{R} \rightarrow \mathbb{R}, g: \mathbb{R} \rightarrow \mathbb{R}, h(x) = f(g(x))$$

$$\nabla h(x) = g'(f(x))\nabla f(x)$$

$$g: \mathbb{R}^m \rightarrow \mathbb{R}, g(x) = f(Ax + b)$$

$$\nabla g(x) = A^T \nabla f(Ax + b)$$

2nd derivative:

$$\nabla^2 f(x) = \begin{bmatrix} \partial^2 f / \partial x_1 \partial x_1 & \dots \\ \vdots & \partial^2 f / \partial x_n \partial x_n \end{bmatrix}$$

$$\nabla f(x) = Px + g$$

$$\nabla^2 f(x) = P$$

Hessian gives the 2nd order approximation:

$$f(x) \approx f(x_0) + \nabla^T f(x_0)(x - x_0) +$$

$$\frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

Matrices:

$A \in \mathbb{R}^{m \times n}$ : set of all real matrices

$$\text{inner product: } \sum_i \sum_j x_{ij} y_{ij} = \text{trace}(XY^T) =$$

$$\text{trace}(Y^T X) = \sum_i (XY)_{ii}$$

note trace has cyclic property

$$\text{frobenius norm: } \|X\|_F = (\sum_i \sum_j X_{ij}^2)^{\frac{1}{2}}$$

range:  $R(A) = \{Ax : x \in \mathbb{R}^n\} = \sum_i a_i x_i$ , where  $a_i$  is

ith column (column space of A)

null space:  $N(A) = \{x : Ax = 0\}$

SVD:

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

U and V are left and right eigenvector matrixes

U and V are orthogonal matrixes ( $BB^T = B^T B = I$ )

$\Sigma$  is rectangular diagonal matrix of eigenvalues

$$A_{m \times n} x_n$$

linear transformation:  $U \Sigma V^T x$

rotation - scaling - rotation

PSD matrix:

$$A \text{ PSD} \iff (\forall x) x^T A x \geq 0 \iff (\forall i) \lambda_i(A) \geq 0$$

$$A \text{ PSD} \implies A^{1/2} \text{ exists}$$

Real symmetric matrices have real eigenvalues:

$$Av = \lambda v$$

$$v^* A v = v^* \lambda v = \lambda \|v\|_2^2$$

$$(v^* A v)^* = v^* A^* v = \lambda^* \|v\|_2^2 \implies \lambda = \lambda^*$$

Affine sets

$$\text{A set } C \subseteq \mathbb{R}^n \text{ is affine if } (\forall x_1, x_2 \in C)(\forall \theta \in \mathbb{R}) \implies \theta x_1 + (1 - \theta)x_2 \in C$$

Convex sets

$$\text{A set } C \subseteq \mathbb{R}^n \text{ is convex if } (\forall x_1, x_2 \in C)(\forall \theta \in \mathbb{R}) 0 \leq \theta \leq 1 \implies \theta x_1 + (1 - \theta)x_2 \in C$$

Operations preserving convex sets:

- partial sum
- sum
- coordinate projection
- scaling
- translation
- intersection between any convex sets

Separating Hyperplanes: if  $S, T \subset \mathbb{R}^n$  are convex and disjoint, then  $\exists a \neq 0, b$  such that:

$$a^T x \geq b, \forall x \in S$$

$$a^T x \leq b, \forall x \in T$$

Supporting Hyperplane:

if  $S$  is convex,  $\forall x_0 \in \partial S$  (boundary of S), then  $\exists a \neq 0$  such that  $a^T x \leq a^T x_0, \forall x \in S$

Convex combination:

$$\sum_i \theta_i x_i, \forall \theta_i \in \mathbb{R}, \sum_i \theta_i = 1, \theta_i \geq 0$$

Convex hull:

The set of all convex combinations of points in  $C$ , the hull is convex

Hyperplane

$$C = \{x : a^T x = b\}, a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$$

### Halfspaces

$$C = \{x : a^T x \leq b\}, a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$$

$$\text{let } a^T x_c = b$$

$$C = \{x : a^T (x - x_c) \leq 0\}, a \in \mathbb{R}^n, a \neq 0$$

### Ellipse

$$E(x_c, P) = \{x : (x - x_c)^T P^{-1} (x - x_c) \leq 1\}, P > 0$$

$$P = r^2 I \implies \text{Euclidean Ball}$$

$$P = Q \begin{bmatrix} \lambda_1 & \dots \\ \dots & \lambda_n \end{bmatrix} Q^T$$

$$(x - x_c)^T (Q \begin{bmatrix} \lambda_1 & \dots \\ \dots & \lambda_n \end{bmatrix} Q^T)^{-1} (x - x_c) \leq 1$$

