SVD review problems for 18.06

Steven G. Johnson, MIT Applied Mathematics

November 13, 2022

1 SVD variants

One of the confusing things about the SVD is that there are a few different conventions for the sizes of the various matrices, and this can make it more difficult to compare exam and homework questions from different textbooks, semesters, and instructors. If A is an $m \times n$ matrix (of real numbers) of rank r, then the main variants of the SVD $A = U\Sigma V^T$ are:

1. The **"full" SVD**: U is $m \times m$ and V is $n \times n$ (i.e. both are *unitary*), and Σ is $m \times n$.
2. The **"compact" SVD**: U is $m \times r$, V is $n \times r$, and Σ is $r \times r$. (Sometimes denoted $\hat{U}\hat{\Sigma}\hat{V}^T$ to distinguish it from the full SVD.)
3. The **"thin" SVD**: Σ is $m \times m$ or $n \times n$, **whichever is smaller**. Hence, there are two cases:
   (a) A is "tall" ($m \geq n$): U is $m \times n$, V is $n \times n$, Σ is $n \times n$.
   (b) A is "wide" ($m \leq n$): U is $m \times m$, V is $n \times m$, Σ is $m \times m$.

For example, suppose that A is a 4×3 matrix with rank 2. Then the three variants would look like:

1. **full**: $A = \underbrace{\begin{pmatrix} u_1 & u_2 & u_3 & u_4 \end{pmatrix}}_{U=4\times 4 \text{ unitary}} \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \\ & & 0 \end{pmatrix}}_{\Sigma = 4\times 3} \underbrace{\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}^T}_{(V=3\times 3 \text{ unitary})^T}.$

2. **compact**: $A = \underbrace{\begin{pmatrix} u_1 & u_2 \end{pmatrix}}_{U=4\times 2} \underbrace{\begin{pmatrix} \sigma_1 & \\ & \sigma_2 \end{pmatrix}}_{\Sigma = 2\times 2} \underbrace{\begin{pmatrix} v_1 & v_2 \end{pmatrix}^T}_{(V=3\times 2)^T}.$

3. **thin**: $A = \underbrace{\begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix}}_{U=4\times 3} \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \end{pmatrix}}_{\Sigma = 3\times 3} \underbrace{\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}^T}_{(V=3\times 3 \text{ unitary})^T}.$

Computer software for the SVD tends to return either the "thin" or "full" forms of the SVD, with the latter being smaller and hence more efficient. (The "compact" form would be even smaller yet, but it is difficult for a computer with finite precision to distinguish singular values that are very small from actual zeros, i.e. roundoff errors make it hard to reliably shrink the thin to the compact form.)

1.1 Important features of *all* SVD variants

Every variant of the SVD has a number of important features in common:

1. Columns of U ("left" singular vectors $u_1, u_2, \ldots$) are **orthonormal**, and columns of V ("right" singular vectors $v_1, v_2, \ldots$) are **orthonormal**.

2. The **rank** of A is equal to the number of singular values $\sigma_k > 0$, which lie on the **diagonal** of Σ.

3. The **first** r **columns** $u_1, \ldots, u_r$ of U (sometimes denoted $\hat{U}$) are an orthonormal basis for $C(A)$. The **first** r **columns** $v_1, \ldots, v_r$ of V (sometimes denoted $\hat{V}$) are an orthonormal basis for $C(A^T)$. This gives:

   (a) $A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$. (Any other singular vectors are multiplied by 0's.)

   (b) A sends these r orthonormal "input" vectors v_k into orthogonal "outputs" parallel to u_k: $Av_k = \sigma_k u_k$. (This makes it an especially ***nice*** **orthonormal basis** for understanding A, because the orthogonality is *preserved* by A.)

4. You can get a low-rank **approximation** for A (in some sense the "best" possible low-rank approximation) by simply **dropping** smaller singular values, i.e. just setting those σ_k terms to **zero**. (For convenience, the singular values are conventionally sorted in descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.)

