Linear Algebra (part 5): Bilinear and Quadratic Forms (by Evan Dummit, 2022, v. 2.00)

Contents

5 Bilinear and Quadratic Forms

In this chapter, we will discuss bilinear and quadratic forms. Bilinear forms are simply linear transformations that are linear in more than one variable, and they will allow us to extend our study of linear phenomena. They are closely related to quadratic forms, which are (classically speaking) homogeneous quadratic polynomials in multiple variables. Despite the fact that quadratic forms are not linear, we can (perhaps surprisingly) still use many of the tools of linear algebra to study them. We will discuss the basic properties of bilinear and quadratic forms and in particular highlighting the notions of positive definiteness and positive semidefiniteness, along with some of their applications in linear algebra, calculus, and geometry. We finish with a discussion of singular value decomposition, which ties together many of the threads we have discussed.

5.1 Bilinear Forms

- We begin by discussing basic properties of bilinear forms on an arbitrary vector space.

5.1.1 Definition, Associated Matrices, Basic Properties

- Let V be a vector space over the field F.
- Definition: A function $\Phi : V \times V \to F$ is a bilinear form on V if it is linear in each variable when the other variable is fixed. Explicitly, this means $\Phi(\mathbf{v}_1 + \alpha\mathbf{v}_2, y) = \Phi(\mathbf{v}_1, \mathbf{w}) + \alpha\Phi(\mathbf{v}_2, \mathbf{w})$ and $\Phi(\mathbf{v}, \mathbf{w}_1 + \alpha\mathbf{w}_2) = \Phi(\mathbf{v}, \mathbf{w}_1) + \alpha\Phi(\mathbf{v}, \mathbf{w}_2)$ for arbitrary $\mathbf{v}_i, \mathbf{w}_i \in V$ and $\alpha \in F$.
  - It is easy to see that the set of all bilinear forms on V forms a vector space under componentwise addition and scalar multiplication.

- Example: An inner product on a real vector space is a bilinear form, but an inner product on a complex vector space is not, since it is conjugate-linear in the second component rather than (actually) linear.
- Example: If $V = F[x]$ and $a, b \in F$, then $\Phi(p, q) = p(a)q(b)$ is a bilinear form on V.
- Example: If $V = C[a, b]$ is the space of continuous functions on $[a, b]$, then $\Phi(f, g) = \int_a^b f(x)g(x)\,dx$ is a bilinear form on V.

- A large class of examples of bilinear forms arise as follows: if $V = F^n$, then for any matrix $A \in M_{n\times n}(F)$, the map $\Phi_A(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T A\mathbf{w}$ is a bilinear form on V.
  - Example: The matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ yields the bilinear form $\Phi_A\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = [x_1\ y_1]\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = x_1x_2 + 2x_1y_2 + 3x_2y_1 + 4y_1y_2$.

- Indeed, if V is finite-dimensional, then by choosing a basis of V we can see that every bilinear form arises in the manner described above:

- Definition: If V is a finite-dimensional vector space, $\beta = \{\beta_1, \dots, \beta_n\}$ is a basis of V, and Φ is a bilinear form on V, the associated matrix of Φ with respect to β is the matrix $[\Phi]_\beta \in M_{n\times n}(F)$ whose (i, j)-entry is the value $\Phi(\beta_i, \beta_j)$.
  - This is the natural analogue of the matrix associated to a linear transformation, for bilinear forms.
  - We will remark that we could give a more general formulation of an associated matrix with respect to two different bases β in the first coordinate and γ in the second coordinate, but we will not need this general formulation at any point.

- Example: For the bilinear form $\Phi((a, b), (c, d)) = 2ac + 4ad - bc$ on F^2, find $[\Phi]_\beta$ for the standard basis $\beta = \{(1, 0), (0, 1)\}$.
  - We simply calculate the four values $\Phi(\beta_i, \beta_j)$ for $i, j \in \{1, 2\}$, where $\beta_1 = (1, 0)$ and $\beta_2 = (0, 1)$.
  - This yields $\Phi(\beta_1, \beta_1) = 2$, $\Phi(\beta_1, \beta_2) = 4$, $\Phi(\beta_2, \beta_1) = -1$, and $\Phi(\beta_2, \beta_2) = 0$.
  - Thus, the associated matrix is $[\Phi]_\beta = \boxed{\begin{bmatrix} 2 & 4 \\ -1 & 0 \end{bmatrix}}$.

- Example: For the bilinear form $\Phi((a, b), (c, d)) = 2ac + 4ad - bc$ on F^2, find $[\Phi]_\gamma$ for the basis $\gamma = \{(2, 1), (-1, 4)\}$.
  - We simply calculate the four values $\Phi(\gamma_i, \gamma_j)$ for $i, j \in \{1, 2\}$, where $\gamma_1 = (2, 1)$ and $\gamma_2 = (-1, 4)$.
  - This yields $\Phi(\gamma_1, \gamma_1) = 14$, $\Phi(\gamma_1, \gamma_2) = 29$, $\Phi(\gamma_2, \gamma_1) = -16$, and $\Phi(\gamma_2, \gamma_2) = -10$.
  - Thus, the associated matrix is $[\Phi]_\gamma = \boxed{\begin{bmatrix} 14 & 29 \\ -16 & -10 \end{bmatrix}}$.

- Example: For the bilinear form $\Phi(p, q) = \int_0^1 p(x)q(x)\,dx$ on $P_2(\mathbb{R})$, find $[\Phi]_\beta$ for the basis $\beta = \{1, x, x^2\}$.
  - We simply calculate the nine values $\Phi(\beta_i, \beta_j)$ for $i, j \in \{1, 2, 3\}$, where $\beta_1 = 1$, $\beta_2 = x$, $\beta_3 = x^2$.
  - For example, $\Phi(\beta_1, \beta_3) = \int_0^1 1 \cdot x^2\,dx = \dfrac{1}{3}$ and $\Phi(\beta_3, \beta_2) = \int_0^1 x^2 \cdot x\,dx = \dfrac{1}{4}$.
  - The resulting associated matrix is $[\Phi]_\beta = \boxed{\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}}$.

- Like with the matrices associated with linear transformations, we can describe how the matrices associated to bilinear forms relate to coordinate vectors, how they transform under a change of basis, and how to use them to translate back and forth between bilinear forms and matrices:

* Proposition (Associated Matrices): Suppose that V is a finite-dimensional vector space, $\beta = \{\beta_1, \dots, \beta_n\}$ is a basis of V, and Φ is a bilinear form on V. Then the following hold:

  1. If $\mathbf{v}$ and $\mathbf{w}$ are any vectors in V, then $\Phi(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_\beta^T [\Phi]_\beta [\mathbf{w}]_\beta$.
     * Proof: If $\mathbf{v} = \beta_i$ and $\mathbf{w} = \beta_j$ then the result follows immediately from the definition of matrix multiplication and the matrix $[\Phi]_\beta$. The result for arbitrary $\mathbf{v}$ and $\mathbf{w}$ then follows by linearity.
  2. The association $\Phi \mapsto [\Phi]_\beta$ of a bilinear form with its matrix representation yields an isomorphism of the space $\mathcal{B}(V)$ of bilinear forms on V with $M_{n\times n}(F)$. In particular, $\dim_F \mathcal{B}(V) = n^2$.
     * Proof: The inverse map is defined as follows: given a matrix $A \in M_{n\times n}(F)$, define a bilinear form Φ_A via $\Phi_A(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_\beta^T A [\mathbf{w}]_\beta$.
     * It is easy to verify that this map is a well-defined linear transformation and that it is inverse to the map given above. The dimension calculation is immediate.
  3. If Φ^T is the reverse form of Φ defined via $\Phi^T(\mathbf{v}, \mathbf{w}) = \Phi(\mathbf{w}, \mathbf{v})$, then $[\Phi^T]_\beta = [\Phi]_\beta^T$.
     * Proof: By definition we have $\Phi^T(\mathbf{v}, \mathbf{w}) = [\mathbf{w}]_\beta^T [\Phi]_\beta [\mathbf{v}]_\beta$. Since the matrix product on the right is a scalar, it is equal to its transpose, which is $[\mathbf{v}]_\beta^T [\Phi]_\beta^T [\mathbf{w}]_\beta$.
     * This means $[\Phi^T]_\beta$ and $[\Phi]_\beta^T$ agree, as bilinear forms, on all pairs of vectors $[\mathbf{v}]_\beta$ and $[\mathbf{w}]_\beta$ in F^n, so they are equal.
  4. If γ is another basis of V and $Q = [I]_\beta^\gamma$ is the change-of-basis matrix from β to γ, then $[\Phi]_\gamma = Q^T [\Phi]_\beta Q$.
     * Proof: By definition, $[\mathbf{v}]_\gamma = Q[\mathbf{v}]_\beta$. Hence $[\mathbf{v}]_\beta^T Q^T [\Phi]_\beta Q [\mathbf{w}]_\beta = [\mathbf{v}]_\gamma^T [\Phi]_\beta [\mathbf{w}]_\gamma$.
     * This means that $Q^T [\Phi]_\beta Q$ and $[\Phi]_\gamma$ agree, as bilinear forms, on all pairs of vectors $[\mathbf{v}]_\beta$ and $[\mathbf{w}]_\beta$ in F^n, so they are equal.

* The last result of the proposition above tells us how bilinear forms behave under change of basis: rather than the more familiar conjugation relation $B = QAQ^{-1}$, we instead have a slightly different relation $B = Q^T AQ$.

* We record one other property of bilinear forms that we will need to make use of later:

* Definition: If Φ is a bilinear form on V and there exists a nonzero vector $\mathbf{x} \in V$ such that $\Phi(\mathbf{x}, \mathbf{v}) = 0$ for all $\mathbf{v} \in V$, we say Φ is degenerate. Otherwise, if there is no such $\mathbf{x}$, we say Φ is nondegenerate.

5.1.2 Symmetric Bilinear Forms and Diagonalization

* In the same way that we classified the linear operators on a vector space that can be diagonalized, we would also like to classify the diagonalizable bilinear forms.

* Definition: If V is finite-dimensional, a bilinear form Φ on V is diagonalizable if there exists a basis β of V such that $[\Phi]_\beta$ is a diagonal matrix.
  * The matrix formulation of this question is as follows: we say that matrices B and C are congruent if there exists an invertible matrix Q such that $C = Q^T BQ$.
  * Then the matrices B and C are congruent if and only if they represent the same bilinear form in different bases (the translation being $B = [\Phi]_\beta$ and $C = [\Phi]_\gamma$, with $Q = [I]_\beta^\gamma$ being the corresponding change-of-basis matrix).
  * Warning: Although we use the same word, diagonalizability for bilinear forms is *not* the same as diagonalizability for linear transformations! Make sure to keep straight the difference between the corresponding matrix versions: two matrices are similar when we can write $B = Q^{-1}AQ$, whereas they are congruent when we can write $B = Q^T AQ$.

* It turns out that when $\text{char}(F) \neq 2$, there is an easy criterion for diagonalizability.

* Definition: A bilinear form Φ on V is symmetric if $\Phi(\mathbf{v}, \mathbf{w}) = \Phi(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in V$.
  * Notice that Φ is symmetric if and only if it equals its reverse form Φ^T.

- ◦ By taking associated matrices, we see immediately that if V is finite-dimensional with basis β, then Φ is a symmetric bilinear form if and only if $[\Phi]_\beta$ is equal to its transpose, which is to say, when it is a symmetric matrix.
  - ◦ Now observe that if Φ is diagonalizable, then $[\Phi]_\beta$ is a diagonal matrix hence symmetric, and thus Φ must be symmetric.

- When the characteristic of F is not equal to 2, the converse holds also:

- Theorem (Diagonalization of Bilinear Forms): Let V be a finite-dimensional vector space over a field F of characteristic not equal to 2. Then a bilinear form on V is diagonalizable if and only if it is symmetric.
  - ◦ Proof: The forward direction was established above. For the reverse, we show the result by induction on $n = \dim_F V$. The base case $n = 1$ is trivial, so suppose the result holds for all spaces of dimension less than n, and let Φ be symmetric on V.
  - ◦ If Φ is the zero form, then clearly Φ is diagonalizable. Otherwise, suppose Φ is not identically zero: we claim there exists a vector $\mathbf{x}$ with $\Phi(\mathbf{x}, \mathbf{x}) \neq 0$.
  - ◦ By hypothesis, Φ is not identically zero so suppose that $\Phi(\mathbf{v}, \mathbf{w}) \neq 0$. If $\Phi(\mathbf{v}, \mathbf{v}) \neq 0$ or $\Phi(\mathbf{w}, \mathbf{w}) \neq 0$ we may take $\mathbf{x} = \mathbf{v}$ or $\mathbf{x} = \mathbf{w}$. Otherwise, we have $\Phi(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = \Phi(\mathbf{v}, \mathbf{v}) + 2\Phi(\mathbf{v}, \mathbf{w}) + \Phi(\mathbf{w}, \mathbf{w}) = 2\Phi(\mathbf{v}, \mathbf{w}) \neq 0$ by the assumption that $\Phi(\mathbf{v}, \mathbf{w}) \neq 0$ and $2 \neq 0$ in F (here is where we require the characteristic not to equal 2), and so we may take $\mathbf{x} = \mathbf{v} + \mathbf{w}$.
  - ◦ Now consider the linear functional $T : V \to F$ given by $T(\mathbf{v}) = \Phi(\mathbf{x}, \mathbf{v})$. Since $T(\mathbf{x}) = \Phi(\mathbf{x}, \mathbf{x}) \neq 0$, we see that $\operatorname{im}(T) = F$, so $\dim_F \ker(T) = n - 1$ by the nullity-rank theorem.
  - ◦ Then the restriction of Φ to $\ker(T)$ is clearly a symmetric bilinear form on $\ker(T)$, so by induction, there exists a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$ of $\ker(T)$ such that the restriction of Φ is diagonalized by this basis, which is to say, $\Phi(\mathbf{v}_i, \mathbf{v}_j) = 0$ for $i \neq j$.
  - ◦ Now observe that since $\mathbf{x} \notin \ker(T)$, the set $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}, \mathbf{x}\}$ is a basis of V. Since $\Phi(\mathbf{x}, \mathbf{v}_i) = \Phi(\mathbf{v}_i, \mathbf{x}) = 0$ for all $i < n$ by definition of T, we conclude that β diagonalizes Φ, as required.