$$\tilde{x}^T \begin{bmatrix} \frac{1}{\lambda_1} & \dots \\ \dots & \frac{1}{\lambda_n} \end{bmatrix} \tilde{x} \leq 1$$

$$\tilde{x}^T \begin{bmatrix} \frac{1}{\lambda_1} & \dots \\ \dots & \frac{1}{\lambda_n} \end{bmatrix} \tilde{x} = \frac{\tilde{x}_1^2}{\lambda_1} + \dots + \frac{\tilde{x}_n^2}{\lambda_n} \leq 1$$

volum of ellipsoid proportional to  $\sqrt{\det(P)} = \sqrt{\prod_i \lambda_i}$

### 0.3 Problem Types

#### LP

standard, inequality, general forms

$$\begin{aligned} \min_x c^T x \text{ s.t. :} \\ Ax = b \\ x \succeq 0 \end{aligned}$$

$$\begin{aligned} \min_x c^T x \text{ s.t. :} \\ Ax \preceq b \end{aligned}$$

$$\begin{aligned} \min_x c^T x + d \text{ s.t. :} \\ Gx \preceq h \\ Ax = b \end{aligned}$$

#### QP

$$\begin{aligned} \min_x \frac{1}{2} x^T P x + q^T x + r \text{ s.t. :} \\ Gx \leq h \\ Ax = b \end{aligned}$$

#### QCQP

$$\begin{aligned} \min_x \frac{1}{2} x^T P_0 x + q_0^T x + r_0, \text{ s.t. :} \\ \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \forall i \\ Ax = b \end{aligned}$$

#### SOCP

$$\begin{aligned} \min_x f^T x \text{ s.t. :} \\ \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \forall i \\ Fx = g \end{aligned}$$

$$\begin{aligned} (\forall i) b_i = 0 &\implies LP \\ (\forall i) c_i = 0 &\implies QCQP \end{aligned}$$

#### GP

$$\min_x f_0(x) \text{ s.t. :}$$

$$f_i(x) \leq 1, \forall i$$

$$h_i(x) = 1, \forall i$$

$$f_i \text{ is a posynomial} := \sum_i h_i$$

$$h_i \text{ is a monomial} := cx_1^{a_1} x_2^{a_2} \dots, c > 0, a_i \in \mathbb{R}$$

Use transform of objective and constraint functions:

$$y_i = \log x_i, x_i = e^{y_i}$$

$h_i$  becomes exponential of affine function

$\tilde{f}_i = \log(f_i)$  becomes log sum exp (convex)

If all constraints and objective are monomials, reduces to LP after transform.

#### SDP

general, standard, inequality forms

$$\min_x c^T x \text{ s.t. :}$$

$$LMI : \sum_i^n x_i F_i + G \preceq_K 0$$

$$Ax = b$$

$$x \in \mathbb{R}^n$$

$$F_i, G \in S^m, K \in S_+^m$$

$$\min_X \text{tr}(CX) \text{ s.t. :}$$

$$\text{tr}(A_i X) = b_i, \forall i$$

$$X \succeq 0$$

$$\min_x c^T x \text{ s.t. :}$$

$$\sum_i^n x_i A_i \preceq_K B$$

$$Ax = b$$

$$B, A_i \in S^m, K \in S_+^m$$

concatenating constraints:

$$F^{(i)}(x) = \sum_j x_j F_i^{(j)} + G^{(i)} \preceq 0$$

$$Gx \preceq h$$

$$\implies$$

$$\text{diag}(Gx - h, F^{(1)}(x), \dots, F^{(m)}(x)) \preceq 0$$

if all matrices are diagonal, reduces to LP

## 0.4 Convex/Concave Functions

- Affine
- Pointwise supremum of convex function
  - distance to farthest point in a set
  - support function of set
- Norm
- Non-negative weighted sum of convex functions

### 0.4.1 log det X, concave

$$\begin{aligned}
 \text{let } X &= Z + tV \succ 0 \\
 f &= \log \det(Z + tV) \\
 f &= \log \det(Z^{-0.5}(I + tZ^{-0.5}VZ^{0.5})Z^{0.5}) \\
 f &= \log(\det(Z^{-0.5})\det(I + tZ^{-0.5}VZ^{0.5})\det(Z^{0.5})) \\
 f &= \log(\det(Z^0)\det(I + tZ^{-0.5}VZ^{0.5})) \\
 f &= \log \det(I + tZ^{-0.5}VZ^{0.5}) \\
 f &= \log \Pi_i(1 + \lambda_i t) \\
 f &= \sum_i \log(1 + \lambda_i t) \\
 \frac{\partial f}{\partial t} &= \sum_i \frac{\lambda_i}{1 + \lambda_i t} \\
 \frac{\partial^2 f}{\partial t^2} &= \sum_i \frac{-\lambda_i^2}{(1 + \lambda_i t)^2} = -\sum_i \frac{\lambda_i^2}{(1 + \lambda_i t)^2} \leq 0 \\
 \nabla^2 f &\leq 0 \iff f \text{ concave}
 \end{aligned}$$

