1. MSRI Graduate Summer School Notes
Jeff Blanchard, 2018
The Singular Value Decomposition

1.1. Orthogonal and Unitary Matrices.

From the linear algebra review, we recall a few things:

- a matrix A is symmetric if $A^T = A$;
- a matrix $A \in \mathbb{R}^{n\times n}$ is orthogonal if $A^T = A^{-1}$ so that $A^T A = I_{n\times n} = AA^T$; in this case the columns of A are an orthonormal basis of $\mathbb{R}^n$;
- a matrix $A \in \mathbb{C}^{n\times n}$ is unitary if $A^* = A^{-1}$ so that $A^*A = I_{n\times n} = AA^*$; again the columns of A are an orthonormal basis of $\mathbb{C}^n$.

Nearly everything we will discuss in my lectures will be based in real spaces, but it all applies to complex spaces with the right modifications. Just keep that in mind when you wonder, "Is this true in $\mathbb{C}^n$?" It probably is and you should try to prove it to yourself. We will do the special case of $\mathbb{R}^n$ so that when you go back one day and prove the case for $\mathbb{C}^n$ you'll get our special case for free.

Exercise 1.1. Prove that if $Q \in \mathbb{R}^{n\times n}$ is orthogonal, then $\|Qx\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^n$. (Use the inner product!)

1.2. Singular Value Decomposition.

Matrix factorizations play a critical role in numerical linear algebra and matrix analysis; among the most important is the *Singular Value Decomposition.* The singular value decomposition, or SVD, of a matrix $A \in \mathbb{R}^{m\times n}$ is a factorization of A into two orthogonal (unitary) matrices and a diagonal matrix,

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

where U, V are orthogonal (unitary) matrices and Σ is diagonal.

Note: we are going to develop the SVD with $m \geq n$. In compressed sensing, we appear on the surface to be concerned with underdetermined linear systems $y = AX$ with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ where $m < n$. However, we first consider the overdetermined case $m \geq n$ which comes up regularly in compressed sensing in terms of intuition, subspaces, iterative algorithms, etc. Ultimately, this choice of $m \geq n$ has no impact on us since we could simply consider the transpose of A if didn't like this choice.

Exercise 1.2. Suppose $A \in \mathbb{R}^{m\times n}$ has the singular value decomposition $A = U\Sigma V^T$. Write A^T in terms of U, Σ, and V. Does A^T have a singular value decomposition?

1.2.1. *Geometric Interpretation.*

Now what exactly is the SVD?[1] The SVD describes how a matrix $A \in \mathbb{R}^{m\times n}$ distorts the unit sphere in $\mathbb{R}^n$ into a hyperellipse in $\mathbb{R}^m$. The image of the unit sphere under the action of the linear transformation defined by A is a hyperellipse. A hyperellipse is a generalization of an ellipse to $\mathbb{R}^m$; a hyperellipse has several principal semiaxes on which the unit sphere is stretched or contracted. To form a hyperellipse, take the unit sphere in $\mathbb{R}^m$ and stretch/contract it by some factors $\sigma_1, \dots, \sigma_m$ (some of them might be zero) in orthonormal directions $u_1, \dots, u_m \in \mathbb{R}^m$. Then the hyperellipse has principal semiaxes $\sigma_i u_i$ of length σ_i.

Definition 1.1 Geometric Definition of Singular Values and Singular Vectors. Suppose $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ and suppose A is full rank, $\text{rank}(A) = r = n$. Let $S \subset \mathbb{R}^n$ be the unit sphere.

The n *singular values* of A are the lengths of the n principal semiaxes of AS written $\sigma_1, \dots \sigma_n$. By convention, the singular values are always indexed in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

The n *left singular vectors* of A are the orthonormal vectors $u_1, u_2, \dots, u_n$ defining the directions of the principal semiaxes of AS.

The n *right singular vectors* of A are unit vectors $v_1, v_2, \dots, v_n \in S$ which are preimages of the principal semiaxes in AS, namely $Av_i = \sigma_i u_i$.

