Determinant

The determinant of $\mathbf{A} \in \mathbb{R}^{m \times m}$, denoted by $\det(\mathbf{A})$, is defined by induction

- For $m = 1$: $\det(\mathbf{A}) = a_{11}$ $\quad A = a_{11}$
- For $m \geq 2$:
  - Let $\mathbf{A}_{ij} \in \mathbb{R}^{(m-1)\times(m-1)}$ be a submatrix of $\mathbf{A}$ obtained by deleting the ith row and jth column of $\mathbf{A}$
  - Let $c_{ij} = (-1)^{i+j} \det(\mathbf{A}_{ij})$
  - Cofactor expansion:

$$\det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij} c_{ij}, \quad \text{for any } i = 1, \ldots, m$$

$$\det(\mathbf{A}) = \sum_{i=1}^{m} a_{ij} c_{ij}, \quad \text{for any } j = 1, \ldots, m$$

where c_{ij}'s are the cofactors and $\det(\mathbf{A}_{ij})$'s are the minors

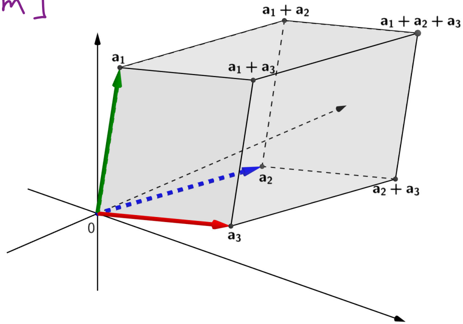
Determinant (Cont'd)

$\Longleftrightarrow$ A singular

- **Fact**: $\mathbf{Ax} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$ if and only if $\det(\mathbf{A}) = 0$

- Interpretation: $|\det(\mathbf{A})|$ is the volume of the parallelepiped $\mathcal{P} = \{\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{a}_i \mid \alpha_i \in [0,1] \; \forall i = 1, \ldots, m\}$

$A = [a_1 \cdots a_m]$



Determinant (Cont'd)

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times m}$

- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ $= \det(B)\det(A) = \det(BA)$
- $\det(\mathbf{A}) = \det(\mathbf{A}^T)$
- $\det(\alpha\mathbf{A}) = \alpha^m \det(\mathbf{A})$ for any $\alpha \in \mathbb{R}$
- $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ for any nonsingular $\mathbf{A}$

  $\det(A) \cdot \det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1$
- $\det(\mathbf{B}^{-1}\mathbf{AB}) = \det(\mathbf{A})$ for any nonsingular $\mathbf{B}$

  $= \det(B^{-1})\det(A)\det(B) = \det(B^{-1})\det(B)\det(A)$
- $\mathbf{A}^{-1} = \dfrac{1}{\det(\mathbf{A})}\tilde{\mathbf{A}}$, where $\tilde{a}_{ij} = c_{ji}$ (the cofactor) for all i, j ($\mathbf{A}$ is nonsingular)
  - $\tilde{\mathbf{A}}$ is the adjoint or adjugate matrix of $\mathbf{A}$

Determinant (Cont'd)

- If $\mathbf{A} \in \mathbb{R}^{m \times m}$ is triangular, either upper or lower,

$$\det(\mathbf{A}) = \prod_{i=1}^{m} a_{ii}$$

  - Proof: Apply cofactor expansion inductively

- If $\mathbf{A} \in \mathbb{R}^{m \times m}$ is *block* upper or lower triangular

$$\mathbf{A} = \begin{bmatrix} \boxed{\mathbf{B}} & \mathbf{C} \\ \mathbf{0} & \boxed{\mathbf{D}} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

where $\mathbf{B}$ and $\mathbf{D}$ are square (and can be of different sizes), then

$$\det(\mathbf{A}) = \det(\mathbf{B})\det(\mathbf{D})$$

Vector Norms

A function $f : \mathbb{R}^n \to \mathbb{R}$ is a vector norm if all of the following hold:

1. $f(\mathbf{x}) \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$
2. $f(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$
3. $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
4. $f(\alpha \mathbf{x}) = |\alpha| f(\mathbf{x})$ for any $\alpha \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$

- Usually $\| \cdot \|$ denotes a norm
- $\|\mathbf{x}\|$ represents the "length" of vector $\mathbf{x}$
- $\|\mathbf{x} - \mathbf{y}\|$ represents the "distance" of vectors $\mathbf{x}, \mathbf{y}$

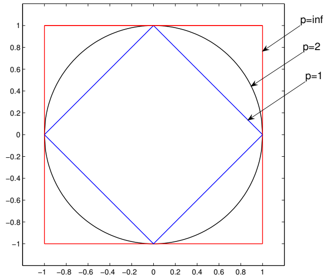
Vector Norms (Cont'd)

Examples:

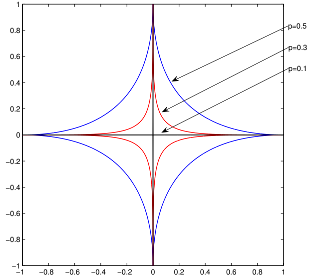
- 2-norm or Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = (\mathbf{x}^T\mathbf{x})^{1/2}$
- 1-norm or Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- ∞-norm: $\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$
- p-norm, $p \geq 1$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

ℓ_p Function

$$f_p(\mathbf{x}) = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \qquad p > 0$$



(a) Region of $f_p(\mathbf{x}) = 1$, $p \geq 1$.



