Understanding Matrices
Eigensystems

• **Eigenvectors** - special directions $\mathbf{v}_k$ in which a matrix only applies scaling

• **Eigenvalues** - the amount $\lambda_k$ of that scaling

• **Right Eigenvectors** (or simply eigenvectors) satisfy $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$
  - Eigenvectors represent directions, so $A(\alpha \mathbf{v}_k) = \lambda_k (\alpha \mathbf{v}_k)$ is also true for all $\alpha$

• **Left Eigenvectors** satisfy $\mathbf{u}_k^T A = \lambda_k \mathbf{u}_k^T$ (or $A^T \mathbf{u}_k = \lambda_k \mathbf{u}_k$)

• Diagonal matrices have eigenvalues on the diagonal, and eigenvectors $\mathbf{e}_k$

\[
\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\]

• Upper/lower triangular matrices also have eigenvalues on the diagonal

\[
\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}
\]
Complex Numbers

• Complex numbers may appear in both eigenvalues and eigenvectors
  \[
  \begin{pmatrix}
  0 & 1 \\
  -1 & 0 \\
  \end{pmatrix}
  \begin{pmatrix}
  1 \\
  i \\
  \end{pmatrix}
  =
  i
  \begin{pmatrix}
  1 \\
  i \\
  \end{pmatrix}
  \]

• Recall: complex conjugate: \((a + bi)^* = a - bi\)

• Hermitian Matrix: \(A^* = A\) (often, \(A^*\) is written as \(A^H\))
  • \(Av = \lambda v\) implies \((Av)^* = (\lambda v)^*\) or \(v^* A = \lambda^* v^*\)
  • Using this, \(Av = \lambda v\) implies \(v^* A v = v^* \lambda v\) or \(\lambda^* v^* v = \lambda v^* v\) or \(\lambda^* = \lambda\)
  • Thus, Hermitian matrices have \(\lambda \in R\) (no complex eigenvalues)

• Symmetric real-valued matrices have real-valued eigenvalues/eigenvectors
  • However, complex eigenvectors work too, e.g. \(A(\alpha v_k) = \lambda_k (\alpha v_k)\) with \(\alpha\) complex
Spatial Deformation

- Let \( c = \sum_k \alpha_k v_k \), so that \( Ac = \sum_k \alpha_k Av_k = \sum_k (\alpha_k \lambda_k) v_k \)

- Then, \( A \) tilts \( c \) away from directions with smaller eigenvalues and towards directions with larger eigenvalues

- Large \( \lambda_k \) stretch components in their associated \( v_k \) directions
- Small \( \lambda_k \) squish components in their associated \( v_k \) directions
- Negative \( \lambda_k \) flip the sign of components in their associated \( v_k \) directions
Spatial Deformation

• Consider every point on the unit circle (green) as a vector $c = \sum_k \alpha_k v_k$, and remap each point via $Ac = \sum_k (\alpha_k \lambda_k) v_k$

• The remapped shape (blue) is more elliptical than the original circle (green)
• The circle is stretched along the axis with the larger eigenvalue, and compressed along the axis with the smaller eigenvalue
• The larger the ratio of eigenvalues, the more elliptical the circle becomes
• This is true for all circles (representing all points in the plane)
Solving Linear Systems

- Perturb the right hand side from $b$ to $\hat{b}$, and solve $A\hat{c} = \hat{b}$ to find $\hat{c}$
- Note how $c$ and $\hat{c}$ are more separated than $b$ and $\hat{b}$, i.e. the solution is more perturbed than the right hand side is

- Small changes in the right hand side lead to larger changes in the solution
- **Small algorithmic errors are also amplified**, since they change $A^{-1}b$ to $\hat{A}^{-1}b$ which is similar to changing $A^{-1}b$ to $A^{-1}\hat{b}$
- The amount of amplification is proportional to the ratio of the eigenvalues
Preconditioning

• Suppose $A$ has large eigenvalue ratios that make $Ac = b$ difficult to solve
• Suppose one had an approximate guess for the inverse, i.e. an $\hat{A}^{-1} \approx A^{-1}$
• Then, transform $Ac = b$ into $\hat{A}^{-1}Ac = \hat{A}^{-1}b$ or $\hat{I}c = \tilde{b}$
  • Typically, a bit more involved than this, but conceptually the same
• $\hat{I}$ is not the identity, so there is still more work to do in order to find $c$
• However, $\hat{I}$ has eigenvalues with similar magnitudes (clusters work too), making $\hat{I}c = \tilde{b}$ far easier to solve than a poorly conditioned $Ac = b$

Preconditioning works GREAT!

