Rather than calculating $A^{-1}$ directly (more work than LU decomposition!), the computer instead estimates it using $O(n^2)$ algorithms.

Let's look at some properties:

What's $\text{cond}(\frac{I}{A})$? 

We have $\frac{\|Ix\|}{\|Ax\|} = \frac{\|x\|}{\|x\|} = 1$ for all $x$. 

So $M = m = 1 \implies \text{cond}(\frac{I}{A}) = 1$.

The same goes for permutation matrix $P$:

$\text{cond}(\frac{P}{A}) = 1$

Note: $\text{cond}(\frac{A}{A}) \geq 1$ by definition!

What about diagonal matrices?
\[ D \approx \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix} \]

Thus \( \text{cond}(D) = \frac{\max |d_{ii}|}{\min |d_{ii}|} \)

Ok, for an error in \( A \) we get the relation:

\[
\frac{\|Ax\|}{\|x\|} \leq \text{cond}(A) \frac{\|Ab\|}{\|b\|}
\]

What if the error is in \( \approx \) instead?

This is a bit more tricky!

Let \( x^* \) be the solution of:

\[
(\approx + E) \approx x^* = b
\]

where \( E \) is some error in \( \approx \). We expect \( \frac{\|E\|}{\|A\|} = C \cdot E_{\text{mech}} \) where
where $C$ is some $O(n)$ cost where $n$ is the matrix size.

Let's look at the residual first. We have:

$$\| b - A\hat{x}^* \| = \| E\hat{x}^* \| \leq \| E\| \| \hat{x}^* \|$$

by the definition of $\| E^* \|$

Dividing by $\| \hat{x}^* \|$, we get:

$$\frac{\| b - A\hat{x}^* \|}{\| A \| \| \hat{x}^* \|} = \frac{\| E\hat{x}^* \|}{\| A \| \| \hat{x}^* \|} \leq \frac{\| E\|}{\| A\|} = C \varepsilon_{mach}$$

so the residual is pretty small. Now for the error:

$$\tilde{A}\hat{x} - A\hat{x}^* = \tilde{b} - A\hat{x}^*$$

$$\Rightarrow \quad \hat{x} - \hat{x}^* = \tilde{A}^{-1}(\tilde{b} - A\hat{x}^*)$$

$$\| \hat{x} - \hat{x}^* \| = \| \tilde{A}^{-1}(\tilde{b} - A\hat{x}^*) \|$$
\[
\leq \|A^{-1}\|_\infty \|b - Ax^*\|_\infty \\
= \|A^{-1}\|_\infty \|E_{\infty} x^*\|_\infty \\
\leq \|A^{-1}\|_\infty \|E\|_\infty \|x^*\|_\infty \\
\]

Thus,
\[
\frac{\|x - x^*\|}{\|x^*\|} \leq \|A^{-1}\|_\infty \|E\|_\infty = \|A^{-1}\|_\infty \|A\|_\infty \sqrt{\text{cond}(A)} C \varepsilon_{\text{mach}} \\
= \text{cond}(A) C \varepsilon_{\text{mach}}
\]

So again the error is on the order of the condition, but is typically \(O(n)\) larger than for error in the RHS vector.
So far we have looked at problems:

\[ A \mathbf{x} = \mathbf{b} \]

where \( A \) is a square non-singular matrix. There are many problems where \( A \) is not square.

\[ [n \ m] = \text{size} \ (A) \]

if \( n > m \) problem is over-determined.

This usually occurs in regression:

You are fitting a curve to data points \( n \) where the number of parameters \( m \) is less than the number of data points \( n \).

Thus, there is no exact solution!

If \( n < m \), problem is under-determined.

This often happens in control: You want to reach a set temperature, and you have a heater and a cooler. You can heat it, you...
can cool it, but you can also do both at the same time!
Thus, there are an infinite of solutions!

Okay, how do we solve these problems? We shall use QR factorization - row & column reduction via orthogonal transformations.

Remember the permutation matrix? It is an example of an orthogonal matrix:

\[ P^T P = I \]

\( \sim x \sim \)

(switch rows back!)
There are many examples of orthogonal matrices:

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}, \quad \frac{1}{2} \begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}, \quad \begin{pmatrix}
\frac{3}{5} & -\frac{4}{5} \\
-\frac{4}{5} & -\frac{3}{5}
\end{pmatrix}
\]

all satisfy \( P^T P \approx I \).

An important property of orthogonal matrices is that they preserve the 2-Norm:

\[
\| P x \|_2 = \left[ (P x)^T (P x) \right]^{\frac{1}{2}}
\]

\[
= \left[ x^T P^T P x \right]^{\frac{1}{2}} = \left[ x^T x \right]^{\frac{1}{2}} = \| x \|_2
\]

This property makes orthogonal matrices useful in regression.

The regression problem was:

\[
\min_{x} \| A x - b \|_2
\]
we can multiply this by an orthogonal matrix $\mathbf{P}$:

$$\| \mathbf{A} \mathbf{x} - \frac{1}{2} \|_2^2 = \| \mathbf{P} (\mathbf{A} \mathbf{x} - \frac{1}{2}) \|_2^2$$

$$= \| (\mathbf{P} \mathbf{A}) \mathbf{x} - (\mathbf{P} \frac{1}{2}) \|_2^2$$

We want to find $\mathbf{P}$ s.t. $\mathbf{P} \mathbf{A}$ is an upper triangular form. Why??

Suppose we have a matrix $\mathbf{R}$ which is of the form:

$$\mathbf{R} = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 7 \end{pmatrix} \quad \text{and} \quad \mathbf{c} = \begin{pmatrix} 14 \\ 10 \\ 21 \end{pmatrix}$$

we want to minimize:

$$\| \mathbf{R} \mathbf{x} - \mathbf{c} \|_2 \quad \text{where} \quad \mathbf{x} \text{ is } 3 \times 1$$
Note that we can solve the first three equations exactly and that no $\bar{x}$ has any effect on the last 2!

\[
\| R \bar{x} - c \|_2^2 = \left\| \begin{pmatrix} 3 & 4 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} 14 \\ 10 \\ 21 \end{pmatrix} \right\|_2^2
\]

\[
+ \| 6 \|_2^2
\]

We can solve the first part via back substitution:

\[
x_3 = \frac{21}{7} = 3, \quad x_2 = \frac{10 - 6}{2} = 2,
\]

\[
x_1 = \frac{14 - 3 - 8}{3} = 1
\]

The residual is just:

\[
\| c \|_2^2 = \sqrt{40}
\]

So this sort of problem is easy to solve!
How do we find some matrix $P$ s.t. $PA \sim R$ in upper triangular form?? We use Householder Transformations!

You use a sequence of orthogonal transformations $P$ which reduces $A$ to upper triangular form one column at a time. Thus:

$$P_n \ldots P_3 P_2 P_1 A = R$$

Because $P_i$ is orthogonal we have:

$$(P_n \ldots P_2 P_1)^T R = P_n^T P_2^T \ldots P_1^T P_2 P_1 A \sim_I$$

$$= A$$

We define $Q = (P_n \ldots P_2 P_1)^T$