### 0.4.2 log $\sum_i \exp(x_i)$ , convex

$$\begin{aligned}
 \nabla^2 f &= \frac{1}{(1^T z)^2} (1^T z \text{diag}(z) - zz^T) \\
 v^T zz^T v &= \det(v^T zz^T v) = \det(vv^T zz^T) \\
 v^T zz^T v &= \sum_j \sum_i z_j z_i v_j v_i \\
 v^T zz^T v &= (\sum_j z_j z_j) (\sum_i z_i v_i) \\
 v^T zz^T v &= (\sum_i z_i v_i)^2
 \end{aligned}$$

use Holder's Inequality :

$$\begin{aligned}
 \|a\|_2^2 \|b\|_2^2 &\geq |a^T b|^2 \\
 \text{let } a &= z_i^{0.5}, b = v_i z_i^{0.5} \\
 1^T z (\sum_i v_i^2 z_i) - (\sum_i z_i v_i)^2 &\geq 0 \\
 v^T \nabla^2 f v &= \frac{1}{(1^T z)^2} \left( 1^T z (\sum_i v_i^2 z_i) - (\sum_i z_i v_i)^2 \right) \geq 0 \\
 \nabla^2 f &\geq 0 \iff f \text{ convex}
 \end{aligned}$$

### 0.4.3 geometric mean on $R_{++}^n$ , concave

$$\begin{aligned}
 f &= (\Pi_i x_i)^{\frac{1}{n}} \\
 \frac{\partial}{\partial x_i} f &= \frac{1}{n} (\Pi_i x_i)^{\frac{1}{n}-1} \Pi_{j \neq i} x_j \\
 \frac{\partial^2}{\partial x_i^2} f &= \frac{1}{n} \left( \frac{1}{n} - 1 \right) (\Pi_i x_i)^{\frac{1}{n}-2} (\Pi_{j \neq i} x_j)^2 \\
 \frac{\partial^2}{\partial x_i^2} f &= \frac{1}{n} \left( \frac{1}{n} - 1 \right) \frac{(\Pi_i x_i)^{\frac{1}{n}}}{x_i^2} \\
 \frac{\partial^2}{\partial x_i \partial x_k} f &= \frac{1}{n^2} \frac{(\Pi_i x_i)^{\frac{1}{n}}}{x_i x_k}, i \neq k \\
 \frac{\partial^2}{\partial x_i \partial x_k} f &= \frac{1}{n^2} \frac{(\Pi_i x_i)^{\frac{1}{n}}}{x_i x_k} - \delta_{ik} \frac{1}{n} \frac{(\Pi_i x_i)^{\frac{1}{n}}}{x_i^2} \\
 v^T \nabla^2 f v &= \frac{-(\Pi_i x_i)^{\frac{1}{n}}}{n^2} \left( n \sum_i \frac{v_i^2}{x_i^2} - \left( \sum_i \frac{v_i}{x_i} \right)^2 \right) \\
 \text{apply Cauchy Schwartz Inequality :} \\
 \text{let } a &= \mathbf{1}, b_i = \frac{v_i}{x_i} \\
 \|\mathbf{1}\|_2^2 \left( \sum_i \frac{v_i^2}{x_i} \right) &\geq \left( \sum_i \frac{v_i}{x_i} \right)^2 \\
 n \sum_i \frac{v_i^2}{x_i^2} - \left( \sum_i \frac{v_i}{x_i} \right)^2 &\geq 0 \\
 v^T \nabla^2 f v &\leq 0 \iff f \text{ concave}
 \end{aligned}$$

#### 0.4.4 quadratic over linear, convex

$$f(x, y) = \frac{h(x)}{g(y)}, g(y) \text{ linear}, g(y) \in R_+$$

$$\nabla^2 f = vv^T \text{ is PSD} \iff f \text{ convex}$$

#### 0.5 Composition of functions

Mnemonic derivation from scalar composite function

$$f = h(g(x))$$

$$f' = g'(x)h'(g(x))$$

$$f'' = g''(x)h'(g(x)) + (g'(x))^2h''(g(x))$$

$$h \text{ convex \& non-decreasing, } g \text{ convex} \implies f \text{ convex}$$

$$h'' \geq 0, g''(x) \geq 0, h'(g(x)) \geq 0 \implies f'' \geq 0$$

$$h \text{ convex \& non-increasing, } g \text{ concave} \implies f \text{ convex}$$

