Algorithms and Data Structures

Numerics
Condition in systems of linear equations (LES)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$

Learning goals

- Matrix multiplication
- LES
- Sherman-Morrison formula
- Woodbury formula

CONDITION NUMBER: MATRIX MULTIPLICATION

Although we are actually interested in $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, we first consider the direct problem: $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Since

$$y_i = \sum_{j=1}^{n} A_{ij} x_j,$$

it is to be expected that the high condition number of the addition is also transferred to $\mathbf{A}\mathbf{x}$.

CONDITION NUMBER: MATRIX MULTIPLICATION / 2

When $\mathbf{A}$ and $\mathbf{x}$ are disturbed by $\Delta\mathbf{A}$ and $\Delta\mathbf{x}$, the **absolute error** in the result $\mathbf{y}$ is:

$$
\begin{aligned}
\mathbf{y} + \Delta\mathbf{y} &= (\mathbf{A} + \Delta\mathbf{A})(\mathbf{x} + \Delta\mathbf{x}) \\
&= \mathbf{A}\mathbf{x} + \Delta\mathbf{A}\mathbf{x} + \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\Delta\mathbf{x} \qquad \mid -\, \mathbf{y} \\
\Delta\mathbf{y} &= \Delta\mathbf{A}\mathbf{x} + \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\Delta\mathbf{x}
\end{aligned}
$$

The absolute error is therefore estimated as follows

$$
\begin{aligned}
\rightarrow \quad \|\Delta\mathbf{y}\| &= \|\Delta\mathbf{A}\mathbf{x} + \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\Delta\mathbf{x}\| \\
&\leq \|\Delta\mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta\mathbf{x}\| + \|\Delta\mathbf{A}\|\|\Delta\mathbf{x}\| \\
&\approx \|\Delta\mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta\mathbf{x}\|
\end{aligned}
$$

CONDITION NUMBER: MATRIX MULTIPLICATION / 3

From the estimate for the absolute error, we obtain

$$
\begin{aligned}
\|\Delta \mathbf{y}\| &\leq \|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta \mathbf{x}\| \\
&= (\|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta \mathbf{x}\|)\frac{\|\mathbf{x}\|}{\|\mathbf{x}\|} \\
&= (\|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta \mathbf{x}\|)\frac{\|\mathbf{A}^{-1}\mathbf{y}\|}{\|\mathbf{x}\|} \\
&\leq (\|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta \mathbf{x}\|)\frac{\|\mathbf{A}^{-1}\|\|\mathbf{y}\|}{\|\mathbf{x}\|} \\
&= \left(\frac{\|\mathbf{A}\|}{\|\mathbf{A}\|}\|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\Delta \mathbf{x}\|\right)\frac{\|\mathbf{A}^{-1}\|\|\mathbf{y}\|}{\|\mathbf{x}\|} \\
&= \left(\frac{\|\mathbf{A}\|\|\mathbf{A}^{-1}\|}{\|\mathbf{A}\|\|\mathbf{x}\|}\|\Delta \mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|}\right)\|\mathbf{y}\| \\
\frac{\|\Delta \mathbf{y}\|}{\|\mathbf{y}\|} &\leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\left(\frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} + \frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|}\right).
\end{aligned}
$$

CONDITION NUMBER: MATRIX MULTIPLICATION / 4

The condition number

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$$

describes the propagation of relative errors both in the matrix and on the right side of the LES.

Example:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1+\epsilon \end{pmatrix}, \quad \mathbf{x}^T = \begin{pmatrix} 1 & -1 \end{pmatrix}$$

$$\|\mathbf{A}\|_\infty = 2+\epsilon, \quad \mathbf{A}^{-1} = \frac{1}{1+\epsilon-1}\begin{pmatrix} 1+\epsilon & -1 \\ -1 & 1 \end{pmatrix}, \quad \|\mathbf{A}^{-1}\|_\infty = \frac{2+\epsilon}{\epsilon}$$

$$\kappa(\mathbf{A}) = (2+\epsilon)\frac{2+\epsilon}{\epsilon} = \frac{4+4\epsilon+\epsilon^2}{\epsilon} \approx \frac{4}{\epsilon} \quad \text{and} \quad \epsilon \ll 1.$$

$\epsilon = 10^{-8}$ and machine epsilon $\texttt{eps} = 10^{-16}$ in **x** result in a relative precision of just about $4/10^{-8} \cdot 10^{-16} = 4 \cdot 10^{-8}$ in **y**.

CONDITION NUMBER: LES

The solution of the LES $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ is (numerically) equivalent to the results of the matrix multiplication. The following applies

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \kappa(\mathbf{A}^{-1}).$$

Note:

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| \geq \|\mathbf{A}\mathbf{A}^{-1}\| = \|\mathbf{I}\| = 1.$$

Example 1:

Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 + 10^{-10} & 1 - 10^{-10} \end{pmatrix}$

The matrix is ill-conditioned with condition number $\kappa \approx 2 \times 10^9$.
For the solution of the LES $\mathbf{Ax} = \mathbf{y}$ this means: a small variation in the input data (e.g. $\mathbf{y} = (1, 1) \to \tilde{\mathbf{y}} = (1, 1.00001)$) leads to a big change in the solution.

