Chapter 1: Introduction

1.2 Optimization Models

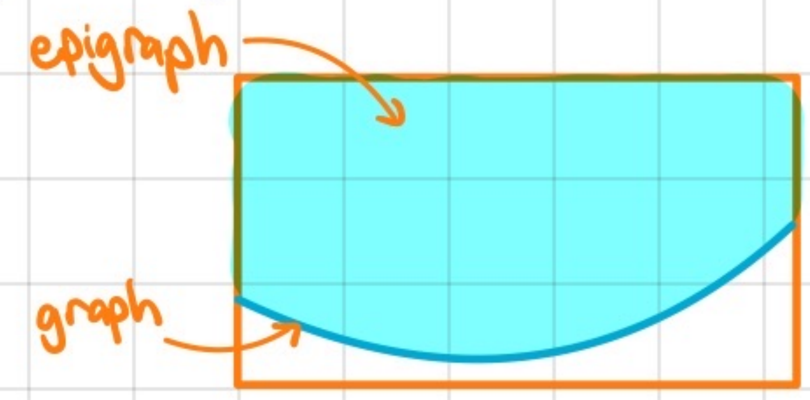
Functions: $f: \mathbb{R}^n \to \mathbb{R}$ take arg $\mathbb{R}^n$ return value in $\mathbb{R}$

Domain: dom f set of points where the function is finite

Map: $f: \mathbb{R}^n \to \mathbb{R}^m$ vector valued functions, return more than one value

Graph: $G(f) := \{(x, f(x)) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n\}$ set of input/output pairs f can have

Epigraph: $\text{epi } f := \{(x,t) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n, t \geq f(x)\}$ input/ouput pairs f can achieve, or anything above



Level Set: $L_t(f) := \{x \in \mathbb{R}^n : x \in \mathbb{R}^n, t = f(x)\}$ set of points achieving exactly some value for function f

Sub-level Set: $S_t(f) := \{x \in \mathbb{R}^n : x \in \mathbb{R}^n, t \geq f(x)\}$ set of points achieving at most a certain value for f, or below

Functional Form: $p^* := \min_x f_0(x)$ subject to $f_i(x) \leq 0, \; i=1, \dots m$

Epigraph Form: $p^* := \min_{x,t} t$ subject to $f_0(x) - t \leq 0, f_i(x) \leq 0, \; i=1 \dots m$

$x \in \mathbb{R}^n$: decision var
$f_0 : \mathbb{R}^n \to \mathbb{R}$: objective func, cost
$f_i : \mathbb{R}^n \to \mathbb{R}, i=1..m$: constraints
p^*: optimal value

Feasible Set: $X := \{x \in \mathbb{R}^n : f_i(x) \leq 0, i=1 \dots m\}$ satisfies the constraints

Optimal Value: p^* value of objective at optimum

Optimal Set: $X^{opt} := \arg\min_{x \in X} f_0(x)$ set of feasible points where obj func is optimal

Chapter 2: Linear Algebra

2.1 Vectors

Basics

Vector: $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\mathbb{R}^n$ real components single point in n-dimensional space

Independence: set of vectors $\{x_1 \dots x_m\}$ in $\mathbb{R}^n$ $i=1,\dots m$ independent iff $\sum_{i=1}^m \lambda_i x_i = 0$ $\lambda \in \mathbb{R}^m$ $\lambda = 0$, no vector in set can be expressed as linear combination of others

Subspace: subset closed under addition and scalar multiplication, flat, through origin

Subspace S can be represented as span of set of vectors $x_i \in \mathbb{R}^n$, $i=1,\dots m$ $\quad S = \text{span}(x_1 \dots x_m) := \left\{ \sum_{i=1}^{m} \lambda_i x_i : \lambda \in \mathbb{R}^m \right\}$

Basis: set of n independent vectors. If vectors $u_1, \dots, u_n$ form basis, can express any vector as linear combination of u_i's: $x = \sum_{i=1}^{n} \lambda_i u_i$

standard basis: $e_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $e_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $e_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Basis of Subspace: basis of subspace $S \subseteq \mathbb{R}^n$ is any independent set of vectors whose span is S

Dimension: choice of the basis is independent of # vectors in basis

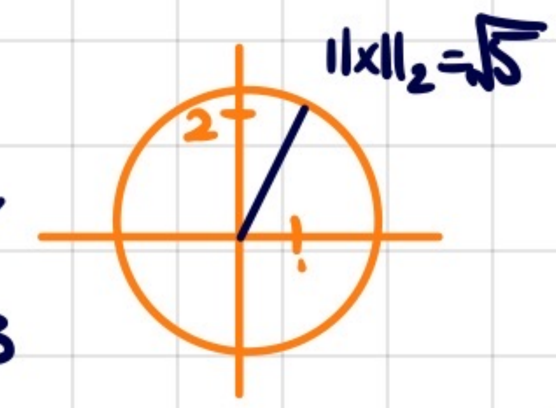
Scalar Product, Norms

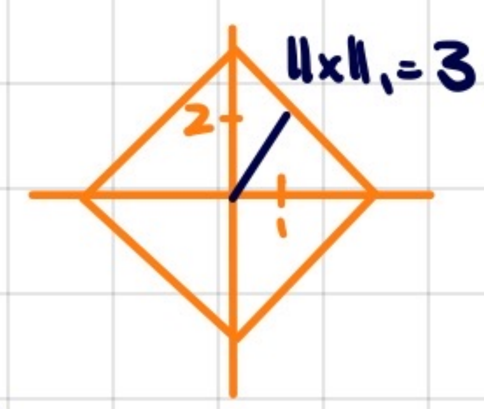
Scalar Product: $x^T y = \sum_{i=1}^{n} x_i y_i$ also $\langle x, y \rangle$

