SOLUTION OF SYSTEMS OF LINEAR EQUATIONS

Notes prepared for EE 6481

by

Professor Cyrus D. Cantrell

May–August 2005

© C. D. Cantrell (02/2005)

OUTLINE

- Direct methods
  - ▷ Gaussian elimination
    - ◦ Back-substitution in upper-triangular matrices
    - ◦ Forward elimination in general square matrices
  - ▷ LU and Cholesky decompositions
- Condition number
  - ▷ Example: Linear least-squares problem
- Iterative methods for large linear systems
  - ▷ Contraction mapping theorem
  - ▷ Gauss-Jordan iteration
  - ▷ Methods for accelerating convergence
  - ▷ Conguate-gradient methods
  - ▷ Krylov-subspace methods

© C. D. Cantrell (02/2005)

SYSTEM OF LINEAR EQUATIONS

- A system of linear equations

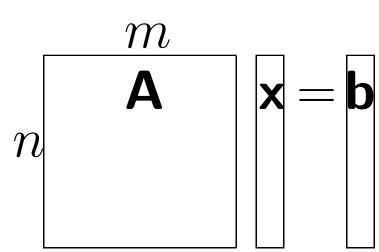
$$
\begin{aligned}
a^1_1 x^1 + a^1_2 x^2 + \cdots + a^1_m x^m &= b^1 \\
&\vdots \\
a^n_1 x^1 + a^n_2 x^2 + \cdots + a^n_m x^m &= b^n
\end{aligned}
$$

in the m unknowns x^1, ... , x^m can be expressed as the matrix equation

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A}$ is the $n \times m$ **coefficient matrix**

- Block diagram when $m = n$:



© C. D. Cantrell (05/2005)

CLASSIFICATION OF LINEAR-EQUATION PROBLEMS

- In matrix form, a system of linear equations is

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

  where $\mathbf{A}$ is the coefficient matrix

- **Direct problem**: We know $\mathbf{A}$ and $\mathbf{x}$ and must find $\mathbf{b}$
  - ▷ Computationally, this is a GAXPY with $\mathbf{y} = \mathbf{0}$
- **Inverse problem**: We know $\mathbf{A}$ and $\mathbf{b}$ and must find the solution vector $\mathbf{x}$
  - ▷ A solution exists only if

$$\mathbf{b} \in \mathrm{range}\,[\mathbf{A}]$$

    If $\mathbf{b} \notin \mathrm{range}\,[\mathbf{A}]$, the linear system is inconsistent and there is no solution

- **Fitting problem**: We know $\mathbf{A}$ and $\mathbf{b}$ and must find a solution vector $\mathbf{x}$ that minimizes the norm $\|\mathbf{r}\|$ of the residual vector

$$\mathbf{r} := \mathbf{b} - \mathbf{A}\mathbf{x}$$

© C. D. Cantrell (02/2005)

UNDERDETERMINED SYSTEMS (1)

- A system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

is **underdetermined** if

$$\mathbf{x} \in \mathcal{V} \text{ where } \dim[\mathcal{V}] > \operatorname{rank}[\mathbf{A}]$$

▷ Example: The single equation $ax + by = c$ does not determine x and y uniquely. In this example,

$$\mathbf{A} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix},$$

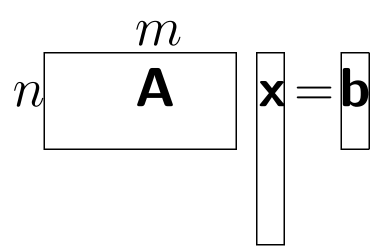
$$\mathcal{V} = \mathbb{R}^2, \quad \operatorname{rank}[\mathbf{A}] = 1$$

▷ The high-school criterion of "more unknowns than equations" translates to rank $[\mathbf{A}]$ (=1 in this example) being too small for the dimension of the space to which $\mathbf{x}$ belongs (=2 in this example)

© C. D. Cantrell (05/2005)

UNDERDETERMINED SYSTEMS (2)

- Block diagram of an underdetermined system:



