Applications of the Spectral Theorem
Tim Smits

The Spectral Theorem

In chapter 7 we answered the question of when an $n \times n$ matrix A is diagonalizable, i.e. when $\mathbb{R}^n$ has an eigenbasis consisting of eigenvectors of A. Is it possible to find an *orthonormal* eigenbasis of $\mathbb{R}^n$ using eigenvectors of A? In some sense, this is the "best" general case basis you can hope for to understand a linear transformation: an orthonormal basis lets you very easily compute the coordinates of some vector x with respect to the eigenbasis, and working in an eigenbasis lets you easily determine the image of said vector x under A.
Unfortunately, it's not always (indeed, there's no reason to even expect it *should* be possible at all!). However, the (perhaps shocking) condition is the following:

Theorem 1 (Real Spectral Theorem)**.** *Let A be an $n \times n$ matrix with real entries. A has an orthonormal basis of eigenvectors if and only if A is symmetric, i.e. $A = A^t$.*

The key to the Spectral Theorem lies in the following facts:

Proposition 2. *Let A be a symmetric $n \times n$ matrix with distinct eigenvalues λ_1 and λ_2, with associated eigenvectors v_1 and v_2. Then $v_1 \cdot v_2 = 0$.*

Proof. $\lambda_2(v_1 \cdot v_2) = v_1 \cdot Av_2 = v_2^t A^t v_1 = v_2^t A v_1 = \lambda_1(v_1 \cdot v_2)$. Since $\lambda_1 \neq \lambda_2$, this forces $v_1 \cdot v_2 = 0$. $\square$

Proposition 3. *Let A be a symmetric $n \times n$ matrix with real entries. Then A has n real eigenvalues counted with multiplicity.*

Proof. Let v be an eigenvector of A with eigenvalue λ. From $Av = \lambda v$, taking complex conjugates says $A\bar{v} = \bar{\lambda}\bar{v}$ (note $A = \bar{A}$ because A has real entries!). We then have $Av \cdot \bar{v} = \bar{v}^t A v = \lambda(v \cdot \bar{v})$. On the other hand, $v \cdot A\bar{v} = \bar{\lambda}(v \cdot \bar{v})$, and since A is symmetric, $Av \cdot \bar{v} = v \cdot A\bar{v}$. This says $\lambda(v \cdot \bar{v}) = \bar{\lambda}(v \cdot \bar{v})$. The quantity $v \cdot \bar{v}$ is non-negative, and in particular is 0 if and only if $v = 0$. This then says $\lambda = \bar{\lambda}$ so λ is real. This says every eigenvalue of A must be real, so the characteristic polynomial of A has n real roots, i.e. A has n eigenvalues with multiplicity. $\square$

Both of these facts are absolutely essential if there is any hope of finding an orthonormal eigenbasis, as the following shows:

Example 4. Let $A = \begin{pmatrix} 0 & 0 & -2 \\ -1 & 2 & -1 \\ 1 & 0 & 3 \end{pmatrix}$. One can check that A is diagonalizable with eigenvalues $1, 2$ and bases of the eigenspaces E_1 and E_2 are given by $\{(2, 1, -1)\}$ and $\{(0, 1, 0), (-1, -1, 1)\}$ respectively. The matrix A is not symmetric, so the Spectral Theorem says that A is not orthogonally diagonalizable. What goes wrong? An orthogonal basis of E_2 is given by $\{(0, 1, 0), (1, 0, -1)\}$. However, $(2, 1, -1) \cdot (0, 1, 0) = 1 \neq 0$. Since any eigenvector $v \in E_2$ is of the form $(c_2, c_1, -c_2)$ for $c_1, c_2 \in \mathbb{R}$, we see that $(2, 1, -1) \cdot (c_2, c_1, -c_2) = 2c_1 + 2c_2$ is 0 only when $c_2 = -c_1$, i.e. the eigenvector is of the form $(-c_1, c_1, c_1)$. Therefore it's impossible to find two linearly independent eigenvectors orthogonal to $(2, 1, -1)$.

The issue of course, is that just because you have an eigenbasis of $\mathbb{R}^n$, running Gram-Schmit to turn it into an orthonormal basis of $\mathbb{R}^n$ doesn't mean that the vectors the algorithm spits out remain eigenvectors (in fact, they usually are not – try it!). However, performing Gram-Schmit inside some eigenspace E_λ *does* preserve these vectors being eigenvectors (by definition, an orthonormal basis of E_λ has to consist of vectors of E_λ, which *are* eigenvectors!). This is where the symmetric condition comes in: the eigenspaces of a *symmetric* matrix are orthogonal to each other, so you can find orthonormal bases of each eigenspace and concatenate them to get an orthonormal basis of $\mathbb{R}^n$ where all vectors are eigenvectors.