$$h'' \geq 0, g''(x) \leq 0, h'(g(x)) \leq 0 \implies f'' \geq 0$$

$$h \text{ concave \& non-decreasing, } g \text{ concave} \implies f \text{ concave}$$

$$h'' \leq 0, g''(x) \leq 0, h'(g(x)) \geq 0 \implies f'' \leq 0$$

$$h \text{ concave \& non-increasing, } g \text{ convex} \implies f \text{ concave}$$

$$h'' \leq 0, g''(x) \geq 0, h'(g(x)) \leq 0 \implies f'' \leq 0$$

## 0.6 Convexity Preservation of Sets

### 0.6.1 Intersection

$$(\forall \alpha \in A) S_\alpha \text{ is convex cone} \implies \\ \cap_{\alpha \in A} S_\alpha \text{ is convex cone}$$

Any closed convex set can be represented by possibly infinitely many half spaces.

### 0.6.2 Affine functions

let  $f(x) = Ax + b, f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

then if  $S$  is a convex set we have:

- project forward:  $f(S) = \{f(X) : X \in S\}$  is convex
- project back:  $f^{-1}(S) = \{X : f(X) \in S\}$  is convex

Example:

$$C = \{y : y = Ax + b, \|x\| \leq 1\}$$

$\|x\| \leq 1$  is convex,  $Ax + b$  is affine  $\implies C$  is convex

Example:

$$C = \{x : \|Ax + b\| \leq 1\}$$

$\{y : \|y\| \leq 1\}$  is convex  $\wedge y$  is an affine function of  $x \implies C$  is convex

## 0.7 Constraint Qualifications

### 0.7.1 Slater's Constraint Qual.

Optimal solution is in relative interior:  $x^* \in \text{relint}(S)$

Inequalities  $(\forall i) f_i(x)$  convex  $\wedge f_i(x) < 0 \implies$   
Slater's constraint satisfied.

Inequalities  $(\forall i) f_i(x)$  affine  $\implies (\forall i) f_i(x) \leq 0 \wedge (\exists i) f_i(x) < 0 \implies$  Slater's constraint satisfied.

Achieving Slater's constraint implies 0 duality gap.

### 0.7.2 KKT

Assumes optimality achieved with 0 duality gap:  
 $\nabla L(x^*, \lambda^*, v^*) = 0$

$$L(x^*, \lambda^*, v^*) = f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i v_i h_i(x^*)$$

$$\nabla L(x^*, \lambda^*, v^*) = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_i v_i \nabla h_i(x^*)$$

We have the constraints:

$$f_i(x^*) \leq 0$$

$$\lambda_i^* \geq 0$$

$$h_i(x^*) = 0$$

$$\lambda_i^* f_i(x^*) = 0$$

$$\nabla L(x^*, \lambda^*, v^*) = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_i v_i \nabla h_i(x^*)$$

Primal inequality constraints convex and equality constraints affine and KKT satisfied  $\implies$  0 duality gap with specified points for primal and dual. (sufficient).

If Slater's constraint satisfied then the above is sufficient and necessary:

Primal inequality constraints convex and equality constraints affine and KKT satisfied  $\iff$  0 duality gap with specified points for primal and dual.



## 0.8 Definitions

### 0.8.1 Convex Function

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \forall \theta = [0, 1]$$

For convenience we sometimes define an extended value function:

$$\tilde{f}(x) = \begin{cases} f(x), & x \in \text{dom}(f) \\ \infty, & \text{other wise} \end{cases}$$

if  $f(x)$  convex, then  $\tilde{f}$  is also convex

Sublevel set of a function

$$C(\alpha) = \{x \in \text{dom}(f) : f(x) \leq \alpha\}$$

For convex function, all sublevel sets are convex ( $\forall \alpha$ ).  
Converse is not true.

Quasi-convex function: if its sublevel sets are all convex.

Epigraph of functions:

$$\text{epi}(f) = \{(x, t) : x \in \text{dom}(f), f(x) \leq t\} \in \mathbb{R}^{n+1}, \\ f \in \mathbb{R}^n \rightarrow \mathbb{R}.$$

$$f \text{ is convex function} \iff \text{epi}(f) \text{ is convex set}$$

### 0.8.2 First order condition

Suppose  $f$  is differentiable and domain of  $f$  is convex.  
Then:

$$f \text{ convex} \iff$$

$$(\forall x, x_0 \in \text{dom}(f)) f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

rough proof:

$$\text{suppose } f(x) \text{ is convex but } (\exists x, x_0) f(x) < f(x_0) + \nabla f(x_0)^T (x - x_0)$$

then this means the function should bend across the tangent line which violates the convexity

proof for converse direction:

$$\text{suppose that } (\exists x, x_0) f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

to show that  $f(x)$  is convex lets take  $x, y \in \text{dom}(f), z = \theta x + (1 - \theta)y$   
 $\theta f(x) + (1 - \theta)f(y) \geq f(z) + \nabla f(z)^T (\theta x - \theta z + (1 - \theta)y - (1 - \theta)z)$   
 $\theta f(x) + (1 - \theta)f(y) \geq f(\theta x + (1 - \theta)y)$   
 $f(x)$  is convex