• It is best to re-scale ellipsoids along eigenvector axes, but scaling along the coordinate axes (diagonal/Jacobi preconditioning) can work well too
Rectangular Matrices (Rank)

- An $m \times n$ rectangular matrix has $m$ rows and $n$ columns
- (Note: these comments also hold for square matrices with $m = n$)

- The columns span a space, and the unknowns are weights on each column (recall $Ac = \sum_k c_k a_k$)
- A matrix with $n$ columns has maximum rank $n$
- The actual rank depends on how many of the columns are linearly independent from one another

- Each column has length $m$ (which is the number of rows)
- Thus, the columns live in an $m$ dimensional space, and at best can span that whole space
- That is, there is a maximum of $m$ independent columns (that could exist)

- Overall, a matrix at most has rank equal to the minimum of $m$ and $n$
- Both considerations are based on looking at the columns (which are scaled by the unknowns)
Rows vs. Columns

• One can find discussions on rows, row spaces, etc. that are used for various purposes

• Although these are fine discussions in regards to matrices/mathematics, they are unnecessary for an intuitive understanding of high dimensional vector spaces (and as such can be ignored)

• The number of columns is identical to number of variables, which depends on the parameters of the problem
  • E.g. the unknown parameters that govern a neural network architecture

• The number of rows depends on the amount of data used, and adding/removing data does not intrinsically effect the nature of the problem
  • E.g. it does not change the network architecture, but merely perturbs the ascertained values of the unknown parameters
Singular Value Decomposition (SVD)

- Factorization of any size \(mxn\) matrix: \(A = U\Sigma V^T\)
- \(\Sigma\) is \(mxn\) diagonal with non-negative diagonal entries (called singular values)
- \(U\) is \(mxm\) orthogonal, \(V\) is \(nxn\) orthogonal (their columns are called singular vectors)
  - Orthogonal matrices have orthonormal columns (an orthonormal basis), so their transpose is their inverse. They preserve inner products, and thus are rotations, reflections, and combinations thereof
  - If \(A\) has complex entries, then \(U\) and \(V\) are unitary (conjugate transpose is their inverse)

- Introduced and rediscovered many times: Beltrami 1873, Jordan 1875, Sylvester 1889, Autonne 1913, Eckart and Young 1936. Pearson introduced principle component analysis (PCA) in 1901, which uses SVD. Numerical methods by Chan, Businger, Golub, Kahan, etc.
(Rectangular) Diagonal Matrices

• All off-diagonal entries are 0
  • Diagonal entries are $a_{kk}$, and off diagonal entries are $a_{ki}$ with $k \neq i$
  
  E.g. \[
  \begin{pmatrix}
  5 & 0 \\
  0 & 2
  \end{pmatrix}
  \begin{pmatrix}
  c_1 \\
  c_2
  \end{pmatrix}
  =
  \begin{pmatrix}
  10 \\
  -1
  \end{pmatrix}
  \] has $5c_1 = 10$ and $2c_2 = -1$, so $c_1 = 2$ and $c_2 = -0.5$
  
  • Note that $\alpha \neq 0$ imposes a “no solution” condition (even though $c_1$ and $c_2$ are well-specified)

• E.g. \[
  \begin{pmatrix}
  5 & 0 & 0 \\
  0 & 2 & 0
  \end{pmatrix}
  \begin{pmatrix}
  c_1 \\
  c_2 \\
  c_3
  \end{pmatrix}
  =
  \begin{pmatrix}
  10 \\
  -1
  \end{pmatrix}
  \] has $5c_1 = 10$ and $2c_2 = -1$, so $c_1 = 2$ and $c_2 = -0.5$
  
  • Note that there are “infinite solutions” for $c_3$ (even though $c_1$ and $c_2$ are well-specified)

• A zero on the diagonal indicates a singular system, which has either no solution (e.g. $0c_1 = 10$) or infinite solutions (e.g. $0c_1 = 0$)
Singular Value Decomposition (SVD)

- \( A^T A = V \Sigma^T U^T U \Sigma V^T = V (\Sigma^T \Sigma) V^T, \) so \( (A^T A) v = \lambda v \) gives \( (\Sigma^T \Sigma)(V^T v) = \lambda (V^T v) \)
- \( \Sigma^T \Sigma \) is \( n \times n \) diagonal with eigenvectors \( \hat{e}_k \), so \( \hat{e}_k = V^T v \) and \( v = V \hat{e}_k \)
- That is, the columns of \( V \) are the eigenvectors of \( A^T A \)

