3. Jacobian Inverse Methods for IK

Game Engineering & XR technologies
Prof. HyeongYeop Kang
siamiz@khu.ac.kr
IIIXR LAB

Singular value decomposition

SVD Introduction

- To understand the calculation of pseudo-inverse matrix, you should understand the singular value decomposition (SVD).

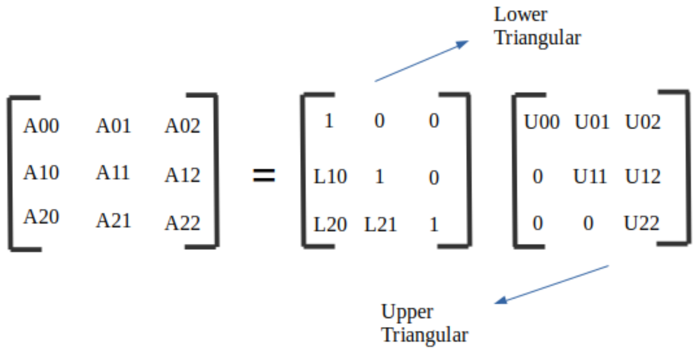
- This is related to the eigen decomposition and polar decomposition.

- In linear algebra, the SVD of a matrix is a factorization of that matrix into three matrices.

- We can think of a matrix A as a transformation that acts on a vector x by multiplication to produce a new vector Ax.
  - We use $[A]_{ij}$ or a_{ij} to denote the element of matrix A at row i and column j.
  - A vector is a quantity which has both magnitude and direction.
  - The general effect of matrix A on the vectors in x is a combination of rotation and stretching.

Matrix factorization (decomposition)?

Matrix factorization is the process of breaking a matrix into multiple parts that multiply together to return the original matrix. There are many ways of doing this, including LU decomposition, Cholesky decomposition, and QR decomposition.



Singular value decomposition

Let's look at the basic first.

Transpose

- The transpose of the column vector is the row vector.
- The element in the i-th row and j-th column of the transposed matrix is equal to the element in the j-th column of the original matrix.
  - For example, $[A^T]_{ij} = [A]_{ji}$.
- The transpose of an $m \times n$ matrix is an $n \times m$ matrix.
  - For example, the transpose of C = $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ is $C^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$.
- The transpose of the transpose of A is A.
  - This means, $(A^T)^T = A$.

Partitioned Matrix

- When calculating the transpose of a matrix, it is usually useful to show it as a partitioned matrix.
  - For example, the matrix B = $\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$ can be written as B = $[u_1 \quad u_2 \quad u_3]$ where $u_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $u_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ $u_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$
  - To write the transpose of B, we can simply turn three column vectors into row vectors: $u_1^T = [1 \quad 2]$ $u_2^T = [3 \quad 4]$ $u_3^T = [5 \quad 6]$. So, $B^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$

Singular value decomposition

Eigenvalues & Eigenvectors

- The only way to change the magnitude of a vector without changing its direction is by multiplying it with a scalar.
  - If we have a vector u, and a scalar quantity λ, λu has the same direction and a different magnitude.
  - The eigenvector of a matrix A is defined as a nonzero vector u such that: $Au = \lambda u$, where λ is called the eigenvalue of A.
  - For example, the eigenvector of B = $\begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix}$ are $u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and their corresponding eigenvalues are λ_1 = -1 *and* λ_2 = -2. (Note that $Bu_1 = \lambda_1 u_1$ and $Bu_2 = \lambda_2 u_2$)

Eigen Decomposition of a matrix (AP = PD)[6]

- The matrix decomposition of a square matrix A into eigenvalues and eigenvectors is a kind of "matrix diagonalization."
- Let's do this. Assume A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ and corresponding linearly independent eigenvectors $X_1, X_2, \ldots, X_k$ which can be denoted as:

$$\begin{array}{ccccc} X_1 & & X_2 & & X_k \\ \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1k} \end{bmatrix}, & & \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2k} \end{bmatrix}, & \ldots & \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kk} \end{bmatrix}. \end{array}$$

Singular value decomposition

Eigen Decomposition of a matrix (AP = PD)

- Then, we can define a matrix P composed of eigenvectors and a diagonal matrix D composed of eigenvalues:

$$\mathrm{P} \equiv [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k] = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix}$$

$$\mathrm{D} \equiv \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix},$$

- Consequently, we can write AP = PD as follows:

$$\mathrm{AP} = \mathrm{A}[\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k] = \underbrace{\boxed{[\mathrm{A}\mathbf{X}_1 \quad \mathrm{A}\mathbf{X}_2 \quad \cdots \quad \mathrm{A}\mathbf{X}_k]}}_{\mathrm{A}u}$$

$$\mathrm{PD} = \overbrace{\boxed{[\lambda_1\mathbf{X}_1 \quad \lambda_2\mathbf{X}_2 \quad \cdots \quad \lambda_k\mathbf{X}_k]}}^{\lambda u} = \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \cdots & \lambda_k x_{k1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \cdots & \lambda_k x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{1k} & \lambda_2 x_{2k} & \cdots & \lambda_k x_{kk} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