(b) Region of $f_p(\mathbf{x}) = 1$, $p \leq 1$.

- Note that f_p is *not* a norm for $0 < p < 1$
- when $p \to 0$, f_p is like the cardinality function $\mathrm{card}(\mathbf{x}) = \sum_{i=1}^{n} \mathbb{1}\{x_i \neq 0\}$, where $\mathbb{1}\{x \neq 0\} = 1$ if $x \neq 0$ and $\mathbb{1}\{x \neq 0\} = 0$ if $x = 0$

Inner Product

The inner product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} y_i x_i = \mathbf{y}^T \mathbf{x} = x^T y = \langle y, x \rangle$$

- $\mathbf{x}, \mathbf{y}$ are said to be orthogonal to each other if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
- $\mathbf{x}, \mathbf{y}$ are said to be parallel if $\mathbf{x} = \alpha \mathbf{y}$ for some α
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \pm \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$ for parallel $\mathbf{x}, \mathbf{y}$

The angle between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\theta = \cos^{-1} \left(\frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \right)$$

- $\mathbf{x}, \mathbf{y}$ are orthogonal if $\theta = \pm \pi/2$
- $\mathbf{x}, \mathbf{y}$ are parallel if $\theta = 0$ or $\theta = \pm \pi$

Hölder Inequality

Hölder Inequality: For any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q$$

Proof. **Young's Inequality**: For any $a, b \geq 0$ and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$x=0$ or $y=0$ trivial

Let $x, y \neq 0$.

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Let $a_i = \frac{|x_i|}{\|x\|_p}$ $b_i = \frac{|y_i|}{\|y\|_q}$ $i = 1, \cdots, n$.

$$\sum_{i=1}^n a_i b_i = \sum_{i=1}^n \frac{|x_i|}{\|x\|_p} \cdot \frac{|y_i|}{\|y\|_q} \leq \sum_{i=1}^n \frac{|x_i|^p}{\|x\|_p^p \cdot p} + \sum_{i=1}^n \frac{|y_i|^q}{\|y\|_q^q\, q}$$

$$= \frac{1}{\|x\|_p^p \cdot p} \underbrace{\sum_{i=1}^n |x_i|^p}_{\|x\|_p^p} + \frac{1}{\|y\|_q^q\, q} \underbrace{\sum_{i=1}^n |y_i|^q}_{\|y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1$$

$$\Rightarrow 1 \geq \sum_{i=1}^n a_i b_i = \frac{\sum_{i=1}^n |x_i y_i|}{\|x\|_p \cdot \|y\|_q} \geq \frac{|\sum_{i=1}^n x_i y_i|}{\|x\|_p \cdot \|y\|_q} = \frac{|x^T y|}{\|x\|_p \|y\|_q}$$

Hölder Inequality (Cont'd)

Hölder Inequality: For any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q$$

$\langle x, y \rangle$

$\shortparallel$

$x^T y = \|x\|_2 \|y\|_2 \cos\theta$

- Cauchy-Schwartz Inequality: Let $p = q = 2$ in Hölder Inequality

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

where the equality holds if and only if $\mathbf{x} = \alpha\mathbf{y}$ for some $\alpha \in \mathbb{R}$

- Hölder Inequality holds for $p = 1$ and $q = \infty$

$$|\mathbf{x}^T\mathbf{y}| \leq \sum_{i=1}^n |x_i y_i| \leq \max_j |y_j| (\sum_{i=1}^n |x_i|) = \|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty.$$

Equivalence of Norms

All norms on $\mathbb{R}^n$ are equivalent in the sense that if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are norms on $\mathbb{R}^n$, then there exist $c1, c_2 > 0$ such that

$$c_1\|\mathbf{x}\|_\alpha \le \|\mathbf{x}\|_\beta \le c_2\|\mathbf{x}\|_\alpha, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

- $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2$
- $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{n}\|\mathbf{x}\|_\infty$
- $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_1 \le n\|\mathbf{x}\|_\infty$

