Univ. of Washington

Lecture : Matrix Inverses and Systems of Equations

Winter 2021

Lecturer: Dan Calderone

Systems of Equations

Matrices are used to represent and solve systems of linear equations. Suppose we $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ that satisfy.

$$y = Ax \tag{1}$$

Note that this equation is slightly more complicated than it first appears. Depending on the shape of A it may have a unique solution, no solution, or a whole subspace of solutions.

Unique Solution

The simplest case is that A is square, ie. $x, y \in \mathbb{R}^n$ and the columns are linearly independent. This means there is a unique linear commbination of the columns that reaches every individual point y in the co-domain. We can compute this exact linear combination by doing *Gaussian elimination* also known as *row reduction*. Each step of Gaussian elimination, each *elementary row operation* can be represented by left-multiplication of Equation (1) by a specific type of matrix called *elementary matrices*. These elementary matrices come in three types: row-multiplying, row-swapping, and row-adding demonstrated below

$$\underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \alpha & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{multiplying a row by } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & \alpha & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{adding a row times } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{swapping rows}} \tag{2}$$

When we perform Gaussian elimination on Equation (1) to transform A into the identity, we left-multiply by the appropriate set of elementary matrices $\{E_1, \ldots, E_k\}$

$$\underbrace{(E_k \cdots E_1)}_{A_l^{-1}} y = \underbrace{(E_k \cdots E_1) A}_{I} x \tag{3}$$

These elementary matrices multiplied together are called the *left-inverse* $A_l^{-1} = (E_k \cdots E_1)$, ie. the matrix that transforms A into the identity by left-multiplying. Note that we could have performed a similar procedure to solve the equation $y^\top = x^\top A$ except we would multiply on the right by *elementary column matrices*. This procedure would construct the *right inverse* of A, denoted A_r^{-1}. $y^\top A_r^{-1} = x^\top A A_r^{-1} = x^\top$. Assuming A is square and invertible, these two left and right inverses are the same and we simply denote them as $A^{-1} = A_l^{-1} = A_r^{-1}$. This can be seen from

$$\begin{aligned} A_l^{-1} \cdot A &= I \\ A_l^{-1} \cdot A \cdot A_r^{-1} &= I \cdot A_r^{-1} \\ A_l^{-1} &= A_r^{-1} \end{aligned} \tag{4}$$

No solution (Least Squares)

If $m > n$, ie. A is "tall", then it is unlikely that there is any solution at all. The columns of A span a subspace of the co-domain called the range of A. There will only be a solution for x if y happens to lie in this subspace. If the columns of A are linearly independent, then A will still have a left-inverse. This is based on the fact that the linear independence of the columns of implies that the matrix $A^\top A$ will be invertible. This in turn implies that we can construct a left-inverse as $A_l^{-1} = (A^\top A)^{-1} A^\top$. Supposing that y is actually in the range of A, ie. there does exist an x solving (1), we can find this x using this left-inverse.

$$\begin{aligned} \text{Assume } y \text{ in range of } A\text{...} \qquad\qquad y &= Ax \\ (A^\top A)^{-1} A^\top y &= (A^\top A)^{-1} A^\top \cdot Ax = x \end{aligned} \tag{5}$$

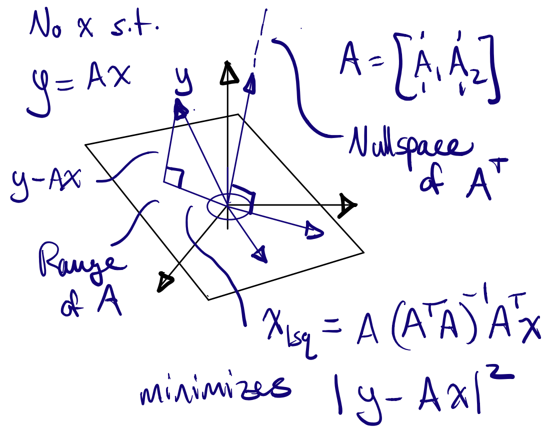
Now suppose y is not in the range of A. We can still try to find an x that makes Ax as close to y as possible, ie. we can try to minimize

$$||y - Ax||_2^2 = (y - Ax)^\top (y - Ax) = y^\top y + y^\top Ax + x^\top A^\top Ax = \sum_i (y_i - A_{i:} x)^2 \tag{6}$$

x that minimizes this quantity is called the *least squares solution*, $x_{\text{lsq}} = A(A^\top A)^{-1} A^\top y$ which is the projection of y onto the range of A. We can derive the least squares solution by computing the derivative of (6) and set it equal to 0.

