Example 160. Show that the eigenvalues of A^TA are all nonnegative.

Proof. Suppose that λ is an eigenvalue of A^TA. Then $A^TA\boldsymbol{v} = \lambda\boldsymbol{v}$ (where $\boldsymbol{v}$ is a λ-eigenvector).
It follows that $\underbrace{\boldsymbol{v}^TA^TA\boldsymbol{v}}_{=\|A\boldsymbol{v}\|^2 \geqslant 0} = \lambda\boldsymbol{v}^T\boldsymbol{v} = \lambda\|\boldsymbol{v}\|^2$. Finally, $\lambda\|\boldsymbol{v}\|^2 \geqslant 0$ implies that $\lambda \geqslant 0$. □

The **pseudoinverse** of an $m \times n$ matrix A is the matrix A^+ such that the system $A\boldsymbol{x} = \boldsymbol{b}$ has "optimal" solution $\boldsymbol{x} = A^+\boldsymbol{b}$.

Here, "optimal" means that $\boldsymbol{x}$ is the smallest least squares solution.
In particular:

- If $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is that solution.
- If $A\boldsymbol{x} = \boldsymbol{b}$ has many solutions, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is the one of smallest norm (the "optimal" one; and there is indeed only one such optimal solution).
- If $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent but has a unique least squares solution, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is that least squares solution.
- If $A\boldsymbol{x} = \boldsymbol{b}$ has many least squares solutions, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is the one with smallest norm.

When there is a unique (least squares) solution, we know how to find the pseudoinverse:

- If A is invertible, then $A^+ = A^{-1}$.
- If A has full column rank, then $A^+ = (A^TA)^{-1}A^T$.
  Recall. If $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, a least squares solution can be determined by solving $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$.
  If A has full column rank (i.e. the columns of A are independent; in this context, the typical case), then $\boldsymbol{x} = (A^TA)^{-1}A^T\boldsymbol{b}$ is the **unique** least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$.

Example 161.

(a) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$?

(b) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$?

(c) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$?

(d) In each case, compute $\Sigma^+\Sigma$ and $\Sigma\Sigma^+$.

Armin Straub
straub@southalabama.edu

Solution.

(a) Recall that, if A has full column rank, then $A^+ = (A^T A)^{-1} A^T$.

Here, $\Sigma^T \Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$, so that $\Sigma^+ = (\Sigma^T \Sigma)^{-1} \Sigma^T = \begin{bmatrix} 1/4 & \\ & 1/9 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}$.

Alternative. Let us think about the optimal solution to $\Sigma \boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

The (unique) least squares solution is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ b_2/3 \end{bmatrix}$. (Review if this is not obvious!)

Since $\begin{bmatrix} b_1/2 \\ b_2/3 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix} \boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}$.

(b) Let us think about the smallest norm ("optimal") solution to $\Sigma \boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

The general solution is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ b_2/3 \\ t \end{bmatrix}$, where t is a free parameter.

Clearly, the smallest norm solution is $\begin{bmatrix} b_1/2 \\ b_2/3 \\ 0 \end{bmatrix}$.

Since $\begin{bmatrix} b_1/2 \\ b_2/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} \boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix}$.

(c) Now, $\Sigma \boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ has no solution (unless $b_2 = 0$).

We therefore need to think about least squares solutions.

The general least squares solution (why?!) is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ s \\ t \end{bmatrix}$, where s, t are free parameters.

Clearly, the smallest norm least squares solution is $\begin{bmatrix} b_1/2 \\ 0 \\ 0 \end{bmatrix}$.

Since $\begin{bmatrix} b_1/2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

(d) Firstly, $\Sigma^+ \Sigma = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\Sigma \Sigma^+ = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Secondly, $\Sigma^+ \Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\Sigma \Sigma^+ = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

[Note how the pseudoinverse tries to behave like the regular inverse. But since Σ has only 2 columns, $\Sigma^+ \Sigma$ and $\Sigma \Sigma^+$ can have rank at most 2 (so cannot be the full 3×3 identity).]

Thirdly, $\Sigma^+ \Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\Sigma \Sigma^+ = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

[Here, Σ has rank 1, so that $\Sigma^+ \Sigma$ and $\Sigma \Sigma^+$ can have rank at most 1.]

