ECON 201B

Author: Wenxiao Yang

Institute: Haas School of Business, University of California Berkeley

Date: 2024

Seeking what is true is not seeking what is desirable.

Contents

Chapter 1 Geometric Programming (GP)

1.1 Arithmetic Mean-Geometric Mean Inequality

1.1.1 AM-GM inequality: $\frac{x_1+x_2+...+x_n}{n} \geq \sqrt[n]{x_1x_2...x_n}$

Theorem 1.1 (AM-GM inequality)

For any $x_1, x_2, .., x_n \geq 0$,

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1x_2 \cdots x_n}$$

Equality is only achieved when $x_1 = x_2 = \cdots = x_n$

♡

- The LHS is the arithmetic mean (average) of $x_1, x_2, ..., x_n$.
- The RHS is the geometric mean of $x_1, x_2, ..., x_n$.

1.1.2 Weighted AM-GM inequality: $\sum_{i=1}^{n} \delta_i x_i \geq \prod_{i=1}^{n} x_i^{\delta_i}$

Theorem 1.2 (Weighted AM-GM inequality)

For any $x_1, x_2, ..., x_n \geq 0$ *with* $\delta_1, \delta_2, \cdots, \delta_n > 0$ *with* $\delta_1 + \cdots + \delta_n = 1$,

$$\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n \geq x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}$$

Equality is only achieved if $x_1 = x_2 = \cdots = x_n$.

♡

When $\delta_1 = \cdots = \delta_n = \frac{1}{n}$, the inequality recovers to unweighted AM-GM inequality.

Proof 1.1

Prove by Jensen's Inequality:

Let $f(t) = -\ln(t)$ *which is strictly convex in* $(0, \infty)$. *Take* $\lambda_1, \lambda_2, \cdots, \lambda_n >$ *such that* $\delta_1+\delta_2+\cdots+\delta_n = 1$. *According to Jensen's Inequality:*

$$f(\sum_{i=1}^{n} \delta_i x_i) \leq \sum_{i=1}^{n} \delta_i f(x_i)$$

By substituting f*:*

$$-\ln(\sum_{i=1}^{n} \delta_i x_i) \leq -\sum_{i=1}^{n} \delta_i \ln(x_i)$$

$$e^{\ln(\sum_{i=1}^{n} \delta_i x_i)} \geq e^{\sum_{i=1}^{n} \delta_i \ln(x_i)}$$

$$\sum_{i=1}^{n} \delta_i x_i \geq x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}$$

1.2 Unconstrained Geometric Programs

1.2.1 Def: Posynomial

Definition 1.1

*A **posynomial term** in variables $t_1, ..., t_m$ is a function of the form*

$$Ct_1^{\alpha_1} t_2^{\alpha_2} \cdots t_m^{\alpha_m}$$

where $\alpha_1, ..., \alpha_m \in \mathbb{R}$ and $C > 0$ is a positive real number. ♣

Definition 1.2

*A **posynomial** is a sum of posynomial terms.* ♣

1.2.2 General Strategy: AM-GM inequality

Definition 1.3

An unconstrained geometric program (GP) is the problem of minimizing a posynomial over positive real inputs.

$$\min_{(t_1,\cdots,t_m \in \mathbb{R}^m_{>0}} g(t_1, \cdots, t_m)$$

where $g(t_1, \cdots, t_m)$ is a sum of posynomial terms. $g(t_1, \cdots, t_m) = \sum_{i=1}^n Term_i(t_1, \cdots, t_m)$, where $Term_i(t_1, \cdots, t_m) = C_i t_1^{\alpha_{i,1}} t_2^{\alpha_{i,2}} \cdots t_m^{\alpha_{i,m}}$ ♣

General Strategy:

Choose weights $\delta_1, ..., \delta_n > 0$ with $\delta_1 + \cdots + \delta_n = 1$ and use the inequality

$$\begin{aligned}\sum_{i=1}^n Term_i(t_1, \cdots, t_m) &= \sum_{i=1}^n \delta_i \left(\frac{Term_i(t_1, \cdots, t_m)}{\delta_i}\right) \\ &\geq \left(\frac{Term_1(t_1, \cdots, t_m)}{\delta_1}\right)^{\delta_1} \cdots \left(\frac{Term_n(t_1, \cdots, t_m)}{\delta_n}\right)^{\delta_n}\end{aligned}$$

1.2.3 Dual of the Unconstrained GP

Example: Suppose we want to find the minimum of $f(x, y) = 2xy + \frac{y}{x^2} + \frac{3x}{y}$.