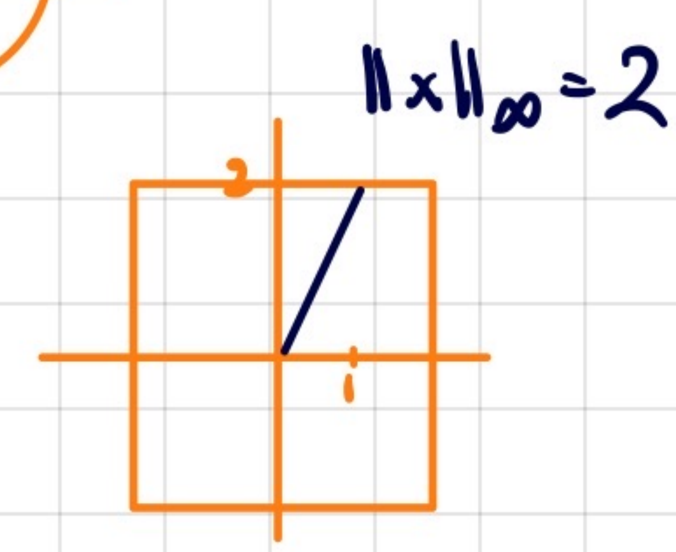
Orthogonality: $x, y \in \mathbb{R}^n$ orthogonal if $x^T y = 0$

3 Popular Norms

1. L2 Norm (Euclidean): $\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$ hypersphere
2. L1 Norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
3. L∞ Norm: $\|x\|_\infty := \max_{1 \le i \le n} |x_i|$ useful for peak values







Cauchy-Schwartz Inequality: For any two vectors $x, y \in \mathbb{R}^n$,

$x^T y \le \|x\|_2 \cdot \|y\|_2$ equality iff x, y are colinear

$\max_{x : \|x\|_2 \le 1} x^T y = \|y\|_2$ optimal $x^* = \frac{y}{\|y\|_2}$ if y non-zero

Angle between vectors $\cos\theta = \frac{x^T y}{\|x\|_2 \|y\|_2}$

Projection on a Line

Projection: for line in $\mathbb{R}^n$ passing through $x_0 \in \mathbb{R}^n$, direction

$u \in \mathbb{R}^n$ $\{x_0 + tu : t \in \mathbb{R}\}$, projection of point x on the line is vector z on the line closest to x

$$\min_t \|x - x_0 - tu\|_2 \qquad t^* = u^T(x - x_0)$$

projected vector: $z^* = x_0 + \frac{u^T(x-x_0)}{u^T u} u$

Orthogonalization

Orthonormal: orthogonal $u_i^T u_j = 0$, $i \neq j$ and normal $\|u_i\|_2 = 1$

orthogonalization: procedure to find orthonormal basis of span of given vectors

vectors $a_1, \dots a_k \in \mathbb{R}^n$, compute vectors $q_1, \dots q_n \in \mathbb{R}^n$ st

$S := \text{span}\{a_1 \dots a_k\} = \text{span}\{q_1 \dots q_r\}$ $\quad r$: dimension S

$q_i^T q_j = 0$ $(i \neq j)$, $q_i^T q_i = 1$, $1 \leq i, j \leq r$

Gram-Schmidt Procedure

orthogonalization algo, orthogonalize each vector wrt previous ones, normalize result

Vectors Independent $(a_1 \dots a_n)$

1) $\tilde{q}_1 = a_1$
2) normalize: set $q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|_2}$
3) remove component of q_1 in a_2: $\tilde{q}_2 = a_2 - (a_2^T q_1) q_1$
4) normalize: $q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$
5) remove components of $q_1 q_2$ in a_3: $\tilde{q}_3 = a_3 - (a_3^T q_1) q_1 - (a_3^T q_2) q_2$

⋮

Vectors Dependent

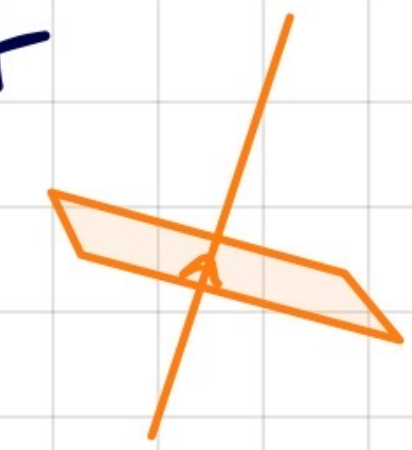
1) set $r = 0$
2) for $i = 1 \dots n$:

   set $\tilde{q} = a_i - \sum_{j=1}^{r} (q_j^T a_i) q_j$

   if $\tilde{q} \neq 0$, $r = r + 1$; $q_r = \frac{\tilde{q}}{\|\tilde{q}\|_2}$

Hyperplanes

Hyperplane: $H = \{x : a^T x = b\}$ set described by scalar product



$b=0$: hyperplane orthogonal to a

$b \neq 0$: hyperplane is translation along direction a, of set

Half space: $H = \{x : a^T x \geq b\}$ set of points $a^T(x - x_0) \geq 0$

Linear Functions

Linear Functions: Preserve scaling and addition of input

for every $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, $f(\alpha x) = \alpha f(x)$ &