▷ Here

$$\mathbf{x} \in \mathcal{V} \text{ where } \dim[\mathcal{V}] = m$$

and

$$\operatorname{rank}[\mathbf{A}] \leq n < m$$

© C. D. Cantrell (05/2005)

RANK-NULLITY THEOREM

- The rank-nullity theorem states that

$$\text{rank}\,[\mathbf{A}] + \text{nullity}\,[\mathbf{A}] = \dim\,[\text{domain}\,[\mathbf{A}]]$$

where the **nullity** of $\mathbf{A}$ is the dimension of the null space $(\dim\,[\text{null}\,[\mathbf{A}]])$

▷ Example: If

$$\mathbf{A} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \quad \Rightarrow \quad \text{rank}\,[\mathbf{A}] = 1$$

then

$$\text{nullity}\,[\mathbf{A}] = 1$$

▷ The 1-dimensional null space in this example is spanned by the vector

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ -a/b \end{pmatrix},$$

provided that $b \neq 0$

© C. D. Cantrell (05/2005)

UNDERDETERMINED SYSTEMS (3)

- If

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

is underdetermined, then the rank-nullity theorem guarantees that

$$\text{nullity}\,[\mathbf{A}] > 0$$

▷ The **homogeneous system of equations**

$$\mathbf{A}\mathbf{x}_0 = \mathbf{0}$$

has non-trivial solutions $\mathbf{x}_0 \in \text{null}\,[\mathbf{A}]$

▷ The most general solution of the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_0$$

where $\mathbf{x}_p$ is a particular solution, i.e.,

$$\mathbf{A}\mathbf{x}_p = \mathbf{b},$$

and $\mathbf{x}_0 \in \text{null}\,[\mathbf{A}]$

© C. D. Cantrell (05/2005)

OVERDETERMINED SYSTEMS (1)

- A system of linear equations

$$\mathbf{Ax} = \mathbf{b}$$

is **overdetermined** if

$$\mathbf{b} \in \mathcal{V}' \text{ where } \dim[\mathcal{V}'] > \operatorname{rank}[\mathbf{A}]$$

▷ Example: The system

$$\begin{pmatrix} a^1_1 & a^1_2 \\ a^2_1 & a^2_2 \\ a^3_1 & a^3_2 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix}$$

is overdetermined, because the row rank of $\mathbf{A}$ is 2 and $\mathbf{b} \in \mathbb{R}^3$

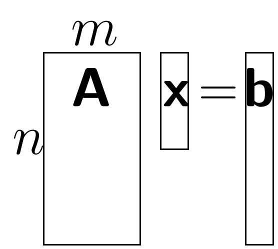
▷ The high-school criterion of "more equations than unknowns" translates to $\operatorname{rank}[\mathbf{A}]$ ($= 2$ in this example) being too small for the dimension of the space to which $\mathbf{b}$ belongs ($= 3$ in this example)

▷ An overdetermined system leads to a fitting problem; see notes on least squares

© C. D. Cantrell (05/2005)

OVERDETERMINED SYSTEMS (2)

- Block diagram of an overdetermined system:



▷ Here

$$\mathbf{b} \in \mathcal{V}' \text{ where } \dim\left[\mathcal{V}'\right] = n$$

and

$$\operatorname{rank}\left[\mathbf{A}\right] \leq m < n$$

© C. D. Cantrell (05/2005)

THE INVERSE PROBLEM

- In the system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

we know $\mathbf{A}$ and $\mathbf{b}$ and must find $\mathbf{x}$

- ▷ Assume that the system is not inconsistent ($i.e.$, $\mathbf{b} \in \text{range}\,[\mathbf{A}]$)
- ▷ Also assume that $\mathbf{A}$ is non-singular
- ▷ If $\mathbf{A}$ is square and non-singular, then the unique solution vector is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

  - ◦ The direct evaluation of $\mathbf{A}^{-1}$ using Cramer's rule requires $O(n \times n!)$ operations, and is computationally ill-advised unless $\mathbf{A}$ is 2×2
  - ◦ Gaussian elimination or iterative methods are recommended

