DSA3102 Lecture 10 (Reading Sections 7.1 & 7.5)

- Convex Sets (2) / Functions (2) / Optimization Problems (2)
- Duality (3) (Lagrangian, Dual Opt. Problem, KKT)
- Statistical Applications (1)
- Optimization Algorithms (Unconstrained (2), ~~Equality~~, ~~Inequality~~).
  - Gradient / Steepest Descent
  - Newton's method.

Family of probability distributions on $\mathbb{R}^m$ indexed by parameter vector $x \in \mathbb{R}^n$, $\{p_x(\cdot)\}_{x \in \mathbb{R}^n}$

↖ argument is y here

For each $x \in \mathbb{R}^n$, $p_x(\cdot)$ is a probability distribution.

Continuous case: $p_x(y) \geq 0 \ \forall y \in \mathbb{R}^m$, $\int_{\mathbb{R}^m} p_x(y)\,dy = 1 \ \forall x \in \mathbb{R}^n$ (hold for every parameter)

Discrete case: $p_x(y) \geq 0 \ \forall y \in \mathbb{R}^m$, $\sum_y p_x(y) = 1$.

More convenient to work with log-likelihood

(make dependence on y explicit)

$l_y(x) \triangleq \log p_x(y)$

y: observations (measurements)

x: unknown parameters.

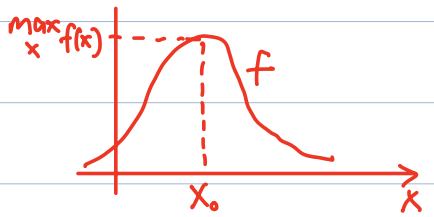
Maximum likelihood estimation (MLE)

$$\hat{x}_{ml}(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \; p_x(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \; \log p_x(y)$$

opt prob?

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \; l_y(x) \quad (*)$$

log likelihood of x



If we have prior information about x, say $x \in C \subseteq \mathbb{R}^n$, this can be incorporated as constraints in $(*)$ as follows

make constrain explicit in the opt prob itself

$$\hat{x}_{ml}(y) = \underset{x \in C}{\operatorname{arg\,max}} \; l_y(x)$$

Examples: Linear measurements in i.i.d. noise.

$$y_i = a_i^T x + v_i, \quad i = 1, \dots, m$$

$x \in \mathbb{R}^n$: unknown vector of parameters

$y_i \in \mathbb{R}$: observations. (known)

$a_i \in \mathbb{R}^n$: feature vectors (known) (attributes)

v_i's independent

$v_i \in \mathbb{R}$: are measurement errors or noise with distribution $p_v(\cdot)$.

Likelihood: $p_x(y) = \prod_{i=1}^{m} p_v(y_i - a_i^T x)$ where $(y_1, \dots, y_n) = y$

probability

randomness is v_i (express as that)

likelihood of observing measurement vector y = likelihood of observing that measurement errors all $y_i - a_i^T x$

Log-likelihood: $l_y(x) = \log p_x(y) = \log \left(\prod_{i=1}^{m} p_v(y_i - a_i^T x) \right)$

$$= \sum_{i=1}^{m} \log p_V(y_i - a_i^T x)$$

ML estimate is the optimal solution for

$$\max_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \log p_V(y_i - a_i^T x)$$

f need not be convex

Def: A function $f: \mathbb{R}^n \to (0,\infty) = \mathbb{R}_{++}$ is <u>log-convex</u> if $g: \mathbb{R}^n \to \mathbb{R}$ defined as $g(x) = \log f(x)$ is a convex function.

Properties: f_1, f_2 are log-convex, $f = f_1^{\alpha} f_2^{1-\alpha}$ is log-convex for any $\alpha \in [0,1]$.

Pf: $g(x) = \log f(x) = \log\left((f_1(x))^{\alpha}(f_2(x))^{1-\alpha}\right)$ ← convex combination

$= \alpha \log f_1(x) + (1-\alpha) \log f_2(x) \to$ convex.

