Lecture 21
Operator Equations II
Short Summary of the Previous Lecture

The problems $f(v) = 0$, $f(v) = v$, $f(v) = \lambda v$ are essentially identical.

The fixed-point problem ($f(v) = v$) in the case of contraction mapping (Lipschitz mapping with $\mu < 1$) of a complete metric space (Banach space) into itself always has a unique solution, that can be found as a limit element of the Picard sequence.

$$v_0, \ v_1 = f(v_0), \ v_2 = f(v_1), \ ... \ , \ v_n = f(v_{n-1}), \ ... \ \lim_{n \to \infty} v_n = \tilde{v} = f(\tilde{v})$$

The Contraction Mapping Theorem (Banach Fixed-Point Theorem) is a tool to prove existence and uniqueness of a solution for practical mathematical problems, e.g., differential equations - but this is not necessarily the easiest or the fastest way to solve the problem.
Does it always work?
Singularity in an Ordinary Differential Equation

Point of singularity for a function $f(t)$: $\lim_{t \to t_0} f(t) = \infty$

**Example:** consider the equation $t \cdot y' = y$, $y(0) = 0$ \quad \left( y' = \frac{y}{t}, \quad \lim_{t \to 0} y(t) = 0 \right)$

In the operator equation $y(t) = y_0 + \int_0^t \frac{y(t')}{t'} dt'$ there is a singularity at $t' \to 0$ and, $\frac{dg}{dy} = \frac{1}{t} \to \infty$

Thus the theorem predicting uniqueness of the solution of 1\textsuperscript{st} order differential equation (even linear in this case) with one initial condition specified is not applied.

However, beyond the singularity point $\ln y = \ln t + c' \quad y = ct$

Any value of $c$ would satisfy the initial condition $y(0) = 0$ – infinite number of solutions.

And, there is no $c$ to satisfy initial condition such as $y(0) = 1$ – no solution.

**Another example** $t \cdot y' = -y$, $y(0) = 0$ \quad \left( y' = -\frac{y}{t}, \quad \lim_{t \to 0} y(t) = 0 \right)$

Solution, similarly to the previous case, $\ln y = -\ln t + c' \quad y = \frac{c}{t}$

There is no constant $c$ to satisfy the initial condition $y(0) = 0$

Yet $y(t) = 0$ satisfies the differential equation $t \cdot y' = -y$, $y(0) = 0$ -- a unique solution

It is important to spot the singularity points when solving differential equations.
Singularity, one more example

\[ \frac{dy}{dt} = 1 \quad y(0) = 0 \]

\[ \left( \frac{dy}{dt} = \frac{1}{y}, \quad \left. \frac{dy}{dt} \right|_{y=0^+} = \frac{1}{y} \right|_{y=0^+} \quad \text{and} \quad \left. \frac{dy}{dt} \right|_{y=0^-} = \frac{1}{y} \right|_{y=0^-} \]

\[ y \, dy = dt \]

\[ \frac{y^2}{2} = t + C \]

\[ y = \pm \sqrt{2(t + C)} \]

\[ y(0) = 0 \implies y = \sqrt{2t} \quad \text{or} \quad y = -\sqrt{2t} \]

two different solutions for this initial condition

\[ y(0) = 1 \implies y = \sqrt{2t - 1} \quad \text{unique solution when there is no singularity} \]
Residual Principle

The Picard sequence in case of $\mu \approx 1$ ($\mu < 1$ to guarantee the convergence) may be converging too slowly. Although it is always an exponential convergence, and any prescribed accuracy $\varepsilon$ can be achieved in a finite and predictable number of iterations $N(\varepsilon)$, this finite number is scaled as $1/(1 - \mu)$, and it may be way too large for practical applications.

The Residual Principle is an alternative algorithm for consecutive approximations.

Consider an operator equation $f(v) = w$, $v \in V$, $w \in W$, $V$ and $W$ are Banach spaces.

Let $V_1 \subset V$ be a Banach subspace in $V$ (“$V_1$ is easy to work with”)

Now consider a task: find $\tilde{v} \in V_1$, which is the best approximation to the solution of $f(v) = w$.