© C. D. Cantrell (06/2005)

GAUSSIAN ELIMINATION (1)

- Upper-triangular matrices:
  - ▷ Determinant = product of main-diagonal elements:

$$\det[\mathbf{U}] = u^1_1 \cdots u^n_n$$

  - ▷ Solution of the linear system

$$\mathbf{Ux} = \mathbf{b}$$

    is obtained by back-substitution, starting with the last row, $u^n_n\, x^n = b^n$
  - ▷ General row:

$$x^i = \frac{b^i - \displaystyle\sum_{k=i+1}^{n} u^i_k\, x^k}{u^i_i}$$

  - ▷ Strategy of Gaussian elimination: Transform a general coefficient matrix $\mathbf{A}$ into an upper-triangular matrix by a process of forward elimination

© C. D. Cantrell (05/2005)

GAUSSIAN ELIMINATION (2)

- Lower-triangular matrices:
  - ▷ Determinant = product of main-diagonal elements:

$$\det[\mathbf{L}] = l^1_1 \cdots l^n_n$$

  - ▷ Solution of the linear system

$$\mathbf{Lx} = \mathbf{b}$$

    is obtained by forward-substitution, starting with the first row, $l^1_1\, x^1 = b^1$
  - ▷ Strategy for efficiently making repeated Gaussian-elimination solutions of $\mathbf{Ax} = \mathbf{b}$ with different right-hand sides $\mathbf{b}$:
    Express $\mathbf{A}$ as the product of lower- and upper-triangular matrices,

$$\mathbf{A} = \mathbf{LU}$$

    Solve $\mathbf{Ly} = \mathbf{b}$ by forward substitution
    Solve $\mathbf{Ux} = \mathbf{y}$ by back-substitution

© C. D. Cantrell (05/2005)

GAUSSIAN ELIMINATION (3)

- Goal: Transform the coefficient matrix $\mathbf{A}$ to upper-triangular form
- Start with the system

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

  ▷ Subtract l_1^j times the first equation from the j^{th} equation for $j = 2, \ldots, n$, where

$$l_1^j = \frac{a_1^j}{a_1^1}$$

  ▷ Equivalently, premultiply both $\mathbf{A}$ and $\mathbf{b}$ with the unit lower-triangular matrix

$$\mathbf{L}^{(1)} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -l_1^2 & 1 & 0 & \cdots & 0 \\ -l_1^3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -l_1^n & 0 & 0 & \cdots & 1 \end{pmatrix}$$

© C. D. Cantrell (05/2005)

GAUSSIAN ELIMINATION (4)

- Transformed system of linear equations:

$$\mathbf{L}^{(1)}\mathbf{A}\mathbf{x} = \mathbf{L}^{(1)}\mathbf{b}$$

  ▷ The transformed coefficient matrix

$$\mathbf{A}^{(2)} = \mathbf{L}^{(1)}\mathbf{A}$$

  has zeros in the first column below the first row:

$$\mathbf{A}^{(2)} = \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 & \cdots & a_n^1 \\ 0 & a_2^{(2)2} & a_3^{(2)2} & \cdots & a_n^{(2)2} \\ 0 & a_2^{(2)3} & a_3^{(2)3} & \cdots & a_n^{(2)3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_2^{(2)n} & a_3^{(2)n} & \cdots & a_n^{(2)n} \end{pmatrix}$$

  where

$$a_i^{(2)j} := a_i^j - l_1^j\, a_i^1$$

GAUSSIAN ELIMINATION (5)

- If the original coefficient matrix $\mathbf{A}$ is non-singular, then the transformed coefficient matrix $\mathbf{A}^{(2)}$ is also non-singular, because

$$\det\left[\mathbf{L}^{(1)}\right] = 1$$

where

$$\mathbf{L}^{(1)} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -l_1^2 & 1 & 0 & \cdots & 0 \\ -l_1^3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -l_1^n & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad l_1^j = \frac{a_1^j}{a_1^1}$$

and therefore

$$\det\left[\mathbf{A}^{(2)}\right] = \det\left[\mathbf{L}^{(1)}\right]\det[\mathbf{A}] = \det[\mathbf{A}] \neq 0$$