We want

$$2xy + \frac{y}{x^2} + \frac{3x}{y} \geq \left(\frac{2xy}{\delta_1}\right)^{\delta_1} \left(\frac{y}{\delta_2 x^2}\right)^{\delta_2} \left(\frac{3x}{\delta_3 y}\right)^{\delta_3}$$

which requires

(1) **Power of x:** $\delta_1 - 2\delta_2 + \delta_3 = 0$

(2) **Power of y:** $\delta_1 + \delta_2 - \delta_3 = 0$

(3) **Sum:** $\delta_1 + \delta_2 + \delta_3 = 1$

(4) **Positive:** $\delta_1, \delta_2, \delta_3 > 0$

In general, we want to eliminate all $t_1, ..., t_n$ is the RHS of the inequality, then the RHS can be transformed into constant $V(\delta) = \left(\frac{C_1}{\delta_1}\right)^{\delta_1} \left(\frac{C_2}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{C_n}{\delta_n}\right)^{\delta_n}$ which is a lower bound of $g(\vec{t}), \vec{t} \in \mathbb{R}^m_{>0}$

Dual Geometric Problem

$$\max_{\vec{\delta} \in \mathbb{R}^n_{>0}} \quad V(\vec{\delta}) = \left(\frac{C_1}{\delta_1}\right)^{\delta_1} \left(\frac{C_2}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{C_n}{\delta_n}\right)^{\delta_n}$$

$$\begin{aligned} s.t. \qquad \delta_1 \alpha_{1,1} + \delta_2 \alpha_{2,1} + \cdots + \delta_n \alpha_{n,1} &= 0 \quad \text{(power of } t_1) \\ &\vdots \\ \delta_1 \alpha_{1,m} + \delta_2 \alpha_{2,m} + \cdots + \delta_n \alpha_{n,m} &= 0 \quad \text{(power of } t_m) \\ \delta_1 + \cdots + \delta_n &= 1 \\ \delta_1, \delta_2, ..., \delta_n &> 0 \end{aligned}$$

Suppose $\vec{\delta}^*$ is the solution to the dual GP.

$$\sum_{i=1}^{n} \delta_i^* \left(\frac{Term_i(\vec{t})}{\delta_i^*}\right) \geq \left(\frac{Term_1(\vec{t})}{\delta_1^*}\right)^{\delta_1^*} \cdots \left(\frac{Term_n(\vec{t})}{\delta_n^*}\right)^{\delta_n^*}$$

The inequality holds only if

$$\frac{Term_1(\vec{t})}{\delta_1^*} = \frac{Term_2(\vec{t})}{\delta_2^*} = \cdots = \frac{Term_m(\vec{t})}{\delta_m^*} = V(\vec{\delta}^*)$$

where $V(\vec{\delta}^*)$ is a function only related to $\vec{\delta}^*$.

Note: It is possible that system of equations for $\vec{t}$ has no solution.

Dual $\Rightarrow$ Primal

Theorem 1.3

Given a feasible point $\vec{\delta}^*$ of the dual program. If the equations

$$\frac{Term_1(\vec{t})}{\delta_1^*} = \frac{Term_2(\vec{t})}{\delta_2^*} = \cdots = \frac{Term_m(\vec{t})}{\delta_m^*} = V(\vec{\delta}^*)$$

have a solution $\vec{t}^*$ with $t_i^* > 0, i = 1, 2, ..., m$, then $\vec{t}^*$ is a primal solution, $\vec{\delta}^*$ is a dual solution, and $g(\vec{t}^*) = V(\vec{\delta}^*)$

Proof 1.2

If a solution $\vec{t}^*$ with $t_i^* > 0, i = 1, 2, ..., m$ exists, then $g(\vec{t}^*) = V(\vec{\delta}^*)$ (by AM-GM inequality).