$V_1$ may or may not contain the exact solution $v$. 
Residual Principle, Continued

Re-formulate the problem as a root-finding problem: \( f(v) - w = \theta. \)

Let \( \bar{v} \in V \) be the exact solution \( f(\bar{v}) - w = \theta. \)

For any other element \( v_s \) (\( v_s \) is called a substitution for the equation \( f(\bar{v}) - w = \theta \))
\[
f(v_s) - w = r(v_s) \neq \theta
\]

\( r(v_s) \) is the residual corresponding to the substitution \( v_s \).

If \( \bar{v} \notin V_1 \), all elements in \( V_1 \) yield some non-zero residual.

As long as \( V_1 \) is a complete subset, there is an element \( \bar{v} \in V_1 \)
which corresponds to the minimal residual:
\[
\forall v_1 \in V_1 \quad \| f(v_1) - w \| > \| f(\bar{v}) - w \|
\]

Notice: \( \min_{v_1 \in V_1} \| f(v_1) - w \| \) does not necessarily coincide with \( \min_{v_1 \in V_1} \| v_1 - \bar{v} \| \)

That is, the residual minimization problem, generally speaking, is different from
the problem of finding the best approximation to the exact solution.

In some special cases, however, the residual minimization problem is technically easier to deal with,
and, moreover, it may yield the best approximation - a subject of the Residual Theorem.
Example: How Bad Could it Be

By substituting the root finding problem with the consecutive approximations within a smaller set $V_1 \subset V$, and then by switching to the minimization of the residual, one may get rather useless solutions. The approximated solutions have to be always considered carefully. Consider the following example.

\[ z^2 = i \]
\[ f(v) - w = \theta \quad V = \mathbb{C} \quad V_1 = \mathbb{R} \]
\[ z^2 - i = 0 \]

Exact solutions
\[ \tilde{z}_1 = \frac{\sqrt{2}}{2} + i \frac{\sqrt{2}}{2} \quad \tilde{z}_2 = -\frac{\sqrt{2}}{2} - i \frac{\sqrt{2}}{2} \]

The best approximation in $V_1 = \mathbb{R}$
\[ \tilde{z}_1 = \frac{\sqrt{2}}{2} \quad \tilde{z}_2 = -\frac{\sqrt{2}}{2} \]

\[ z_1 \in \mathbb{R} \quad \| z_1 - \tilde{z}_1 \| = \sqrt{(z_1 - \text{Re}(\tilde{z}_1))^2 + (\text{Im}(\tilde{z}_1))^2} \to \min \quad \text{at} \quad z_1 = \tilde{z}_1 \]

\[ z_2 \in \mathbb{R} \quad \| z_2 - \tilde{z}_2 \| = \sqrt{(z_2 - \text{Re}(\tilde{z}_2))^2 + (\text{Im}(\tilde{z}_2))^2} \to \min \quad \text{at} \quad z_2 = \tilde{z}_2 \]

The smallest residual, however, is achieved for $z_3 = \tilde{z} = 0$

\[ z_3 \in \mathbb{R} \quad \| r(z_3) \| = \| z_3^2 - i \| = \sqrt{z_3^4 + 1} \to \min \quad \text{at} \quad z_2 = \tilde{z} = 0 \]
Example: It May Be Rather Good

Consider the following equation.

\[ z^2 = 10000 + i \]
\[ f(v) - w = \theta \quad V = C \quad V_1 = R \]
\[ z^2 - 10000 - i = 0 \]

Exact solutions

\[ z_1 = \sqrt{10000 + i} \approx 100.000000012 + 0.005i \quad z_2 = -\sqrt{10000 + i} \approx -100.000000012 - 0.005i \]

The best approximation in \( V_1 = R \) \( z_1 \approx 100.000000012 \quad z_2 \approx -100.000000012 \)

\[ z_1 \in R \quad \|z_1 - \tilde{z}_1\| = \sqrt{(z_1 - \text{Re}(\tilde{z}_1))^2 + (\text{Im}(\tilde{z}_1))^2} \to \min \quad \text{at } z_1 = \tilde{z}_1 \]
\[ z_2 \in R \quad \|z_2 - \tilde{z}_2\| = \sqrt{(z_2 - \text{Re}(\tilde{z}_2))^2 + (\text{Im}(\tilde{z}_2))^2} \to \min \quad \text{at } z_2 = \tilde{z}_2 \]