### 0.8.3 Second order condition

Suppose  $f$  is twice differentiable and  $\text{dom}(f)$  is convex,  
then  $f(x)$  is convex  $\iff \nabla^2 f(x) \geq 0$  (PSD, eg: wrt.  $S_+^n$ )

proof for scalar case:

suppose that  $f(x)$  is convex, then the first-order condition holds

$$\text{for } x, y \in \text{dom}(f) : f(x) \geq f(y) + f'(y)(x - y)$$

$$\text{for } y, x \in \text{dom}(f) : f(y) \geq f(x) + f'(x)(y - x)$$

$$f'(x)(y - x) \leq f(y) - f(x) \leq f'(y)(y - x)$$

$$f'(x)(y - x) \leq f'(y)(y - x) \implies 0 \leq (y - x)(f'(x) - f'(y))$$

$$\text{take } y \rightarrow x : 0 \leq f''(x)$$

$$f''(x) \geq \frac{f'(x+\delta x) - f'(x)}{\delta x}$$

conversely, suppose that  $f'(z) \geq 0, \forall z \in \text{dom}(f)$ ,  
take  $x, y \in \text{dom}(f)$  WLOG  $x < y$

$$\int_x^y f''(z)(y - z) dz \geq 0$$

$$f''(z) \geq 0, (y - z) \geq 0 = I_1 + I_2$$

$$I_1 = \int_x^y f''(z)y dz - y f'(z)|_x^y = y(f'(y) - f'(x))$$

$$I_2 = - \int_x^y f''(z) dz$$

$$dv = f''(z) dz \implies v = f'(z)$$

$$u = z \implies du = dz$$

$$I_2 = -z f'(z)|_x^y + \int_x^y f(z) dz = -y f'(y) + x f'(x) + f(y) - f(x)$$

$$I_1 + I_2 = y f'(y) - y f'(x) - y f'(y) + x f'(x) + f(y) - f(x) \geq 0$$

$$\implies f(y) \geq f(x) + f'(x)(y - x) \text{ first order condition:}$$

$$x < y$$

$$\text{first order condition holds} \implies f(x) \text{ convex}$$

### 0.8.4 Inequalities

$$x \preceq_K y \iff y - x \in K$$

$x$  is a minimum in  $S$  wrt. cone  $K$ :

$$x \in S : (\forall y \in S) f(y) \succeq_K f(x) \iff f(y) - f(x) \in S$$

$$S \subseteq x + K$$

$x$  is a minimal in  $S$  wrt. cone  $K$ :

$$x \in S : (\forall y \in S) f(y) \preceq_K f(x) \implies x = y$$

$$x \in S : (\forall y \in S) f(x) - f(y) \in K \implies x = y$$

$$(x - K) \cap S = \{x\}$$

### 0.8.5 Cone

$$(\forall x \in C, \forall \theta \geq 0) \theta x \in C$$

### 0.8.6 Convex Cone

Eg:  $S^n, S_+^n$  are convex cones  
convexity check for  $S_+^N$ :

$$x_1 \in S_+^n \implies v^T x_1 v \geq 0$$

$$x_2 \in S_+^n \implies v^T x_2 v \geq 0$$

$$v^T (\theta x_1 + (1 - \theta) x_2) v \geq 0$$

$$v^T \theta x_1 v + (1 - \theta) v^T x_2 v \implies x \in S_+^n$$

convexity check for cone:

$$x_1 \in S_+^n \implies \theta x_1 \in S_+^n, \theta \geq 0$$

$$(\forall v) v^T x v \geq 0 \implies v^T (\theta x) v \geq 0 \implies \text{cone}$$

### 0.8.7 Proper Cone

Definition:

- convex
- closed (contains all limit points)
- solid (non-empty interior)
- pointed (contains no line):  $x \in K \implies -x \notin K$

Then the proper cone  $K$  defines a generalized inequality ( $\leq_K$ ) in  $\mathbb{R}^n$

$$x \leq_K y \implies y - x \in K$$

$$x <_K y \implies y - x \in \text{int}(K)$$

Example:  $K = R_+^n$  (non-negative orthant):

$$n = 2$$

$$x \leq_{R_+^2} y \implies y - x \in R_+^2$$

Cone provides partial ordering using difference of 2 objects.

