

Om Gram-matriser

Gemini AI Pro 2.5

Q:

Linear Algebra and Gram Matrix. Please give an overview of what the Gram Matrix is, how it is constructed. I know it has some useful properties, but lack the full overview of this. Is it possible to explain why the Gram Matrix has these properties? I am also interested in some use cases for the Gram Matrix.

A:

Here is an overview of the Gram matrix, its construction, properties, and applications.

What is a Gram Matrix?

The Gram matrix (or Gramian matrix) is a fundamental tool in linear algebra that captures the **geometric relationships** between a set of vectors. At its core, a Gram matrix is a square matrix where each entry G_{ij} is the **inner product** (or dot product) of the i -th and j -th vectors in a given set.

This matrix essentially encodes all information about the lengths of the vectors and the angles between them.

How to Construct a Gram Matrix

There are two primary ways to think about its construction:

1. The Inner Product Definition

Given a set of n vectors $\{v_1, v_2, \dots, v_n\}$ in an inner product space (like the standard Euclidean space \mathbb{R}^m), the Gram matrix G is an $n \times n$ matrix where the entry in the i -th row and j -th column is the inner product $\langle v_i, v_j \rangle$.

$$G = \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \dots & \langle v_1, v_n \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & \dots & \langle v_2, v_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \langle v_n, v_2 \rangle & \dots & \langle v_n, v_n \rangle \end{pmatrix}$$

- The diagonal entries $G_{ii} = \langle v_i, v_i \rangle = \|v_i\|^2$ are the **squared lengths** of the vectors.
- The off-diagonal entries $G_{ij} = \langle v_i, v_j \rangle$ relate to the **angle** between vectors v_i and v_j .

2. The Matrix Multiplication Definition

This is the most common way to compute the Gram matrix in practice.

1. Create a matrix A by placing your vectors v_1, \dots, v_n as its **columns**.
- If each v_i is in \mathbb{R}^m , then A will be an $m \times n$ matrix.

$$A = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix}$$

2. The Gram matrix G is simply the product of the transpose of A with A :

$$G = A^T A$$

Why this works: The (i, j) entry of G is (row i of A^T) \times (column j of A).

- Row i of A^T is just the vector v_i^T .
- Column j of A is the vector v_j .
- Therefore, $G_{ij} = v_i^T v_j$, which is the standard dot product $\langle v_i, v_j \rangle$.

Key Properties and *Why* They Hold

The Gram matrix has several powerful properties that stem directly from its construction.

1. It is Symmetric

- **Property:** The Gram matrix G is always symmetric, meaning $G = G^T$.
- **Why?**
 - **Reason 1 (Inner Product):** The definition of a real inner product is symmetric: $\langle v_i, v_j \rangle = \langle v_j, v_i \rangle$. Therefore, the entry G_{ij} is always equal to the entry G_{ji} .

- **Reason 2 (Matrix):** The transpose of a product $(XY)^T$ is $Y^T X^T$. Applying this to $G = A^T A$:

$$G^T = (A^T A)^T = A^T (A^T)^T = A^T A = G$$

2. It is Positive Semi-Definite

- **Property:** G is always **positive semi-definite** (PSD). This is its most important property. It means that for *any* vector x (of size $n \times 1$), the resulting scalar $x^T G x$ is always non-negative ($x^T G x \geq 0$).
- **Why?** This has a very elegant proof using the $G = A^T A$ construction:
 1. Start with the expression $x^T G x$.
 2. Substitute $G = A^T A$:

$$x^T G x = x^T (A^T A) x$$

3. Use the associativity of matrix multiplication to regroup:

$$x^T (A^T A) x = (x^T A^T) (A x)$$

4. Recognize that $(A x)^T = x^T A^T$:

$$(x^T A^T) (A x) = (A x)^T (A x)$$

5. Let's define a new vector $v = A x$. The expression is now $v^T v$.
6. $v^T v$ is just the dot product of v with itself, which is the **squared norm** (or squared length) of v : $v^T v = \|v\|^2$.
7. The squared length of *any* real vector is, by definition, greater than or equal to zero ($\|v\|^2 \geq 0$).
8. Therefore, $x^T G x = \|A x\|^2 \geq 0$ for all x . This is the definition of a positive semi-definite matrix.