Singular value decomposition

Eigen Decomposition of a matrix

- From AP = PD, we can derive the decomposition of A into a similarity transformation involving P and D: $A = PDP^{-1}$.
  - This decomposition is always possible for a square matrix A.
  - Squaring both sides of equation gives:

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)DP^{-1} \\ &= PD^2P^{-1}. \end{aligned}$$

  - For general positive integer powers: $A^n = PD^nP^{-1}$

  - The inverse of A is $A^{-1} = (PDP^{-1})^{-1} = PD^{-1}P^{-1}$

$$D^{-1} = \begin{bmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^{-1} \end{bmatrix}.$$

Eigen Decomposition - Practice

Q. Suppose that A = $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. It has two eigenvectors u_1 = $\begin{bmatrix} 0.8507 \\ 0.5257 \end{bmatrix}$ and u_2 = $\begin{bmatrix} -0.5257 \\ 0.8507 \end{bmatrix}$. The corresponding eigenvalues were $\lambda_1 = 3.618$ and $\lambda_2 = 1.382$. Then complete the following decomposition:

$$A = PDP^{-1} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

Eigen Decomposition – Practice Solution

Q. Suppose that A = $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. It has two eigenvectors $u_1 = \begin{bmatrix} 0.8507 \\ 0.5257 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -0.5257 \\ 0.8507 \end{bmatrix}$. The corresponding eigenvalues were $\lambda_1 = 3.618$ and $\lambda_2 = 1.382$. Then complete the following decomposition:

$$A = PDP^{-1} = PDP^{T} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

A. $P = [u_1 \quad u_2] = \begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}$

$D = \begin{bmatrix} 3.618 & 0 \\ 0 & 1.382 \end{bmatrix}$

$P^{T} = P^{-1} = \begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}$ (since, P is an orthogonal matrix)

Singular value decomposition

Singular value

- The singular values of a matrix are a set of numbers that can provide useful information about the matrix
  - Information? – matrix rank, pseudoinverse, etc.
- Given an $m \times n$ matrix A, the singular values of A are the positive square roots of the eigenvalues of the matrix A^TA or AA^T.
- The singular values are denoted by $\sigma_{1} \ldots \sigma_i$, where i is the rank of A.

Finding Singular Values Example

- Let $A = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$.
  - $A^TA = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$
  - To find eigenvalues of A^TA, let's get use *characteristic equation*: $\det(A^TA - \lambda I) = 0$.
  - $(5-\lambda)(5-\lambda) - 9 = 0 \rightarrow \lambda^2 - 10\lambda + 16 = 0 \rightarrow \lambda_1 = 2$ and $\lambda_2 = 8 \rightarrow \sigma_1 = \sqrt{2}$ and $\sigma_2 = 2\sqrt{2}$

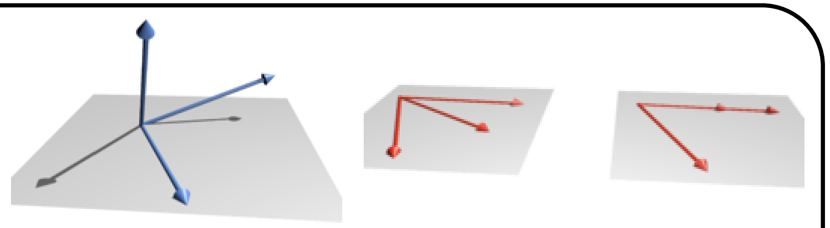
Singular value decomposition

Rank

- In linear algebra, the rank of a matrix A is the dimension of the vector space generated by its columns. This corresponds to the maximal number of linearly independent columns of A.
- A matrix is said to have **full rank** if its rank equals the largest possible for a matrix of the same dimensions. Otherwise, a matrix is said to be rank-deficient.

Linearly Dependent?

A set of vectors is said to be linearly dependent if there is a nontrivial linear combination of the vectors that equals the zero vector. If no such linear combination exists, then the vectors are said to be linearly independent.



linearly independent vectors (left) and linearly dependent vectors (right)

Singular value decomposition

Rank

- Let's take a look at the example. The matrix $C = \begin{bmatrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{bmatrix}$ has row rank 2: the first two rows are linearly independent, but the third is linearly dependent.
  - To determine if these vectors are linearly independent, we can create a matrix with the vectors as its columns and perform *row reduction* to determine if any row can be reduced to all zeros. If no row can be reduced in this way, the vectors are linearly independent.
  - We can perform row operations to reduce the matrix C to row echelon form:

$$\text{row}_2 = \text{row}_2 + 2\text{row}_1 \rightarrow C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 3 & 3 & 0 \end{bmatrix} \rightarrow \text{row}_3 = \text{row}_3 - 3\text{row}_1 \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{bmatrix}$$

$$\rightarrow \text{row}_3 = \text{row}_3 + \text{row}_2 \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 0 & 0 & 0 \end{bmatrix} = \text{rank } 2$$

(The rank is the number of non-zero rows after you've performed row reduction)

