Singular value decomposition

(Singular value decomposition)

Every $m \times n$ matrix A can be decomposed as $A = U\Sigma V^T$, where

- Σ is a (rectangular) diagonal matrix with nonnegative entries, $(m \times n)$

The diagonal entries σ_i are called the **singular values** of A.

- U is orthogonal, $(m \times m)$
- V is orthogonal. $(n \times n)$

Comment. If A is symmetric, then the singular value decomposition is already provided by the spectral theorem (the diagonalization of A). Moreover, in that case, $V = U$.

Important observations. If $A = U\Sigma V^T$, then $A^T A = V\Sigma^T \Sigma V^T$.

- Note that $\Sigma^T \Sigma$ is an $n \times n$ diagonal matrix. Its entries are σ_i^2 (the squares of the entries in Σ).
- $A^T A$ is a symmetric matrix! (Why?!) Hence, by the spectral theorem, we are able to find V and $\Sigma^T \Sigma$.

In other words, V is obtained from the (orthonormally chosen) eigenvectors of $A^T A$. Likewise, the entries of $\Sigma^T \Sigma$ are the eigenvalues of $A^T A$; their square roots are the entries of Σ, the singular values.

Finally, the equation $AV = U\Sigma$ allows us to determine U. How?! (Hint: $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$)

This results in the following **recipe** to determine the SVD $A = U\Sigma V^T$ for any matrix A.

Find an orthonormal basis of eigenvectors $\boldsymbol{v}_i$ of $A^T A$. Let λ_i be the eigenvalue of $\boldsymbol{v}_i$.

- V is the matrix with columns $\boldsymbol{v}_i$.
- Σ is the diagonal matrix with entries $\sigma_i = \sqrt{\lambda_i}$.
- U is the matrix with columns $\boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i$. If needed, fill in additional columns to make U orthogonal.

Example 154. Determine the SVD of $A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$.

Solution. $A^T A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ has 8-eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and 2-eigenvector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Since $A^T A = V\Sigma^2 V^T$ (here, $\Sigma^T \Sigma = \Sigma^2$), we conclude that $V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sqrt{8} & \\ & \sqrt{2} \end{bmatrix}$.

From $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$, we find $\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \frac{1}{\sqrt{8}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Likewise, $\boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Hence, $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Check that, indeed, $A = U\Sigma V^T$!

Comment. For applications, it is common to arrange the singular values in decreasing order like we did.

Comment. If we had chosen $V = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$ instead, then $U = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sqrt{8} & \\ & \sqrt{2} \end{bmatrix}$.

As with diagonalization, there are choices! (A lot fewer choices though.) This is another perfectly fine SVD. In fact, it's what Sage computes below.

Armin Straub
straub@southalabama.edu

Sage. Let's have Sage do the work for us. In Sage, the SVD is currently only implemented for floating point numbers. (RDF is the real numbers as floating point numbers with double precision)

```
Sage] A = matrix(RDF, [[2,2],[-1,1]])
Sage] U,S,V = A.SVD()
Sage] U
```

$$\begin{bmatrix} -1.0 & 1.11022302463 \times 10^{-16} \\ 8.64109131471 \times 10^{-17} & 1.0 \end{bmatrix}$$

```
Sage] S
```

$$\begin{bmatrix} 2.82842712475 & 0.0 \\ 0.0 & 1.41421356237 \end{bmatrix}$$

```
Sage] V
```

$$\begin{bmatrix} -0.707106781187 & -0.707106781187 \\ -0.707106781187 & 0.707106781187 \end{bmatrix}$$

Armin Straub
straub@southalabama.edu

Review. SVD

Example 155. Determine the SVD of $A=\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$.

Comment. In contrast to our previous example, $\operatorname{rank}(A)=1$. It follows that A^TA has eigenvalue 0, so that 0 is a singular value of A.

Solution. $A^TA=\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$ has 10-eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and 0-eigenvector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

We conclude that $V=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $\Sigma=\begin{bmatrix} \sqrt{10} & \\ & 0 \end{bmatrix}$.

$\boldsymbol{u}_1=\frac{1}{\sigma_1}A\boldsymbol{v}_1=\frac{1}{\sqrt{10}}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}=\frac{1}{\sqrt{20}}\begin{bmatrix} 4 \\ 2 \end{bmatrix}=\frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

We cannot obtain $\boldsymbol{u}_2$ in the same way because $\sigma_2=0$. Since for every vector $\boldsymbol{u}_2$, $A\boldsymbol{v}_2=\sigma_2\boldsymbol{u}_2$, we can choose $\boldsymbol{u}_2$ as we wish, as long as the columns of U are orthonormal in the end.