The smallest residual, however, is achieved for \( \tilde{z}_1 = 100 \) and \( \tilde{z}_2 = -100 \)

\[ \tilde{z} \in R \quad \|r(\tilde{z})\| = \|\tilde{z}^2 - 10000 - i\| = \sqrt{(\tilde{z} - 1)(\tilde{z} + 1) + 1} \to \min \quad \text{at } \tilde{z} = 100 \quad \text{or} \quad \tilde{z} = -100 \]

The best approximation is different from the residual minimization solution by \( \frac{0.000000012}{100} = 1.2 \cdot 10^{-9} \)
The Residual Theorem

Let the mapping $f(v) = w$ be an injection (a one-to-one mapping, an inverse mapping $v = f^{-1}(w)$ can be defined within the range of $f(v)$).

Let the inverse mapping be a Lipschitz mapping with constant $\mu$.

Then the approximation error for a substitution $v_s$ is bounded by

$$||v_s - \bar{v}|| \leq \mu||v_s||$$

Proof

As long as the inverse mapping exists, the exact solution can be formally written as $\bar{v} = f^{-1}(w)$.

For a substitution $v_s$

$$f(v_s) - w = r(v_s)$$

$$f(v_s) = w + r(v_s)$$

$$v_s = f^{-1}(w + r(v_s))$$

The approximation error:

$$||v_s - \bar{v}|| = ||f^{-1}(w + r(v_s)) - f^{-1}(w)||$$

For the Lipschitz mapping: $||f^{-1}(w + r(v_s)) - f^{-1}(w)|| \leq \mu||w + r(v_s) - w|| = \mu||r(v_s)||$

Thus

$$||v_s - \bar{v}|| \leq \mu||r(v_s)||$$

Note: $f^{-1}$ does not have to be a contraction mapping – no restriction on $\mu$
The Residual Theorem, Continued

Notice that

1) If $f^{-1}$ is known, there is no need for approximation. The exact solution can be found as $f^{-1}(w)$.

2) If $f^{-1}$ is unknown, it may be very hard to check if it is a Lipschitz mapping in the first place. Namely, the Lipschitz formula $||f^{-1}(v_1) - f^{-1}(v_2)|| \leq \mu ||v_1 - v_2||$ must be analyzed without knowledge of the function $f^{-1}$, just by using properties of the direct mapping $f$. 
Example

Consider the following ordinary differential equation  \( \frac{dx}{dt} + tx = y(t) \quad t \in [0, 1] \quad x(0) = x_0 \)

In the operator form
\[
\hat{f}(x(t)) = y(t), \quad \text{where} \quad \hat{f}(x(t)) = \left( \frac{d}{dt} + t \right)x(t)
\]

Is the inverse mapping \( \hat{f}^{-1} \) a Lipschitz mapping?

For the mappings \( \hat{f}(x(t)) = y(t) \) and \( x(t) = \hat{f}^{-1}(y(t)) \) compare the norms \( \|x_1(t) - x_2(t)\| \) and \( \|y_1(t) - y_2(t)\| \)

\[
\frac{dx}{dt} + tx = y(t)
\]
\[
\frac{dx}{dt} = y(t) - tx
\]
\[
\int_0^t \frac{dx}{dt} dt = \int_0^t y'(t) dt - \int_0^t t x'(t) dt
\]
\[
x_1(t) = x_0 + \int_0^t y_1'(t) dt - \int_0^t t x_1'(t) dt
\]
\[
x_2(t) = x_0 + \int_0^t y_2'(t) dt - \int_0^t t x_2'(t) dt
\]
\[
x_1(t) - x_2(t) = \int_0^t (y_1(t) - y_2(t)) dt - \int_0^t t (x_1(t) - x_2(t)) dt
\]
Example, Continued

$$\| x_1(t) - x_2(t) \| = \int_0^1 | x_1(t) - x_2(t) | dt = \int_0^1 \left| \int_0^t (y_1(t') - y_2(t')) dt' - \int_0^t (x_1(t') - x_2(t')) dt' \right| dt$$

From the triangular axiom

$$\int_0^t (y_1(t') - y_2(t')) dt' - \int_0^t (x_1(t') - x_2(t')) dt' \leq \int_0^t (y_1(t') - y_2(t')) dt' + \int_0^t (x_1(t') - x_2(t')) dt' = I_1(t) + I_2(t)$$