for every $x_1, x_2 \in \mathbb{R}^n$, $f(x_1 + x_2) = f(x_1) + f(x_2)$

affine iff function $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ $\tilde{f}(x) = f(x) - f(0)$ is linear

can be expressed w/ scalar product $f(x) = a^T x + b$

affine func gradient: $\nabla f(x) = a$ for $f(x) = a^T x + b$

First Order expansion of function:

first-order approx of differentiable function f at point x_0

form $f(x) \simeq l(x) = f(x_0) + \nabla f(x_0)^T (x - x_0)$

2.2 Matrices

Basics

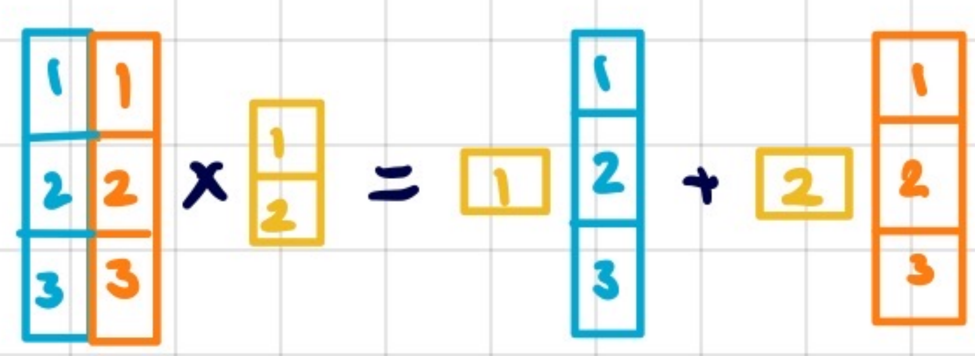
Matrix: collection of n column vectors in $\mathbb{R}^m$, n points m dimensional space

$A = (a_1 \dots a_n)$

Transpose: A^T matrix w/ (i,j) element A_{ji}

Matrix Products

Matrix-vector Product: $Ax = \sum_{i=1}^{n} x_i a_i$



Matrix-Matrix Product: $(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$ $\quad (AB)^T = B^T A^T$

Blocks: $AB = (A_1 \; A_2) \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = A_1 B_1 + A_2 B_2$ $\quad BA = \begin{bmatrix} B_1 A_1 & B_1 A_2 \\ B_2 A_1 & B_2 A_2 \end{bmatrix}$

Trace: $\mathrm{Tr}\,A = \sum_{i=1}^{n} A_{ii}$ trace of square matrix A is sum of diagonal elements

$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$

Special Matrices

Identity Matrix: I ones on diagonals, zero elsewhere

Diagonal Matrices: $\mathrm{diag}(a)$: $A_{ij} = 0$ when $i \neq j$

$$\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$$

Symmetric Matrices: $A_{ij} = A_{ji}$ for every pair (i,j)

Upper Triangular: $A \in M^{m \times n}$, $A_{ij} = 0$ when $i < j$

Orthogonal Matrices: $U^T U = I_n$ colums form orthonormal basis

cooresspond to rotations around point, reflections across line

Dyad: $Ax = (uv^T)x = (v^T x)u$ rank 1 matrices, output always points in same direction u in output space

$(A^T)^T = A$

$(A+B)^T = A^T + B^T$

$(AB)^T = B^T A^T$

QR

QR: factor matrix as product of two matrices

A full column rank

1) $a_i = (a_i^T q_1)q_1 + \ldots + (a_i^T q_{i-1})q_{i-1} + \|\tilde{q}_i\|_2 q_i$, $i = 1, \ldots n$

2) $A = QR$, $Q = (q_1 \cdots q_n)$, $R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & r_{nn} \end{pmatrix}$

Matrix Inverses

Matrix Inverses: invertible only if columns are independent

Full Column Rank: $m \times n$ matrix if columns are independent, $m \geq n$

Full Row Rank: $m \times n$ matrix if rows are independent

Linear Maps

Linear map: map linear iff every one of components are linear

can be expressed as $f(x) = Ax$

Matrix Norm

Frobenius Norm: $\|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{Tr(A^T A)}$ avg of squared error norm

Peak Gain: $\|A\|_{LSV} := \max_v \|Av\|_2 : \|v\|_2 \leq 1$

2.3 Linear Equations

Existence, Unicity

Linear Equation: $Ax = y$

$A : \in \mathbb{R}^{m \times n}$
$x : \in \mathbb{R}^n$
$y : \in \mathbb{R}^m$

Solutions in affine subspace
- determine if sol exists
- if unique
- compute solution x_0
- find orthonormal basis

Range: $R(A) := \{Ax : x \in \mathbb{R}^n\}$ span of columns of A

describes vectors $y = Ax$ can be attained in output space by arbitrary choice of a vector x in the input space

Rank: dimension of the range, rank cannot exceed dimensions $A: r \leq \min(m,n)$
full rank if $r = \min(m,n)$

Nullspace: $N(A) := \{x \in \mathbb{R}^n : Ax = 0\}$ $A \in \mathbb{R}^{m \times n}$

Nullity: dimension of the nullspace

Rank - Nullity Theorem
nullity and rank of $m \times n$ matrix A add up to column dimension of A, n
nullity = $n - r$

Fundamental Theorem of Linear Algebra
Range of a matrix is the orthogonal complement of the nullspace of its transpose

$$R(A)^{\perp} = N(A^T)$$

Solving Linear Eq via QR decomposition

When solving linear equations, $Ax = y$, once you have upper triangular, you can back substitute

QR Decomposition: express $m \times n$ matrix A as $A = QR$

Q: $m \times m$, orthogonal $Q^T Q = I_m$
R: upper triangular

Solving linear Equation:

1) $A = QR$ QR decomposition

2) $QRx = y \Leftrightarrow Rx = Q^T y$ pre multiply Q with both sides

QR decomposition process

1) $AP = QR = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix}$

Q: $[Q_1\ Q_2]$ $m \times m$ orthogonal ($Q^T Q = I_m$)

Q_1: $m \times r$ orthonormal columns ($Q^T Q = I_r$)

Q_2: $m \times (m-r)$, orthonormal columns ($Q^T Q = I_{m-r}$)

r: rank of A

R_1: $r \times r$ upper triangular, invertible

R_2: $r \times (n-r)$ matrix

P: $m \times m$ permutation mat ($P^T = P^{-1}$)

0: zero submatrix $m-r$ rows

2) $A = QRP^T$ $Rz = Q^T y$ $z := P^T x$

$$\begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} Q_1^T y \\ Q_2^T y \end{pmatrix}$$

Assume $Q_2^T y = 0$,

3) $R_1 z_1 + R_2 z_2 = Q_1^T y$ set $z_2 = 0$, invertible triangular matrix R_1

$$x_0 := P \begin{pmatrix} R_1^{-1} Q_1^T y \\ 0 \end{pmatrix}$$

Set of all solutions

$$x = P \begin{pmatrix} R_1^{-1} Q_1^T (y - R_2 z_2) \\ 0 \end{pmatrix} = x_0 + L z_2 \qquad L := -P \begin{pmatrix} R_1^{-1} Q_1^T R_2 \\ 0 \end{pmatrix}$$

2.4 Ordinary Least Squares (OLS)

Ordinary Least Squares

Ordinary Least Squares (OLS/LS): minimizes Euclidean norm of residual error

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

$A \in \mathbb{R}^{m \times n}$ design, input

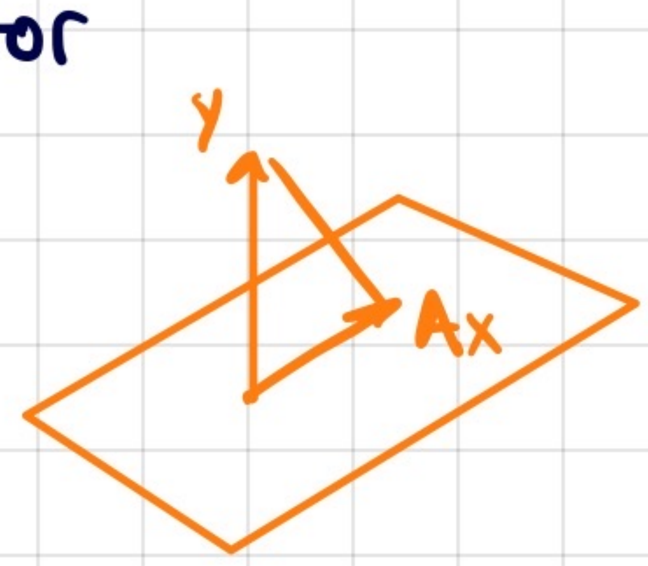
$y \in \mathbb{R}^m$ measurement, output

(A, y) problem data

$r := y - Ax$ residual error vector

best approx of y in linear combination of columns of A

project vector y on the span of the vectors q_i's



Solving by QR Decomposition

$x^* = (A^TA)^{-1}A^Ty$ if $A \in m\times n$ tall $(m \geq n)$ full column rank

Variants

Linearly Constrained Least Squares

$$\min_x \|Ax - y\|_2^2 : Cx = d \quad C \in \mathbb{R}^{p\times n} \quad d \in \mathbb{R}^p$$

Solution

$$\min_z \|\tilde{A}z - \tilde{y}\|_2 \quad \tilde{A} := AN \quad \tilde{y} = y - Ax_0$$

Kernels

$$\min_w \|X^T w - y\|_2^2 + \lambda\|w\|_2^2$$

w in $\text{span}\{x_1 \ldots x_m\}$ $\quad w = Xv$

$$\Downarrow$$

$$\min_v \|Kv - y\|_2^2 + \lambda v^T K v$$

depends only on $K := X^TX$

when $n \gg m$, dramatic reduction in problem size

Gaussian Kernals: $K(x,z) = \exp\left(-\frac{\|x-z\|_2^2}{2\sigma^2}\right)$

2.5 Eigenvalues

Symmetric Matrix Definitions

Quadratic function: $q: \mathbb{R}^n \to \mathbb{R}$

$$q(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j + 2\sum_{i=1}^{n} b_ix_i + c$$

$c, i, j \in \{1, \ldots, n\}$

$$q(x) = x^TAx + 2b^Tx + c$$

2nd order approx

$$f(x) \approx q(x) = f(x_0) + f(x_0)'(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

Spectral Theorem

Eigenvalue: λ $A \in \mathbb{R}^{n\times n}$ symmetric Matrix, λ eigenvalue if

$Au = \lambda u$ nonzero vector $u \in \mathbb{R}^n$

Eigenvector: u eigenvector normalized if $\|u_2\| = 1$

$\det(\lambda I - A) = 0$

Spectral Theorem (Symmetric eigenvalue decomposition (SED))

for any symmetric matrix, exactly n real eigenvalues, eigenvectors form orthonormal basis

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U \Lambda U^T \quad , \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

