First Half of the Linear Algebra Course (**wi1 277LR**)

Row Reduction and Echelon Forms

A system of linear equations can be represented by an *augmented* matrix $[A|\mathbf{b}]$ or by a *matrix equation* $A\mathbf{x} = \mathbf{b}$ or by a *vector equation*: $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ with $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$.

The augmented matrix can be changed into an equivalent *echelon form*: in this form you can easily decide if the system is consistent or not, and if a general solution has *free variables* (look at the pivots).

To solve the system: change the augmented matrix into its (unique) *reduced echelon form*.

If $\mathbf{v}_h$ is the general solution of the *homogeneous* system $A\mathbf{x} = \mathbf{0}$, and $\mathbf{v}_p$ is one (particular) solution of $A\mathbf{x} = \mathbf{b}$, then $\mathbf{v}_p + \mathbf{v}_h$ is the *general solution* of $A\mathbf{x} = \mathbf{b}$.

Linear Transformation

T **is a linear transformation** if $T(c\,\mathbf{u} + d\,\mathbf{v}) = c\,T(\mathbf{u}) + d\,T(\mathbf{v})$ for all $\mathbf{u}$ and $\mathbf{v}$ and all scalars c and d.

Example The matrix transformation $T : \mathbf{x} \mapsto A\mathbf{x}$.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $A = [T(\mathbf{e}_1)\, T(\mathbf{e}_2) \cdots T(\mathbf{e}_n)]$ is the standard matrix for T, that is $T(\mathbf{x}) = A\mathbf{x}$.

A mapping (general function, transformation) $T : V \to W$ is **one-to-one** if each b in W is the image of at most one x in V. The mapping T is **onto** W if each b in W is the image of at least one x in V.

Span

$\mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is the collection of all the linear combinations of the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ [= the set of all vectors that can be written in the form $c_1\,\mathbf{v}_1 + \cdots + c_p\,\mathbf{v}_p$].

Subspace

H **is a** ***subspace*** **of** $\mathbb{R}^n$ if H is a set of $\mathbb{R}^n$ with the properties:
1) $\mathbf{0}$ is in H and **2)** for each $\mathbf{u}$ and $\mathbf{v}$ in H and for each scalar c and d, the linear combination $c\mathbf{u} + d\mathbf{v}$ is also in H.

Examples $\mathrm{Span}(\mathbf{v}_1, \cdots \mathbf{v}_p)$, $\mathrm{Col}(A)$, $\mathrm{Nul}(A)$.

Given: $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$ is a $m \times n$-matrix. Then $\mathrm{Col}A \equiv \mathrm{Span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$ is a subspace of $\mathbb{R}^m$ and $\mathrm{Nul}(A)$ ($\equiv$ the set of all solutions of $A\mathbf{x} = \mathbf{0}$) is a subspace of $\mathbb{R}^n$.

$$
\begin{aligned}
\mathrm{Col}A = \mathbb{R}^m \quad &\Leftrightarrow \quad A\mathbf{x} = \mathbf{b} \text{ is consistent for all } \mathbf{b} \\
&\Leftrightarrow \quad \text{all rows of (the echelon form of) } A \text{ contain a pivot} \\
&\Leftrightarrow \quad \text{the matrix transformation } T : \mathbf{x} \mapsto A\mathbf{x} \text{ is onto } \mathbb{R}^m \\
\mathrm{Nul}A = \{\mathbf{0}\} \quad &\Leftrightarrow \quad A\mathbf{x} = \mathbf{0} \text{ has only the trivial solution} \\
&\Leftrightarrow \quad \text{all columns of (the echelon form of) } A \text{ contain a pivot} \\
&\Leftrightarrow \quad \text{the matrix transformation } T : \mathbf{x} \mapsto A\mathbf{x} \text{ is one-to-one}
\end{aligned}
$$

Linear Independent

The set $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linearly independent if the vector equation $x_1\mathbf{v}_1 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$ has *only* the trivial solution (that is: $x_1 = x_2 = \cdots = x_p = 0$).