Finally,

$$\| x_1(t) - x_2(t) \| \leq \int_0^1 (I_1(t) + I_2(t)) dt$$

$$I_1(t) = \int_0^t | y_1(t') - y_2(t') | dt' \leq \int_0^t | y_1(t') - y_2(t') | dt' \leq \int_0^t | y_1(t') - y_2(t') | dt' = \| y_1(t) - y_2(t) \|$$

$$I_2(t) = \int_0^t | x_1(t') - x_2(t') | dt' \leq \int_0^t | x_1(t') - x_2(t') | dt' \leq \int_0^t | x_1(t') - x_2(t') | dt' = t \| x_1(t) - x_2(t) \|$$

Thus

$$\| x_1(t) - x_2(t) \| \leq \| y_1(t) - y_2(t) \| t + \| x_1(t) - x_2(t) \| \int_0^1 t dt = \| y_1(t) - y_2(t) \| + \frac{1}{2} \| x_1(t) - x_2(t) \|$$

$$\| x_1(t) - x_2(t) \| \leq 2 \| y_1(t) - y_2(t) \|$$

$$\hat{f}^{-1}$$ is a Lipschitz mapping with $$\mu = 2.$$
Approximate solution

A particular case \( y(t) = t, \ x_0 = 0 \)

The exact solution for \( \frac{dx}{dt} + tx = t \) is \( x(t) = 1 + Ce^{-\frac{t^2}{2}}, \ C = -1 \) (to satisfy \( x_0 = 0 \))

Let us pretend we don't know it and find the best approximation to the solution among the polynomials \( x_n(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_n t^n \)

Start with the linear function \( x_n(t) = \alpha + \beta t. \ x_0 = 0 \Rightarrow \alpha = 0, \ x_n(t) = \beta t. \ \beta = ? \)

\[
\frac{d}{dt}(\beta t) + t \cdot \beta - \frac{t}{y(t)} = r(t, \beta) \rightarrow \min
\]

\[
r(t, \beta) = \beta + \beta t^2 - t
\]

\[
\|r(t, \beta)\| = \sqrt{\int_0^1 (\beta + \beta t^2 - t)^2 \, dt} = \sqrt{\frac{28}{15} \beta^2 - \frac{3}{2} \beta + \frac{1}{3}}
\]

\[
\frac{d}{d\beta} \|r(t, \beta)\| = 0 \Rightarrow \frac{d}{d\beta} \|r(t, \beta)\|^2 = 0 \Rightarrow \beta = \frac{45}{112}
\]

\( x_1(t) = \frac{45}{112} t \) is the approximation to the exact solution

that provides the smallest residual among all linear functions
Incompletely Specified Equations

The equations may contain hidden variables, not available for direct analysis.

Sometimes the problem is referred to as a measurement problem. The error of a particular measurement is defined by the hidden variables. Mathematically, the equation \( f(v) = w \) is now changed to a set of equations

\[
\begin{align*}
  f(v, h_1) &= w + \varepsilon_1 \\
  f(v, h_2) &= w + \varepsilon_2 \\
  &\quad \vdots \\
  f(v, h_n) &= w + \varepsilon_n
\end{align*}
\]

In the presence of hidden variables, the correspondence between \( v \) and \( w \) is no longer a mapping. The inverse mapping also cannot be defined.

Once the hidden variables are omitted, the system of equations looks over-specified.

The exact solution may not exist at all.

Finding the best approximation is still a legitimate task: even though there are unknown hidden variables that affect the results of measurements, try to get the best from the available measurement data, find the best approximation to the solution.
Measurements, Continued

Consider a linear mapping from m-dimensional Hilbert space V into n-dimensional Hilbert space W (n > m). \( \hat{A}\tilde{v} = \bar{w} \). The corresponding system of equations is overspecified. In general, such systems are inconsistent and do not have a solution. It is possible, however, to minimize the residual:

\[
\| r \| = \| \hat{A}\tilde{v} - \bar{w} \| \rightarrow \min \quad \text{or} \quad \| r \|^2 = \langle \hat{A}\tilde{v} - \bar{w}, \hat{A}\tilde{v} - \bar{w} \rangle \rightarrow \min
\]

Find such \( \tilde{v} \) so that \( \forall \tilde{v} \in V \quad \| \hat{A}\tilde{v} - \bar{w} \| \geq \| \hat{A}\tilde{v} - \bar{w} \|
\]

The Least-Square Theorem:
The least-square solution \( \tilde{v} \) always exists, and this solution satisfies the equation \( \hat{A}^+\hat{A}\tilde{v} = \hat{A}^+\bar{w} \), where \( \hat{A}^+ = (\hat{A}^T)^* \) is the Hermitian conjugate to the matrix \( \hat{A} \).

