CSE/Math 455
Lecture # 10

Algorithms for back and forward substitution

Back substitution solves a linear system of the form

$$\begin{aligned} u_{11}x_1 + \quad u_{12}x_2 + \quad u_{13}x_3 &= y_1 \\ u_{22}x_2 + \quad u_{23}x_3 &= y_2 \\ u_{33}x_3 &= y_3 \end{aligned}$$

assuming $u_{11}, u_{22}, u_{33} \neq 0$.

For a 3×3 matrix this takes the form

$$\begin{aligned} x_3 &= y_3/u_{33} \\ x_2 &= (y_2 - u_{23}x_3)/u_{22} \\ x_1 &= (y_1 - u_{12}x_2 - u_{13}x_3)/u_{11} \end{aligned}$$

In the form of an algorithm it is

$x_n = y_n/u_{nn}$;
for $i = n-1{:}\,{-1}{:}\,1$
    $x_i = (y_i - \sum_{j=i+1}^{n} u_{ij}x_j)/u_{ii}$;
end

This is a rowwise algorithm and the sum can be replaced by one MATLAB statement! The algorithm can also be written columnwise. I will let you figure that out for yourselves.

Forward substitution is similar. The matrix is lower triangular. In the 3×3 case we are solving

$$\begin{aligned} \ell_{11}y_1 \qquad\qquad\qquad\qquad &= b_1 \\ \ell_{21}y_1 \quad + \ell_{22}y_2 \qquad\qquad &= b_2 \\ \ell_{31}y_1 \quad + \ell_{32}y_2 \quad + \ell_{33}y_3 &= b_3 \end{aligned}$$

The solution is, of course,

$$\begin{aligned} y_1 &= b_1/\ell_{11} \\ y_2 &= (b_2 - \ell_{21}y_1)/\ell_{22} \\ y_3 &= (b_3 - \ell_{31}y_1 - \ell_{32}y_2)/\ell_{33} \end{aligned}$$

For our application (Gaussian elimination) $\ell_{ii} = 1$ for all i.

The big topic today is NORMS and inner products. This is not in your book, look in www.cse.psu.edu/~barlow/cse455/norm_notes.pdf

Definition 1 *An* inner product *in* R^n *is a function* $(\cdot,\cdot)$ *mapping* $\mathrm{R}^n \times \mathrm{R}^n$ *into* R^n *that satisfies the following four axioms*

1. $(\mathbf{x},\mathbf{x}) \geq 0$;
   $(\mathbf{x},\mathbf{x}) = 0$, *if and only if* $\mathbf{x} = 0, \quad \mathbf{x} \in \mathrm{R}^{\mathbf{n}}$,

2. $(\alpha\mathbf{x},\mathbf{y}) = \alpha(\mathbf{x},\mathbf{y}), \quad \mathbf{x},\mathbf{y} \in \mathrm{R}^n, \alpha \in \mathrm{R}$,

3. $(\mathbf{x},\mathbf{y}) = (\mathbf{y},\mathbf{x}), \quad \mathbf{x},\mathbf{y} \in \mathrm{R}^n$,

4. $(\mathbf{x}+\mathbf{z},\mathbf{y}) = (\mathbf{x},\mathbf{y}) + (\mathbf{z},\mathbf{y}), \quad \mathbf{x},\mathbf{y},\mathbf{z} \in \mathrm{R}^n$.

We note that if we define an inner product for complex vectors, the third axiom becomes

$$(\mathbf{x},\mathbf{y}) = \overline{(\mathbf{y},\mathbf{x})} \quad \mathbf{x},\mathbf{y} \in \mathbf{C}^n \tag{1}$$

where $\overline{a}$ denotes the complex conjugate of a.

The inner product that is used most often is the Euclidean dot product

$$(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

Others take the form

$$(\mathbf{x},\mathbf{y}) = \mathbf{x}^T A\mathbf{y} \tag{2}$$

for an $A \in \mathrm{R}^{n\times n}$ that is symmetric and positive definite. Properties (2) and (4) imply that an inner product must have the form (2) for some matrix A. Property (3) implies that A must be symmetric, that is $A = A^T$. Property (1) implies that A must be positive definite, that is for all $\mathbf{x} \neq 0$,

$$\mathbf{x}^T A\mathbf{x} > 0.$$

An equivalent definition of symmetric positive definite is that for some non-singular matrix B, A can be written

$$A = B^T B.$$

Thus

$$(\mathbf{x},\mathbf{y}) = \mathbf{x}^T A\mathbf{y} = \mathbf{x}^T B^T B\mathbf{y} = (B\mathbf{x})^T(B\mathbf{y}).$$

Example 1 *Let*

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

Then

$$\mathbf{x}^T\mathbf{y} = 1 \cdot 1 + 1 \cdot (-2) = -1.$$

whereas

$$\mathbf{x}^T\mathbf{x} = 2, \quad \mathbf{y}^T\mathbf{y} = 5.$$

Another possible inner product is

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \mathbf{y}.$$

Note that

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = B^T B$$

where

$$B = \begin{pmatrix} \sqrt{2} & 0 \\ -\sqrt{0.5} & \sqrt{1.5} \end{pmatrix}.$$

With $\mathbf{x}$ and $\mathbf{y}$ in the above example,

$$(\mathbf{x}, \mathbf{y}) = -1, \quad (\mathbf{x}, \mathbf{x}) = 2, \quad (\mathbf{y}, \mathbf{y}) = 14.$$

The *Cauchy-Schwarz* inequality given below is quite useful for all inner products.