Projections on Subspaces

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set (not necessarily a subspace)
Given $\mathbf{y} \in \mathbb{R}^m$, a projection of $\mathbf{y}$ onto $\mathcal{S}$ is any solution to

$$\min_{\mathbf{z} \in \mathcal{S}} \|\mathbf{z} - \mathbf{y}\|_2^2$$

- a point in $\mathcal{S}$ that is closest to $\mathbf{y}$
    - Projection of $\mathbf{y} \in \mathcal{S}$ onto $\mathcal{S}$ is $\mathbf{y}$ itself
- If for *any* $\mathbf{y} \in \mathbb{R}^m$, there always exists a *unique* projection of $\mathbf{y}$ onto $\mathcal{S}$, then we denote

$$\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z} \in \mathcal{S}} \|\mathbf{z} - \mathbf{y}\|_2^2$$

and $\Pi_{\mathcal{S}}$ is called the projection (or projection operator) of $\mathbf{y}$ onto $\mathcal{S}$

Projection Theorem

Theorem (Projection Theorem)

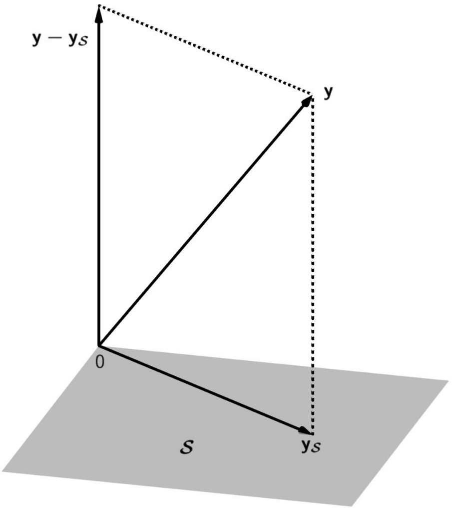
Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$.

1. *For any $\mathbf{y} \in \mathbb{R}^m$, there exists a unique vector $\mathbf{y}_s \in \mathcal{S}$ that minimizes $\|\mathbf{z} - \mathbf{y}\|_2^2$ over $\mathbf{z} \in \mathcal{S}$ (so that we can use the notation $\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z}\in\mathcal{S}} \|\mathbf{z} - \mathbf{y}\|_2^2$).*
2. *Given $\mathbf{y} \in \mathbb{R}^m$,*

$$\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}) \iff \mathbf{y}_s \in \mathcal{S}, \quad \mathbf{z}^T(\mathbf{y}_s - \mathbf{y}) = 0 \text{ for all } \mathbf{z} \in \mathcal{S}.$$

- Statement 1 of Projection Theorem also holds for closed convex set (more general)
  - Very important to convex optimization

Projection Theorem (Cont'd)



Orthogonal Complement

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set

The orthogonal complement of $\mathcal{S}$ is defined as

$$\mathcal{S}^\perp = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{z}^T\mathbf{y} = 0 \text{ for all } \mathbf{z} \in \mathcal{S}\} \ni 0$$

- $\mathcal{S}^\perp$ is a subspace (Why?) no matter S is subspace or not
- Any $\mathbf{z} \in \mathcal{S}$ and any $\mathbf{y} \in \mathcal{S}^\perp$ are orthogonal
- Either $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$ or $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$

  $\rightarrow 0 \in S$
- **Facts**:
  - $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$
  - $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp$
  - Recall that range and nullspace of a matrix are subspaces

For any $x \in \mathcal{N}(A^T)$ and $y \in \mathcal{R}(A)$,

$A^T x = 0$ $\quad\quad \exists z$ s.t. $y = Az$.

$\langle y, x \rangle = y^T x = (Az)^T x = z^T A^T x = 0$

Orthogonal Complement of Subspace

Theorem

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. For any $\mathbf{y} \in \mathbb{R}^m$, there uniquely exists $(\mathbf{y}_s, \mathbf{y}_c) \in \mathcal{S} \times \mathcal{S}^\perp$ such that

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c.$$

In particular, $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}), \mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y}) = \Pi_{\mathcal{S}^\perp}(\mathbf{y})$.

- Proof sketch: From Statement 2 of the Projection Theorem,

$$\mathbf{y}_s \in \mathcal{S},\ \mathbf{y} - \mathbf{y}_s \in \mathcal{S}^\perp \quad \Longleftrightarrow \quad \mathbf{y}_s \in \Pi_{\mathcal{S}}(\mathbf{y})$$

Orthogonal Complement of Subspace (Cont'd)

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. It follows from the above theorem that

- $\mathcal{S} + \mathcal{S}^\perp = \mathbb{R}^m$
- $\dim \mathcal{S} + \dim \mathcal{S}^\perp = m$
  - Proof: $\dim \mathcal{S} + \dim \mathcal{S}^\perp = \dim(\mathcal{S} + \mathcal{S}^\perp) + \dim(\mathcal{S} \cap \mathcal{S}^\perp) = \dim(\mathcal{S} + \mathcal{S}^\perp) + 0 = \dim \mathbb{R}^m$
- $(\mathcal{S}^\perp)^\perp = \mathcal{S}$