Practice

Q. What is the rank of matrix A = $\begin{bmatrix} 1 & 1 & 0 & 2 \\ -1 & -1 & 0 & -2 \end{bmatrix}$

Practice - Solution

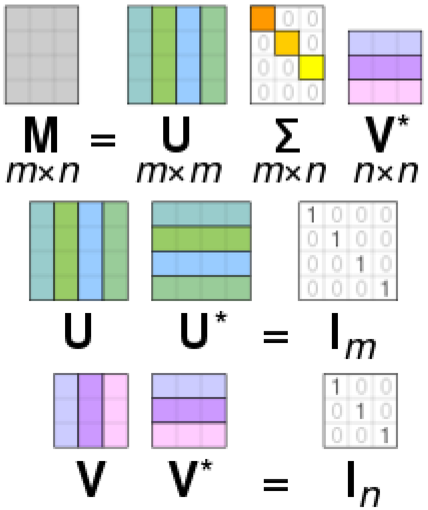
Q. What is the rank of matrix $A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ -1 & -1 & 0 & -2 \end{bmatrix}$

A. $\text{row}_2 = \text{row}_2 + \text{row}_1 \rightarrow A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ = rank 1

Singular value decomposition

Singular value decomposition

- The SVD of $m \times n$ matrix A is given by the formula, $A = U\Sigma V^*$ where:
  - U: $m \times m$ complex unitary matrix.
  - Σ: $m \times n$ rectangular diagonal matrix with non-negative real numbers of the diagonal.
  - V: $n \times n$ complex unitary matrix.
- If A is real, U and V can also be guaranteed to be real *orthogonal* matrices. In this context, the SVD is often denoted $U\Sigma V^T$.
  - The diagonal entries $\sigma_i = \Sigma_{ii}$ of Σ are uniquely determined by A and are known as the singular values of A.
    - The singular values are arranged in decreasing order.
  - The number of non-zero singular values is equal to the rank of A.
  - The columns of U and the columns of V are called left-singular vectors and right-singular vectors of A, respectively.
  - They form two sets of orthonormal bases $u_1, \ldots, u_m$ and $v_1, \ldots, v_n$.



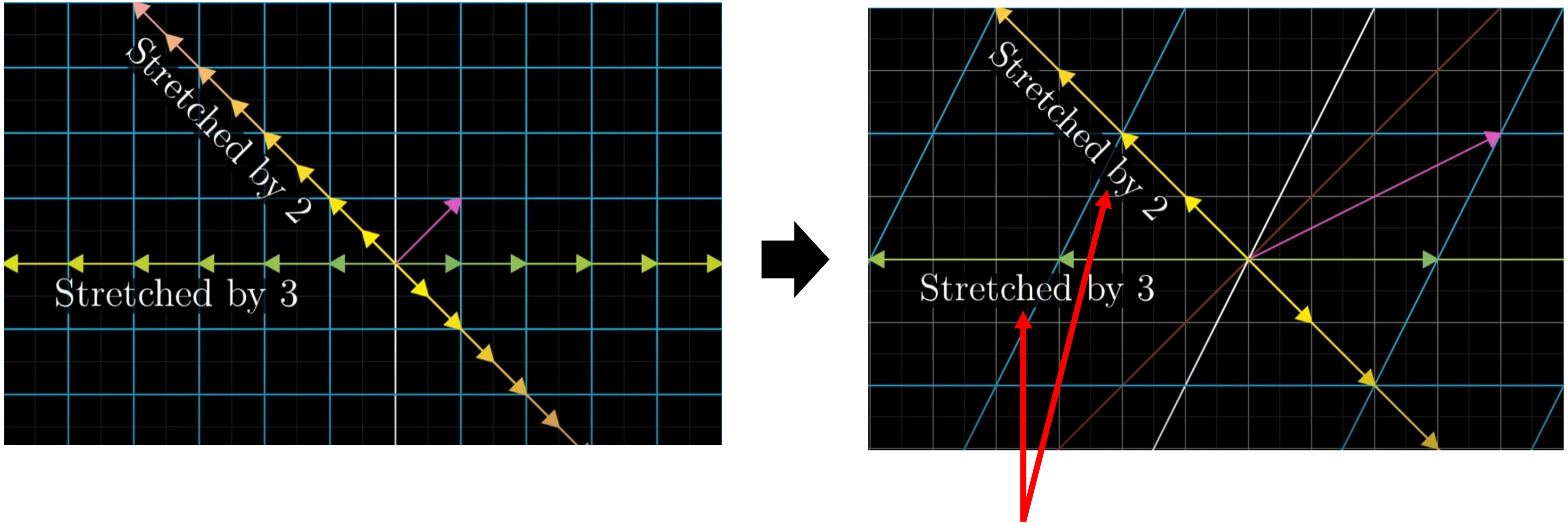
Unitary matrix?

A complex square matrix U is unitary if its conjugate transpose U^* is also its inverse: $U^*U = UU^* = UU^{-1} = I$.

Singular value decomposition

Geometric meaning of SVD

- Eigenvector points in a direction in which it is stretched by the transformation.
- Eigenvalue is the factor by which it is stretched (if the eigenvalue is negative, the direction is reversed.)

