9.3 Advanced Topics in Linear Algebra

Diagonalization and Jordan's Theorem

A system of differential equations $\mathbf{x}' = A\mathbf{x}$ can be transformed to an uncoupled system $\mathbf{y}' = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{y}$ by a change of variables $\mathbf{x} = P\mathbf{y}$, provided P is invertible and A satisfies the relation

$$(1) \qquad AP = P\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n).$$

A matrix A is said to be **diagonalizable** provided (1) holds. This equation is equivalent to the system of equations $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$, $k = 1, \ldots, n$, where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are the columns of matrix P. Since P is assumed invertible, each of its columns are nonzero, and therefore $(\lambda_k, \mathbf{v}_k)$ is an eigenpair of A, $1 \le k \le n$. The values λ_k need not be distinct (e.g., all $\lambda_k = 1$ if A is the identity). This proves:

Theorem 10 (Diagonalization)
An $n \times n$ matrix A is diagonalizable if and only if A has n eigenpairs $(\lambda_k, \mathbf{v}_k)$, $1 \le k \le n$, with $\mathbf{v}_1, \ldots, \mathbf{v}_n$ independent. In this case,

$$A = PDP^{-1}$$

where $D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$ and the matrix P has columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Theorem 11 (Jordan's theorem)
Any $n \times n$ matrix A can be represented in the form

$$A = PTP^{-1}$$

where P is invertible and T is upper triangular. The diagonal entries of T are eigenvalues of A.

Proof: We proceed by induction on the dimension n of A. For $n = 1$ there is nothing to prove. Assume the result for dimension n, and let's prove it when A is $(n+1) \times (n+1)$. Choose an eigenpair $(\lambda_1, \mathbf{v}_1)$ of A with $\mathbf{v}_1 \ne \mathbf{0}$. Complete a basis $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ for $\mathcal{R}^{n+1}$ and define $V = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$. Then $V^{-1}AV = \left(\begin{array}{c|c} \lambda_1 & B \\ \hline \mathbf{0} & A_1 \end{array}\right)$ for some matrices B and A_1. The induction hypothesis implies there is an invertible $n \times n$ matrix P_1 and an upper triangular matrix T_1 such that $A_1 = P_1 T_1 P_1^{-1}$. Let $R = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & P_1 \end{array}\right)$ and $\mathbf{T} = \left(\begin{array}{c|c} \lambda_1 & BT_1 \\ \hline 0 & T_1 \end{array}\right)$. Then T is upper triangular and $(V^{-1}AV)R = RT$, which implies $A = PTP^{-1}$ for $P = VR$. The induction is complete.

Cayley-Hamilton Identity

A celebrated and deep result for powers of matrices was discovered by Cayley and Hamilton (see [**?**]), which says that an $n \times n$ matrix A satisfies its own characteristic equation. More precisely:

Theorem 12 (Cayley-Hamilton)
Let $\det(A - \lambda I)$ be expanded as the nth degree polynomial

$$p(\lambda) = \sum_{j=0}^{n} c_j \lambda^j,$$

for some coefficients $c_0, \ldots, c_{n-1}$ and $c_n = (-1)^n$. Then A satisfies the equation $p(\lambda) = 0$, that is,

$$p(A) \equiv \sum_{j=0}^{n} c_j A^j = 0.$$

In factored form in terms of the eigenvalues $\{\lambda_j\}_{j=1}^n$ (duplicates possible), the matrix equation $p(A) = 0$ can be written as

$$(-1)^n (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I) = 0.$$

Proof: If A is diagonalizable, $AP = P\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n)$, then the proof is obtained from the simple expansion

$$A^j = P\,\mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j) P^{-1},$$

because summing across this identity leads to

$$\begin{aligned} p(A) &= \textstyle\sum_{j=0}^{n} c_j A^j \\ &= P \left(\textstyle\sum_{j=0}^{n} c_j\, \mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j) \right) P^{-1} \\ &= P\,\mathbf{diag}(p(\lambda_1), \ldots, p(\lambda_n)) P^{-1} \\ &= P\,\mathbf{diag}(0, \ldots, 0) P^{-1} \\ &= 0. \end{aligned}$$

If A is not diagonalizable, then this proof fails. To handle the general case, we apply **Jordan's theorem**, which says that $A = PTP^{-1}$ where T is *upper triangular* (instead of *diagonal*) and the not necessarily distinct eigenvalues λ_1, $\ldots$, λ_n of A appear on the diagonal of T. Using Jordan's theorem, define

$$A_\epsilon = P(T + \epsilon\,\mathbf{diag}(1, 2, \ldots, n)) P^{-1}.$$

For small $\epsilon > 0$, the matrix A_ϵ has distinct eigenvalues $\lambda_j + \epsilon j$, $1 \le j \le n$. Then the diagonalizable case implies that A_ϵ satisfies its characteristic equation. Let $p_\epsilon(\lambda) = \det(A_\epsilon - \lambda I)$. Use $0 = \lim_{\epsilon \to 0} p_\epsilon(A_\epsilon) = p(A)$ to complete the proof.

An Extension of Jordan's Theorem

Theorem 13 (Jordan's Extension)
Any $n \times n$ matrix A can be represented in the block triangular form

$$A = PTP^{-1}, \quad T = \textbf{diag}(T_1, \ldots, T_k),$$

where P is invertible and each matrix T_i is upper triangular with diagonal entries equal to a single eigenvalue of A.

The proof of the theorem is based upon Jordan's theorem, and proceeds by induction. The reader is invited to try to find a proof, or read further in the text, where this theorem is presented as a special case of the Jordan decomposition $A = PJP^{-1}$.

Solving Block Triangular Differential Systems

A matrix differential system $\mathbf{y}'(t) = T\mathbf{y}(t)$ with T block upper triangular splits into scalar equations which can be solved by elementary methods for first order scalar differential equations. To illustrate, consider the system

$$\begin{aligned} y_1' &= 3y_1 + x_2 + y_3, \\ y_2' &= 3y_2 + y_3, \\ y_3' &= 2y_3. \end{aligned}$$

The techniques that apply are the growth-decay formula for $u' = ku$ and the integrating factor method for $u' = ku + p(t)$. Working backwards from the last equation with back-substitution gives

$$\begin{aligned} y_3 &= c_3 e^{2t}, \\ y_2 &= c_2 e^{3t} - c_3 e^{2t}, \\ y_1 &= (c_1 + c_2 t) e^{3t}. \end{aligned}$$

What has been said here applies to any triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$, in order to write an exact formula for the solution $\mathbf{y}(t)$.

If A is an $n \times n$ matrix, then Jordan's theorem gives $A = PTP^{-1}$ with T block upper triangular and P invertible. The change of variable $\mathbf{x}(t) = P\mathbf{y}(t)$ changes $\mathbf{x}'(t) = A\mathbf{x}(t)$ into the block triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$.

There is no special condition on A, to effect the change of variable $\mathbf{x}(t) = P\mathbf{y}(t)$. The solution $\mathbf{x}(t)$ of $\mathbf{x}'(t) = A\mathbf{x}(t)$ is a product of the invertible matrix P and a column vector $\mathbf{y}(t)$; the latter is the solution of the block triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$, obtained by growth-decay and integrating factor methods.

