Definiteness

A positive definite $\iff x^T Ax > 0 \;\forall x \neq 0$
A positive semidefinite $\iff x^T Ax \geq 0 \;\forall x$
If A is positive definite:

1. A is invertible
2. $\forall$ eigenvalues λ_i, $\lambda_i > 0$
3. n linearly independent vectors such that $A_{ij} = x_i^T x_j$

If A is positive semidefinite:

1. $A + \gamma I$ is positive definite for any $\gamma > 0$.

Diagonalization and EVD

If $n \times n$ matrix A has n linearly independent eigenvectors, $A = PDP^{-1}$. If A is also symmetric, $P^{-1} = P^T$, so $A = PDP^T$.
$A = U\Lambda U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$ where u_i's form an orthonormal basis.
Example:

$$S = \begin{bmatrix}2\\6\end{bmatrix}\begin{bmatrix}2\\6\end{bmatrix}^T + \begin{bmatrix}-3\\1\end{bmatrix}\begin{bmatrix}-3\\1\end{bmatrix}^T$$

S is in the form $S = v_1 v_1^T + v_2 v_2^T$. The eigenvalue decomposition: $u_1 = \frac{v_1}{r_1} = \frac{1}{\sqrt{40}}\begin{bmatrix}2\\6\end{bmatrix}$, $u_2 = \frac{v_2}{r_2} = \frac{1}{\sqrt{10}}\begin{bmatrix}-3\\1\end{bmatrix}$, $\lambda_1 = r_1^2 = 40$, $\lambda_2 = r_2^2 = 10$

Ellipsoids

Bounded ellipsoid centered around origin where P is PD:

$$E = \{x \in \mathcal{R}^n : x^T P^{-1} x \leq 1\}$$

u_i's of P define the directions of the semi-axes and their lengths are given by $\sqrt{\lambda_i}$ where $\lambda_i > 0$ (since P is PD). There exists $P^{-1} = A^T A$ so $xP^{-1}x = x^T A^T Ax = ||Ax||_2^2$ so we get the following equivalent formulation: $E = \{x \in \mathcal{R}^n : ||Ax||_2 \leq 1\}$

Example

Shapes constrained by $||Ax||_2 \leq 1$

1. $A = \begin{bmatrix}-1 & 0 & 0\\0 & 1 & 0\end{bmatrix}$, $||Ax||_2 = x^T A^T Ax \leq 1$,
$A^T A = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$.

$$\begin{bmatrix}x_1 & x_2 & x_3\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = x_1^2 + x_2^2 \leq 1$$

This describes a cylinder of radius one centered around the x_3 axis.

2. $A = \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}$, $||Ax||_2 = x^T A^T Ax \leq 1$, $A^T A = \begin{bmatrix}2 & -2\\-2 & 2\end{bmatrix}$.

$$\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}2 & -2\\-2 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = 2(x_1 - x_2)^2 \leq 1 \to (x_1 - x_2)^2 \leq \frac{1}{2}$$

$x_1 - x_2 \leq \frac{1}{\sqrt{2}}$ and $x_1 - x_2 \geq \frac{-1}{\sqrt{2}}$. We see this gives us a plane, the region between the two parallel lines.

3. $$A = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$$

We are given the eigenvalues are 3 and 1. $x^T A^2 x \leq 1$. We know that A is PD so A^2 is has eigenvalues 9 and 1 with eigenvectors $u_1 = (1,1)/\sqrt{2}$ and $u_2 = (1,-1)/\sqrt{2}$. Let $x = U\bar{x}$.

$$1 \geq x^T A^2 x = x^T U \Lambda^2 U^T x = \lambda_1^2 \bar{x}_1^2 + \lambda_2^2 \bar{x}_2 = 9\bar{x}_1^2 + \bar{x}_2^2$$

We get the ellipse with principle axes u_1, u_2 and semi-axis lengths 1 and $\frac{1}{3}$.