We illustrate the orthogonal diagonalization algorithm in the following example:

Example 5. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. Then A is symmetric, so A is orthogonally diagonalizable by the Spectral Theorem. A has rank 1, so $\dim(\ker(A)) = 2$. In particular, A is not invertible, so A has 0 as an eigenvalue, and $\dim(E_0) = \dim(\ker(A)) = 2$. Vectors (x, y, z) in the kernel satisfy $x + y + z = 0$, i.e. $\ker(A) = \mathrm{Span}\{(-1, 1, 0), (-1, 0, 1)\}$. A has 3 eigenvalues with multiplicity, and we know 2 of them are 0. Since $\mathrm{Tr}(A)$ is the sum of the eigenvalues of A, this says the remaining eigenvalue must be 3. (here is why trace is useful!) We then see that $E_3 = \mathrm{Span}\{(1, 1, 1)\}$. Alternatively, one could compute $p_A(\lambda) = 3\lambda^2 - \lambda^3$ and find the eigenvalues this way, but the above sort of argument shows you the types of computational tricks that you can sometimes use to make life easier.

We now run Gram-Schmidt on each eigenspace: after doing so, we find an orthonormal basis of E_0 is given by $\{(-1/\sqrt{2}, 1/\sqrt{2}, 0), (-1/\sqrt{6}, 1/\sqrt{6}, 2/\sqrt{6})\}$ and an orthonormal basis of E_3 is given by $\{(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\}$. The change of basis matrix from the orthonormal eigenbasis $\beta = \{(-1/\sqrt{2}, 1/\sqrt{2}, 0), (-1/\sqrt{6}, 1/\sqrt{6}, 2/\sqrt{6}), (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\}$ to the standard basis $\mathcal{E}$ is given by $S_\beta^\mathcal{E} = \begin{pmatrix} -1\sqrt{2} & -1\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix}$. Since β is an orthonormal basis, $S_\beta^\mathcal{E}$ is orthogonal, so $S_\mathcal{E}^\beta = (S_\beta^\mathcal{E})^{-1} = (S_\beta^\mathcal{E})^t$. We then have

$$A = S_\beta^\mathcal{E} A_\beta (S_\beta^\mathcal{E})^t = \begin{pmatrix} -1\sqrt{2} & -1\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} -1\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}$$

.

Application 1: Quadratic forms

Definition 6. A **quadratic form** $q(x_1, \dots, x_n) : \mathbb{R}^n \to \mathbb{R}$ is a polynomial where all terms have degree 2. If $A = [a_{ij}]$ is an $n \times n$ symmetric matrix, the quadratic form associated to A is the function defined by $q_A(x) = x^t A x = \displaystyle\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$.

Given a quadratic form q, from the coefficients you can read off the symmetric matrix associated to it: if $q = \displaystyle\sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j$, define a symmetric matrix A by $A_{ij} = A_{ji} = \dfrac{1}{2} c_{ij}$ and $A_{ii} = c_{ii}$ for all i, j. It's then easy to see that $x^t A x = q(x)$ for all $x \in \mathbb{R}^n$.

Example 7. The function $q(x, y) = x^2 - xy + y^2$ is a quadratic form. The matrix associated to q is given by $A = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$.

Since the matrix A is symmetric, the Spectral Theorem says it's orthogonally diagonalizable. Let $\beta = \{v_1, \dots, v_n\}$ be an orthonormal eigenbasis of $\mathbb{R}^n$ consisting of eigenvectors of A. Denote the eigenvalues by $\lambda_1, \dots, \lambda_n$. For a vector $x \in \mathbb{R}^n$, suppose that the β-coordinates of x are given by $[x]_\beta = (c_1, \dots, c_n)$. If we work entirely in β-coordinates, we have $[q(x)]_\beta = [x]_\beta^t D [x]_\beta$, where $D = A_\beta$ is the diagonal matrix of eigenvalues of A. Expanding this out (and writing q as a function of coordinates), we find $q(c_1, \dots, c_n) = \lambda_1 c_1^2 + \ldots + \lambda_n c_n^2$.