© C. D. Cantrell (05/2005)

GAUSSIAN ELIMINATION (6)

- If the original coefficient matrix $\mathbf{A}$ is non-singular, continue for $k = 2, \ldots, n$ to eliminate x^k from rows $k+1, \ldots, n$, finally obtaining the transformed system

$$\mathbf{U}\mathbf{x} = \mathbf{b}^{(n)}$$

where

$$\mathbf{U} := \mathbf{A}^{(n)} = \mathbf{L}'\mathbf{A} \quad \text{and} \quad \mathbf{b}^{(n)} := \mathbf{L}'\mathbf{b}$$

where

$$\mathbf{L}' := \mathbf{L}^{(n-1)} \cdots \mathbf{L}^{(1)}$$

and

$$\mathbf{L}^{(k)} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ 0 & \cdots & -l_k^{k+1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & -l_k^n & \cdots & 1 \end{pmatrix}, \quad l_k^j = \frac{a_k^{(k)j}}{a_k^{(k)k}}$$

© C. D. Cantrell (05/2005)

GAUSSIAN ELIMINATION (7)

- In the transformed system

$$\mathbf{U}\mathbf{x} = \mathbf{b}^{(n)}$$

the transformed coefficient matrix $\mathbf{U}$ is upper-triangular:

$$\mathbf{U} = \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 & \cdots & a_n^1 \\ 0 & a_2^{(2)2} & a_3^{(2)2} & \cdots & a_n^{(2)2} \\ 0 & 0 & a_3^{(3)3} & \cdots & a_n^{(3)3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n^{(n)n} \end{pmatrix}$$

where $\forall\, i \geq k : \forall\, j > k : \; a_i^{(k+1)j} := a_i^{(k)j} - l_k^j \, a_i^{(k)k}$ and $l_k^j := \dfrac{a_k^{(k)j}}{a_k^{(k)k}}$

- ▷ The diagonal elements $a_k^{(k)k}$ are called **pivots**
- ▷ The system $\mathbf{U}\mathbf{x} = \mathbf{b}^{(n)}$ can be solved by back-substitution

© C. D. Cantrell (05/2005)

LU DECOMPOSITION (1)

- Goal: Express $\mathbf{A}$ as the product of a lower-triangular and an upper-triangular matrix, $\mathbf{A} = \mathbf{LU}$
- To obtain $\mathbf{L}'$ (where $\mathbf{L}'\mathbf{A} = \mathbf{U}$):
  - ▷ Augment $\mathbf{A}$ with the $n \times n$ identity matrix to form an $n \times 2n$ matrix

$$(\mathbf{A}, \mathbf{1})$$

  - ▷ Apply Gaussian elimination to the augmented matrix; the result is

$$(\mathbf{U}, \mathbf{L}')$$

  - ▷ In the next slide we show how to compute

$$\mathbf{L} = (\mathbf{L}')^{-1}$$

  - ▷ From this equation and $\mathbf{L}'\mathbf{A} = \mathbf{U}$ it follows that

$$\mathbf{A} = \mathbf{LU}$$

© C. D. Cantrell (05/2005)

LU DECOMPOSITION (2)

- The inverse of

$$\mathbf{L}' = \mathbf{L}^{(n-1)} \cdots \mathbf{L}^{(1)}$$

is

$$\mathbf{L} = (\mathbf{L}^{(1)})^{-1} \cdots (\mathbf{L}^{(n-1)})^{-1}$$

$$(\mathbf{L}^{(k)})^{-1} = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & l_k^{k+1} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & l_k^n & 0 & \cdots & 1 \end{pmatrix} \Rightarrow \mathbf{L} = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ l_1^k & \cdots & 1 & 0 & \cdots & 0 \\ l_1^{k+1} & \cdots & l_k^{k+1} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ l_1^n & \cdots & l_k^n & l_{k+1}^n & \cdots & 1 \end{pmatrix}$$