Suppose there exists another solution $\vec{t}'$ to the primal problem. Because $V(\vec{\delta}^*)$ is a lower bound of $g(\vec{t})$, $g(\vec{t}') \geq V(\vec{\delta}^*) = g(\vec{t}^*) \Rightarrow \vec{t}^*$ is an optimal solution minimizing g.

Suppose there exists feasible $\vec{\delta}'$, $V(\vec{\delta}')$ is a lower bound of $g(\vec{t}) \Rightarrow V(\vec{\delta}^*) = g(\vec{t}^*) \geq V(\vec{\delta}') \Rightarrow \vec{t}^*$ is also optimal maximizing V.

Primal $\Rightarrow$ Dual

Theorem 1.4

If $\vec{t}^*$ is an optimal primal solution, then

$$\vec{\delta}^* = \left(\frac{Term_1(\vec{t}^*)}{g(\vec{t}^*)}, \frac{Term_2(\vec{t}^*)}{g(\vec{t}^*)}, \cdots, \frac{Term_n(\vec{t}^*)}{g(\vec{t}^*)}\right)$$

is an optimal dual solution and $g(\vec{t}^*) = V(\vec{\delta}^*)$.

♡

Proof 1.3

If $\vec{t}^*$ is an optimal primal solution, it's a critical point and $\nabla g(\vec{t}^*) = \vec{0}$. Recall $g(\vec{t}^*) = \sum_{i=1}^{m} Term_i(\vec{t}^*) = \sum_{i=1}^{m} C_i {t_1^*}^{\alpha_{i,1}} {t_2^*}^{\alpha_{i,2}} \cdots {t_m^*}^{\alpha_{i,m}}$

$\nabla g(\vec{t}^*) = \vec{0}$ implies, for each $1 \leq j \leq n$,

$$\frac{\partial g(\vec{t}^*)}{\partial t_j} = \sum_{i=1}^{n} \frac{\alpha_{i,j}}{t_j} Term_i(\vec{t}^*) = 0$$

Then, we can check $\vec{\delta}^* = \left(\frac{Term_1(\vec{t}^*)}{g(\vec{t}^*)}, \frac{Term_2(\vec{t}^*)}{g(\vec{t}^*)}, \cdots, \frac{Term_n(\vec{t}^*)}{g(\vec{t}^*)}\right)$ is feasible in dual problem and can get the equality in AM-GM inequality.

Chapter 2 Polynomial Interpolation

Suppose we are given a collection of points $\{(x_1, y_1), (x_2, y_2), ..., (x_k, y_k)\}$. We want to find **polynomial** f that passes through all the points.

We have two methods to solve this problem:

(1) Set up and solve the system $M\vec{a} = \vec{y}$

(2) Use the Lagrange interpolation formula.

2.1 Method 1: $M\vec{a} = \vec{y}$

If $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ passes through all (x_i, y_i), then

$$\begin{gathered}a_n x_1^n + \cdots + a_1 x_1 + a_0 = y_1 \\ a_n x_2^n + \cdots + a_1 x_2 + a_0 = y_2 \\ \vdots \\ a_n x_k^n + \cdots + a_1 x_k + a_0 = y_k\end{gathered}$$

in matrix form as $M\vec{a} = \vec{y}$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & \ddots & & \vdots \\ 1 & x_k & x_k^2 & \cdots & x_k^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$$

In order to avoid no solution or multi solutions, we would let to set $n = k - 1$.

Theorem 2.1

If $x_i \neq x_j$ for all $1 \le i < j \le k$, then there is a unique polynomial of degree at most $k-1$ that passes through the points $\{(x_1, y_1), (x_2, y_2), \cdots, (x_k, y_k)\}$.