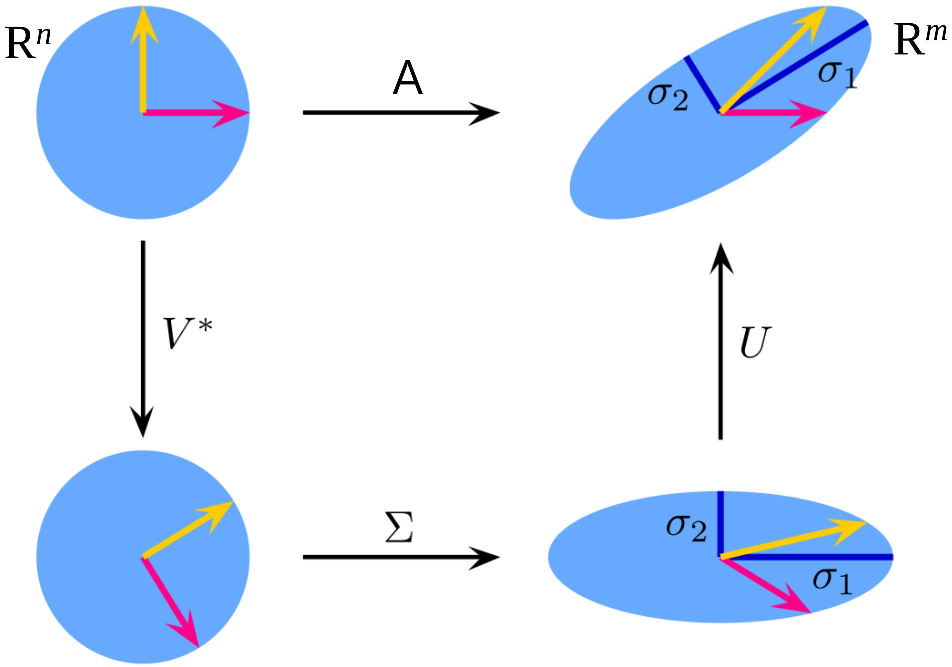


Eigenvectors with eigenvalues of 2 and 3,respectively[7]

Singular value decomposition

Geometric meaning of SVD

- Therefore, SVD can be summarized as follows:
  - Let $T: K^n \rightarrow K^m$ denotes a linear transformation.
  - The linear map (performed by a transformation matrix A) maps the sphere in R^n (shown left below) onto an ellipsoid in R^m(also shown right below).
  - Non-zero singular values are simply the lengths of the semi-axes of this ellipsoid.



Singular value decomposition

SVD calculation ($A = U\Sigma V^T$)

- Calculating the SVD consists of finding the eigenvalues and eigenvectors of AA^T and A^TA.
  - The eigenvectors of A^TA make up the columns of V.
  - The eigenvectors of AA^T make up the columns of U.
  - The singular values in Σ are square roots of eigenvalues from AA^T and A^TA.

Singular value decomposition

SVD example 1 ($A = U\Sigma V^T$)

- Let matrix A = $\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$.
- To find V^T, let's compute $AA^T = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix} = \begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix}$.
- Solving the characteristic equation of AA^T: $\det(A^TA - \lambda I) = 0$.
  - The corresponding eigenvectors are $v_1 = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.
  - These have eigenvalues $\lambda_1 = (\sigma_1)^2 = 40$ and $\lambda_2 = (\sigma_2)^2 = 10$. Then, $\Sigma = \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix}$.
  - The columns of the matrix U are the normalized vectors: $U = \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}$.
- Now, $v_i = \frac{1}{\sigma_1}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}^T u_i$
  - $v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$ and $v_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$. $V = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$
- Thus, the SVD of A is:

$$\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}\begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Singular value decomposition

SVD example 2 ($A = U\Sigma V^T$)

- Let matrix A = $\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$.
- To find V^T, let's compute $A^TA = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$.
  - The eigenvectors of this matrix will give us the vectors v_i, and the eigenvalues will give us the numbers σ_i.
  - Two orthogonal eigenvectors of A^TA are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.
  - To get an orthonormal basis, let $v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$.
  - These have eigenvalues $(\sigma_1)^2 = 32$ *and* $(\sigma_2)^2 = 18$.

$Au = \lambda u$

- To find U, $AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$.
  - Then, $u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$.
- Thus, the SVD of A is:

$$\underbrace{\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}_{V^T}.$$

The Moore-Penrose Pseudo Inverse

Moore-Penrose inverse[1]

- The Moore-Penrose inverse A^+ of a matrix A is the most widely known generalization of the inverse matrix.

- When referring to a matrix, the term pseudoinverse, without further specification, is often used to indicate the Moore-Penrose inverse.

- The term generalized inverse is sometimes used as a synonym for pseudoinverse.

- A common use of the pseudoinverse is to compute a “best fit” (least squares) solution to a system of linear equations that lacks a solution.

- The pseudoinverse is defined and unique for all matrices whose entries are real or complex numbers.

- It can be computed using the singular value decomposition.

The Moore-Penrose Pseudo Inverse

Pseudo-Inverse and Least Squares

- The method of least squares is a way of “solving” an overdetermined system of linear equations $Ax = b$, a system in which A is a rectangular $m \times n$ matrix with more equations than unknowns (when $m > n$).

$$\begin{cases} y_1 = ax_1 + b \\ y_2 = ax_2 + b \\ y_3 = ax_3 + b \\ y_4 = ax_4 + b \\ y_5 = ax_5 + b \end{cases} \Rightarrow \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} b \\ a \end{pmatrix}}_{\mathbf{x}}$$

- The reason why more equations than unknowns arise in such problems is that repeated measurements are taken to minimize errors.
  - This produces an overdetermined and often inconsistent system of linear equations.