The *importance of this idea* is to provide a theoretical method for solving any system $\mathbf{x}'(t) = A\mathbf{x}(t)$. We show in the Jordan Form section, *infra*, how to find matrices P and T in Jordan's extension $A = PTP^{-1}$, using computer algebra systems.

Symmetric Matrices and Orthogonality

Described here is a process due to Gram-Schmidt for replacing a given set of independent eigenvectors by another set of eigenvectors which are of unit length and **orthogonal** (dot product zero or 90 degrees apart). The process, which applies to any independent set of vectors, is especially useful in the case of eigenanalysis of a **symmetric** matrix: $A^T = A$.

Unit eigenvectors. An eigenpair $(\lambda, \mathbf{x})$ of A can always be selected so that $\|\mathbf{x}\| = 1$. If $\|\mathbf{x}\| \neq 1$, then replace eigenvector $\mathbf{x}$ by the scalar multiple $c\mathbf{x}$, where $c = 1/\|\mathbf{x}\|$. By this small change, it can be assumed that the eigenvector has unit length. If in addition the eigenvectors are *orthogonal*, then the eigenvectors are said to be **orthonormal**.

Theorem 14 (Orthogonality of Eigenvectors)
Assume that $n \times n$ matrix A is **symmetric**, $A^T = A$. If $(\alpha, \mathbf{x})$ and $(\beta, \mathbf{y})$ are eigenpairs of A with $\alpha \neq \beta$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal: $\mathbf{x} \cdot \mathbf{y} = 0$.

Proof: To prove this result, compute $\alpha\mathbf{x} \cdot \mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \mathbf{x}^T A\mathbf{y}$. Analagously, $\beta\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T A\mathbf{y}$. Subtracting the relations implies $(\alpha - \beta)\mathbf{x} \cdot \mathbf{y} = 0$, giving $\mathbf{x} \cdot \mathbf{y} = 0$ due to $\alpha \neq \beta$. The proof is complete.

Theorem 15 (Real Eigenvalues)
If $A^T = A$, then all eigenvalues of A are real. Consequently, matrix A has n real eigenvalues counted according to multiplicity.

Proof: The second statement is due to the fundamental theorem of algebra. To prove the eigenvalues are real, it suffices to prove $\lambda = \overline{\lambda}$ when $A\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \neq \mathbf{0}$. We admit that $\mathbf{v}$ may have complex entries. We will use $\overline{A} = A$ (A is real). Take the complex conjugate across $A\mathbf{v} = \lambda\mathbf{v}$ to obtain $A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$. Transpose $A\mathbf{v} = \lambda\mathbf{v}$ to obtain $\mathbf{v}^T A^T = \lambda\mathbf{v}^T$ and then conclude $\mathbf{v}^T A = \lambda\mathbf{v}^T$ from $A^T = A$. Multiply this equation by $\overline{\mathbf{v}}$ on the right to obtain $\mathbf{v}^T A\overline{\mathbf{v}} = \lambda\mathbf{v}^T\overline{\mathbf{v}}$. Then multiply $A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$ by $\mathbf{v}^T$ on the left to obtain $\mathbf{v}^T A\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^T\overline{\mathbf{v}}$. Then we have

$$\lambda\mathbf{v}^T\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^T\overline{\mathbf{v}}.$$

Because $\mathbf{v}^T\overline{\mathbf{v}} = \sum_{j=1}^n |v_j|^2 > 0$, then $\lambda = \overline{\lambda}$ and λ is real. The proof is complete.

Theorem 16 (Independence of Orthogonal Sets)
Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be a set of nonzero orthogonal vectors. Then this set is independent.

Proof: Form the equation $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$, the plan being to solve for $c_1, \ldots, c_k$. Take the dot product of the equation with $\mathbf{v}_1$. Then

$$c_1\mathbf{v}_1 \cdot \mathbf{v}_1 + \cdots + c_k\mathbf{v}_1 \cdot \mathbf{v}_k = \mathbf{v}_1 \cdot \mathbf{0}.$$

All terms on the left side except one are zero, and the right side is zero also, leaving the relation

$$c_1\mathbf{v}_1 \cdot \mathbf{v}_1 = 0.$$

Because $\mathbf{v}_1$ is not zero, then $c_1 = 0$. The process can be applied to the remaining coefficients, resulting in

$$c_1 = c_2 = \cdots = c_k = 0,$$

which proves independence of the vectors.

The Gram-Schmidt process

The eigenvectors of a symmetric matrix A may be constructed to be orthogonal. First of all, observe that eigenvectors corresponding to distinct eigenvalues are orthogonal by Theorem 14. It remains to construct from k independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$, corresponding to a single eigenvalue λ, another set of independent eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ for λ which are pairwise orthogonal. The idea, due to Gram-Schmidt, applies to any set of k independent vectors.

Application of the Gram-Schmidt process can be illustrated by example: let $(-1, \mathbf{v}_1)$, $(2, \mathbf{v}_2)$, $(2, \mathbf{v}_3)$, $(2, \mathbf{v}_4)$ be eigenpairs of a 4×4 symmetric matrix A. Then $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$. The eigenvectors $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ belong to eigenvalue $\lambda = 2$, but they are not necessarily orthogonal. The Gram-Schmidt process replaces eigenvectors $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ by $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{y}_4$ which are pairwise orthogonal. The result is that eigenvectors $\mathbf{v}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{y}_4$ are pairwise orthogonal and the eigenpairs of A are replaced by $(-1, \mathbf{v}_1)$, $(2, \mathbf{y}_2)$, $(2, \mathbf{y}_3)$, $(2, \mathbf{y}_4)$.

Theorem 17 (Gram-Schmidt)
Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be independent n-vectors. The set of vectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ constructed below as linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are pairwise orthogonal and independent.

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{x}_1 \\
\mathbf{y}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 \\
\mathbf{y}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_2}{\mathbf{y}_2 \cdot \mathbf{y}_2}\mathbf{y}_2 \\
&\vdots \\
\mathbf{y}_k &= \mathbf{x}_k - \frac{\mathbf{x}_k \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 - \cdots - \frac{\mathbf{x}_k \cdot \mathbf{y}_{k-1}}{\mathbf{y}_{k-1} \cdot \mathbf{y}_{k-1}}\mathbf{y}_{k-1}
\end{aligned}$$

Proof: Induction will be applied on k to show that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are nonzero and orthogonal. If $k = 1$, then there is just one nonzero vector constructed $\mathbf{y}_1 = \mathbf{x}_1$. Orthogonality for $k = 1$ is not discussed because there are no pairs to test. Assume the result holds for $k - 1$ vectors. Let's verify that it holds for k vectors, $k > 1$. Assume orthogonality $\mathbf{y}_i \cdot \mathbf{y}_j = 0$ for $i \neq j$ and $\mathbf{y}_i \neq \mathbf{0}$ for