Lemma 1 (Cauchy-Schwarz inequality) *Let* $(\cdot, \cdot)$ *be an inner product in* R^n. *Then for all* $\mathbf{x}, \mathbf{y} \in \mathrm{R}^n$ *we have*

$$|(\mathbf{x}, \mathbf{y})| \leq (\mathbf{x}, \mathbf{x})^{1/2} (\mathbf{y}, \mathbf{y})^{1/2}. \qquad (3)$$

Moreover, equality in (3) *holds if and only if* $\mathbf{x} = \alpha\mathbf{y}$ *for some* $\alpha \in \mathrm{R}$.

This inequality leads to the following definition of the angle between two vectors relative to an inner product.

Definition 2 *The angle* θ *between the two nonzero vectors* $\mathbf{x}, \mathbf{y} \in \mathrm{R}^n$ *with respect to an inner product* $(\cdot, \cdot)$ *is given by*

$$\cos\,\theta = \frac{(\mathbf{x}, \mathbf{y})}{(\mathbf{x}, \mathbf{x})^{1/2} (\mathbf{y}, \mathbf{y})^{1/2}}. \qquad (4)$$

Example 2 *Using the Euclidean inner product, for the vectors*

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

we have

$$\cos\,\theta = \frac{\mathbf{x}^T\mathbf{y}}{(\mathbf{x}^T\mathbf{x})^{1/2}\,(\mathbf{y}^T\mathbf{y})^{1/2}} = -1/(\sqrt{2}\sqrt{5}) \approx -0.3162.$$

Thus

$$\theta \approx 1.8925^R \approx 108.4349°.$$

This is the angle you get if you draw these two vectors on a piece of paper.

If we use the other inner product,

$$\cos\,\theta = \frac{\mathbf{x}^T A\mathbf{y}}{(\mathbf{x}^T A\mathbf{x})^{1/2}\,(\mathbf{y}^T A\mathbf{y})^{1/2}} = -1/(\sqrt{2}\sqrt{14}) \approx 0.1890.$$

In that case,

$$\theta \approx 1.7609^R \approx 100.8934°.$$

The next important definition is that of a vector norm.

The following are important examples of norms.

$$\|\mathbf{x}\|_2 = \left(\mathbf{x}^T\mathbf{x}\right)^{1/2} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}. \tag{5}$$

This is called the Euclidean norm or the two-norm.

The one-norm is is given by

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|. \tag{6}$$

Its nickname is the Manhattan norm.

The ∞-norm given by

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|. \tag{7}$$

It is also called the maximum norm.

Example 3 *Take* $\mathbf{x} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$. *Then*

$$\|\mathbf{x}\|_2 = 5, \quad \|\mathbf{x}\|_1 = 7, \quad \|\mathbf{x}\|_\infty = 4.$$

A norm is just any function that satisfies the following properties.

Definition 3 *A norm in* R^n *is a function* $\|\cdot\|$ *mapping* R^n *into* R *satisfying the following three axioms*

1. $\|\mathbf{x}\| \geq 0$;
   $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = 0, \quad \mathbf{x} \in \mathrm{R}^n$.
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|, \quad \mathbf{x} \in \mathrm{R}^n, \alpha \in \mathrm{R}$.
3. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \quad \mathbf{x}, \mathbf{y} \in \mathrm{R}^n$.

There are two important classes of norms.

Inner Product Norms

$$\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})^{1/2}$$

is a norm for any inner product. The two-norm is the inner product norm for the Euclidean inner product. Neither the one-norm nor the ∞-norms is an inner product norm. The general form is

$$\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^T A \mathbf{x}}$$

where A is some positive definite matrix. Since $A = B^T B$ for some nonsingular B, this is equivalent to

$$\begin{aligned} \|\mathbf{x}\|_A &= \sqrt{\mathbf{x}^T A \mathbf{x}} = \sqrt{\mathbf{x}^T B^T B \mathbf{x}} \\ &= \|B\mathbf{x}\|_2 \end{aligned}$$

The p-norms

These are also called Hölder norms.

They are given by

$$\|\mathbf{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}, \quad p \geq 1. \tag{8}$$

Here $p = 2$ is the two-norm and $p = 1$ is the one-norm. The ∞-norm is so named because

$$\|\mathbf{x}\|_\infty = \lim_{p \to \infty} \|\mathbf{x}\|_p.$$