- \( A A^T = U \Sigma V^T V \Sigma^T U^T = U (\Sigma \Sigma^T) U^T, \) so \( (AA^T) v = \lambda v \) gives \( (\Sigma \Sigma^T)(U^T v) = \lambda (U^T v) \)
- \( \Sigma \Sigma^T \) is \( m \times m \) diagonal with eigenvectors \( \hat{e}_k \), so \( \hat{e}_k = U^T v \) and \( v = U \hat{e}_k \)
- That is, the columns of \( U \) are the eigenvectors of \( AA^T \)

- When \( m \neq n \), either \( \Sigma^T \Sigma \) or \( \Sigma \Sigma^T \) is larger and contains extra zeros on the diagonal
- Their other diagonal entries are the squares of the singular values
- That is, the singular values are the (non-negative) square roots of the non-extra eigenvalues of \( A^T A \) and \( AA^T \)

- Note that both \( A^T A \) and \( AA^T \) are symmetric positive semi-definite, and thus easy to work with
- E.g. symmetry means their eigensystem (and thus the SVD) has no complex numbers when \( A \) doesn’t
Example (Tall Matrix)

- Consider size 4x3 matrix \( A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \)

- Label the columns \( a_1 = \begin{pmatrix} 1 \\ 4 \\ 7 \\ 10 \end{pmatrix} \), \( a_2 = \begin{pmatrix} 2 \\ 5 \\ 8 \\ 11 \end{pmatrix} \), \( a_3 = \begin{pmatrix} 3 \\ 6 \\ 9 \\ 12 \end{pmatrix} \)

- Since \( a_1 \) and \( a_2 \) point in different directions, \( A \) is at least rank 2

- \( a_3 = 2a_2 - a_1 \), so the third column is in the span of the first two columns

- Thus, \( A \) is only rank 2 (not rank 3)
Example (SVD)

\[ A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} .141 & .825 & -.420 & -.351 \\ .344 & .426 & .298 & .782 \\ .547 & .028 & .644 & -.509 \\ .750 & -.371 & -.542 & .079 \end{pmatrix} \begin{pmatrix} 25.5 & 0 & 0 \\ 0 & 1.29 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .504 & .574 & .644 \\ -.761 & -.057 & .646 \\ .408 & -.816 & .408 \end{pmatrix} \]

- Singular values are 25.5, 1.29, and 0
- Singular value of 0 indicates that the matrix is rank deficient
- The rank of a matrix is equal to its number of nonzero singular values
Derivation from $A^T A$ and $AA^T$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} .141 & .825 & -.420 & -.351 \\ .344 & .426 & .298 & .782 \\ .547 & .028 & .644 & -.509 \\ .750 & -.371 & -.542 & .079 \end{pmatrix} \begin{pmatrix} 25.5 & 0 & 0 \\ 0 & 1.29 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .504 & .574 & .644 \\ -.761 & -.057 & .646 \\ .408 & -.816 & .408 \end{pmatrix}$$

- $A^T A$ is size $3 \times 3$ and has 3 eigenvectors (seen in $V$)
- The square roots of the 3 eigenvalues of $A^T A$ are seen in $\Sigma$ (color coded to the eigenvectors)
- $AA^T$ is size $4 \times 4$ and has 4 eigenvectors (seen in $U$)
- The square roots of 3 of the eigenvalues of $AA^T$ are seen in $\Sigma$
  - The 4th eigenvalue of $AA^T$ is an extra eigenvalue of 0
Understanding $A c$

$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} .141 & .825 & -.420 & -.351 \\ .344 & .426 & .298 & .782 \\ .547 & .028 & .644 & -.509 \\ .750 & -.371 & -.542 & .079 \end{pmatrix} \begin{pmatrix} 25.5 & 0 & 0 \\ 0 & 1.29 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .504 & .574 & .644 \\ -.761 & -.057 & .646 \\ .408 & -.816 & .408 \end{pmatrix}$