2.2 Method 2: Lagrange Interpolation Formula

Suppose

$$l_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}, i = 1, 2, ..., k$$

The $l_1, \cdots, l_k$ has the following properties:

- Each has degree $k - 1$

- $l_i(x_i) = 1$
- $l_i(x_j) = 0$ when $i \neq j$

Then the Lagrange Interpolation Formula is

$$f(x) = y_1 l_1(x) + y_2 l_2(x) + \cdots + y_k l_k(x)$$

where $f(x_i) = y_1 \cdot 0 + y_2 \cdot 0 + \cdots + y_i \cdot 1 + \cdots + y_k \cdot 0 = y_i$

2.3 Lines of Best Fit

Suppose we use a linear function $y = ax + b$ to fit the collection of points (x_1, y_1), ..., (x_k, y_k).

We use error to measure the accuracy of line's accuracy of fitting.

1. $Error(a, b) = |(ax_1 + b) - y_1| + \cdots + |(ax_k + b) - y_k|$

   Pros: 1. Convex; 2. Minimizing problem is linear.

   Cons: not differentiable

2. $Error(a, b) = [(ax_1 + b) - y_1]^2 + \cdots + [(ax_k + b) - y_k]^2 = \|a\vec{x} + b\vec{1} - \vec{y}\|^2$ which is also convex

$$\frac{\partial Error(a, b)}{\partial a} = 2\sum_{i=1}^{n}(ax_i + b - y_i)x_i = 2(a\vec{x} + b\vec{1} - \vec{y})^T \cdot \vec{x}$$
$$\frac{\partial Error(a, b)}{\partial b} = 2\sum_{i=1}^{n}(ax_i + b - y_i) = 2(a\vec{x} + b\vec{1} - \vec{y})^T \cdot \vec{1}$$

   The critical point is the global minimizer

2.4 Least-Square Problem (Overconstrainted $A\vec{x} = \vec{b}$)

For $A \in \mathbb{R}^{m \times n}, \vec{y} \in \mathbb{R}^m, m > n$. We want to solve $A\vec{x} = \vec{y}$. However, this equation system is overconstrained, we can only find a $\vec{x}$ to minimize the error between $A\vec{x}$ and $\vec{y}$.

The least-square error problem can be written as find $\vec{x} \in \mathbb{R}^n$ to minimize $\|A\vec{x} - \vec{y}\|$.

If A is an $m \times n$ matrix, then the set $V = \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$.

Then, "minimizing $\|A\vec{x} - \vec{y}\|$" means "finding the point of V closest to $\vec{y} \in \mathbb{R}^m$".

2.4.1 Lemma: closest point $\Leftrightarrow (A\vec{x}^* - \vec{y}) \perp \vec{a},\ \forall \vec{a} \in V$

Lemma 2.1

If $V = \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$, then the point $A\vec{x}^* \in V$ is the closest point of V to $\vec{y} \in \mathbb{R}^m$ if and only if

$$(A\vec{x}^* - \vec{y}) \perp \vec{a},\ \forall \vec{a} \in V$$

Proof 2.1

We look at restrictions to lines. $\vec{x}^*$ is the global minimizer of $f(\vec{x}) = ||A\vec{x} - \vec{y}||^2$. $\vec{u} \in \mathbb{R}^n$ is an arbitrary element of V. Let $\vec{a} = A\vec{u}$.

$$\begin{aligned}\phi_{\vec{u}}(t) &= ||A(\vec{x}^* + t\vec{u}) - \vec{y}||^2 \\ &= ||A\vec{x}^* - \vec{y} + t\vec{a}||^2 \\ &= ||A\vec{x}^* - \vec{y}||^2 + 2t(A\vec{x}^* - \vec{y})^T\vec{a} + t^2||\vec{a}||^2\end{aligned}$$

Since $\vec{x}^*$ is the global minimizer of $f(\vec{x})$, $t = 0$ is the global minimizer of $\phi_{\vec{u}}$. For $C_1t^2 + C_2t + C_3$, $t = 0$ is global minimizer when $C_1 \geq 0, C_2 = 0 \Rightarrow (A\vec{x}^* - \vec{y})^T\vec{a} = 0$