($\log f_1(x)$: convex; $\log f_2(x)$: convex)

Def: A function $f: \mathbb{R}^n \to (0,\infty) = \mathbb{R}_{++}$ is <u>log-concave</u> if $g: \mathbb{R}^n \to \mathbb{R}$ defined as $g(x) = \log f(x)$ is a concave function.

so that we can take log

<u>Folklore</u>: Most densities (distributions) are log-concave.

(most of the time it's true)

<u>Examples</u>: Suppose v_i's are Gaussian with zero mean and variance σ^2 ($\mathbb{E}[v_i] = 0$, $\mathrm{Var}(v_i) = \sigma^2$), the density

$$p_V(v_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-v_i^2/(2\sigma^2)\right), \quad v_i \in \mathbb{R}.$$

probability density function

density of Gaussian

log concave

arg max (-ve of quadratic)

$v_i = y_i - a_i^T x$

$$\text{MLE: } \hat{x}_{ml} = \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \sum_{i=1}^{m} \log p_v(y_i - a_i^T x)$$

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \sum_{i=1}^{m} \log \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - a_i^T x)^2}{2\sigma^2} \right) \right)$$

(v_i)

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \; -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Ax - y\|^2$$

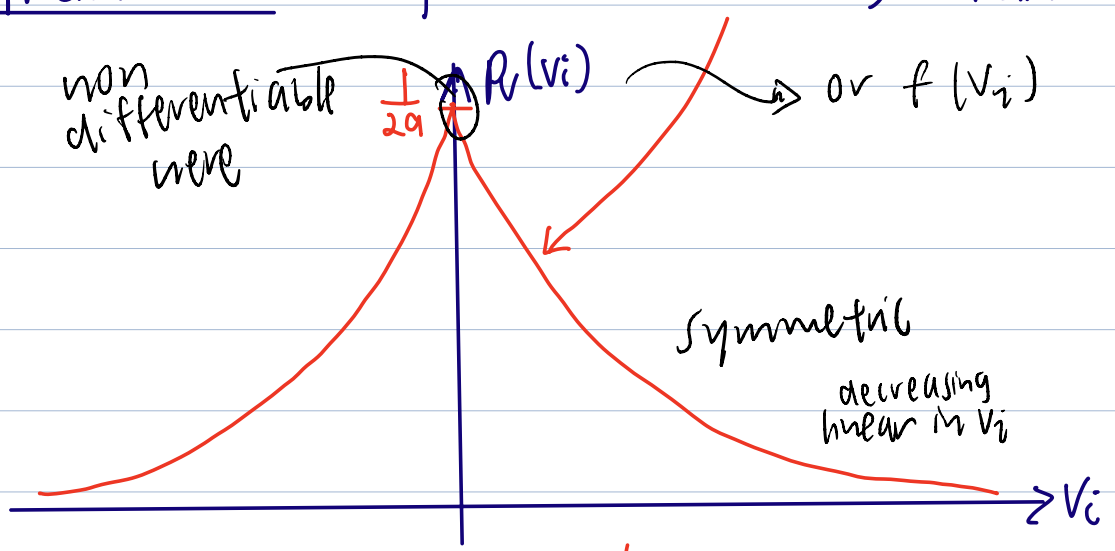
does not involve x (not important)

$A \in \mathbb{R}^{m \times n}$ is the matrix with rows $a_1^T, \ldots, a_m^T$.

ML estimate of x: $\hat{x}_{ml} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - y\|^2$ convex opt prob

$\Rightarrow$ least squares problem.

