E1 244: Detection and Estimation

Preliminaries

Linear Algebra, Random Processes, and Optimization Theory

INDIAN INSTITUTE OF SCIENCE

Vectors

- A N-dimensional vector is assumed to be a column vector:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

- Complex conjugate (Hermitian) transpose

$$\mathbf{x}^{\mathrm{H}} = (\mathbf{x}^{\mathrm{T}})^* = [x_1^*, x_2^*, \ldots, x_N^*]$$

Matrices

- An $N \times M$ matrix has N rows and M columns:

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{bmatrix}$$

- Complex conjugate (Hermitian) transpose

$$\mathbf{A}^{\mathrm{H}} = (\mathbf{A}^{\mathrm{T}})^* = (\mathbf{A}^*)^{\mathrm{T}}$$

- Hermitian matrix

$$\mathbf{A} = \mathbf{A}^{\mathrm{H}}$$

E.g.,

$$\mathbf{A} = \begin{bmatrix} 1 & 1+j \\ 1-j & 1 \end{bmatrix}, \quad \text{then} \quad \mathbf{A}^{\mathrm{H}} = \begin{bmatrix} 1 & 1+j \\ 1-j & 1 \end{bmatrix} = \mathbf{A}$$

Vectors

Vector norms: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{N} |x_i|^p\right)^{1/p}$, for $p = 1, 2, \ldots$.

Examples:

Euclidean (2-norm): $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{N} x_i^* x_i\right)^{1/2} = (\mathbf{x}^{\mathrm{H}}\mathbf{x})^{1/2}$

1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i|$

∞-norm: $\|\mathbf{x}\|_\infty = \max_i |x_i|$

Inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y} = \sum_{i=1}^{N} x_i^* y_i$$

- Two vectors are *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$; if the vectors have unit norm, then they are *orthonormal*

Matrices

For $\mathbf{A} \in \mathbb{C}^{M \times N}$

- **2-norm** (spectral norm, operator norm):

$$\|\mathbf{A}\| := \max_{\mathbf{x}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \quad \text{or} \quad \|\mathbf{A}\|^2 := \max_{\mathbf{x}} \frac{\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{H}}\mathbf{x}}$$

  Largest magnification that can be obtained by applying $\mathbf{A}$ to any vector

- **Forbenius norm**

$$\|\mathbf{A}\|_{\mathrm{F}} := \left(\sum_{i=1}^{M} \sum_{j=1}^{N} |a_{ij}|^2 \right)^{1/2} = \sqrt{\mathrm{trace}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}$$

  Represents energies in its entries

Rank of a matrix

Rank

- The rank of $\mathbf{A}$ is the number of independent columns or rows of $\mathbf{A}$

Prototype rank-1 matrix: $\mathbf{A} = \mathbf{a}\mathbf{b}^{\mathrm{H}}$

- The ranks of $\mathbf{A}, \mathbf{A}\mathbf{A}^{\mathrm{H}}$, and $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ are the same

- If $\mathbf{A}$ is square and full rank, there is a unique inverse $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- An $N \times N$ matrix $\mathbf{A}$ has rank N, then $\mathbf{A}$ is invertible $\Leftrightarrow \det(\mathbf{A}) \neq 0$

Linear independence, vector spaces, and basis vectors

Linear independence

- A collection of N vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ is called *linearly independent* if

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_N \mathbf{x}_N = 0 \quad \Leftrightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$$

Subspaces

- The space $\mathcal{H}$ spanned by a collection of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$

$$\mathcal{H} := \{\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_N \mathbf{x}_N | \alpha_i \in \mathbb{C}, \forall i\}$$

  is called a *linear subspace*

  - If the vectors are linearly independent they are called a *basis* for the subspace
  - The number of basis vectors is called the *dimension* of the subspace
  - If the vectors are orthogonal, then we have an *orthogonal basis*
  - If the vectors are orthonormal, then we have an *orthonormal basis*