The Moore-Penrose Pseudo Inverse

Pseudo-Inverse and Least Squares

- Suppose that we observed the motion of a small object, assimilated to a point, in the plane.
  - From our observations, we suspect that this point moves along a straight line $y = dx + c$.

  - We observed the moving point at three different locations (x_1, y_1), (x_2, y_2) and (x_3, y_3).
  - Then we have equations:

$$c + dx_1 = y_{1,}$$
$$c + dx_2 = y_{2,}$$
$$c + dx_3 = y_{3.}$$

  - If there were no errors in our measurements, these equations would be compatible, and c and d would be determined by only two of the equations. However, in the presence of errors, the system may be inconsistent.

  - The idea of the method of least squares is to determine (c, d) so that it minimizes the *sum of the squares of the errors* (SSE):

$$c + dx_1 + e_1 = y_{1,}\ c + dx_2 + e_2 = y_{2,}\ c + dx_3 + e_3 = y_{3.}$$
$$SSE = (c + dx_1 - y_1)^2 + (c + dx_2 - y_2)^2 + (c + dx_3 - y_3)^2$$

The Moore-Penrose Pseudo Inverse

Pseudo-Inverse and Least Squares

- Suppose that we observed the motion of a small object, assimilated to a point, in the plane.
  - We can rewrite the problem as $y = Ax + e$.
  - Our goal is to minimize $e^T e$:

$$\begin{bmatrix} e_1 & e_2 & \cdots & e_N \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \sum_{i=1}^{N} e_i^2$$

  - Therefore we should find

$min_x\, e^T e = (y-Ax)^T(y-Ax)$

$\rightarrow min_x\, e^T e = y^T y - y^T Ax - x^T A^T y + x^T A^T Ax$

$\rightarrow min_x\, e^T e = y^T y - 2x^T A^T y + x^T A^T Ax$ (note that $(y^T Ax)^T = x^T A^T y$ and both are scalar terms. Therefore, $y^T Ax = x^T A^T y$)

  - Using the mathematical statements ($\frac{\partial}{\partial b} b^T X^T X b = 2XTXb$)[2], let's minimize the expression:

$\rightarrow \frac{\partial}{\partial x} e^T e = -2A^T y + 2A^T Ax = 0$

$\rightarrow \frac{\partial}{\partial x} e^T e = -2A^T y + 2A^T Ax = 0$

$\rightarrow A^T Ax = A^T y \rightarrow x = (A^T A)^{-1} A^T y \rightarrow x = A^{-1} y$

  - Therefore, we can say $A^{\dagger} = (A^T A)^{-1} A^T$

Practice

Q. Consider 3 linear equations:

- $x_1 + 3x_2 = 17$
- $5x_1 + 7x_2 = 19$
- $11x_1 + 13x_2 = 23$

We can write above equations in matrix form as $M\vec{x} = \vec{y}$.
Then, find the Moore-Penrose Pseudoinverse of M.
(Please use matrix calculator such as: https://matrixcalc.org/en/)
(Hint: $A^{\dagger} = (A^T A)^{-1} A^T$)

Solution

Q. Consider 3 linear equations:

- $x_1 + 3x_2 = 17$
- $5x_1 + 7x_2 = 19$
- $11x_1 + 13x_2 = 23$

We can write above equations in matrix form as $M\vec{x} = \vec{y}$.
Then, find the Moore-Penrose Pseudoinverse of M.

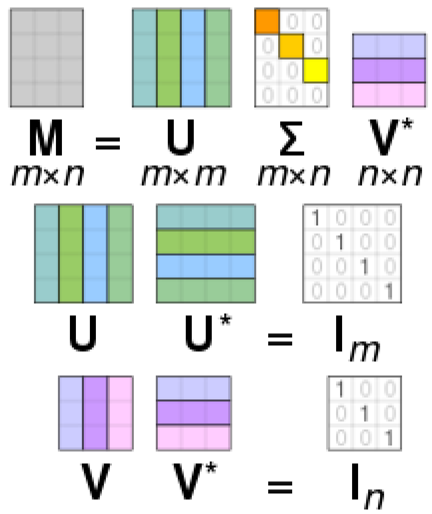
$$M = \begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix}$$ Then, $M^\dagger = (M^T M)^{-1} M^T$

$$M^T = \begin{pmatrix} 1 & 5 & 11 \\ 3 & 7 & 13 \end{pmatrix}$$

$$(M^T M)^{-1} = \begin{pmatrix} \frac{227}{608} & \frac{-181}{608} \\ \frac{-181}{608} & \frac{147}{608} \end{pmatrix} \qquad (M^T M)^{-1} M^T = \begin{pmatrix} \frac{-79}{152} & \frac{-33}{152} & \frac{9}{38} \\ \frac{65}{152} & \frac{31}{152} & \frac{-5}{38} \end{pmatrix}$$

SVD and Pseudo Inverse

Pseudo Inverse and SVD



- The SVD also can be used to compute the pseudo-inverse of a matrix. Let's take a look at the computing inverse matrix using SVD first.