Laplacian noise $\quad p_v(v_i) = \frac{1}{2a} e^{-|v_i|/a}, \quad v_i \in \mathbb{R}$



log-concave

$$\text{MLE: } \hat{x}_{ml} = \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \sum_{i=1}^{m} \log p_v(y_i - a_i^T x)$$

$-|y_i - a_i^T x|/a$

↳ minus ℓ_1 norm

concave as ℓ_1 norm is convex

constant constant

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left(\frac{1}{2a} e^{-|y_i - a_i^T x|/a} \right)$$

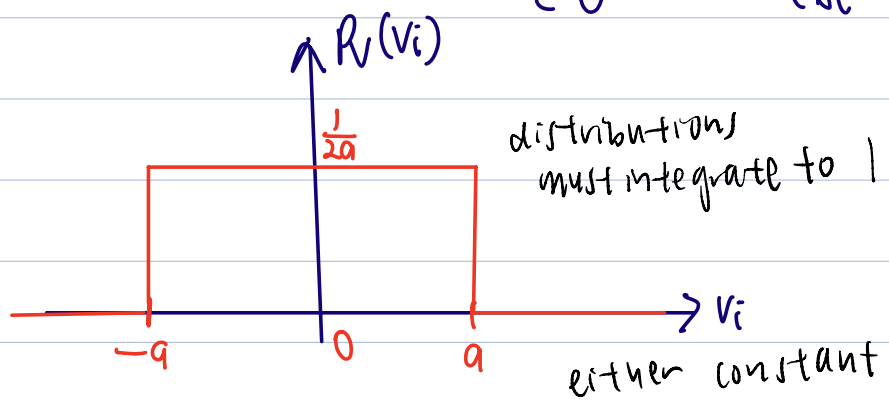
$$\hat{x}_{ml} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - y\|_1$$

When the noise is Laplacian, the MLE is an l_1-norm approximation problem.

$$\|z\|_1 = \sum_{i=1}^{m} |z_i|.$$

Uniform noise

uniform dist $p_V(v_i) = \begin{cases} \frac{1}{2a}, & v_i \in [-a, a] \\ 0 & \text{else} \end{cases}$



distributions must integrate to 1

holds true for any dist

MLE: $\hat{x}_{ml} = \underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \sum_{i=1}^{m} \log p_V(y_i - a_i^T x)$

either constant or 0 ($\log 0 \to -\infty$, never achieve the arg max)

ML estimate: $\hat{x}_{ml}$ is any vector x that satisfies $\|Ax - y\|_\infty \leq a$.

feasibility problem

$$\|z\|_\infty = \max_{1 \leq i \leq m} |z_i|$$

If form $Ax - y$, every component should have abs value not more than a

if this satisfied, probability is the same

Counting Problems with Poisson distribution

$y \in \{0, 1, \dots\}$ observations. — observations take values on a non-negative integer

$$\Pr(y=k) = \frac{e^{-\mu}\mu^k}{k!}, \quad k = 0, 1, 2, 3, \dots \qquad \mathbb{E}[y] = \mu$$

y: represent the count or number of events of a Poisson process over a fixed period of time.

Model the mean $\mu = \mathbb{E}[y] = a^T u + b \qquad u \in \mathbb{R}^n$

u: vector of explanatory variables.

$a \in \mathbb{R}^n$, $b \in \mathbb{R}$: unknown parameters.

$\mu = a^T u_i + b$.

$\frac{e^{-\mu}\mu^{y_i}}{y_i!}$

pairs of explanatory variables and measurements $= u, y$

Given # of observations (u_i, y_i), $i = 1, \dots, m$.

Aim: Find parameters $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ from these data.

put μ into poisson formula

Likelihood function: $L_{u,y}(a,b) = \prod_{i=1}^{m} \left[\frac{e^{-(a^T u_i + b)} (a^T u_i + b)^{y_i}}{y_i!} \right]$

Log-likelihood function:

$$l_{u,y}(a,b) = \sum_{i=1}^{m} \left[y_i \log(a^T u_i + b) - (a^T u_i + b) - \log(y_i!) \right]$$

$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} u_i \\ 1 \end{bmatrix}$

ML estimate:

$$(\hat{a}_{ml}, \hat{b}_{ml}) = \underset{a,b}{\operatorname{argmax}} \sum_{i=1}^{m} \left[y_i \log(a^T u_i + b) - (a^T u_i + b) \right]$$

$y_i \geq 0$

concave

log (linear function) is concave

This is a convex optimization problem. taking max & the objective fn is concave.

probability mass function of poisson is log-concave

Logistic Regression.