- We will note that the assumption that $\operatorname{char}(F) \neq 2$ in the theorem above cannot be removed.
  - ◦ Explicitly, if $F = \mathbb{F}_2$ is the field with 2 elements and Φ is the bilinear form on F^2 with associated matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then Φ is symmetric but cannot be diagonalized.
  - ◦ Explicitly, suppose $Q = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$: then $Q^T A Q = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & ad+bc \\ ad+bc & 0 \end{bmatrix}$, so the only possible diagonalization of Φ would be the zero matrix, but that is impossible because Φ is not the zero form.
  - ◦ In this example we can see that[1] $\Phi(\mathbf{x}, \mathbf{x}) = 0$ for all $\mathbf{x} \in F^2$, which causes the inductive argument to fail.

- As an immediate corollary, we see that every symmetric matrix is congruent to a diagonal matrix in characteristic $\neq 2$:

- Corollary: If $\operatorname{char}(F) \neq 2$, then every symmetric matrix over F is congruent to a diagonal matrix.
  - ◦ Proof: The result follows immediately by diagonalizing the corresponding bilinear form.

- We can give an explicit procedure for writing a symmetric matrix S in the form $D = Q^T S Q$ that is similar to the algorithm for computing the inverse of a matrix.
  - ◦ Recall that if E is an elementary row matrix (obtained by performing an elementary row operation on the identity matrix), then EA is the matrix obtained by performing that elementary row operation on A.
  - ◦ Likewise, if C is an elementary column matrix, then AC is the matrix obtained by performing that elementary column operation on A.

[1]This follows by noting that $\Phi(\beta_i, \beta_i) = 0$ for each basis element β_i. Then if $\mathbf{v} = a_1\beta_1 + \cdots + a_n\beta_n$, expanding $\Phi(\mathbf{v}, \mathbf{v})$ linearly and applying symmetry shows that every term $a_i a_j \Phi(\beta_i, \beta_j)$ for $i \neq j$ has a coefficient of 2, so $\Phi(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v}$.

- Hence if E is an elementary row matrix, then EAE^T is the matrix obtained by performing the elementary row operation on A (given by E) and then the corresponding elementary column operation (given by E^T).
- Since the invertible matrix Q is a product $E_1 \cdots E_d$ of elementary row matrices, we see that $Q^T SQ = E_d^T \cdots E_1^T S E_1 \cdots E_d$ is obtained from S by performing a sequence of these paired row-column operations.
- Our result on diagonalization above ensures that there is a sequence of these operations that will yield a diagonal matrix.
- We may find the proper sequence of operations by performing these "paired" operations using a method similar to row-reduction: using the (1,1)-entry, we apply row operations to clear out all the entries in the first column below it. (If this entry is zero, we add an appropriate multiple of another row to the top row to make it nonzero.)
- This will also clear out the column entries to the right of the (1,1)-entry, yielding a matrix whose first row and column are now diagonalized. We then restrict attention to the smaller $(n-1) \times (n-1)$ matrix excluding the first row and column, and repeat the procedure recursively until the matrix is diagonalized.
- Then we may obtain the matrix $Q^T = E_d^T \cdots E_1^T I$ by applying all of the elementary row operations (in the same order) starting with the identity matrix.

- We may keep track of these operations using a "double matrix" as in the algorithm for computing the inverse of a matrix: on the left we start with the symmetric matrix S, and on the right we start with the identity matrix I.
  - At each step, we select a row operation and perform it, and its corresponding column operation, on the left matrix. We also perform the row operation (but *only* the row operation!) on the right matrix.
  - When we are finished, we will have transformed the double-matrix $[S|I]$ into the double-matrix $[D|Q^T]$, and we will have $Q^T SQ = D$.

- Example: For $S = \begin{bmatrix} 1 & 3 \\ 3 & -4 \end{bmatrix}$, find an invertible matrix Q and diagonal matrix D such that $Q^T SQ = D$.
  - We set up the double matrix and perform row/column operations as listed (to emphasize again, we perform the row and then the corresponding column operation on the left side, but *only the row operation* on the right side):

$$\left[\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 3 & -4 & 0 & 1 \end{array}\right] \xrightarrow[C_2-3C_1]{R_2-3R_1} \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & -13 & -3 & 1 \end{array}\right]$$

  - The matrix on the left is now diagonal.
  - Thus, we may take $D = \begin{bmatrix} 1 & 0 \\ 0 & -13 \end{bmatrix}$ with $Q^T = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ and thus $Q = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$. Indeed, one may double-check that $Q^T SQ = D$, as claimed.

- Example: For $S = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 3 & 0 & 2 \end{bmatrix}$, find an invertible matrix Q and diagonal matrix D such that $Q^T SQ = D$.
  - We set up the double matrix and perform row/column operations as listed:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 3 & 0 & 2 & 0 & 0 & 1 \end{array}\right] \xrightarrow[C_2-2C_1]{R_2-2R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -2 & 1 & 0 \\ 3 & -6 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[C_3-3C_1]{R_3-3R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -3 & -6 & -2 & 1 & 0 \\ 0 & -6 & -7 & -3 & 0 & 1 \end{array}\right] \xrightarrow[C_3-2C_2]{R_3-2R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & -2 & 1 & 0 \\ 0 & 0 & 5 & 1 & -2 & 1 \end{array}\right]$$

  - The matrix on the left is now diagonal.
  - Thus, we may take $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ with $Q^T = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$ and thus $Q = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$. Indeed, one may double-check that $Q^T SQ = D$, as claimed.

5.2 Quadratic Forms

- In the proof that symmetric bilinear forms are diagonalizable, the existence of a vector $\mathbf{x} \in V$ such that $\Phi(\mathbf{x}, \mathbf{x}) \neq 0$ played a central role. We now examine this non-linear function $\Phi(\mathbf{x}, \mathbf{x})$ more closely.

5.2.1 Definition and Basic Properties

- Definition: If Φ is a symmetric bilinear form on V, the function $Q : V \to F$ given by $Q(\mathbf{v}) = \Phi(\mathbf{v}, \mathbf{v})$ is called the quadratic form associated to Φ.
  - Example: If Φ is the symmetric bilinear form with matrix $A = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}$ over F^2, then the corresponding quadratic form has $Q(\begin{bmatrix} x \\ y \end{bmatrix}) = x^2 + 6xy + 4y^2$. (The fact that this is a homogeneous[2] quadratic function of the entries of the input vector is the reason for the name "quadratic form".)
  - Example: If Φ is an inner product $\langle \cdot, \cdot \rangle$ on a real vector space, then the associated quadratic form is $Q(\mathbf{v}) = ||\mathbf{v}||^2$, the square of the norm of $\mathbf{v}$.

- Clearly, Q is uniquely determined by Φ. When $\mathrm{char}(F) \neq 2$, the reverse holds as well.
  - Explicitly, since $Q(\mathbf{v} + \mathbf{w}) = \Phi(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = Q(\mathbf{v}) + 2\Phi(\mathbf{v}, \mathbf{w}) + Q(\mathbf{w})$, we can write $\Phi(\mathbf{v}, \mathbf{w}) = \dfrac{1}{2}[Q(\mathbf{v} + \mathbf{w}) - Q(\mathbf{v}) - Q(\mathbf{w})]$, and so we may recover Φ from Q.
  - Also, observe that for any scalar $\alpha \in F$, we have $Q(\alpha\mathbf{v}) = \Phi(\alpha\mathbf{v}, \alpha\mathbf{v}) = \alpha^2\Phi(\mathbf{v}, \mathbf{v}) = \alpha^2 Q(\mathbf{v})$.
  - This last two relations provide us a way to define a quadratic form without explicit reference to the underlying symmetric bilinear form.

- Definition: If V is a vector space, a quadratic form is a function $Q : V \to F$ such that $Q(\alpha\mathbf{v}) = \alpha^2 Q(\mathbf{v})$ for all $\alpha \in F$, and the function $Q(\mathbf{v} + \mathbf{w}) - Q(\mathbf{v}) - Q(\mathbf{w})$ is a bilinear form in $\mathbf{v}$ and $\mathbf{w}$.
  - By setting $\alpha = 0$ we see $Q(\mathbf{0}) = 0$, and by setting $\alpha = -1$ we see $Q(-\mathbf{v}) = Q(\mathbf{v})$.
  - Like with bilinear forms, the set of all quadratic forms on V forms a vector space.

- Example (again): Show that the function $Q[(x, y)] = x^2 + 6xy + 4y^2$ is a quadratic form on F^2.
  - First observe that $Q[\alpha(x, y)] = (\alpha x)^2 + 6(\alpha x)(\alpha y) + 4(\alpha y)^2 = \alpha^2(x^2 + 6xy + 4y^2) = \alpha^2 Q(x, y)$.
  - We also see that $Q[(x_1, y_1) + (x_2, y_2)] - Q[(x_1, y_1)] - Q[(x_2, y_2)] = 2x_1x_2 + 6x_1y_2 + 6x_2y_1 + 8y_1y_2$. It is straightforward to verify that this is a bilinear form by checking the linearity explicitly.
  - Alternatively (at least when $\mathrm{char}(F) \neq 2$) we can write down the associated bilinear form $\Phi((a, b), (c, d)) = \dfrac{1}{2}[Q[(a + c, b + d)] - Q[(a, c)] - Q[(b, d)]] = ac + 3ad + 3bc + 4bd$, and this is the bilinear form associated to the matrix $\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}$, as indeed we saw earlier.

- Example: If $V = C[a, b]$, show that the function $Q(f) = \int_a^b f(x)^2\,dx$ is a quadratic form on V.
  - First, we have $Q(\alpha f) = \int_a^b [\alpha f(x)]^2\,dx = \alpha^2 \int_a^b f(x)^2\,dx = \alpha^2 Q(f)$.
  - Also, we have $Q(f+g) - Q(f) - Q(g) = \int_a^b [f(x) + g(x)]^2\,dx - \int_a^b f(x)^2\,dx - \int_a^b g(x)^2\,dx = \int_a^b 2f(x)g(x)\,dx$, and this is indeed a bilinear form in f and g.

- If $\mathrm{char}(F) \neq 2$, then the function $\dfrac{1}{2}[Q(\mathbf{v} + \mathbf{w}) - Q(\mathbf{v}) - Q(\mathbf{w})]$ is the bilinear pairing associated to Q. It is not hard to see that we obtain a correspondence between quadratic forms and bilinear pairings in this case, since we may recover a bilinear pairing from each quadratic form and a quadratic form from each bilinear pairing.

[2] A polynomial in several variables is called homogeneous if all of the terms have the same total degree. For example, $x^3 + 3xy^2 - 2y^3$ and $2xyz$ are both homogeneous of degree 3, while $x^4z + 2w^2y^3 - p^5$ is homogeneous of degree 5. But $x^2 + y$ and $x^3y + y^4 - x^2$ are not homogeneous because they both have terms of different degrees.

- In particular, any homogeneous quadratic function on F^n (i.e., any polynomial function all of whose terms have total degree 2) is a quadratic form on F^n: for variables $x_1, \dots, x_n$, such a function has the general form $\sum_{1 \le i \le j \le n} a_{i,j} x_i x_j$. [3]
- Then we can see that the associated matrix A for the corresponding bilinear form has entries $a_{i,j} = a_{j,i} = \begin{cases} a_{i,i} & \text{for } i = j \\ a_{i,j}/2 & \text{for } i \ne j \end{cases}$; this of course requires $\text{char}(F) \ne 2$ in order to be able to divide by 2.
- Example: The function $Q(x_1, x_2) = 7x_1^2 - 4x_1x_2 + 3x_2^2$ is a quadratic form on F^2. The matrix for the associated symmetric bilinear form is $\begin{bmatrix} 7 & -2 \\ -2 & 3 \end{bmatrix}$.
- Example: The function $Q(x_1, x_2, x_3) = x_1^2 + 2x_1x_3 - 3x_2x_3 + 4x_3^2$ is a quadratic form on F^3. When $\text{char}(F) \ne 2$, the matrix for the associated symmetric bilinear form is $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & -\frac{3}{2} \\ 1 & -\frac{3}{2} & 4 \end{bmatrix}$.
- Example: The function $Q(x_1, \dots, x_n) = x_1^2 + 2x_2^2 + 3x_3^2 + \cdots + nx_n^2$ is a quadratic form on F^n. Its associated matrix is the diagonal matrix with entries $1, 2, \dots, n$.

5.2.2 Quadratic Forms Over $\mathbb{R}^n$: Diagonalization of Quadratic Varieties

- In the event that V is a finite-dimensional vector space over $F = \mathbb{R}$, quadratic forms are particularly pleasant. By choosing a basis we may assume that $V = \mathbb{R}^n$ for concreteness.
  - Then, per the real spectral theorem, any real symmetric matrix is orthogonally diagonalizable, meaning that if S is any real symmetric matrix, then there exists an orthogonal matrix Q (with $Q^T = Q^{-1}$) such that $QSQ^{-1} = D$ is diagonal.
  - But since $Q^T = Q^{-1}$, if we take $R = Q^T$ then this condition is the same as saying $R^T SR = D$ is diagonal. This is precisely the condition we require in order to diagonalize a symmetric bilinear form.
  - Hence: we may diagonalize a symmetric bilinear form over $\mathbb{R}$ by computing the (regular) diagonalization of the corresponding matrix: this is quite efficient as it only requires finding the eigenvalues and eigenvectors.
  - The corresponding diagonalization represents "completing the square" in the quadratic form via a change of variables that is orthogonal (i.e., arises from an orthonormal basis), which corresponds geometrically to a rotation of the standard coordinate axes, possibly also with a reflection.

- Example: Find the matrix associated to the quadratic form $Q(x, y) = 4x^2 - 4xy + 7y^2$, and also find an orthonormal basis of $\mathbb{R}^2$ that diagonalizes Q.
  - We can read off the associated matrix from the coefficients as $A = \boxed{\begin{bmatrix} 4 & -2 \\ -2 & 7 \end{bmatrix}}$.
  - To diagonalize Q, we diagonalize A by finding the eigenvalues and eigenvectors of A.
  - The characteristic polynomial is $p(t) = (t-4)(t-7) - 4 = t^2 - 11t + 24 = (t-3)(t-8)$, so the eigenvalues are $\lambda = 3, 8$.
  - We can compute eigenvectors $(2, 1)$ and $(-1, 2)$ for $\lambda = 3, 8$ respectively, so upon normalizing these eigenvectors, we see that we can take $\gamma = \boxed{\left\{ \frac{1}{\sqrt{5}}(2, 1), \frac{1}{\sqrt{5}}(-1, 2) \right\}}$
  - Explicitly, with $Q = \dfrac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$, we have $[\Phi]_\gamma = Q^T [\Phi]_\beta Q = \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}$.
  - In terms of the quadratic form, this says for $x' = \frac{1}{\sqrt{5}}(2x + y)$ and $y' = \frac{1}{\sqrt{5}}(-x + 2y)$, we have $Q(x, y) = 4x^2 - 4xy + 7y^2 = 3(x')^2 + 8(y')^2$. Note that by changing basis in this manner, we have eliminated the cross-term $-4xy$ in the original quadratic form Q.