A set $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is a basis for H if
1) $H = \mathsf{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_p)$ and **2)** $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linear independent.

If $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is a basis for subspace H then $\mathsf{dim}\, H = p$.

- The set of vectors $\mathcal{B}$ is linear independent if and only if no vector of $\mathcal{B}$ is a linear combination of the other vectors of $\mathcal{B}$
- If $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linearly independent, then it spans a p-dimensional subspace.
- The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ of vectors of $\mathbb{R}^n$ is linearly independent if $A = [\mathbf{v}_1 \cdots \mathbf{v}_p]$ has in each column a pivot.
- Suppose there is some linear dependence relation between $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$ is row equivalent to $B = [\mathbf{b}_1 \cdots \mathbf{b}_n]$. Then the same linear dependence relation exists between $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

Rank

The **rank** of matrix A is equal to $\mathsf{dimCol}A$.
The pivot columns of A form a basis of $\mathsf{Col}A$.
A basis of $\mathsf{Nul}A$ has as many vectors as the solution of $A\mathbf{x} = \mathbf{0}$ has free variables.
Therefore: $\mathsf{dimCol}A + \mathsf{dimNul}A =$ number of columns of A

Coordinate vector

If $\mathcal{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_p\}$ is a basis for H and $\mathbf{x}$ is in H, then *the coordinate vector of* $\mathbf{x}$ *relative to* the basis $\mathcal{B}$ is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \quad \text{with } \mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p$$

Matrix Multiplication

If $B = [\mathbf{b}_1 \cdots \mathbf{b}_p]$ then $A\,B = [A\mathbf{b}_1\; A\mathbf{b}_2 \cdots A\mathbf{b}_p]$.

If A is a square $n \times n$-matrix with $AB = BA = I_n$ for some matrix B then A is invertible (A is also called *non singular*) and $B \equiv A^{-1}$ is the inverse of A.

$(A\,B)^T = B^T A^T$. $(A\,B)^{-1} = B^{-1}A^{-1}$ if A^{-1} and B^{-1} exist.

If A^{-1} then $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. A^{-1} exists only if $\mathsf{det}A = ad - bc \neq 0$. $A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Inner product in $\mathbb{R}^n$

$\mathbf{u} \bullet \mathbf{v} \equiv u_1v_1 + \cdots + u_nv_n = [\mathbf{u}]^T[\mathbf{v}]$. Some properties:
$\mathbf{u} \bullet \mathbf{v} = \mathbf{v} \bullet \mathbf{u}$, $(a\,\mathbf{u} + b\,\mathbf{v}) \bullet \mathbf{w} = a\,\mathbf{u} \bullet \mathbf{w} + b\,\mathbf{v} \bullet \mathbf{w}$, $\mathbf{u} \bullet \mathbf{u} = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$.

Norm (or length): $\|\mathbf{u}\| \equiv \sqrt{u_1^2 + \cdots + u_n^2} = \sqrt{\mathbf{u} \bullet \mathbf{u}}$.
Distance: $\mathsf{dist}(\mathbf{u}, \mathbf{v}) \equiv \|\mathbf{u} - \mathbf{v}\|$.

Orthogonality: $\mathbf{u} \perp \mathbf{v}$ if $\mathbf{u} \bullet \mathbf{v} = 0$.

$\mathbf{u} \perp \mathbf{v} \;\Leftrightarrow\; \|\mathbf{u}-\mathbf{v}\| = \|\mathbf{u}+\mathbf{v}\|$.
$\mathbf{u} \perp \mathbf{v} \;\Leftrightarrow\; \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ (Pythagorean Theorem for $\mathbb{R}^n$).

$\mathbf{u} \perp W$ if $\mathbf{u} \perp \mathbf{w}$ for all $\mathbf{w}$ from W.
$W^\perp$= orthogonal complement of $W \equiv$ the set of all $\mathbf{u}$ with $\mathbf{u} \perp W$.