$1 \le i, j \le k-1$. It remains to test $\mathbf{y}_i \cdot \mathbf{y}_k = 0$ for $1 \le i \le k-1$ and $\mathbf{y}_k \ne \mathbf{0}$. The test depends upon the identity

$$\mathbf{y}_i \cdot \mathbf{y}_k = \mathbf{y}_i \cdot \mathbf{x}_k - \sum_{j=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{y}_j}{\mathbf{y}_j \cdot \mathbf{y}_j} \mathbf{y}_i \cdot \mathbf{y}_j,$$

which is obtained from the formula for $\mathbf{y}_k$ by taking the dot product with $\mathbf{y}_i$. In the identity, $\mathbf{y}_i \cdot \mathbf{y}_j = 0$ by the induction hypothesis for $1 \le j \le k-1$ and $j \ne i$. Therefore, the summation in the identity contains just the term for index $j = i$, and the contribution is $\mathbf{y}_i \cdot \mathbf{x}_k$. This contribution cancels the leading term on the right in the identity, resulting in the orthogonality relation $\mathbf{y}_i \cdot \mathbf{y}_k = 0$. If $\mathbf{y}_k = \mathbf{0}$, then $\mathbf{x}_k$ is a linear combination of $\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$. But each $\mathbf{y}_j$ is a linear combination of $\{\mathbf{x}_i\}_{i=1}^{j}$, therefore $\mathbf{y}_k = \mathbf{0}$ implies $\mathbf{x}_k$ is a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$, a contradiction to the independence of $\{\mathbf{x}_i\}_{i=1}^{k}$. The proof is complete.

Orthogonal Projection

Reproduced here is the basic material on shadow projection, for the convenience of the reader. The ideas are then extended to obtain the orthogonal projection onto a subspace V of $\mathcal{R}^n$. Finally, the orthogonal projection formula is related to the Gram-Schmidt equations.

The **shadow projection** of vector $\vec{X}$ onto the direction of vector $\vec{Y}$ is the number d defined by

$$d = \frac{\vec{X} \cdot \vec{Y}}{|\vec{Y}|}.$$

The triangle determined by $\vec{X}$ and $d\dfrac{\vec{Y}}{|\vec{Y}|}$ is a right triangle.

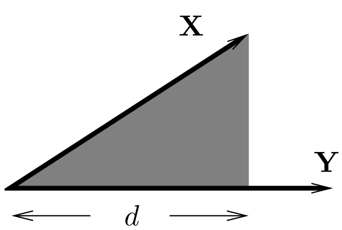

Figure 3. Shadow projection d of vector X onto the direction of vector Y.

The **vector shadow projection** of $\vec{X}$ onto the line L through the origin in the direction of $\vec{Y}$ is defined by

$$\mathbf{proj}_{\vec{Y}}(\vec{X}) = d\frac{\vec{Y}}{|\vec{Y}|} = \frac{\vec{X} \cdot \vec{Y}}{\vec{Y} \cdot \vec{Y}}\vec{Y}.$$

Orthogonal Projection for Dimension 1. The extension of the shadow projection formula to a subspace V of $\mathcal{R}^n$ begins with unitizing $\vec{Y}$ to isolate the vector direction $\mathbf{u} = \vec{Y}/\|\vec{Y}\|$ of line L. Define the subspace $V = \mathbf{span}\{\mathbf{u}\}$. Then V is identical to L. We define the **orthogonal projection** by the formula

$$\mathbf{Proj}_V(\mathbf{x}) = (\mathbf{u}\cdot\mathbf{x})\mathbf{u}, \quad V = \mathbf{span}\{\mathbf{u}\}.$$

The reader is asked to verify that

$$\mathbf{proj}_{\vec{Y}}(\mathbf{x}) = d\mathbf{u} = \mathbf{Proj}_V(\mathbf{x}).$$

These equalities imply that the orthogonal projection is identical to the vector shadow projection when V is one dimensional.

Orthogonal Projection for Dimension k. Consider a subspace V of $\mathcal{R}^n$ given as the span of orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Define the **orthogonal projection** by the formula

$$\mathbf{Proj}_V(\mathbf{x}) = \sum_{j=1}^{k}(\mathbf{u}_j\cdot\mathbf{x})\mathbf{u}_j, \quad V = \mathbf{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}.$$

Justification of the Formula. The definition of $\mathbf{Proj}_V(\mathbf{x})$ seems to depend on the choice of the orthonormal vectors. Suppose that $\{\mathbf{w}_j\}_{j=1}^k$ is another orthonormal basis of V. Define $\mathbf{u} = \sum_{i=1}^k(\mathbf{u}_i\cdot\mathbf{x})\mathbf{u}_j$ and $\mathbf{w} = \sum_{j=1}^k(\mathbf{w}_j\cdot\mathbf{x})\mathbf{w}_j$. It will be established that $\mathbf{u} = \mathbf{w}$, which justifies that the projection formula is independent of basis.

Lemma 1 (Orthonormal Basis Expansion) Let $\{\mathbf{v}_j\}_{j=1}^k$ be an orthonormal basis of a subspace V in $\mathcal{R}^n$. Then each vector $\mathbf{v}$ in V is represented as

$$\mathbf{v} = \sum_{j=1}^{k}(\mathbf{v}_j\cdot\mathbf{v})\mathbf{v}_j.$$

Proof: First, $\mathbf{v}$ has a basis expansion $\mathbf{v} = \sum_{j=1}^k c_j\mathbf{v}_j$ for some constants c_1, $\ldots$, c_k. Take the inner product of this equation with vector $\mathbf{v}_i$ to prove that $c_i = \mathbf{v}_i\cdot\mathbf{v}$, hence the claimed expansion is proved.

Lemma 2 (Orthogonality) Let $\{\mathbf{u}_i\}_{i=1}^k$ be an orthonormal basis of a subspace V in $\mathcal{R}^n$. Let $\mathbf{x}$ be any vector in $\mathcal{R}^n$ and define $\mathbf{u} = \sum_{i=1}^k(\mathbf{u}_i\cdot\mathbf{x})\mathbf{u}_i$. Then $\mathbf{y}\cdot(\mathbf{x}-\mathbf{u}) = 0$ for all vectors $\mathbf{y}$ in V.

Proof: The first lemma implies $\mathbf{u}$ can be written a second way as a linear combination of $\mathbf{u}_1$, ... $\mathbf{u}_k$. Independence implies equal basis coefficients, which gives $\mathbf{u}_j\cdot\mathbf{u} = \mathbf{u}_j\cdot\mathbf{x}$. Then $\mathbf{u}_j\cdot(\mathbf{x}-\mathbf{u}) = 0$. Because $\mathbf{y}$ is in V, then $\mathbf{y} = \sum_{j=1}^k c_j\mathbf{u}_j$, which implies $\mathbf{y}\cdot(\mathbf{x}-\mathbf{u}) = \sum_{j=1}^k \mathbf{u}_j(\mathbf{x}-\mathbf{u}) = 0$.