$$\frac{\partial}{\partial x}\Big(y^\top y - y^\top Ax - x^\top Ay + x^\top A^\top Ax\Big) = -2y^\top A + 2x^\top A^\top A = 0 \tag{7}$$

$$\Rightarrow \qquad x = (A^\top A)^{-1} A^\top y \tag{8}$$



Subspace/Continuum of Solutions

Suppose $n > m$, ie. A is "fat", and there are more than m linearly independent columns. In . this case, we have more columns than we need to span the space. If we pick any m linear independent columns, we can compute a solution. Suppose the first m columns of A are linearly independent, $A = [\bar{A} \ \cdots]$ where $\bar{A} \in \mathbb{R}^{m \times m}$. We can then compute one solution as $x^1 = [\bar{A}^{-1}y \ \mathbf{0}]^\top$ where $\mathbf{0}$ is the appropriate size vector of zeros. The same procedure with different sets of columns produces up to $n - m + 1$ linearly independent solutions which we can organize as the columns of $X = [x^1 \ \cdots \ x^{n-m+1}]$. Note that $A(x^i - x^j) = 0$, ie. $x^i - x^j$ is in the nullspace of A. A basis for the nullspace of A can be computed as the columns of XW where the matrix $W \in \mathbb{R}^{(n-m+1)\times(n-m)}$ is given by $W = [\mathbf{1} - I]^\top$ where $\mathbf{1}$ is a vector of ones of the appropriate size. (Note that W computes differences between the columns of X. A different W that computes column differences could be used.) Any solution of (1) has the form

$$x = x^0 + x_{\text{NS}} = x^0 + XWz$$

for some $z \in \mathbb{R}^{n-m}$, ie. any solution consists of some specific solution x^0 plus some component in the nullspace of A. We can compute a specific solution using the method above (selecting m linearly independent columns). However, assuming the rows of A are linearly independent and if we want a specific solution x^0 that is orthogonal to the nullspace of A, then we can select x as a linear combination of the rows of A. Assume x^0 has the form $x^0 = A^\top w$ with $w \in \mathbb{R}^m$. Plugging into (1), gives

$$y = AA^\top w \qquad \Rightarrow \qquad w = (AA^\top)^{-1}y \qquad \Rightarrow \qquad x^0 = A^\top(AA^\top)^{-1}y \tag{9}$$

Note that x^0 is y times a right-inverse of A. Note also that x^0 is orthogonal to the nullspace of A since $x_{NS}^\top A^\top(AA^\top)^{-1} = 0$. Note also that x^0 computed in this way is the solution with the *minimum 2-norm*. To see this, note that adding some component from the nullspace only increases the square of the 2-norm.

$$
\begin{aligned}
|x^0 + x_{\text{NS}}|^2 &= (x^0 + x_{\text{NS}})^\top (x^0 + x_{\text{NS}}) && (10)\\
&= (x^0)^\top x^0 + 2x_{\text{NS}}^\top x^0 + x_{\text{NS}}^\top x_{\text{NS}} && (11)\\
&= (x^0)^\top x^0 + x_{\text{NS}}^\top x_{\text{NS}} = |x^0|^2 + |x_{\text{NS}}|^2 \geq= |x^0|^2 && (12)
\end{aligned}
$$

General Case

Minimum-Norm, Least Squares (Moore-Penrose Pseudoinverse)