Fundamental subspaces of $\mathbf{A}$

- Range (column span) of $\mathbf{A} \in \mathbb{C}^{M \times N}$

$$\operatorname{ran}(\mathbf{A}) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{C}^N\} \subset \mathbb{C}^M$$

The dimension of $\operatorname{ran}(\mathbf{A})$ is rank of $\mathbf{A}$, denoted by $\rho(\mathbf{A})$

- Kernel (row null space) of $\mathbf{A} \in \mathbb{C}^{M \times N}$

$$\ker(\mathbf{A}) = \{\mathbf{x} \in \mathbb{C}^N : \mathbf{A}\mathbf{x} = \mathbf{0}\} \subset \mathbb{C}^N$$

The dimension of $\ker(\mathbf{A})$ is $N - \rho(\mathbf{A})$

- Four fundamental subspaces

$$\operatorname{ran}(\mathbf{A}) \oplus \ker(\mathbf{A}^{\mathrm{H}}) = \mathbb{C}^M$$

$$\operatorname{ran}(\mathbf{A}^{\mathrm{H}}) \oplus \ker(\mathbf{A}) = \mathbb{C}^N$$

direct sum: $\mathcal{H}_1 \oplus \mathcal{H}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 | \mathbf{x}_1 \in \mathcal{H}_1, \mathbf{x}_2 \in \mathcal{H}_2\}$

Unitary and Isometry

- A square matrix $\mathbf{U}$ is called *unitary* if $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$ and $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}$
  - Examples are rotation or reflection matrices
  - $\|\mathbf{U}\| = 1$; its rows and columns are orthonormal
- A tall rectangular matrix $\hat{\mathbf{U}}$ is called an *isometry* if $\hat{\mathbf{U}}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$
  - Its columns are orthonormal basis of a subspace (not the complete space)
  - $\|\hat{\mathbf{U}}\| = 1$;
  - There is an orthogonal complement $\hat{\mathbf{U}}^{\perp}$ of $\hat{\mathbf{U}}$ such that $[\hat{\mathbf{U}} \quad \hat{\mathbf{U}}^{\perp}]$ is unitary

Projection

- A square matrix $\mathbf{P}$ is a *projection* if $\mathbf{PP} = \mathbf{P}$
- It is an orthogonal projection if $\mathbf{P}^{\mathrm{H}} = \mathbf{P}$
  - The norm of an orthogonal projection is $\|\mathbf{P}\| = 1$
  - For an isometry $\hat{\mathbf{U}}$, the matrix $\mathbf{P} = \hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}}$ is an orthogonal projection onto the space spanned by the columns of $\hat{\mathbf{U}}$.
- Suppose $\mathbf{U} = [\underbrace{\hat{\mathbf{U}}}_{d} \quad \underbrace{\hat{\mathbf{U}}^{\perp}}_{N-d}]$ is unitary. Then, from $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}_N$:

$$\hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}} + \hat{\mathbf{U}}^{\perp}(\hat{\mathbf{U}}^{\perp})^{\mathrm{H}} = \mathbf{I}_N, \quad \hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}} = \mathbf{P}, \quad \hat{\mathbf{U}}^{\perp}(\hat{\mathbf{U}}^{\perp})^{\mathrm{H}} = \mathbf{P}^{\perp} = \mathbf{I}_N - \mathbf{P}$$

- Any vector $\mathbf{x} \in \mathbb{C}^N$ can be decomposed as $\mathbf{x} = \hat{\mathbf{x}} + \hat{\mathbf{x}}^{\perp}$ with $\hat{\mathbf{x}} \perp \hat{\mathbf{x}}^{\perp}$:

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{x} \in \mathrm{ran}(\hat{\mathbf{U}}) \quad \hat{\mathbf{x}}^{\perp} = \mathbf{P}^{\perp}\mathbf{x} \in \mathrm{ran}(\hat{\mathbf{U}}^{\perp})$$

Singular value decomposition

- For any matrix $\mathbf{X}$, there is a decomposition

  $$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}$$

  Here, $\mathbf{U}$ and $\mathbf{V}$ are unitary, and $\boldsymbol{\Sigma}$ is diagonal with positive real entries.