The proof that $\mathbf{w} = \mathbf{u}$ has these details:

$$\begin{aligned}\mathbf{w} &= \textstyle\sum_{j=1}^k(\mathbf{w}_j\cdot\mathbf{x})\mathbf{w}_j \\ &= \textstyle\sum_{j=1}^k(\mathbf{w}_j\cdot\mathbf{u})\mathbf{w}_j && \text{Because } \mathbf{w}_j\cdot(\mathbf{x}-\mathbf{u}) = 0 \text{ by the second lemma.}\end{aligned}$$

$$
\begin{array}{ll}
= \sum_{j=1}^{k} \left(\mathbf{w}_j \cdot \sum_{i=1}^{k} (\mathbf{u}_i \cdot \mathbf{x}) \mathbf{u}_i \right) \mathbf{w}_j & \text{Definition of } \mathbf{u}. \\
= \sum_{j=1}^{k} \sum_{i=1}^{k} (\mathbf{w}_j \cdot \mathbf{u}_i)(\mathbf{u}_i \cdot \mathbf{x}) \mathbf{w}_j & \text{Dot product properties.} \\
= \sum_{i=1}^{k} \left(\sum_{j=1}^{k} (\mathbf{w}_j \cdot \mathbf{u}_i) \mathbf{w}_j \right) (\mathbf{u}_i \cdot \mathbf{x}) & \text{Switch summations.} \\
= \sum_{i=1}^{k} \mathbf{u}_i (\mathbf{u}_i \cdot \mathbf{x}) & \text{First lemma with } \mathbf{v} = \mathbf{u}_i. \\
= \mathbf{u} & \text{Definition of } \mathbf{u}.
\end{array}
$$

Orthogonal Projection and Gram-Schmidt. Define $\mathbf{y}_1, \ldots, \mathbf{y}_k$ by the Gram-Schmidt relations on page 662. Define

$$\mathbf{u}_j = \mathbf{y}_j / \|\mathbf{y}_j\|$$

for $j = 1, \ldots, k$. Then $V_{j-1} = \mathbf{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}\}$ is a subspace of $\mathcal{R}^n$ of dimension $j-1$ with orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$ and

$$
\begin{aligned}
\mathbf{y}_j &= \mathbf{x}_j - \left(\frac{\mathbf{x}_j \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1 + \cdots + \frac{\mathbf{x}_k \cdot \mathbf{y}_{j-1}}{\mathbf{y}_{j-1} \cdot \mathbf{y}_{j-1}} \mathbf{y}_{j-1} \right) \\
&= \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j).
\end{aligned}
$$

The Near Point Theorem

Developed here is the characterization of the orthogonal projection of a vector $\mathbf{x}$ onto a subspace V as the unique point $\mathbf{v}$ in V which minimizes $\|\mathbf{x} - \mathbf{v}\|$, that is, the point in V which is nearest to $\mathbf{x}$.

In remembering the Gram-Schmidt formulas, and in the use of the orthogonal projection in proofs and constructions, the following key theorem is useful.

Theorem 18 (Orthogonal Projection Properties)
Let V be the span of orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

(a) Every vector in V has an orthogonal expansion $\mathbf{v} = \sum_{j=1}^{k} (\mathbf{u}_j \cdot \mathbf{v}) \mathbf{u}_j$.

(b) The vector $\mathbf{Proj}_V(\mathbf{x})$ is a vector in the subspace V.

(c) The vector $\mathbf{w} = \mathbf{x} - \mathbf{Proj}_V(\mathbf{x})$ is orthogonal to every vector in V.

(d) Among all vectors $\mathbf{v}$ in V, the minimum value of $\|\mathbf{x} - \mathbf{v}\|$ is uniquely obtained by the orthogonal projection $\mathbf{v} = \mathbf{Proj}_V(\mathbf{x})$.

Proof: Properties (a), (b) and (c) are proved in the preceding lemmas. Details are repeated here, in case the lemmas were skipped.

(a): Write a basis expansion $\mathbf{v} = \sum_{j=1}^{k} c_j \mathbf{v}_j$ for some constants $c_1, \ldots, c_k$. Take the inner product of this equation with vector $\mathbf{v}_i$ to prove that $c_i = \mathbf{v}_i \cdot \mathbf{v}$.

(b): Vector $\mathbf{Proj}_V(\mathbf{x})$ is a linear combination of basis elements of V.

(c): Let's compute the dot product of $\mathbf{w}$ and $\mathbf{v}$. We will use the orthogonal expansion from (a).

$$\begin{aligned}
\mathbf{w}\cdot\mathbf{v} &= (\mathbf{x} - \mathbf{Proj}_V(\mathbf{x}))\cdot\mathbf{v} \\
&= \mathbf{x}\cdot\mathbf{v} - \left(\sum_{j=1}^{k}(\mathbf{x}\cdot\mathbf{u}_j)\mathbf{u}_j\right)\cdot\mathbf{v} \\
&= \sum_{j=1}^{k}(\mathbf{v}\cdot\mathbf{u}_j)(\mathbf{u}_j\cdot\mathbf{x}) - \sum_{j=1}^{k}(\mathbf{x}\cdot\mathbf{u}_j)(\mathbf{u}_j\cdot\mathbf{v}) \\
&= 0.
\end{aligned}$$

(d): Begin with the Pythagorean identity

$$\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 = \|\mathbf{a}+\mathbf{b}\|^2$$

valid exactly when $\mathbf{a}\cdot\mathbf{b} = 0$ (a right triangle, $\theta = 90^\circ$). Using an arbitrary $\mathbf{v}$ in V, define $\mathbf{a} = \mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}$ and $\mathbf{b} = \mathbf{x} - \mathbf{Proj}_V(\mathbf{x})$. By (b), vector $\mathbf{a}$ is in V. Because of (c), then $\mathbf{a}\cdot\mathbf{b} = 0$. This gives the identity

$$\|\,\mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}\|^2 + \|\mathbf{x} - \mathbf{Proj}_V(\mathbf{x})\|^2 = \|\mathbf{x} - \mathbf{v}\|^2,$$

which establishes $\|\mathbf{x} - \mathbf{Proj}_V(\mathbf{x})\| < \|\mathbf{x} - \mathbf{v}\|$ except for the unique $\mathbf{v}$ such that $\|\,\mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}\| = 0$.

The proof is complete.

Theorem 19 (Near Point to a Subspace)
Let V be a subspace of $\mathcal{R}^n$ and $\mathbf{x}$ a vector not in V. The **near point** to $\mathbf{x}$ in V is the orthogonal projection of $\mathbf{x}$ onto V. This point is characterized as the minimum of $\|\mathbf{x} - \mathbf{v}\|$ over all vectors $\mathbf{v}$ in the subspace V.

Proof: Apply **(d)** of the preceding theorem.

Theorem 20 (Cross Product and Projections)
The cross product direction $\mathbf{a}\times\mathbf{b}$ can be computed as $\mathbf{c} - \mathbf{Proj}_V(\mathbf{c})$, by selecting a direction $\mathbf{c}$ not in $V = \mathbf{span}\{\mathbf{a},\mathbf{b}\}$.

Proof: The cross product makes sense only in $\mathcal{R}^3$. Subspace V is two dimensional when $\mathbf{a}$, $\mathbf{b}$ are independent, and Gram-Schmidt applies to find an orthonormal basis $\mathbf{u}_1$, $\mathbf{u}_2$. By (c) of Theorem 18, the vector $\mathbf{c} - \mathbf{Proj}_V(\mathbf{c})$ has the same or opposite direction to the cross product.