[3] It can also be verified directly from the definition that this is a quadratic form via some mild calculations; this also shows the statement is true even when $\text{char}(F) = 2$.

- Example: Find an orthonormal change of basis that diagonalizes the quadratic form $Q(x, y, z) = 5x^2 + 4xy + 6y^2 + 4yz + 7z^2$ over $\mathbb{R}^3$.
  - We simply diagonalize the matrix for the corresponding bilinear form, which is $A = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{bmatrix}$. The characteristic polynomial is $p(t) = \det(tI_3 - A) = t^3 - 18t^2 + 99t - 162 = (t-3)(t-6)(t-9)$, so the eigenvalues are $\lambda = 3, 6, 9$.
  - Computing a basis for each eigenspace yields eigenvectors $(2, -2, 1)$, $(-2, -1, 2)$, and $(1, 2, 2)$ for $\lambda = 3, 6, 9$ respectively.
  - Hence, after normalizing, we may take $Q = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ -2 & -1 & 2 \\ 1 & 2 & 2 \end{bmatrix}$, so that $Q^T = Q^{-1}$ and $QAQ^{-1} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9 \end{bmatrix} = D$.
  - Therefore the desired change of basis is $x' = \frac{1}{3}(2x - 2y + z)$, $y' = \frac{1}{3}(-2x - y + 2z)$, $z' = \frac{1}{3}(x + 2y + 2z)$, and with this change of basis it is not hard to verify that, indeed, $Q(x, y, z) = 3(x')^2 + 6(y')^2 + 9(z')^2$.
- One application of the existence of such a diagonalization is to classify the conic sections in $\mathbb{R}^2$, and the quadric surfaces in $\mathbb{R}^3$. These curves (in $\mathbb{R}^2$) and surfaces (in $\mathbb{R}^3$)
  - For conics in $\mathbb{R}^2$, the general equation is $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. By diagonalizing, we may eliminate the xy term, and so the quadratic term can be put into the form $Ax^2 + Cy^2$. We then have various cases depending on the signs of A and C.
  - If A and C are both zero then the conic degenerates to a line. If one is zero and the other is not, then by rescaling and swapping variables we may assume $A = 1$ and $C = 0$, in which case the equation $x^2 + Dx + Ey + F = 0$ yields a parabola upon solving for y.
  - If both A, C are nonzero, then we may complete the square to eliminate the linear terms, and then rescale so that $F = -1$. The resulting equation then has the form $A'x^2 + C'y^2 = 1$. If A', C' have the same sign, then the curve is an ellipse, while if A', C' have the opposite sign, the curve is a hyperbola.
  - For quadric surfaces in $\mathbb{R}^3$ we may likewise eliminate cross-terms by diagonalizing, which yields a reduced equation $Ax^2 + By^2 + Cz^2 + Dx + Ey + Fz + G = 0$.
  - We can then perform a similar analysis (based on how many of A, B, C are zero and the relative signs of the nonzero coefficients and the linear terms) to obtain all of the possible quadric surfaces in $\mathbb{R}^3$.
  - In addition to the "degenerate" surfaces (e.g., a point, a plane, two planes), after rescaling the variables, one obtains 9 different quadric surfaces: the ellipsoid (e.g., $x^2 + y^2 + z^2 = 1$), the elliptic, parabolic, and hyperbolic cylinders (e.g., $x^2 + y^2 = 1$, $y = x^2$, and $x^2 - y^2 = 1$), the hyperboloid of one sheet (e.g., $z^2 - x^2 - y^2 = 1$), the elliptical cone (e.g., $z^2 = x^2 + y^2$), the hyperboloid of two sheets (e.g., $x^2 + y^2 - z^2 = 1$), the elliptic paraboloid (e.g., $z = x^2 + y^2$), and the hyperbolic paraboloid (e.g., $z = x^2 - y^2$).
  - Seven of the quadric surfaces are plotted in Figure 1 (the parabolic and hyperbolic cylinders are omitted).
- All of the conics and quadric surfaces are examples of <u>algebraic varieties</u>, which are the solution sets of polynomial equations in several variables.
  - If we have a general quadratic variety (i.e., a quadratic polynomial equation in n variables), we can make an appropriate translation and rescaling to convert it to the form $Q(x_1, \dots, x_n) = 1$ or 0, where Q is a quadratic form.
  - By diagonalizing the corresponding quadratic form using an orthonormal change of basis (which corresponds to a rotation of the coordinates axes and possibly also a reflection), we can then determine the shape of the variety's graph in $\mathbb{R}^n$.

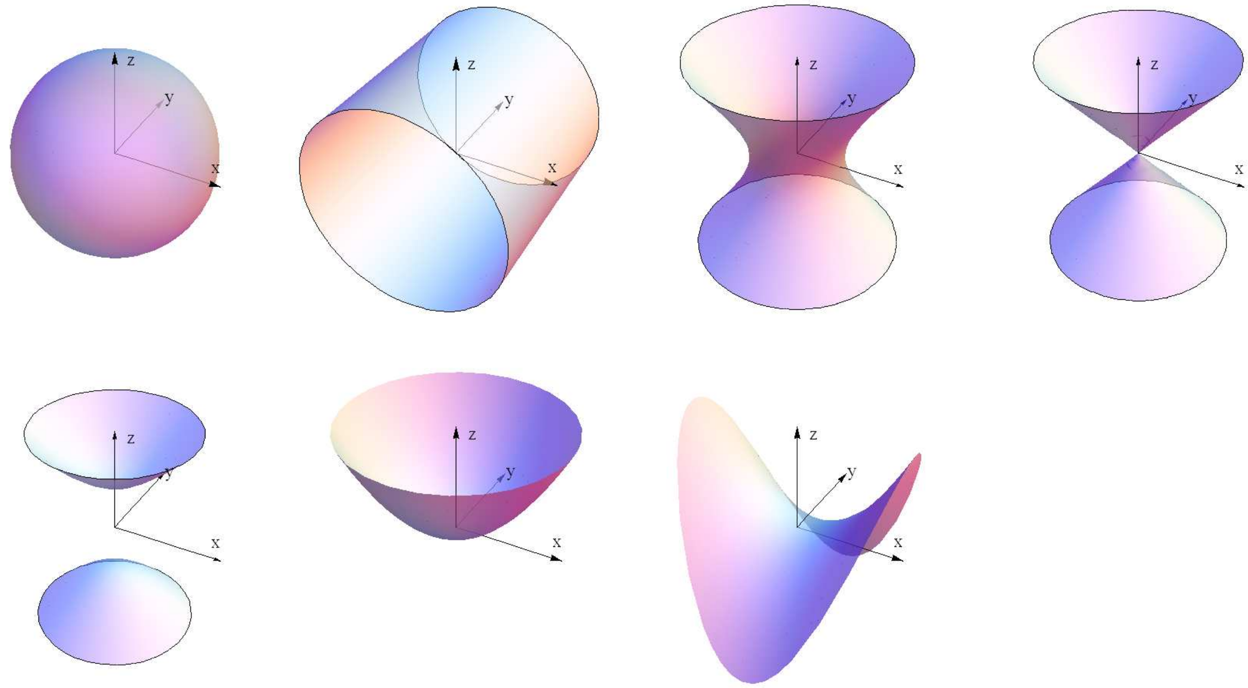


Figure 1: (Top row) Ellipsoid, Circular Cylinder, Hyperboloid of One Sheet, Cone,
(Bottom row) Hyperboloid of Two Sheets, Elliptic Paraboloid, Hyperbolic Paraboloid

- Example: Diagonalize the quadratic form $Q(x, y) = 2x^2 - 4xy - y^2$. Use the result to describe the shape of the conic section $Q(x, y) = 1$ in $\mathbb{R}^2$.
  - The matrix associated to the corresponding bilinear form is $A = \begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}$.
  - The characteristic polynomial is $p(t) = \det(tI_2 - A) = t^3 - t + 6$ with eigenvalues $\lambda = 3, -2$.
  - We need to diagonalize A using an orthonormal basis of eigenvectors. Since the eigenvalues are distinct, we simply compute a basis for each eigenspace: doing so yields eigenvectors $(-2, 1)$ and $(1, 2)$ for $\lambda = 3, -2$ respectively.
  - Thus, we may diagonalize A via the orthogonal matrix $Q = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$, and the resulting diagonalization is $Q(x, y, z) = 3(x')^2 - 2(y')^2$.
  - In particular, since the change of basis is orthonormal, in the new coordinate system the equation $Q(x, y, z) = 1$ reads simply as $3(x')^2 - 2(y')^2 = 1$.
  - By rescaling again, with $x'' = \sqrt{3}x'$, $y'' = \sqrt{2}y'$, this is equivalent to $(x'')^2 - (y'')^2 = 1$, which is a $\boxed{\text{hyperbola}}$.

- Example: Diagonalize the quadratic form $Q(x, y, z) = 2x^2 + 4xy - 20xz + 11y^2 + 16yz + 5z^2$. Use the result to describe the shape of the quadric surface $Q(x, y, z) = 1$ in $\mathbb{R}^3$.
  - The matrix associated to the corresponding bilinear form is $A = \begin{bmatrix} 2 & 2 & -10 \\ 2 & 11 & 8 \\ -10 & 8 & 5 \end{bmatrix}$.
  - The characteristic polynomial is $p(t) = \det(tI_3 - A) = t^3 - 18t^2 - 81t + 1458 = (t+9)(t-9)(t-18)$ with eigenvalues $\lambda = 9, 18, -9$.
  - We need to diagonalize A using an orthonormal basis of eigenvectors. Since the eigenvalues are distinct, we simply compute a basis for each eigenspace: doing so yields eigenvectors $(-2, -2, 1)$, $(-1, 2, 2)$, $(2, -1, 2)$, for $\lambda = 9, 18, -9$ respectively.

- Thus, we may diagonalize A via the orthogonal matrix $Q = \frac{1}{3}\begin{bmatrix} -2 & -1 & 2 \\ -2 & 2 & -1 \\ 1 & 2 & 2 \end{bmatrix}$, and the resulting diagonalization is $Q(x, y, z) = 9(x')^2 + 18(y')^2 - 9(z')^2$.
- In particular, since the change of basis is orthonormal, in the new coordinate system the equation $Q(x, y, z) = 1$ reads simply as $9(x')^2 + 18(y')^2 - 9(z')^2 = 1$.
- By rescaling again, with $x'' = 3x'$, $y'' = 3\sqrt{2}y'$, $z'' = 3z'$, this is equivalent to $(x'')^2 + (y'')^2 - (z'')^2 = 1$, which is a $\boxed{\text{hyperboloid of one sheet}}$.

5.2.3 Definiteness of Real Quadratic Forms

- One of the main properties of a real quadratic form that characterizes its behavior is whether it takes positive values, negative values, or both:
- Definition: If V is a real vector space, a quadratic form Q on V is positive definite if $Q(\mathbf{v}) > 0$ for every nonzero vector $\mathbf{v} \in V$, it is negative definite if $Q(\mathbf{v}) < 0$ for every nonzero vector $\mathbf{v} \in V$, and it is indefinite if Q takes both positive and negative values.
  - Example: If V is a real inner product space, then the square of the norm function $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle$ is a positive-definite quadratic form on V. Indeed, it is not hard to see that the underlying bilinear pairing Φ associated with Q is an inner product precisely when Q is a positive-definite quadratic form.
  - Example: The quadratic form $Q(x, y) = x^2 + 2y^2$ is positive definite, since $Q(x, y) > 0$ for all $(x, y) \neq (0, 0)$.
  - Example: The quadratic form $Q(x, y, z) = -2x^2 - 2xy - 5y^2 = -(x - y)^2 - (x + 2y)^2$ is negative definite, since the second expression shows that $Q(x, y) < 0$ for all $(x, y) \neq (0, 0)$.
  - Example: The quadratic form $Q(x, y) = xy$ is indefinite, since $Q(1, 1) = 1$ and $Q(-1, 1) = -1$, so Q takes both positive and negative values.
  - There are also useful weaker versions of these conditions: we say Q is positive semidefinite if $Q(\mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$ and negative semidefinite if $Q(\mathbf{v}) \leq 0$ for all $\mathbf{v} \in V$.
  - Example: The quadratic form $Q(x, y) = x^2$ is positive semidefinite, since $Q(x, y) \geq 0$ for all (x, y), but it is not positive definite because $Q(0, 1) = 0$.
  - It is easy to see that Q is positive (semi)definite if and only if $-Q$ is negative (semi)definite, so for example by the above we see that $Q(x, y) = -x^2$ is negative semidefinite.
  - Notice that every nonzero quadratic form lies in exactly one of the following five classes: positive-definite, positive-semidefinite but not positive-definite, indefinite, negative-semidefinite but not negative definite, negative definite.
- By diagonalizing a quadratic form, we can easily determine its definiteness:
- Proposition (Definiteness and Eigenvalues): If Q is a quadratic form on a finite-dimensional real vector space V with associated matrix A, then Q is positive definite if and only if all eigenvalues of A are positive, Q is positive semidefinite if and only if all eigenvalues of A are nonnegative, Q is negative definite if and only all eigenvalues of A are negative, Q is negative semidefinite if and only if all eigenvalues of A are nonpositive, and Q is indefinite if and only if it has both a positive and a negative eigenvalue.
  - Note that all of the eigenvalues of A are real, by the spectral theorem.
  - Proof: Observe that definiteness is unaffected by changing basis, because each of the definiteness conditions $Q(\mathbf{v}) > 0$, $Q(\mathbf{v}) \geq 0$, $Q(\mathbf{v}) < 0$, $Q(\mathbf{v}) \leq 0$ are statements about all vectors $\mathbf{v}$ in the vector space V.
  - Therefore, we may diagonalize Q without affecting its definiteness. After diagonalizing, we have an expression of the form $Q(x_1, x_2, \ldots, x_n) = a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2$.
  - If any of the coefficients are negative, then Q necessarily takes negative values (specifically, if $a_i < 0$, then Q evaluated at the ith standard unit vector will be a_i).

- Likewise, if any of the coefficients are positive then Q necessarily takes positive values, and if any coefficients are zero or have opposite signs then Q takes the value 0 at some nonzero vector.
- Assuming we use an orthogonal diagonalization, then since the coefficients a_i are simply the eigenvalues of A, all of the claimed statements then follow immediately.
- Explicitly, if Q takes only positive values on nonzero vectors then no coefficients a_i can be zero or negative (so they are all positive), if Q takes only nonnegative values then no coefficients a_i can be negative, and likewise in the other two cases.