3. It Detects Linear Independence

- **Property:** G is **positive definite** (PD), which is a stronger condition (meaning $x^T G x > 0$ for all $x \neq 0$), *if and only if* the vectors v_1, \dots, v_n are **linearly independent**.
- **Why?**
 1. We continue from the proof above: $x^T G x = \|A x\|^2$.
 2. For G to be positive *definite*, we need $x^T G x = 0$ *only* when $x = 0$.
 3. This means we need $\|A x\|^2 = 0$ *only* when $x = 0$.
 4. The norm $\|A x\|$ is zero if and only if the vector $A x$ is the zero vector: $A x = 0$.
 5. What does $A x = 0$ mean? If $x = (c_1, \dots, c_n)^T$, then $A x$ is the linear combination $c_1 v_1 + c_2 v_2 + \dots + c_n v_n$.
 6. So, the condition for G to be positive definite is: $c_1 v_1 + \dots + c_n v_n = 0$ is true *only when* all $c_i = 0$ (i.e., $x = 0$).

7. This is precisely the **definition of linear independence** for the vectors v_1, \dots, v_n .
- **Corollary:** A matrix is invertible if and only if its determinant is non-zero. A positive definite matrix always has a positive determinant. Therefore, G is invertible if and only if the vectors $\{v_i\}$ are linearly independent.

4. Its Determinant is a (Squared) Volume

- **Property:** The determinant of G , $\det(G)$, is called the **Gram determinant**. It is equal to the **squared n -dimensional volume** of the parallelepiped (or parallelotope) spanned by the vectors v_1, \dots, v_n .
- **Why?**
 - **Simple Case:** If we have 2 vectors v_1, v_2 in \mathbb{R}^2 , they form a parallelogram.
 - * $\det(G) = \det \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{pmatrix} = \|v_1\|^2 \|v_2\|^2 - \langle v_1, v_2 \rangle^2$
 - * Using the dot product formula $\langle v_1, v_2 \rangle = \|v_1\| \|v_2\| \cos \theta$:
 - * $\det(G) = \|v_1\|^2 \|v_2\|^2 - (\|v_1\| \|v_2\| \cos \theta)^2$
 - * $\det(G) = \|v_1\|^2 \|v_2\|^2 (1 - \cos^2 \theta) = \|v_1\|^2 \|v_2\|^2 \sin^2 \theta$
 - * The area of a parallelogram is base \times height $= \|v_1\| (\|v_2\| \sin \theta)$.
 - * Therefore, $\det(G) = (\text{Area of parallelogram})^2$.
 - This property generalizes to n dimensions. If the vectors are linearly dependent (Property 3), they are “flat” and span a parallelepiped with zero volume. This lines up perfectly: $\det(G) = 0$ if and only if the volume is 0.

Use Cases for the Gram Matrix

1. **Checking Linear Independence:** This is a classic application. Given a set of vectors, compute their Gram matrix G and find its determinant.
 - If $\det(G) > 0$, the vectors are **linearly independent**.
 - If $\det(G) = 0$, the vectors are **linearly dependent**.
2. **Calculating Volumes:** As seen above, if you need the volume of the parallelepiped spanned by n vectors, you can compute $V = \sqrt{\det(G)}$.
3. **Machine Learning: The Kernel Trick** This is one of the most powerful modern applications of the Gram matrix.
 - Many algorithms, like **Support Vector Machines (SVMs)**, depend only on the dot products of data points, not the data points themselves.
 - We can map our data x to a much higher-dimensional (even infinite-dimensional) space using a function $\phi(x)$.
 - In this new space, the algorithm would need to compute inner products $\langle \phi(x_i), \phi(x_j) \rangle$. This is computationally expensive or impossible.

- The “**kernel trick**” is to use a *kernel function* $k(x_i, x_j)$ that computes this high-dimensional inner product *directly* from the original data points: $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$.
- The resulting matrix K , where $K_{ij} = k(x_i, x_j)$, is a **Gram matrix** (often called a *kernel matrix*). The fact that it’s positive semi-definite (by Mercer’s theorem) is crucial for guaranteeing that the optimization problem in the SVM is convex and solvable.