The only difference between the variants is whether there are "extra" columns in U and V. Any additional columns beyond the r-th must be *orthogonal* to $C(A)$ (for U) or $C(A^T)$ (for V), and hence must be in $C(A)^\perp = N(A^T)$ or $C(A^T)^\perp = N(A)$, respectively:

1. Any additional columns of U (beyond u_r) are in $N(A^T)$. If U is $m \times m$ (square/unitary), then these additional $m - r$ columns are an orthonormal *basis* for $N(A^T)$.

2. Any additional columns of V (beyond v_r) are in $N(A)$. If V is $n \times n$ (square/unitary), then these additional $n - r$ columns are an orthonormal *basis* for $N(A)$.

2 “Thin” SVD practice problems

The following practice problems from previous exams are adapted to the “thin” SVD.

Spring 2019 quiz 2, problem 1

A random 4×3 matrix A has a thin SVD computed with Julia. The singular values are (to 3 digits):

$$\sigma_1, \sigma_2, \sigma_3 = 2.07, 0.996, 0.485$$

and the corresponding singular vectors are the columns of

$$U = \begin{pmatrix} -0.534 & 0.697 & 0.397 \\ -0.325 & -0.691 & 0.539 \\ -0.650 & -0.156 & -0.108 \\ -0.431 & -0.108 & -0.735 \end{pmatrix}, \qquad V = \begin{pmatrix} -0.392 & 0.466 & 0.793 \\ -0.730 & 0.367 & -0.577 \\ -0.560 & -0.805 & 0.197 \end{pmatrix}.$$

Questions:

1. Is $A^T A$ invertible? Why or why not?
2. If y is a vector perpendicular to **every** column of A, find and circle vectors in U and/or V above that must **also** be $\perp y$.
3. Is AA^T invertible? Why or why not?
4. How many solutions to $Ax = b$ are likely to exist for a randomly generated b such as $b = \begin{pmatrix} 1.26 \\ -0.649 \\ -1.87 \\ -1.67 \end{pmatrix}$.

   Explain briefly.
5. What is the dimension of the **orthogonal complement** of the **row space** of A?

Solutions:

1. **Yes**. A has 3 nonzero singular values, so it is **rank 3** and hence **full column rank**. We saw in class that $A^T A$ and A have the **same rank**, so $A^T A$ is a 3×3 matrix of rank 3 and hence must be invertible.

   In fact, we can explicitly write $(A^T A)^{-1}$ in terms of the SVD above, since

   $$A^T A = (U\Sigma V^T)^T U \Sigma V^T = V \Sigma^T \cancelto{I}{U^T U} \Sigma V^T = V \underbrace{\begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{pmatrix}}_{\Sigma^T\Sigma = \Sigma^2} V^T,$$

   using the orthonormal columns of U. Since V is a *square* matrix with orthonormal columns, it is *unitary* ($V^T = V^{-1}$), while diagonal matrices like Σ^2 are easy to invert, and thus the inverse is:

   $$(A^T A)^{-1} = (V^T)^{-1} \Sigma^{-2} V^{-1} = V \underbrace{\begin{pmatrix} \sigma_1^{-2} & & \\ & \sigma_2^{-2} & \\ & & \sigma_3^{-2} \end{pmatrix}}_{\Sigma^{-2}} V^T.$$

2. y is in $C(A)^\perp$, and the first $r = 3$ columns of U (i.e. all the columns in this case) are a **basis** for $C(A)$, so y must be **orthogonal to all three columns of** U. (Equivalently, $U^T y = \vec{0}$.)
3. **No**. As in the first part, $\text{rank}(AA^T) = \text{rank}(A^T) = \text{rank}(A) = 3$, so AA^T is a 4×4 matrix with rank 3 and hence is singular/non-invertible.