$\boldsymbol{u}_2=\frac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ (but $\boldsymbol{u}_2=\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ works just as well)

Hence, $U=\frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$.

In summary, $A=U\Sigma V^T$ with $U=\frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$, $\Sigma=\begin{bmatrix} \sqrt{10} & \\ & 0 \end{bmatrix}$, $V=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

Check. Do check that, indeed, $A=U\Sigma V^T$.

Example 156. Determine the SVD of $A=\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Solution. $A^TA=\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ has 3-eigenvector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and 1-eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Since $A^TA=V\Sigma^T\Sigma V^T$, we conclude that $V=\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\Sigma=\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$.

$\boldsymbol{u}_1=\frac{1}{\sigma_1}A\boldsymbol{v}_1=\frac{1}{\sqrt{3}}\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}=\frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$

$\boldsymbol{u}_2=\frac{1}{\sigma_2}A\boldsymbol{v}_2=\frac{1}{1}\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$\boldsymbol{u}_3$ is chosen so that the matrix U is orthogonal. Hence, $\boldsymbol{u}_3=\frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ (or $\boldsymbol{u}_3=\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$).

Hence, $U=\begin{bmatrix} -2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$.

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

How did we find $\boldsymbol{u}_3$? We already have the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, and need a vector orthogonal to both.

That is, we need to find the vector spanning $\operatorname{span}\left\{\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix},\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right\}^{\perp}=\operatorname{col}\left(\begin{bmatrix} -2 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}\right)^{\perp}=\operatorname{null}\left(\begin{bmatrix} -2 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}\right)$.

[Without the intermediate steps, can you see why the null space consists of precisely the vectors orthogonal to both $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$?]

More generally, proceeding like this, we can always fill in "missing" vectors $\boldsymbol{u}_i$ to obtain an orthonormal basis $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$ that we can use as the columns of U.

Example 157. Determine the SVD of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Solution. $A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ has characteristic polynomial $(1-\lambda)(2-\lambda) - 1 = \lambda^2 - 3\lambda + 1$.

The eigenvalues of $A^T A$ are $\lambda_{1,2} = \frac{3 \pm \sqrt{5}}{2}$.

$\frac{3+\sqrt{5}}{2}$-eigenvector $\begin{bmatrix} 2 \\ 1+\sqrt{5} \end{bmatrix}$ and $\frac{3-\sqrt{5}}{2}$-eigenvector $\begin{bmatrix} 2 \\ 1-\sqrt{5} \end{bmatrix}$.

It would be rather painful to continue with exact expressions, and that is not how applications typically proceed. Numerically:

- 2.618-eigenvector $\begin{bmatrix} 0.526 \\ 0.851 \end{bmatrix}$ and 0.382-eigenvector $\begin{bmatrix} -0.851 \\ 0.526 \end{bmatrix}$. These eigenvectors are normalized, and it is now actually immediately obvious that they are orthogonal. (Of course, they had to be!)

- Hence, $\Sigma = \begin{bmatrix} \sqrt{2.618} & \\ & \sqrt{0.382} \end{bmatrix} = \begin{bmatrix} 1.618 & \\ & 0.618 \end{bmatrix}$ and $V = \begin{bmatrix} 0.526 & -0.851 \\ 0.851 & 0.526 \end{bmatrix}$.

  [We chose $\begin{bmatrix} -0.851 \\ 0.526 \end{bmatrix}$ instead of $\begin{bmatrix} 0.851 \\ -0.526 \end{bmatrix}$, so that, for the resulting V, $\det V = +1$.]

- $\boldsymbol{u}_1 = \frac{1}{\sigma_1} A \boldsymbol{v}_1 = \frac{1}{1.618} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.526 \\ 0.851 \end{bmatrix} = \begin{bmatrix} 0.851 \\ 0.526 \end{bmatrix}$

  $\boldsymbol{u}_2 = \frac{1}{\sigma_2} A \boldsymbol{v}_1 = \frac{1}{0.618} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -0.851 \\ 0.526 \end{bmatrix} = \begin{bmatrix} -0.526 \\ 0.851 \end{bmatrix}$.

  Hence, $U = \begin{bmatrix} 0.851 & -0.526 \\ 0.526 & 0.851 \end{bmatrix}$. (Again, notice the obvious orthogonality!)