2.4.2 Theorem: $\vec{x}^* = (A^TA)^{-1}A^T\vec{y} = A^+\vec{y}$

We can use this characterization to write down a **normal equation**. (”normal” is another word for ”perpendicular”)

Theorem 2.2

A point $\vec{x}^* \in \mathbb{R}^n$ minimizes $||A\vec{x} - \vec{y}||$ if and only if $A^TA\vec{x}^* = A^T\vec{y}$

♡

Proof 2.2

Let $\vec{a}^{(i)} = A\vec{e}^{(i)}$ be the basis vector of V, then any vector $\vec{a}$ can be linear combination of $\{\vec{a}^{(i)}\}_{i=1,...,n}$. Hence, $\vec{x}^*$ is the global minimizer $\Leftrightarrow (A\vec{x}^* - \vec{y})\perp\vec{a}^{(i)}, i = 1, ..., n \Leftrightarrow A^T(A\vec{x}^* - \vec{y}) = 0$

We can solve the optimal x^* as

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

we call $A^+ = (A^TA)^{-1}A^T$ the pseudoinverse of A because A^+ is a ”pseudo-solution” to the overconstrained system.

2.4.3 Def: Projection Matrix: $P = AA^+$; Projection of $\vec{y}$ on V: $A\vec{x}^* = P\vec{y}$

Definition 2.1

The matrix $P = AA^+$ is called a ***projection matrix****. (It maps a point $\vec{y} \in \mathbb{R}^n$ to $P\vec{y}$, the closest point in V to $\vec{y}$, $P\vec{y}$ is also called projection of $\vec{y}$ on V.)*

♣

We have $P\vec{y} = A\vec{x}^*$, $(P\vec{y} - \vec{y})\perp\vec{a}$, $\forall\vec{a} \in V$.

Properties of P:

- $P^2 = P$ (i.e., P is idempotent)
- $P^T = P$ (i.e., P is symmetric)

2.4.4 Special Case: Projection on vector $Proj_{\vec{a}}(\vec{y}) = \frac{(\vec{a}\cdot\vec{y})\vec{a}}{\|\vec{a}\|^2}$

A special case: $A = \vec{a} \in \mathbb{R}^m$.

$$P = \vec{a}(\vec{a}^T\vec{a})^{-1}\vec{a}^T = \frac{\vec{a}\vec{a}^T}{\|\vec{a}\|^2}$$

The projection of $\vec{y}$ onto $\vec{a}$

$$Proj_{\vec{a}}(\vec{y}) = P\vec{y} = \frac{\vec{a}\vec{a}^T}{\|\vec{a}\|^2}\vec{y} = \frac{\vec{a}\cdot\vec{y}}{\|\vec{a}\|^2}\vec{a}$$

where $\frac{\vec{a}\cdot\vec{y}}{\|\vec{a}\|^2}$ is a scalar to measure how much of $\vec{y}$ is "pointing in the same direction as" $\vec{a}$.

If we normalize $\vec{a}$ to $\vec{u}$, the projection of $\vec{y}$ onto vector $\vec{u}$ is

$$Proj_{\vec{u}}(\vec{y}) = (\vec{u}\cdot\vec{y})\vec{u}$$

2.4.5 Theorem: Projection Matrix = Sum of outer products of orthonormal basis

Let $\vec{u}^{(1)}, ..., \vec{u}^{(n)}$ be the **orthonormal basis** of $V \in \mathbb{R}^m$ (which requires $\|\vec{u}^{(i)}\| = 1, i = 1, ..., n$ and $\vec{u}^{(i)}\cdot\vec{u}^{(j)} = 0, \forall i \neq j$)

Theorem 2.3

Suppose that $V \subseteq \mathbb{R}^m$ with an orthonormal basis $\{\vec{u}^{(1)}, ..., \vec{u}^{(n)}\}$. Then the projection matrix onto V is given by the formula

$$P = \vec{u}^{(1)}(\vec{u}^{(1)})^T + \vec{u}^{(2)}(\vec{u}^{(2)})^T + \cdots + \vec{u}^{(n)}(\vec{u}^{(n)})^T$$

♡

Note: sometimes $\vec{u}\vec{v}^T$ is called the **outer product** of $\vec{u}$ and $\vec{v}$.