Traces

$Tr(A) = a_{11} + \cdots + a_{nn} = \sum_{i=1}^n a_{ii}$
$Tr(A+B) = Tr(A) + Tr(B) | Tr(cA) = cTr(A) | Tr(A) = Tr(A^T)$
Cyclic permutations:
$Tr(ABCD) = Tr(BCDA) = Tr(CDAB) = Tr(DABC)$
$Tr(XY^T) = \sum_{i,j} X_{ij} Y_{ij}$
If A is $n \times n$, $Tr(A) = \sum_i \lambda_i$, $det(A) = \prod_i \lambda_i$, $Tr(A^k) = \sum_i \lambda_i^k$
Let x be a scalar. Then $x = Tr(x)$.
Matrix Inversion Lemma
$w = (X^T X + \lambda I)^{-1} X^T y = X^T (XX^T + \lambda I)^{-1} y$
Rayleigh Quotient: $R(M, x) = \frac{x^T Mx}{x^T x}$
Range and Nullspaces $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp = \mathcal{N}(A^T A)^\perp = \mathcal{R}(A^T A)$

Calculus

$$\frac{df}{dx} \approx \frac{1}{h}(f(x+h, y) - f(x, y))$$

Principle Components Analysis

Singular Value Decomposition

$$X = USV^T = \sum_{i=1}^d \sigma_i u_i v_i^T$$

1. U is $d \times d$ with the left singular vectors
2. S is $d \times d$ with the singular values on the diagonal.
3. V is $n \times d$ with the right singular vectors
4. $U^T U = I$ and $V^T V = I$ since both U and V contain orthogonal vectors.
5. $S = Diag(\sigma_i)$ where singular values ordered from greatest to least $\sigma_1 \geq \cdots \geq 0$.

The singular values of X are the square roots of the eigenvalues for $X^T X$ or XX^T. The left singular vectors are the eigenvectors of XX^T and the right singular vectors are the eigenvectors of $X^T X$.
Recovering U or V given A, Σ, and the other

$$AV\Sigma^{-1} = U\Sigma V^T V \Sigma^{-1} = U$$
$$\Sigma^{-1} U^T A = \Sigma^{-1} U^T U \Sigma V^T = V^T$$

Norms in terms of SVD
$||A||_F^2 = Trace(A^T A) = Trace(VS^T U^T USV^T) = Trace(VSSV^T) = Trace(V^T VSS) = Trace(SS) = \sum_i \sigma_i^2$
$||A||_2^2 = sup_{||x||=1} ||Ax||_2 = sup_{||x||=1} x^T A^T Ax$ We see that we are just finding the suprenum of the Raleigh quotient, which is maximized with the largest eigenvalue of $A^T A$ which is equivalent to σ_1^2.
PCA Algorithm
Center X, compute SVD of X, then return $\hat{X} = S_r V_r^T, U_r, \mu_x$
Latent Factor Analysis
Goal: Factor X in AB^T. $min_{A,B} ||X - AB^T||_F^2$. The solution is $A = U_r S_r^{\frac{1}{2}}$ and $B = V_r S_r^{\frac{1}{2}}$.
Low Rank Approximation

$$min_{\hat{D}} ||D - \hat{D}||_F : rank(D) \leq r$$

Solved by $\hat{D}^* = \sum_{i=1}^r \sigma_i u_i v_i^T$.
EVD to SVD

$$A = \sum_{i=1}^n \lambda_i v_i v_i^T$$
$$= \sum_{i=1}^n |\lambda_i| (sgn(\lambda_i) v_i) v_i^T$$

Choose $\sigma_i = |\lambda_i|$ and $u_i = sgn(\lambda_i) v_i$. Order such that $|\lambda_1| > \cdots > |\lambda_r|$.

Vectors and Matrices

Norms

- $||A||_F^2 = Tr(A^T A)$
- $||x||_2^2 = \sum_{i=1}^n x_i^2 = x^T x$
- $||x||_1 = \sum_{i=1}^n |x_i|$

Norm Properties

- $||\alpha A|| = |\alpha| ||A||$ for every A and scalar α.
- Triangle inequality holds for any two matrices A, B of same size, and u of appropriate size with $||u||_1 = 1$. o

Triangle Inequality

$$||a + b|| \leq ||a|| + ||b||$$

Cachy-Schwarz Inequality: $|x^T y| = ||x||_2 ||y||_2$
Cosine Similarity

$$ab = ||a|| ||b|| \cos\theta$$
$$\cos\theta = \frac{ab}{||a|| ||b||}$$

If two vectors are orthogonal, then the are linearly independent.
Derivatives

- $\frac{\partial x^T a}{\partial x} = \frac{\partial a^T x}{x} = a$
- $\frac{\partial a^T Xb}{\partial X} = ab^T$
- $\frac{\partial b^T X^T Xc}{\partial X} = X(bc^T + cb^T)$
- $\frac{\partial x^T Bx}{\partial x} = (B + B^T)x$
- $\frac{\partial Tr(F(X))}{\partial X} = f(X)^T$
- $\frac{\partial}{\partial x} ||x - a||_2 = 2x$
- $\frac{\partial}{\partial X} ||X||_F^2 = 2X$