- Properties:
  - The columns $\mathbf{u}_i$ of $\mathbf{U}$ are called the left singular vectors
  - The columns $\mathbf{v}_i$ of $\mathbf{V}$ are called the right singular vectors
  - The diagonal entries σ_i of $\boldsymbol{\Sigma}$ are called the singular values
  - They are positive, real, and sorted

    $$\sigma_1 \geq \sigma_2 \geq \cdots 0$$

Singular value decomposition

- For an $M \times N$ tall matrix $\mathbf{X}$, there is a decomposition

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}} = [\hat{\mathbf{U}}\ \hat{\mathbf{U}}^{\perp}] \left[\begin{array}{cc|cc} \sigma_1 & & & \\ & \sigma_d & & \\ \hline & & 0 & \\ & & & 0 \\ \hline 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{V}} \\ (\hat{\mathbf{V}}^{\perp})^{\mathrm{H}} \end{array}\right]$$

$$\mathbf{U} : M \times M, \quad \boldsymbol{\Sigma} : M \times N, \mathbf{V} : N \times N$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_d > \sigma_{d+1} = \cdots \sigma_N 0$$

- Economy size SVD: $\mathbf{X} = \hat{\mathbf{U}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}^{\mathrm{H}}$, where $\hat{\boldsymbol{\Sigma}} : d \times d$ is a diagonal matrix containing $\sigma_1, \cdots, \sigma_d$ along the diagonals.

Singular value decomposition

- The rank of $\mathbf{X}$ is d, the number of nonzero singular values
- $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}} \;\Leftrightarrow\; \mathbf{X}^{\mathrm{H}} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathrm{H}} \;\Leftrightarrow\; \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \;\Leftrightarrow\; \mathbf{X}^{\mathrm{H}}\mathbf{U} = \mathbf{V}\mathbf{\Sigma}$
  - The columns of $\hat{\mathbf{U}}$ ($\hat{\mathbf{U}}^{\perp}$) are the orthonormal basis for $\mathrm{ran}(\mathbf{X})$ ($\ker(\mathbf{X}^{\mathrm{H}})$)
  - The columns of $\hat{\mathbf{V}}$ ($\hat{\mathbf{V}}^{\perp}$) are the orthonormal basis for $\mathrm{ran}(\mathbf{X}^{\mathrm{H}})$ ($\ker(\mathbf{X})$)
- $\mathbf{X} = \sum_{i=1}^{d} \sigma_i(\mathbf{u}_i\mathbf{v}_i^{\mathrm{H}})$; $\mathbf{u}_i\mathbf{v}_i^{\mathrm{H}}$ is a rank-1 isometry matrix.
- $\mathbf{X}\mathbf{v}_i = \sigma_i\mathbf{u}_i$
- $\|\mathbf{X}\| = \|\mathbf{X}^{\mathrm{H}}\| = \sigma_1$, the largest singular value.

Eigenvalue decomposition

- The eigenvalue problem is $(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$
- Any λ that makes $\mathbf{A} - \lambda \mathbf{I}$ singular is called an eigenvalue and the corresponding invariant vector is called the eigenvector
- Stacking

$$\mathbf{A}[\mathbf{x}_1 \ \mathbf{x}_2 \cdots] = [\mathbf{x}_1 \ \mathbf{x}_2 \cdots] \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{bmatrix}$$

$$\mathbf{AT} = \mathbf{T\Lambda} \Leftrightarrow \mathbf{A} = \mathbf{T\Lambda T}^{-1}$$

(might exist when $\mathbf{T}$ is invertible and when eigenvalues are distinct)

Eigenvalue decomposition and SVD

- Suppose the SVD of $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}$. Therefore

$$\mathbf{X}\mathbf{X}^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{V}^{\mathrm{H}}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}}$$

  - The eigenvalues of $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ are singular values of $\mathbf{X}$ squared.
  - Eigenvectors of $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ are the left singular vectors of $\mathbf{X}$
  - Eigenvalue decomposition of $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ always exits and SVD always exists.