The QR Decomposition

The Gram-Schmidt formulas can be organized as matrix multiplication $A = QR$, where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the independent columns of A, and Q has columns equal to the Gram-Schmidt orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, which are the unitized Gram-Schmidt vectors.

Theorem 21 (The QR-Decomposition)
Let the $m \times n$ matrix A have independent columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then there is an upper triangular matrix R with positive diagonal entries and an orthonormal matrix Q such that

$$A = QR.$$

Proof: Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be the Gram-Schmidt orthogonal vectors given by relations on page 662. Define $\mathbf{u}_k = \mathbf{y}_k/\|\mathbf{y}_k\|$ and $r_{kk} = \|\mathbf{y}_k\|$ for $k = 1, \ldots, n$, and otherwise $r_{ij} = \mathbf{u}_i \cdot \mathbf{x}_j$. Let $Q = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Then

$$(2) \qquad \begin{aligned} \mathbf{x}_1 &= r_{11}\mathbf{u}_1, \\ \mathbf{x}_2 &= r_{22}\mathbf{u}_2 + r_{21}\mathbf{u}_1, \\ \mathbf{x}_3 &= r_{33}\mathbf{u}_3 + r_{31}\mathbf{u}_1 + r_{32}\mathbf{u}_2, \\ &\vdots \\ \mathbf{x}_n &= r_{nn}\mathbf{u}_n + r_{n1}\mathbf{u}_1 + \cdots + r_{nn-1}\mathbf{u}_{n-1}. \end{aligned}$$

It follows from (2) and matrix multiplication that $A = QR$. The proof is complete.

Theorem 22 (Matrices Q and R in $A = QR$)
Let the $m \times n$ matrix A have independent columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be the Gram-Schmidt orthogonal vectors given by relations on page 662. Define $\mathbf{u}_k = \mathbf{y}_k/\|\mathbf{y}_k\|$. Then $AQ = QR$ is satisfied by $Q = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and

$$R = \begin{pmatrix} \|y_1\| & \mathbf{u}_1 \cdot \mathbf{x}_2 & \mathbf{u}_1 \cdot \mathbf{x}_3 & \cdots & \mathbf{u}_1 \cdot \mathbf{x}_n \\ 0 & \|y_2\| & \mathbf{u}_2 \cdot \mathbf{x}_3 & \cdots & \mathbf{u}_2 \cdot \mathbf{x}_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \|y_n\| \end{pmatrix}.$$

Proof: The result is contained in the proof of the previous theorem.
Some references cite the diagonal entries as $\|\mathbf{x}_1\|$, $\|\mathbf{x}_2^\perp\|$, …, $\|\mathbf{x}_n^\perp\|$, where $\mathbf{x}_j^\perp = \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j)$, $V_{j-1} = \mathbf{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}\}$. Because $\mathbf{y}_1 = \mathbf{x}_1$ and $\mathbf{y}_j = \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j)$, the formulas for the entries of R are identical.

Theorem 23 (Uniqueness of Q and R)
Let $m \times n$ matrix A have independent columns and satisfy the decomposition $A = QR$. If Q is $m \times n$ orthogonal and R is $n \times n$ upper triangular with positive diagonal elements, then Q and R are uniquely determined.

Proof: The problem is to show that $A = Q_1R_1 = Q_2R_2$ implies $R_2R_1^{-1} = I$ and $Q_1 = Q_2$. We start with $Q_1 = Q_2R_2R_1^{-1}$. Define $P = R_2R_1^{-1}$. Then $Q_1 = Q_2P$. Because $I = Q_1^TQ_1 = P^TQ_2^TQ_2P = P^TP$, then P is orthogonal. Matrix P is the product of square upper triangular matrices with positive diagonal elements, which implies P itself is square upper triangular with positive diagonal elements. The only matrix with these properties is the identity matrix I. Then $R_2R_1^{-1} = P = I$, which implies $R_1 = R_2$ and $Q_1 = Q_2$. The proof is complete.

Theorem 24 (The QR Decomposition and Least Squares)
Let $m\times n$ matrix A have independent columns and satisfy the decomposition $A = QR$. Then the normal equation

$$A^TA\mathbf{x} = A^T\mathbf{b}$$

in the theory of least squares can be represented as

$$R\mathbf{x} = Q^T\mathbf{b}.$$

Proof: The theory of orthogonal matrices implies $Q^TQ = I$. Then the identity $(CD)^T = D^TC^T$, the equation $A = QR$, and R^T invertible imply

$A^TA\mathbf{x} = A^T\mathbf{b}$	Normal equation
$R^TQ^TQR\mathbf{x} = R^TQ^T\mathbf{x}$	Substitute $A = QR$.
$R\mathbf{x} = Q^T\mathbf{x}$	Multiply by the inverse of R^T.

The proof is complete.

The formula $R\mathbf{x} = Q^T\mathbf{b}$ can be solved by back-substitution, which accounts for its popularity in the numerical solution of least squares problems.

Theorem 25 (Spectral Theorem)
Let A be a given $n \times n$ real matrix. Then $A = QDQ^{-1}$ with Q orthogonal and D diagonal if and only if $A^T = A$.

Proof: The reader is reminded that Q orthogonal means that the columns of Q are **orthonormal**. The equation $A = A^T$ means A is **symmetric**.

Assume first that $A = QDQ^{-1}$ with $Q = Q^T$ orthogonal ($Q^TQ = I$) and D diagonal. Then $Q^T = Q = Q^{-1}$. This implies $A^T = (QDQ^{-1})^T = (Q^{-1})^TD^TQ^T = QDQ^{-1} = A$.

Conversely, assume $A^T = A$. Then the eigenvalues of A are real and eigenvectors corresponding to distinct eigenvalues are orthogonal. The proof proceeds by induction on the dimension n of the $n \times n$ matrix A.

For $n = 1$, let Q be the 1×1 identity matrix. Then Q is orthogonal and $AQ = QD$ where D is 1×1 diagonal.

Assume the decomposition $AQ = QD$ for dimension n. Let's prove it for A of dimension $n + 1$. Choose a real eigenvalue λ of A and eigenvector $\mathbf{v}_1$ with $\|\mathbf{v}_1\| = 1$. Complete a basis $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ of $\mathcal{R}^{n+1}$. By Gram-Schmidt, we assume as well that this basis is orthonormal. Define $P = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$. Then P is orthogonal and satisfies $P^T = P^{-1}$. Define $B = P^{-1}AP$. Then B is symmetric ($B^T = B$) and $\mathbf{col}(B, 1) = \lambda\,\mathbf{col}(I, 1)$. These facts imply that B is a block matrix

$$B = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right)$$

where C is symmetric ($C^T = C$). The induction hypothesis applies to C to obtain the existence of an orthogonal matrix Q_1 such that $CQ_1 = Q_1D_1$ for some diagonal matrix D_1. Define a diagonal matrix D and matrices W and Q as follows:

$$
\begin{aligned}
D &= \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right), \\
W &= \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right), \\
Q &= PW.
\end{aligned}
$$

Then Q is the product of two orthogonal matrices, which makes Q orthogonal. Compute

$$W^{-1}BW = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1^{-1} \end{array}\right)\left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right)\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right) = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right).$$

Then $Q^{-1}AQ = W^{-1}P^{-1}APW = W^{-1}BW = D$. This completes the induction, ending the proof of the theorem.