Projections
We solve the following optimization problem: $\min_t ||x - z||_2$ where $x = x_0 + tu$. Given vector x passing, find projection of z on line through origin x_0 with normalized direction u. We have a arg min over t, which gives us the magnitude of u. General solution:

$$t = \frac{u^T(z - x_0)}{||u||_2^2}$$
$$z^* = x_0 + \frac{u^T(x - x_0)u}{u^T u}$$

Convex Sets

Set C convex if for any $\theta \in [0, 1]$ and for any $x1, x2 \in C$, $\theta x_1 + (1-\theta)x_2 \in C$.
Operations That Preserve Convexity

1. Intersection: If S_1 and S_2 convex, $S_1 \cap S_2$ convex.
2. Affine Functions: Affine functions are linear functions plus a constant, $f(x) = Ax + b$. Let S be convex. Then $f(S) = \{f(x) | x \in S\}$ is convex.

Convex Functions

Function f is convex if dom$f$ is convex set and $\forall x, y \in$ dom$f$ and $\theta \in [0,1]$, $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$.
Extended-value Extensions:

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathrm{dom} f \\ \infty & x \notin \mathrm{dom} f \end{cases}$$

First Order Conditions: $f(y) \geq f(x) + \nabla f(x)^T (y - x)$
Second Order Conditions: $\nabla^2 f(x) \succeq 0$
Pointwise Max/Supremum: If $f_1, \ldots, f_m$ convex, then $f(x) = \max f_1(x), \ldots, f_m(x)$ also convex.

Linear Programs

General Form of LPs

$$\min_x c^T x + d$$
$$Gx \preceq h$$
$$Ax = b$$

Standard Form of LPs

Only inequalities are component-wise non-negativity constraints $x \succeq 0$.

$$\min_x c^T x + d$$
$$Ax = b$$
$$x \succeq 0$$

Transformation to Standard Form

Step 1: Introduce slack variable s_i.

$$\min_x c^T x + d$$
$$Gx + s = h$$
$$Ax = b$$
$$s \succeq 0$$

Step 2: Express $x = x^+ - x^-$, $x^+, x^- \succeq 0$.

$$\min c^T x^+ - c^T x^- + d$$
$$Gx^+ - Gx^- + s = h$$
$$Ax^+ - Ax^- = b$$
$$s \succeq 0, x^+ \succeq 0, x^- \succeq 0$$

LP Formulation Example

$y(t) = h_0 u(t) + h_1 u(t-1) + \cdots + h_m u(t-m),\ t = 0, 1, 2, \ldots$ and we want inputs $u(0), \ldots, u(T)$. We track desired output $y_{des}(t)$

1. Minimize $\max_{0 \le t \le T} |y(t) - y_d es(t)|$.
2. Inputs zero for $t < 0$, $t > M$, where $M < T$ given.
3. $|u(t) \le U,\ |u(t+1) - u(t)| \le S$

We want to formulate an LP. Using the given constraints we get:

$$\min_{u(0),\ldots,u(T)} \max_{0 \le t \le T} |y(t) - y_{des}(t)|$$
$$|u(t)| \le U,\ |u(t+1) - u(t)| \le S,\ t = 0, \ldots, M$$
$$u(t) = 0,\ t = M+1, \ldots, T$$

We first deal with the absolute value constraints.
$|u(t)| \le U \to \max(u(t), -u(t)) \le U \to -u(t) \le U,\ u(t) \le U \to -U \le u(t) \le U,\ t = 0, \ldots, M$ and
$|u(t+1) - u(t) \le S| \to \max(u(t+1) - u(t), -(u(t+1) - u(t))) \to -(u(t+1) - u(t)) \le S \le u(t+1) - u(t)$
We also know that $y(t) = \sum_{i=1}^m h_i u(t-i)$. Thus we get the following objective:

$$\min_{u(0),\ldots,u(T)} \max_{0 \le t \le T} |\sum_{i=1}^m h_i u(t-i) - y_{des}(t)|$$
$$w = \max_{0 \le t \le T} \sum_{i=1}^m h_i u(t-i) - y_{des}(t)$$
$$\min_{u(0),\ldots,u(T)} w$$
$$-w \le \sum_{i=1}^m h_i u(t-i) - y_{des}(t) \le w,\ t = 0, \ldots, T$$

We also include the other constraints defined above.