In the general case, $A \in \mathbb{R}^{m \times n}$ may not be full column or row rank. In this case neither $A^\top A$ or $AA^\top$ or In this case, for there are many possible x's that are all equally bad at reaching y. Perhaps the most sensible x to choose in this case is the *minimum-norm, least squares* solution. Here, we look for the least squares solution that does not include any element in the nullspace of A. This can be computed using the *Moore-Penrose* pseudoinverse denoted $A^\dagger$. This is best understood using the singular-value decomposition. (Here we assume the matrix A is real and so we use the real SVD; an exactly analogous formula works in the complex case). Given that the SVD of A, $A^\dagger$ can be written as follows

$$
A = \underbrace{\begin{bmatrix} U_1 \; U_2 \end{bmatrix}}_{U} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -V_1^\top - \\ -V_2^\top - \end{bmatrix}}_{V^\top} \qquad \Longrightarrow \qquad A^\dagger = \underbrace{\begin{bmatrix} V_1 \; V_2 \end{bmatrix}}_{V} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -U_1^\top - \\ -U_2^\top - \end{bmatrix}}_{U^\top} \tag{13}
$$

Note that in this formula, we've followed the standard rules for taking an inverse—reversing the order and inverting V and U (since $U^{-1} = U^\top$ and $V^{-1} = V^\top$) — but we only inverted the part of the center matrix that is invertible. As detailed in the SVD lecture, U_1, U_2, V_1, V_2 have the following interpretations.

$$
\begin{aligned}
U_1 &: \text{ orthonormal basis for the range of } A\\
U_2 &: \text{ orthonormal basis for the nullspace of } A^\top\\
V_1 &: \text{ orthonormal basis for the range of } A^\top\\
V_2 &: \text{ orthonormal basis for the nullspace of } A
\end{aligned}
$$

Intuitively, A contains an invertible map between the range of A and the range of $A^\top$ and we've inverted this part of A while ignoring the part in the nullspace. Expanding out, we get that (13) could be written as

$$
A = U_1 \Sigma V_1^\top \qquad \Longrightarrow \qquad A^\dagger = V_1 \Sigma^{-1} U_1^\top \tag{14}
$$

Note that here this looks like a simple formula except U_1 and V_1 are tall so they can't simply be inverted.

Exercise:. Show that for an equation $y = Ax$ (for general A), $x = A^\dagger y$ gives the least squares solution with the minimum norm.

Gaussian Elimination: Row Reduction

We now consider what happens if we perform Gaussian elimination on a general matrix with rank k where $k < m$, $k < n$. For a matrix $A \in \mathbb{R}^{m\times n}$ with rank k, assuming the first k columns of A are linearly independent we can find an invertible $E \in \mathbb{R}^{m\times m}$ that is a composition of elementary matrices $E = E_\ell \cdots E_1$ such that

$$EA = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \tag{15}$$

with $I \in \mathbb{R}^{k\times k}$ and $B \in \mathbb{R}^{k\times n-k}$. It will be helpful to decompose E and also E^{-1} as

$$E = \begin{bmatrix} - & E' & - \\ - & E'' & - \end{bmatrix}, \qquad E^{-1} = \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix}$$

where

$$E' \in \mathbb{R}^{k\times m}, \quad E'' \in \mathbb{R}^{(m-k)\times m}, \qquad\qquad F' \in \mathbb{R}^{m\times k}, \quad F'' \in \mathbb{R}^{m\times(m-k)}$$

Note: since A has rank k, there will always be at least k linearly independent columns, if the first k columns aren't linearly independent then the above formula must be changed to be

$$EAP = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \tag{16}$$

where P is some permutation matrix that reorders the columns so that the first k are linearly independent. This is the most general form Gaussian elimination can take. In this case, we solve the linear system $y = APx'$ where $x = Px' \iff x' = P^\top x$. Once we've solved for x', we can recover x. For simplicity, we will consider equation (15).

Note that since the columns (or rows) of E', E'', F', F'' are all columns (or rows) of invertible matrices, they must be linearly independent. We note also that

$$I = EE^{-1} = \begin{bmatrix} - & E' & - \\ - & E'' & - \end{bmatrix} \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix} = \begin{bmatrix} E'F' & E'F'' \\ E''F' & E''F'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Specifically, note which submatrices must be orthogonal. We also have that $I = E^{-1}E = F'E' + F''E''$.