Suppose A is not full rank but has rank $r < n$. Then the dimension of the column space of A has dimension r so that the image of the unit sphere under the mapping by A must lie in an r dimensional subspace. In other words, if A has rank r, then exactly r of the principal semiaxes have nonzero length, i.e. $\sigma_i > 0$ for exactly r values of $i \in \{1, \dots, m\}$. Of course, when $m \neq n$, then $r \leq \min\{m, n\}$.

[1] This discussion is based on *Numerical Linear Algebra* by Trefethen and Bau.

1.2.2. *Reduced SVD.* We begin with the *reduced singular value decomposition* and establish the existence of the reduced SVD for every matrix.

Definition 1.2 Reduced SVD. The reduced singular value decomposition of a full rank matrix $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ is given by

$$A = \hat{U}\hat{\Sigma}\hat{V}^T,$$

where

$$\begin{aligned} \hat{U} &\in \mathbb{R}^{m\times n} \quad \text{has orthonormal columns,} \\ \hat{\Sigma} &\in \mathbb{R}^{n\times n} \quad \text{is diagonal,} \\ \hat{V} &\in \mathbb{R}^{n\times n} \quad \text{is orthogonal.} \end{aligned}$$

By reordering the columns of $\hat{U}$, $\hat{\Sigma}$, and $\hat{V}^T$, we can always write the singular values in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. One factor in the importance of the SVD is that every matrix, even rectangular matrices, have a singular value decomposition.

Theorem 1.2.1. *Any matrix $A \in \mathbb{R}^{m\times n}$ has a reduced singular value decomposition $A = \hat{U}\hat{\Sigma}\hat{V}^T$, where $\hat{U} \in \mathbb{R}^{m\times n}$ has orthonormal columns, $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ is diagonal with nonnegative diagonal entries, and $\hat{V} \in \mathbb{R}^{n\times n}$ is orthogonal.*

To establish theorem, we need to recall the spectral theorem for symmetric matrices.

Theorem 1.2.2. *If $B \in \mathbb{R}^{n\times n}$ is symmetric, then:*

- *B has n real eigenvalues counting multiplicities;*
- *the dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation* $\det(A - \lambda I) = 0$*;*
- *the eigenspaces are mutually orthogonal (i.e. eigenvectors corresponding to different eigenvalues are orthogonal);*
- *B is orthogonally diagonalizable (i.e. there exists and orthogonal matrix Q and a diagonal matrix D such that $B = QDQ^T$).*

Exercises 1.3–1.7 will prove Theorem 1.2.1.

Exercise 1.3. Show that for any matrix $A \in \mathbb{R}^{m\times n}$, A^TA is a symmetric matrix in $\mathbb{R}^{n\times n}$.

Exercise 1.4. Using Thm. 1.2.2 and Ex. 1.3, argue that there exists an orthogonal matrix $\hat{V} = [v_1|v_2|\cdots|v_n]$ and a diagonal matrix D with diagonal entries λ_i where the columns of $\hat{V}$ and the diagonal entries of D form eigenvalue-eigenvector pairs for A^TA. (In other words, show that for $\lambda_i = d_{ii}$, $A^TAv_i = \lambda_i v_i$.)

Exercise 1.5. With the notation from Ex. 1.4, prove that $\|Av_i\|_2 = \lambda_i^{\frac{1}{2}}$ and $\lambda_i \geq 0$ for all $i = 1, \ldots, n$.

Exercise 1.6. Define $u_i = \frac{Av_i}{\|Av_i\|_2}$ and let $\hat{U} = [u_1|u_2|\cdots|u_n] \in \mathbb{R}^{m\times n}$. Prove that the columns of $\hat{U}$ are orthonormal.

Exercise 1.7. Define $\sigma_i = \sqrt{\lambda_i}$. Show that for all $i = 1, \ldots, n$, $Av_i = \sigma_i u_i$. Finally, formulate this set of equations as a matrix equation to complete the proof of Thm. 1.2.1.

From our proof of Thm. 1.2.1, we know that $\hat{V}$ is orthogonal so that $\hat{V}^T\hat{V} = I^{n\times n} = \hat{V}\hat{V}^T$. We also know that $\hat{U}$ has orthonormal columns. If $m = n$, then $\hat{U}$ is also orthogonal. If $m > n$, then $\hat{U}$ is rectangular and thus has no inverse; $\hat{U}$ can't possibly be orthogonal.