Quadratic Programs

$$p^* = \min_x \frac{1}{2} x^T P x + q^T x$$
$$Gx \preceq h,\ Ax = b$$

Convex Optimization Strategies

Standard Form

$$\min f_0(x)$$
$$f_i(x) \ge 0,\ i = 1, \ldots, m$$
$$h_i(x) = 0,\ i = 1, \ldots, m$$

Change of Variables

$\phi : \mathbb{R}^n \to \mathbb{R}^n$. Define $\tilde{f}_i(z) = f_i(\phi(z)),\ i = 0, \ldots, m$, $\tilde{h}_i(z) = h_i(\phi(z)), i = 1, \ldots, p$. We have the following equivalent formulation:

$$\min \tilde{f}_0(z)$$
$$\tilde{f}_i(z) \ge 0,\ i = 1, \ldots, m$$
$$\tilde{h}_i(z) = 0,\ i = 1, \ldots, m$$

where we substitute $x = \phi(z)$ and $z = \phi^{-1}(x)$ solves the problem with the changed variables.

Transformation of Objective and Constraint Functions

$v_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing, $v_1, \ldots, v_m : \mathbb{R} \to \mathbb{R}$ satisfy $v_i(u) \le 0 \iff u \le 0$ and $v_{m+1}, \ldots, v_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy $v_i(u) = 0 \iff u = 0$. Define: $\tilde{f}_i(x) = v_i(f_i(x)),\ i = 0, \ldots, m$, $\tilde{h}_i(x) = v_{m+i}(h_i(x))), i = 1, \ldots, p$. We have the following equivalent formulation:

$$\min \tilde{f}_0(x)$$
$$\tilde{f}_i(x) \ge 0,\ i = 1, \ldots, m$$
$$\tilde{h}_i(x) = 0,\ i = 1, \ldots, m$$

An example of this transformation is least-norm-squared, where we square the least-norm value.

Slack Variables

Notice $f_i(x) \le 0 \iff s_i \ge 0$ s.t. $f_i(x) + s_i = 0$.

$$\min f_0(x)$$
$$s_i \ge 0,\ i = 1, \ldots, m$$
$$f_i(x) + s_i = 0,\ i = 1, \ldots, m$$
$$h_i(x) = 0,\ i = 1, \ldots, m$$

The optimal solution is equivalent since we know that $s_i = -f_i(x) \ge 0$.

Optimizing Over Some Variables

$\inf_{x,y} f(x, y) = \inf_x \tilde{f}(x)$ where $\tilde{f}(x) = \inf_y f(x, y)$. We can always minimize function by first minimizing over some of the variables, then minimizing over the remaining ones.

Epigraph Problem Form

$$\min t$$
$$t \ge f_0(x)$$
$$f_i(x) \ge 0,\ i = 1, \ldots, m$$
$$h_i(x) = 0,\ i = 1, \ldots, m$$

The epigraph form is an optimization problem minimizing t over the epigraph of f_0 subject to constraints on x.

Simple subproblems

- $p^* = \min_x a^T x = \begin{cases} -\infty & a \neq 0 \\ 0 & o.w. \end{cases}$
- $p^* = \min_x \frac{1}{2} x^T A x + b^T x$: Unconstrained convex quadratic has solution $p^* = -\frac{1}{2} b^T A^{-1} b$
- Scalara variable $x \in \mathbb{R}$: $p^* = \min_{x \ge c} ax = \begin{cases} -\infty & a < 0 \\ ac & o.w. \end{cases}$
- scalar variable, quadratic objective, linear constraint: $p^* = \min_{x \ge 0} \frac{1}{2} ax^2 - bx = \begin{cases} 0 & b/a < 0 \\ -\frac{b^2}{2a} & o.w. \end{cases}$
- vector variable $x \in \mathbb{R}^n$ with linear objective and constraint:
$$p^* = \min_x a^T x$$
$$b^T x \ge c$$
$$p^* = \begin{cases} c||a||_2 / ||b||_2 & a = \gamma b,\ for \gamma > 0 \\ -\infty & o.w. \end{cases}$$
- Rayleigh quotient: $\lambda_{max}(A) = \max_x x^T A x\ :\ ||x||_2 = 1$. The Rayleigh quotient problem is nonconvex even when $A \succeq 0$. (Because constraint region is the boundary of a hypersphere).
- Relaxed rayleigh quotient problem:
$$\max(0, \lambda max(A)) = \max_x x^T A x\ :\ ||x||_2 \le 1$$
If $\lambda_{max}(A) > 0$, then optimum value is $\lambda_{max}(A)$. Otherwise, when $\lambda_{max}(A) \le 0$, A is negative definite so for any $x \neq 0$, $x^T A x < 0$. Thus optimized when $x^* = 0$ so $x^{*T} A x^* = 0$.