- $A$ maps from $\mathbb{R}^3$ to $\mathbb{R}^4$
- $A c$ first projects $c \in \mathbb{R}^3$ onto the 3 basis vectors in $V$
- Then, the associated singular values (diagonally) scale the results
- Lastly, those scaled results are used as weights on the basis vectors in $U$
Understanding $Ac$

$$Ac = \begin{pmatrix}
.141 & .825 & -.420 & -.351 \\
.344 & .426 & .298 & .782 \\
.547 & .028 & .644 & -.509 \\
.750 & -.371 & -.542 & .079
\end{pmatrix} \begin{pmatrix} 25.5 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix}
.504 & .574 & .644 \\
-.761 & -.057 & .646 \\
.408 & -.816 & .408
\end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

$$= \begin{pmatrix}
.141 & .825 & -.420 & -.351 \\
.344 & .426 & .298 & .782 \\
.547 & .028 & .644 & -.509 \\
.750 & -.371 & -.542 & .079
\end{pmatrix} \begin{pmatrix} 25.5 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix}
\sigma_1 v_1^T c \\
\sigma_2 v_2^T c \\
\sigma_3 v_3^T c \\
0
\end{pmatrix}$$

$$= u_1 \sigma_1 v_1^T c + u_2 \sigma_2 v_2^T c + u_3 \sigma_3 v_3^T c + u_4 0$$

* $Ac$ projects $c$ onto the basis vectors in $V$, scales by the associated singular values, and uses those results as weights on the basis vectors in $U$
Extra Dimensions

\[ A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \]

\[
\begin{pmatrix}
.141 & .825 & -.420 & -.51 \\
.344 & .426 & .298 & .782 \\
.547 & .028 & .644 & -.59 \\
.750 & -.371 & -.542 & .09 \\
\end{pmatrix}
\begin{pmatrix}
25.5 & 0 & 0 \\
0 & 1.29 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
.504 & .574 & .644 \\
-.761 & -.057 & .646 \\
.408 & -.816 & .408 \\
\end{pmatrix}
\]

- The 3D space of vector inputs can only span a 3D subspace of \( \mathbb{R}^4 \)
- The last (green) column of \( U \) represents the unreachable dimension, orthogonal to the range of \( A \), and is always multiplied by 0
- One can delete this column and the associated portion of \( \Sigma \) (and still obtain a valid factorization)
Zero Singular Values

The 3\textsuperscript{rd} singular value is 0, so $A$ has a 1D null space that reduces the 3D input vectors to only 2 dimensions.

The associated (pink) terms make no contribution to the final result, and can also be deleted (still obtaining a valid factorization).

The first 2 columns of $U$ span the 2D subset of $R^4$ that comprises the range of $A$. 

\[
A = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
10 & 11 & 12
\end{pmatrix} = \begin{pmatrix}
.141 & .825 & -4.0 & -3.61 \\
.344 & .426 & .238 & .781 \\
.547 & .028 & .644 & -5.9 \\
.750 & -.371 & -.512 & .09
\end{pmatrix} \begin{pmatrix}
.504 & .574 & .644 \\
-.761 & -.057 & .646 \\
.183 & -.623 & .779
\end{pmatrix}
Approximating $A$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \approx \begin{pmatrix} .141 & .815 & -.410 & -.561 \\ .344 & .415 & .273 & .785 \\ .547 & .013 & .646 & -.519 \\ .750 & -.571 & -.542 & .019 \end{pmatrix} \begin{pmatrix} 25.5 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} .504 & .574 & .644 \\ -.981 & .007 & .816 \\ .108 & .092 & .770 \end{pmatrix}$$

- The first singular value is much bigger than the second, and so represents the vast majority of what $A$ does (note, the vectors in $U$ and $V$ are unit length)
- Thus, one could approximate $A$ quite well by only using the terms associated with the largest singular value
- This is not a valid factorization, but an approximation (and the idea behind PCA)
Summary

• The columns of $V$ that do not correspond to “nonzero” singular values form an orthonormal basis for the null space of $A$

• The remaining columns of $V$ form an orthonormal basis for the space perpendicular to the null space of $A$ (parameterizing meaningful inputs)

• The columns of $U$ corresponding to “nonzero” singular values form an orthonormal basis for the range of $A$

• The remaining columns of $U$ form an orthonormal basis for the (unattainable) space perpendicular to the range of $A$

• One can drop the columns of $U$ and $V$ that do not correspond to “nonzero” singular values and still obtain a valid factorization of $A$

• One can drop the columns of $U$ and $V$ that correspond to “small/smaller” singular values and still obtain a reasonable approximation of $A$
Example (Wide Matrix)

\[ A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} = \begin{pmatrix} .504 & -.761 & .408 \\ .574 & -.057 & -.816 \\ .644 & .646 & .408 \end{pmatrix} \begin{pmatrix} 25.5 & 0 & 0 & 0 \\ 0 & 1.29 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .141 & .344 & .547 & .750 \\ .825 & .426 & .028 & -.371 \\ -.420 & .298 & .644 & -.542 \\ -.351 & .782 & -.509 & .079 \end{pmatrix} \]