- Example: Determine the definiteness of the quadratic form $Q(x, y) = 2x^2 + 4xy + 5y^2$ on $\mathbb{R}^2$.
  - The associated matrix is $\begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$, whose characteristic polynomial is $p(t) = (t-2)(t-5) - (2)(2) = t^2 - 7t + 6 = (t-6)(t-1)$, so its eigenvalues are $\lambda = 1, 6$.
  - Since both eigenvalues are positive, Q is $\boxed{\text{positive definite}}$.

- Example: Determine the definiteness of the quadratic form $Q(x, y) = 3x^2 - 6xy - 5y^2$ on $\mathbb{R}^2$.
  - The associated matrix is $\begin{bmatrix} 3 & -3 \\ -3 & -5 \end{bmatrix}$, whose characteristic polynomial is $p(t) = (t-3)(t+5) - (-3)(-3) = t^2 - 2t - 24 = (t+6)(t-4)$, so its eigenvalues are $\lambda = -6, 4$.
  - Since one eigenvalue is positive and the other is negative, Q is $\boxed{\text{indefinite}}$.

- Example: Determine the definiteness of the quadratic form $Q(x, y) = 3x^2 - 2xy + 4xz + 3y^2 - 4yz + 2z^2$ on $\mathbb{R}^3$.
  - The associated matrix is $\begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 2 \end{bmatrix}$, whose characteristic polynomial is $p(t) = t^3 - 8t^2 + 12t = t(t-2)(t-6)$, so its eigenvalues are $\lambda = 0, 2, 6$.
  - Since one eigenvalue is zero and the others are positive, Q is $\boxed{\text{positive semidefinite}}$.

- We can also view definiteness as a property of symmetric matrices themselves by considering the definiteness of the associated quadratic form. In this lens, we can give another way to identify definiteness using determinants:

- Theorem (Sylvester's Criterion): Suppose A is an $n \times n$ symmetric real matrix. For each $1 \le k \le n$, define the kth principal minor A_k to be the upper-left $k \times k$ corner of A. Then A is positive definite if and only if $\det(A_k) > 0$ for all k.
  - There is also a version for checking positive semidefiniteness, but it is more complicated: one requires $\det(A_k) \ge 0$ for every $k \times k$ symmetric submatrix of A, rather than just the principal minors.
  - Also, since A is positive definite if and only if $-A$ is negative definite, one can also use Sylvester's criterion to identify negative definite matrices.
  - We will not prove Sylvester's criterion, although it is not hard to see that the given condition is necessary, since if A is positive definite we must have $\mathbf{x}^T A \mathbf{x} > 0$ for all vectors $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_k \; 0 \; \cdots \; 0]$: this means the matrix A_k must also be positive definite and therefore must have positive determinant.

- Example: Determine the definiteness of the matrix $A = \begin{bmatrix} 2 & -1 & 2 \\ -1 & 4 & 2 \\ 2 & 2 & 5 \end{bmatrix}$.
  - The principal minors are $[2]$, $\begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$, and $\begin{bmatrix} 2 & -1 & 2 \\ -1 & 4 & 2 \\ 2 & 2 & 5 \end{bmatrix}$, whose determinants respectively are 2, 7, and 3.
  - Since all of the principal minors have positive determinants, the given matrix is $\boxed{\text{positive definite}}$.
  - Remark: To four decimal places, the eigenvalues are 6.8004, 4.0917, and 0.1078, so we see A is indeed positive definite.

5.2.4 Quadratic Forms Over $\mathbb{R}^n$: The Second Derivatives Test

- We can also use quadratic forms and an analysis of definiteness to establish the famous "second derivatives test" from multivariable calculus:
- Theorem (Second Derivatives Test in $\mathbb{R}^n$): Suppose f is a function of n variables $x_1, \dots, x_n$ that is twice-differentiable and P is a critical point of f, so that $f_{x_i}(P) = 0$ for each i. Let H be the Hessian matrix, whose (i,j)-entry is the second-order partial derivative $f_{x_i x_j}(P)$. If all eigenvalues of H are positive then f has a local minimum at P, if all eigenvalues of H are negative then f has a local maximum at P, if H has at least one eigenvalue of each sign then f has a saddle point at P, and in all other cases (where H has at least one zero eigenvalue and does not have one of each sign) the test is inconclusive.
  - Proof (outline): By translating appropriately, assume for simplicity that P is at the origin.
  - Then by the multivariable version of Taylor's theorem in $\mathbb{R}^2$, the function $f(x_1, \dots, x_n) - f(P)$ will be closely approximated by its degree-2 Taylor polynomial T, which has the form $T = \sum_{1 \le i \le j \le n} a_{i,j} x_i x_i$, where $a_{i,j} = \begin{cases} f_{x_i, x_i}(P)/2 & \text{for } i = j \\ f_{x_i, x_j}(P) & \text{for } i \neq j \end{cases}$.
  - Specifically, Taylor's theorem says that $\lim_{(x_1,\dots,x_n)\to P} \dfrac{f(x_1, \dots, x_n) - T - f(P)}{x_1^2 + \cdots + x_n^2} = 0$, which we can write more compactly as $f(x_1, \dots, x_n) - f(P) = T + O(x_1^2 + \cdots + x_n^2)$.
  - Now observe T is a quadratic form whose associated bilinear form has matrix $H/2$. By using an orthonormal change of basis, we may diagonalize this quadratic form, and the entries on the diagonal of the diagonalization are the eigenvalues of $H/2$.
  - If $x_1', \dots, x_n'$ is the new coordinate system, this means $f(x_1, \dots, x_n) - f(P) = \frac{1}{2}\lambda_1 (x_1')^2 + \cdots + \frac{1}{2}\lambda_n (x_n')^2 + O[(x_1')^2 + \cdots + (x_n')^2]$.
  - If all of the λ_i are positive (i.e., when $\frac{1}{2}\lambda_1 (x_1')^2 + \cdots + \frac{1}{2}\lambda_n (x_n')^2$ is positive definite) then the error term is smaller than the remaining terms, and so we see that $f(x_1, \dots, x_n) - f(P) > 0$ sufficiently close to P, meaning that P is a local minimum.
  - Likewise, if all of the λ_i are negative (i.e., when $\frac{1}{2}\lambda_1 (x_1')^2 + \cdots + \frac{1}{2}\lambda_n (x_n')^2$ is negative definite) then the error term is smaller than the remaining terms, and so we see that $f(x_1, \dots, x_n) - f(P) < 0$ sufficiently close to P, meaning that P is a local maximum.
  - If there is at least one positive eigenvalue λ_i and one negative eigenvalue λ_j (i.e., when $\frac{1}{2}\lambda_1 (x_1')^2 + \cdots + \frac{1}{2}\lambda_n (x_n')^2$ is indefinite) then approaching P along the direction x_i' yields values of f less than P, while approaching P along the direction x_j' yields values of f greater than P, so P is a saddle point.
  - The other cases (i.e., when $\frac{1}{2}\lambda_1 (x_1')^2 + \cdots + \frac{1}{2}\lambda_n (x_n')^2$ is positive or negative semidefinite) are inconclusive: if $\lambda_1 = 0$ then as we approach P along the direction x_1', the quadratic form T is constant, and so the sign of $f(x_1, \dots, x_n) - f(P)$ along that path will be determined by the error term. Except in the case where the function is known to take both positive and negative values (guaranteeing a saddle point), any of the other possible behaviors not ruled out by the existence of positive or negative eigenvalues could occur.
  - As explicit examples, consider the functions $f_1 = x_1^2 + x_2^4$ and $f_2 = x_1^2 - x_2^4$: for both functions H has a single positive eigenvalue and a zero eigenvalue, but at $(0,0)$ f_1 has a local minimum while f_2 has a saddle point. Likewise, for $g_1 = x_1^4 + x_2^4$, $g_2 = x_1^4 - x_2^4$, and $g_3 = -x_1^4 - x_2^4$, we see that for all three functions H has both eigenvalues equal to zero, but at $(0,0)$ g_1 has a local minimum, g_2 has a saddle point, and g_3 has a local maximum.
- Example: Classify the critical point at $(0,0)$ for the function $f(x,y) = 2x^2 + xy + 4y^2$.
  - We compute the Hessian matrix: we have $f_{xx} = 4$, $f_{xy} = f_{yx} = 1$, and $f_{yy} = 8$, so evaluating these at $(0,0)$ yields $H = \begin{bmatrix} 4 & 1 \\ 1 & 8 \end{bmatrix}$.
  - The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^2 - 12t + 31$, whose roots are $\lambda = 6 \pm \sqrt{5}$.
  - Since the eigenvalues are both positive, the critical point is a $\boxed{\text{local minimum}}$.

* Example: Classify the critical point at $(0,0)$ for the function $f(x,y) = x^2 + 3xy - 6y^2 + x^5y^3$.
    * We compute the Hessian matrix: we have $f_{xx} = 2+20x^3y^3$, $f_{xy} = f_{yx} = 3+15x^4y^2$, and $f_{yy} = -6+6x^5y$, so evaluating these at $(0,0)$ yields $H = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.
    * The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^2 - 4t - 21 = (t-7)(t+3)$.
    * Since the eigenvalues are -7 and 3, there is an eigenvalue of each sign, so the critical point is a $\boxed{\text{saddle point}}$.
* Example: Classify the critical point at $(0,0,0)$ for the function $f(x,y,z) = 3x^2 + 2xy - xz + y^2 - yz + z^2$.
    * We compute the Hessian matrix: we have $f_{xx} = 6$, $f_{xy} = f_{yx} = 2$, $f_{xz} = f_{zx} = -1$, $f_{yy} = 2$, $f_{yz} = f_{zy} = -1$, and $f_{zz} = 2$, so $H = \begin{bmatrix} 6 & 2 & -1 \\ 2 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$.
    * The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^3 - 10t^2 + 22t - 12 = (t-2)(t^2 - 8t + 6)$, whose roots are $\lambda = 2, 4 \pm \sqrt{10}$.
    * Since the eigenvalues are all positive, the critical point is a $\boxed{\text{local minimum}}$.

5.2.5 Quadratic Forms Over $\mathbb{R}^n$: Sylvester's Law of Inertia

* We now discuss another fundamental result (which was, in fact, somewhat implicit in our earlier discussion of conics) regarding the possible diagonal entries for the diagonalization of a real quadratic form.
    * By making different choices for the matrix P (e.g., by rescaling it or selecting different row operations), we may obtain different diagonalizations of a given real quadratic form.
    * For example, with the quadratic form $Q(x,y) = x^2 + 2y^2$, which is already diagonal, if we change basis to $x' = x/2$, $y' = y/3$, then we obtain $Q(x,y) = 4(x')^2 + 18(y')^2$.
    * Indeed, it is clear that given any diagonalization, if we scale the ith row of the diagonalizing matrix by the scalar α, then the coefficient of the ith variable will be scaled by α^2.
    * Hence, by rescaling, we may change any positive coefficient to an arbitrary positive value and any negative coefficient to an arbitrary negative value.
    * It turns out that this is essentially the only possible change we may make to the diagonalization over $\mathbb{R}$.
* Theorem (Sylvester's Law of Inertia): Suppose V is a finite-dimensional real vector space and Q is a quadratic form on V. Then the numbers of positive diagonal entries, zero diagonal entries, and negative diagonal entries in any diagonalization of Q is independent of the choice of diagonalization.
    * The idea of this result is that we may decompose V as a direct sum of three spaces: one on which Q acts as a positive-definite quadratic form (corresponding to the positive diagonal entries), one on which Q acts as the zero map (corresponding to the zero entries), and one on which Q acts as a negative-definite quadratic form (corresponding to the negative diagonal entries).
    * Since this decomposition of V depends only on Q, these three spaces (and thus their dimensions) are independent of the choice of diagonalizing basis, and so the number of positive, zero, and negative diagonal entries in any diagonalization is necessarily fixed.
    * Proof: Since we are over a field of characteristic not 2, we may equivalently work with the symmetric bilinear form Φ associated to Q.
    * Let V_0 be the subspace of V given by $V_0 = \{\mathbf{v}_0 \in V : \Phi(\mathbf{v}_0, \mathbf{v}) = 0 \text{ for all } \mathbf{v} \in V\}$: then Φ acts as the zero map on V_0. Now write $V = V_0 \oplus V_1$: we claim that Φ is nondegenerate on V_1.
    * To see this, suppose $\mathbf{y} \in V_1$ has $\Phi(\mathbf{y}, \mathbf{v}_1) = 0$ for all $\mathbf{v}_1 \in V$: then for any $\mathbf{w} \in V$ we may write $\mathbf{w} = \mathbf{v}_0 + \mathbf{v}_1$ for $\mathbf{v}_i \in V_i$, in which case $\Phi(\mathbf{y}, \mathbf{w}) = \Phi(\mathbf{y}, \mathbf{v}_0) + \Phi(\mathbf{y}, \mathbf{v}_1) = 0$. But this would imply $\mathbf{y} \in V_0$ whence $\mathbf{y} = \mathbf{0}$.

- Now we will show that if Φ is nondegenerate on V_1, then V_1 decomposes as a direct sum $V_1 = V_+ \oplus V_-$, where Φ is positive-definite on V_+ and negative-definite on V_-.
- Let V_+ be the maximal subspace of V_1 on which Φ is positive-definite (since the condition is defined only on individual vectors, this subspace is well-defined), and define $V_- = \{\mathbf{w} \in V : \Phi(\mathbf{v}_+, \mathbf{w}) = 0 \text{ for all } \mathbf{v}_+ \in V_+\}$. Then by an application of Gram-Schmidt[4] (via Φ, rather than an inner product), we see that $V_1 = V_+ \oplus V_-$.
- It remains to show that Φ is negative-definite on V_-, so let $\mathbf{z} \in V_-$ be nonzero. Then by assumption, Φ is not positive-definite on $V_+ \oplus \langle \mathbf{z} \rangle$, so there exists some nonzero $\mathbf{v} = \mathbf{v}_+ + \alpha \mathbf{z}$ with $\mathbf{v}_+ \in V_+$ and $\alpha \in \mathbb{R}$ such that $\Phi(\mathbf{v}, \mathbf{v}) \leq 0$.
- We cannot have $\alpha = 0$ since then positive-definiteness would imply $\mathbf{v}_+ = 0$. Since $\Phi(\mathbf{v}, \mathbf{v}) = \Phi(\mathbf{v}_+, \mathbf{v}_+) + 2\alpha\Phi(\mathbf{v}_+, \mathbf{z}) + \alpha^2\Phi(\mathbf{z}, \mathbf{z}) = \Phi(\mathbf{v}_+, \mathbf{v}_+) + \alpha^2\Phi(\mathbf{z}, \mathbf{z})$, we have $\Phi(\mathbf{z}, \mathbf{z}) = \dfrac{1}{\alpha^2}[\Phi(\mathbf{v}, \mathbf{v}) - \Phi(\mathbf{v}_+, \mathbf{v}_+)]$.
- Then both terms are less than or equal to zero, and both cannot be zero. Hence $\Phi(\mathbf{z}, \mathbf{z}) < 0$ for all nonzero $\mathbf{z} \in V_-$ and so Φ is negative-definite on V_-.
- The desired result then follows from the direct sum decomposition $V = V_0 \oplus V_+ \oplus V_-$: if we select any diagonalization, then the restriction to the subspace generated by the basis vectors with diagonal entries 0, positive, negative (respectively) is trivial, positive-definite, negative-definite (respectively), and thus the number of diagonal elements is at least $\dim(V_0)$, $\dim(V_+)$, $\dim(V_-)$ (respectively). But since the total number of diagonal elements is $\dim(V) = \dim(V_0) + \dim(V_+) + \dim(V_-)$, we must have equality everywhere.
- Hence the numbers of positive diagonal entries, zero diagonal entries, and negative diagonal entries in any diagonalization of Q is independent of the choice of diagonalization, as claimed.