4. **None**. A is a 4×3 matrix with rank 3, so $C(A)$ is a 3d subspace of $\mathbb{R}^4$ — it's vanishingly unlikely that a vector b with independently chosen random entries will lie in this subspace.

5. **0**. A is a 4×3 matrix with rank 3, so its row space $C(A^T)$ is a 3-dimensional subspace of $\mathbb{R}^3$, hence $C(A^T) = \mathbb{R}^3$ and its complement is the 0-dimensional subspace $\{\vec{0}\}$.

Spring 2019 practice exam 2, problem 9

Suppose we are given the thin SVD of an $m \times n$ matrix A on a computer. **Questions:**

1. How would you obtain the rank r?
2. How would you check if $A^T A$ is invertible?
3. How would you check, given a vector b, if $Ax = b$ has a solution?
4. If $Ax = b$ has at least one solution, how could use use the thin SVD to obtain a solution x?

Solutions:

1. The rank r is the number of (nonzero) singular values σ_k.

   (Of course, roundoff errors on a compute throw a wrench into this on a real computer. If σ_1 is the biggest singular value, we might treat any singular values $\lesssim 10^{-15}\sigma_1$ as being indistinguishable from zero.)

2. $A^T A$ is invertible if $\text{rank}(A^T A) = \text{rank}(A) = r$ is equal to n, i.e. if A is **full column rank**.

3. We need to check whether $b \in C(A)$. If we let $\hat{U}$ denote the first r columns of U, then these are an **orthonormal basis** for $C(A)$ and hence projection onto $C(A)$ is given by $\hat{U}\hat{U}^T$. So, $Ax = b$ is solvable if $\boxed{\hat{U}\hat{U}^T b = b}$. (On a real computer we would check approximate equality up to many digits, due to roundoff errors.)

4. If $Ax = b$ is solvable, then we can write

   $$b = \hat{U} \underbrace{\hat{U}^T b}_{c} = c_1 u_1 + c_2 u_2 + \cdots + c_r u_r$$

   in the basis of the first r left singular vectors (from the previous part, where $c = \hat{U}^T b$ are the coefficients obtained via dot products). We can look for a particular solution x expressed in the $(v_1 \cdots v_r) = \hat{V}$ basis for $C(A^T)$:

   $$x = \hat{V} \underbrace{\hat{V}^T x}_{y} = y_1 v_1 + y_2 v_2 + \cdots + y_r v_r.$$

   Then

   $$Ax = \underbrace{\sigma_1 y_1}_{c_1} u_1 + \underbrace{\sigma_2 y_2}_{c_2} u_2 + \cdots + \underbrace{\sigma_r y_r}_{c_r} u_r = b$$

   since $Av_k = \sigma_k u_k$, so by inspection we have $y_k = c_k/\sigma_k$, i.e.

   $$\boxed{x = \sigma_1^{-1} v_1 \underbrace{u_1^T b}_{c_1} + \cdots + \sigma_r^{-1} v_r \underbrace{u_r^T b}_{c_r} = \hat{V}\hat{\Sigma}^{-1}\hat{U}^T b},$$

   where $\hat{\Sigma}$ is the $r \times r$ diagonal matrix of the first r singular values (ala the compact SVD).

   In fact, $\hat{V}\hat{\Sigma}^{-1}\hat{U}^T$ is something called the "pseudo-inverse" of A, oddly denoted A^+, which generally gives a least-squares solution. When $b \in C(A)$, the least-squares solution is an *exact* solution, which is why this works. But 18.06 doesn't always cover the pseudo-inverse. **Probably this problem is too hard for a real exam** unless the pseudo-inverse was covered in lecturel.

Spring 2019 practice exam 2, problem 11

Questions:

1. Use $Q^T Q = I$ to show that $\det Q = \pm 1$ for a (real) unitary matrix Q.
2. Show that $\det(A)$ is $\pm$ the product of its singular values for a non-singular A.