1.2.3. *Full SVD.* Often, in applications, the reduced SVD is all that is needed or desired. However, for theoretical reasons, we want to be able have a matrix U that is also orthogonal.

Definition 1.3 Full SVD. The full singular value decomposition of a matrix $A \in \mathbb{R}^{m\times n}$ is given by

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

where

$$\begin{aligned} U \in \mathbb{R}^{m\times m} &\quad \text{is orthogonal;} \\ \Sigma \in \mathbb{R}^{m\times n} &\quad \text{is diagonal;} \\ V \in \mathbb{R}^{n\times n} &\quad \text{is orthogonal.} \end{aligned}$$

(By convention, the singular values are decreasing $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.)

Notice that here we did not require $m \geq n$. Why not?

The reason $\hat{U}$ was not orthogonal was that it was a rank n matrix in $\mathbb{R}^{m\times n}$ with $m > n$ rows. It can't possibly have an inverse let alone satisfy $\hat{U}^T = \hat{U}^{-1}$. The nice feature of $\hat{U}$ is the columns are orthonormal. The problem is that it does not have enough columns. We are missing $m-n$ orthogonal columns. What should we do? We simply append $m-n$ orthogonal columns. How do we do this? Well let's say we do this arbitrarily or with some black box. In practice we can do this by appending $m-n$ vectors $u_{n+1}, \ldots, u_m$ so that $u_1, \ldots, u_m$ are linearly independent, and then perform a Gram-Schmidt orthogonalization or Householder reflectors. We'll omit this at least for now.

So, after finding $\hat{U}$ we define the matrix $U = [\hat{U}|u_{n+1}|\cdots|u_m]$ where U is orthogonal.

Exercise 1.8. Describe the conditions we need on $u_{n+1}, \ldots, u_m$ so that U is orthogonal.

Exercise 1.9. How do we define V?

Now we know $A \in \mathbb{R}^{m\times n}$ so it must be that $U\Sigma V^T \in \mathbb{R}^{m\times n}$ as well since they are equal. Since $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$, we require $\Sigma \in \mathbb{R}^{m\times n}$. In other words, the diagonal matrix must have the same dimension as A.

Exercise 1.10. How do we define Σ so that $\hat{U}\hat{\Sigma}\hat{V}^T = A = U\Sigma V^T$?

The SVD is perfectly valid for complex matrices $A \in \mathbb{C}^{m\times n}$.

Definition 1.4 Full SVD (General). The full singular value decomposition of a matrix $A \in \mathbb{C}^{m\times n}$ is given by

$$A = U\Sigma V^*,$$

where $V^* = \overline{V^T}$ is the conjugate transpose of V and

$$\begin{aligned} U \in \mathbb{C}^{m\times m} &\quad \text{is unitary;} \\ \Sigma \in \mathbb{R}^{m\times n} &\quad \text{is diagonal;} \\ V \in \mathbb{C}^{n\times n} &\quad \text{is unitary.} \end{aligned}$$

Again, no assumption like $m \geq n$ was made in this definition. From now on, we'll call the Full SVD just the SVD. If we want to specify, we will refer the reduced SVD in its full name.

Theorem 1.2.3 (SVD: Existence and Uniqueness)**.** *Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition given by Def. 1.4. Furthermore, the singular values $\{\sigma_i\}$ are uniquely determined.*

For the real case, our proof of Thm. 1.2.1 combined with Exercises 1.8, 1.9, and 1.10 prove the first part of Thm. 1.2.3. The orthogonal matrices U, V are not uniquely determined unless A is square with distinct singular values. The singular values, however, **are** uniquely determined because they are the eigenvalues of the matrix A^TA. In Ex. 1.4, we showed that the set of eigenvector-eigenvalue pairs of A^TA is $\{v_i, \sigma_i^2\}_{i=1}^n$.

Exercise 1.11. Prove that the set $\{u_i, \sigma_i^2\}_{i=1}^n$ is a set of eigenvector-eigenvalue pairs of AA^T.

Exercises 1.4 and 1.11 show that the nonzero singular values of A are uniquely determined as the square roots of the nonzero eigenvalues of A^TA and AA^T. (f an eigenvalue of A^TA has multiplicity greater than 1 any orthonormal basis of the eigenspace associated with that eigenvalue would be acceptable columns of V; this is why the orthogonal matrices are not necessarily unique.