- Let $A \in \mathrm{R}^{m\times m}$. Then there exists orthogonal matrices $\mathrm{U} \in \mathrm{R}^{m\times m}$ and $\mathrm{V} \in \mathrm{R}^{m\times m}$ such that the matrix A can be decomposed as $A = U\Sigma V^T$ where Σ is an $m\times m$ diagonal matrix having the form:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m \end{bmatrix}$$

- In the above equation, the σ_i are termed the *singular values* of the matrix A, the columns of U are termed the *left singular vectors*, and the columns of V are termed the *right singular vectors*.

- U and V are orthogonal, meaning that their transposes are equal to their inverses ($U^TU = I$). Then, the inverse of A can be determined by $A^{-1} = V\Sigma^{-1}U^T$, where

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_m} \end{bmatrix}$$

SVD and Pseudo Inverse

Pseudo Inverse and SVD

- Although the SVD is a powerful tool for obtaining inverse matrix, it cannot be used when the target matrix A is not square (overdetermined systems) or when A has zero singular values.
  - In overdetermined systems, the matrix A does not have a unique inverse.
  - If any of the $\sigma_i = 0$, then Σ^{-1} doesn't exist, because the corresponding diagonal entry would be $\frac{1}{\sigma i} = 1/0$.
- More specifically, if a matrix A has any zero singular values (let's say $\sigma_i = 0$), then multiplying by A destroys information because it takes the component of the vector along the right singular vector $\vec{v_j}$ and multiplies it by zero.
  - We can't recover this information, so there's no way to *invert* the mapping $A\vec{x}$ to recover the original $\vec{x}$.
  - The best we can do is to recover the components of $\vec{x}$ that weren't destroyed via multiplication with zero.
- The matrix that recovers all recoverable information is called the pseudo-inverse and is often denoted $A^\dagger$. We can obtain the pseudoinverse from the SVD by inverting all singular values that are non-zero and leaving all zero singular values at zero.
- Let $A \in R^{m\times n}$. Then there exists orthogonal matrices $U \in R^{m\times m}$ and $V \in R^{n\times n}$ such that the matrix A can be decomposed as $A = U\Sigma V^T$. Then, we can write $A^\dagger = V\Sigma^\dagger U^T$ where

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_p & 0 \end{bmatrix} \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0. \qquad p = \min\{m, n\} \qquad \Sigma^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_p} & 0 \end{bmatrix}$$

Jacobian Inverse Methods

Jacobian Pseudo-inverse

- The pseudo-inverse of the Jacobian matrix is used when the matrix is not square or not invertible.

- Let $J^{\dagger}$ is an $n \times m$ matrix called the pseudo-inverse of J. This pseudo-inverse gives the best possible solution to the equation $J\Delta\theta = \vec{e}$ in the least-squares sense.

- The pseudo-inverse method can be derived by using Moore-Penrose inverse shown as follows:
  - Depending on the dimensions of the Jacobian matrix, you will choose one of the following products:
    - If $m \geq n$ (tall or square matrix) and J is full column rank (rank(J) = n), compute J^TJ.
    - If $m < n$ (wide matrix), and J is full row rank (rank(J) = m), compute JJ^T.
    - If the matrix J is not full row rank or full column rank, SVD can be used to obtain $J^{\dagger}$.
  - Let $J\Delta\theta = \vec{e} \;\rightarrow\; J^TJ\Delta\theta = J^T\vec{e}$.
  - In this case, the minimum magnitude solution $\Delta\theta$ can be expressed as $\Delta\theta = (J^TJ)^{-1}J^T\vec{e}$ (Recall that $J^{\dagger} = (J^TJ)^{-1}J^T$.)

Full row rank? ($\leftrightarrow$ Full column rank)

When a rank r of $m \times n$ matrix equals m ($r = m$), the matrix is said to have a full row rank. In other words, A matrix is full row rank when each of the rows of the matrix are linearly independent. A^TA is always invertible when A is a full row rank (Rank(A^TA) = Rank(A), and the dimensions of Jacobian in IK are $n \times m$ ($n < m$).

Damped Least Squares

Damped Least Squares (DLS)[3]

- The damped least squares method avoids many of the pseudoinverse method's problems with singularities and can give a numerically stable method of selecting $\Delta\theta$.
  - It is also called Levenberg-Margquardt method.
  - This was first used for inverse kinematics by Wampler[4] and Nakamura and Hanafusa[5].

- Rather than just finding the minimum vector $\Delta\theta$ that gives a best solution to $J\Delta\theta = \vec{e}$, DLS find the value of $\Delta\theta$ that minimizes the quantity $\|J\Delta\theta - \vec{e}\|^2 + \lambda^2\|\Delta\theta\|^2$, where $\lambda \in R$ is a non-zero damping constant.
  - The damping constant must be chosen carefully to make the system numerically stable.
  - The damping constant should large enough so that the solutions for $\Delta\theta$ are well-behaved near singularities.
  - However, the convergence rate is too slow, if it is chosen too large.