Quadratic forms often arise in various applications (economics, calculus, physics, statistics, number theory, etc). In many applications, one is interested in optimization problems involving quadratic forms. If you've taken math 32A, the following type of problem probably looks familiar:

Example 8. Consider $f(x, y) = x^2 + 4xy + y^2$. What is the minimal value of f with the constraint $x^2 + y^2 = 1$? If you've taken multivariable calculus, this is a standard exercise in Langrange multipliers. However, since f is a quadratic form, we can use linear algebra. The matrix associated to f is given by $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, and has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$. If $\beta = \{v_1, v_2\}$ is an orthonormal eigenbasis for $\mathbb{R}^2$ of eigenvectors of A, using the above result says in β-coordinates that $f(c_1, c_2) = 3c_1^2 - c_2^2$, where the constraint is $c_1^2 + c_2^2 = 1$. We then clearly see that $f(c_1, c_2) = 3c_1^2 - c_2^2 \geq -c_2^2$. The minimal value is obviously attained at the point $(0, 1)$ in the β-coordinate system. With respect to the standard basis, this says v_2 is a minimizer of $f(x, y)$. One can then check that $v_2 = (-1/\sqrt{2}, 1/\sqrt{2})$.

Definition 9. An $n \times n$ symmetric matrix A is called **positive semi-definite** (denoted $A \geq 0$) if $x^t Ax \geq 0$ for all $x \in \mathbb{R}^n$. Similarly, A is **negative semi-definite** (denoted $A \leq 0$) if $x^t Ax \leq 0$ for all $x \in \mathbb{R}^n$. A is **indefinite** if it's neither positive semi-definite or negative semi-definite. Positive definiteness and negative definiteness are defined how you think they should be.

Another way of phrasing the above is that the quadratic form $q_A(x)$ only takes on non-negative or non-positive values. If $\beta = \{v_1, \ldots, v_n\}$ is an orthonormal eigenbasis, working in β coordinates, we saw that we can write q_A as a function of the β-coordinates by $q_A(c_1, \ldots, c_n) = \lambda_1 c_1^2 + \ldots + \lambda_n c_n^2$. The following is then immediately obvious:

Proposition 10. *A symmetric $n \times n$ matrix A is positive (negative) semi-definite if and only if $\lambda_i \geq 0$ (≤ 0) for all eigenvalues λ_i of A.*

The following is a test for positive definiteness:

Theorem 11 (Sylvester's criterion)**.** *An $n \times n$ symmetric matrix A is positive definite if and only if $\det(A^{(m)}) > 0$ for all m, where $A^{(m)}$ is the $m \times m$ submatrix of A starting from the upper left corner.*

Note that A is negative definite if and only if $-A$ is positive definite, so you can also use Sylvester's criterion to test for negative definiteness.

Example 12. Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$. We have $A^{(1)} = (1)$, so $\det(A^{(1)}) = 1 > 0$. We have $A^{(2)} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$, and $\det(A^{(2)}) < 0$. Finally, $A^{(3)} = A$, and $\det(A) = 0$, so the matrix A is not positive definite (which was already known from A not being invertible).

Example 13. (If you've taken math 32A) Note that a quadratic form associated to an $n \times n$ matrix defines a 2-dimensional surface in $\mathbb{R}^{n+1}$, e.g. $f(x, y) = x^2 + 4xy + y^2$ defines a paraboloid in $\mathbb{R}^3$. The geometry of the associated quadratic form can be classified based on whether the quadratic form is positive semi-definite, negative semi-definite, or indefinite, resulting in an upward facing paraboloid, downward facing paraboloid, or saddle surface respectively. Consider a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$. Recall that f attains a local maximum or minimum at a critical point, i.e. a point $P = (a, b)$ where $\nabla f(a, b) = (0, 0)$. You may also recall the second derivative test: if you look at the expression $D(a, b) = f_{xx}(a, b) f_{yy}(a, b) - (f_{xy}(a, b))^2$, then P is a local minimum if $D(a, b) > 0$ and $f_{xx}(a, b) > 0$, and P is a local maximum if $D(a, b) > 0$ and $f_{xx}(a, b) < 0$. If $D(a, b) < 0$ then P is called a saddle point. This is, of course, a higher dimensional generalization of the second derivative test you learn in single variable calculus: if $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a local minimum, and if $f'(x_0) = 0$ and $f''(x_0) < 0$ then x_0 is a local maximum.