SVD Additional Exercises

Exercise 1.12. Two matrices $A, B \in \mathbb{C}^{m\times m}$ are *unitarily equivalent* if there exists a unitary matrix $Q \in \mathbb{C}^{m\times m}$ such that $A = QBQ^*$. Prove that any two unitarily equivalent matrices must have the same singular values? (Hint: Use Thm. 1.2.3.)

Exercise 1.13. Prove that if A is symmetric ($A^T = A$) then the singular values of A are the absolute values of the eigenvalues of A.

The next two exercises show a typical application of the SVD to solving a linear system of equations.

Exercise 1.14 Change of Basis. Suppose $A = U\Sigma V^T$ and let $b' = U^T b$ and $x' = V^T x$. Prove $Ax = b$ if and only if $\Sigma x' = b'$.

Exercise 1.15. Use the previous exercise to formulate an algorithm for using the SVD to solve a system of linear equations in $\mathbb{R}^{n\times n}$. (Implement the algorithm using `svd` in Matlab or using `np.linalg.svd` from the numpy.linalg package in Python.)

In the following two exercises, we see that the rank of a matrix A can be read directly from the SVD $A = U\Sigma V^T$. Recall the definition of the rank of a matrix A, $\text{rank}(A) = \dim(\text{Col}A)$.

Exercise 1.16. Suppose $A \in \mathbb{R}^{m\times n}$ has the SVD $A = U\Sigma V^T$. Show that if $\{z_1, \dots, z_t\}$ is a basis for $\text{null}(A)$, then $\{V^T z_1, \dots, V^T z_t\}$ is a basis for $\text{null}(\Sigma)$.

Exercise 1.17. Use the rank-nullity theorem and the previous exercise to prove that $\text{rank}(A)$ is the number of nonzero singular values of A.

Finally, we relate the ℓ_2-norm of a matrix to the largest singular value.

Theorem 1.2.4. *For any matrix $A \in \mathbb{R}^{m\times n}$, the ℓ_2-norm of A is equal to the largest singular value, i.e. $\|A\|_2 = \sigma_1$.*

Exercise 1.18. Prove Thm. 1.2.4. (Hint: Use the result of Exercise 1.1 to show $\|A\|_2 = \|\Sigma\|_2$.)

Although a rectangular matrix does not have an inverse, we will see in Chapter 1.3 that we can find a best fitting solution (in terms of ℓ_2-norm of the error) using something called a pseudo inverse, denoted $A^\dagger$. The SVD offers an efficient way to work with the pseudo inverse.

Exercise 1.19. Let $A \in \mathbb{R}^{m\times n}$.

(i) Suppose $m \geq n$. Define $A^\dagger = (A^T A)^{-1} A^T$. If we have the singular value decomposition $A = U\Sigma V^T$, find $A^\dagger$ in terms of U, Σ, V.

(ii) Suppose $m < n$. Define $A^\dagger = A^T (AA^T)^{-1}$. If we have the singular value decomposition $A = U\Sigma V^T$, find $A^\dagger$ in terms of U, Σ, V.

Here's a good discussion question (for a bunch of math people, not on a first date or anything).

Exercise 1.20. Why do you think they are called *singular values*? What do singular values have to do with singular matrices? Can you think of a way you might be able to say that a matrix was close to being singular? What would you have called them?

Do this crazy problem taken from Trefethen and Bau last ...

Exercise 1.21. Let $A \in \mathbb{R}^{m\times n}$ and let $B \in \mathbb{R}^{n\times m}$ be the matrix obtained by rotating A on the page by ninety degrees clockwise on the page. Do A and B have the same singular values? Either prove that they do or construct a counter example.[2]

[2] This exercise is taken from *Numerical Linear Algebra* by Trefethen and Bau.

1.3. **Least Squares Solutions.** We now investigate solving an overdetermined linear system of equations in the matrix form $y = Ax$ where $A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n, y \in \mathbb{R}^m$ with $m \geq n$. The problem of "solving" a linear system of equations can be stated as follows:

> *Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $y \in \mathbb{R}^m$, can we find a vector $x \in \mathbb{R}^n$ so that $y = Ax$?*

The case $m = n$ was likely a major focus of your first linear algebra course.