4. Solving Linear Least Squares:

- To find the best solution to an overdetermined system $Ax = b$ (where A is $m \times n$ and $m > n$), we often solve the **normal equations**:

$$(A^T A)x = A^T b$$

- The matrix $G = A^T A$ is the Gram matrix of the columns of A . This transforms the problem from an $m \times n$ system to a smaller $n \times n$ system. Because G is symmetric and (if A ’s columns are independent) positive definite, this system can be solved efficiently (e.g., using Cholesky decomposition).
5. **Finite Element Method (FEM):** In engineering and physics, when solving partial differential equations numerically, the “stiffness matrix” and “mass matrix” are often Gram matrices. They are formed from the inner products of a set of basis functions (like polynomials or splines) used to approximate the solution. The positive definite property is essential for ensuring a unique, stable solution exists.

Q:

Please explain the relationship between the dot product and the angle between the vectors. And based on this explain the entries in the Gram matrix.

A:

Here’s a breakdown of the relationship between the dot product and angles, and how that directly explains the entries of a Gram matrix.

1. The Dot Product and the Angle

The dot product (also called the inner product) is an operation that takes two vectors and returns a single number (a scalar). This number has a powerful geometric meaning.

For two vectors, v and w , the dot product $\langle v, w \rangle$ is defined by the geometric formula:

$$\langle v, w \rangle = \|v\| \|w\| \cos(\theta)$$

Where: $\|v\|$ is the length (or norm) of vector v . $\|w\|$ is the length of vector w . θ is the angle between v and w . $\cos(\theta)$ is the cosine of that angle.

This formula tells us that the dot product encodes both the lengths of the vectors and the angle between them.

What the Dot Product's Value Means:

We can interpret the relationship by looking at the value of $\cos(\theta)$:

- **If $\langle v, w \rangle = 0$:** Since the lengths $\|v\|$ and $\|w\|$ are (usually) positive, this means $\cos(\theta) = 0$. This only happens when the angle $\theta = 90^\circ$.
Conclusion: A zero dot product means the vectors are **orthogonal** (perpendicular).
- **If $\langle v, w \rangle > 0$ (Positive):** This means $\cos(\theta) > 0$. This happens when the angle θ is **acute** (less than 90°). The vectors point generally in the same direction.
- **If $\langle v, w \rangle < 0$ (Negative):** This means $\cos(\theta) < 0$. This happens when the angle θ is **obtuse** (greater than 90°). The vectors point generally in opposite directions.

2. Explanation of Gram Matrix Entries

Now, let's apply this to the Gram matrix.

Recall that the Gram matrix G for a set of vectors $\{v_1, v_2, \dots, v_n\}$ is defined by its entries:

$$G_{ij} = \langle v_i, v_j \rangle$$

Based on the dot product formula, we can explain every single entry in the matrix. We just need to split them into two cases:

Case 1: The Diagonal Entries (where $i = j$)

These are the entries $G_{11}, G_{22}, \dots, G_{nn}$.

- **Entry:** $G_{ii} = \langle v_i, v_i \rangle$
- **Angle:** The angle between a vector v_i and itself is $\theta = 0^\circ$.
- **Cosine:** $\cos(0^\circ) = 1$.

Let's plug this into the formula:

$$G_{ii} = \langle v_i, v_i \rangle = \|v_i\| \|v_i\| \cos(0^\circ) = \|v_i\| \|v_i\| \cdot 1 = \|v_i\|^2$$

Conclusion: The diagonal entries of a Gram matrix are the **squared lengths** of the vectors.

Case 2: The Off-Diagonal Entries (where $i \neq j$)

These are all the other entries, like G_{12}, G_{13}, G_{21} , etc.

- **Entry:** $G_{ij} = \langle v_i, v_j \rangle$
- **Angle:** The angle is θ_{ij} , the angle between vector v_i and vector v_j .

Plugging this into the formula gives:

$$G_{ij} = \|v_i\| \|v_j\| \cos(\theta_{ij})$$

Conclusion: The off-diagonal entries G_{ij} encode the **geometric relationship between pairs of different vectors**. The value is a combination of their two lengths and the angle between them.

- If $G_{ij} = 0$, you know v_i and v_j are **orthogonal**.
- If $G_{ij} > 0$, they form an **acute** angle.
- If $G_{ij} < 0$, they form an **obtuse** angle.