Theorem 26 (Schur's Theorem)
Given any real $n \times n$ matrix A, possibly non-symmetric, there is an upper triangular matrix T, whose diagonal entries are the eigenvalues of A, and a complex matrix Q satisfying $\overline{Q}^T = Q^{-1}$ (Q is unitary), such that

$$AQ = QT.$$

If $A = A^T$, then Q is real orthogonal ($Q^T = Q$).

Schur's theorem can be proved by induction, following the induction proof of Jordan's theorem, or the induction proof of the Spectral Theorem. The result can be used to prove the Spectral Theorem in two steps. Indeed, Schur's Theorem implies Q is real, T equals its transpose, and T is triangular. Then T must equal a diagonal matrix D.

Theorem 27 (Eigenpairs of a Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs $(\lambda_1, \mathbf{v}_1)$, ..., $(\lambda_n, \mathbf{v}_n)$, with independent eigenvectors $\mathbf{v}_1$, ..., $\mathbf{v}_n$.

Proof: The preceding theorem applies to prove the existence of an orthogonal matrix Q and a diagonal matrix D such that $AQ = QD$. The diagonal entries of D are the eigenvalues of A, in some order. For a diagonal entry λ of D appearing in row j, the relation $A\,\mathbf{col}(Q, j) = \lambda\,\mathbf{col}(Q, j)$ holds, which implies that A has n eigenpairs. The eigenvectors are the columns of Q, which are orthogonal and hence independent. The proof is complete.

Theorem 28 (Diagonalization of Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs. For each distinct eigenvalue λ, replace the eigenvectors by orthonormal eigenvectors, using the Gram-Schmidt process. Let $\mathbf{u}_1$, ..., $\mathbf{u}_n$ be the orthonormal vectors so obtained and define

$$Q = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_n), \quad D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n).$$

Then Q is orthogonal and $AQ = QD$.

Proof: The preceding theorem justifies the eigenanalysis result. Already, eigenpairs corresponding to distinct eigenvalues are orthogonal. Within the set of eigenpairs with the same eigenvalue λ, the Gram-Schmidt process produces a replacement basis of orthonormal eigenvectors. Then the union of all the eigenvectors is orthonormal. The process described here does not disturb the ordering of eigenpairs, because it only replaces an eigenvector. The proof is complete.

The Singular Value Decomposition

The decomposition has been used as a data compression algorithm. A geometric interpretation will be given in the next subsection.

Theorem 29 (Positive Eigenvalues of A^TA)
Given an $m \times n$ real matrix A, then A^TA is a real symmetric matrix whose eigenpairs $(\lambda, \mathbf{v})$ satisfy

$$\lambda = \frac{\|A\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \geq 0. \tag{3}$$

Proof: Symmetry follows from $(A^TA)^T = A^T(A^T)^T = A^TA$. An eigenpair $(\lambda, \mathbf{v})$ satisfies $\lambda\overline{\mathbf{v}}^T\mathbf{v} = \overline{\mathbf{v}}^TA^TA\mathbf{v} = (\overline{A\mathbf{v}})^T(A\mathbf{v}) = \|A\mathbf{v}\|^2$, hence (3).

Definition 4 (Singular Values of A)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$. The numbers

$$\sigma_k = \sqrt{\lambda_k}, \quad 1 \leq k \leq n,$$

are called the **singular values** of the matrix A. The ordering of the singular values is always with decreasing magnitude.

Theorem 30 (Orthonormal Set $\mathbf{u}_1, \ldots, \mathbf{u}_m$)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$ and corresponding orthonormal eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$, obtained by the Gram-Schmidt process. Define the vectors

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1, \ldots, \mathbf{u}_r = \frac{1}{\sigma_r}A\mathbf{v}_r.$$

Because $\|A\mathbf{v}_k\| = \sigma_k$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is orthonormal. Gram-Schmidt can extend this set to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathcal{R}^m$.

Proof of Theorem 30: Because $A^TA\mathbf{v}_k = \lambda_k\mathbf{v}_k \neq \mathbf{0}$ for $1 \leq k \leq r$, the vectors $\mathbf{u}_k$ are nonzero. Given $i \neq j$, then $\sigma_i\sigma_j\mathbf{u}_i \cdot \mathbf{u}_j = (A\mathbf{v}_i)^T(A\mathbf{v}_j) = \lambda_j\mathbf{v}_i^T\mathbf{v}_j = 0$, showing that the vectors $\mathbf{u}_k$ are orthogonal. Further, $\|\mathbf{u}_k\|^2 = \mathbf{v}_k \cdot (A^TA\mathbf{v}_k)/\lambda_k = \|\mathbf{v}_k\|^2 = 1$ because $\{\mathbf{v}_k\}_{k=1}^n$ is an orthonormal set.

The extension of the $\mathbf{u}_k$ to an orthonormal basis of $\mathcal{R}^m$ is not unique, because it depends upon a choice of independent spanning vectors $\mathbf{y}_{r+1}, \ldots, \mathbf{y}_m$ for the set $\{\mathbf{x} \ : \ \mathbf{x}\cdot\mathbf{u}_k = 0, \quad 1 \le k \le r\}$. Once selected, Gram-Schmidt is applied to $\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{y}_{r+1}, \ldots, \mathbf{y}_m$ to obtain the desired orthonormal basis.

Theorem 31 (The Singular Value Decomposition (svd))
Let A be a given real $m \times n$ matrix. Let $(\lambda_1, \mathbf{v}_1)$,…,$(\lambda_n, \mathbf{v}_n)$ be a set of orthonormal eigenpairs for A^TA such that $\sigma_k = \sqrt{\lambda_k}$ $(1 \le k \le r)$ defines the positive singular values of A and $\lambda_k = 0$ for $r < k \le n$. Complete $\mathbf{u}_1 = (1/\sigma_1)A\mathbf{v}_1, \ldots, \mathbf{u}_r = (1/\sigma_r)A\mathbf{v}_r$ to an orthonormal basis $\{\mathbf{u}_k\}_{k=1}^m$for $\mathcal{R}^m$. Define

$$\begin{aligned} &U = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_m), \quad \Sigma = \left(\begin{array}{c|c} \mathbf{diag}(\sigma_1, \ldots, \sigma_r) & 0 \\ \hline 0 & 0 \end{array}\right), \\ &V = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_n). \end{aligned}$$

Then the columns of U and V are orthonormal and

$$\begin{aligned} A &= U\Sigma V^T \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \\ &= A(\mathbf{v}_1)\mathbf{v}_1^T + \cdots + A(\mathbf{v}_r)\mathbf{v}_r^T \end{aligned}$$