In general. Proceeding, as in this example, we find that the pseudoinverse of any $m \times n$ diagonal matrix Σ is the $n \times m$ (transposed dimensions!) diagonal matrix whose nonzero entries are the inverses of the entries of Σ.

Comment. Observe that, in all three cases, $\Sigma^{++} = \Sigma$.

Comment. Note that $\begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}^+ = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon^{-1} \end{bmatrix}$ for small $\varepsilon \neq 0$, while $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. This shows that the pseudoinverse is not a continuous operation.

It turns out that the pseudoinverse A^+ can be easily obtained from the SVD of A:

Theorem 162. The **pseudoinverse** of an $m\times n$ matrix A with SVD $A=U\Sigma V^T$ is

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

where Σ^+, the pseudoinverse of Σ, is the $n\times m$ diagonal matrix, whose nonzero entries are the inverses of the entries of Σ.

Proof. The equation $A\boldsymbol{x}=\boldsymbol{b}$ is equivalent to $U\Sigma V^T\boldsymbol{x}=\boldsymbol{b}$ and, thus, $\Sigma V^T\boldsymbol{x}=U^T\boldsymbol{b}$.
Write $\boldsymbol{y}=V^T\boldsymbol{x}$ and note that $\boldsymbol{y}$ and $\boldsymbol{x}$ have the same norm (why?!).
We already know that the equation $\Sigma\boldsymbol{y}=U^T\boldsymbol{b}$ has optimal solution $\boldsymbol{y}=\Sigma^+U^T\boldsymbol{b}$.
Since $\boldsymbol{y}$ and $\boldsymbol{x}$ have the same norm, it follows that $\boldsymbol{x}=V\boldsymbol{y}=V\Sigma^+U^T\boldsymbol{b}$ is the optimal solution to $A\boldsymbol{x}=\boldsymbol{b}$.
Hence, $A^+=V\Sigma^+U^T$. □

Lemma 163. The pseudoinverse of A^+ is $A^{++}=A$.

Proof. Starting with the SVD $A=U\Sigma V^T$, we have $A^+=V\Sigma^+U^T$, which is the SVD of A^+.
Therefore, $A^{++}=U\Sigma^{++}V^T$. The claim thus follows from $\Sigma^{++}=\Sigma$. □

Example 164. Determine the pseudoinverse of $A=\begin{bmatrix}1&-1\\0&1\\1&0\end{bmatrix}$ in two ways.

First, using the SVD and, second, using the fact that A has full column rank.

Solution. (SVD) We have computed the SVD of this matrix before.

Since $A=U\Sigma V^T$ with $U=\begin{bmatrix}-2/\sqrt{6}&0&-1/\sqrt{3}\\1/\sqrt{6}&1/\sqrt{2}&-1/\sqrt{3}\\-1/\sqrt{6}&1/\sqrt{2}&1/\sqrt{3}\end{bmatrix}$, $\Sigma=\begin{bmatrix}\sqrt{3}&0\\0&1\\0&0\end{bmatrix}$, $V=\frac{1}{\sqrt{2}}\begin{bmatrix}-1&1\\1&1\end{bmatrix}$,

the pseudoinverse is $A^+=V\Sigma^+U^T$ where $\Sigma^+=\begin{bmatrix}1/\sqrt{3}&0&0\\0&1&0\end{bmatrix}$.

Multiplying these matrices, $A^+=\frac{1}{3}\begin{bmatrix}1&1&2\\-1&2&1\end{bmatrix}$.

Comment. For many applications, it may be neither necessary nor helpful to multiply V,Σ^+,U^T.

Solution. (full column rank) Since A clearly has full column rank, we also have $A^+=(A^TA)^{-1}A^T$.

Indeed, $A^+=(A^TA)^{-1}A^T=\begin{bmatrix}2&-1\\-1&2\end{bmatrix}^{-1}\begin{bmatrix}1&0&1\\-1&1&0\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2&1\\1&2\end{bmatrix}\begin{bmatrix}1&0&1\\-1&1&0\end{bmatrix}=\frac{1}{3}\begin{bmatrix}1&1&2\\-1&2&1\end{bmatrix}$.

