MATHEMATICS PRIMER FOR NERC DATA ASSIMILATION & VISUALIZATION SCHOOL

Univ. of Reading, September 2014

---

1. Vectors and matrices

1.1. Vector representation of information (n elements).

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix}, \quad v_i = (\mathbf{v})_i.$$

1.2. Matrix ($m \times n$ elements).

$$\mathbf{N} = \begin{pmatrix} N_{11} & \cdots & N_{1j} & \cdots & N_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ N_{i1} & \cdots & N_{ij} & \cdots & N_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N_{m1} & \cdots & N_{mj} & \cdots & N_{mn} \end{pmatrix}, \quad N_{ij} = (\mathbf{N})_{ij}.$$

1.3. Matrix ($m \times n$) acting on a vector.

$$\mathbf{v}^{\mathrm{b}} = \mathbf{N}\mathbf{v}^{\mathrm{a}}, \quad v_i^{\mathrm{b}} = \sum_{j=1}^{n} N_{ij} v_j^{\mathrm{a}}, \quad 1 \leq i \leq m \text{ as below:}$$

$$\begin{pmatrix} v_1^{\mathrm{b}} \\ \vdots \\ v_i^{\mathrm{b}} \\ \vdots \\ v_m^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} N_{11} & \cdots & N_{1j} & \cdots & N_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ N_{i1} & \cdots & N_{ij} & \cdots & N_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N_{m1} & \cdots & N_{mj} & \cdots & N_{mn} \end{pmatrix} \begin{pmatrix} v_1^{\mathrm{a}} \\ \vdots \\ v_j^{\mathrm{a}} \\ \vdots \\ v_n^{\mathrm{a}} \end{pmatrix} = \begin{pmatrix} N_{11}v_1^{\mathrm{a}} + \ldots + N_{1j}v_j^{\mathrm{a}} + \ldots + N_{1n}v_n^{\mathrm{a}} \\ \vdots \\ N_{i1}v_1^{\mathrm{a}} + \ldots + N_{ij}v_j^{\mathrm{a}} + \ldots + N_{in}v_n^{\mathrm{a}} \\ \vdots \\ N_{m1}v_1^{\mathrm{a}} + \ldots + N_{mj}v_j^{\mathrm{a}} + \ldots + N_{mn}v_n^{\mathrm{a}} \end{pmatrix}.$$

1.4. Identity/unit matrix ($p \times p$ elements).

$$\mathbf{I}_p = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \quad (\mathbf{I}_p)_{ij} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

1.5. Matrix addition ($m \times n$ elements).

$$\mathbf{N} = \mathbf{N}^{\mathrm{a}} + \mathbf{N}^{\mathrm{b}}, \quad N_{ij} = N^{\mathrm{a}}_{ij} + N^{\mathrm{b}}_{ij} \text{ as below:}$$

$$\begin{pmatrix} N^{\mathrm{a}}_{11} & \cdots & N^{\mathrm{a}}_{1n} \\ \vdots & \vdots & \vdots \\ N^{\mathrm{a}}_{m1} & \cdots & N^{\mathrm{a}}_{mn} \end{pmatrix} + \begin{pmatrix} N^{\mathrm{b}}_{11} & \cdots & N^{\mathrm{b}}_{1n} \\ \vdots & \ddots & \vdots \\ N^{\mathrm{b}}_{m1} & \cdots & N^{\mathrm{b}}_{mn} \end{pmatrix} = \begin{pmatrix} N^{\mathrm{a}}_{11} + N^{\mathrm{b}}_{11} & \cdots & N^{\mathrm{a}}_{1n} + N^{\mathrm{b}}_{1n} \\ \vdots & \ddots & \vdots \\ N^{\mathrm{a}}_{m1} + N^{\mathrm{b}}_{m1} & \cdots & N^{\mathrm{a}}_{mn} + N^{\mathrm{b}}_{mn} \end{pmatrix}.$$