Comment. The matrix A itself has eigenvalues $1, 1$, but the 1-eigenspace is only 1-dimensional. We are missing an eigenvector, which renders A not diagonalizable.

Comment. If we had continued symbolically, there are some magical simplifications like $\sqrt{\frac{3+\sqrt{5}}{2}} = \frac{1+\sqrt{5}}{2}$ going on. By the way, this is the golden ratio!

Sage. In Sage, the SVD is currently only implemented for floating point numbers (`RDF` is the real numbers as floating point numbers with double precision). Here's our computation:

```
Sage] A = matrix(RDF, [[1,1],[0,1]])
Sage] U,S,V = A.SVD()
Sage] U
```

$$\begin{bmatrix} 0.850650808352 & -0.525731112119 \\ 0.525731112119 & 0.850650808352 \end{bmatrix}$$

```
Sage] S
```

$$\begin{bmatrix} 1.61803398875 & 0.0 \\ 0.0 & 0.61803398875 \end{bmatrix}$$

```
Sage] V
```

$$\begin{bmatrix} 0.525731112119 & -0.850650808352 \\ 0.850650808352 & 0.525731112119 \end{bmatrix}$$

Example 158. (continued) The matrices U and V are rotation matrices. By what angle?

Why rotations? Recall that orthogonal matrices have determinant $+1$ or -1.
Since $\det U = +1$ and $\det V = +1$, the orthogonal matrices U, V are rotations.

Solution. Being rotation matrices, each of them equals $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for some angle θ.

To find the angle θ_V for V, we compute $\arccos(0.526) = 1.017$. This means that $\theta_V = 1.017$ or $\theta_V = 2\pi - 1.017$ (make a sketch of $\cos(\theta)$ if that's unclear!). Since $\sin(1.017) = 0.851$ (whereas $\sin(2\pi - 1.017) = -0.851$), we conclude that V is a rotation by $\theta_V = 1.017 = 58.3°$. Keep that angle in mind for the next example!
Likewise, U is a rotation by $\theta_U = 0.554 = 31.7°$.

Comment. The two angles add up to $90°$. That's a consequence of the (atypical) fact that the matrices U and V have essentially the same entries.

Example 159. Explain the geometric meaning of the SVD in the previous example.

- The map $\boldsymbol{x} \mapsto A\boldsymbol{x}$ with $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ sends the (orthogonal) grid spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to the (nonorthogonal) grid spanned by $A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

  Make a sketch! The two grids are overlayed in the first plot on the next page.
- Likewise, for instance, (orthogonal) grid spanned by $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ (this is the $45°$ degree rotated version of the previous grid) is sent to the (again, nonorthogonal) grid spanned by $\frac{1}{\sqrt{2}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
  Make a sketch! The two grids are overlayed in the second plot on the next page.
- Can we find an orthogonal grid which is sent to another orthogonal grid by A?
  Solution. Yes! The SVD $A = U\Sigma V^T$ is equivalent to $AV = U\Sigma$. That is, $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$.
  In other words, the orthogonal grid spanned by $\boldsymbol{v}_1, \boldsymbol{v}_2$ is sent to the orthogonal grid spanned by $\sigma_1\boldsymbol{u}_1, \sigma_2\boldsymbol{u}_2$.
  As we observed earlier, the grid spanned by $\boldsymbol{v}_1, \boldsymbol{v}_2$ is the $58.3°$ degree rotated version of the standard grid)
  While the input grid consists of little squares, the output grid consists of rectangles with sides σ_1, σ_2.
  Make a sketch! The two grids are overlayed in the third plot on the next page.

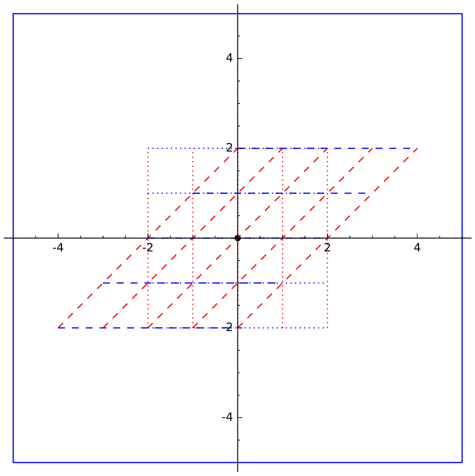
The following Sage code prepares the plots on the next page. Even if you have no coding background, can you see, roughly, what is happening?

```
Sage] def grid_lines(v1, v2, n, args={}):
          lines = Graphics()
          for i in [-n..n]:
              lines += line([i*v1-n*v2, i*v1+n*v2], color='red', **args)
              lines += line([i*v2-n*v1, i*v2+n*v1], color='blue', **args)
          return lines
```

```
Sage] def svd_rotate(angle):
          A = matrix([[1,1],[0,1]])
          t = angle*2*pi/360
          R = matrix([[cos(t),-sin(t)],[sin(t),cos(t)]])
          G1 = grid_lines(R*vector([1,0]), R*vector([0,1]), 2, {'linestyle':':'})
          G2 = grid_lines(A*R*vector([1,0]), A*R*vector([0,1]), 2, {'linestyle':'--'})
          B = polygon([(-5,-5), (-5,5), (5,5), (5,-5)], fill=False)
          O = point((0,0), pointsize=30,color='black')
          return B+O+G1+G2
```