Proof 2.3

For any $\vec{y}$, let $\vec{x} = P\vec{y}$ and $\vec{z} = \vec{y} - \vec{x}$, we know $\vec{z} \perp V$.

Because $\vec{x} \in V$, we can write

$$\vec{x} = a_1\vec{u}^{(1)} + \cdots + a_n\vec{u}^{(n)}$$

So we can write

$$\vec{y} = \vec{x} + \vec{z} = a_1\vec{u}^{(1)} + \cdots + a_n\vec{u}^{(n)} + \vec{z}$$

For any $i = 1, ..., n$, we can compute by orthonormal property and

$$\vec{u}^{(i)}\cdot\vec{y} = a_i$$

Then,

$$\begin{aligned}\vec{x} &= (\vec{u}^{(1)} \cdot \vec{y})\vec{u}^{(1)} + \cdots + (\vec{u}^{(n)} \cdot \vec{y})\vec{u}^{(n)} \\ &= \vec{u}^{(1)}(\vec{u}^{(1)})^T\vec{y} + \cdots + \vec{u}^{(n)}(\vec{u}^{(n)})^T\vec{y} \\ &= \left[\vec{u}^{(1)}(\vec{u}^{(1)})^T + \vec{u}^{(2)}(\vec{u}^{(2)})^T + \cdots + \vec{u}^{(n)}(\vec{u}^{(n)})^T\right]\vec{y} \\ &= P\vec{y}\end{aligned}$$

2.4.6 Corollary: Q has orthonormal columns $\Rightarrow \vec{x}^* = Q^T\vec{y}$. ($Q^+ = Q^T$)

Corollary 2.1

When columns of Q are orthonormal, the vector $\vec{x}^*$ that minimizes $\|Q\vec{x} - \vec{y}\|$ can be computed as $\vec{x}^* = Q^T\vec{y}$.

♡

Proof 2.4

$Q^+ = (Q^TQ)^{-1}Q^T = IQ^T = Q^T$

2.4.7 The Gram-Schmidt process

Now we know that if Q has orthonormal columns, then we get a much nicer formula for the projection matrix and for the least-squares minimization problem. How do we make Q have orthonormal columns?

One method for doing this is the Gram-Schmidt process. We want to input vectors $\vec{a}^{(1)}, ..., \vec{a}^{(n)}$ and output orthonormal vectors $\vec{u}^{(1)}, ..., \vec{u}^{(n)}$.

We assume $\vec{a}^{(1)}, ..., \vec{a}^{(n)}$ are linearly independent. In the algorithm, we will produce $\vec{v}^{(1)}, ..., \vec{v}^{(n)}$ that are orthogonal but not orthonormal, then we get $\vec{u}^{(1)}, ..., \vec{u}^{(n)}$ by $\vec{u}^{(i)} = \frac{\vec{v}^{(i)}}{\|\vec{v}^{(i)}\|}$.

Gram-Schmidt process:

(1) Let $\vec{v}^{(1)} = \vec{a}^{(1)}$ and $\vec{u}^{(1)} = \frac{\vec{v}^{(1)}}{\|\vec{v}^{(1)}\|}$

(2) For $j = 1, ..., n$

$$\begin{aligned}\vec{v}^{(j)} &= \vec{a}^{(j)} - \sum_{i=1}^{j-1}(\vec{u}^{(i)} \cdot \vec{a}^{(j)})\vec{u}^{(i)} \\ &= \vec{a}^{(j)} - \sum_{i=1}^{j-1}\frac{(\vec{v}^{(i)} \cdot \vec{a}^{(j)})\vec{v}^{(i)}}{\|\vec{v}^{(i)}\|^2} \\ &= \vec{a}^{(j)} - \sum_{i=1}^{j-1} Proj_{\vec{v}^{(i)}}(\vec{a}^{(j)})\end{aligned}$$

and

$$\vec{u}^{(j)} = \frac{\vec{v}^{(j)}}{\|\vec{v}^{(j)}\|}$$

Note: In the general case, where $\vec{a}^{(1)}, ...\vec{a}^{(n)}$ are not linearly independent, step (2) will sometimes give us $\vec{v}^{(j)} = 0$. In that case, we omit the j^{th} vector: what this tells us is that $\vec{a}^{(j)}$ is not necessary to span the subspace.