- We will also mention that there is some classical terminology associated with this result: the <u>index</u> of a quadratic form is the number of positive diagonal entries (in any diagonalization) and the <u>signature</u> is the difference between the number of positive and negative diagonal entries.
  - Equivalently, by our discussion of the spectral theorem, the index is equal to the number of positive eigenvalues of the matrix associated to the symmetric bilinear form, while the signature is the difference between the number of positive eigenvalues and the number of negative eigenvalues.
  - <u>Remark</u>: Some authors instead refer to the triple $(\dim V_+, \dim V_-, \dim V_0)$, or some appropriate permutation, as the signature of the quadratic form. These three values themselves are called the <u>invariants</u> of the form, and the value of any two of them (along with the dimension of the ambient space V) is sufficient to find the value of the other one.
  - For nondegenerate forms, where there are no 0 entries (so $\dim V_0 = 0$), the dimension of the space along with the value of $\dim V_+ - \dim V_-$ is sufficient to recover the two values.
  - <u>Example</u>: The quadratic form $Q(x, y, z) = x^2 - y^2 - z^2$ on $\mathbb{R}^3$ has index 1 and signature -1.
  - <u>Example</u>: The quadratic form $Q(x, y, z) = x^2 - z^2$ on $\mathbb{R}^3$ has index 1 and signature 0.
  - <u>Example</u>: The quadratic form $Q(x, y, z) = 5x^2 + 4xy + 6y^2 + 4yz + 7z^2$ on $\mathbb{R}^3$ has index 3 and signature 3, since we computed its diagonalization to have diagonal entries $3, 6, 9$.

- <u>Example</u>: Find the index and signature of the quadratic form $Q(x, y, z) = -x^2 - 8xy + 4xz - y^2 + 4yz + 2z^2$.
  - The matrix associated to the corresponding bilinear form is $A = \begin{bmatrix} -1 & -4 & 2 \\ -4 & -1 & 2 \\ 2 & 2 & 2 \end{bmatrix}$.
  - The characteristic polynomial is $p(t) = \det(tI_3 - A) = t^3 - 27t + 54 = (t-3)^2(t+6)$.
  - Thus, since the eigenvalues are $\lambda = 3, 3, -6$, we see that the diagonalization will have two positive diagonal entries and one negative diagonal entry.
  - This means that the index is $\boxed{2}$ and the signature is $\boxed{1}$.

[4] The argument here is the same as for showing that $\dim(W) + \dim(W^\perp) = \dim(V)$ for an inner product. The Gram-Schmidt algorithm does not use the positive-definiteness of the inner product (it requires only linearity and symmetry), so the same argument also works for any bilinear form.

- As a corollary of Sylvester's law of inertia, we can read off the shape of a conic section or quadric surface (in all nondegenerate cases, and also in many degenerate cases) simply by examining the signs of the eigenvalues of the underlying quadratic form.

- Example: Determine the shape of the quadric surface $13x^2 - 4xy + 10y^2 - 8xz + 4yz + 13z^2 = 1$.
  - If $Q(x, y, z)$ is the quadratic form above, the bilinear form has associated matrix $A = \begin{bmatrix} 13 & -2 & -4 \\ -2 & 10 & 2 \\ -4 & 2 & 13 \end{bmatrix}$.
  - The characteristic polynomial is $p(t) = \det(tI_3 - A) = t^3 - 144t^2 + 6480t - 93312 = (t - 36)^2(t - 72)$.
  - This means, upon diagonalizing $Q(x, y, z)$, we will obtain the equation $36(x')^2 + 36(y')^2 + 72(z')^2 = 1$. This is the equation of an $\boxed{\text{ellipsoid}}$.
  - Note that the only information we needed here was the fact that all three eigenvalues were positive to make this observation: the quadric surfaces $Q(x, y, z) = 1$ that are ellipsoids are precisely those for which $Q(x, y, z)$ is a positive-definite quadratic form.

- We will close our discussion by observing that the study of quadratic forms touches on nearly every branch of mathematics: we have already examined some of its ties to linear algebra (in the guise of bilinear forms and diagonalization), analysis (in the classification of critical points), and geometry (in the analysis of quadratic varieties and the action of matrices on quadratic forms).
  - We will not discuss it much here, since the requisite tools do not really belong to linear algebra, but the study of quadratic forms over $\mathbb{Q}$ turns out to be intimately tied with many topics in number theory.
  - A very classical problem in elementary number theory is to characterize, in as much detail as possible, the integers represented by a particular quadratic form. For example: which integers are represented by the quadratic form $Q(x, y) = x^2 + y^2$ (i.e., which integers can be written as the sum of two squares)?
  - This family of problems, while seemingly quite simple, is actually intimately related to a number of very deep results in modern number theory, and (historically speaking) was a major motivating force in the development of a branch of algebraic number theory known as class field theory.

5.3 Singular Values and Singular Value Decomposition

- Diagonalization is a very useful tool, but it suffers from two main drawbacks: first, not all linear transformations $T : V \to V$ are diagonalizable, and second, we cannot diagonalize general linear transformations $T : V \to W$ when V and W are different.
  - The Jordan canonical form allows us to give a "near diagonalization" for non-diagonalizable linear transformations $T : V \to V$. However, it still cannot be used to describe general transformations $T : V \to W$.

- We will now discuss a decomposition that is in some sense a modified diagonalization, called singular value decomposition, that extends naturally to transformations $T : V \to W$ when V and W are different.

5.3.1 Singular Values and Singular Value Decomposition

- The main idea of singular value decomposition is as follows: if we have a linear transformation $T : V \to W$ where V and W are finite-dimensional inner product spaces, then we may construct orthonormal bases β of V and γ of W such that the associated matrix $A = [T]_\beta^\gamma$ has its only nonzero entries in the "diagonal" positions $a_{i,i}$.
  - This procedure combines the ideas of diagonalization, in that we obtain a representation of T by an essentially diagonal matrix (up to not being square), and the QR factorization, in that we perform orthonormal changes of basis to simplify the form of the transformation.
  - The main idea is to note that if $T : V \to W$ is linear, then $T^*T : V \to V$ is Hermitian, since $(T^*T)^* = T^*T^{**} = T^*T$ again.

- Moreover, the quadratic form $Q_{T^*T}(\mathbf{v}) = \langle T^*T\mathbf{v}, \mathbf{v}\rangle = \langle T\mathbf{v}, T\mathbf{v}\rangle = ||T\mathbf{v}||^2$ on V is positive semidefinite since $||T\mathbf{v}||^2 \geq 0$ for all $\mathbf{v}$. Thus, from our characterization of the definiteness of quadratic forms, we see that all of the eigenvalues of T^*T are nonnegative.
- Therefore, since T^*T is a Hermitian operator with nonnegative eigenvalues, it can be orthogonally diagonalized with respect to an orthonormal basis β, and the diagonal entries of its diagonalization D are $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ for some nonnegative real numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.
- We can then use the orthonormal basis β of V to write down an appropriate orthonormal basis γ for W so that $[T]_\beta^\gamma$ will have its only nonzero entries in the positions $a_{i,i}$.

- To make this precise, we introduce some terminology:
- Definition: If V and W are finite-dimensional inner product spaces and $T : V \to W$ is linear, the singular values of T are the nonnegative real numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ such that $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ are the eigenvalues of T^*T.
  - We have previously shown that $\ker(T^*T) = \ker(T)$, so the rank of T^*T equals the rank of T. Thus, if T has rank r, the singular values $\sigma_1, \dots, \sigma_r$ will be positive and the remaining ones $\sigma_{r+1}, \dots, \sigma_n$ will be zero.
  - It is a matter of convention whether to include the zero eigenvalues on the list of singular values.
  - For a matrix A, the singular values are the eigenvalues of A^*A. If β is an orthonormal basis of V and γ is an orthonormal basis of W, then the singular values of T are the singular values of $[T]_\beta^\gamma$. (Note that if we make an orthonormal change of basis in V or W, the eigenvalues of A^*A are not changed.)

- Example: Find the singular values of the matrix $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$.
  - We have $A^*A = \begin{bmatrix} 2 & 2 & -1 & 1 \\ 2 & 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$ with characteristic polynomial $p(t) = \det(tI_3 - A) = (t-10)(t-10) - (6)(6) = t^2 - 20t + 64 = (t-4)(t-16)$.
  - Since the eigenvalues of A^*A are $\lambda = 16, 4$, we see that the singular values of A are $\boxed{4, 2}$.

- Example: Find the singular values of the matrix $A = \begin{bmatrix} 1-i & 3i \\ 2+i & i \\ 2+i & 1-2i \end{bmatrix}$.
  - We have $A^*A = \begin{bmatrix} 1+i & 2-i & 2-i \\ -3i & -i & 1+2i \end{bmatrix} \begin{bmatrix} 1-i & 3i \\ 2+i & i \\ 2+i & 1-2i \end{bmatrix} = \begin{bmatrix} 12 & -2 \\ -2 & 15 \end{bmatrix}$ with characteristic polynomial $p(t) = \det(tI_2 - A) = (t-12)(t-15) - (-2)^2 = (t-11)(t-16)$.
  - Since the eigenvalues of A^*A are $\lambda = 16, 11$, we see that the singular values of A are $\boxed{4, \sqrt{11}}$.

- Example: Find the singular values of the matrix $A = \begin{bmatrix} 1+i & 3 & 1 & i \\ 1+i & 1 & 2 & -i \end{bmatrix}$.
  - We have $A^*A = \begin{bmatrix} 1+i & 1+i \\ 3 & 1 \\ 1 & 2 \\ i & -i \end{bmatrix} \begin{bmatrix} 1+i & 3 & 1 & i \\ 1+i & 1 & 2 & -i \end{bmatrix} = \begin{bmatrix} 4 & 4-4i & 3-3i & 0 \\ 4+4i & 10 & 5 & 2i \\ 3+3i & 5 & 5 & -i \\ 0 & -2i & i & 2 \end{bmatrix}$ with characteristic polynomial $p(t) = \det(tI_4 - A) = t^4 - 21t^3 + 68t^2 = t^2(t-4)(t-17)$.
  - Since the eigenvalues of A^*A are $\lambda = 17, 4, 0, 0$, we see that the singular values of A are $\boxed{\sqrt{17}, 2, 0, 0}$.

- Our general result is that we can use the singular values of a matrix to write down a matrix associated to T with respect to orthonormal bases that is particularly nice:

- Theorem (Singular Value Bases): Suppose V and W are finite-dimensional inner product spaces and $T : V \to W$ is a linear transformation of rank r with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then there exist orthonormal bases $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of W such that $T(\mathbf{v}_1) = \sigma_1 \mathbf{w}_1$, $T(\mathbf{v}_2) = \sigma_2 \mathbf{w}_2$, ... , $T(\mathbf{v}_r) = \sigma_r \mathbf{w}_r$, and $T(\mathbf{v}_{r+1}) = T(\mathbf{v}_{r+2}) = \cdots = T(\mathbf{v}_n) = \mathbf{0}$.

  - Proof: Recall that we proved in our discussion of least squares that T^*T and T have the same rank, so T^*T also has rank r.
  - Additionally, as noted above, T^*T is a Hermitian operator on V, so by the spectral theorem, there exists an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V consisting of eigenvectors of T^*T, where the basis is ordered the associated eigenvalues are $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0$.
  - Then for $i \neq j$, we have $\langle T(\mathbf{v}_i), T(\mathbf{v}_j) \rangle = \langle \mathbf{v}_i, T^*T\mathbf{v}_j \rangle = \sigma_i^2 \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ since $\mathbf{v}_i, \mathbf{v}_j$ are orthogonal, so $T(\mathbf{v}_i)$ and $T(\mathbf{v}_j)$ are also orthogonal.
  - Additionally, for $1 \leq i \leq r$ we have $||T(\mathbf{v}_i)||^2 = \langle T(\mathbf{v}_i), T(\mathbf{v}_i) \rangle = \langle \mathbf{v}_i, T^*T\mathbf{v}_i \rangle = \sigma_i^2 \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \sigma_i^2$, so since $\sigma_i > 0$ we see $||T(\mathbf{v}_i)|| = \sigma_i$.
  - Therefore, if we define $\mathbf{w}_i = T(\mathbf{v}_i)/\sigma_i$ for each $1 \leq i \leq r$, we see that $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ is an orthonormal set. By extending this set to an orthonormal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of W, we obtain the required result.

- To calculate a singular value decomposition, we simply need to compute an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for T^*T, and then compute their images under T and divide by the singular values (potentially then extending to a basis of W using Gram-Schmidt) to get the basis $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$.

  - To compute T^*, if we set $A = [T]_\alpha^\delta$ where α and δ are arbitrary orthonormal bases for V and W (respectively) then $[T^*]_\delta^\alpha$ is simply the adjoint matrix A^*.

- Example: Find singular value bases for the map $T : P_1(\mathbb{R}) \to P_2(\mathbb{R})$ with $T(p) = p(0)$, under the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$ on both $P_1(\mathbb{R})$ and $P_2(\mathbb{R})$.