Proof of Theorem 31: The product of U and Σ is the $m \times n$ matrix

$$\begin{aligned} U\Sigma &= \mathbf{aug}(\sigma_1\mathbf{u}_1, \ldots, \sigma_r\mathbf{u}_r, \mathbf{0}, \ldots, \mathbf{0}) \\ &= \mathbf{aug}(A(\mathbf{v}_1), \ldots, A(\mathbf{v}_r), \mathbf{0}, \ldots, \mathbf{0}). \end{aligned}$$

Let $\mathbf{v}$ be any vector in $\mathcal{R}^n$. It will be shown that $U\Sigma V^T\mathbf{v}$, $\sum_{k=1}^r A(\mathbf{v}_k)(\mathbf{v}_k^T\mathbf{v})$ and $A\mathbf{v}$ are the same column vector. We have the equalities

$$\begin{aligned} U\Sigma V^T\mathbf{v} &= U\Sigma \begin{pmatrix} \mathbf{v}_1^T\mathbf{v} \\ \vdots \\ \mathbf{v}_n^T\mathbf{v} \end{pmatrix} \\ &= \mathbf{aug}(A(\mathbf{v}_1), \ldots, A(\mathbf{v}_r), \mathbf{0}, \ldots, \mathbf{0}) \begin{pmatrix} \mathbf{v}_1^T\mathbf{v} \\ \vdots \\ \mathbf{v}_n^T\mathbf{v} \end{pmatrix} \\ &= \sum_{k=1}^r (\mathbf{v}_k^T\mathbf{v})A(\mathbf{v}_k). \end{aligned}$$

Because $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal basis of $\mathcal{R}^n$, then $\mathbf{v} = \sum_{k=1}^n (\mathbf{v}_k^T\mathbf{v})\mathbf{v}_k$. Additionally, $A(\mathbf{v}_k) = \mathbf{0}$ for $r < k \le n$ implies

$$\begin{aligned} A\mathbf{v} &= A\left(\sum_{k=1}^n (\mathbf{v}_k^T\mathbf{v})\mathbf{v}_k\right) \\ &= \sum_{k=1}^r (\mathbf{v}_k^T\mathbf{v})A(\mathbf{v}_k) \end{aligned}$$

Then $A\mathbf{v} = U\Sigma V^T\mathbf{v} = \sum_{k=1}^r A(\mathbf{v}_k)(\mathbf{v}_k^T\mathbf{v})$, which proves the theorem.

Singular Values and Geometry

Discussed here is how to interpret singular values geometrically, especially in low dimensions 2 and 3. First, we review conics, adopting the viewpoint of eigenanalysis.

Standard Equation of an Ellipse. Calculus courses consider ellipse equations like

$$85x^2 - 60xy + 40y^2 = 2500$$

and discuss removal of the cross term $-60xy$. The objective is to obtain a **standard** ellipse equation

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1.$$

We re-visit this old problem from a different point of view, and in the derivation establish a connection between the ellipse equation, the symmetric matrix A^TA, and the singular values of A.

9 **Example (Image of the Unit Circle)** Let $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$.
Verify that the invertible matrix A maps the unit circle into the ellipse

$$85x^2 - 60xy + 40y^2 = 2500.$$

Solution: The unit circle has parameterization $\theta \to (\cos\theta, \sin\theta)$, $0 \le \theta \le 2\pi$. Mapping of unit circle by the matrix A is formally the set of dual relations

$$\begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = A^{-1} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The Pythagorean identity $\cos^2\theta + \sin^2\theta = 1$ used on the second relation implies

$$85x^2 - 60xy + 40y^2 = 2500.$$

10 **Example (Removing the xy-Term in an Ellipse Equation)** After a rotation $(x, y) \to (X, Y)$ to remove the xy-term in

$$85x^2 - 60xy + 40y^2 = 2500,$$

verify that the ellipse equation in the new XY-coordinates is

$$\frac{X^2}{100} + \frac{Y^2}{25} = 1.$$

Solution: The xy-term removal is accomplished by a change of variables $(x, y) \to (X, Y)$ which transforms the ellipse equation $85x^2 - 60xy + 40y^2 = 2500$ into the ellipse equation $25X^2 + 100Y^2 = 2500$, details below. It's standard form is obtained by dividing by 2500, to give

$$\frac{X^2}{100} + \frac{Y^2}{25} = 1.$$

Analytic geometry says that the semi-axis lengths are $\sqrt{100} = 10$ and $\sqrt{25} = 5$.

In previous discussions of the ellipse, the equation $85x^2 - 60xy + 40y^2 = 2500$ was represented by the vector-matrix identity

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2500.$$

The program used earlier to remove the xy-term was to diagonalize the coefficient matrix $B = \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix}$ by calculating the eigenpairs of B:

$$\left(25, \begin{pmatrix} -2 \\ 1 \end{pmatrix}\right), \quad \left(100, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right).$$

Because B is symmetric, then the eigenvectors are orthogonal. The eigenpairs above are replaced by unitized pairs:

$$\left(25, \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right), \quad \left(100, \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right).$$

Then the diagonalization theory for B can be written as

$$BQ = QD, \quad Q = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix}, \quad D = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix}.$$

The single change of variables

$$\begin{pmatrix} x \\ y \end{pmatrix} = Q \begin{pmatrix} X \\ Y \end{pmatrix}$$

then transforms the ellipse equation $85x^2 - 60xy + 40y^2 = 2500$ into $25X^2 + 100Y^2 = 2500$ as follows:

$85x^2 - 60xy + 40y^2 = 2500$	Ellipse equation.
$\mathbf{u}^T B\mathbf{u} = 2500$	Where $B = \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix}$ and $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$.
$(Q\mathbf{w})^T B(Q\mathbf{w}) = 2500$	Change $\mathbf{u} = Q\mathbf{w}$, where $\mathbf{w} = \begin{pmatrix} X \\ Y \end{pmatrix}$.
$\mathbf{w}^T(Q^T BQ)\mathbf{w} = 2500$	Expand, ready to use $BQ = QD$.
$\mathbf{w}^T(D)\mathbf{w} = 2500$	Because $D = Q^{-1}BQ$ and $Q^{-1} = Q^T$.
$25X^2 + 100Y^2 = 2500$	Expand $\mathbf{w}^T D\mathbf{w}$.

Rotations and Scaling. The 2×2 singular value decomposition $A = U\Sigma V^T$ can be used to decompose the change of variables $(x, y) \to (X, Y)$ into three distinct changes of variables, each with a geometrical meaning:

$$(x, y) \longrightarrow (x_1, y_1) \longrightarrow (x_2, y_2) \longrightarrow (X, Y).$$

Table 1. Three Changes of Variable

Domain	Equation	Image	Meaning
Circle 1	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = V^T \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$	Circle 2	Rotation
Circle 2	$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \Sigma \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$	Ellipse 1	Scale axes
Ellipse 1	$\begin{pmatrix} X \\ Y \end{pmatrix} = U \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$	Ellipse 2	Rotation

Geometry. We give in Figure 4 a geometrical interpretation for the singular value decomposition

$$A = U\Sigma V^T.$$

For illustration, the matrix A is assumed 2×2 and invertible.