2.5 Minimum-norm problems (Underconstrainted $A\vec{x} = \vec{b}$)

Consider systems of equations $A\vec{x} = \vec{b}$ with infinitely many solutions. We want to find the solution $\vec{x}$ with the smallest norm.

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \|\vec{x}\|$$
$$s.t. \quad A\vec{x} = \vec{b}$$

i.e., we want to find the projection of $\vec{0}$ on $S = \{\vec{x} \in \mathbb{R}^n : A\vec{x} = \vec{b}\}$

2.5.1 Applying the least-squares technique

The solution set $S = \{\vec{x} \in \mathbb{R}^n : A\vec{x} = \vec{b}\}$ which is an example of an **affine subspace**, i.e., a vector subspace of $\mathbb{R}^n$ that does not contain $\vec{0}$.

We may write $S = S' + \vec{x}_0$, where:

- $\vec{x}_0$ is an arbitrary element of S (s.t. $A\vec{x}_0 = \vec{b}$)
- $S' = \{\vec{y} \in \mathbb{R}^n : A\vec{y} = \vec{0}\}$

"Finding the point in S closest to $\vec{0}$" is equivalent to "Finding the point in S' closest to $-\vec{x}_0$"

Example:

$$\min_{\vec{x}\in\mathbb{R}^4} \quad \|\vec{x}\|$$
$$s.t. \quad 2x_1 - x_2 + x_3 - x_4 = 3$$
$$x_2 - x_3 - x_4 = 1$$

$x_2 = 1 + x_3 + x_4 \Rightarrow 2x_1 - 1 - x_3 - x_4 + x_3 - x_4 = 3 \Rightarrow x_1 - x_4 = 2.$

The general solution has the form

$$\vec{x} = \begin{bmatrix} 2 + x_4 \\ 1 + x_3 + x_4 \\ x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} x_4$$

Let $\vec{x}_0 = [2, 1, 0, 0]^T$, the solution set can be $S = S' + \vec{x}_0$ where S' is the set of linear combinations of $[0, 1, 1, 0]^T$ and $[1, 1, 0, 1]^T$.

Then, the problem becomes

$$\min_{(x_3,x_4)\in\mathbb{R}^2} \left\| \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} - \begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \end{bmatrix} \right\|$$

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$. From what we have already known, we can solve this problem by solving

$$A^T AX = A^T(-x_0) \Leftrightarrow \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} X = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Then, we can solve $(x_3, x_4) = (0, -1)$, then the optimal solution is $(1, 0, 0, -1)$.

2.5.2 The short cut method

We can find that $S = \{\vec{x} \in \mathbb{R}^n : A\vec{x} = \vec{b}\}$ and $S' = \{\vec{y} \in \mathbb{R}^n : A\vec{y} = \vec{0}\}$ are parallel. Then the vector $\vec{x}^* - \vec{0}$ (i.e. $\vec{x}^*$) should be perpendicular to S'.

Lemma 2.2

A vector $\vec{x}^*$ satisfying $A\vec{x}^* = \vec{b}$ is the minimum-norm solution to the system of equations $A\vec{x} = \vec{b}$ if and only if $\vec{x}^* \cdot \vec{y} = 0$ for all solutions $\vec{y}$ of the homogeneous system $A\vec{y} = \vec{0}$.