  - First, we must compute T^*. Using Gram-Schmidt we construct orthonormal bases $\alpha = \{1, \sqrt{3}(2x-1)\}$ for $P_1(\mathbb{R})$ and $\delta = \{1, \sqrt{3}(2x-1), \sqrt{5}(6x^2-6x+1)\}$ for $P_2(\mathbb{R})$.
  - Then since $T(1) = 1$ and $T(\sqrt{3}(2x-1)) = -\sqrt{3}$, we obtain $[T]_\alpha^\delta = \begin{bmatrix} 1 & -\sqrt{3} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, hence because α and δ are orthonormal bases, we have $[T^*]_\delta^\alpha = \begin{bmatrix} 1 & 0 & 0 \\ -\sqrt{3} & 0 & 0 \end{bmatrix}$.
  - To find the desired singular value bases, we compute an eigenbasis for T^*T, whose associated matrix is $[T^*T]_\alpha^\alpha = [T^*]_\delta^\alpha [T]_\alpha^\delta = \begin{bmatrix} 1 & 0 & 0 \\ -\sqrt{3} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\sqrt{3} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & 3 \end{bmatrix}$.
  - The eigenvalues are $\lambda = 4, 0$ with corresponding orthonormal eigenvectors $[\mathbf{v}_1]_\alpha = \frac{1}{2} \begin{bmatrix} 1 \\ -\sqrt{3} \end{bmatrix}$ and $[\mathbf{v}_2]_\alpha = \frac{1}{2} \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$, whence $\mathbf{v}_1 = \frac{1}{2}\left[-1 + \sqrt{3} \cdot \sqrt{3}(2x-1)\right] = 3x - 2$ and $\mathbf{v}_2 = \frac{1}{2}\left[\sqrt{3} + 1 \cdot \sqrt{3}(2x-1)\right] = x\sqrt{3}$.
  - The nonzero singular value is $\sigma_1 = \sqrt{4} = 2$, so that $\mathbf{w}_1 = T(\mathbf{v}_1)/\sigma_1 = -1$, and then using Gram-Schmidt we can extend this to an orthonormal basis of W (indeed, we can just reuse the other vectors from δ above).
  - Thus we obtain bases $\{\mathbf{v}_1, \mathbf{v}_2\} = \boxed{\{3x-2, x\sqrt{3}\}}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} = \boxed{\{-1, \sqrt{3}(2x-1), \sqrt{5}(6x^2-6x+1)\}}$.

- We can recast the theorem above in terms of matrices, as follows:

- Theorem (Singular Value Decomposition): Suppose $F = \mathbb{R}$ or $\mathbb{C}$ and that $A \in M_{m\times n}(F)$ has rank r. If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values of A, then A can be written as a matrix product $A = U\Sigma V^*$ where $U \in M_{n\times n}(F)$ and $V \in M_{m\times m}(F)$ are unitary and $\Sigma \in M_{n\times m}(F)$ is the matrix whose first r diagonal entries are $\sigma_1, \dots, \sigma_r$ and whose remaining entries are 0.
  - Proof: Let $T : V \to W$ be the linear transformation with $T(\mathbf{v}) = A\mathbf{v}$.
  - By the theorem above, we have orthonormal bases $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of F^n and $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of F^m such that $T(\mathbf{v}_1) = \sigma_1\mathbf{w}_1$, $T(\mathbf{v}_2) = \sigma_2\mathbf{w}_2$, ... , $T(\mathbf{v}_r) = \sigma_r\mathbf{w}_r$, and $T(\mathbf{v}_{r+1}) = T(\mathbf{v}_{r+2}) = \cdots = T(\mathbf{v}_n) = \mathbf{0}$.
  - This means the associated matrix $[T]_\beta^\gamma$ is the $n \times m$ matrix Σ whose first r diagonal entries are $\sigma_1, \dots, \sigma_r$ and whose remaining entries are 0.
  - Now let α be the standard basis of F^n and δ be the standard basis of F^m and note that $[T]_\alpha^\delta = A$. Furthermore, since β is orthonormal the change-of-basis matrix $V = [I]_\beta^\alpha$ is unitary so $[I]_\alpha^\beta = V^{-1} = V^*$, and since γ is orthonormal the change-of-basis matrix $U = [I]_\gamma^\delta$ is also unitary.
  - Then $A = [T]_\alpha^\delta = [I]_\gamma^\delta [T]_\beta^\gamma [I]_\alpha^\beta = U\Sigma V^*$, as claimed.

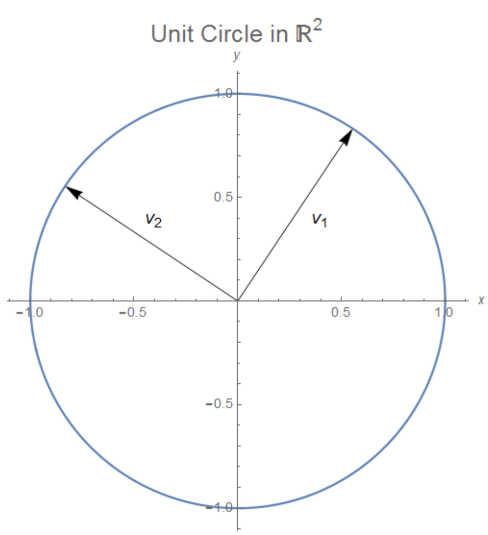
- Example: Find the singular values, and a singular value decomposition, of $A = \begin{bmatrix} 0 & 6 \\ 6 & 5 \end{bmatrix}$.
  - We have $A^*A = \begin{bmatrix} 36 & 30 \\ 30 & 61 \end{bmatrix}$ with characteristic polynomial $p(t) = (t-36)(t-61) - (30)(30) = t^2 - 97t + 1296 = (t-16)(t-81)$, so the singular values of A are $\sigma_1 = \sqrt{81} = 9$ and $\sigma_2 = \sqrt{16} = 4$.
  - We can then find a basis for the 81-eigenspace of A^*A as $\left\{\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right\}$ and a basis for the 16-eigenspace as $\left\{\begin{bmatrix} -3 \\ 2 \end{bmatrix}\right\}$, so after normalizing we can take $\mathbf{v}_1 = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{v}_2 = \frac{1}{\sqrt{13}}\begin{bmatrix} -3 \\ 2 \end{bmatrix}$.
  - We also have $\mathbf{w}_1 = \frac{1}{9}A\mathbf{v}_1 = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{w}_2 = \frac{1}{4}A\mathbf{v}_2 = \frac{1}{\sqrt{13}}\begin{bmatrix} 3 \\ -2 \end{bmatrix}$; as expected we see that $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthonormal set (and in fact an orthonormal basis) of $\mathbb{R}^2$.
  - Putting all of this together, we get $U = \boxed{\frac{1}{\sqrt{13}}\begin{bmatrix} 2 & 3 \\ 3 & -2 \end{bmatrix}}$, $\Sigma = \boxed{\begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}}$, and $V^* = \boxed{\frac{1}{\sqrt{13}}\begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}}$.

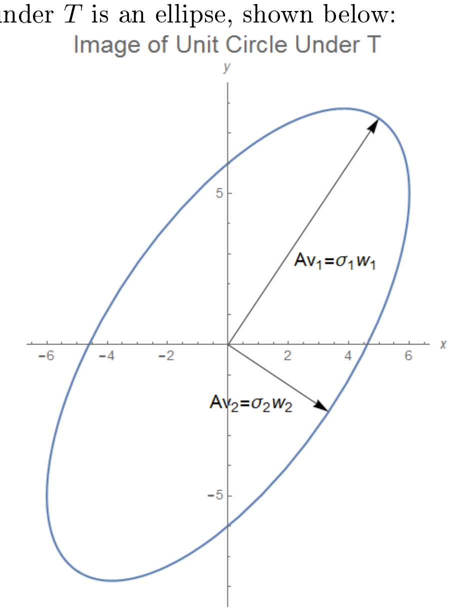
- Example: Find a singular value decomposition of $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$.
  - We previously found that the eigenvalues of $A^*A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$ are $\lambda = 16, 4$ and so the singular values of A are $\sigma_1 = 4$ and $\sigma_2 = 2$.
  - We can then calculate a basis for the 16-eigenspace of A^*A as $\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ and a basis for the 4-eigenspace of A^*A as $\left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$, so after normalizing we can take $\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.
  - We also have $\mathbf{w}_1 = \frac{1}{4}A\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{w}_2 = \frac{1}{2}A\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$; as expected we see that $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthonormal set.
  - By using Gram-Schmidt, we can extend $\{\mathbf{w}_1, \mathbf{w}_2\}$ to an orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}$ of $\mathbb{R}^4$ with $\mathbf{w}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{w}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

- Thus, we get $U = \boxed{\dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{bmatrix}}$, $\Sigma = \boxed{\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}$, and $V^* = \boxed{\dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}$.
- Remark: Note that the singular value decomposition is not unique here, since we could choose other vectors $\mathbf{w}_3, \mathbf{w}_4$ to complete the orthonormal basis of $\mathbb{R}^4$.

- Example: Find a singular value decomposition of the matrix $A = \begin{bmatrix} 1+i & 3 & 1 & i \\ 1+i & 1 & 2 & -i \end{bmatrix}$.
  - We previously calculated the eigenvalues of A^*A as 17, 4, 0, and 0 so that the positive singular values are $\sigma_1 = \sqrt{17}$ and $\sigma_2 = 2$.
  - We can then calculate an orthonormal basis of eigenvectors with $\lambda = 17, 4, 0, 0$ respectively as $\mathbf{v}_1 = \dfrac{1}{\sqrt{221}}\begin{bmatrix} 5+5i \\ 11i \\ 7i \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \dfrac{1}{\sqrt{52}}\begin{bmatrix} -1-i \\ 3i \\ -4i \\ 5 \end{bmatrix}$, $\mathbf{v}_3 = \dfrac{1}{2}\begin{bmatrix} 1+i \\ -i \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_4 = \dfrac{1}{\sqrt{34}}\begin{bmatrix} -2+2i \\ -1 \\ 4 \\ -3i \end{bmatrix}$.
  - We also have $\mathbf{w}_1 = \dfrac{1}{\sqrt{17}}A\mathbf{v}_1 = \dfrac{1}{\sqrt{13}}\begin{bmatrix} 3i \\ 2i \end{bmatrix}$ and $\mathbf{w}_2 = \dfrac{1}{2}A\mathbf{v}_2 = \dfrac{1}{\sqrt{13}}\begin{bmatrix} 2i \\ -3i \end{bmatrix}$; as expected we see that $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthonormal set.
  - Putting all of this together, we get $U = \boxed{\dfrac{1}{\sqrt{13}}\begin{bmatrix} 3i & 2i \\ 2i & -3i \end{bmatrix}}$, $\Sigma = \boxed{\begin{bmatrix} \sqrt{17} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}}$, and $V^* = \boxed{\begin{bmatrix} (5+5i)/\sqrt{221} & (-1-i)/\sqrt{52} & (1+i)/2 & (-2+2i)/\sqrt{34} \\ 11i/\sqrt{221} & 3i/\sqrt{52} & -i/2 & -1/\sqrt{34} \\ 7i/\sqrt{221} & -4i/\sqrt{52} & 0 & 4/\sqrt{34} \\ 1/\sqrt{221} & 5/\sqrt{52} & 1/2 & -3i/\sqrt{34} \end{bmatrix}}$.
  - Remark: Note that the singular value decomposition is not unique here, since for example we could have chosen any other orthonormal basis $\mathbf{v}_3, \mathbf{v}_4$ of the 0-eigenspace of A^*A.

- The singular value basis and associated decomposition have a convenient geometric interpretation in terms of the action of the transformation $T : V \to W$ on the "unit sphere" $||\mathbf{v}|| = 1$ in V.
  - To illustrate, consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with associated standard matrix $A = \begin{bmatrix} 0 & 6 \\ 6 & 5 \end{bmatrix}$ from the first example above.
  - The image of the unit circle $||\mathbf{v}|| = 1$ (i.e., $x^2 + y^2 = 1$) under T is an ellipse, shown below:

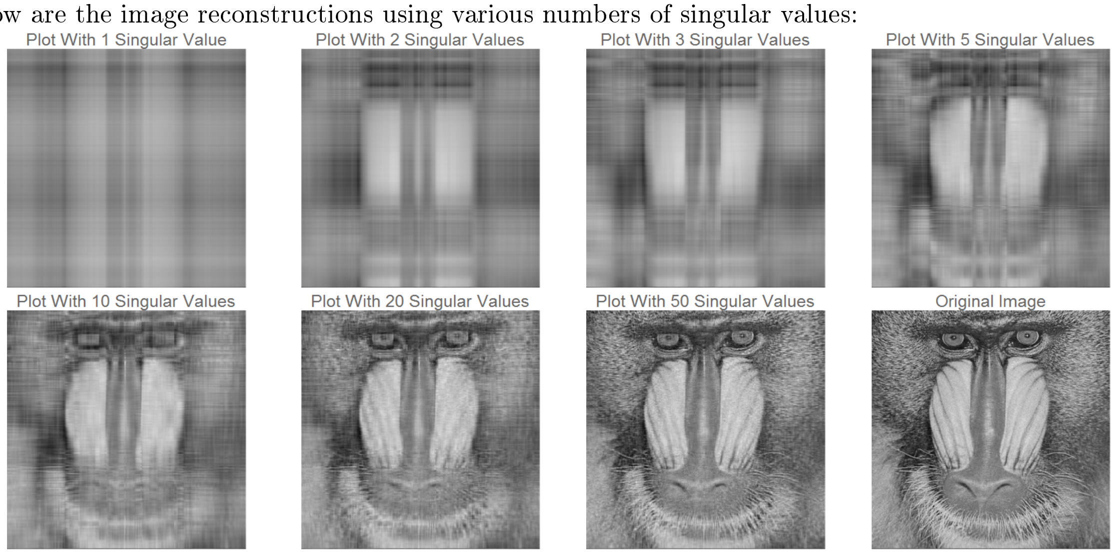




* We can see quite clearly from the second picture that the vectors $A\mathbf{v}_1$ and $A\mathbf{v}_2$ give the principal axes of the ellipse.
* This observation can be verified algebraically from the facts that $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\gamma = \{\mathbf{w}_1, \mathbf{w}_2\}$ are orthonormal bases of $\mathbb{R}^2$ and the fact that $[T]_\beta^\gamma$ is the matrix $\begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$: then the image $9a\mathbf{w}_1 + 4b\mathbf{w}_2$ of any linear combination $a\mathbf{v}_1 + b\mathbf{v}_2$ on the unit circle (i.e., with $a^2 + b^2 = 1$) has norm $81a^2 + 16b^2$, and the norm is clearly maximized when $b = 0$ and minimized when $a = 0$.
* This means that the major axis of the ellipse is parallel to $\mathbf{w}_1$ and has length σ_1, while the minor axis of the ellipse is parallel to $\mathbf{w}_2$ and has length σ_2.
* It is not hard to see that analogous results hold in higher dimensions, for the same reasons: in general, the image of the unit sphere $||\mathbf{v}|| = 1$ under a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ of rank r will be an r-dimensional ellipsoid whose principal axes are the vectors $\sigma_1\mathbf{w}_1$, $\sigma_2\mathbf{w}_2$, ... , $\sigma_r\mathbf{w}_r$.