1.6. Matrix multiplication ($m \times n$ elements - an $m \times p$ matrix multiplied by an $p \times n$ matrix).

$$\mathbf{N} = \mathbf{N}^{\mathrm{a}}\mathbf{N}^{\mathrm{b}}, \quad N_{ij} = \sum_{k=1}^{p} N^{\mathrm{a}}_{ik} N^{\mathrm{b}}_{kj} \text{ as below:}$$

$$\begin{pmatrix} N^{\mathrm{a}}_{11} & \cdots & N^{\mathrm{a}}_{1p} \\ \vdots & \ddots & \vdots \\ N^{\mathrm{a}}_{m1} & \cdots & N^{\mathrm{a}}_{mp} \end{pmatrix} \begin{pmatrix} N^{\mathrm{b}}_{11} & \cdots & N^{\mathrm{b}}_{1n} \\ \vdots & \ddots & \vdots \\ N^{\mathrm{b}}_{p1} & \cdots & N^{\mathrm{b}}_{pn} \end{pmatrix} = \begin{pmatrix} N^{\mathrm{a}}_{11}N^{\mathrm{b}}_{11} + \ldots + N^{\mathrm{a}}_{1p}N^{\mathrm{b}}_{p1} & \cdots & N^{\mathrm{a}}_{11}N^{\mathrm{b}}_{1n} + \ldots + N^{\mathrm{a}}_{1p}N^{\mathrm{b}}_{pn} \\ \vdots & \ddots & \vdots \\ N^{\mathrm{a}}_{m1}N^{\mathrm{b}}_{11} + \ldots + N^{\mathrm{a}}_{mp}N^{\mathrm{b}}_{p1} & \cdots & N^{\mathrm{a}}_{m1}N^{\mathrm{b}}_{1n} + \ldots + N^{\mathrm{a}}_{mp}N^{\mathrm{b}}_{pn} \end{pmatrix}.$$

In general, matrices are non-commutative $\mathbf{N}^{\mathrm{a}}\mathbf{N}^{\mathrm{b}} \neq \mathbf{N}^{\mathrm{b}}\mathbf{N}^{\mathrm{a}}$.

1.7. Matrix transpose ($\mathbf{N}^{\mathrm{a}}$ has $m \times n$ elements, $\mathbf{N}^{\mathrm{b}}$ has $n \times m$ elements).

$$\text{If } \mathbf{N}^{\mathrm{b}} = \mathbf{N}^{\mathrm{aT}}, \quad N^{\mathrm{b}}_{ij} = N^{\mathrm{a}}_{ji}, \text{ e.g.:}$$

$$\mathbf{N}^{\mathrm{a}} = \begin{pmatrix} N^{\mathrm{a}}_{11} & N^{\mathrm{a}}_{12} & N^{\mathrm{a}}_{13} \\ N^{\mathrm{a}}_{21} & N^{\mathrm{a}}_{22} & N^{\mathrm{a}}_{23} \end{pmatrix}, \quad \mathbf{N}^{\mathrm{b}} = \begin{pmatrix} N^{\mathrm{a}}_{11} & N^{\mathrm{a}}_{21} \\ N^{\mathrm{a}}_{12} & N^{\mathrm{a}}_{22} \\ N^{\mathrm{a}}_{13} & N^{\mathrm{a}}_{23} \end{pmatrix}.$$

If $\mathbf{N}^{\mathrm{a}} = \mathbf{N}^{\mathrm{aT}}$ then matrix $\mathbf{N}^{\mathrm{a}}$ is symmetric (only square matrices ($m = n$) can be symmetric).

1.8. Transpose of a product of matrices.

$$(\mathbf{N}^{\mathrm{a}}\mathbf{N}^{\mathrm{b}})^{\mathrm{T}} = \mathbf{N}^{\mathrm{bT}}\mathbf{N}^{\mathrm{aT}}.$$

1.9. Matrix inversion ($n \times n$ elements). Let $\mathbf{N}$ be a square ($n \times n$) non-singular matrix.

$$\text{If } \mathbf{v}^{\mathrm{b}} = \mathbf{N}\mathbf{v}^{\mathrm{a}}, \text{ then } \mathbf{v}^{\mathrm{a}} = \mathbf{N}^{-1}\mathbf{v}^{\mathrm{b}}.$$

$$\text{In general } (\mathbf{N}^{-1})_{ij} \neq (\mathbf{N})^{-1}_{ij}.$$

For $n = 2$, $\quad \mathbf{N} = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix}, \quad \mathbf{N} = \dfrac{1}{\det(\mathbf{N})} \begin{pmatrix} N_{22} & -N_{12} \\ -N_{21} & N_{11} \end{pmatrix}, \quad \det(\mathbf{N}) = N_{11}N_{22} - N_{12}N_{21}.$

If $\mathbf{N}$ is singular then it has a zero determinant and the inverse cannot be found in general.