Obviously, **all** vectors in null space ($N(A) = \{\vec{y} : A\vec{y} = \vec{0}\}$) are orthogonal to a vector **if and only if** it is a linear combination of A's rows. $\vec{x}^* = A^T\vec{w}$, for some $\vec{w} \in \mathbb{R}^n$. $(\vec{x}^* \cdot (\vec{y}) = (\vec{x}^*)^T\vec{y} = \vec{w}^T A\vec{y} = 0)$

Theorem 2.4

A vector $\vec{x}^*$ satisfying $A\vec{x}^* = \vec{b}$ is the minimum-norm solution to the system of equations $A\vec{x} = \vec{b}$ if and only if it can be written as $\vec{x}^* = A^T\vec{w}$ for some $\vec{w} \in \mathbb{R}^n$

Hence, we cam find the minimum-norm solution $\vec{x}^*$ by solving

$$AA^T\vec{w} = \vec{b}$$

for $\vec{w}$ and then computing $\vec{x}^* = A^T\vec{w}$.

Same Example:

$$\min_{\vec{x}\in\mathbb{R}^4} \quad \|\vec{x}\|$$

$$s.t. \quad 2x_1 - x_2 + x_3 - x_4 = 3$$

$$x_2 - x_3 - x_4 = 1$$

Solve the system

$$AA^T\vec{w} = \vec{b}$$

$$\begin{bmatrix} 2 & -1 & 1 & -1 \\ 0 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 7 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

We can solve $(w_1, w_2) = (\frac{1}{2}, \frac{1}{2})$, then $\vec{x}^* = A^T\vec{w} = [1, 0, 0, -1]^T$

2.5.3 The short cut method with H-norm

Definition 2.2

Given a positive definite (symmetric) matrix H, let the associated inner product be $\vec{x} \cdot_H \vec{y} = \vec{x}^T H\vec{y}$ and the associated norm be $\|\vec{x}\|_H = \sqrt{\vec{x}^T H\vec{x}}$ ♣

Solve the optimization problem

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \|\vec{x}\|_H^2 = \vec{x}^T H\vec{x}$$

$$s.t. \quad A\vec{x} = \vec{b}$$

Lemma 2.3

*A point $\vec{x}^*$ that satisfies $A\vec{x}^* = \vec{b}$ is the **minimum-H-norm solution** to $A\vec{x} = \vec{b}$ if and only if*

$$\vec{x}^* \cdot_H \vec{y} = 0$$

for all $\vec{y}$ for which $A\vec{y} = \vec{0}$ ♡

Proof 2.5

We proved that the $\vec{x}^*$ being perpendicular with $\{y \in \mathbb{R}^n : A\vec{y} = \vec{0}\}$ is equivalent to $\vec{x}^*$ is the minimum-norm solution to $A\vec{x}^* = \vec{b}$

A vector that is orthogonal to $\{y \in \mathbb{R}^n : A\vec{y} = \vec{0}\}$ if and only if $\vec{x}^* = H^{-1}A^T\vec{w}$, for some $\vec{w}$. ($\vec{x}^* \cdot_H \vec{y} =$

$(\vec{x}^*)^T H\vec{y} = \vec{w}^T AH^{-1}H\vec{y} = 0)$

The same as the short cut method we can solve the $\vec{x}^*$ by computing $AH^{-1}A^T\vec{w} = \vec{b}$

Theorem 2.5

*The **minimum-H-norm solution** $\vec{x}^*$ of the Underconstrainted system $A\vec{x} = \vec{b}$ can be found by sovling*

$$AH^{-1}A^T\vec{w} = \vec{b}$$

for $\vec{w}$ and then computing $\vec{x}^* = H^{-1}A^T\vec{w}$

Example 2.1

$$\min_{(x,y)\in\mathbb{R}^2} 3x^2 + 2xy + 2y^2$$

$$s.t. \quad 3x - y = 3$$

$3x^2 + 2xy + 2y^3$ is the square of the H-norm of the point $\begin{bmatrix} x \\ y \end{bmatrix}$ for the matrix $H = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. We can also compute $H^{-1} = \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}$.
Then, we can solve $\vec{x}^*$ by computing

$$AH^{-1}A^T\vec{w} = \vec{b}$$

$$\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.6 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \vec{w} = 3$$

We can compute $\vec{w} = \frac{5}{9}$, then the optimal solution is $\begin{bmatrix} x \\ y \end{bmatrix} = H^{-1}A^T\vec{w} = \begin{bmatrix} \frac{7}{9} \\ -\frac{2}{3} \end{bmatrix}$