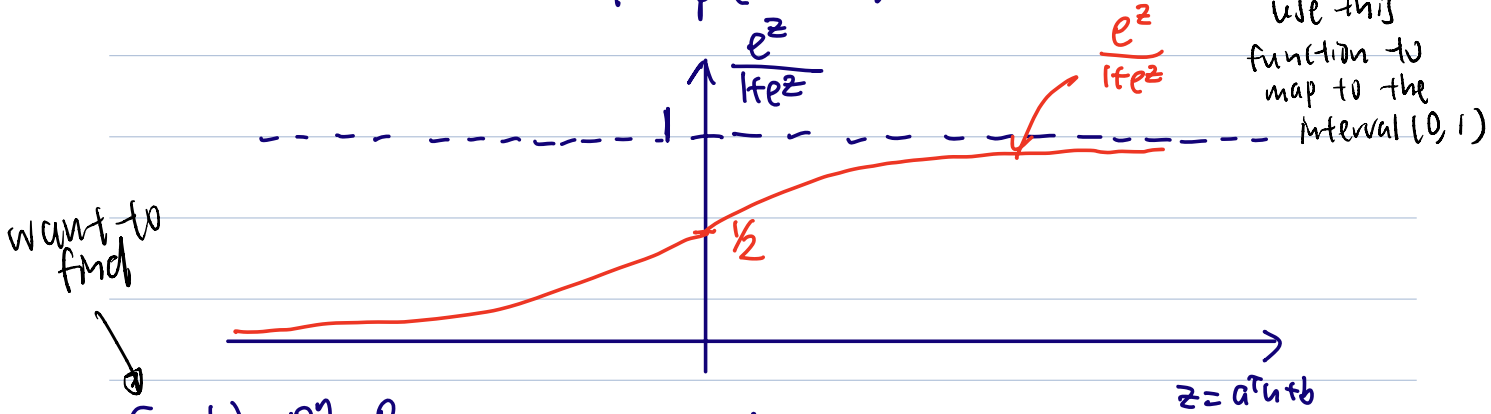
Consider a random variable $y \in \{0,1\}$

$Pr(y=1) = p, \quad Pr(y=0) = 1-p.$

$p \in [0,1]$ depends on a vector of explanatory variables $u \in \mathbb{R}^n$

Logistic model $p = \dfrac{\exp(a^T u + b)}{1+\exp(a^T u + b)} \in (0,1)$

real number



$(a,b) \in \mathbb{R}^n \times \mathbb{R}$ model parameters.

explanatory variables

Data: $u_1, \ldots, u_m \in \mathbb{R}^n$ and corresponding outcomes $y_1, \ldots, y_m \in \{0,1\}$.

Given data $\{(u_i, y_i)\}_{i=1}^m$, find (a,b). (want a good estimate for the model parameter)

This task is called logistic regression.

Reorder the data s.t. $y_1 = \cdots = y_q = 1, \; y_{q+1} = \cdots = y_m = 0$

Likelihood: $L_{u,y}(a,b) = \prod_{i=1}^{q} p_i \prod_{i=q+1}^{m} (1-p_i)$

$$= \underbrace{\left(\prod_{i=1}^{q} \frac{\exp(a^T u_i + b)}{1+\exp(a^T u_i + b)} \right)}_{y_i = 1} \underbrace{\left(\prod_{i=q+1}^{m} \frac{1}{1+\exp(a^T u_i + b)} \right)}_{y_i = 0}$$