$$X \leq_{S_+^n} Y \iff Y - X \in S_+^n \iff Y - X \text{ is PSD}$$

### 0.8.8 Norm Cone

$$K = \{(x, t) \in \mathbb{R}^{n+1} : \|x\| \leq t\}, x \in \mathbb{R}^n$$

### 0.8.9 Dual norm

$$\|z\|_* := \sup_x \{z^T x : \|x\|_p \leq 1\}$$

Dual of L1-norm:

$$\|z\|_* := \sup_x \{z^T x : \|x\|_1 \leq 1\}$$

$$\max \sum_i z_i x_i,$$

$$\text{subject to : } \sum_i \|x_i\| \leq 1$$

select  $x_i$  corresponding to  $z_i$  with maximum absolute value

$$\text{equivalent to } \|z\|_* = \|z\|_\infty$$

Dual of L- $\infty$ -norm:

$$\|z\|_* := \sup_x \{z^T x : \|x\|_\infty \leq 1\}$$

$$\max \sum_i z_i x_i,$$

$$\text{subject to : } \|x_i\| \leq 1, \forall i$$

choose  $x_i = 1$  if  $z_i \geq 0$  and  $x_i = 0$  if  $z_i < 0$

$$\text{equivalent to } \|z\|_* = \|z\|_1$$

Dual norm of Lp-norm: Lq-norm where  $1/p + 1/q = 1$

Properties:

- $K^*$  closed and convex
- $K_1 \subseteq K_2 \implies K_1^* \subseteq K_2^*$
- $K$  has non-empty interior  $\implies K^*$  pointed
- $cl(K)$  pointed  $\implies K^*$  has non-empty interior
- $K^{**} = cl(\text{convhull}(K))$
- $K$  convex and closed  $\implies K = K^{**}$

### 0.8.10 Operator norm

$$\|X\|_{a,b} = \sup\{\|Xu\|_a : \|u\|_b \leq 1\}, X \in \mathbb{R}^{m \times n}$$

### 0.8.11 Dual cone

$$K \text{ is a cone}$$

$$K^* = \{y : x^T y \geq 0, \forall x \in K\}$$

### 0.8.12 Dual norm cone

$$K^* = \{(u, v) : \|u\|_* \leq v\}$$

$$\text{where } K = \{(x, t) : \|x\| \leq t\}$$

### 0.8.13 support function of a set

$$S_C(x) = \sup\{x^T y : y \in C\}$$

$$\text{dom}(S_C) = \{x : \sup_{y \in C} x^T y < \infty\}$$

It is pointwise supremum of convex function, so it is convex.

## 0.9 Regularized Approximation

Noise sensitivity of different objectives:

- robust least squares / Huber penalty
- log barrier
- deadzone linear
- quadratic

Least norm problems

- L2 norm objective with equality constraint
- sparsity inducing norms (eg: L1)
- norm ball constraint
- probability distribution
  - convex comb. of columns of  $A$  to fit  $b$
- variable constraints
  - box
  - one sided bound

Multicriterion Formulation

Tikhonov Regularization  
todo..

## 0.10 Descent Methods

---

### Algorithm 1: Descent Overview

---

```

1 init  $x_0 \in \text{dom} f$ ;
2 while stopping criterion is not satisfied do
3    $\Delta x \leftarrow$  Compute Descent Direction;
4    $t \leftarrow$  Compute Descent Step Size;
5    $x_{k+1} \leftarrow x_k + t\Delta x$ 

```

---

### 0.10.1 Search Step Size

Given  $\Delta x$ , step direction

Exact Line Search:  $t = \operatorname{argmin}_{s \geq 0} f(x + s\Delta x)$

---

### Algorithm 2: Backtracking Line Search

---

$\Delta x$ : Search direction  
 $t$  : Step size

```

1  $\alpha \in (0, 0.5)$ ;
2  $\beta \in (0, 1)$ ;
3  $t \leftarrow 1$ ;
4 while  $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$  do
5    $t \leftarrow \beta t$ ;

```

---

## 0.10.2 Search Direction - 1st Order Methods

Steepest Descent

$$\operatorname{argmin}_{\Delta x} f(x) + \nabla f(x)^T \Delta x_{sd}$$

$$\operatorname{argmin}_{\Delta x} \nabla f(x)^T \Delta x_{sd}$$

Normalized Steepest Descent

$$\Delta x_{nsd} = \operatorname{argmin}_v \{ \nabla f(x)^T v : \|v\| \leq 1 \}$$

$$\Delta x_{nsd} = \frac{\Delta x_{sd}}{\|\nabla f(x)\|_*}$$

$$\|\nabla f(x)\|_* = \sup_y \{ \nabla f(x)^T y : \|y\| \leq 1 \} \text{ (dualnorm)}$$

$$\nabla f(x)^T \Delta x_{sd} = \nabla f(x)^T \Delta x_{nsd} \|\nabla f(x)\|_* = \|\nabla f(x)\|_*^2$$