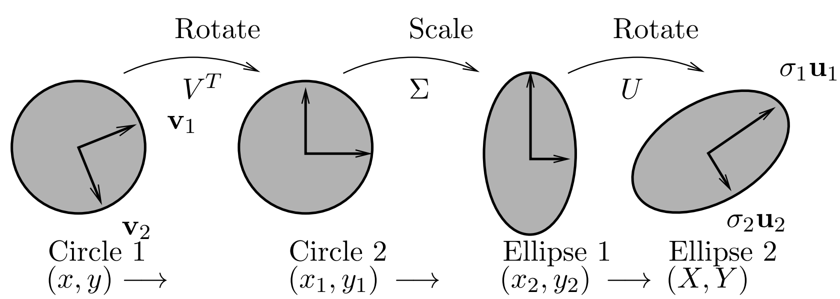

Figure 4. Mapping the unit circle.

- Invertible matrix A maps Circle 1 into Ellipse 2.
- Orthonormal vectors $\mathbf{v}_1$, $\mathbf{v}_2$ are mapped by matrix $A = U\Sigma V^T$ into orthogonal vectors $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1$, $A\mathbf{v}_2 = \sigma_2\mathbf{u}_2$, which are exactly the semi-axes vectors of Ellipse 2.
- The semi-axis lengths of Ellipse 2 equal the singular values σ_1, σ_2 of matrix A.
- The semi-axis directions of Ellipse 2 equal the basis vectors $\mathbf{u}_1$, $\mathbf{u}_2$.
- The process is a rotation $(x, y) \to (x_1, y_1)$, followed by an axis-scaling $(x_1, y_1) \to (x_2, y_2)$, followed by a rotation $(x_2, y_2) \to (X, Y)$.

Summary for the Example. The singular value decomposition for $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$ is $A = U\Sigma V^T$, where

$$U = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}, \quad V = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

- Invertible matrix $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$ maps the unit circle into an ellipse.
- The columns of V are orthonormal vectors $\mathbf{v}_1$, $\mathbf{v}_2$, computed as eigenpairs $(\lambda_1, \mathbf{v}_1)$, $(\lambda_2, \mathbf{v}_2)$ of A^TA, ordered by $\lambda_1 \geq \lambda_2$.

$$\left(100, \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right), \quad \left(25, \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right).$$

- The singular values are $\sigma_1 = \sqrt{\lambda_1} = 10$, $\sigma_2 = \sqrt{\lambda_2} = 5$.
- The image of $\mathbf{v}_1$ is $A\mathbf{v}_1 = U\Sigma V^T\mathbf{v}_1 = U\begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} = \sigma_1\mathbf{u}_1$.
- The image of $\mathbf{v}_2$ is $A\mathbf{v}_2 = U\Sigma V^T\mathbf{v}_2 = U\begin{pmatrix} 0 \\ \sigma_2 \end{pmatrix} = \sigma_2\mathbf{u}_2$.

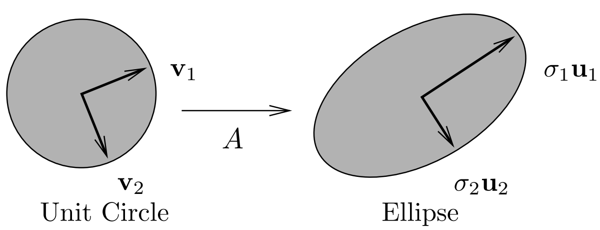

Figure 5. Mapping the unit circle into an ellipse.

The Four Fundamental Subspaces. The subspaces appearing in the Fundamental Theorem of Linear Algebra are called the **Four Fundamental Subspaces**. They are:

Subspace	**Notation**
Row Space of A	**Image**(A^T)
Nullspace of A	**kernel**(A)
Column Space of A	**Image**(A)
Nullspace of A^T	**kernel**(A^T)

The singular value decomposition $A = U\Sigma V^T$ computes orthonormal bases for the row and column spaces of of A. In the table below, symbol $r = \mathbf{rank}(A)$. Matrix A is assumed $m \times n$, which implies A maps $\mathcal{R}^n$ into $\mathcal{R}^m$.

Table 2. Four Fundamental Subspaces and the SVD

Orthonormal Basis	Span	Name
First r columns of V	**Image**(A^T)	Row Space of A
Last $n-r$ columns of V	**kernel**(A)	Nullspace of A
First r columns of U	**Image**(A)	Column Space of A
Last $m-r$ columns of U	**kernel**(A^T)	Nullspace of A^T

A Change of Basis Interpretation of the SVD. The singular value decomposition can be described as follows:

> For every $m \times n$ matrix A of rank r, orthonormal bases
>
> $$\{\mathbf{v}_i\}_{i=1}^n \text{ and } \{\mathbf{u}_j\}_{j=1}^m$$
>
> can be constructed such that
>
> - Matrix A maps basis vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ to non-negative multiples of basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$, respectively.
> - The $n-r$ left-over basis vectors $\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$ map by A into the zero vector.
> - With respect to these bases, matrix A is represented by a real diagonal matrix Σ with non-negative entries.

Exercises 9.3

Diagonalization. Find the matrices P and D in the relation $AP = PD$, that is, find the eigenpair packages.

1.

2.

3.

4.

Jordan's Theorem. Find matrices P and T in the relation $AP = PT$, where T is upper triangular.

5.

6.

7.

8.

Cayley-Hamilton Theorem.

9.

10.

11.

12.

Gram-Schmidt Process.

13.

14.

15.

16.

17.

18.

19.

20.

Shadow Projection.

21.

22.

23.

24.

Orthogonal Projection.

25.

26.

27.

28.

29. Shadow Projection Prove that

$$\mathbf{proj}_{\vec{Y}}(\mathbf{x}) = d\mathbf{u} = \mathbf{Proj}_V(\mathbf{x}).$$

The equality means that the orthogonal projection is the vector shadow projection when V is one dimensional.

30. Gram-Schmidt Construction Prove these properties, where we define

$$\mathbf{x}_j^{\perp} = \mathbf{x}_j - \mathbf{Proj}_{W_{j-1}}(\mathbf{x}_j),$$

$$W_{j-1} = \mathbf{span}(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}).$$

(a) Subspace W_{j-1} is equal to the Gram-Schmidt $V_{j-1} = \mathbf{span}(\mathbf{u}_1, \ldots, \mathbf{u}_j)$.

(b) Vector $\mathbf{x}_j^{\perp}$ is orthogonal to all vectors in W_{j-1}.

(c) The vector $\mathbf{x}_j^{\perp}$ is not zero.

(d) The Gram-Schmidt vector is

$$\mathbf{u}_j = \frac{\mathbf{x}_j^{\perp}}{\|\mathbf{x}_j^{\perp}\|}.$$

Near Point Theorem.
Find the near point to subspace V of $\mathcal{R}^n$.

31.

32.

33.

34.

QR-Decomposition.

35.

36.

37.

38.

39.

40.

41.

42.

Least Squares.

43.

44.

45.

46.

Orthonormal Diagonal Form.

47.

48.

49.

50.

Eigenpairs of Symmetric Matrices.

51.

52.

53.

54.

Singular Value Decomposition.

55.

56.

57.

58.

59.

60.

61.

62.

Ellipse and the SVD.

63.

64.

65.

66.