Proof.
\( \forall \tilde{v} \in V \quad \text{find} \quad \bar{w} = \hat{A}\tilde{v} \quad \hat{A}\tilde{v} \in \mathcal{R}(\hat{A}) \) (range of the matrix \( \hat{A} \))
\( \mathcal{R}(\hat{A}) \) is a closed subset in \( W \), \( \mathcal{R}(\hat{A}) \subset W \).

According to the projection theorem,
1) \( \forall \bar{w} \in W \quad \exists! \bar{r}_0 \in \mathcal{R}(\hat{A}) \subset W \quad \| \bar{w} - \bar{r}_0 \| \leq \| \bar{w} - \bar{r} \| \quad \forall \bar{r} \in \mathcal{R}(\hat{A}) \)
2) \( (\bar{w} - \bar{r}_0) \perp \mathcal{R}(\hat{A}) \) (That is, \( \langle \bar{w} - \bar{r}_0, r \rangle = 0 \quad \forall \bar{r} \in \mathcal{R}(\hat{A}) \)
The Least Square Theorem, Continued

Substitute $\vec{r}_0 = \hat{A}\tilde{v}$ and $\vec{r} = \hat{A}\tilde{v}$

$\langle \tilde{w} - \hat{A}\tilde{v}, \hat{A}\tilde{v} \rangle = 0$

$\langle \hat{A}^+(\tilde{w} - \hat{A}\tilde{v}), \tilde{v} \rangle = 0 \quad \forall \tilde{v}$, so take $\tilde{v} = \hat{A}^+(\tilde{w} - \hat{A}\tilde{v})$

$\langle \hat{A}^+(\tilde{w} - \hat{A}\tilde{v}), \hat{A}^+(\tilde{w} - \hat{A}\tilde{v}) \rangle = 0$

$\|\hat{A}^+(\tilde{w} - \hat{A}\tilde{v})\| = 0$

$\hat{A}^+(\tilde{w} - \hat{A}\tilde{v}) = \theta$

$\hat{A}^+\hat{A}\tilde{v} = \hat{A}^+\tilde{w}$
Example: Multiple Measurements

\[
\begin{align*}
    x &= y_1 \\
x &= y_2 \\
    &\vdots \\
x &= y_n \quad \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \hat{A} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}
\end{align*}
\]

\[
\hat{A}^+ = (\hat{A}^T)^* = (1 \ 1 \ \ldots \ 1) \quad \hat{A}^+ \hat{A} = (1 \ 1 \ \ldots \ 1) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 1 + 1 + \ldots + 1 = n
\]

\[
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = y_1 + y_2 + \ldots + y_n
\]

The approximation that minimizes the residual \( \tilde{x} \)

\[
\hat{A}^+ \hat{A} \tilde{x} = \hat{A}^+ \tilde{y}
\]

\( n\tilde{x} = y_1 + y_2 + \ldots + y_n \)

\[
\tilde{x} = \frac{1}{n}(y_1 + y_2 + \ldots + y_n)
\]

Notice, even though it was not referred to directly, by the construction, \( \tilde{x} \) is such that

\[
\sqrt{(\tilde{x} - y_1)^2 + (\tilde{x} - y_2)^2 + \ldots + (\tilde{x} - y_n)^2} \rightarrow \min
\]
Another Example

\[ z = f(x, y) = \alpha x + \beta y \]

"Results of measurements" \( f(1, 2) = 1, \ f(3, 4) = 1, \ f(5, 6) = 2 \)

Or, using the hidden variable terminology

\[ z = f(x, y, h) = \alpha x + \beta y + \delta(x, y, h), \] and there is a reason to believe that \( \delta(x, y, h) \) is small

\[ f(1, 2, h_1) = 1, \ f(3, 4, h_2) = 1, \ f(5, 6, h_3) = 2 \]

\[
\begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6 \\
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\end{pmatrix}
= 
\begin{pmatrix}
1 \\
1 \\
2 \\
\end{pmatrix}
\]