Damped Least Squares

Damped Least Squares (DLS)[3]

- Minimizing $\|J\Delta\theta - \vec{e}\|^2 + \lambda^2\|\Delta\theta\|^2$ is equivalent to minimizing the quantity:

$$\left\| \begin{pmatrix} J \\ \lambda I \end{pmatrix} \Delta\theta - \begin{pmatrix} \vec{e} \\ 0 \end{pmatrix} \right\|$$

- The corresponding normal equation is

$$\begin{pmatrix} J \\ \lambda I \end{pmatrix}^T \begin{pmatrix} J \\ \lambda I \end{pmatrix} \Delta\theta = \begin{pmatrix} J \\ \lambda I \end{pmatrix}^T \begin{pmatrix} \vec{e} \\ 0 \end{pmatrix}$$

- This can be equivalently rewritten as

$$(J^T J + \lambda^2 I)\Delta\theta = J^T \vec{e}$$

- Since $J^T J + \lambda^2 I$ is non-singular, the DLS solution is equal to

$$\Delta\theta = (J^T J + \lambda^2 I)^{-1} J^T \vec{e} = J^T (J J^T + \lambda^2 I)^{-1} \vec{e}$$

Normal Equation?

Normal equation is an analytical approach to Linear Regression with a Least Square Cost Function. Given a matrix equation $y = Ax + e$, the normal equation is that which minimizes $e^T e = \sum_{i=1}^{n} e_i^2$. This is written by $min_x\, e^T e = (y - Ax)^T (y - Ax) = min_x\, e^T e = y^T y - 2x^T A^T y - x^T A^T A x$.

Using the mathematical statements $\left(\frac{\partial}{\partial b} b^T X^T X b = 2XTXb\right)$[2], we can solve: $\frac{\partial}{\partial b} e^T e = -2A^T y + 2A^T A x = 0$ and can be written by $A^T A x = A^T y$ or $x = (A^T A)^{-1} A^T y$. is called normal equation and $A^T A$ is a normal matrix.

Damped Least Squares

Pseudo-inverse Damped Least Squares[3]

- The Pseudo-inverse Damped Least Squares uses the singular value decomposition (SVD) under the damped least squares method. Hence, the matrix $JJ^T+\lambda^2 I$ can be rewritten as:

$$JJ^T+\lambda^2 I = (U\Sigma V^T)(V\Sigma^T U^T) + \lambda^2 I$$

  where U and V are orthogonal matrices and Σ is diagonal.

- If J is $m \times n$, then U is $m \times m$, Σ is $m \times n$, and V is $n \times n$.
  - U and V are orthogonal matrices: $U^T = U^{-1}$, $V^T = V^{-1}$
  - The transpose Σ^T of Σ is the $n \times m$ diagonal matrix with diagonal entries σ_i.
  - The product $\Sigma\Sigma^T$ is the $m \times m$ matrix with diagonal entries σ_i^2

- Therefore, $JJ^T+\lambda^2 I = (U\Sigma V^T)(V\Sigma^T U^T) + \lambda^2 I = U(\Sigma\Sigma^T + \lambda^2 I)U^T$.
  - The matrix $\Sigma\Sigma^T + \lambda^2 I$ is the diagonal matrix with diagonal entries $\sigma_i^2 + \lambda^2$.

Damped Least Squares

Pseudo-inverse Damped Least Squares[3]

- In here, $JJ^T+\lambda^2 I$ is non-singular, and its inverse is the $m \times m$ diagonal matrix with non-zero entries $(\sigma_i^2 + \lambda^2)^{-1}$
- Therefore, we can write

$$\begin{aligned} J^T (JJ^T+\lambda^2 I)^{-1} &= J^T (U(\Sigma\Sigma^T + \lambda^2 I)U^T)^{-1} \\ &= (V\Sigma^T U^T)(U(\Sigma\Sigma^T + \lambda^2 I)^{-1} U^T) \\ &= V\Sigma^T (U(\Sigma\Sigma^T + \lambda^2 I)^{-1} U^T \\ &= VEU^T \end{aligned}$$

where E is an n × m diagonal matrix with entries equal to $e_{i,i} = \frac{\sigma_i}{\sigma_i^2 + \lambda^2}$.

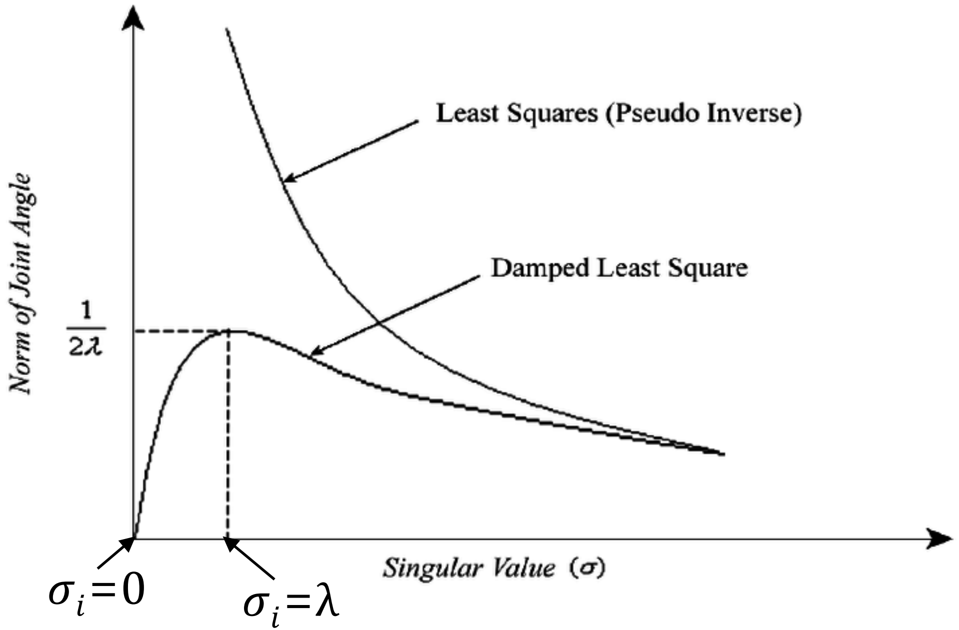
- Then, the damped least squares solution can be expressed in the form