Example 165. What is the pseudoinverse of $A=\begin{bmatrix}2&2\\1&1\end{bmatrix}$?

Solution. Recall (or compute) that $A=U\Sigma V^T$ with $U=\frac{1}{\sqrt{5}}\begin{bmatrix}2&-1\\1&2\end{bmatrix}$, $\Sigma=\begin{bmatrix}\sqrt{10}&\\&0\end{bmatrix}$, $V=\frac{1}{\sqrt{2}}\begin{bmatrix}1&-1\\1&1\end{bmatrix}$.
Hence, $A^+=V\Sigma^+U^T$ where $\Sigma^+=\begin{bmatrix}1/\sqrt{10}&0\\0&0\end{bmatrix}$.

Multiplying these matrices (which may not be necessary or helpful for applications), $A^+=\frac{1}{10}\begin{bmatrix}2&1\\2&1\end{bmatrix}$.

Note. Since A does not have full column rank, $A^+=(A^TA)^{-1}A^T$ cannot be used. That's because A^TA is not invertible.

Comment. Here, $A^+A=\boldsymbol{v}_1\boldsymbol{v}_1^T=\frac{1}{2}\begin{bmatrix}1&1\\1&1\end{bmatrix}$ and $AA^+=\boldsymbol{u}_1\boldsymbol{u}_1^T=\frac{1}{5}\begin{bmatrix}4&2\\2&1\end{bmatrix}$ are not visually like the identity. However, note that these are the (orthogonal) projections onto $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$ respectively (in particular, the eigenvalues are $1,0$).

Review.

- If the $m\times n$ matrix A has SVD $A=U\Sigma V^T$, then its pseudoinverse is $A^+=V\Sigma^+U^T$.

  Here, Σ^+, the pseudoinverse of Σ, is the $n\times m$ diagonal matrix, whose nonzero entries are the inverses of the entries of Σ.

- The system $A\boldsymbol{x}=\boldsymbol{b}$ has "optimal" solution $\boldsymbol{x}=A^+\boldsymbol{b}$.

  Here, "optimal" means that $\boldsymbol{x}$ is the smallest least squares solution.

Example 166.

(a) Find the pseudoinverse of $A=\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$.

(b) Find the smallest solution to $x_1+2x_2+3x_3=6$.

As before, smallest solutions means the solution $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|$ is as small as possible. One obvious solution is $[\,1,1,1\,]^T$, but is it the smallest?

Solution.

(a) $A^TA=\begin{bmatrix} 1\\2\\3 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}=\begin{bmatrix} 1&2&3\\2&4&6\\3&6&9 \end{bmatrix}$ has 14-eigenvector $\begin{bmatrix} 1\\2\\3 \end{bmatrix}$ and 0-eigenvectors $\begin{bmatrix} -2\\1\\0 \end{bmatrix},\begin{bmatrix} -3\\0\\1 \end{bmatrix}$.

$$\boldsymbol{u}_1=\frac{1}{\sigma_1}A\boldsymbol{v}_1=\frac{1}{\sqrt{14}}\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\frac{1}{\sqrt{14}}\begin{bmatrix} 1\\2\\3 \end{bmatrix}=1$$

Hence, $A=U\Sigma V^T$ with $U=[\,1\,]$, $\Sigma=\begin{bmatrix} \sqrt{14} & 0 & 0 \end{bmatrix}$, $V=\begin{bmatrix} 1/\sqrt{14} & * & *\\ 2/\sqrt{14} & * & *\\ 3/\sqrt{14} & * & * \end{bmatrix}$.

$$A^+=V\Sigma^+U^T=\begin{bmatrix} 1/\sqrt{14} & * & *\\ 2/\sqrt{14} & * & *\\ 3/\sqrt{14} & * & * \end{bmatrix}\begin{bmatrix} 1/\sqrt{14}\\ 0\\ 0 \end{bmatrix}[\,1\,]=\frac{1}{14}\begin{bmatrix} 1\\2\\3 \end{bmatrix}$$

Comment. No surprise on U. The only options for U are $U=[\,1\,]$ and $U=[\,-1\,]$.