This system of equations is overspecified. Moreover, it is easy to see that there is no exact solution:

\[
(3 - 1)\alpha + (4 - 2)\beta = 1 - 1 \quad 2\alpha + 2\beta = 0 \\
(5 - 3)\alpha + (6 - 4)\beta = 2 - 1 \quad 2\alpha + 2\beta = 1
\]

The least-squares approximation nevertheless exists.

\[
\hat{A} =
\begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6 \\
\end{pmatrix} \quad \hat{A}^+ = 
\begin{pmatrix}
1 & 3 & 5 \\
2 & 4 & 6 \\
\end{pmatrix} \quad \hat{A}^+ \hat{\mathbf{w}} = 
\begin{pmatrix}
1 & 3 & 5 \\
2 & 4 & 6 \\
\end{pmatrix}
\begin{pmatrix}
1 \\
2 \\
\end{pmatrix}
= 
\begin{pmatrix}
14 \\
18 \\
\end{pmatrix}
\]

\[
\hat{A}^+ \hat{A} = 
\begin{pmatrix}
1 & 3 & 5 \\
2 & 4 & 6 \\
\end{pmatrix}
\begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6 \\
\end{pmatrix}
= 
\begin{pmatrix}
35 & 44 \\
44 & 56 \\
\end{pmatrix} \quad \begin{pmatrix}
35 & 44 \\
44 & 56 \\
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\end{pmatrix}
= 
\begin{pmatrix}
14 \\
18 \\
\end{pmatrix} \quad \begin{pmatrix}
\alpha \\
\beta \\
\end{pmatrix}
= 
\begin{pmatrix}
1 \\
12 \\
\end{pmatrix}
\begin{pmatrix}
-4 \\
7 \\
\end{pmatrix}
\]
Example, continued

\[
\begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6
\end{pmatrix}
\frac{1}{12}
\begin{pmatrix}
-4 \\
7
\end{pmatrix}
= \begin{pmatrix}
5/6 \\
4/3 \\
11/6
\end{pmatrix}
\approx \begin{pmatrix}
0.833 \\
1.333 \\
1.833
\end{pmatrix}
\neq \begin{pmatrix}
1 \\
1
\end{pmatrix}
\]

However, \( \begin{pmatrix}
\alpha \\
\beta
\end{pmatrix} = \frac{1}{12}
\begin{pmatrix}
-4 \\
7
\end{pmatrix} \) provides minimization of the norm

\[
\left\| \begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta
\end{pmatrix}
- \begin{pmatrix}
1 \\
1
\end{pmatrix} \right\|
\]
Uniqueness

Is the least-square solution unique?

As a rule, no, it is not.

If nullspace $\mathcal{N}(\hat{A}) \neq \{\theta\}$, \( \tilde{v} + v_{\perp} \) will also provide minimization of the residual

Nullity $\nu(\hat{A}) = 0$ (same as rank $\rho(\hat{A}) = n$) is the criterion for uniqueness of the solution.

Theorem

The matrix $\hat{A}^+ \hat{A}$ is non-singular if, and only if, the nullspace of the matrix $\hat{A}$ is zero-dimensional.

The unique solution to the least-square approximation problem is then found as $\tilde{v} = \left(\hat{A}^+ \hat{A}\right)^{-1} \hat{A}^+ \tilde{w}$

Proof

Sufficiency.

Assume the nullspace of $\hat{A}$ is zero-dimensional.

Let $\nu$ be such element that $\hat{A}^+ \hat{A} \nu = \theta$

Take the inner product $\langle \hat{A}^+ \hat{A} \nu, \nu \rangle = \langle \theta, \nu \rangle = 0$

From the definition of the Hermitian conjugate $\langle \hat{A}^* \hat{A} \nu, \nu \rangle = \langle \hat{A} \nu, \hat{A} \nu \rangle$

$\langle \hat{A} \nu, \hat{A} \nu \rangle = 0$ means $\hat{A} \nu = \theta$, thus $\nu \in \mathcal{N}(\hat{A})$

Once the nullspace is zero-dimensional, the only element in it is $\theta$.