Orthogonal sets

$\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is a orthogonal basis for W: a basis for W with $\mathbf{u}_i \perp \mathbf{u}_j$ for all $i \neq j$.

$\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is an orthonormal basis for W: orthogonal basis with $\|\mathbf{u}_i\| = 1$ for all i.

The $m \times n$-matrix U has orthonormal columns $\Leftrightarrow U^T U = I_n$.

The *square* matrix A is an orthogonal matrix if $A^{-1} = A^T$. An orthogonal matrix has orthonormal columns.

Special case: $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is a orthogonal basis for W and $\mathbf{y}$ in W. Then

$$\mathbf{y} = c_1\,\mathbf{u}_1 + \cdots + c_p\,\mathbf{u}_p \quad \text{with} \quad c_j = \frac{\mathbf{y} \bullet \mathbf{u}_j}{\mathbf{u}_j \bullet \mathbf{u}_j}$$

Orthogonal projection

Orthogonal projection of vector $\mathbf{y}$ on $L = \mathsf{Span}\{\mathbf{u}\}$: $\quad \hat{\mathbf{y}} \equiv \mathsf{proj}_L \mathbf{y} = \left(\frac{\mathbf{y}\bullet\mathbf{u}}{\mathbf{u}\bullet\mathbf{u}}\right)\mathbf{u}$.
$\hat{\mathbf{y}}$ is the vector with the property: $\hat{\mathbf{y}}$ in L and $\mathbf{z} = (\mathbf{y} - \hat{\mathbf{y}}) \perp L$.

Orthogonal projection of vector $\mathbf{y}$ on subspace W is the unique vector $\hat{\mathbf{y}} = \mathsf{proj}_W \mathbf{y}$ defined by the decomposition $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ with $\hat{\mathbf{y}}$ in W and $\mathbf{z} = (\mathbf{y} - \hat{\mathbf{y}}) \perp W$.

Remarks:

- If $\mathbf{y}$ in W then $\hat{\mathbf{y}} = \mathbf{y}$
- $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$ (best approximation in W).

Special case: $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is an orthogonal basis of W. Then

$$\hat{\mathbf{y}} = \mathsf{proj}_W \mathbf{y} = c_1\,\mathbf{u}_1 + \cdots + c_p\,\mathbf{u}_p \quad \text{with} \quad c_j = \frac{\mathbf{y} \bullet \mathbf{u}_j}{\mathbf{u}_j \bullet \mathbf{u}_j}$$

Gram-Schmidt process

The Gram-Schmidt process is an algorithm: starting from a basis $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ of subspace W it constructs an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ for W.

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1\,, & W_1 &= \mathsf{Span}\{\mathbf{x}_1\} = \mathsf{Span}\{\mathbf{v}_1\} \\
\mathbf{v}_2 = \mathbf{x}_2 - \mathsf{proj}_{W_1}\mathbf{x}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{x}_2\bullet\mathbf{v}_1}{\mathbf{v}_1\bullet\mathbf{v}_1}\right)\mathbf{v}_1\,, & W_2 &= \mathsf{Span}\{\mathbf{x}_1, \mathbf{x}_2\} = \mathsf{Span}\{\mathbf{v}_1, \mathbf{v}_2\} \\
\mathbf{v}_3 = \mathbf{x}_3 - \mathsf{proj}_{W_2}\mathbf{x}_3 &= \mathbf{x}_3 - \left[\left(\frac{\mathbf{x}_3\bullet\mathbf{v}_1}{\mathbf{v}_1\bullet\mathbf{v}_1}\right)\mathbf{v}_1 + \left(\frac{\mathbf{x}_3\bullet\mathbf{v}_2}{\mathbf{v}_2\bullet\mathbf{v}_2}\right)\mathbf{v}_2\right]\,, & W_3 &= \mathsf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = \mathsf{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \\
&\vdots & &\vdots
\end{aligned}$$

Least-squares problems

Given the matrix equation $A\mathbf{x} = \mathbf{b}$ (may be inconsistent).