Steepest Descent for L2-Norm

$$\Delta x = -\nabla f(x)$$

$$\Delta x_{nsd} = \frac{-\nabla f(x)}{\|\nabla f(x)\|_*}$$

Steepest Descent for Quadratic Norm

$$\|z\|_P = (z^T P z)^{\frac{1}{2}} = \|P^{\frac{1}{2}} z\|_2, P \in S_{++}^n$$

$$\Delta x_{nsd} = -(\nabla f(x)^T P^{-1} \nabla f(x))^{-\frac{1}{2}} P^{-1} \nabla f(x)$$

$$\Delta x_{sd} = -P^{-1} \nabla f(x)$$

Normalized Steepest Descent for L1-norm

$$\Delta x_{nsd} = \operatorname{argmin}_v \{ \nabla f(x)^T v : \|v\|_1 \leq 1 \}$$

$$\Delta x_{nsd} = -\operatorname{sign}\left(\frac{\partial f(x)}{\partial x_i}\right) e_i, i : \|f(x)\|_\infty = |(\nabla f(x))_i|$$

## Search Direction - 2nd Order Methods

Newton's Method

$$\begin{aligned}\Delta x_{nt} &= -\nabla^2 f(x)^{-1} \nabla f(x) \\ \nabla^2 f(x)^{-1} &\succ 0 \implies \\ \nabla f(x)^T \Delta x_{nt} &= -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0\end{aligned}$$

Affine invariance of Newton's method: auto-scaling of level curves to enable better descent direction

2nd order Taylor approximation of function is minimized with  $\Delta x_{nt}$ , so quadratic model gives good approximation near minimizer.

Newton Decrement: used for convergence/stopping criterion

$$\begin{aligned}\lambda(x) &= (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{\frac{1}{2}} \\ \frac{1}{2} \lambda(x)^2 &\approx f(x) - p^* \\ \nabla f(x)^T \Delta x_{nt} &= -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \\ \nabla f(x)^T \Delta x_{nt} &= -\lambda(x)^2 = \frac{d}{dt} f(x + \Delta x_{nt} t)|_{t=0}\end{aligned}$$

---

### Algorithm 3: Newton Method Descent

---

```

1 init  $x_0 \in \text{dom} f$ ;
2  $\Delta x_{nt} \leftarrow -\nabla^2 f(x)^{-1} \nabla f(x)$ ;
3  $\lambda(x)^2 \leftarrow \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$ ;
4 while  $\frac{1}{2} \lambda(x)^2 > \epsilon$  do
5    $t \leftarrow \text{Compute Step Size (line search)}$ ;
6    $x_{k+1} \leftarrow x_k + t \Delta x_{nt}$ 
```

---

Convergence: split into damped phase and quadratically convergent phase. Fast convergence once step size is 1.

BFGS: a Quasi-newton method  
Direction:

$$p_k = -B_k^{-1} \nabla f_k$$

Properties for  $B$ :

$$\begin{aligned}B &\succ 0 \\ B &= B^T\end{aligned}$$

Satisfaction of Secant equation:

$$\begin{aligned}B_{k+1}(x_{k+1} - x_k) &= \nabla f_{k+1} - \nabla f_K \\ B_{k+1} s_k &= y_k\end{aligned}$$

Curvature condition:

$$\begin{aligned}B_{k+1} s_k &= y_k \\ s_k^T B_{k+1} s_k &= s_k^T y_k \\ B_{k+1} \succ 0 &\implies s_k^T y_k > 0\end{aligned}$$

This need to be enforced if doing optimization on a non-convex function

General problem of find  $B$

$$\begin{aligned}\min_B &\|B - B_k\| \\ \text{s.t. } &B = B^T, Bs - K = y_k\end{aligned}$$

$$\text{BFGS Update: } B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k}$$

---

### Algorithm 4: BFGS Descent

---

```

1 init  $x_0 \in \text{dom} f$ 
2 init  $B_0$  (eg:  $I$ )
3 while  $\|y_k\| > \epsilon$  do
4   solve for  $p_k$  in  $B_k p_k = -\nabla f(x_k)$ 
5    $t \leftarrow \text{argmin}_s f(x_k + s p_k)$  (or backtrack search)
6    $s_k \leftarrow t p_k$ 
7    $x_{k+1} \leftarrow x_k + s_k$ 
8    $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ 
9    $B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k}$ 
```

---

Inversion of  $B$  via Sherman-Morrison:

$$B_{k+1}^{-1} = \left( I - \frac{s_k y_k^T}{y_k^T s_k} \right) B_k^{-1} \left( I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \frac{s_k s_k^T}{y_k^T s_k}.$$