The *Hessian matrix* of f is the matrix $H(a,b) = \begin{pmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{pmatrix}$. Notice that $D(a,b) = \det(H(a,b))$. Using the Hessian matrix, we see that the second partial derivative test is nothing other than using Sylvester's criterion to test whether H is positive or negative definite, and P being a saddle says that H is indefinite. Without going into technical details, a multivariable version of a Taylor series tells you that near the point (a,b), $f(x,y)$ "looks like" the quadratic form defined by $H(a,b)$. The corresponding geometry of the surface then tells you the behavior of f. As a consequence, this view point makes it much easier to imagine how the second derivative test generalizes to higher dimensions (something you would not have seen in 32A!).

Application 2: Singular Value Decomposition

We've seen why diagonalization is great. Unfortunately, not every matrix you care about is square. Is there some sort of analogous matrix factorization? The answer is yes. To begin, we need some theory.

If $T : \mathbb{R}^m \to \mathbb{R}^n$ is the linear transformation $T(x) = Ax$ for A an $n \times m$ matrix, then T is determined entirely by what it does to unit vectors u. This is because $x = \|x\| \frac{x}{\|x\|}$, so $T(x) = \|x\| T(\frac{x}{\|x\|})$. The set of unit vectors in $\mathbb{R}^m$ is called the **unit sphere**, and is denoted by S^{m-1}. Therefore, to understand T, we just need to know what the image of S^{m-1} looks like under T.

The matrix A may not be square, but the matrix A^tA *is* an $m \times m$ matrix, and moreover, A^tA is symmetric. By the Spectral Theorem, A^tA is orthogonally diagonalizable. Let $\beta = \{v_1, \dots, v_m\}$ be an orthonormal eigenbasis of $\mathbb{R}^m$, with eigenvalues $\lambda_1, \dots, \lambda_m$.

Proposition 14. *Let A be an $n \times m$ matrix. Then $A^tA \geq 0$.*

Proof. We have $x^tA^tAx = (Ax)^tAx = (Ax) \cdot (Ax) = \|Ax\|^2 \geq 0$ for all x. □

In particular, this tells us that A^tA has non-negative eigenvalues.

Proposition 15. *Let A be an $n \times m$ matrix and let v be a unit eigenvector of A^tA of eigenvalue λ. Then $\|Av\| = \sqrt{\lambda}$.*

Proof. By the above, $\|Av\|^2 = (Av) \cdot (Av) = v^tA^tAv = v^t\lambda v = \lambda\|v\|^2 = \lambda$, since v is a unit vector. This says $\|Av\| = \sqrt{\lambda}$. □

Definition 16. The **singular values** of an $n \times m$ matrix A are defined by $\sigma_i = \sqrt{\lambda_i}$, where λ_i is an eigenvalue of A^tA. By convention, we will always make sure to write the singular values in **descending** order: $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_m$.

The above then says that $\|Av_i\| = \sigma_i$.

Proposition 17. *Let $\beta = \{v_1, \dots, v_m\}$ be an orthonormal eigenbasis of A^tA. Then $(Av_i)\cdot(Av_j) = 0$ for all $i \neq j$.*

Proof. $(Av_i) \cdot (Av_j) = v_j^tA^tAv_i = v_j^t\lambda_i v_i = \lambda_i(v_i \cdot v_j) = 0$. □

Proposition 18. *Suppose that A is an $n \times m$ matrix with rank$(A) = r$. Then $\sigma_1, \dots, \sigma_r \neq 0$ and $\sigma_{r+1} = \ldots = \sigma_m = 0$.*

Proof. Let $\sigma_1, \dots, \sigma_m$ be the singular values of A, and let $\beta = \{v_1, \dots, v_m\}$ be an orthonormal eigenbasis of $\mathbb{R}^m$ with eigenvectors of A^tA. If $y = Ax \in \mathrm{Im}(A)$, write $x = c_1v_1 + \ldots + c_mv_m$. We have $Ax = c_1Av_1 + \ldots + c_mAv_m$, so $\mathrm{Im}(A) = \mathrm{Span}\{Av_1, \dots, Av_m\}$. By the above, the vectors Av_i are orthogonal, and therefore this is a linearly independent set that spans $\mathrm{Im}(A)$, so it's a basis of $\mathrm{Im}(A)$. Since rank$(A) = r$, we therefore must have only r vectors in this set. The only way this can happen if is $Av_{r+1}, \dots, Av_m = 0$ (because of the ordering we imposed on the singular values), so that $\|Av_{r+1}\| = \ldots = \|Av_m\| = 0$ says $\sigma_{r+1} = \ldots = \sigma_m = 0$. □

We can now prove the following:

Theorem 19. *Let A be an $n \times m$ matrix, and $T(x) = Ax$ the associated linear transformation. The image of S^{m-1} under T is an ellipsoid in $\mathbb{R}^n$.*

Proof. Let $\beta = \{v_1, \ldots, v_m\}$ be our orthonormal eigenbasis of $\mathbb{R}^m$ with eigenvectors of A^tA with eigenvalues $\lambda_1, \ldots, \lambda_m$. Let $x \in S^{m-1}$ and suppose the β-coordinates of x are given by $[x]_\beta = (c_1, \ldots, c_m)$, so that $c_1^2 + \ldots + c_m^2 = 1$. Then $Ax = A(c_1v_1 + \ldots + c_mv_m) = c_1Av_1 + \ldots + c_mAv_m$. Suppose that $\lambda_1, \ldots, \lambda_r$ are non-zero while $\lambda_{r+1}, \ldots, \lambda_m$ are 0, so that we may write $Ax = c_1Av_1 + \ldots + c_rAv_r$. Then for $1 \leq i \leq r$, $\sigma_i \neq 0$, so $Ax = c_1\sigma_1\frac{1}{\sigma_1}Av_1 + \ldots + c_r\sigma_r\frac{1}{\sigma_r}Av_r$. Extend the orthonormal set $\{\frac{1}{\sigma_1}Av_1, \ldots, \frac{1}{\sigma_r}Av_r\} = \{u_1, \ldots, u_r\}$ to an orthonormal basis $\gamma = \{\frac{1}{\sigma_1}Av_1, \ldots, \frac{1}{\sigma_r}Av_r, u_{r+1}, \ldots, u_n\}$ of $\mathbb{R}^n$. We then have $[Ax]_\gamma = (c_1\sigma_1, \ldots, c_r\sigma_r, 0, \ldots, 0)$. Since $c_1^2 + \ldots + c_n^2 = 1$, in particular, $c_1^2 + \ldots + c_r^2 \leq 1$. This says Ax is a point in the ellipsoid (which may be a solid ellipsoid) $E = \{(x_1, \ldots, x_r, 0, \ldots, 0) : (x_1/\sigma_1)^2 + \ldots + (x_r/\sigma_r)^2 \leq 1\}$ in the subspace spanned by $u_1, \ldots, u_r$. $\square$

Example 20. Let $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. We check that $A^tA = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, which has eigenvalues $\lambda_1 = 3$, $\lambda_2 = 1$, and $\lambda_3 = 0$. The singular values are given by $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$, and $\sigma_3 = 0$. An orthonormal eigenbasis of $\mathbb{R}^3$ is given by
$\beta = \{(1/\sqrt{6}, 2/\sqrt{6}, 1/\sqrt{6}), (1/\sqrt{2}, 0, -1/\sqrt{2}), (1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})\}$. Let $x \in S^2$ be a point on the unit sphere. If $[x]_\beta = (c_1, c_2, c_3)$, we have $c_1^2 + c_2^2 + c_3^2 = 1$, and $Ax = c_1Av_1 + c_2Av_2 = c_1\sqrt{3}\frac{1}{\sigma_1}Av_1 + c_2\frac{1}{\sigma_2}Av_2$. With $u_1 = \frac{1}{\sigma_1}Av_1$ and $u_2 = \frac{1}{\sigma_2}Av_2$, we have $\gamma = \{u_1, u_2\}$ is a basis of $\mathbb{R}^2$. Then $[Ax]_\gamma = (c_1\sqrt{3}, c_2)$ is a point on the ellipse $E = \{(x_1, x_2) : (x_1/\sqrt{3})^2 + x_2^2 \leq 1\}$ in the γ-coordinate system. Another way of putting it, E is the solid ellipse in $\mathbb{R}^2$ with semi-major and semi-minor axes u_1 and u_2 of lengths $\sqrt{3}$ and 1 respectively.

The quadratic form associated to this ellipse in the γ-coordinate system is given by the matrix $D = \begin{pmatrix} 1/3 & 0 \\ 0 & 1 \end{pmatrix}$, and since the principal axes are given by $u_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $u_2 = (-1/\sqrt{2}, 1/\sqrt{2})$, these are the eigenvectors of the quadratic form (in standard coordinates). Change of basis then says that the matrix of the quadratic form in standard coordinates is given by $A = S_\gamma^\mathcal{E} D S_\mathcal{E}^\gamma = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}$. In standard coordinates, this is given by $\frac{2}{3}x^2 - \frac{2}{3}xy + \frac{2}{3}y^2 \leq 1$.