Replace this equation by $A\mathbf{x} = \hat{\mathbf{b}} = \text{proj}_W\mathbf{b}$ with $W = \text{Col}A$. The solution $\hat{\mathbf{x}}$ of the later equation is called the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

Remarks

- $\|\mathbf{b} - \hat{\mathbf{b}}\| = \|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{u}\|$ for all $\mathbf{u}$. A least-square solution $\hat{\mathbf{x}}$ is such that $A\hat{\mathbf{x}}$ is the closest you can get to $\mathbf{b}$ (best approximation).
- You can determine the orthogonal projection of $\mathbf{b}$ on $W = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ by first calculating a least square solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (with $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$) and then calculating $A\hat{\mathbf{x}}$.
- If $A\mathbf{x} = \mathbf{b}$ is consistent then $\hat{\mathbf{b}} = \mathbf{b}$ and a least-squares solution is a solution of $A\mathbf{x} = \mathbf{b}$.
- The least-squares error $= \|\mathbf{b} - \hat{\mathbf{b}}\|$.
- The least-squares solution may not be unique. It is only unique if $A\mathbf{x} = \hat{\mathbf{b}}$ has a unique solution, that is: all the columns of A have a pivot (so the columns of A are linear independent; also: $\text{Nul}(A) = \{\mathbf{0}\}$).

Theorem The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is equal to the general solution of the *normal equations* $A^T A\mathbf{x} = A^T\mathbf{b}$.

Application

At time $t = 0$ a certain mixture of radioactive substances contains M_A grams of substance A and M_B grams of substance B. A model for the total amount y of the mixture present at time t is $y = M_A\, e^{-0.02t} + M_B\, e^{-0.07t}$.

The following (t, y)-data are available:

$$(10, 21.34), \quad (11, 20.68), \quad (14, 18.87), \quad (15, 18.30)$$

Determine the equation $y = M_A\, e^{-0.02t} + M_B\, e^{-0.07t}$ which is the best fit to these data according to the least-squares method.

Solution By inserting the data in the (model) equation you get a system of equations, linear in the unknown parameters M_A and M_B. For example:

$$\begin{aligned} (10, 21.34) \quad &\Rightarrow \quad M_A\, e^{-0.20} + M_B\, e^{-0.70} = 21.34 \\ (11, 20.68) \quad &\Rightarrow \quad M_A\, e^{-0.22} + M_B\, e^{-0.77} = 20.68 \end{aligned}$$

Etc. The parameter vector $\mathbf{x} = \begin{bmatrix} M_A \\ M_B \end{bmatrix}$ has then to satisfy the matrix equation

$$\begin{bmatrix} e^{-0.20} & e^{-0.7} \\ e^{-0.22} & e^{-0.77} \\ e^{-0.28} & e^{-0.98} \\ e^{-0.30} & e^{-1.05} \end{bmatrix} \begin{bmatrix} M_A \\ M_B \end{bmatrix} = \overbrace{\begin{bmatrix} 0.8187 & 0.4966 \\ 0.8025 & 0.4630 \\ 0.7558 & 0.3753 \\ 0.7408 & 0.3499 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} M_A \\ M_B \end{bmatrix}}^{\mathbf{x}} = \overbrace{\begin{bmatrix} 21.34 \\ 20.68 \\ 18.87 \\ 18.30 \end{bmatrix}}^{\mathbf{b}}$$

The least-squares solution of the (inconsistent) equation $A\mathbf{x} = \mathbf{b}$ can be computed by solving the normal equation $A^T A\mathbf{x} = A^T \mathbf{b}$, or by computing $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.
Solution: $M_A = 19.9411$, $M_B = 10.0996$.
Therefore the equation $y = 19.94\, e^{-0.02t} + 10.10\, e^{-0.07t}$ is the best fit to these data according to the least-squares method.