Log-likelihood:

$$\ell(a,b) = \log \left[\left(\prod_{i=1}^{q} \frac{\exp(a^T u_i + b)}{1+\exp(a^T u_i + b)} \right) \left(\prod_{i=q+1}^{m} \frac{1}{1+\exp(a^T u_i + b)} \right) \right]$$

$$= \sum_{i=1}^{q} \log \left(\frac{\exp(a^T u_i + b)}{1+\exp(a^T u_i + b)} \right) + \sum_{i=q+1}^{m} \log \left(\frac{1}{1+\exp(a^T u_i + b)} \right)$$

$$= \sum_{i=1}^{q} (a^T u_i + b) - \sum_{i=1}^{m} \log \left(1+\exp(a^T u_i + b) \right).$$

MLE of a, b is

$$(\hat{a}_{mle}, \hat{b}_{mle}) = \underset{a,b}{\operatorname{argmax}} \underbrace{\sum_{i=1}^{q} (a^T u_i + b)}_{\text{Concave in } a,b.} - \underbrace{\sum_{i=1}^{m} \log \left(1+\exp(a^T u_i + b) \right)}_{\text{Concave in } (a,b)}$$

$\ell(a,b) = \sum_{i=1}^{q} (a^T u_i + b) - \sum_{i=1}^{m} \log \left(1+\exp(a^T u_i + b) \right)$

$-\log(\cdot)$: decreasing convex.

$\ell(a,b)$ is concave in (a,b)

Hence, the problem is a convex opt. problem & can be solved numerically & efficiently.

Covariance Estimation of Gaussian random vectors.

$y \in \mathbb{R}^n$: Gaussian random vector.

Zero-mean $\mathbb{E}[y] = \underline{0}$, covariance matrix (unknown) $\mathbb{E}[yy^T] = R$.

Probability density function:

$$p_R(y) = \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \exp\left(-y^T R^{-1} y / 2\right), \quad R \in S_{++}^n.$$

$|R| = \det(R)$

Estimate R from N independent samples $y_1, \dots, y_N \in \mathbb{R}^n$.

$$l_y(R) = \log p_R(y_1, \dots, y_N)$$

$$= \log \left[\prod_{k=1}^{N} \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \exp\left(-y_k^T R^{-1} y_k / 2\right) \right].$$

$$= -\frac{Nn}{2} \log(2\pi) - \frac{N}{2} \log\det(R) - \frac{1}{2} \sum_{k=1}^{N} y_k^T R^{-1} y_k$$

Define $Y = \frac{1}{N} \sum_{k=1}^{N} y_k y_k^T$ sample covariance matrix of observations $y_1, \dots, y_N$. ('based on this, want to estimate R)

$$l_y(R) = -\frac{Nn}{2} \log(2\pi) - \frac{N}{2} \log\det(R) - \frac{N}{2} \operatorname{tr}(R^{-1} Y)$$

$$\frac{1}{2}\sum_{k=1}^{N} y_k^T R^{-1} y_k = \frac{N}{2}\left(\frac{1}{N}\sum_{k=1}^{N} y_k^T R^{-1} y_k\right)$$

$y_k^T R^{-1} y_k$: 1×1 matrix

$$= \frac{N}{2}\frac{1}{N}\sum_{k=1}^{N} \operatorname{tr}\left(y_k^T R^{-1} y_k\right)$$

the trace is itself

$$= \frac{N}{2}\frac{1}{N}\sum_{k=1}^{N} \operatorname{tr}\left(R^{-1} y_k y_k^T\right) = \frac{N}{2}\operatorname{tr}\left(R^{-1}\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T\right)$$

$$= \frac{N}{2}\operatorname{tr}(R^{-1}Y).$$

need this to be concave but it is convex!

$$\max_{R \in S_{++}^n} l_y(R) = -\frac{Nn}{2}\log(2\pi) - \frac{N}{2}\log\det(R) - \frac{N}{2}\operatorname{tr}(R^{-1}Y)$$

$\det(S^{-1}) = \frac{1}{\det(S)}$.

This is not a convex opt. problem in R.