$U := [u_1 \ldots u_n]$ orthogonal
$U^T U = U U^T = I_n$
eigenvectors

Rayleigh Quotients: symmetrix matrix A, smallest, largest eigenvalue as variational form:

$$\lambda_{min}(A) = \min_x \{ x^T A x : x^T x = 1 \}$$

$$\lambda_{max}(A) = \max_x \{ x^T A x : x^T x = 1 \}$$

or $\dfrac{x^T A x}{x^T x}$

Positive-Semi Definite (PSD) Matrices

Quadratic form: $q(x) = x^T A x$ $\qquad q: \mathbb{R}^n \to \mathbb{R}$

Positive Semi Definite (PSD): $A \succeq 0$ iff associated quadratic form q non-negative everywhere

$q(x) \geq 0 \;\; \forall \; x \in \mathbb{R}^n$

Positive Definite (PD): quadratic form is non negative, definite $q(x) = 0$ iff $x = 0$

Eigenvalues of PSD Matrices

$q(x) = x^T A x$, $A \in S^n$ is non-negative iff every eigenvalue of symmetric matrix A is non-negative

Principal Component Analysis (PCA)

Variance Maximization Problem: Find direction $u \in \mathbb{R}^n$ st sample variance of coorresponding vector $u^T X = (u^T x_1, \ldots u^T x_n)$ maximal

$$\max_{u: \|u\|_2 = 1} u^T \Sigma u \qquad \Sigma := \frac{1}{n} \sum_{j=1}^{n} (x_j - \hat{x})(x_j - \hat{x})^T \quad \text{(sample covariance matrix)}$$

Principal Component Analysis (PCA):

1) Find direction for maximal variance between the data points
2) Project data on hyperplane orthogonal of that direction
3) Using new dataset, find new direction

4) Repeat until enough directions

directions are eigen vectors of data's covariance matrix, or finding eigenvalue decomposition of PSD (covariance matrix)

Total Variance: $\mathrm{Tr}\,\Sigma = \mathrm{Tr}(U\Lambda U^T) = \mathrm{Tr}(U^T U \Lambda) = \mathrm{Tr}\,\Lambda = \lambda_1 + \dots + \lambda_n$

Ratio of variance explained: $\dfrac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_n}$

2.6 Singular Values

Singular Value Decomposition (SVD) Theorem

complexity: $O(nm\min(n,m))$

SVD generalizes spectral theorem. Three steps: 1) rotation in the input space 2) positive scaling so that input space → output space 3) rotation in output space

Decomposition of form

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = U\tilde{S}V^T, \quad \tilde{S} := \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$$

$r \le \min(m,n)$
$U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ orthogonal matrices
$S := \mathrm{diag}(\sigma_1 \dots \sigma_r)$

singular values (σ_i): unique

left singular vectors of A: first r columns of U: u_i

$$Av_i = \sigma_i u_i, \quad u_i^T A = \sigma_i v_i \qquad i = 1..r$$

To get Ax

1) form $\tilde{x} := V^T x \in \mathbb{R}^n$ — $\tilde{x}$ is rotated x
2) first r elements scaled by $\sigma_1, \dots \sigma_r$ — $\tilde{y}$ now in output space $\mathbb{R}^m$
3) Rotate vector $\tilde{y}$ by orthogonal matrix U — $y = U\tilde{y} = Ax$

SED

$$AA^T = U\Lambda_m U^T, \quad A^T A = V\Lambda_n V^T$$

Matrix Properties via SVD

Nullspace via SVD: null space of matrix A w/ SVD

$$A = U\tilde{S}V^T, \quad \tilde{S} := \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}, \quad S = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$$

admits last $n-r$ columns as orthonormal basis

Full column rank: $A = U\begin{pmatrix} S \\ 0 \end{pmatrix}V^T$

Range via SVD: range of matrix A w/ SVD

$$A = U\tilde{S}V^T, \quad \tilde{S} = \mathrm{diag}(\sigma_1, \dots \sigma_r, 0 \dots 0)$$

admits the first r columns of U as an orthonormal basis

Fundmental Theorem of Linear Algebra

Let $A \in \mathbb{R}^{m \times n}$ $N(A)$ and $R(A^T)$ form orthonormal decomposition of $\mathbb{R}^n$

x can be written as $x = y + z, \; y \in N(A), \; z \in R(A^T), \; y^T z = 0$

condition on vector x to be orthagonal to any vector in nullspace means in range: $x^T y = 0$ whenever $Ay = 0 \Leftrightarrow \exists \lambda \in \mathbb{R}^m : x = A^T \lambda$

Frobenius Norm via SVD:

$$\|A\|_F = Tr(V\tilde{S}^T\tilde{S}V^T) = \sum_{i=1}^{r} \sigma_i^2$$

Largest Singular Value Norm:

$$\|A\|_{LSV} := \max_{x : \|x\|_2 < 1} \|Ax\|_2 = \sigma_1(A)$$

$\sigma_1(A)$: largest singular value of A

condition number: ratio between largest and smallest singular values

$$K(A) = \frac{\sigma_1}{\sigma_n} = \|A\| \, \|A^{-1}\|$$

Solving Linear Equations via SVD

Solution set

$$Ax = y$$

1) $A = U\tilde{S}V^T, \quad U^T U \tilde{S} V^T x = U^T y$

$$\tilde{S}\hat{x} = \hat{y} \qquad \hat{x} = V^T x$$
$$\hat{y} = U^T y$$

2) $\sigma_i \bar{x}_i = \bar{y}_i \quad , i=1..r \qquad 0 = \check{y}_i \; , \quad i = r+1, \dots m$

a) Infeasible if last $m-r$ components $\check{y}$ not zero

b) If y in range of A

$$\hat{x}_i = \frac{\hat{y}_i}{\sigma_i} \; , i=1\dots r$$

Pseudo-inverse A^+:

$$A^+ := V\tilde{S}^+ U^T \, , \; \tilde{S} = \mathrm{diag}(\sigma_1^{-1}, \dots \sigma_r^{-1}, 0, \dots 0)$$