$\dim \mathcal{R}(A^T) + \dim \mathcal{N}(A) = n$

Example: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$

- $\dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{R}(\mathbf{A})^\perp = \dim \mathcal{R}(\mathbf{A}) + \dim \mathcal{N}(\mathbf{A}^T) = m$
- **Rank-Nullity Theorem**: $\dim \mathcal{N}(\mathbf{A}) = n - \dim \mathcal{R}(\mathbf{A}^T) = n - \mathrm{rank}(\mathbf{A})$

Orthogonal and Orthonormal Vectors

A collection of *nonzero* vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are said to be

- orthogonal if $\mathbf{a}_i^T \mathbf{a}_j = 0$ for all i, j with $i \neq j$
- orthonormal if they are orthogonal and $\|\mathbf{a}_i\|_2 = 1$ for all i

Same definition applies to complex $\mathbf{a}_i$'s by replacing transpose (T) with Hermitian transpose (H)

$e_i = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

Example: Any vectors from $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ are orthonormal and $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset \mathbb{R}^m$ is an orthonormal basis for $\mathbb{R}^m$

Orthonormal vectors are linearly independent

orthogonal

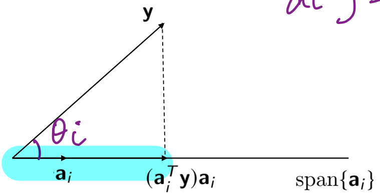
Orthogonal and Orthonormal Vectors (Cont'd)

Fact: Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be an orthonormal set of vectors and $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then, the coefficient α for the representation

$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$

is uniquely given by $\alpha_i = \mathbf{a}_i^T \mathbf{y}$, $i = 1, \ldots, n$

$$a_i^T y = \overset{=1}{\|a_i\|_2} \cdot \underbrace{\|y\|_2 \cdot \cos\theta_i}$$



Fact: Every subspace $\mathcal{S}$ with $\mathcal{S} \neq \{\mathbf{0}\}$ has an orthonormal basis

- It can be shown using Gram-Schmidt

Orthogonal Matrix

A real matrix $\mathbf{Q}$ is said to be

- orthogonal if it is square and its columns are orthonormal
- semi-orthogonal if its columns are orthonormal
  - a semi-orthogonal $\mathbf{Q}$ must be tall or square

A complex matrix $\mathbf{Q}$ is said to be unitary if it is square and its columns are orthonormal, and semi-unitary if its columns are orthonormal

Example: Consider the transformation $\mathbf{y} = \mathbf{Q}\mathbf{x}$ with

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad \text{rotation counterclock-wise by } \theta \in [0, 2\pi)$$

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \quad \text{reflection about the } \theta/2 \text{ line, } \theta \in [0, 2\pi)$$

The rotation and reflection matrices are orthogonal

Orthogonal Matrix (Cont'd)

Facts:

- $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ and $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ for orthogonal $\mathbf{Q}$

$Q^T = Q^{-1}$

- $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ (but *not* necessarily $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) for semi-orthogonal $\mathbf{Q}$

- $\|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for orthogonal $\mathbf{Q}$
  - For example, rotation and reflection do not change the vector length

- For any tall and semi-orthogonal matrix $\mathbf{Q}_1 \in \mathbb{R}^{n\times k}$, there exists a matrix $\mathbf{Q}_2 \in \mathbb{R}^{n\times(n-k)}$ such that $[\ \mathbf{Q}_1\ \mathbf{Q}_2\ ]$ is orthogonal

Matrix Product Representations

Let $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$. Consider

$$\mathbf{C} = \mathbf{A}\mathbf{B}$$

- Column representation:

  $$[c_1 \cdots c_n] = A[b_1 \cdots b_n]$$

  $$\mathbf{c}_i = \mathbf{A}\mathbf{b}_i, \quad i = 1, \dots, n$$

  where $\mathbf{c}_i$ and $\mathbf{b}_i$ are the ith columns of $\mathbf{C}$ and $\mathbf{B}$

- Inner-product representation: Let $\tilde{\mathbf{a}}_i^T \in \mathbb{R}^{1\times k}$ be the ith row of $\mathbf{A}$

  $$\mathbf{A}\mathbf{B} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T\mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_1^T\mathbf{b}_n \\ \vdots & & \vdots \\ \tilde{\mathbf{a}}_m^T\mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_m^T\mathbf{b}_n \end{bmatrix}$$

  Thus,

  $$c_{ij} = \tilde{\mathbf{a}}_i^T\mathbf{b}_j, \quad \text{for all } i, j$$