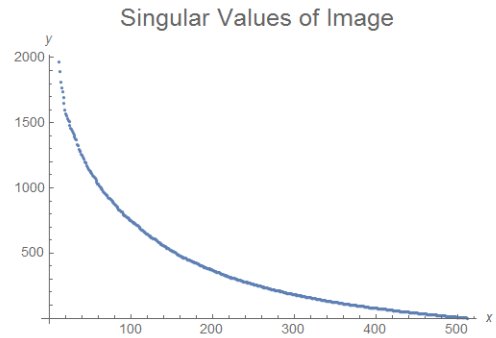
- This geometric interpretation of singular value decomposition has many practical applications, such as performing principal component analysis and doing data compression.
  * The main idea is that for an $m \times n$ matrix A with singular values $\sigma_1, \dots, \sigma_r$ and corresponding orthonormal bases $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$, if we multiply out the matrix product $A = U\Sigma V^*$, we can rephrase the singular value decomposition as giving a sum $A = \sigma_1\mathbf{v}_1\mathbf{w}_1^* + \sigma_2\mathbf{v}_2\mathbf{w}_2^* + \cdots + \sigma_r\mathbf{v}_r\mathbf{w}_r^*$ of a total of r $m \times n$ matrices $\mathbf{v}_i\mathbf{w}_i^*$ each of which has rank 1.
  * Therefore, if we want to approximate A by a matrix of rank less than r, the best approximation will be given by deleting the terms of the sum above that have the smallest norm, which are the terms with smallest σ_i.
  * In other words[5], the best approximation to A by a matrix of rank d is obtained by taking the initial terms of the singular value sum above: $\sigma_1\mathbf{v}_1\mathbf{w}_1^* + \sigma_2\mathbf{v}_2\mathbf{w}_2^* + \cdots + \sigma_d\mathbf{v}_d\mathbf{w}_d^*$.
  * In the situation where we have a set of data that is high-dimensional (i.e., lies inside F^n where n is large), this gives an explicit procedure for projecting onto a smaller-dimensional subspace that loses as little information as possible.

- We can illustrate these ideas by calculating the singular value decomposition of a matrix representing the 2-dimensional grid of color intensity from a black-and-white photograph (taken from the standard set of sample data included with Mathematica).
  * The photograph used here is 512 pixels by 512 pixels, corresponding to a 512×512 matrix A.
  * We can then give compressed versions of the image by taking the initial terms of the singular value sum.
  * Below are the image reconstructions using various numbers of singular values:



[5]The formal statement is the Eckart-Young-Mirsky theorem, as follows: let $A \in M_{m\times n}(F)$ for $F = \mathbb{R}$ or $\mathbb{C}$ have rank r and singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, with corresponding singular value bases $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of F^n and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of F^m. Then for any $1 \leq d \leq r$, the best rank-d approximation to A is the sum $A_d = \sigma_1\mathbf{v}_1\mathbf{w}_1^* + \sigma_2\mathbf{v}_2\mathbf{w}_2^* + \cdots + \sigma_d\mathbf{v}_d\mathbf{w}_d^*$, in the sense that if $B : V \to W$ is any rank-d transformation, then $||A_d - A|| \leq ||B - A||$ with equality only for $B = A_d$, where the norm is the Frobenius norm on $M_{m\times n}(F)$.

- The total amount of data required to store the full image is the equivalent to 512^2 data points (one per pixel). To store the decomposition with k singular values, on the other hand, requires only storing about $2k \cdot 512$ data points (each singular value matrix $\mathbf{v}_i \mathbf{w}_i^*$ requires just the values of the vectors $\mathbf{v}_i$ and $\mathbf{w}_i$).
- So, for example, to store and reconstruct the compressed image using 20 singular values only requires about $40/512 \approx 8\%$ of the total amount of uncompressed data in the original image.
- The reason this sort of procedure works is because most of the information in the image is carried by the first few singular values, which are much larger than the later ones. For this image, the first ten singular values are 66679, 10490, 5904, 4144, 3501, 2853, 2664, 2420, 2384, and 2188, with most of the remaining values being smaller:



- Therefore, taking just the first few singular values will capture the vast majority of information contained in the data set.

5.3.2 The Moore-Penrose Pseudoinverse

- To conclude our discussion, we will study the Moore-Penrose pseudoinverse of a linear transformation $T : V \to W$, which has pleasant applications to computing orthogonal projections and to solving systems of linear equations.
  - The main idea is to construct a linear transformation $T^\dagger : W \to V$ that captures as much of the spirit of an inverse transformation as possible, even when T is not necessarily invertible or even one-to-one.
  - A natural approach is simply to restrict attention to the "piece" of T that is invertible: clearly, T is not invertible on $\ker(T)$ (since it is the zero map there), so there is nothing to be gained by considering T on $\ker(T)$.
  - However, if we take any subspace V_1 such that $V = \ker(T) \oplus V_1$ (we may find such a subspace simply by extending a basis of $\ker(T)$ to a basis of V), then $\ker(T|_{V_1}) = \ker(T) \cap V_1 = \{\mathbf{0}\}$, so T is one-to-one when restricted to the subspace V_1.
  - It is easy to see that $\mathrm{im}(T|_{V_1}) = \mathrm{im}(T)$, so since $T|_{V_1} : V_1 \to \mathrm{im}(T)$ is one-to-one, it has an inverse map $S : \mathrm{im}(T) \to V$. We can then extend this inverse map to be defined on all of W (rather than just $\mathrm{im}(T)$) by selecting a subspace W_1 of W with $W = \mathrm{im}(T) \oplus W_1$ (we may find such a subspace simply by extending a basis of $\mathrm{im}(T)$ to a basis of W) and declaring that S maps W_1 to zero, so that $\ker(S) = W_1$.
  - The resulting map $S : W \to V$ then behaves like a "partial inverse" of $T : V \to W$, in the sense that for any $\mathbf{w} \in \mathrm{im}(T)$ and any $\mathbf{v} \in V_1$, we have $S(T(\mathbf{v})) = \mathbf{v}$ and $T(S(\mathbf{w})) = \mathbf{w}$, and also for any $\tilde{\mathbf{v}} \in \ker(T)$ and $\tilde{\mathbf{w}} \in W_1$ we have $T(S(T(\mathbf{v})) = \mathbf{0} = T(\mathbf{v})$ and $S(T(S(\mathbf{w})) = \mathbf{0} = S(\mathbf{w})$.
  - In particular, although S and T are not full inverses of one another, the calculations above show that they do satisfy the relations $STS = S$ and $TST = T$.
  - Of course, we have made two arbitrary choices above; namely, we have chosen the complement subspace V_1 to $\ker(T)$ in V arbitrarily, and we have also chosen the complement subspace W_1 to $\mathrm{im}(T)$ in W arbitrarily.
  - When V and W are finite-dimensional inner product spaces, there are natural choices for these complements: specifically, we could take V_1 to be the orthogonal complement of $\ker(T)$ in V and W_1 to be the orthogonal complement of $\mathrm{im}(T)$ in W.
  - The resulting construction yields the Moore-Penrose pseudoinverse of T:

- Definition: Suppose that V and W are finite-dimensional inner product spaces and $T : V \to W$ is linear, and let $S : \ker(T)^\perp \to \mathrm{im}(T)$ be the restriction of T to $\ker(T)^\perp$. The (Moore-Penrose) pseudoinverse of T is the linear transformation $T^\dagger : W \to V$ defined so that $T^\dagger(\mathbf{w}) = S^{-1}(\mathbf{w})$ for all $\mathbf{w} \in \mathrm{im}(T)$ and $T^\dagger(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w} \in \mathrm{im}(T)^\perp$.
    - We extend the definition of the pseudoinverse to matrices in the natural way: if $A \in M_{m\times n}(F)$ where $F = \mathbb{R}$ or $\mathbb{C}$, then the pseudoinverse of A is the associated matrix (with respect to the standard bases) of the pseudoinverse $T^\dagger$ of the linear transformation with $T(\mathbf{v}) = A\mathbf{v}$.
    - By the discussion above, the pseudoinverse of T is well defined, since S is one-to-one (hence has a valid inverse) and the union of bases of $\mathrm{im}(T)$ and $\mathrm{im}(T)^\perp$ yields a basis for W, so the two conditions on $T^\dagger$ characterize the value of $T^\dagger$ on all of W.
    - If T is an isomorphism, then the pseudoinverse of T is merely T^{-1}, since in that case $\ker(T)^\perp = V$ so that $S = T$, and $\mathrm{im}(T) = W$, so that $T^\dagger(\mathbf{w}) = S^{-1}(\mathbf{w}) = T^{-1}(\mathbf{w})$ for all $\mathbf{w} \in W$.
    - If T is merely one-to-one, then $T^\dagger$ is obtained by extending the domain of $T^{-1} : \mathrm{im}(T) \to V$ to all of W, where the "extra vectors" in W (namely, $\mathrm{im}(T)^\perp$) are sent to zero.
- In principle, we can use the definition of the pseudoinverse to compute it directly, although this requires computing bases for $\ker(T)^\perp$, $\mathrm{im}(T)$, and $\mathrm{im}(T)^\perp$ and then evaluating T^{-1} on $\mathrm{im}(T)$:
- Example: Find the pseudoinverse of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(x, y, z) = (x, 0)$, where the inner products on $\mathbb{R}^3$ and $\mathbb{R}^2$ are the dot product.
    - We have the obvious basis $\{(0, 1, 0), (0, 0, 1)\}$ for $\ker(T)$, and $\ker(T)^\perp$ is easily seen to have basis $\{(1, 0, 0)\}$.
    - Likewise, we have the obvious basis $\{(1, 0)\}$ for $\mathrm{im}(T)$, and $\mathrm{im}(T)^\perp$ is easily seen to have basis $\{(0, 1)\}$.
    - Therefore, since $S(1, 0, 0) = (1, 0)$, we see that $T^\dagger(1, 0) = S^{-1}(1, 0) = (1, 0, 0)$, and also $T^\dagger(0, 1) = (0, 0, 0)$.
    - Thus, the map $T^\dagger : \mathbb{R}^2 \to \mathbb{R}^3$ has $\boxed{T^\dagger(x, y) = (x, 0, 0)}$.
- Example: Find the pseudoinverse of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(x, y, z) = (x + y - z, x + z)$, where the inner products on $\mathbb{R}^3$ and $\mathbb{R}^2$ are the dot product.
    - We can compute the basis $\{(-1, 2, 1)\}$ for $\ker(T)$, and $\ker(T)^\perp$ is easily seen to have basis $\{(1, 1, -1), (1, 0, 1)\}$.
    - Likewise, we have the obvious basis $\{(1, 0), (0, 1)\}$ for $\mathrm{im}(T)$, while $\mathrm{im}(T)^\perp$ has empty basis.
    - Therefore, with $S = T|_{\ker(T)^\perp}$, we compute $S(1, 1, -1) = (3, 0)$ and $S(1, 0, 1) = (0, 2)$, so $S(1/3, 1/3, -1/3) = (1, 0)$ and $S(1/2, 0, 1/2) = (0, 1)$.
    - Thus, we have $T^\dagger(1, 0) = S^{-1}(1, 0) = (1/3, 1/3, -1/3)$, and also $T^\dagger(0, 1) = (1/2, 0, 1/2)$.
    - This means the map $T^\dagger : \mathbb{R}^2 \to \mathbb{R}^3$ has $\boxed{T^\dagger(x, y) = (\frac{1}{3}x + \frac{1}{2}y, \frac{1}{3}x, -\frac{1}{3}x + \frac{1}{2}y)}$.
- Using singular value decomposition, we can give another procedure for computing pseudoinverses.
    - Explicitly, for $T : V \to W$ suppose we have computed the singular value bases $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of W such that $T(\mathbf{v}_i) = \sigma_i \mathbf{w}_i$ for $1 \le i \le r$ and $T(\mathbf{v}_i) = \mathbf{0}$ for $i > r$.
    - Then $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is a basis for $\ker(T)$ so $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is a basis for $\ker(T)^\perp$, and $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ is a basis for $\mathrm{im}(T)$ so $\{\mathbf{w}_{r+1}, \dots, \mathbf{w}_m\}$ is a basis for $\mathrm{im}(T)^\perp$.
    - We then see $S(\mathbf{v}_i) = \sigma_i \mathbf{w}_i$ and so $T^\dagger(\mathbf{w}_i) = S^{-1}(\mathbf{w}_i) = \mathbf{v}_i/\sigma_i$ for each $1 \le i \le r$, and also $T^\dagger(\mathbf{w}_i) = \mathbf{0}$ for $i > r$, which gives the values of $T^\dagger$ on a basis for W.
- The above calculations in fact yield a singular value decomposition for the pseudoinverse of a matrix:
- Proposition (Pseudoinverses and SVDs): Suppose $F = \mathbb{R}$ or $\mathbb{C}$ and that $A \in M_{m\times n}(F)$ has rank r and a singular value decomposition $A = U\Sigma V^*$ where $U \in M_{n\times n}(F)$ and $V \in M_{m\times m}(F)$ are unitary and $\Sigma \in M_{n\times m}(F)$ is the matrix whose first r diagonal entries are $\sigma_1, \dots, \sigma_r$ and whose remaining entries are 0. If $\Sigma^\dagger \in M_{m\times n}(F)$ is the matrix whose first r diagonal entries are $\sigma_1^{-1}, \dots, \sigma_r^{-1}$ and whose remaining entries are 0, then the pseudoinverse $A^\dagger$ has a singular value decomposition $A^\dagger = V\Sigma^\dagger U^*$.