The above theorem can be used to give a matrix factorization as follows: let's write down the image Ax of a vector $x \in S^{m-1}$ with respect to the γ-coordinate system, where γ is the basis as in the above theorem. Set β the orthonormal eigenbasis of A^tA, $\{v_1, \ldots, v_m\}$. We could first start in the standard coordinate system of $\mathbb{R}^m$, map to the standard coordinate system of $\mathbb{R}^n$, and then change coordinates to the γ-coordinate system. In terms of matrices, this says $A_\mathcal{E}^\gamma = S_\mathcal{E}^\gamma A$. On the other hand, we could first change coordinates from $\mathcal{E}$ to β, and then convert the images of Av_i into γ-coordinates. In matrix land, this says $A_\mathcal{E}^\gamma = A_\beta^\gamma S_\mathcal{E}^\beta$, so that $S_\mathcal{E}^\gamma A = A_\beta^\gamma S_\mathcal{E}^\beta$. Multiplying though gives $A = S_\gamma^\mathcal{E} A_\beta^\gamma S_\mathcal{E}^\beta$. Set $S_\mathcal{E}^\gamma = U$, $A_\beta^\gamma = \Sigma$, and $S_\mathcal{E}^\beta = V^t$, which then gives $A = U\Sigma V^t$.

The matrix U is the change of basis matrix from the γ basis to the standard basis, and since γ is orthonormal, U is orthogonal. Similar, V is the change of basis matrix from β to $\mathcal{E}$, and since β is orthonormal, V is orthogonal, and $V^t = V^{-1}$ is also orthogonal. Finally, The way we defined u_i was through the relation $Av_i = \sigma_iu_i$. This says $[Av_i]_\gamma = \sigma_ie_i$, and so $A_\beta^\gamma = \Sigma$ is "diagonal" $n \times m$ matrix with non-zero entries $\sigma_1, \ldots, \sigma_r$. This proves the following:

Theorem 21 (Singular Value Decomposition)**.** *Let A be an $n \times m$ matrix with rank$(A) = r$. Then there exist orthogonal $n \times n$ U, $m \times m$ V, and an $n \times m$ matrix Σ such that $A = U\Sigma V^t$. Explicitly, the matrices are given as follows:*

- $U = \begin{pmatrix} u_1 & \dots & u_n \end{pmatrix}$ *where* $u_i = \dfrac{1}{\sigma_i} A v_i$ *for* $1 \leq i \leq r$*, and* $\{u_1, \dots, u_r, u_{r+1}, \dots, u_n\}$ *forms an orthonormal basis of* $\mathbb{R}^n$*.*
- $V = \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix}$ *where* $\{v_1, \dots, v_m\}$ *form an orthonormal eigenbasis of* $\mathbb{R}^m$ *of eigenvectors of* A^tA*. By convention,* v_i *is the eigenvector associated to* λ_i *of the matrix* A^tA *where the eigenvalues are in descending order:* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$*.*
- Σ *has* r *non-zero diagonal entries* $\sigma_1, \dots, \sigma_r$ *and all other entries* 0*.*

Example 22. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$. Then $A^t = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, and $A^tA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. We have $p_{A^tA}(\lambda) = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$, so A has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$, with singular values $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$. A basis of the eigenspace E_3 is given by $\{(1,1)\}$ and a basis of E_1 is given by $\{(-1,1)\}$, so an orthonormal eigenbasis of $\mathbb{R}^2$ is given by $\beta = \{v_1, v_2\} = \{(1/\sqrt{2}, 1/\sqrt{2}), (-1/\sqrt{2}, 1/\sqrt{2})\}$. Set $u_1 = \dfrac{1}{\sigma_1} A v_1 = \dfrac{1}{\sqrt{3}}(1/\sqrt{2}, 2/\sqrt{2}, 1/\sqrt{2}) = (1/\sqrt{6}, \sqrt{2/3}, 1/\sqrt{6})$ and $u_2 = \dfrac{1}{\sigma_2} A v_2 = (1/\sqrt{2}, 0, -1/\sqrt{2})$. We now need to complete this to an orthonormal basis $\{u_1, u_2, u_3\}$ of $\mathbb{R}^3$. To get a unit vector orthogonal to u_1 and u_2, take $u_3 = u_1 \times u_2 = (-1/\sqrt{3}, 1/\sqrt{3}, -1/\sqrt{3})$. This gives $U = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ \sqrt{2/3} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \end{pmatrix}$, $V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.