Domain-splitting

$p^* = \min_{x \in X} f(x)$ for $X = Y \cup Z$ then
$p^* = \min\Big(\min_{x \in Y} f(x), \min_{x \in Z} f(x)\Big)$.

Reparameterization

For any vector $x \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n \times m}$, write $x = Az + v$ for some $z \in \mathbb{R}^m$ and $v \in \mathcal{N}(A^T)$.

$$p^* = \min_{x \in X} f(x) = \min_{z,v} f(Az + v)\ :\ Az + v \in X,\ A^T v = 0$$

Max-forall Trick

$$0 \ge \max_{x \in X} f(x) \iff \forall x \in X,\ 0 \ge f(x)$$

Absolute Value Trick

$$|x_i| \le s_i \iff x_i \le s_i,\ -x_i \le s_i[=]$$

Duality

Weak Duality

For every set $\mathcal{X}$, $\mathcal{Y}$

$$p^* = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) \ge \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x, y) = d^*$$

Strong Duality

Strong duality generally does not hold

$$p^* = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x, y) = d^*$$

Lagrange Dual Problem

$$p^* = \min_x f_0(x)$$
$$\text{s.t. } f_i(x) \le 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p$$
$$p^* = \min_x \max_{\lambda \succeq 0,\ v} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p v_j h_j(x)$$
$$p^* \ge d^* = \max_{\lambda \succeq 0,\ v} \min_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p v_j h_j(x)$$

Lagrange Dual Function:
$g(\lambda, v) = \min_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p v_j h_j(x)$
Dual Problem: $d^* = \max_{\lambda \succeq 0,\ v} g(\lambda, v)$ **For maximization primal problems, we have a minus sign in front of λ_i's when constructing the max-min problem**

Slater's Theorem

Given primal problem, if $\mathcal{D}$ is the domain of the problem, $\mathcal{D} = dom(f) \cap \bigcap_{i=1}^m dom(f_i) \cap \bigcap_{i=1}^n dom(h_i)$, $x \in \text{relint}\mathcal{D}$ such that $f_i(x) < 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, n$. Basically, the $f_i(x)$ constraint is met without equality, and then we have strong duality. A weaker form exists if first k inequality constraints are affine. Then, $f_i(x) < 0,\ i = k+1, \ldots, m$ for the inequality constraint for strong duality.

Sion's Theorem

If X is a compact and convex subset of a linear topological space and Y is a convex subset of linear topological space,

$$\forall x \in X, L(x, \cdot) \text{is (quasi)concave and upper semicontinuous}$$
$$\forall y \in y, L(\cdot, y) \text{is (quasi)convex and lower semicontinuous}$$

then $\min_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} L(x, y) = \sup_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} L(x, y)$. All concave functions are quasiconcave and all convex functions are quasiconvex. Compact sets contain limit points. Examples include $[a, b] \subset \mathbb{R}$ and $\{x : ||x||_p \le 1\}$. If Lagrangian concave in y, convex in x, continuous, then satisfies Sion's. If X, Y convex and one compact, sufficient.

KKT Conditions

Let x^* and (λ^*, v^*) be primal and dual optimal points with zero duality gap ($p^* = d^*$).

- $0 \ge f_i(x^*)\ i = 1, \ldots, m$
- $0 = h_i(x^*)\ i = 1, \ldots, n$
- $0 \preceq \lambda^*$
- $0 = \lambda_i^* f_i(x^*)\ i = 1, \ldots, m$
- $0 = \nabla_x f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^n v_i^* \nabla h_i(x^*)$

First three conditions require x^*, λ^* to be feasible. Last condition comes from solving unconstrained minimization and setting it to 0. Fourth condition is **complementary slackness**.

Duality Tricks

$$||x||_p = \max_z z^T x\ :\ ||z||_q \le 1$$

p-norm and q-norm dual norms for $1/p + 1/q = 1$ and $||z||_p = (\sum_i z_i^p)^{1/p}$.

$$\Big(\min_{z \in Z} z^T x\Big) \le b \iff \exists z \in \mathcal{Z}\ :\ z^T x \le b$$
$$\Rightarrow \min_{x \in \mathcal{X}} f(x)\ :\ \Big(\min_{z \in Z} z^T x\Big) \le b = \min_{x \in \mathcal{X}} f(x)\ :\ z^T x \le b$$