Comment. Realizing what we did here allows us to write down A^+ immediately for all $1\times n$ matrices A. See Example 167.

Homework. Complete the SVD of A. That is, find an option for the two missing columns of V, so that V is an orthogonal matrix. In other words, find an orthonormal basis for the 0-eigenspace.

Comment. An even better approach would be to compute AA^T first (instead of A^TA) which would allow us to compute U first (rather than V first). Can you fill in the blanks?

(b) We are solving $A\boldsymbol{x}=[\,6\,]$ with $A=\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ as in the previous example.

We conclude that the smallest solution is $\boldsymbol{x}=A^+[\,6\,]=\frac{3}{7}\begin{bmatrix} 1\\2\\3 \end{bmatrix}$.

Compare. $\left\|\frac{3}{7}\begin{bmatrix} 1\\2\\3 \end{bmatrix}\right\|=\frac{3}{7}\sqrt{14}\approx 1.604$ is indeed smaller than, say, $\left\|\begin{bmatrix} 1\\1\\1 \end{bmatrix}\right\|=\sqrt{3}\approx 1.732$.

Geometric picture. The equation $x_1+2x_2+3x_3=6$ describes a plane (not through the origin), and we are asking for the point on that plane which is closest to the origin. That's a typical question in Calculus III. Note that $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$ is the normal vector of the plane. Explain why the answer had to be a multiple of that normal vector!

Example 167. More generally, find the pseudoinverse of $A=\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$.

Solution. As in the previous example, we see that the answer will be $A^+=\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|^2}$ with $\boldsymbol{a}=\begin{bmatrix} a_1\\a_2\\a_3 \end{bmatrix}$.

Comment. Likewise for $A=\begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}$.

Armin Straub
straub@southalabama.edu

Example 168. How is the rank of A reflected in its singular value decomposition $A = U\Sigma V^T$?

Solution. The rank of A is equal to the number of nonzero singular values.

> **Theorem 169. (matrix approximation lemma)** Suppose A is a $m \times n$ matrix, and we want to approximate A using a matrix B of rank s (smaller than the rank of A).
> Let $A = U\Sigma V^T$ be the SVD of A (with singular values in decreasing order).
> Then, the best such approximation is $B = U_s \Sigma_s V_s^T$, where Σ_s is the $s \times s$ diagonal matrix with entries $\sigma_1, \sigma_2, \ldots, \sigma_s$ and U_s, V_s are obtained from U, V by only taking the first s columns.

Comment. Note that, by choosing s small compared to r, we can store an approximation of A using much less data. This approximation will be good if the omitted singular values $\sigma_{s+1}, \sigma_{s+2}, \ldots, \sigma_r$ are all "small".

Comment. Equivalently, $B = U\Sigma_s V^T$, where Σ_s is now obtained from Σ by setting all but the largest s singular values to 0. In other words, Σ_s has the values $\sigma_1, \sigma_2, \ldots, \sigma_s$ on its diagonal, followed by zeros.

In other words. Here is another common way to say the same thing:

- Observe that $A = U\Sigma V^T$ is equivalent to $A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$.
- Each matrix $\boldsymbol{u}_i \boldsymbol{v}_i^T$ has rank 1.
- The best rank s approximation to A is $B = \sum_{i=1}^{s} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$.

Advanced comment. Here, "best" approximation is measured using the Frobenius norm of a matrix A (which is the same as the norm of a vector with all the entries of A).

Example 170. Determine the best rank 1 approximation of $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}$.

Solution. We determine (do it!) that A has the SVD

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

Hence, the best rank 1 approximation of A is (that is, we keep 1 singular value only) is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Comment. Equivalently, $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

Example 171. Determine the best rank 1 approximation of $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Solution. Recall that $A = U\Sigma V^T$ with $U = \begin{bmatrix} -2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $V = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

Hence, the best rank 1 approximation of A is $\frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}^T = \frac{1}{2} \begin{bmatrix} 2 & -2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$.

Example 172. (image compression) Let us load a 341x512 grayscale photo and store it as a matrix A. Each entry of the matrix is a value between 0 (black) and 1 (white).