- We will remark that the nonzero diagonal entries of $\Sigma^\dagger$ are in increasing order rather than decreasing order; one may put the singular values in the usual decreasing order simply by reversing the order of the corresponding basis elements.
- Proof: Let $T : F^n \to F^m$ have $T(\mathbf{v}) = A\mathbf{v}$, and suppose we have singular value bases $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of $V = F^n$ and $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ of $W = F^m$ such that $T(\mathbf{v}_i) = \sigma_i \mathbf{w}_i$ for $1 \le i \le r$ and $T(\mathbf{v}_i) = \mathbf{0}$ for $i > r$.
- Then as calculated above, we have $T^\dagger(\mathbf{w}_i) = S^{-1}(\mathbf{w}_i) = \mathbf{v}_i/\sigma_i$ for each $1 \le i \le r$, and also $T^\dagger(\mathbf{w}_i) = \mathbf{0}$ for $i > r$.
- Thus, the associated matrix $[T^\dagger]_\gamma^\beta$ is the $m \times n$ rectangular diagonal matrix $\Sigma^\dagger$.
- Now, if we let α be the standard basis of F^n and δ be the standard basis of F^m, then $[T]_\alpha^\delta = A$, $V = [I]_\beta^\alpha$, and $U = [I]_\gamma^\delta$ so that $[I]_\delta^\gamma = U^{-1} = U^*$.
- Then the pseudoinverse $A^\dagger = [T^\dagger]_\delta^\alpha = [I]_\beta^\alpha [T^\dagger]_\gamma^\beta [I]_\delta^\gamma = V\Sigma^\dagger U^*$, as claimed, and V and U are unitary and $\Sigma^\dagger$ is a rectangular diagonal matrix, this is in fact a singular value decomposition of $A^\dagger$.

- Example: Find the pseudoinverse of the matrix $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$.
  - We previously calculated a singular value decomposition $A = U\Sigma V^*$ with

    $$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{bmatrix},\ \Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ \text{and } V^* = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

  - Then $A^\dagger = V\Sigma^\dagger U^* = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{bmatrix} \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \boxed{\dfrac{1}{8} \begin{bmatrix} 1 & 1 & -2 & 2 \\ 1 & 1 & 2 & -2 \end{bmatrix}}$.

- Example: Find the pseudoinverse of the matrix $A = \begin{bmatrix} 1+i & 3 & 1 & i \\ 1+i & 1 & 2 & -i \end{bmatrix}$.
  - We previously calculated a singular value decomposition $A = U\Sigma V^*$ with $U = \dfrac{1}{\sqrt{13}} \begin{bmatrix} 3i & 2i \\ 2i & -3i \end{bmatrix}$,

    $$\Sigma = \begin{bmatrix} \sqrt{17} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix},\ \text{and } V^* = \begin{bmatrix} (5+5i)/\sqrt{221} & (-1-i)/\sqrt{52} & (1+i)/2 & (-2+2i)/\sqrt{34} \\ 11i/\sqrt{221} & 3i/\sqrt{52} & -i/2 & -1/\sqrt{34} \\ 7i/\sqrt{221} & -4i/\sqrt{52} & 0 & 4/\sqrt{34} \\ 1/\sqrt{221} & 5/\sqrt{52} & 1/2 & -3i/\sqrt{34} \end{bmatrix}.$$

  - Then $A^\dagger = V \begin{bmatrix} 1/\sqrt{17} & 0 \\ 0 & 1/2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \dfrac{1}{\sqrt{13}} \begin{bmatrix} -3i & -2i \\ -2i & 3i \end{bmatrix} = \boxed{\dfrac{1}{68} \begin{bmatrix} 2-2i & 7-7i \\ 18 & -5 \\ -4 & 20 \\ -14i & 19i \end{bmatrix}}$.

- Example: Find the pseudoinverse of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(x, y, z) = (x + y - z, x + z)$, where the inner products on $\mathbb{R}^3$ and $\mathbb{R}^2$ are the dot product.
  - We use the singular value decomposition formula on the associated matrix $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$.
  - We compute $A^*A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ with eigenvalues $\lambda = 3, 2, 0$ and corresponding unit eigenvectors $\mathbf{v}_1 = \dfrac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \dfrac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$. The singular values are $\sigma_1 = \sqrt{3}$, $\sigma_2 = \sqrt{2}$, and then $\mathbf{w}_1 = A\mathbf{v}_1/\sigma_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{w}_1 = A\mathbf{v}_2/\sigma_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- The SVD for A is $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$ so the SVD for $A^\dagger$ is $A^\dagger = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/2 \\ 1/3 & 0 \\ -1/3 & 1/2 \end{bmatrix}$.
- This means the map $T^\dagger : \mathbb{R}^2 \to \mathbb{R}^3$ has $\boxed{T^\dagger(x, y) = (\frac{1}{3}x + \frac{1}{2}y, \frac{1}{3}x, -\frac{1}{3}x + \frac{1}{2}y)}$, just as calculated before.

- The pseudoinverse is defined in terms of orthogonal complements, and as such we can use it to calculate orthogonal projections and thereby also compute least-squares solutions to inconsistent systems. Explicitly:
- Proposition (Pseudoinverses and Projections): Suppose that V and W are finite-dimensional inner product spaces, $T : V \to W$ is linear, and $F = \mathbb{R}$ or $\mathbb{C}$. Then the following hold:

1. The composition $T^\dagger T$ is the orthogonal projection of V onto $\ker(T)^\perp$.
   - Proof: By definition, the orthogonal projection of V onto $\ker(T)^\perp$ is the linear transformation that is the identity on $\ker(T)^\perp$ and zero on $\left[\ker(T)^\perp\right]^\perp = \ker(T)$.
   - But for $\mathbf{v} \in \ker(T)$ we have $T^\dagger T(\mathbf{v}) = T^\dagger(\mathbf{0}) = \mathbf{0}$, while for $\mathbf{v} \in \ker(T)^\perp$ we have $T^\dagger T(\mathbf{v}) = \mathbf{v}$ since $T^\dagger$ is the inverse of T when restricted to $\operatorname{im}(T)$. Thus, $T^\dagger T$ is the orthogonal projection of V onto $\ker(T)^\perp$.
2. The composition $TT^\dagger$ is the orthogonal projection of W onto $\operatorname{im}(T)$.
   - Proof: By definition, the orthogonal projection of W onto $\operatorname{im}(T)$ is the linear transformation that is the identity on $\operatorname{im}(T)$ and zero on $\operatorname{im}(T)^\perp$.
   - But for $\mathbf{w} \in \operatorname{im}(T)$ we have $TT^\dagger(\mathbf{w}) = \mathbf{w}$ since $T^\dagger$ is the inverse of T when restricted to $\operatorname{im}(T)$, while for $\mathbf{w} \in \operatorname{im}(T)^\perp$ we have $TT^\dagger(\mathbf{w}) = T(\mathbf{0}) = \mathbf{0}$ since $T^\dagger$ is zero on $\operatorname{im}(T)^\perp$. Thus, $TT^\dagger$ is the orthogonal projection of W onto $\operatorname{im}(T)$.
3. We have $TT^\dagger T = T$ and $T^\dagger TT^\dagger = T^\dagger$.
   - Proof: For $\mathbf{v} \in \ker(T)$ we have $TT^\dagger T(\mathbf{v}) = \mathbf{0} = T(\mathbf{v})$ and for $\mathbf{v} \in \ker(T)^\perp$ we have $TT^\dagger T(\mathbf{v}) = T(\mathbf{v})$ since $T^\dagger T$ is the identity on $\mathbf{v}$ by (1). Thus $TT^\dagger T$ and T agree on $\ker(T)$ and $\ker(T)^\perp$, hence are equal.
   - Similarly, $\mathbf{w} \in \operatorname{im}(T)^\perp$ we have $T^\dagger TT^\dagger(\mathbf{w}) = \mathbf{0} = T(\mathbf{w})$ and for $\mathbf{w} \in \operatorname{im}(T)$ we have $T^\dagger TT^\dagger(\mathbf{w}) = T^\dagger(\mathbf{w})$ since $TT^\dagger$ is the identity on $\mathbf{w}$ by (2). Thus $T^\dagger TT^\dagger$ and $T^\dagger$ agree on $\operatorname{im}(T)$ and $\operatorname{im}(T)^\perp$, hence are equal.
4. For any $A \in M_{m\times n}(F)$, the product $A^\dagger A$ represents orthogonal projection onto the row space of A, while $AA^\dagger$ represents orthogonal projection onto the column space of A.
   - Proof: This follows immediately from (1) and (2) by taking associated matrices.
5. For any $A \in M_{m\times n}(F)$ and $\mathbf{c} \in F^m$, if the system $A\mathbf{x} = \mathbf{c}$ is consistent, then the vector $\hat{\mathbf{x}} = A^\dagger \mathbf{c}$ is the unique solution of minimum norm, meaning that any other solution $\mathbf{y} \in F^n$ has $||\hat{\mathbf{x}}|| \le ||\mathbf{y}||$ with equality only for $\mathbf{y} = \hat{\mathbf{x}}$.
   - Proof: Let $T : F^n \to F^m$ have $T(\mathbf{v}) = A\mathbf{v}$. If $A\mathbf{x} = \mathbf{c}$ is consistent then $\mathbf{c} \in \operatorname{im}(T)$ and then $A\hat{\mathbf{x}} = AA^\dagger \mathbf{c} = \mathbf{c}$ by (2): thus $\hat{\mathbf{x}}$ is a solution to the system.
   - For the second part suppose that $A\mathbf{y} = \mathbf{c}$: then $A^\dagger A\mathbf{y} = A^\dagger \mathbf{c} = \hat{\mathbf{x}}$, and so by (1), $\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{y}$ onto $\ker(T)^\perp$.
   - Then for $\mathbf{y}^\perp = \mathbf{y} - \hat{\mathbf{x}} \in [\ker(T)^\perp]^\perp = \ker(T)$ we have the Pythagorean relation $||\mathbf{y}||^2 = ||\hat{\mathbf{x}}||^2 + \left|\left|\mathbf{y}^\perp\right|\right|^2$, which immediately yields $||\hat{\mathbf{x}}|| \le ||\mathbf{y}||$ with equality only for $\mathbf{y} = \hat{\mathbf{x}}$.
6. For any $A \in M_{m\times n}(F)$ and $\mathbf{c} \in F^m$, the vector $\hat{\mathbf{x}} = A^\dagger \mathbf{c}$ is the unique least-squares solution of minimum norm to the system $A\mathbf{x} = \mathbf{c}$, meaning that for any $\mathbf{y} \in F^n$ it is true that $||A\hat{\mathbf{x}} - \mathbf{c}|| \le ||A\mathbf{y} - \mathbf{c}||$ with equality only if $A\hat{\mathbf{x}} = A\mathbf{y}$, and in that case $||\hat{\mathbf{x}}|| \le ||\mathbf{y}||$ with equality only for $\mathbf{y} = \hat{\mathbf{x}}$.
   - Proof: Let $T : F^n \to F^m$ have $T(\mathbf{v}) = A\mathbf{v}$. By our discussion of least squares, the unique vector $\mathbf{w} \in \operatorname{im}(T)$ minimizing the distance $||\mathbf{w} - \mathbf{c}||$ will be the orthogonal projection $\mathbf{c}$ into $\operatorname{im}(T)$, which by (1) is $\mathbf{w} = AA^\dagger \mathbf{c} = A\hat{\mathbf{x}}$.

- In other words, for any $\mathbf{y} \in F^n$ we see that $||A\hat{\mathbf{x}} - \mathbf{c}|| \le ||A\mathbf{y} - \mathbf{c}||$ with equality only if $A\hat{\mathbf{x}} = A\mathbf{y}$.
- For the second part, let $A\hat{\mathbf{x}} = A\mathbf{y} = \mathbf{c}'$. Then by (3) we have $A^\dagger \mathbf{c}' = A^\dagger A\hat{\mathbf{x}} = A^\dagger A A^\dagger \mathbf{c} = A^\dagger \mathbf{c} = \hat{\mathbf{x}}$, and so by applying (5) to the system $A\mathbf{x} = \mathbf{c}'$ we have $||\hat{\mathbf{x}}|| \le ||\mathbf{y}||$ with equality only for $\mathbf{y} = \hat{\mathbf{x}}$.

- Example: Find the matrix associated to orthogonal projection onto the subspace of $\mathbb{C}^4$ spanned by the vectors $(1+i, 3, 1, i)$ and $(1+i, 1, 2, -i)$, with respect to the standard basis under the standard inner product.
  - For $A = \begin{bmatrix} 1+i & 3 & 1 & i \\ 1+i & 1 & 2 & -i \end{bmatrix}$, we wish to compute the orthogonal projection onto the row space of A, which by (4) of the theorem above is given by the product $A^\dagger A$.
  - Since $A^\dagger = \dfrac{1}{68}\begin{bmatrix} 2-2i & 7-7i \\ 18 & -5 \\ -4 & 20 \\ -14i & 19i \end{bmatrix}$, the projection is $A^\dagger A = \boxed{\dfrac{1}{68}\begin{bmatrix} 18 & 13-13i & 16-16i & -5-5i \\ 13+13i & 49 & 8 & 23i \\ 16+16i & 8 & 36 & -24i \\ -5+5i & -23i & 24i & 33 \end{bmatrix}}$.

- Example: Find the solution of minimal norm to the system $x + y - z = 6$, $x + z = 2$.
  - By (5) of the theorem above, the unique solution of minimum norm is $\hat{\mathbf{x}} = A^\dagger \mathbf{c}$ for $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$.
  - Since $A^\dagger = \begin{bmatrix} 1/3 & 1/2 \\ 1/3 & 0 \\ -1/3 & 1/2 \end{bmatrix}$, the solution is $\hat{\mathbf{x}} = A^\dagger \mathbf{c} = \begin{bmatrix} 1/3 & 1/2 \\ 1/3 & 0 \\ -1/3 & 1/2 \end{bmatrix}\begin{bmatrix} 6 \\ 2 \end{bmatrix} = \boxed{\begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}}$.
  - Remark: Solving the system with row-reduction yields $(x, y, z) = (2-t, 4+2t, t)$ which as a vector has squared norm $(2-t)^2 + (4+2t)^2 + t^2 = 6t^2 + 12t + 18 = 6(t+1)^2 + 12$, so the minimum norm indeed occurs when $t = -1$, as claimed.

- Example: Find the least-squares solution of minimal norm to the inconsistent system $2x+2y = 5$, $2x+2y = 3$, $-x + y = 3$, $x - y = -5$.
  - By (6) of the theorem above, for $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 5 \\ 3 \\ 3 \\ -5 \end{bmatrix}$ the least-squares solution of minimal norm is $\hat{\mathbf{x}} = A^\dagger \mathbf{c}$. Earlier we found $A^\dagger = \dfrac{1}{8}\begin{bmatrix} 1 & 1 & -2 & 2 \\ 1 & 1 & 2 & -2 \end{bmatrix}$, so the desired solution is $\hat{\mathbf{x}} = A^\dagger \mathbf{c} = \dfrac{1}{8}\begin{bmatrix} 1 & 1 & -2 & 2 \\ 1 & 1 & 2 & -2 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \\ 3 \\ -5 \end{bmatrix} = \boxed{\begin{bmatrix} -1 \\ 3 \end{bmatrix}}$.

Well, you're at the end of my handout. Hope it was helpful.