Solution Set: $S = \{A^+ y + z : z \in N(A)\}$

Full col rank: $A^+ = (A^TA)^{-1}A^T$

Full row rank: $A^+ = A^T(AA^T)^{-1}$

Least Squares and SVD

Optimal Set of Ordinary least squares

Optimal Set of the OLS problem

$$p^* := \min_x \|Ax - y\|_2$$ expressed as $$x^{opt} = \underbrace{A^+ y}_{\text{min norm point}} + N(A)$$

$$x^* = A^+ y = (A^TA)^{-1}A^T y$$

Sensitivity Analysis: $\delta y \in \mathbb{R}^m$: measurement noise

Given $y + \delta y = Ax$

want $\min_{x, \delta y} \|\delta y\|_2 : y + \delta y = Ax \qquad \min_x \|Ax - y\|_2$

$$A^+ = (A^TA)^{-1}A^T = V(\Sigma^{-1} \ 0)U^T$$

Set possible errors: $E = \{A^+\delta y : \|\delta y\|_2 \le 1\}$

Best Linear Unbiased Estimator (BLUE):

For family of linear estimators:

$\hat{x} = By$ $\qquad B \in \mathbb{R}^{n \times m}$

Errors on solution δx

$E = \{B\delta y : \|\delta\|_2 \le 1\}$ $\qquad BB^T \succeq A^+(A^+)^T$

Low Rank Approximation of a Matrix

best k-rank approximation $\hat{A}_K$ is given by zeroing out $r-k$ trailing singular values of A

$$\hat{A}_K = U\hat{S}_K V^T, \quad \hat{S}_K = \mathrm{diag}(\sigma_1, \dots \sigma_K, 0, \dots 0)$$

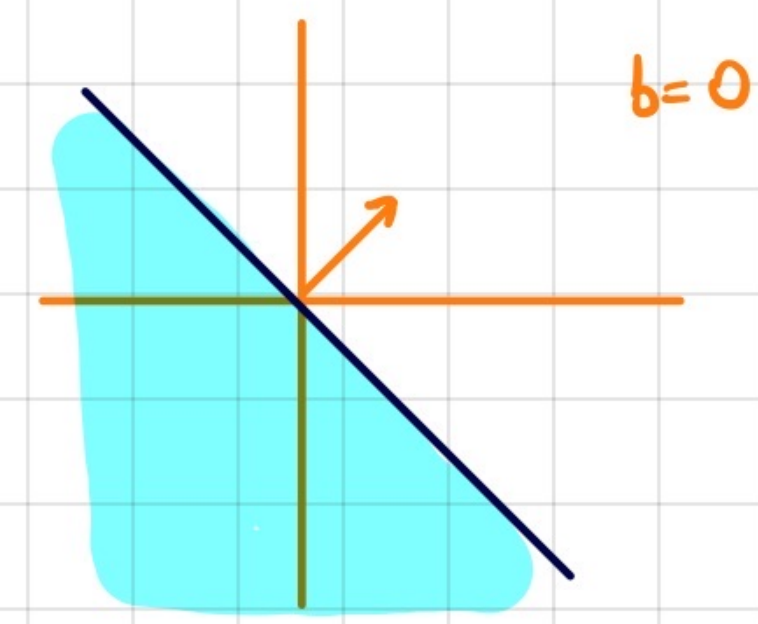
minimal error: $\|A - \hat{A}_K\|_F = \sqrt{\sigma_{K+1}^2 + \dots + \sigma_r^2}$

Explained variance: $\dfrac{\|\hat{A}_K\|_F^2}{\|A\|_F^2} = \dfrac{\sigma_1^2 + \dots + \sigma_K^2}{\sigma_1^2 + \dots + \sigma_n^2}$

Chapter 3 : Convex Models

3.2 Linear & Quadratic Programming

Polyhedra



Half Spaces: set defined by single affine inequality

$$H = \{x : a^T x \leq b\} \qquad a \in \mathbb{R}^n \quad b \in \mathbb{R} \qquad Ax \leq b \quad A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$$

separates the whole space into two halves

Polyhedra: set described finitely many affine inequalities

$$P = \{x : a_i^T x \leq b_i \ , i = 1 \dots m\} \qquad a_i \in \mathbb{R}^n \quad b_i \in \mathbb{R},\ i=1,\dots m$$

Polyhedron: intersection of finitely many half-spaces

$$P = \bigcap_{i=1}^{m} H_i, \quad H_i := \{x : a_i^T x \leq b_i\} \quad i = 1, \dots m$$

Standard Forms

Linear Programs: optimization problem w/ linear objective and affine inequality constraints.

$$\min_x \ f_0(x) : f_i(x) \leq 0 \ , i=1 \dots m$$

Quadratic Program: optimization problem where
1) constraint functions are affine 2) objective function f_0 quadratic convex

$$f_0(x) = c^T x + x^T Q x \ : \qquad \text{vector } c \in \mathbb{R}^n \quad Q = Q^T \succeq 0$$

Standard Inequality form: function $f : \mathbb{R}^n \to \mathbb{R}$ affine iff

$$f(x) = a^T x - b$$

linear program: $\min_x \ c^T x : Ax \leq b \qquad A := \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \in \mathbb{R}^{m \times n} \qquad b := \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^n$