Then $\nu = \theta$, $\hat{A} \nu = \theta$, $\hat{A}^+ \hat{A} \nu = \theta$, the nullspace of $\hat{A}^+ \hat{A}$ is also zero-dimensional.

$\nu(\hat{A}^+ \hat{A}) = 0$ means that $\hat{A}^+ \hat{A}$ is non-singular.
Uniqueness, Continued

Necessity.

Assume that $\hat{A}^+\hat{A}$ is non-singular, $\nu(\hat{A}^+\hat{A}) = 0$.

Let vector $v'$ be a vector in a nullspace of $\hat{A}$: $\hat{A}v' = \theta$

Then $\hat{A}^+(\hat{A}v') = \hat{A}^+\theta$

$\hat{A}^+\hat{A}v' = \theta$

Thus $v'$ must belong to the nullspace of $\hat{A}^+\hat{A}$.

According to the assumption, there is only one such vector $v' = \theta$, so that thenullspace of $\hat{A}$ is also zero-dimensional.

In practice, it might be easier to check for the rank of the matrix $\left(\rho(\hat{A}) = n\right)$

rather than for the nullity $\left(\nu(\hat{A}) = 0\right)$. 
Last example, revisited

\[ \hat{A}^+ \hat{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 35 & 44 \\ 44 & 56 \end{pmatrix} \]

\[ \det(\hat{A}^+ \hat{A}) = 35 \cdot 56 - 44^2 = 24 \neq 0 \]

\[ \exists! (\hat{A}^+ \hat{A})^{-1} = \frac{1}{24} \begin{pmatrix} 56 & -44 \\ -44 & 35 \end{pmatrix} \]

And thus there is a unique solution

\[ \tilde{v} = (\hat{A}^+ \hat{A})^{-1} \hat{A}^+ w = \frac{1}{24} \begin{pmatrix} 56 & -44 \\ -44 & 35 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \frac{1}{12} \begin{pmatrix} -4 \\ 7 \end{pmatrix} \]

On the other hand, nullspace of \( \hat{A} \): \( \hat{A}v = \theta \)

\[ \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \]

\[ \begin{cases} (3+1)v_1 + (4+2)v_2 = 0 \\ 5v_1 + 6v_2 = 0 \end{cases} \]

\[ \begin{cases} 4v_1 + 6v_2 = 0 \\ 5v_1 + 6v_2 = 0 \end{cases} \]

\[ \begin{cases} v_2 = 0 \\ v_1 = 0 \end{cases} \]

The nullspace of \( \hat{A} \) contains only one element \( \theta \): \( \mathcal{N}(\hat{A}) = \{\theta\} \), \( v(\hat{A}) = 0 \)

Also, the rank of the matrix \( \hat{A} \): \( \rho(\hat{A}) = \rho \left( \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \right) = 2 \)
Pseudo-Inverse Operator

What if $\rho(\hat{A}) < n$?

The system of equations $\hat{A}\vec{v} = \vec{w}$ is under-specified.

The solution will contain free parameters: take one particular solution and add any vector from the nullspace of $\hat{A}$:

$$\hat{A}(\vec{v}_0 + \vec{v}_N) = \hat{A}\vec{v}_0 + \hat{A}\vec{v}_N = \vec{w} + \theta = \vec{w}$$

Among the solutions $\vec{v}_0 + \vec{v}_N$ (they form a closed space), there is a unique vector that has a minimal norm. Let us call it $\vec{v}_0$.

The mapping $W \rightarrow V$ that transforms a vector $\vec{w}$ into $\vec{v}_0$ is called pseudo-inverse mapping

$$\hat{A}^\#(\vec{w}) = \vec{v}_0$$

In the case of $\rho(\hat{A}) = n$, the pseudo-inverse mapping coincides with the least-square mapping.

$$\hat{A}^\# = (\hat{A}^+\hat{A})^{-1}\hat{A}^+$$

Properties of $\hat{A}^\#$

1) Linearity $\hat{A}^\#(\alpha_1\vec{w}_1 + \alpha_2\vec{w}_2) = \alpha_1\hat{A}^\#\vec{w}_1 + \alpha_2\hat{A}^\#\vec{w}_2$

2) $(\hat{A}^\#)^\# = \hat{A}$

3) $\hat{A}\hat{A}^\#\hat{A} = \hat{A}$