Let $S = R^{-1}$ (precision / information / inverse covariance matrix).

$$\max_{S \in S_{++}^n} l_y(S) = -\frac{Nn}{2}\log(2\pi) + \frac{N}{2}\log\det(S) - \frac{N}{2}\operatorname{tr}(SY).$$

(inner product (linear))

$\log\det(S)$ is concave in S & $\operatorname{tr}(SY)$ is also concave in S.

Hence the ML problem for S is a convex optimization problem.

$$S_{ml} = \underset{S \in \mathcal{A}}{\operatorname{argmax}} \log\det(S) - \operatorname{tr}(SY).$$

convex opt prob

$\mathcal{A}$: encode some prior knowledge on S.

R-L is +ve semi definite

i) $L \preceq R \preceq U \Leftrightarrow U^{-1} \preceq S \preceq L^{-1}$

ii) $\frac{\lambda_{max}(R)}{\lambda_{min}(R)} \leq K$: condition number

$\Leftrightarrow \frac{\lambda_{max}(S)}{\lambda_{min}(S)} \leq K$

iii) S to be sparse matrix $\Rightarrow$ Gaussian graphical models

$$S_{ML} = \underset{S \in \mathcal{A}}{\operatorname{argmax}} \log\det(S) - \operatorname{tr}(SY) - \lambda \|S\|_1$$

inverse covariance matrix sparse by adding regularisation parameter

$\|S\|_1 = \sum_{i,j} |S_{ij}|$ induces sparsity in matrix S.

Experiment Design

Regression: Estimate unknown parameter vector $x \in \mathbb{R}^n$ given some measurements or experiments.

$$y_i = (a_i^T) x + w_i, \quad i = 1, \dots, m.$$

feature vectors (choose from a set?)

w_i : measurement noise $w_i \sim N(0, 1)$.

Assume $a_1, \dots, a_m$ span $\mathbb{R}^n$. Set

$$A = \begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Then we know that $A^TA \in \mathbb{R}^{n \times n}$ is invertible.

$$A^TA = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_m \\ | & & | \end{bmatrix} \begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} = \sum_{i=1}^m a_i a_i^T \text{ has rank } n.$$

MLE of x: $x_{ml} = \underset{x}{\operatorname{argmin}} \|Ax - y\|_2^2$ (least sq; minimize sum of residual)

$$\hat{x} = x_{ml} = (A^TA)^{-1}A^Ty = \left(\sum_{i=1}^m a_i a_i^T\right)^{-1} \sum_{i=1}^m a_i y_i.$$

Estimation error covariance matrix $\mathbb{E}[ee^T]$ where

$e = \hat{x} - x$.

Error vector

$$e = \hat{x} - x = (A^TA)^{-1}A^Ty - x = (A^TA)^{-1}A^T(Ax + w) - x$$
$$= \cancel{(A^TA)^{-1}(A^TA)x} + (A^TA)^{-1}A^Tw - \cancel{x}$$
$$= (A^TA)^{-1}A^Tw.$$

Estimation error covariance matrix $\mathbb{E}[ee^T]$

$$= \mathbb{E}[(A^TA)^{-1}A^Tw((A^TA)^{-1}A^Tw)^T]$$
$$= \mathbb{E}[(A^TA)^{-1}A^Tww^TA(A^TA)^{-1}].$$

components of w indep and all have variance 1

$$= (A^TA)^{-1}\cancel{A^T}\ \underbrace{\cancel{E[ww^T]}}_{I}\ \cancel{A}\ \cancel{(A^TA)^{-1}} = (A^TA)^{-1} = E.$$

E characterizes the accuracy of the estimation or the informativeness of the experiments.
↳ the smaller the E is, the better (choose A huge?)