1.10. **Diagonal matrix ($n \times n$ elements).** A matrix is diagonal if $N_{ij} = 0$ if $i \neq j$. $\mathbf{N}$ may be written:

$$\mathbf{N} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

The inverse of a square diagonal matrix is $(\mathbf{N}^{-1})_{ii} = (\mathbf{N})_{ii}^{-1}$, $(\mathbf{N}^{-1})_{ij} = 0$ for $i \neq j$:

$$\begin{pmatrix} N_{11} & 0 & \cdots \\ 0 & N_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}^{-1} = \begin{pmatrix} 1/N_{11} & 0 & \cdots \\ 0 & 1/N_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

1.11. **Euclidean vector inner product (scalar product/dot product) (n elements).** Two different vectors:

$$a = \mathbf{v}^{\mathrm{a}} \cdot \mathbf{v}^{\mathrm{b}} = \mathbf{v}^{\mathrm{aT}} \mathbf{v}^{\mathrm{b}} = \left\langle \mathbf{v}^{\mathrm{a}}, \mathbf{v}^{\mathrm{b}} \right\rangle = \sum_{i=1}^{n} v_i^{\mathrm{a}} v_i^{\mathrm{b}} \text{ as below:}$$

$$\begin{pmatrix} v_1^{\mathrm{a}} \\ \vdots \\ v_i^{\mathrm{a}} \\ \vdots \\ v_n^{\mathrm{a}} \end{pmatrix} \cdot \begin{pmatrix} v_1^{\mathrm{b}} \\ \vdots \\ v_i^{\mathrm{b}} \\ \vdots \\ v_n^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} v_1^{\mathrm{a}} & \cdots & v_i^{\mathrm{a}} & \cdots & v_n^{\mathrm{a}} \end{pmatrix} \begin{pmatrix} v_1^{\mathrm{b}} \\ \vdots \\ v_i^{\mathrm{b}} \\ \vdots \\ v_n^{\mathrm{b}} \end{pmatrix} = v_1^{\mathrm{a}} v_1^{\mathrm{b}} + \cdots + v_i^{\mathrm{a}} v_i^{\mathrm{b}} + \cdots + v_n^{\mathrm{a}} v_n^{\mathrm{b}}.$$

The same vector:

$$b = \mathbf{v} \cdot \mathbf{v} = \mathbf{v}^{\mathrm{T}} \mathbf{v} = \langle \mathbf{v}, \mathbf{v} \rangle = \sum_{i=1}^{n} v_i^2 = \|\mathbf{v}\|^2 \text{ as below:}$$

$$\begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_i & \cdots & v_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} = v_1 v_1 + \cdots + v_i v_i + \cdots + v_n v_n.$$

1.12. **Non-Euclidean vector inner product.** Two different vectors (m elements and n elements):

$$a = \mathbf{v}^{\mathrm{a}} \cdot (\mathbf{C} \mathbf{v}^{\mathrm{b}}) = \mathbf{v}^{\mathrm{aT}} \mathbf{C} \mathbf{v}^{\mathrm{b}} = \left\langle \mathbf{v}^{\mathrm{a}}, \mathbf{v}^{\mathrm{b}} \right\rangle_{\mathbf{C}} = \sum_{i=1}^{m} v_i^{\mathrm{a}} \sum_{j=1}^{n} C_{ij} v_j^{\mathrm{b}} \text{ as below:}$$

$$\begin{pmatrix} v_1^{\mathrm{a}} & \cdots & v_i^{\mathrm{a}} & \cdots & v_m^{\mathrm{a}} \end{pmatrix} \begin{pmatrix} C_{11} & \cdots & C_{1j} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ C_{i1} & \cdots & C_{ij} & \cdots & C_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mj} & \cdots & C_{mn} \end{pmatrix} \begin{pmatrix} v_1^{\mathrm{b}} \\ \vdots \\ v_j^{\mathrm{b}} \\ \vdots \\ v_n^{\mathrm{b}} \end{pmatrix}.$$

The same vector (n elements):

$$b = \mathbf{v} \cdot (\mathbf{C}\mathbf{v}) = \mathbf{v}^{\mathrm{T}}\mathbf{C}\mathbf{v} = \langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{C}} = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} C_{ij} v_j = \|\mathbf{v}\|_{\mathbf{C}}^2 \text{ as below:}$$

$$\begin{pmatrix} v_1 & \cdots & v_i & \cdots & v_n \end{pmatrix} \begin{pmatrix} C_{11} & \cdots & C_{1j} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ C_{i1} & \cdots & C_{ij} & \cdots & C_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nj} & \cdots & C_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix}.$$