Summary

The Gram matrix is a compact “lookup table” of all pairwise geometric relationships in a set of vectors.

$$G = \begin{bmatrix} \text{Length of } v_1 \text{ squared} & \text{Rel'ship of } v_1 \text{ and } v_2 & \dots & \text{Rel'ship of } v_1 \text{ and } v_n \\ \text{Rel'ship of } v_2 \text{ and } v_1 & \text{Length of } v_2 \text{ squared} & \dots & \text{Rel'ship of } v_2 \text{ and } v_n \\ \vdots & \vdots & \ddots & \vdots \\ \text{Rel'ship of } v_n \text{ and } v_1 & \text{Rel'ship of } v_n \text{ and } v_2 & \dots & \text{Length of } v_n \text{ squared} \end{bmatrix}$$

(Note: $G_{ij} = G_{ji}$ because $\langle v_i, v_j \rangle = \langle v_j, v_i \rangle$, which is why the matrix is symmetric.)

Q:

In a matrix where some of the columns are linearly dependent, and others are not - is the determinant of the Gram matrix still zero? Is this universal?

A:

Yes, the determinant of the Gram matrix is **zero**. This is a **universal** property.

The determinant of the Gram matrix tests the linear independence of the **entire set** of vectors used to build it. It doesn't matter if *some* of the vectors are linearly independent *from each other*. If the *collection as a whole* is linearly dependent, the determinant is zero.

Here's the simple rule and a more intuitive explanation:

The Rule

Let G be the Gram matrix of the vectors $\{v_1, v_2, \dots, v_n\}$.

- $\det(G) = 0 \iff$ The set of vectors $\{v_1, \dots, v_n\}$ is **linearly dependent**.
- $\det(G) > 0 \iff$ The set of vectors $\{v_1, \dots, v_n\}$ is **linearly independent**.

(Note: The determinant can't be negative because the Gram matrix is positive semi-definite).

The Intuitive Explanation: “The Flat Box”

The easiest way to think about this is using the **volume analogy**.

The determinant of the Gram matrix, $\det(G)$, is equal to the **squared n -dimensional volume** of the parallelepiped (a “slanted box”) spanned by your n vectors.

Let's see what this means for your specific case. Imagine you have 3 vectors in 3D space: v_1 , v_2 , and v_3 .

Case 1: All Linearly Independent Your vectors v_1, v_2, v_3 all point in different directions (none lies in the plane of the others). They form a proper 3D parallelepiped with a real, non-zero volume. * **Volume** > 0 * $\det(G) = (\text{Volume})^2 > 0$

Case 2: Your Scenario (Linearly Dependent) You state that “some are dependent, and others are not.” Let's say v_1 and v_2 are independent (they point in different directions and form a 2D parallelogram with a non-zero area).

But, the *set* is dependent, which means v_3 is a linear combination of the others (e.g., $v_3 = c_1 v_1 + c_2 v_2$).

This means v_3 **lies in the exact same plane** as v_1 and v_2 .

What “box” do these three vectors form? They form a **perfectly flat shape**. You have a 2D parallelogram (from v_1 and v_2) and another vector v_3 that just lies on top of it.

What is the 3D volume of a flat 2D shape? **It's zero.**

- **Volume** $= 0$
- $\det(G) = (\text{Volume})^2 = 0^2 = 0$

This is universal. If you have 10 vectors, but one of them (say, v_{10}) can be written as a combination of the others, your 10-dimensional “box” is flattened into a 9-dimensional shape. It has zero 10-dimensional volume, and thus, $\det(G) = 0$.

Q:

The $\cos(\theta)$ between v_1 and v_2 can be seen as the projection of v_1 onto v_2 and vice versa. Please elaborate on the connection between the Gram matrix and a projection matrix.

A:

That's a very common and insightful intuition. You're right that these concepts are all deeply connected, but let's clarify the relationship, as the premise has a slight (but important) inaccuracy.

- **Correction:** $\cos(\theta)$ itself isn't the projection. The **scalar projection** of v_1 onto v_2 is $\|v_1\| \cos(\theta)$. The **vector projection** is a *vector* with length $\|v_1\| \cos(\theta)$ pointing in the direction of v_2 . The **dot product** $\langle v_1, v_2 \rangle$ is what contains this information, as $\langle v_1, v_2 \rangle = \|v_1\| \|v_2\| \cos(\theta)$.