Suppose $a_1, \dots, a_m$ can be only chosen from p possible test vectors $v_1, \dots, v_p \in \mathbb{R}^n$.
(can only choose from these rows; $p \ll m$)

Choose $a_1, \dots, a_m$ from the set $\{v_1, \dots, v_p\}$ s.t. E is "small" in some precise sense. (error covariance should be small)

51 9 0 70 — m_1 m_2 m_3 m_4

$m = 130$. (how to allocate to choose these numbers s.t. E is as small as possible)

Let $m_j \in \mathbb{N}$ denote the # of experiments in which a_i is chosen to be v_j.

want to choose these numbers in a more principled way (convex opt)

$$E = \left(\sum_{i=1}^{m} a_i a_i^T\right)^{-1} = 51 v_1 v_1^T + 9 v_2 v_2^T + 0 v_3 v_3^T + 70 v_4 v_4^T$$

$$= \left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1} \quad \text{sum over } p \ (p \text{ is small})$$

want to minimize this

Expt design: Given the menu of possible choices of the expts $v_1, \dots, v_p$ & the total # of expts m, choose the #s of each type of expt (i.e., $m_1, m_2, \dots, m_p$ s.t. E is small.

$$m_1 + m_2 + \cdots + m_p = m.$$

constraint set already non-convex

$$\min_{\text{w.r.t. } S_{++}^n} \underbrace{\left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1}}_{\text{Estimation error}} \quad \text{s.t.} \quad m_i \in \mathbb{N} \cup \{0\} \ (0,1,2,3\ldots), \quad \sum_{i=1}^{p} m_i = m$$

m_is are integers

Integer program. (not convex)

Relaxation: Let $\lambda_i = \frac{m_i}{m}$ (fraction of times v_i is used). Then $E = \frac{1}{m}\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}$

$$\min_{\text{w.r.t. } S_{++}^n} E = \frac{1}{m}\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}$$

matrix

need to be converted into a function to be minimised

$$\lambda = (\lambda_1, \ldots, \lambda_p) \geq 0, \quad \sum_{i=1}^{p} \lambda_i = 1 \Leftrightarrow \mathbb{1}^T\lambda = 1$$

vector

Rmk: Almost equivalent to the integer program when m is large.

Scalarization: D-optimal design (D = determinant).

$$\min_{\lambda \geq 0,\ \mathbb{1}^T\lambda = 1} \log\det\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}$$

$\log\det\left(\frac{1}{\lambda}\right) = -\log\lambda$ which is convex

X

set of all probability vector (convex)

This is a convex optimization problem.

Rewrite as:

$$\min_{\substack{-\lambda \leq 0 \ (z) \\ \lambda \geq 0,\ \mathbb{1}^T\lambda = 1 \ (\mu)}} \log\det(X^{-1}) \quad \text{—(Primal)}$$

$$X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \quad (Z)$$

introduce X as a new parameter

Z (capital) matrix used as L multiplier

lagrange multipliers

Lagrangian:

$$L(\lambda, X; Z, z, \mu) = \log\det(X^{-1}) + z^T(-\lambda) + \mu(1^T\lambda - 1) + \operatorname{tr}\left(Z\left(X - \sum_{j=1}^{p} \lambda_j v_j v_j^T\right)\right)$$

primal variables

$$= \log\det(X^{-1}) + \operatorname{tr}(ZX) - \sum_j \lambda_j v_j^T Z v_j - z^T\lambda + \mu(1^T\lambda - 1)$$

linear

$$= \log\det(X^{-1}) + \operatorname{tr}(ZX) + \sum_{j=1}^{p} \lambda_j \underbrace{(-v_j^T Z v_j - z_j + \mu)}_{=0} - \mu.$$

$$g(Z, z, \mu) = \inf_{\lambda, X} L(\lambda, X; Z, z, \mu)$$

The min over λ is unbounded below unless $\mu - v_j^T Z v_j = z_j$.