Solutions:

1. Using the properties of determinants, $\det(Q^T Q) = \underbrace{\det(Q^T)}_{=\det Q}\det(Q) = (\det Q)^2 = \det I = 1$, and the only real numbers whose square $= 1$ are $\det Q = \pm 1$.
2. If A is a square non-singular $m \times m$ matrix, then its SVD is

$$A = \underbrace{U}_{m\times m}\underbrace{\Sigma}_{m\times m}\underbrace{V^T}_{m\times m}$$

(and all of the SVD variants are the same!). Since U and V are square matrices with orthonormal columns, they are unitary, so

$$\det A = \cancelto{\pm 1}{\det(U)}\det(\Sigma)\cancelto{\pm 1}{\det(V^T)} = \pm\det\Sigma,$$

and $\det\Sigma$ (being the determinant of a diagonal matrix) is simply the product of the diagonal entries, i.e. the product of the singular values.

Spring 2019 practice exam 2, problem 12

Question: Describe how you can use the thin SVD $A = U\Sigma V^T$ to obtain the projection matrix onto the column space of an $m \times n$ matrix A?

Solution: The first r columns of U, where r is the number of singular values (> 0) from the diagonal of Σ, are an orthonormal basis for $C(A)$. Denote the first r columns by $\hat{U}$. Therefore, the projection matrix onto $C(A) = \hat{U}\hat{U}^T$.

(There is some ambiguity in the rank-0 case. What does a matrix $\hat{U}$ with zero columns mean? In this case the projection is simply the $m \times m$ matrix of zeros, however, since $C(A)$ is 0-dimensional.)

Spring 2019 practice exam 2, problem 13

Question: Describe the solution to the least-squares problem of minimizing $\|Ax - b\|$ for an arbitrary b (not necessarily in the column space of A), in terms of the thin SVD $A = U\Sigma V^T$ of the $m \times n$ matrix A. For simplicity, assume that A has full column rank (independent columns).

Solution: We are told that A has full column rank, i.e. rank $r = n$, which means that it is a "tall" matrix ($m \geq n$), there are n singular values, and the thin SVD is

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times n}\underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}}_{\Sigma = m\times m}\underbrace{V^T}_{n\times n}.$$

Note, in particular, that the columns of U are an orthonormal basis for $C(A)$ and V is a *square* unitary ($V^T = V^{-1}$) matrix. There are various ways to proceed, but one option is just to simplify the usual normal-equations formula for the least-squares solution:

$$
\begin{aligned}
\hat{x} &= (A^T A)^{-1} A^T b \\
&= \left[(U\Sigma V^T)^T U\Sigma V^T\right]^{-1} (U\Sigma V^T)^T b \\
&= \left[V \cancelto{\Sigma}{\Sigma^T} \cancelto{I}{U^T U} \Sigma V^T\right]^{-1} V\Sigma U^T b \\
&= \left[V\Sigma^2 V^T\right]^{-1} V\Sigma U^T b \\
&= \cancelto{V}{(V^T)^{-1}} \Sigma^{-2} \cancelto{I}{V^{-1} V} \Sigma U^T b \\
&= V\Sigma^{-1} U^T b = \boxed{V \underbrace{\begin{pmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{pmatrix}}_{\Sigma^{-1}} U^T b}.
\end{aligned}
$$

That is, we would take the dot products $U^T b$, then divide by the singular values, then multiply by V. Crucially, this derivation relies on V being square, so that we could write $\left[V\Sigma^2 V^T\right]^{-1}$ as a product of the inverses in reverse order.

- Related to a comment for a previous problem above, $V\Sigma^{-1}U^T$ is based on something called the "pseudo-inverse" of A, oddly denoted A^+, which generally gives a least-squares solution from an SVD. But 18.06 doesn't always cover the pseudo-inverse. **Probably this problem is too hard for a real exam**, since it is basically trying to teach you the pseudo-inverse.