Using the above decomposition the row-reduction operations become

$$EA = \begin{bmatrix} - & E' & - \\ - & E'' & - \end{bmatrix} \begin{bmatrix} | & | \\ A' & A'' \\ | & | \end{bmatrix} = \begin{bmatrix} E'A' & E'A'' \\ E''A' & E''A'' \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix}$$

It can also be quite useful to write the above equation as a decomposition of A

$$A = E^{-1} \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix} \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} = F' \begin{bmatrix} I & B \end{bmatrix}$$

It is clear from this that F' must span the range of A. Since the columns are linearly independent it is also a basis. We also have that the rows of E'' form a basis for the nullspace of A^T by a similar linear independence argument, rank-nullity (applied to the co-domain) and the fact that $E''F' = 0$. Note that the rows of $\begin{bmatrix} I & B \end{bmatrix}$ are also linearly independent (since the first subblock is the identity) and thus $\begin{bmatrix} I & B \end{bmatrix}^T$ is a basis for the range of A^T Finally, by arguments given in the discussion on nullspaces, the rows of $\begin{bmatrix} B^T & -I \end{bmatrix}^T$ are a basis for the nullspace of A. We can summarize these insights in a list of bases for the four fundamental subspaces related to A.

$$\underset{of\ A}{Range} : \ F', \quad \underset{of\ A^T}{Nullspace} : \ E''^T \quad \underset{of\ A^T}{Range} : \begin{bmatrix} I \\ B^T \end{bmatrix}, \quad \underset{of\ A}{Nullspace} : \begin{bmatrix} B \\ -I \end{bmatrix}$$

Gaussian Elimination: Column Reduction

We can make a similar argument to the above for column reduction in the general case. For a matrix $A \in \mathbb{R}^{m \times n}$ with rank k assuming the first k rows of A are linearly independent, we can find an invertible $E \in \mathbb{R}^{m \times m}$ that is a composition of elementary matrices $E = E_1 \cdots E_\ell$ such that

$$AE = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix}$$

with $I \in \mathbb{R}^{k \times k}$ and $C \in \mathbb{R}^{k \times m-k}$. (If the first k rows are not linearly independent, replace AE with PAE for a permutation matrix P that shuffles k linearly independent rows into the first k spots.) Note here this composition of elementary matrices E will be different than in the row reduction case. Again it will be helpful to decompose E and also E^{-1} as

$$E = \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix}, \qquad E^{-1} \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix}$$

where $E' \in \mathbb{R}^{n \times k}$, $E'' \in$ ⋉ × ⋉ − ⅂, $F' \in \mathbb{R}^{k \times n}$, and $F'' \in$ ⋉ − ⅂ × ⋉. Note that since the columns (or rows) of E', E'', F', F'' are all columns (or rows) of invertible matrices, they must be linearly independent. We note also that

$$I = E^{-1}E = \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix} \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix} = \begin{bmatrix} F'E' & F'E'' \\ F''E' & F''E'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Specifically, note which submatrices must be orthogonal. We also have that $I = EE^{-1} = E'F' + E''F''$.

Using the above decomposition the column-reduction operations become

$$AE = \begin{bmatrix} - & A' & - \\ - & A'' & - \end{bmatrix} \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix} = \begin{bmatrix} A'E' & A'E'' \\ A''E' & A''E'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix}$$

It can also be quite useful to write the above equation as decomposition of A

$$A = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix} E^{-1} = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix} \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix} = \begin{bmatrix} I \\ C \end{bmatrix} F'$$

It is clear from this that the rows of F' must span the range of A^T. Since the rows are linearly independent it is also a basis. We also have that the columns of E'' form a basis for the nullspace of A by a similar linear independence argument, rank-nullity (applied to the co-domain) and the fact that $F'E'' = 0$. Note that the columns of $\begin{bmatrix} I \\ C \end{bmatrix}$ are also linearly independent (since the first subblock is the identity) and thus $\begin{bmatrix} I \\ C \end{bmatrix}^T$ is a basis for the range of A Finally, by arguments given in the discussion on nullspaces, the rows of $\begin{bmatrix} C & -I \end{bmatrix}$ are a basis for the nullspace of A^T. We can summarize these insights in a list of bases for the four fundamental subspaces related to A.

$$\substack{Range \\ of\ A} : \begin{bmatrix} I \\ C \end{bmatrix}, \quad \substack{Nullspace \\ of\ A^T} : \begin{bmatrix} C^T \\ -I \end{bmatrix} \quad \substack{Range \\ of\ A^T} : \ F'^T, \quad \substack{Nullspace \\ of\ A} : \ E''$$

Inverse Properties

Properties of inverses:

$P, Q \in \mathbb{C}^{n\times n}$ invertible, and $k \in \mathbb{C}$.