The beautiful picture is taken from: http://www.southalabama.edu/departments/publicrelations/brand/photography.html

[The same approach works with color pictures. These are often represented by three matrices: one for the red component of the pixel, one for the green and for the blue component (RGB color scheme).]

```
Sage] import pylab
```

```
Sage] A = matrix(pylab.imread('/home/armin/photo.png'))
```

```
Sage] A.dimensions()
```

$(341, 512)$

```
Sage] A[0,0]
```

0.137254908681

```
Sage] matrix_plot(A, cmap='gray')
```

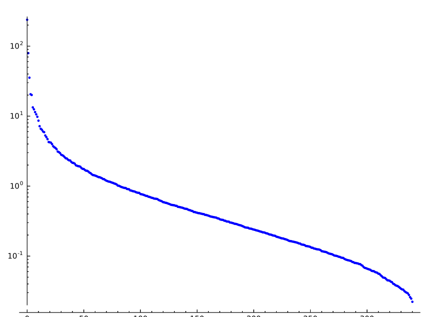
Next, we compute the SVD of A. Despite the size of A that takes the computer only a fraction of a second:

```
Sage] U,S,V = A.SVD()
```

```
Sage] S.diagonal()[:6]
```

$[238.443435709, 79.4429775448, 35.4540786319, 20.5662302846, 20.0697710337, 13.3421216529]$

```
Sage] list_plot(S.diagonal(), scale='semilogy')
```



Armin Straub
straub@southalabama.edu

As we can see, the magnitude of the singular values drops off quickly. We get a good approximation to A (our original photo) by computing a best rank s approximation to A by computing $U_s \Sigma_s V_s$ where Σ_s is the $s \times s$ diagonal matrix with entries $\sigma_1, \sigma_2, \ldots, \sigma_s$ and U_s, V_s are obtained from the corresponding matrices in the SVD $A = U \Sigma V^T$ by only taking the first s columns.

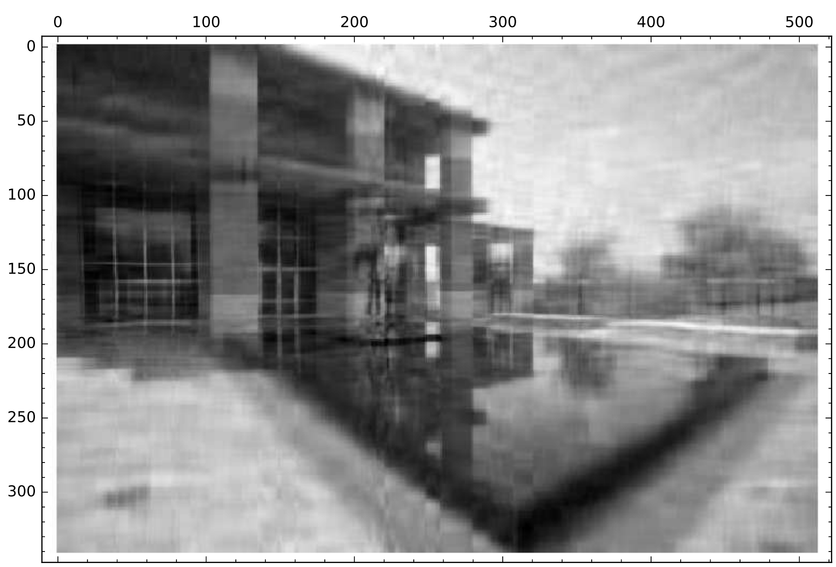
```
Sage] def A_approx(s):
          U0 = U.matrix_from_columns([0..s-1])
          S0 = diagonal_matrix(S.diagonal()[:s])
          V0 = V.matrix_from_columns([0..s-1])
          return U0*S0*V0.transpose()
```

Taking only 100 of the 341 singular values, we get an approximation, which is almost as good as the original:

```
Sage] matrix_plot(A_approx(100), cmap='gray')
```

But notice the development of artifacts. Taking only 20 of the 341 singular values, a lot is lost:

```
Sage] matrix_plot(A_approx(20), cmap='gray')
```



Comment. Image compression is just one (nice visual) example of the power of SVD. A variation of this approach can, for instance, also be used for image denoising. Much more generally, the SVD is able to extract the most important features of any sort of data!