QP: $\min_x \ c^T x + x^T Q x : Ax \leq b$

Polyhedral Functions

Polyhedral: $f: \mathbb{R}^n \to \mathbb{R}$ if epigraph is a polyhedron

$$\text{epi } f = \left\{ (x,t) \in \mathbb{R}^{n+1} : C\begin{pmatrix} x \\ t \end{pmatrix} \leq d \right\}$$

$C \in \mathbb{R}^{m \times (n+1)}$
$d \in \mathbb{R}^m$

- Polyhedral functions include functions that can be expressed as max affine func

$$\text{epi } f = \left\{ (x,t) \in \mathbb{R}^{n+1} : t \geq \max_{1 \leq i \leq m} (a_i^T x + b_i) \right\}$$

- epi f is projection of a polyhedron on (x,t) space, itself is a polyhedron

Minimizing Polyhedral Functions as LP

$$\min_{x,t} t : (x,t) \in \text{epi } f, \ Ax \leq b \quad \text{or} \quad \min_{x,t} t : Cx \leq d, \ t \geq a_i^T x + b_i, \ i = 1 \ldots m$$

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

Minimizing Sum of Maxima of affine functions

$$\min_{x,t} \sum_{j=1}^{p} t_j : t_j \geq a_{ij}^T x + b_{ij}, \ Cx \leq d \qquad i = 1 \ldots m, \ j = 1 \ldots p$$

Min Cardinality

$$\min_x \text{Card}(x) : x \in P$$

P: polyhedron
Card: cardinality (# non-zero elements)

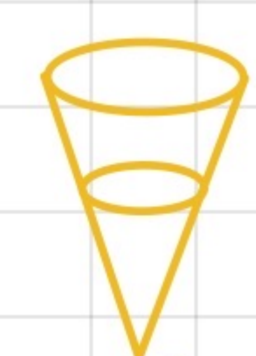
l_1-norm heuristic:

$$\min_x \|x\|_1 : x \in P$$

P: polyhedron $\{x : Ax \leq b\}$

3.3 Second-Order Cone Programming

Second Order Cone Inequalities



Second-order cone: set of vectors (x_1, x_2, y) $\quad y \geq \sqrt{x_1^2 + x_2^2}$

Second-order cone programming: generalization of LP, QP that contrains inside second-order cones

Rotated Second-order cone: $K_p = \left\{ (x,y,z) \in \mathbb{R}^{p+2} : x^T x \leq yz, \ y \geq 0, \ z \geq 0 \right\}$

Second-order cone inequality: $\|Ax + b\|_2 \leq c^T x + d$

$A \in \mathbb{R}^{m \times n}$
$b \in \mathbb{R}^m$
$c \in \mathbb{R}^n$

Standard Forms

Inequality form: $\min_x \ c^T x : \|A_i x + b_i\|_2 \le c_i^T x + d \ , i=1,\dots m$

Conic form: $\min_x c^T x : (A_i x + b_i, c_i^T x + d_i) \in K_p \ , i=1\dots m$

Quadratically constrained quadratic programming:

$$\min_x \ a_0^T x + x^T Q_0 x \ : \ x^T Q_i x + a_i^T x \le b_i \ , \ i=1\dots m$$

Group Sparsity

l_1 norm tricks to address group sparsity problem

$$\min_x \sum_{i=1}^k \|x_i\|_2 : x \in P \quad \text{or} \quad \min_x \|Ax-b\|_2 + \sum_{i=1}^k \|x_i\|_2$$

3.4 Robust Linear Programming

Motivations and Standard forms

Assume data is not known but model of uncertainty is

robust counterpart: $\min_x \ c^T x : \forall a_i \in U_i, \ a_i^T x \le b_i \ , i=1\dots m$ find feasible, independent of a_i solution

Tractable Cases

Uncertainty Model: uncertainty on a coefficient vector a described by finite set pts

$\min_x \ c^T x : (a_i^k)^T x \le b \ , k=1\dots K_i, \ i=1\dots m$ $\quad U = \{a^1, \dots a^k\}$

Box Uncertainty Model: assume every point lies in hyperrectangle $U = \{a : \|a - \hat{a}\|_\infty \le p\}$

$$\min_x \ c^T x : \hat{a}_i^T x + p\|x\|_1 \le b_i \ , i=1,\dots m$$

Ellipsoidal uncertainty Model: $R \in \mathbb{R}^{n \times p}$ describes shape of ellipsoid around center

$$\min_x \ c^T x : \hat{a}_i^T x + \|R_i^T x\|_2 \le b_i \ , i=1\dots m$$

Midterm

3.1 Convex Optimization

Convex Sets

Convex: subset $C \in \mathbb{R}^n$ iff contains line segment between any two points

$$\forall x_1, x_2 \in C, \ \forall \lambda \in [0,1] : \lambda x_1 + (1-\lambda) x_2 \in C$$

Intersection: intersection of family of convex sets is convex

Affine Transformation: map $f: \mathbb{R}^n \to \mathbb{R}^m$ is affine, C is convex set $f(C) := \{f(x) : x \in C\}$

projection of a convex set on a subspace is convex

Convex Functions

Domain: $f: \mathbb{R}^n \to \mathbb{R}$ set $\text{dom } f \subseteq \mathbb{R}^n$ where f is defined

$$\text{dom } f := \{x \in \mathbb{R}^n : -\infty < f(x) < +\infty\}$$

Convex: $f: \mathbb{R}^n \to \mathbb{R}$ convex when 1) dom f is convex

2) $\forall x_1, x_2 \in \text{dom} f$, $\forall \theta \in [0,1]$, $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$

or prove through

2) Epigraph: if epigraph $\text{epi } f := \{(x,t) \in \mathbb{R}^{n+1} : t \geq f(x)\}$ is convex