- $(P^{-1})^{-1} = P$
- $(kP)^{-1} = \frac{1}{k}P^{-1}$
- $(PQ)^{-1} = Q^{-1}P^{-1}$
- $\det(P^{-1}) = \frac{1}{\det(P)}$
- $P^{-1} = \frac{1}{\det(P)}\text{Adj}(P)$

Equivalent Inverse Properties:

- P is invertible, ie. P^{-1} exists.
- $P^\top$ is invertible

- P can be row reduced to the identity (via Gaussian Elimination (GE))
- P can be column reduced to the identity (via GE).
- P is a product of elementary matrices.
- P (square) is full row rank.
- P (square) is full column rank.
- Columns of P (square) are linearly independent, ie. $Px = 0 \ \Rightarrow \ x = 0$.
- Rows of P (square) are linearly independent, ie. $y^\top P = 0 \ \Rightarrow \ y^\top = 0$. Rows of P (square) are linearly independent.
- $y = Px$ has a unique solution for each y.
- P has a trivial nullspace. $\mathcal{N}(P) = \{0\}$
- 0 is not an eigenvalue of P.
- $\det(P) \neq 0$.
- There exists Q such that $PQ = QP = I$ ($P^{-1} = Q$).
- P has a left and a right inverse.

Inverse Formulas

- 2×2 inverse

$$
\begin{aligned}
P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad P^{-1} &= \frac{1}{\det(P)}\text{Adj}(P) = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\
&= \frac{1}{\det(P)}\Big[\text{Tr}(P)I - P\Big]
\end{aligned}
$$

- 3×3 inverse

$$
\begin{aligned}
P^{-1} &= \frac{1}{\det(P)}\text{Adj}(P) \\
&= \frac{1}{\det(P)}\left[\tfrac{1}{2}\Big(\text{Tr}(P)^2 - \text{Tr}(P^2)\Big)I - P\text{Tr}(P) + P^2\right]
\end{aligned}
$$

- **Block Matrix Inversion**

  $P^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}$
  $= \begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$
  $= \begin{bmatrix} A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$ $assuming$ D^{-1} and $(A-BD^{-1}C)^{-1}$ exist or A^{-1} and $(D-CA^{-1}B)^{-1}$ exist.

  Proof:

  $$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \left(\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A-BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \end{aligned}$$

  $$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \left(\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (D-CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \end{aligned}$$

- **Woodbury Matrix Identity**
  Note: this formula is a work horse of matrix algebra and worth memorizing.

  $$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$

  where $A \in \mathbb{C}^{n\times n}$, $U \in \mathbb{C}^{n\times k}$, $C \in \mathbb{C}^{k\times k}$, and $V \in \mathbb{C}^{k\times n}$. This formula is particularly useful when $n > k$ (U is tall and V is fat). In particular, if U is a column vector, V is a row vector, and C is a scalar, then this equation is called the *Sherman-Morrison Formula*.

  Special Cases:

  - **Inverse of $A+B$:**

    $$(A+B)^{-1} = A^{-1} - A^{-1}B(I+A^{-1}B)^{-1}A^{-1}$$

    Note: other forms are possible as well.

  - **Sherman-Morrison:**

    $$(A+uv^\top)^{-1} = A^{-1} - A^{-1}u\frac{1}{1+v^\top A^{-1}u}v^\top A^{-1}$$

- **Neumann Series**

$$A^{-1} = \sum_{n=0}^{\infty} (I - A)^n, \qquad \text{if } \lim_{n \to \infty} (I - A)^n = 0$$

- **Derivative of Inverse**

  For $P(t)$

$$\frac{\partial P^{-1}}{\partial t} = -P^{-1} \frac{\partial P}{\partial t} P^{-1}$$