Chapter 6

Pseudo inverse of A:

$$A^{+} = (A^{*}A)^{-1}A^{*} \tag{6.1}$$

Condition number of $b = Ax$.

$$\kappa(A) = \|A\| \cdot \left\|A^{+}\right\| = \frac{\sigma_{\max}}{\sigma_{\min}} \tag{6.2}$$

Condition number of $(A^{*}A)$: (normal equation)

$$\kappa^{2}(A) \tag{6.3}$$

6.1 Floating Point Arithmetic

$$\begin{aligned}
\beta &= \textit{radix} \qquad (\textit{usually } 2) && (6.4)\\
t &= \textit{precision} \qquad (\textit{usually } 24 \textit{ or } 53 \textit{ in single/double precision}) && (6.5)\\
x &= \pm\left(\frac{m}{\beta^{t}}\right)\beta^{e} && (6.6)\\
m &: \textit{integer} \qquad \beta^{t-1} \leq m \leq \beta^{t+1} && (6.7)\\
e &: \textit{integer} && (6.8)
\end{aligned}$$

- Machine epsilon:

$$\epsilon_{\text{machine}} = \text{half a unit in the last place} = \frac{1}{2}\frac{1}{\beta^{t}}. \tag{6.9}$$

- ± 0 (the need for sign of zero)

- Floating point rounding operator

$$\forall x \in \mathbb{R}, \exists \epsilon, |\epsilon| \leq \epsilon_{\text{machine}} : fl(x) = x(1+\epsilon) \tag{6.10}$$

The distance between x and the closest floating point number is less than $\epsilon_{\text{machine}}$, i.e., less than $\frac{1}{2}$ unit in last place.

- For all practical purposes we say that the result of any floating point operation conforms to:

$$fl(x \odot y) = (x \odot y)(1+\delta) \tag{6.11}$$

where, $|\delta| \leq \epsilon_{\text{machine}}$

- Infinity ($\pm\infty$)
- Double precision floating point numbers

1	11	52
sign	exponent	fraction

exponent	fraction	
0	0	$\pm$ 0
0	$\neq 0$	subnormal
11...1	0	$\pm$ infinity
11...1	$\neq 0$	NAN

Figure 6.1: FloatingPoint.