1.13. **Vector outer product ($m \times n$ elements).**

$$\mathbf{N} = \mathbf{v}^{\mathrm{a}}\mathbf{v}^{\mathrm{bT}}, \quad N_{ij} = v_i^{\mathrm{a}} v_j^{\mathrm{b}} \text{ as below:}$$

$$\begin{pmatrix} v_1^{\mathrm{a}} \\ \vdots \\ v_i^{\mathrm{a}} \\ \vdots \\ v_m^{\mathrm{a}} \end{pmatrix} \begin{pmatrix} v_1^{\mathrm{b}} & \cdots & v_j^{\mathrm{b}} & \cdots & v_n^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} v_1^{\mathrm{a}}v_1^{\mathrm{b}} & \cdots & v_1^{\mathrm{a}}v_j^{\mathrm{b}} & \cdots & v_1^{\mathrm{a}}v_n^{\mathrm{b}} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ v_i^{\mathrm{a}}v_1^{\mathrm{b}} & \cdots & v_i^{\mathrm{a}}v_j^{\mathrm{b}} & \cdots & v_i^{\mathrm{a}}v_n^{\mathrm{b}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_m^{\mathrm{a}}v_1^{\mathrm{b}} & \cdots & v_m^{\mathrm{a}}v_j^{\mathrm{b}} & \cdots & v_m^{\mathrm{a}}v_n^{\mathrm{b}} \end{pmatrix}.$$

1.14. **Schur (or Hadamard) product.** For matrices ($m \times n$ elements):

$$\mathbf{N} = \mathbf{N}^{\mathrm{a}} \circ \mathbf{N}^{\mathrm{b}}, \quad N_{ij} = N_{ij}^{\mathrm{a}} N_{ij}^{\mathrm{b}} \text{ as below:}$$

$$\begin{pmatrix} N_{11}^{\mathrm{a}} & \cdots & N_{1n}^{\mathrm{a}} \\ \vdots & \ddots & \vdots \\ N_{m1}^{\mathrm{a}} & \cdots & N_{mn}^{\mathrm{a}} \end{pmatrix} \circ \begin{pmatrix} N_{11}^{\mathrm{b}} & \cdots & N_{1n}^{\mathrm{b}} \\ \vdots & \ddots & \vdots \\ N_{m1}^{\mathrm{b}} & \cdots & N_{mn}^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} N_{11}^{\mathrm{a}}N_{11}^{\mathrm{b}} & \cdots & N_{1n}^{\mathrm{a}}N_{1n}^{\mathrm{b}} \\ \vdots & \ddots & \vdots \\ N_{m1}^{\mathrm{a}}N_{m1}^{\mathrm{b}} & \cdots & N_{mn}^{\mathrm{a}}N_{mn}^{\mathrm{b}} \end{pmatrix}.$$

For vectors (n elements)

$$\mathbf{v} = \mathbf{v}^{\mathrm{a}} \circ \mathbf{v}^{\mathrm{b}}, \quad v_i = v_i^{\mathrm{a}} v_i^{\mathrm{b}} \text{ as below:}$$

$$\begin{pmatrix} v_1^{\mathrm{a}} \\ \vdots \\ v_n^{\mathrm{a}} \end{pmatrix} \circ \begin{pmatrix} v_1^{\mathrm{b}} \\ \vdots \\ v_n^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} v_1^{\mathrm{a}}v_1^{\mathrm{b}} \\ \vdots \\ v_n^{\mathrm{a}}v_n^{\mathrm{b}} \end{pmatrix}.$$

1.15. **The trace of a matrix ($n \times n$ elements).** The trace of a square matrix $\mathbf{N}$, $\mathrm{tr}(\mathbf{N})$, is the sum of the diagonal elements:

$$\mathrm{tr}(\mathbf{N}) = \sum_{i=1}^{n} N_{ii}.$$

1.16. **The Sherman-Morrison-Woodbury formula.**

$$\left(\mathbf{A} + \mathbf{C}\mathbf{D}^{\mathrm{T}}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\left(\mathbf{I} + \mathbf{D}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}^{-1}.$$

Replacing $\mathbf{C} \rightarrow \mathbf{C}\mathbf{B}$ and then setting $\mathbf{C} = \mathbf{D} = \mathbf{H}$ and $\mathbf{A} = \mathbf{R}$, the following useful formula results:

$$\left(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{B}\mathbf{H}^{\mathrm{T}} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right).$$