Exercise 1.22. State a necessary and sufficient condition for solving $y = Ax$ when $A \in \mathbb{R}^{n \times n}$, and $x, y \in \mathbb{R}^n$. State a few more equivalent conditions.

Now let's focus on the case where $m > n$. In this setting, there are more rows than columns in the matrix A.

Exercise 1.23. What is the largest number of linearly independent rows in the matrix $A \in \mathbb{R}^{m \times n}$?

Suppose the matrix A has full rank, $\text{rank}(A) = n$. The equation $y = Ax$ either has no solution, because the equations are inconsistent, or it has a unique solution because the full set of m equations is consistent. In other words, if a solution exists, this overdetermined system of equations simply has $m - n$ equations which are linear combinations of n linearly independent equations. In this case we can find the unique solution by reducing the system of equations to an $n \times n$ system of equations.

Now we consider the case where we have an inconsistent $m \times n$, $m > n$, set of equations. In this setting a solution does not exist. However, we still want to "solve" this system of equations. The *least squares* solution to an inconsistent system of equations is the vector $x \in \mathbb{R}^n$ which is closest to being a solution in the sense of Euclidean distance. Equivalently, we want to find a vector x so that the residual $r = y - Ax$ has the smallest ℓ_2 norm. This is formulated as an optimization problem.

$$\text{Given } \ y \text{ and } A, \quad \text{minimize} \|y - Ax\|_2 \ \text{ subject to } \ x \in \mathbb{R}^n.$$

Note that since $\text{rank}(A) = n$, the set $\text{col}(A) = \{Ax : x \in \mathbb{R}^n\}$ is an n dimensional subspace of $\mathbb{R}^m$. In other words, the set $\{Ax : x \in \mathbb{R}^n\}$ is an n dimensional hyperplane in $\mathbb{R}^m$. If we want to minimize $\|r\|_2 = \|y - Ax\|_2$, we need to find the vector $\bar{x} \in \mathbb{R}^n$ so that $y - A\bar{x}$ is perpendicular, or orthogonal to the hyperplane $\{Ax : x \in \mathbb{R}^n\}$.

Exercise 1.24 Normal Equations. Follow this outline to derive the normal equations for solving the least squares problem.

- State a condition in terms of the inner product to guarantee $(y - A\bar{x}) \perp \{Ax : x \in \mathbb{R}^n\}$.
- Write this inner product in the form $< w, z >= z^T w$.
- Conclude that $A^T(y - A\bar{x}) = 0$.
- Simply rewrite this fact to form the normal equations:

  Given the inconsistent system $y = Ax$, if

$$A^T A\bar{x} = A^T y$$

  then $\bar{x}$ is the least squares solution to this system of equations.

If $r = y - A\bar{x}$ is the residual of the least squares solution, we measure the error of our solution in four equivalent ways.

- ℓ_2 error: $\|r\|_2 = \sqrt{\sum_{i=1}^m |r_i|^2}$;
- Squared Error (SE): $\|r\|_2^2 = \sum_{i=1}^m |r_i|^2$;
- Mean Squared Error (MSE):

$$\frac{\|r\|_2^2}{m} = \frac{\sum_{i=1}^m |r_i|^2}{m};$$

- Root Mean Squared Error (RMSE):

$$\frac{\|r\|_2}{\sqrt{m}} = \sqrt{\frac{\sum_{i=1}^m |r_i|^2}{m}};$$

1.3.1. *The Pseudoinverse.* If $A \in \mathbb{R}^{m\times n}$, $m > n$ is full rank, then we know that $A^T A$ is nonsingular. In that case the normal equations give rise to a straightforward solution to the least squares problem.

Definition 1.5. If $A \in \mathbb{R}^{m\times n}$, $m > n$ is full rank, the *Moore-Penrose Pseudoinverse* of A is

$$A^\dagger = (A^T A)^{-1} A^T.$$

(Often the notation for the pseudoinverse is A^+ but using a dagger † is too cool to pass up and is almost as common.)

Exercise 1.25. Use Exer. 1.24 to show that if $y = Ax$ is an inconsistent system, the least squares solution is given by $\bar{x} = A^\dagger y$.