$$J^T (JJ^T+\lambda^2 I)^{-1} = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \lambda^2} v_i u_i^T \text{ (For } i > r, \sigma_i = 0, \text{thus } \sigma_r \neq 0, \text{)}$$

Damped Least Squares

Pseudo-inverse Damped Least Squares[3]

- Let's compare the damped least square, $J^T(JJ^T+\lambda^2 I)^{-1} = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2+\lambda^2} v_i u_i^T$, to the solution of pseudoinverse, $J^\dagger = \sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^T$.
  - The pseudoinverse method is unstable as σ_i approaches zero.
  - For values of σ_i which are large compared to λ, the damped least squares method is not very different from the pseudoinverse since $\frac{\sigma_i}{\sigma_i^2+\lambda^2} \approx \frac{1}{\sigma_i}$.
  - But, when σ_i is of the same order of magnitude as λ or smaller, then the values $\frac{\sigma_i}{\sigma_i^2+\lambda^2}$ and $\frac{1}{\sigma_i}$ diverge.
  - This means that the damped least squares method tends to act similarly to the pseudoinverse method away from singularities and effectively smooths out the performance of pseudoinverse method.

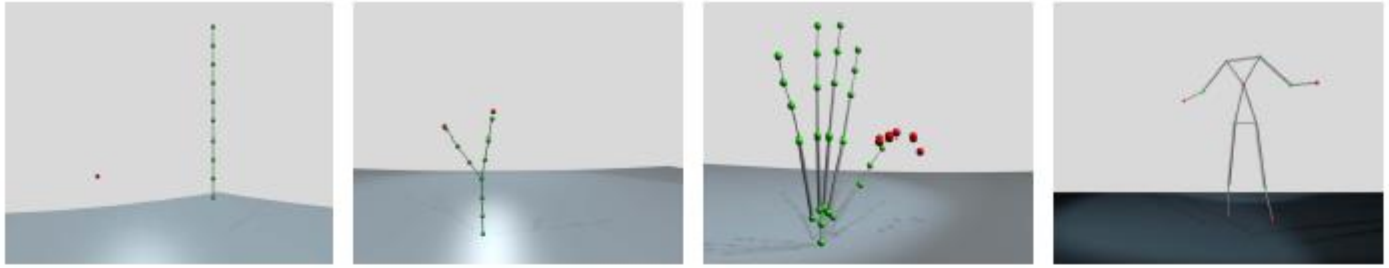


Performance Comparison

Jacobian Transpose vs. Jacobian DLS vs. Jacobian SVD-DLS[5]

- Runtimes are in seconds and were measured with custom MATLAB code on a Pentium 2 Duo 2.2 GHz.

- No optimizations were used for any method reported in the table.



(a) (b) (c) (d)

Figure *The structure of the models used in our experimental examples. (a) A kinematic chain consisting of 10 joints and 1 end effector. There are 2 kinematic chain models, an unconstrained and a constrained version, (b) a kinematic model with 10 unconstrained joints and 2 end effectors, (c) a hand model with 26 unconstrained joints and 5 end effectors, (d) a 13 joint humanoid model, in a constrained and unconstrained version, with 4 end effectors. The target joint positions (end effectors) are shown in red and the joint positions that the IK solvers have to estimate are shown in green.*

Average results when the target is reachable.

	Number of Iterations	Matlab exe. time (sec)	Frames per second
Jacobian Transpose	1311.190	12.98947	0.077
Jacobian DLS	998.648	10.48501	0.095
Jacobian SVD-DLS	808.797	9.29652	0.107

Reference

[1] https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse
[2] https://economictheoryblog.com/2015/02/19/ols_estimator/
[3] http://www.cs.cmu.edu/~15464-s13/lectures/lecture6/iksurvey.pdf
[4] Y. Nakamura and H. Hanafusa, Inverse kinematics solutions with singularity robustness for robot manipulator control, Journal of Dynamic Systems, Measurement, and Control, 108 (1986), pp. 163–171.
[5] C. W. Wampler, Manipulator inverse kinematic solutions based on vector formulations and damped least squares methods, IEEE Transactions on Systems, Man, and Cybernetics, 16 (1986), pp. 93–101.
[6] Inverse Kinematics: a review of existing techniques and introduction of a new fast iterative solver, University of Cambridge, Andreas Aristidou and Joan Lasenby, 2009