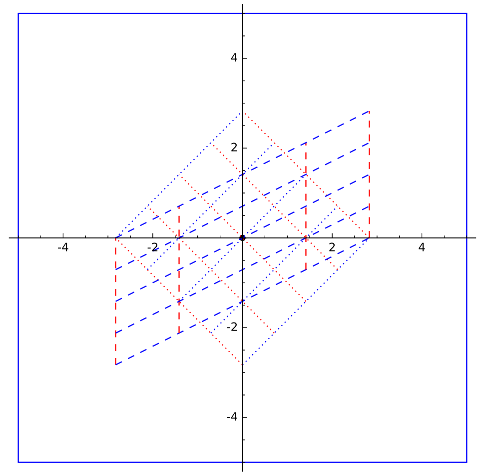
Grid spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (dotted), and grid spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (dashed):

```
Sage] svd_rotate(angle = 0)
```



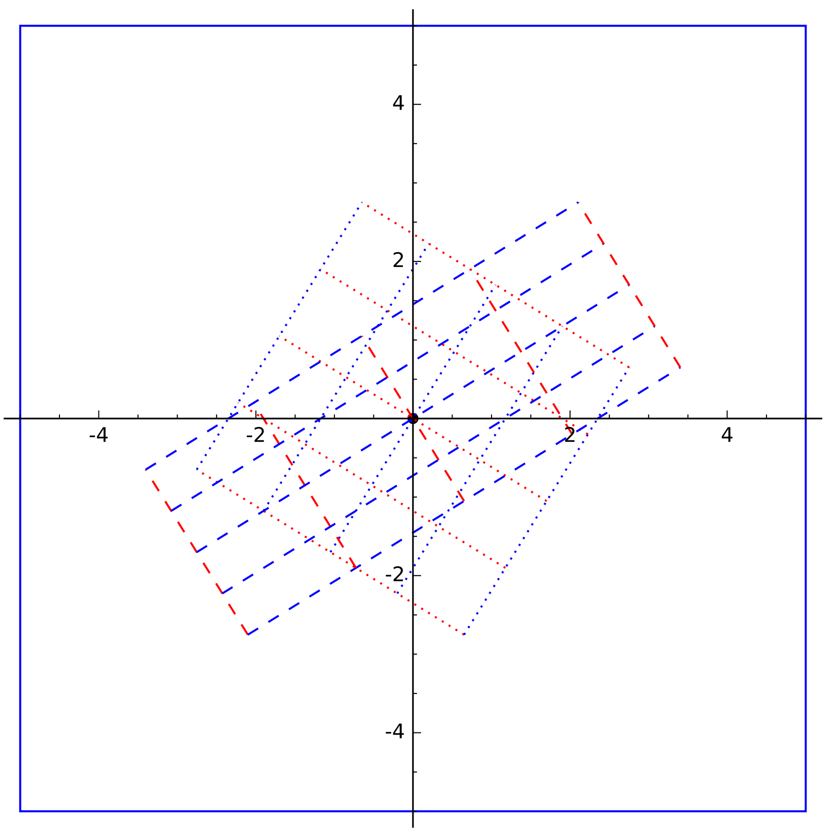
Grid spanned by $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ (dotted), and grid spanned by $\frac{1}{\sqrt{2}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (dashed):

```
Sage] svd_rotate(angle = 45)
```



Finally, here is the special situation (given by the SVD!) which shows an orthogonal grid (rotated by $58.3°$ degree) that is sent to another orthogonal grid (rotated by $31.7°$ degree):

```
Sage] svd_rotate(angle = 58.3)
```



For more pictures and detailed comments see the beautiful article:
http://www.ams.org/samplings/feature-column/fcarc-svd

Armin Straub
straub@southalabama.edu