Sherman-Morrison:

$$\begin{aligned}A^{-1} \text{ exists: } (A + uv^T)^{-1} \text{ exists} &\iff 1 + v^T A^{-1} u \neq 0 \\ (A + uv^T)^{-1} &= A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}\end{aligned}$$

## 0.11 Appendix

### 0.11.1 Gradient of Log Det

$$f(x) = \log(\det X)$$

$$\begin{aligned} f(X + \delta X) &= \log \det(X + \delta X) \\ &= \log \det((X^{\frac{1}{2}}(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})X^{\frac{1}{2}}) \\ &= \log \det(X) + \log \det(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}}) \\ \text{let } M &= X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}} \\ &= \log \det(X) + \log \det(I + M) \end{aligned}$$

claim eigenvalues of  $I + M$ :  $1 + \lambda_i$

$$\begin{aligned} Mv_i &= \lambda_i v_i \\ (I + M)v_i &= (1 + \lambda_i)v_i \\ \det(M) &= \prod_i (1 + \lambda_i) \end{aligned}$$

$$\begin{aligned} f(X + \delta X) &= \log \det(X) + \log \prod_i (1 + \lambda_i) \\ &= \log \det(X) + \sum_i \log(1 + \lambda_i) \\ &\approx \log \det(X) + \sum_i \lambda_i \text{ since } \delta X \text{ is small} \\ &\approx \log \det(X) + \text{trace}(X^{-\frac{1}{2}}\delta X X^{\frac{1}{2}}) \\ &\approx \log \det(X) + \text{trace}(X^{-1}\delta X) \end{aligned}$$

$$\text{trace}(X^{-1}\delta X) = (X^{-T})^T \delta X$$

$$f(X + \delta X) = f(X) + (\nabla f(X))^T \delta X \implies \nabla f(X) = X^{-1}$$

$$\log \det(X) = \log(X) \implies f'(X) = \frac{1}{X}$$

### 0.11.2 2nd order approximation of Log Det

$$f(X + \delta X) = f(X) + \langle \nabla f(X), \delta X \rangle + \frac{1}{2} \langle \delta X, \nabla^2 f(X) \delta X \rangle$$

first look at first order approximation:  $g(X) = X^{-1}$

$$\begin{aligned} g(X + \delta X) &= (X + \delta X)^{-1} = (X^{\frac{1}{2}}(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})X^{\frac{1}{2}})^{-1} \\ &= X^{-\frac{1}{2}}(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})^{-1}X^{-\frac{1}{2}} \\ \text{for small } A (\text{small eigenvalues}): (I + A)^{-1} &\approx I - A \\ &= X^{-\frac{1}{2}}(I - X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})X^{-\frac{1}{2}} \\ &= X^{-1} - X^{-1}\delta X X^{-1} \end{aligned}$$

$$\log \det(X + \delta X) = \log \det(X) + \text{tr}(X^{-1}\delta X) - \frac{1}{2} \text{tr}(\delta X X^{-1}\delta X X^{-1})$$

## 0.12 Miscellaneous Properties

$$(a+x)^{-1} \approx 1-x$$

$$\lim_{t \rightarrow 0} \frac{f(x+\epsilon t) - f(x)}{t} = \frac{\partial f(x)}{\partial x} \epsilon$$

### 0.12.1 Pseudo-inverse

Overconstrained case:

Cast as L2 norm approximation problem

$$\min_x \|Ax - b\|_2^2$$

$$(Ax - b)^T (Ax - b) = x^T A^T Ax - 2x^T A^T b + b^T b$$

$$\frac{\partial}{\partial x} (x^T A^T Ax - 2x^T A^T b + b^T b) = 2A^T Ax - 2A^T b$$

$$x = (A^T A)^{-1} A^T b$$

Underconstrained case:

Cast as a least-norm problem w/ equality constraint

$$\min_x \|x\|_2^2$$

$$s.t. : Ax = b$$

$$\min_x L(x, \lambda, v) = x^T x - v^T (Ax - b)$$

$$\frac{\partial (x^T x - v^T (Ax - b))}{\partial x} = 2x - A^T v = 0$$

$$x = -\frac{1}{2} A^T v$$

$$g(\lambda, v) = [x^T x - v^T (Ax - b)]_{x=-\frac{1}{2} A^T v}$$

$$g(\lambda, v) = -\frac{1}{4} v^T A A^T v - v^T b$$

$$dual : \max_{\lambda, v} g(\lambda, v) = -\min_{\lambda, v} g(\lambda, v)$$

$$\frac{\partial (\frac{1}{4} v^T A A^T v + v^T b)}{\partial v} = 0$$

$$v = -\frac{1}{2} A A^T v + b = 0$$

$$v = -2(AA)^{-1} b$$

$$x = -\frac{1}{2} A^T v|_{v=-2(AA)^{-1}b} = A^T (AA)^{-1} b$$