© C. D. Cantrell (05/2005)

LU DECOMPOSITION (3)

- Solution for p different right-hand sides $\mathbf{b}_1, \ldots, \mathbf{b}_p$:
  - ▷ Augment $\mathbf{A}$ with the right-hand sides to form an $n \times (n+p)$ matrix

$$(\mathbf{A}, \mathbf{b}_1, \ldots, \mathbf{b}_p) = (\mathbf{LU}, \mathbf{b}_1, \ldots, \mathbf{b}_p)$$

  - ▷ Apply Gaussian elimination to the augmented matrix (equivalently, multiply the augmented matrix by $\mathbf{L}$)
    - ∘ Forward elimination is performed only once
  - ▷ The result is

$$\left(\mathbf{U}, \mathbf{L}'\mathbf{b}_1, \ldots, \mathbf{L}'\mathbf{b}_p\right)$$

  - ▷ The systems

$$\mathbf{Ux} = \mathbf{L}'\mathbf{b}_1, \ldots, \mathbf{Ux} = \mathbf{L}'\mathbf{b}_p$$

    can be solved by back-substitution, requiring much fewer operations than forward elimination as long as $p \ll n$

© C. D. Cantrell (05/2005)

OPERATION COUNT FOR GAUSSIAN ELIMINATION (1)

- Operation count for forward elimination applied to the $n \times (n+p)$ matrix $(\mathbf{A}, \mathbf{b}_1, \ldots, \mathbf{b}_p)$:
  - ▷ For a given value of $k \in [1, n]$, calculating the multipliers

$$l_k^j = \frac{a_k^{(k)j}}{a_k^{(k)k}} \quad \text{where} \quad j \in [k+1, n]$$

    requires $(n-k)$ floating-point divisions
  - ▷ For a given value of $k \in [1, n]$, calculating the new matrix elements

$$a_i^{(k+1)j} = a_i^{(k)j} - l_k^j \, a_i^{(k)k}$$

    requires $(n-k)(n-k+p)$ floating-point multiplications and additions (number of rows $\times$ number of columns)
  - ▷ Approximate number of operations for forward elimination for large n:

$$N_f \approx \sum_{k=1}^{n-1} (n-k)(n-k+p)$$

© C. D. Cantrell (05/2005)

OPERATION COUNT FOR GAUSSIAN ELIMINATION (2)

- Operation count for forward elimination applied to $(\mathbf{A}, \mathbf{b}_1, \ldots, \mathbf{b}_p)$:

$$N_f \approx \sum_{k=1}^{n-1}(n-k)(n-k+p) = \sum_{k=1}^{n-1}(n-k)^2 + p\sum_{k=1}^{n-1}(n-k)$$

  ▷ Use the formulas

$$\sum_{k=1}^{m} k = \frac{m(m+1)}{2}, \quad \sum_{k=1}^{m} k^2 = \frac{n(m+1)(2m+1)}{6}$$

  to get the result

$$N_f \approx \frac{n(n-1)(2n-1)}{6} + p\frac{n(n-1)}{2} \approx \frac{n^3}{3}$$

  ▷ The operation count for back-substitution is

$$N_b \approx \frac{n(n-1)}{2} \ll N_f$$

© C. D. Cantrell (05/2005)

COMPUTATION OF THE DETERMINANT

- The determinant of $\mathbf{A}$ is the product of the pivots:

$$\det[\mathbf{A}] = \det[\mathbf{U}] = a_1^1 a_2^{(2)2} a_3^{(3)3} \cdots a_n^{(n)n}$$

  - ▷ If R row interchanges are required in addition to multiplication by $\mathbf{L}' := \mathbf{L}^{(n-1)} \cdots \mathbf{L}^{(1)}$, then the determinant is multiplied by $(-1)^R$
  - ▷ This reduces the number of floating-point operations required to compute the determinant from $(n-1)n!$ to n plus the operation count for Gaussian elimination

© C. D. Cantrell (05/2005)