Pseudo inverse

- For a tall full-column rank matrix $\mathbf{X} : M \times N$

  Pseudo-inverse of $\mathbf{X}$ is $\mathbf{X}^{\dagger} = (\mathbf{X}^{\mathrm{H}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{H}}$.
  - $\mathbf{X}^{\dagger}\mathbf{X} = \mathbf{I}_N$:inverse on the short space
  - $\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{P}_c$: Projector onto $\mathrm{ran}(\mathbf{X})$

- For a tall rank matrix $\mathbf{X} : M \times N$ with rank d, $\mathbf{X}^{\mathrm{H}}\mathbf{X}$ is not invertible.

  Moore-Penrose Pseudo inverse of $\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^{\mathrm{H}}$ is $\mathbf{X}^{\dagger} = \hat{\mathbf{V}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^{\mathrm{H}}$
  1. $\mathbf{X}\mathbf{X}^{\dagger}\mathbf{X} = \mathbf{X}$
  2. $\mathbf{X}^{\dagger}\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{X}^{\dagger}$
  3. $\mathbf{X}\mathbf{X}^{\dagger} = \hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}} = \mathbf{P}_c$:Projector onto $\mathrm{ran}(\mathbf{X})$
  4. $\mathbf{X}^{\dagger}\mathbf{X} = \hat{\mathbf{V}}\hat{\mathbf{V}}^{\mathrm{H}} = \mathbf{P}_r$:Projector onto $\mathrm{ran}(\mathbf{X}^{\mathrm{H}})$

Optimization theory

- The local and global minima of an objective function $f(x)$, with real x, satisfy

$$\frac{\partial f(x)}{\partial x} = \nabla_x f(x) = 0 \quad \text{and} \quad \frac{\partial^2 f(x)}{\partial x^2} = \nabla_x^2 f(x) > 0$$

  If $f(x)$ is convex, then the local minimum is the global minimum

- For $f(z)$ with complex z, we write $f(z)$ as $f(z, z^*)$ and treat $z = x + jy$ and $z^* = x - jy$ as independent variables and define the partial derivatives w.r.t. z and z^* as

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left[\frac{\partial f}{\partial x} - j\frac{\partial f}{\partial y}\right] \quad \text{and} \quad \frac{\partial f}{\partial z^*} = \frac{1}{2}\left[\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}\right]$$

- For an objective function $f(z, z^*)$, the stationary points of $f(z, z^*)$ are found by setting the derivative of $f(z, z^*)$ w.r.t. z or z^* to zero.

Optimization theory

- For an objective function in two or more real variables, $f(x_1, x_2, \ldots, x_N) = f(\mathbf{x})$, the first-order derivative (gradient) and the second-order derivative (Hessian) are given by

$$[\nabla_x f(\mathbf{x})]_i = \frac{\partial f(\mathbf{x})}{\partial x_i} \quad \text{and} \quad [\mathbf{H}(\mathbf{x})]_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

- The local and global minima of an objective function $f(\mathbf{x})$, with real $\mathbf{x}$, satisfy

$$\nabla_x f(\mathbf{x}) = \mathbf{0} \quad \text{and} \quad \mathbf{H}(\mathbf{x}) > 0$$

- For an objective function $f(\mathbf{z}, \mathbf{z}^*)$, the stationary points of $f(\mathbf{z}, \mathbf{z}^*)$ are found by setting the derivative of $f(\mathbf{z}, \mathbf{z}^*)$ w.r.t. $\mathbf{z}$ or $\mathbf{z}^*$ to zero.