- \( A \) maps from \( R^4 \) to \( R^3 \) and so has at least a 1D null space (green)
- The 3\(^{rd}\) singular value is 0, and the associated (pink) terms make no contribution to the final result
Example (Wide Matrix)

\[
A = \begin{pmatrix}
1 & 4 & 7 & 10 \\
2 & 5 & 8 & 11 \\
3 & 6 & 9 & 12
\end{pmatrix}
\]

\[
\begin{pmatrix}
.504 & -.761 & .413 \\
.574 & -.057 & -.816 \\
.644 & .646 & .413
\end{pmatrix}
\begin{pmatrix}
25.5 \\
0 \\
0
\end{pmatrix}
\begin{pmatrix}
.141 & .344 & .547 & .750 \\
.825 & .426 & .028 & -.371 \\
-1.29 & .293 & .641 & -.743
\end{pmatrix}
\]

- Only a 2D subspace of \( \mathbb{R}^4 \) matters, with the rest of \( \mathbb{R}^4 \) in the null space of \( A \)
- Only a 2D subspace of \( \mathbb{R}^3 \) is in the range of \( A \)
Notes

• The SVD is often unwieldy for computational purposes
• However, replacing matrices by their SVD can be quite useful/enlightening for theoretical pursuits
• Moreover, its theoretical underpinnings are often used to devise computational algorithms

• The SVD is unique under certain assumptions, such as all $\sigma_k \geq 0$ and in descending order
• However, one can make both a $\sigma_k$ and its associated column in $U$ negative for an "alternate SVD" (see e.g. "Invertible Finite Elements For Robust Simulation of Large Deformation", Irving et al. 2004)
SVD Construction (Important Detail)

• Let \( A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \) so that \( A^T A = A A^T = I \), and thus \( U = V = \Sigma = I \)

• But \( A \neq U \Sigma V^T = I \) What’s wrong?

• Given a column vector \( v_k \) of \( V \), \( A v_k = U \Sigma V^T v_k = U \Sigma \hat{e}_k = U \sigma_k \hat{e}_k = \sigma_k u_k \) where \( u_k \) is the corresponding column of \( U \)

\[
\begin{align*}
A v_1 &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = u_1 \\
A v_2 &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u_2
\end{align*}
\]

• Since \( U \) and \( V \) are orthonormal, their columns are unit length

• However, there are still two choices for the direction of each column

• Multiplying \( u_2 \) by \(-1\) to get \( u_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \) makes \( U = A \), and thus \( A = U \Sigma V^T \) as desired
SVD Construction (Important Detail)

• An orthogonal matrix has determinant equal to ±1, where −1 indicates a reflection of the coordinate system

• If \( \det V = -1 \), flip the direction of any column to make \( \det V = 1 \) (so \( V \) does not contain a reflection)

• Then, for each \( v_k \), compare \( Av_k \) to \( \sigma_k u_k \) and flip the direction of \( u_k \) when necessary in order to make \( Av_k = \sigma_k u_k \)

• \( \det U = \pm 1 \) and may contain a reflection

• When \( \det U = -1 \), one can flip the sign of the smallest singular value in \( \Sigma \) to be negative, whilst also flipping the direction of the corresponding column in \( U \) so that \( \det U = 1 \)

• This embeds the reflection into \( \Sigma \) and is called the polar-SVD (Irving et al. 2004)
Solving Linear Systems

- $Ac = b$ becomes $U\Sigma V^T c = b$ or $\Sigma(V^T c) = (U^T b)$ or $\Sigma\hat{c} = \hat{b}$
- The unknowns $c$ are remapped into the space spanned by $V$, and the right hand side $b$ is remapped into the space spanned by $U$
- Every matrix is a diagonal matrix, when viewed in the right space
- Solve the diagonal system $\Sigma\hat{c} = \hat{b}$ by dividing the entries of $\hat{b}$ by the singular values $\sigma_k$; then, $c = V\hat{c}$
- The SVD transforms the problem into an inherently diagonal space with eigenvectors along the coordinate axes
- Circles becoming ellipses (discussed earlier) is still problematic
  - Eccentricity is caused by ratios of singular values (since $U$ and $V$ are orthogonal matrices)
Condition Number

• The condition number of $A$ is $\frac{\sigma_{\text{max}}}{\sigma_{\text{min}}}$ and measures closeness to being singular.