CONDITION NUMBER: LES / 2

```
A = matrix(c(1, 1 + 10e-10, 1, 1 - 10e-10), nrow = 2)
y = c(1, 1)
yt = c(1, 1.00001)
```

```
solve(A, y)
## [1] 0.4999999722444252 0.5000000277555748
```

```
solve(A, yt)
## [1] 5000.499858862901 -4999.499858862901
```

CONDITION NUMBER: LES / 3

Example 2: The Hilbert matrix is known to be ill-conditioned!

$$H_{ij} = \frac{1}{i + j - 1},$$

```
hilbert = function(n) {
  i = 1:n
  return(1 / outer(i - 1, i, "+"))
}
hilbert(4)
```

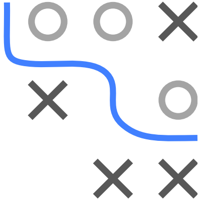
CONDITION NUMBER: LES / 4

```
##                   [,1]              [,2]
## [1,] 1.000000000000000 0.500000000000000
## [2,] 0.500000000000000 0.333333333333333
## [3,] 0.333333333333333 0.250000000000000
## [4,] 0.250000000000000 0.200000000000000
##                   [,3]              [,4]
## [1,] 0.333333333333333 0.250000000000000
## [2,] 0.250000000000000 0.200000000000000
## [3,] 0.200000000000000 0.166666666666667
## [4,] 0.166666666666667 0.142857142857143
```

CONDITION NUMBER: LES / 5

```
foo = function(n) {
  cond = sapply(n, function(i) {
    norm(hilbert(i)) * norm(solve(hilbert(i)))
  })
  return(cbind(n, cond))
}
foo(4:10)
##       n                 cond
## [1,] 4 2.837499999999738e+04
## [2,] 5 9.436559999999363e+05
## [3,] 6 2.907027900294877e+07
## [4,] 7 9.851948897194694e+08
## [5,] 8 3.387279082022739e+10
## [6,] 9 1.099650993366049e+12
## [7,] 10 3.535372424347476e+13
```

WELL- VS. ILL-POSED PROBLEMS



A problem is called well-posed if the following holds:

- There exists a solution for the problem
- The existing solution is unique
- The solution depends continuously on the condition of the problem (stable)

A problem is called ill-posed if it violates at least one of these properties. However, the instability of solutions usually causes the most difficulties.

"DO NOT INVERT THAT MATRIX"

Important: Never solve an LES (numerically) using $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

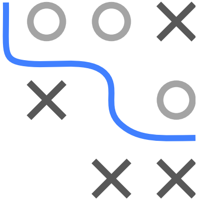
```
solve(A) %*% y
```

Although theoretically correct, internally **two** (possibly ill-posed) problems are solved:

- Inversion of $\mathbf{A}$ (solution of $\mathbf{Ax} = \mathbf{0}$) has a condition of $\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$
- The multiplication of $\mathbf{A}^{-1} \cdot \mathbf{y}$ has a condition of $\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$

The condition inflates: $\|\mathbf{A}\|^2\|\mathbf{A}^{-1}\|^2$

"DO NOT INVERT THAT MATRIX" / 2



Better: Solve directly by

```
solve(A, y)
```

Advantages:

- **Stability**: In the worst case only **one** ill-posed subproblem is solved.
- **Memory**: The n^2 entries of the inverted matrix $\boldsymbol{A}^{-1}$ must be saved. With a direct solution via the LES only $\boldsymbol{x} \in \mathbb{R}^n$ is stored ($\frac{1}{n}$-th of storage space).

Systems of equations can be solved efficiently and numerically stable by means of matrix decompositions (more on this in chapter 7 - matrix decompositions).

SHERMAN-MORRISON FORMULA

If a matrix $\boldsymbol{X}$ can be represented by $\boldsymbol{X} = \mathbf{A} + \boldsymbol{u}\boldsymbol{v}^T$, $\boldsymbol{X}^{-1}$ can be calculated using the **Sherman-Morrison formula** as follows:

$$\boldsymbol{X}^{-1} = (\mathbf{A} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}}{1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}}$$

Proof:

$$\begin{aligned}
\boldsymbol{X} \cdot \boldsymbol{X}^{-1} &= (\mathbf{A} + \boldsymbol{u}\boldsymbol{v}^T)(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}}{1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}}) \\
&= \mathbf{A}\mathbf{A}^{-1} + \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \frac{\mathbf{A}\mathbf{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}}{1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}} \\
&= \boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \frac{\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1}}{1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}} \\
&= \boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \frac{\boldsymbol{u}(1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u})\boldsymbol{v}^T\mathbf{A}^{-1}}{1 - \boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{u}} \\
&= \boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} - \boldsymbol{u}\boldsymbol{v}^T\mathbf{A}^{-1} = \boldsymbol{I}
\end{aligned}$$

WOODBURY FORMULA

If a matrix $\boldsymbol{X}$ can be represented by $\boldsymbol{X} = \mathbf{A} + \boldsymbol{UCV}$, $\boldsymbol{X}^{-1}$ can be calculated using the **Woodbury formula** as:

$$\boldsymbol{X}^{-1} = (\mathbf{A} + \boldsymbol{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\mathbf{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\mathbf{A}^{-1}$$

The formula is especially useful if $\mathbf{A}^{-1}$ is very easy to calculate or has already been calculated.

The Woodbury formula is often used in optimization (low-rank updates, BFGS updates). See Chapter 10 (Multivariate Optimization) for more information.