1.3.2. *Underdetermined Systems of Equations.* In this section, we now consider the case where $A \in \mathbb{R}^{m\times n}$ with $m < n$. This system has either no solution, because the equations are inconsistent, or it has infinitely many solutions. We ignore the case where no solution exists, since we could not extend such a system to an $n \times n$ system with a solution. In other words, the system is not full rank. Instead, we determine a method for selecting a particular solution from the infinitely many choices of a consistent underdetermined system of equations. Throughout this section, let $y = Ax$ with $y \in \mathbb{R}^m, x \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, m < n$.

From all of our previous discussions, we known that $\text{rank}(A) \leq \min(m, n) = m$. We consider only the case where A is full rank, namely $\text{rank}(A) = m$. In this case, we already determined that when A has full rank, then AA^T is nonsingular while $A^T A$ is singular. So, for an underdetermined linear system of equations we need a different pseudoinverse as the one from Definition 1.5 is undefined.

Definition 1.6. If $A \in \mathbb{R}^{m\times n}$ with $m < n$, the pseudoinverse is defined as

$$A^\dagger = A^T (AA^T)^{-1}.$$

Exercise 1.26. If $y = Ax$, prove that $A^T (AA^T)^{-1} y - x \in \text{null}(A)$.

From Exer. 1.26, we see that if $y = Ax$, then $A^\dagger y - x \in \text{null}(A)$. We now want to show that this is the least squares solution.

Exercise 1.27. Suppose $y \in \mathbb{R}^m$ and define $\bar{x} = A^\dagger y$. Prove that $y = A\bar{x}$ so that $\bar{x}$ is a solution to the least squares problem. Determine $\|r\|_2 = \|y - A\bar{x}\|_2$. (That last part is obvious, but describes why we might call it the least squares solution.)

We will call $\bar{x} = A^\dagger y$ the least squares solution to the underdetermined system of equations $y = Ax$.

Exercise 1.28. Is $\bar{x}$ the unique least squares solution? (Hint: If $\bar{r} = y - A\bar{x}$ is the residual for our least squares solution, and for any vector $x \in \mathbb{R}^n$ we define $r = y - Ax$, are there any vectors $x \in \mathbb{R}^n$ with $\|\bar{r}\|_2 = \|r\|_2$?)

While this method provides us a means to find some solution to an underdetermined system of equations, we have no control over its practical value for a given application. One trivial fact is that if $\bar{x} \neq 0$, then $\bar{x}$ is orthogonal to the null space of A.

Least Squares Additional Exercises

Exercise 1.29. Develop a 3×2 system of consistent equations with a unique solution. Then solve this system.

What happens to your consistent system if you add 0.1 to a single entry in the right hand side? Find the least squares solution to this system and compare it with the solution to the consistent system.

This relates directly to applications in signal processing where you want to measure x using A but your measurements are always at least slightly corrupted by noise so that you get $y = Ax + e$ for some noise (error) vector e.

Exercise 1.30. Recall that if we know the SVD of A, namely $A = U\Sigma V^T$, we developed a straightforward expression for the pseudoinverse in Exer. 1.19. Use this to develop an algorithm for solving the least squares problem by transforming the normal equations into a diagonal system via a change of basis.

Exercise 1.31. Write a MATLAB function for solving the least squares problem using the previous exercise. Considering efficiency, do you need the full SVD or will the partial SVD suffice? Use MATLAB's built-in function `svd` to get either the full or reduced SVD. (Use Python if you prefer.)

Exercise 1.32. Prove $\bar{x}$ is orthogonal to the null space of A. What does this mean? (Hint: Rewrite $\bar{x}$ in terms of the pseudoinverse of A.)

Exercise 1.33. Find the underdetermined pseudoinverse in terms of the singular value decomposition. (Hint: It will be useful to consider the diagonal matrix Σ in block form: $\Sigma = \left[S_{m\times m} | 0_{m\times(n-m)}\right]$.)

Exercise 1.34. Develop an algorithm to solve an underdetermined system of equations and implement it in MATLAB or Python.

Better yet, implement an algorithm that uses the `svd` to return the least squares solution for any of $m > n, m = n, m < n$. You should notice that these are closely related and really only requires a minimal use of conditional statements.