3) 1st order cond: if f is differentiable, convex iff $\forall x,y : f(y) \geq f(x) + \nabla f(x)^T (y-x)$

4) Restriction to line: convex iff restriction to line is convex, every $x_0 \in \mathbb{R}^n$, $v \in \mathbb{R}^n$
$g(t) := f(x_0 + tv)$ convex

5) Second-order condition: f twice differentiable, convex iff Hessian $\nabla^2 f$ PSD on domain

<u>Operations that conserve Convexity</u>

- affine preserves: $g(x) = f(Ax+b)$ function is convex

Pointwise max: if $(f_\alpha)_{\alpha \in A}$ $f(x) := \max_{\alpha \in A} f_\alpha(x)$ is convex, largest singular value, nonnegative weighted sum, partial minumum, monotone convex function

Convex Optimization Functions

Standard form: $\min_x f_0(x)$: $f_i(x) \leq 0$, $i = 1 \dots m$
$h_i(x) = 0$, $i = 1 \dots p$

convex optimization if:
objective function f_0 convex
function defining inequality f_i are convex
functions defining equality h_i are affine

if unconstrained, optimality $\nabla f_0(x) = 0$

Property: max over any set is max over convex hull of set

$$\max_{x \in S} f(x) = \max_{x \in \text{Co}S} f(x)$$

Algorithms for Convex Optimization

Unconstrained minimization: Newton's method

$$x_{t+1} = x_t - (\nabla^2 f(x_t))^{-1} \nabla f(x_t)$$

Constrained minimization: interior-point methods

applied to function $\min_x f_0(x)$ subject to $f_i(x) \le 0$, $i=1 \dots m$

- replace constraind problem w/ unconstrained problem

$$\min_x f(x) := f_0(x) - \mu \sum_{i=1}^{m} \log(-f_i(x))$$

Gradient Methods

Unconstrained: $x_{t+1} = x_t - \alpha_t \nabla f(x_t)$ Constrained: $x_{t+1} = P(x_t - \alpha_t \nabla f(x)_t)$

Chapter 4: Duality

4.1 Weak Duality

Dual Problem

Dual Problem: concave max problem providing lower bound on value of original problem

Weak Duality: form bounds on non-convex problems using convex optimization

Strong Duality: If "dual" problem same as "primal", strong duality holds

Primal Problem: $p^* := \min_x f_0(x) : f_i(x) \le 0$, $i=1 \dots m$

Primal value: optimal value to primal problem

Lagrange function: $L: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ depends on primal variables and additional variable $y \in \mathbb{R}^m$: dual variable

$$L(x,y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) = f_0(x) + y^T f(x)$$

Dual Function: For $\forall x$ feasible : $f_0(x) \geq L(x,y)$

dual func: $g: \mathbb{R}^m \to \mathbb{R}$ $g(y) := \min_z L(z,y) = \min_z f_0(z) + \sum_{i=1}^m y_i f_i(z)$

then $L(x,y) \geq g(y)$ $\rightarrow$ $f_0(x) \geq g(y)$

Dual Problem: $p^* \geq d^* := \max_{y \geq 0} g(y)$ d^*: dual optimal value

Minimax Inequality

Minimax Inequality: for any function two var $L: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$

$$\min_{x \in X} \max_{y \in Y} L(x,y) \geq \max_{y \in Y} \min_{x \in X} L(x,y)$$

4.2 Strong Duality

Slater Condition

Convex Constrained Optimization Problem in standard form:

$$p^* := \min_x f_0(x): Ax = b,\ f_i(x) \leq 0,\ i = 1 \dots m$$

Lagrangian: $L(x,\lambda,v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + v^T(Ax-b)$

Dual function: $g(\lambda,v) = \min_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + v^T(Ax-b)$

$$d^* = \max_{\lambda \geq 0, v} g(\lambda,v)$$

Strong Duality: holds if $p^* = d^*$, weak $p^* \geq d^*$

Slater's Sufficient condition for strong duality: if
- primal problem is convex
- strictly feasible, $x_0 \in \mathbb{R}^n$ st $Ax_0 = b,\ f_i(x_0) < 0,\ i = 1 \dots m$

then strong duality holds

Optimality Conditions

Primal Optimum Attainment

When dual problem feasible $\exists \lambda > 0, v: g(\lambda,v) > -\infty$

Strong duality holds exists x, λ, v st
- x is feasible for primal
- λ, v are feasible for dual problem, $\lambda \geq 0$, $(\lambda,v) \in \text{dom } g$

Optimality Conditions

Lagrangian Stationarity: $0 = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} v_i \nabla h_i(x)$

$\boxed{\text{Sensitivity}}$

Nominal Problem: $p^*(u,v) := \min_x f_0(x) : Ax = b + v \ , \ f_i(x) \le u_i$

Dual: $d^* = \max_{\lambda \ge 0, v} g(\lambda, v)$

Perturbed: $p^*(u,v) := \min_x f_0(x) : Ax = b + v, \ f_i(x) \le u_i$

Dual of Perturbed: $d^* = \max_{\lambda \ge 0, v} g(\lambda, v) - \lambda^T u - v^T v$

weak duality applied $\quad p^*(u,v) \ge g(\lambda^*, v^*) - u^T \lambda^* - v^T v^*$