Lecture 9 Linear Equations

Note by Samion Suwito on 2/18/25

Last Time

Example: Application of LS to Time Series Analysis
Autoregressive (AR) - building history in a sequence of operations
AR Model: $y(k) = w_1 y(k-1) + \cdots + w_n y(k-n) + e(k)$
Suppose we know $\phi(k) = (y(k-1), \ldots, y(k-n))^\top$ and $w = (w_1, \ldots w_n)^\top$
How to estimate $y(n)$?
$y(k) \approx w^\top \phi(k)$
Estimation error: $e(k) = y(k) - w^\top \phi(k)$

Problem: How to learn weights w?
Consider weights to be part of the physical system and don't change based on time.
Get Training Data

Suppose we observe past history $y(1), y(2), \ldots, y(N) \iff \phi(1), \phi(2), \ldots, \phi(N)$
since we don't have before $y(1)$ we can just put 0s in the ϕ.
Idea: $y(k) \approx w^\top \phi(k)$ so formulate LS problem as:

$$\min_w \|\Phi w - y\|_2^2 \; \Phi = \begin{bmatrix} - & \phi(1)^\top & - \\ & \vdots & \\ - & \phi(N)^\top & - \end{bmatrix} y = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}$$

We can rephrase this as $\min_w \sum |e(i)|^2$

Question: What if $y(k)$ is quadratic function of previous $y(k-1), y(k-2)$?

$$\begin{aligned} y(k) &= w_1 y(k-1) + w_2 y(k-2) \\ &\quad + w_3 y(k-1) y(k + w_4 y(k-1)^2 + w_5 y(k-2)^2 \\ &\quad + e(k) \end{aligned}$$

The same scheme works, but now

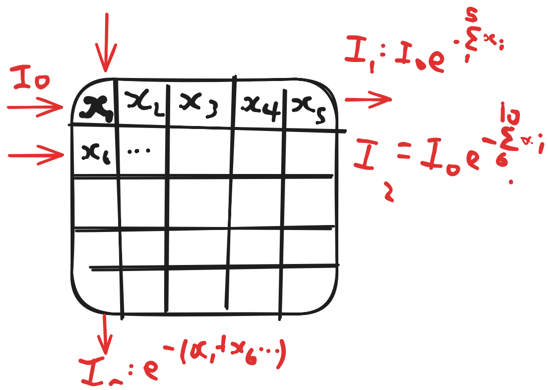
$$\begin{aligned} \phi(k) = (y(k-1), y(k-2), y(k-1)y(k-2), \\ y(k-1)^2, y(k-2)^2) \end{aligned}$$

Linear Equations In Engineering

Friendly Review of Some Examples

Good for modelling constraints in engineering/science.

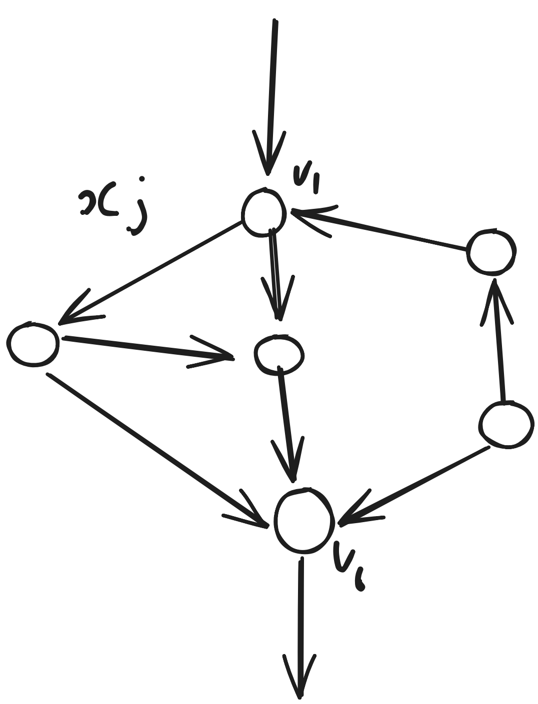
Tomography (Imaging)



Where x_i is density of tissue in pixel
Red arrow is beam of light and we measure the intensity coming out

Take the Log of the measurements to get linear equations
$\log\left(\frac{I_0}{I_1}\right) = x_1 + x_2 + x_3 + x_4 + x_5$

Network Flows



$G = (V, E)$ And (u, v) means from u to v
Flow is represented by x_i for edge i
In general conserved flow, ”Flow in = Flow out”
Feasible flow: $(x_e)_{e \in E}$ s.t. flow in = Flow out for every vertex