Review. matrix approximation and compression

Function spaces

Recall the following:

- We call objects **vectors** if they can be added and scaled (subject to the usual laws).
- A set of vectors is a **vector space** if it is closed under addition and scaling.

  In other words, vector spaces are spans.

We will now discuss spaces of vectors, where the vectors are functions.

Why? Just one example why it is super useful to apply our linear algebra machinery to functions: we discussed the **distance** between vectors and how to find vectors closest to interesting subspaces (i.e. orthogonal projections). These notions are important for functions, too. For instance, given a (complicated) function, we want to find the closest function in a subspace of (simple) functions. In other words, we want to approximate functions using other (typically, simpler) functions.

Comment. Functions $f(x)$ and $g(x)$ can also be multiplied. This is an extra structure (it makes appropriate sets of functions an **algebra**, which is something more special than a **vector space**), which we ignore during our discussion of vector spaces.

An inner product on function spaces

On the space of, say, (piecewise) continuous functions $f\colon [a, b] \to \mathbb{R}$, it is natural to consider the dot product

$$\langle f, g \rangle = \int_a^b f(t) g(t) \mathrm{d}t.$$

Why? A (sensible) dot product provides a (sensible) notion of distance between functions. The dot product above is the continuous analog of the usual dot product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{t=1}^n x_t y_t$ for vectors in $\mathbb{R}^n$. Do you see it?!

As a consequence, once we have the dot product, we can orthogonally project functions onto spaces of simple functions. In other words, we can compute best approximations of functions by simple functions (for instance, best quadratic approximations).

Why continuous? We need that any product $f(x) g(x)$ is integrable. That means we cannot work with all functions. Continuity is certainly sufficient. In fact, the right condition is that $f(x)^2$ should be integrable on $[a, b]$ (i.e. $f(x)$ is square-integrable). Such a function is said to be in $\mathcal{L}^2[a, b]$.

Example 173. What is the orthogonal projection of $f\colon [a, b] \to \mathbb{R}$ onto the space of constant functions (that is, $\mathrm{span}\{1\}$)?

Solution. The orthogonal projection of $f\colon [a, b] \to \mathbb{R}$ onto $\mathrm{span}\{1\}$ is

$$\frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} 1 = \frac{\int_a^b f(t) 1 \mathrm{d}t}{\int_a^b 1^2 \mathrm{d}t} = \frac{1}{b - a} \int_a^b f(t) \mathrm{d}t.$$

This is the average of $f(x)$ on $[a, b]$.

Comment. Makes perfect sense, doesn't it? Intuitively, the best approximation of a function by a constant should indeed be the one where the constant is the average.

Armin Straub
straub@southalabama.edu

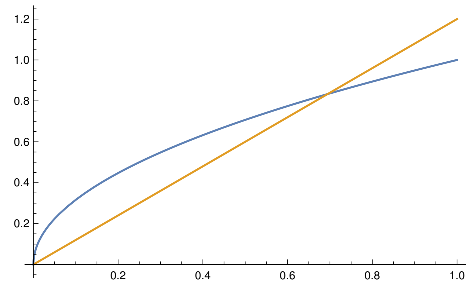
Example 174. Find the best approximation of $f(x) = \sqrt{x}$ on the interval $[0, 1]$ using a function of the form $y = a x$.

Solution. The orthogonal projection of $f: [0, 1] \rightarrow \mathbb{R}$ onto $\operatorname{span}\{x\}$ is

$$\frac{\langle f, x \rangle}{\langle x, x \rangle} x = \frac{\int_0^1 f(t) t \mathrm{d}t}{\int_0^1 t^2 \mathrm{d}t} x = 3x \int_0^1 t f(t) \mathrm{d}t.$$

In our case, the best approximation is

$$3x \int_0^1 t\sqrt{t} \mathrm{d}t = 3x \int_0^1 t^{3/2} \mathrm{d}t = 3x \left[\frac{1}{5/2} t^{5/2}\right]_0^1 = \frac{6}{5} x.$$



Armin Straub
straub@southalabama.edu