Differentiate w.r.t. X and set to 0. This yields $X^{-1} = Z$.

$$h(X) = \log\det(X^{-1}) + \operatorname{tr}(ZX)$$

$$\nabla h(X) = -X^{-1} + Z = 0 \Rightarrow \quad X^{-1} = Z.$$

Dual function:

$$g(Z, z, \mu) = \begin{cases} \log\det Z + n - \mu, & \mu - v_j^T Z v_j = z_j \ \forall j \\ -\infty, & \text{else.} \end{cases}$$

Dual problem:

$$\max_{Z, z, \mu \succeq 0} \log\det Z + n - \mu \quad \text{s.t.} \quad \mu - v_j^T Z v_j = z_j \ \forall j$$

$Z \succ 0$ so $\mu \geq v_j^T Z v_j$

$$\max_{Z,\mu} \log\det Z + n - \mu \quad \text{s.t.} \quad v_j^T Z v_j \le \mu \quad \forall j = 1,\dots,p$$

$\det(\mu W) = \mu^n \det(W)$

$\Leftrightarrow \; v_j^T \frac{Z}{\mu} v_j \le 1$

Eliminate μ by making a change of variables $W = \frac{1}{\mu} Z$.

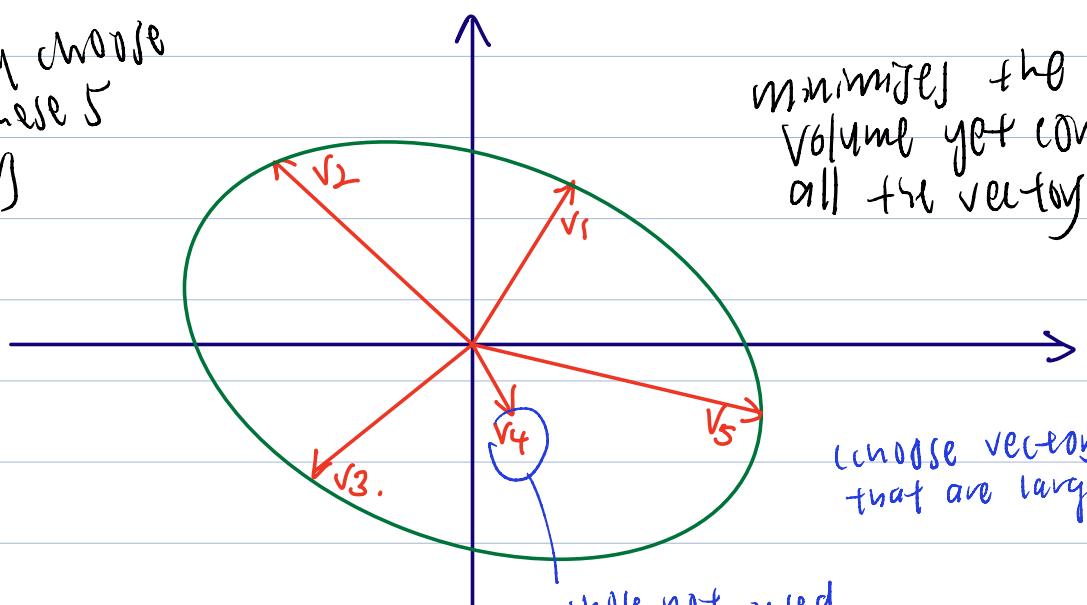
$$\max_{W,\mu} \log\det W + n + n\log\mu - \mu \quad \text{s.t.} \quad v_j^T W v_j \le 1 \;\; \forall j.$$

can split them up and do seperately

Analytically max over μ & the optimum is at $\mu = n$.

$$\max_{W \in S^n_{++}} \log\det W + n\log n \quad \text{s.t.} \quad v_j^T W v_j \le 1 \;\; \forall j.$$

Optimal solution W^* determines the min volume ellipsoid, centered at the origin $\{x : x^T W^* x \le 1\}$ that contains the points $v_1, \dots, v_p$.



can only choose from these 5 vectors

(proportions given by matrix W)

minimize the volume yet contain all the vectors

(choose vectors that are large)

those not used arise from complementary slackness

dual optimal variable

By complementary slackness $\gamma_i^*(1 - v_i^T W^* v_i) = 0 \quad \forall i = 1, \ldots, p$

i.e., the optimal design only uses expts that lie on the surface of the min volume ellipsoid.