Flow in = flow out at vertex v:

$$\sum_{e \in E} x_e = \sum_{\epsilon \in E} x_\epsilon$$

s.t. $e = (\cdot, v)$ and $\epsilon = (v, \cdot)$
This is equivalent to

$$\sum_{u \in V} x_{(u,v)} = \sum_{v \in V} x_{(v,v)}$$

s.t. $(u, v) \in E$ and $(v, v) \in E$. We can represent this in a matrix where the *rows* are indexed by vertices and *columns* are indexed by edges

$$\begin{bmatrix} \quad & \quad & \quad \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_{e_1} \\ \vdots \\ x_{e_E} \end{bmatrix} = 0$$

Where

$$[A]_{v,e} = \begin{cases} 1 \text{ if } e = (\cdot, v) \\ -1 \text{ if } e = (v, \cdot) \\ 0 \text{ otherwise} \end{cases}$$

A is called an adjacency matrix.
The 0 at the end represents no external input/output however if there is external input/output we can control from each vertex how much comes out where for example in the picture we can set it so that the entry corresponding to v_1, v_6 how much flow we want e.g.

$$Ax = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ -5 \end{bmatrix}$$

Linear Equations In Optimisation

$$\min_x f(x) \text{ s.t. } Ax = b$$

Linear Algebra Review

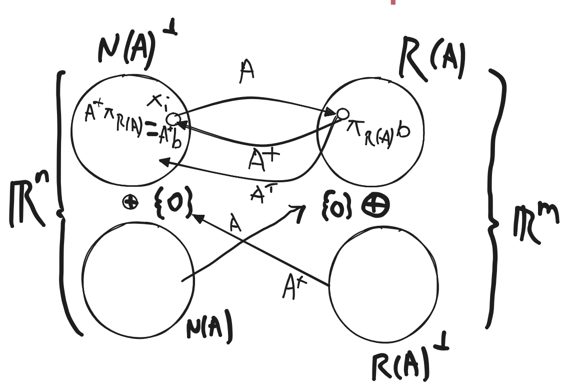
Linear Algebra = The Language of Optimisation
Basic objects: vectors, matrices (linear transformations)
Important Concepts: Subspaces, Bases, Norms, Inner products
^ geometric in nature

Most Important tools:
Projections: Given subspace $\mathcal{S}$ solve

$$\pi_{\mathcal{S}}(x) = \arg\min_{s \in \mathcal{S}} \|s - x\|_2^2$$

Solution characterised by "orthogonally".
$(x - \pi_{\mathcal{S}}(x))^\top y = 0 \quad \forall y \in \mathcal{S}$
Remark: $x \mapsto \pi_{\mathcal{S}}(x)$ is linear transformation.
Application: Gram-Schmidt
$u_{k+1} = x_{k+1} - \pi_{\mathrm{Sp}(u_1 \dots u_k)}(x_{k+1})$
Application of Gram-Schmidt: $A = QR$,
R is upper triangular due to the sequential nature of Gram-Schmidt

Matrices represent Linear transformations:
Four Fundamental Subspaces



This picture can allow us to better understand/reduce problems
Example:

$$\min_x \|Ax - b\|_2^2 \;\text{ s.t. } Cx = d$$

If feasible transform this to

$$\min_z \|Ax_0 + ANz - b\|_2^2$$

where $x_0 = C^+d$ and $R(N) = \mathcal{N}(C)$

Many of the other concepts we saw emerge from optimisation problems:
Example 1 If $A \in \mathbb{S}^n_+$, $\lambda_{\max}(A) = \max_x \frac{x^\top Ax}{x^\top x}$ get **spectral decomposition** the get A, $A = U\Lambda U^\top$.
Applications: PCA=just the spectral decomposition of $XX^\top$

Example 2: **SVD** = for generic $A \in \mathbb{R}^{m\times n}$

$$v_1 = \arg\max_{v:\|v\|=1} \|Av\|_2 \quad \sigma_1 = \|Av_1\|_2 \quad u_1 = \frac{Av_1}{\sigma_1}$$

$\implies$ SVD of $A = \sum \sigma_i u_i v_i^\top = U\Sigma V^\top$.
Applications: Low rank approximation:

$$\sum_{i=1}^k u_i\sigma_i v_i^\top = \arg\min_{A_k:\mathrm{rank}(A_k)=k} \|A_k - A\|_F$$

Application: Pseudo-inverse $A^+ = V\Sigma U^\top$
Example: $\min_x \|Ax-b\|_2 \implies x^* = A^+b + \mathcal{N}(A)$