---

2. Functions

2.1. **Scalar valued function of a vector (n elements) and its derivative.**

$$f(\mathbf{v}), \quad f, \quad \nabla_{\mathbf{v}} f(\mathbf{v}) = \left(\frac{\partial f}{\partial \mathbf{v}}\right)^{\mathrm{T}} = \begin{pmatrix} \partial f/\partial v_1 \\ \partial f/\partial v_2 \\ \vdots \\ \partial f/\partial v_n \end{pmatrix}.$$

---

3. Mean, (co)variance, correlation and Gaussian statistics

3.1. **The variance, standard deviation and mean of a scalar.** Consider a population of N scalars, s^l, $1 \leq l \leq N$. The following are for the variance, $\mathrm{var}(s)$, standard deviation, σ_s, and mean, $\langle s \rangle$ (common notations are given)[1]:

$$\mathrm{var}(s) = \left\langle (s - \langle s \rangle)^2 \right\rangle = \overline{(s - \bar{s})^2} = \mathcal{E}\left((s - \mathcal{E}(s))^2\right) \approx \frac{1}{\tilde{N}} \sum_{l=1}^{N} (s^l - \langle s \rangle)^2,$$

$$\sigma_s = \sqrt{\mathrm{var}(s)}, \quad \langle s \rangle = \bar{s} = \mathcal{E}(s) \approx \frac{1}{N} \sum_{l=1}^{N} s^l.$$

3.2. **The covariance between two scalars.** Consider two populations, each of N scalars, s^l, t^l, $1 \leq l \leq N$. The following is for the covariance, $\mathrm{cov}(s, t)$ (common notations are given)[2]:

$$\mathrm{cov}(s, t) = \langle (s - \langle s \rangle)(t - \langle t \rangle) \rangle = \overline{(s - \bar{s})(t - \bar{t})} = \mathcal{E}\left((s - \mathcal{E}(s))(t - \mathcal{E}(t))\right) \approx \frac{1}{\tilde{N}} \sum_{l=1}^{N} (s^l - \langle s \rangle)(t^l - \langle t \rangle).$$

The covariance between two scalars can be negative, zero or positive.

[1] Sample variance and sample mean are those (approximate expressions) given in terms of a finite population (a sample). In the formula for the sample variance, $\tilde{N} = N$ if $\langle s \rangle$ is taken to be the *exact* mean, but $\tilde{N} = N - 1$ if $\langle s \rangle$ is taken to be the *sample* mean.

[2] Sample covariance and sample mean are those (approximate expressions) given in terms of a finite population (a sample). In the formula for the sample covariance, $\tilde{N} = N$ if $\langle s \rangle$ and $\langle t \rangle$ are taken to be the *exact* means, but $\tilde{N} = N - 1$ if $\langle s \rangle$ and $\langle t \rangle$ are taken to be the *sample* means.

3.3. The correlation between two scalars.

$$\mathrm{cor}(s,t)=\frac{\mathrm{cov}(s,t)}{\sigma_s\sigma_t},\quad -1\leq \mathrm{cor}(s,t)\leq 1,\quad \mathrm{cor}(s,s)=\mathrm{cor}(t,t)=1.$$

3.4. The covariance matrix between two vectors (one with m elements, another with n elements). Consider two populations, each of N vectors, $\mathbf{u}^l$, $\mathbf{v}^l$, $1\leq l\leq N$. The following is for the covariance matrix, $\mathrm{cov}(\mathbf{u},\mathbf{v})$, which uses the outer product (common notations are given):

$$\begin{aligned}
\mathrm{cov}(\mathbf{u},\mathbf{v}) &= \left\langle(\mathbf{u}-\langle\mathbf{u}\rangle)(\mathbf{v}-\langle\mathbf{v}\rangle)^{\mathrm{T}}\right\rangle=\overline{(\mathbf{u}-\bar{\mathbf{u}})(\mathbf{v}-\bar{\mathbf{v}})^{\mathrm{T}}}=\mathcal{E}\left((\mathbf{u}-\mathcal{E}(\mathbf{u}))(\mathbf{v}-\mathcal{E}(\mathbf{v}))^{\mathrm{T}}\right),\\
&\approx \frac{1}{\tilde{N}}\sum_{l=1}^{N}\left(\mathbf{u}^l-\langle\mathbf{u}\rangle\right)\left(\mathbf{v}^l-\langle\mathbf{v}\rangle\right)^{\mathrm{T}},\\
(\mathrm{cov}(\mathbf{u},\mathbf{v}))_{ij} &\approx \frac{1}{\tilde{N}}\sum_{l=1}^{N}\left(u_i^l-\langle u_i\rangle\right)\left(v_j^l-\langle v_j\rangle\right),
\end{aligned}$$

$$\mathrm{cov}(\mathbf{u},\mathbf{v})=\frac{1}{\tilde{N}}\sum_{l=1}^{N}\begin{pmatrix}
\left(u_1^l-\langle u_1\rangle\right)\left(v_1^l-\langle v_1\rangle\right) & \cdots & \left(u_1^l-\langle u_1\rangle\right)\left(v_j^l-\langle v_j\rangle\right) & \cdots & \left(u_1^l-\langle u_1\rangle\right)\left(v_n^l-\langle v_n\rangle\right)\\
\vdots & \ddots & \vdots & \vdots & \vdots\\
\left(u_i^l-\langle u_i\rangle\right)\left(v_1^l-\langle v_1\rangle\right) & \cdots & \left(u_i^l-\langle u_i\rangle\right)\left(v_j^l-\langle v_j\rangle\right) & \cdots & \left(u_i^l-\langle u_i\rangle\right)\left(v_n^l-\langle v_n\rangle\right)\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\left(u_m^l-\langle u_m\rangle\right)\left(v_1^l-\langle v_1\rangle\right) & \cdots & \left(u_m^l-\langle u_m\rangle\right)\left(v_j^l-\langle v_j\rangle\right) & \cdots & \left(u_m^l-\langle u_m\rangle\right)\left(v_n^l-\langle v_n\rangle\right)
\end{pmatrix}.$$

If $\mathbf{u}=\mathbf{v}$, then $\mathrm{cov}(\mathbf{v},\mathbf{v})$ is the auto-covariance matrix of $\mathbf{v}$. Diagonal elements are variances of each element of $\mathbf{v}$, i.e. $(\mathrm{cov}(\mathbf{v},\mathbf{v}))_{ii}=\mathrm{var}(v_i)$:

$$\mathrm{cov}(\mathbf{v},\mathbf{v})=\frac{1}{\tilde{N}}\sum_{l=1}^{N}\begin{pmatrix}
\left(v_1^l-\langle v_1\rangle\right)^2 & \cdots & \left(v_1^l-\langle v_1\rangle\right)\left(v_i^l-\langle v_i\rangle\right) & \cdots & \left(v_1^l-\langle v_1\rangle\right)\left(v_n^l-\langle v_n\rangle\right)\\
\vdots & \ddots & \vdots & \vdots & \vdots\\
\left(v_i^l-\langle v_i\rangle\right)\left(v_1^l-\langle v_1\rangle\right) & \cdots & \left(v_i^l-\langle v_i\rangle\right)^2 & \cdots & \left(v_i^l-\langle v_i\rangle\right)\left(v_n^l-\langle v_n\rangle\right)\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\left(v_n^l-\langle v_n\rangle\right)\left(v_1^l-\langle v_1\rangle\right) & \cdots & \left(v_n^l-\langle v_n\rangle\right)\left(v_i^l-\langle v_i\rangle\right) & \cdots & \left(v_n^l-\langle v_n\rangle\right)^2
\end{pmatrix}.$$

Auto-covariance matrices are symmetric.

3.5. The correlation matrix between two vectors.

$\mathrm{cor}(\mathbf{u},\mathbf{v})=\mathbf{\Sigma}_{\mathbf{u}}^{-1}\mathrm{cov}(\mathbf{u},\mathbf{v})\mathbf{\Sigma}_{\mathbf{v}}^{-1}$, $\quad\mathbf{\Sigma}_{\mathbf{u}}=\mathrm{diag}(\sigma_{u_1},\sigma_{u_2},\cdots\sigma_{u_m})$, $\quad\mathbf{\Sigma}_{\mathbf{v}}=\mathrm{diag}(\sigma_{v_1},\sigma_{v_2},\cdots\sigma_{v_n})$ as below:

$$\mathrm{cor}(\mathbf{u},\mathbf{v}) = \begin{pmatrix}
\sigma_{u_1}^{-1} & 0 & \cdots & 0\\
0 & \sigma_{u_2}^{-1} & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots\\
0 & 0 & \cdots & \sigma_{u_m}^{-1}
\end{pmatrix}\begin{pmatrix}
(\mathrm{cov}(\mathbf{u},\mathbf{v}))_{11} & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{12} & \cdots & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{1n}\\
(\mathrm{cov}(\mathbf{u},\mathbf{v}))_{21} & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{22} & \cdots & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{2n}\\
\vdots & \vdots & \ddots & \vdots\\
(\mathrm{cov}(\mathbf{u},\mathbf{v}))_{m1} & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{m2} & \cdots & (\mathrm{cov}(\mathbf{u},\mathbf{v}))_{mn}
\end{pmatrix}\times$$

$$\begin{pmatrix}
\sigma_{v_1}^{-1} & 0 & \cdots & 0\\
0 & \sigma_{v_2}^{-1} & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots\\
0 & 0 & \cdots & \sigma_{v_n}^{-1}
\end{pmatrix},$$

$$\text{i.e.} \quad (\mathrm{cor}(\mathbf{u},\mathbf{v}))_{ij} = \frac{(\mathrm{cov}(\mathbf{u},\mathbf{v}))_{ij}}{\sigma_{u_i}\sigma_{v_j}}.$$

If $\mathbf{u} = \mathbf{v}$ then $\mathrm{cor}(\mathbf{v},\mathbf{v})$ is the auto-correlation matrix of $\mathbf{v}$:

$$\mathrm{cor}(\mathbf{v},\mathbf{v}) = \mathbf{\Sigma}_{\mathbf{v}}^{-1}\mathrm{cov}(\mathbf{v},\mathbf{v})\mathbf{\Sigma}_{\mathbf{v}}^{-1} \text{ as below:}$$

$$\begin{aligned}
\mathrm{cor}(\mathbf{v},\mathbf{v}) &= \begin{pmatrix} \sigma_{v_1}^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_{v_2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{v_n}^{-1} \end{pmatrix} \begin{pmatrix} (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{11} & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{12} & \cdots & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{1n} \\ (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{21} & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{22} & \cdots & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{n1} & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{n2} & \cdots & (\mathrm{cov}(\mathbf{v},\mathbf{v}))_{nn} \end{pmatrix} \times, \\
&\qquad \begin{pmatrix} \sigma_{v_1}^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_{v_2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{v_n}^{-1} \end{pmatrix}, \\
&= \begin{pmatrix} 1 & (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{12} & \cdots & (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{1n} \\ (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{21} & 1 & \cdots & (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{n1} & (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{n2} & \cdots & 1 \end{pmatrix},
\end{aligned}$$

$$\text{i.e.} \quad (\mathrm{cor}(\mathbf{v},\mathbf{v}))_{ij} = \frac{(\mathrm{cov}(\mathbf{v},\mathbf{v}))_{ij}}{\sigma_{v_i}\sigma_{v_j}}.$$

Auto-correlation matrices are symmetric.

3.6. **Gaussian/normal probability density function.** For n-variables:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{P})}} \exp\left[-\frac{1}{2}\left(\mathbf{x} - \langle\mathbf{x}\rangle\right)^{\mathrm{T}} \mathbf{P}^{-1} \left(\mathbf{x} - \langle\mathbf{x}\rangle\right)\right], \quad \mathbf{P} = \mathrm{cov}(\mathbf{x},\mathbf{x}).$$

For $n = 1$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \langle x\rangle)^2}{2\sigma^2}\right].$$

For $n = 2$:

$$p(\mathbf{x}) = p(x_1, x_2) = \frac{1}{\sqrt{4\pi^2(\sigma_1^2\sigma_2^2 - P_{12}^2)}} \exp\left[-\frac{1}{2}\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & P_{12} \\ P_{12} & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right].$$

Some basic statistical concepts

A.S. Lawless

University of Reading

Random variable A random variable is a variable which takes on values at random.

Probability distribution function A probability distribution function $P(x)$ describes the probability that x will take on a certain value. Thus the probability that x lies between x_1 and x_2 is given by

$$\int_{x_1}^{x_2} P(x)dx. \tag{1}$$

Expectation value Suppose that a random variable x can take on all values between $-\infty$ and ∞. Then the expectation value of x is given by

$$< x >= \int_{-\infty}^{\infty} xP(x)dx, \tag{2}$$

where $P(x)$ is the probability distribution function of x. The expectation value is a generalization of the mean. While the mean is calculated from a sum over a real data sample, the expectation value sums over a theoretical probability distribution. If a data sample is described by a theoretical distribution then as the size of the data sample tends to infinity, the mean tends to the expectation value. The definitions which follow can be applied to a finite data sample by replacing the expectation value with the arithmetic mean.

We note the properties

$$< x + y >=< x > + < y >, \tag{3}$$

but in general $< xy > \neq < x >< y >$.

Gaussian distribution The Gaussian distribution function (also known as the *normal* distribution) is a particularly important probability distribution function. It takes the form

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \tag{4}$$

It is a bell-shaped curve centred on $x = \mu$, with the width determined by σ. We find that μ is equal to the expectation value (or mean) of x and σ is the *standard deviation* of the distribution (see later definition).

The Gaussian distribution is important, since it describes well the distribution of errors, an important part of data assimilation. We often assume that errors have a Gaussian distribution.

Variance The variance of x, $V(x)$, is given by

$$V(x) = \langle (x - \langle x \rangle)^2 \rangle \tag{5}$$

$$= \langle x^2 \rangle - \langle x \rangle^2 . \tag{6}$$

The variance is a measure of the spread of x around the expectation value $\langle x \rangle$.

Standard deviation The standard deviation is simply the square root of the variance and is usually denoted by the symbol σ, so that

$$\sigma = \sqrt{V(x)} \tag{7}$$

$$= \sqrt{\langle (x - \langle x \rangle)^2 \rangle}. \tag{8}$$

For a Gaussian distribution:

68.27% of the area lies with σ of the mean

95.45% of the area lies with 2σ of the mean

99.73% of the area lies with 3σ of the mean

Covariance Let x, y be two random variables. Then the covariance between x and y is defined as

$$\mathrm{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle \tag{9}$$

$$= \langle xy \rangle - \langle x \rangle \langle y \rangle . \tag{10}$$

The covariance measures the dependence between the two variables. If values of x above the expected value have a tendency to occur with values of y above the expected value, then both terms in (9) will have the same sign and the covariance will be positive. A similar situation occurs if both have lower than expected values together. If however values of x above the expected value occur with values of y below the expected value, then the terms will have the opposite sign and the covariance will be negative. If the variables x and y are independent then $x - \langle x \rangle$ has an equal chance of being multiplied by a positive or negative $y - < y >$ and the covariance will be zero.

We note also that

$$\text{cov}(x, x) = V(x). \tag{11}$$

Covariance matrix Suppose we have n random variables $x_{(1)}, \ldots, x_{(n)}$. Then we can define a covariance between any two variables by

$$\text{cov}(x_{(i)}, x_{(j)}) =< (x_{(i)} - < x_{(i)} >)(x_{(j)} - < x_{(j)} >) > . \tag{12}$$

Then we can easily see that these covariances form an $n \times n$ matrix with entries

$$V_{ij} = \text{cov}(x_{(i)}, x_{(j)}). \tag{13}$$

This matrix is known as the *covariance matrix*. We note two important properties of this matrix:

1. The covariance matrix is symmetric, since $\text{cov}(x_{(i)}, x_{(j)}) = \text{cov}(x_{(j)}, x_{(i)})$.
2. Using (11) we see that the diagonal entries of the covariance matrix are just the variances.

Correlation coefficient The correlation coefficient ρ is a version of the covariance, normalized by the standard deviations to give a dimensionless quantity. It is defined for two variables x, y by

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)}. \tag{14}$$

The correlation coefficient varies beteen -1 and 1. If $\rho = 0$ then the variables are independent and are said to be uncorrelated. If $\rho = -1$ or $\rho = 1$ then the variables are completely correlated and one can be determined from the other.

Notes on statistics of errors

Let us suppose that $T_0(\mathbf{r}, t)$ is some variable which we are trying to measure (eg. temperature) and we have an estimate $T_e(\mathbf{r}, t)$ which has error $\epsilon(\mathbf{r}, t)$. Hence

$$T_e(\mathbf{r}, t) = T_0(\mathbf{r}, t) + \epsilon(\mathbf{r}, t). \tag{15}$$

Then we say that

- The measurement is *unbiased* if $< \epsilon(\mathbf{r}, t) >= 0$.

- The error is not spatially correlated if $< \epsilon(\mathbf{r}_i, t)\epsilon(\mathbf{r}_j, t) >= 0$ for $i \neq j$.
- The error is not temporally correlated if $< \epsilon(\mathbf{r}, t_1)\epsilon(\mathbf{r}, t_2) >= 0$ for $t_1 \neq t_2$.

References

Barlow, R.J. (1989), *Statistics - A guide to the use of statistical methods in the physical sciences*, John Wiley and Sons.

Daley, R (1991), *Atmospheric Data Analysis*, Cambridge University Press.