The Gram matrix is built from these dot products. The projection matrix *uses* the Gram matrix to perform its job.

Here is the connection.

The Direct Connection: G is *Inside* P

The most direct relationship is that the Gram matrix G is a crucial component **used to build** the orthogonal projection matrix P .

Let's say you have a set of n basis vectors $\{v_1, \dots, v_n\}$ that span a subspace W . As we did before, let's put them into the columns of a matrix A :

$$A = \begin{bmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{bmatrix}$$

1. **Gram Matrix (G):** This matrix describes the *internal geometry* of your basis vectors v_i .

$$G = A^T A$$

G is an $n \times n$ matrix. Its entries $G_{ij} = \langle v_i, v_j \rangle$ tell you the angles and lengths of your basis vectors.

2. **Projection Matrix (P):** This matrix is an *operator*. It takes *any* vector b from the larger space and maps it to its closest point (its "shadow") p inside the subspace W . The formula for this projection matrix P (which projects onto the column space of A) is:

$$P = A(A^T A)^{-1} A^T$$

If we substitute $G = A^T A$, the formula becomes:

$$P = AG^{-1}A^T$$

This is the connection. **The projection matrix P is built using the inverse of the Gram matrix G .**

Why This Works: An Intuitive Explanation

Why does G^{-1} show up here? It's all about solving the “closest point” problem, which is the heart of projection.

Let's find the projection p of a vector b onto the subspace W (the column space of A).

1. The projection p must be *in* the subspace, so it must be a linear combination of A 's columns. We can write this as $p = Ax$ for some unknown vector of coefficients $x = (c_1, \dots, c_n)^T$.
2. The “closest point” is defined by the property that the error vector $(b - p)$ must be **orthogonal** to the subspace W .
3. This means $(b - p)$ must be orthogonal to *every basis vector* in A .
4. We can write this condition concisely as:

$$A^T(b - p) = 0$$

5. Now substitute $p = Ax$:

$$A^T(b - Ax) = 0$$

6. Distribute A^T :

$$A^T b - A^T A x = 0$$

7. Rearrange and substitute $G = A^T A$:

$$(A^T A)x = A^T b \text{ implies } Gx = A^T b$$

This famous equation, $Gx = A^T b$, is called the **normal equation**.

8. We need to find the coefficients x . We can do that by inverting G (this is only possible if the vectors are linearly independent, which we assume for a basis):

$$x = G^{-1}(A^T b)$$

9. We're not done! x is just the *coefficients*. We need the actual projection p . We go back to $p = Ax$:

$$p = Ax = A(G^{-1}A^T b)$$

10. If we regroup the parentheses, we see what matrix P turns b into p :

$$p = (AG^{-1}A^T)b \text{ implies } P = AG^{-1}A^T$$

What this means: The Gram matrix $G = A^T A$ contains all the information about how the basis vectors $\langle v_i, v_j \rangle$ overlap. Inverting it, G^{-1} , essentially “de-correlates” or “untangles” this non-orthogonal basis. It’s the correction factor needed to account for the fact that your basis vectors might not be a nice, simple, orthonormal set (where G would just be the identity matrix I).

Summary: Gram Matrix vs. Projection Matrix

This table clarifies their different roles.

Property	Gram Matrix ($G = A^T A$)	Projection Matrix ($P = A(A^T A)^{-1} A^T$)
Main Purpose	Stores information. A “lookup table” of all inner products (angles/lengths) of the basis vectors.	Performs an action. An “operator” that maps any vector onto a specific subspace.
What it Represents	The <i>internal geometry</i> of the basis set $\{v_i\}$.	The <i>transformation</i> of projecting onto the space spanned by $\{v_i\}$.
Input / Output	It’s a static matrix. It doesn’t “take” an input.	Input: Vector b in the ambient space. Output: Vector p in the subspace.
Size	$n \times n$ (where n is the number of basis vectors).	$m \times m$ (where m is the dimension of the ambient space, e.g., \mathbb{R}^m).
Key Property	Symmetric, Positive Semi-Definite.	Idempotent ($P^2 = P$) and Symmetric ($P^T = P$).

In short: The **Gram matrix** describes the *subspace itself* (via its chosen basis), while the **projection matrix** provides the *mechanism to get to* that subspace.