Random variables

- A random variable x is a function that assigns a number to each outcome of a random experiment
- Probability distribution function

$$F_x(\alpha) = \Pr\{x \leq \alpha\}$$

- Probability density function

$$f_x(\alpha) = \frac{d}{d\alpha} F_x(\alpha)$$

- Mean or expected value

$$m_x = E\{x\} = \int_{-\infty}^{\infty} \alpha f_x(\alpha) d\alpha$$

- Variance

$$\sigma_x^2 = \mathrm{var}\{x\} = E\{(x - m_x)^2\} = \int_{-\infty}^{\infty} (\alpha - m_x)^2 f_x(\alpha) d\alpha$$

Random variables

- Joint probability distribution function

$$F_{x,y}(\alpha, \beta) = \Pr\{x \le \alpha, y \le \beta\}$$

- Joint density function

$$f_{x,y}(\alpha, \beta) = \frac{\partial^2}{\partial\alpha\partial\beta} F_{x,y}(\alpha, \beta)$$

  - x and y are independent: $f_{x,y}(\alpha, \beta) = f_x(\alpha) f_x(\beta)$

- Correlation

$$r_{xy} = E\{xy^*\}$$

- Covariance

$$c_{xy} = \mathrm{cov}\{x, y\} = E\{(x - m_x)(y - m_y)^*\} = r_{xy} - m_x m_y^*$$

- x and y are uncorrelated: $c_{xy} = 0$ or $E\{xy^*\} = E\{x\}E\{y^*\}$ or $r_{xy} = m_x m_y^*$.

- Independent random variables are always uncorrelated. Converse, is not always true.

Random processes

- A random process $x(n)$ is a sequence of random variables
- Mean and variance:

$$m_x = E\{x\} \quad \text{and} \quad \sigma_x^2(n) = E\{|x(n) - m_x(n)|^2\}$$

- Autocorrelation and autocovariance

$$r_x(k,l) = E\{x(k)x^*(l)\}$$

$$c_x(k,l) = E\{[x(k) - m_x(k)][x(l) - m_x(l)]^*\}$$

Stationarity

- First-order stationarity if $f_{x(n)}(\alpha) = f_{x(n+k)}(\alpha)$. This implies $m_x(n) = m_x(0) = m_x$.

- Second-order stationarity if, for any k, the process $x(n)$ and $x(n+k)$ have the same second-order density function:

$$f_{x(n_1),x(n_2)}(\alpha_1, \alpha_2) = f_{x(n_1+k),x(n_2+k)}(\alpha_1, \alpha_2).$$

  This implies $r_x(k,l) = r_x(k-l,0) = r_x(k-l)$.

Wide-sense stationarity

- Wide-sense stationary (WSS):

$$m_x(n) = m_x; \quad r_x(k,l) = r_{xy}(k-l); \quad c_x(0) < \infty.$$

- Properties of WSS processes:
  - Symmetry: $r_x(k) = r_x^*(-k)$
  - mean-square value: $r_x(0) = E\{|x(n)|^2\} \geq 0.$
  - maximum value: $r_x(0) \geq |r_x(k)|$
  - mean-squared periodic: $r_x(k_0) = r_x(0)$
- Power spectrum: discrete Fourier transform of the deterministic sequence $r_x(k)$

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k) e^{-j\omega}$$

Autocorrelation and autocovariance matrices

- Consider a WSS process $x(n)$ and collect $p+1$ samples in

$$\mathbf{x} = [x(0), x(1), \ldots, x(p)]^T$$

- Autocorrelation matrix:

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \begin{bmatrix} r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(p) \\ r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ \vdots & \vdots & \vdots & \cdots & \cdots \\ r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix}$$

- $\mathbf{R}_x$ is Toeplitz, Hermitian, and nonnegative definite.

- Autocovariance matrix: $\mathbf{C}_x = \mathbf{R}_x - \mathbf{m}_x\mathbf{m}_x^{\mathrm{H}}$,
  where $\mathbf{m}_x = m_x\mathbf{1}$

Gaussian Processes

- Suppose $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ is a vector of n real-valued random variables.

- Then $\mathbf{x}$ is said to be a Gaussian random vector and the random variables x_i are said to be jointly Gaussian if the joint probability density function is

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{R}_x|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_x)^{\mathrm{T}} \mathbf{R}_x^{-1} (\mathbf{x} - \mathbf{m}_x) \right\}$$