• For a square matrix, it measures the difficulty in solving $Ac = b$.

• For a rectangular (and square) matrix, it measures how close the columns are to being linearly dependent.
  • For a wide (rectangular) matrix, it ignores the extra columns that are guaranteed to be linearly dependent (which is fine, because the associated variables lack any data).

• The condition number does not depend on the right hand side.

• The condition number is always bigger than 1, and approaches $\infty$ for nearly singular matrices.

• Singular matrices have condition number equal to $\infty$, since $\sigma_{\text{min}} = 0$. 
Singular Matrices

• Diagonalize $Ac = b$ to $\Sigma (V^T c) = (U^T b)$, e.g. $\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}$ with $\hat{c}_1 = \frac{\hat{b}_1}{\sigma_1}$, $\hat{c}_2 = \frac{\hat{b}_2}{\sigma_2}$

• Suppose $\sigma_1 \neq 0$ and $\sigma_2 = 0$; then, there is no unique solution:
  • When $\hat{b}_2 = 0$, there are infinite solutions for $\hat{c}_2$ (but $\hat{c}_1$ is still uniquely determined)
  • When $\hat{b}_2 \neq 0$, there is no solution for $\hat{c}_2$, and $b$ is not in the range of $A$ (but $\hat{c}_1$ is still uniquely determined)

• Consider: $\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix}$ with $\hat{c}_1 = \frac{\hat{b}_1}{\sigma_1}$, $\hat{c}_2 = \frac{\hat{b}_2}{\sigma_2}$
  • When $\hat{b}_3 = 0$, the last row adds no new information (one has extra redundant data)
  • When $\hat{b}_3 \neq 0$, the last row is false and there is no solution (but $\hat{c}_1$ and $\hat{c}_2$ are still uniquely determined)

• Consider: $\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix}$ with $\hat{c}_1 = \frac{\hat{b}_1}{\sigma_1}$, $\hat{c}_2 = \frac{\hat{b}_2}{\sigma_2}$
  • Infinite solutions work for $\hat{c}_3$ (but $\hat{c}_1$ and $\hat{c}_2$ are still uniquely determined)
Understanding Variables

• Consider any column $k$ of $\Sigma$
  • When $\sigma_k \neq 0$, one can state a value for $\hat{c}_k$
  • When $\sigma_k = 0$ or there is no $\sigma_k$, then there is no information in the data for $\hat{c}_k$
    • This does not mean that other parameters cannot be adequately determined!

• Consider a row $i$ of $\Sigma$ that is identically zero
  • When $\hat{b}_i = 0$, this row indicates that there is extra redundant data
  • When $\hat{b}_i \neq 0$, this row indicates that there is conflicting information in the data
  • Conflicting information doesn’t necessarily imply that all is lost, i.e. “no solution”; rather, it might merely mean that the data contains a bit of noise
  • Regardless, in spite of any conflicting information, the determinable $\hat{c}_k$ represent the “best” that one can do
Norms

- Common norms: \( \| c \|_1 = \sum_k |c_k|, \quad \| c \|_2 = \sqrt{\sum_k c_k^2}, \quad \| c \|_\infty = \max_k |c_k| \)

- "All norms are interchangeable" is a **theoretically** valid statement (only)
- In practice, the “worst case scenario” (\( L^\infty \)) and the “average” (\( L^1, L^2 \), etc.) are not interchangeable

- E.g. \((100 \text{ people} \times 98.6^\circ + 1 \text{ person} \times 105^\circ)/(101 \text{ people}) = 98.66^\circ\)
- Their average temperature is 98.66\(^\circ\), but everyone is not “ok”
Matrix Norms

- Define the norm of a matrix \( \|A\| = \max_{c \neq 0} \frac{\|Ac\|}{\|c\|} \), so:
  - \( \|A\|_1 \) is the maximum absolute value column sum
  - \( \|A\|_\infty \) is the maximum absolute value row sum
  - \( \|A\|_2 \) is the square root of the maximum eigenvalue of \( A^T A \), i.e. the maximum singular value of \( A \)

- The condition number for solving (square matrix) \( Ac = b \) is \( \|A\|_2 \|A^{-1}\|_2 \)

- Since \( A^{-1} = V \Sigma^{-1} U^T \) where \( \Sigma^{-1} \) has diagonal entries \( \frac{1}{\sigma_k} \), \( \|A^{-1}\|_2 = \frac{1}{\sigma_{\text{min}}} \)

- Thus, \( \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \)