Lecture Notes 3
Random Vectors

- Specifying a Random Vector
- Mean and Covariance Matrix
- Coloring and Whitening
- Gaussian Random Vectors

Specifying a Random Vector

- Let $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space. We define a random vector (RV) as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- $\mathbf{X}$ is completely specified by its joint cdf for $\mathbf{x} = (x_1, x_2, \ldots, x_n)$:

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}\{X_1 \le x_1,\, X_2 \le x_2,\, \ldots,\, X_n \le x_n\}\,, \quad \mathbf{x} \in \mathbf{R}^n$$

- If $\mathbf{X}$ is continuous, i.e., $F_{\mathbf{X}}(\mathbf{x})$ is a continuous function of $\mathbf{x}$, then $\mathbf{X}$ can be specified by its joint pdf:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)\,, \quad \mathbf{x} \in \mathbf{R}^n$$

- If $\mathbf{X}$ is discrete then it can be specified by its joint pmf:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)\,, \quad \mathbf{x} \in \mathcal{X}^n$$

- A marginal cdf (pdf, pmf) is the joint cdf (pdf, pmf) for a subset of $\{X_1, \ldots, X_n\}$; e.g., for

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

  the marginals are

$$f_{X_1}(x_1)\,,\; f_{X_2}(x_2)\,,\; f_{X_3}(x_3)$$
$$f_{X_1,X_2}(x_1, x_2)\,,\; f_{X_1,X_3}(x_1, x_3)\,,\; f_{X_2,X_3}(x_2, x_3)$$

- The marginals can be obtained from the joint in the usual way. For the previous example,

$$F_{X_1}(x_1) = \lim_{x_2, x_3 \to \infty} F_{\mathbf{X}}(x_1, x_2, x_3)$$
$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\, dx_3$$

- Conditional cdf (pdf, pmf) can also be defined in the usual way. E.g., the conditional pdf of $\mathbf{X}_{k+1}^n = (X_{k+1}, \ldots, X_n)$ given $\mathbf{X}^k = (X_1, \ldots, X_k)$ is

$$f_{\mathbf{X}_{k+1}^n | \mathbf{X}^k}(\mathbf{x}_{k+1}^n | \mathbf{x}^k) = \frac{f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)}{f_{\mathbf{X}^k}(x_1, x_2, \ldots, x_k)} = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}^k}(\mathbf{x}^k)}$$

- Chain Rule: We can write

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1,X_2}(x_3|x_1, x_2) \cdots f_{X_n|\mathbf{X}^{n-1}}(x_n|\mathbf{x}^{n-1})$$

Proof: By induction. The chain rule holds for $n = 2$ by definition of conditional pdf. Now suppose it is true for $n - 1$. Then

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{X}^{n-1}}(\mathbf{x}^{n-1}) f_{X_n|\mathbf{X}^{n-1}}(x_n|\mathbf{x}^{n-1}) \\ &= f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) \cdots f_{X_{n-1}|\mathbf{X}^{n-2}}(x_{n-1}|\mathbf{x}^{n-2}) f_{X_n|\mathbf{X}^{n-1}}(x_n|\mathbf{x}^{n-1}), \end{aligned}$$

which completes the proof

Independence and Conditional Independence

- Independence is defined in the usual way; e.g., $X_1, X_2, \ldots, X_n$ are independent if

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{for all } (x_1, \ldots, x_n)$$

- Important special case, i.i.d. r.v.s: $X_1, X_2, \ldots, X_n$ are said to be independent, identically distributed (i.i.d.) if they are independent and have the same marginals

  Example: if we flip a coin n times independently, we generate i.i.d. $\mathrm{Bern}(p)$ r.v.s. $X_1, X_2, \ldots, X_n$

- R.v.s X_1 and X_3 are said to be conditionally independent given X_2 if

$$f_{X_1,X_3|X_2}(x_1, x_3|x_2) = f_{X_1|X_2}(x_1|x_2) f_{X_3|X_2}(x_3|x_2) \quad \text{for all } (x_1, x_2, x_3)$$

- Conditional independence neither implies nor is implied by independence; X_1 and X_3 independent given X_2 does not mean that X_1 and X_3 are independent (or vice versa)

- Example: Coin with random bias. Given a coin with random bias $P \sim f_P(p)$, flip it n times independently to generate the r.v.s $X_1, X_2, \ldots, X_n$, where $X_i = 1$ if i-th flip is heads, 0 otherwise
  - $X_1, X_2, \ldots, X_n$ are not independent
  - However, $X_1, X_2, \ldots, X_n$ are conditionally independent given P; in fact, they are i.i.d. $\mathrm{Bern}(p)$ for every $P = p$
- Example: Additive noise channel. Consider an additive noise channel with signal X, noise Z, and observation $Y = X + Z$, where X and Z are independent
  - Although X and Z are independent, they are not in general conditionally independent given Y

Mean and Covariance Matrix

- The mean of the random vector $\mathbf{X}$ is defined as

$$\mathrm{E}(\mathbf{X}) = \begin{bmatrix} \mathrm{E}(X_1) & \mathrm{E}(X_2) & \cdots & \mathrm{E}(X_n) \end{bmatrix}^T$$

- Denote the covariance between X_i and X_j, $\mathrm{Cov}(X_i, X_j)$, by σ_{ij} (so the variance of X_i is denoted by σ_{ii}, $\mathrm{Var}(X_i)$, or $\sigma^2_{X_i}$)

- The covariance matrix of $\mathbf{X}$ is defined as

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

- For $n = 2$, we can use the definition of correlation coefficient to obtain

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2_{X_1} & \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} \\ \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} & \sigma^2_{X_2} \end{bmatrix}$$

Properties of Covariance Matrix $\Sigma_{\mathbf{X}}$

- $\Sigma_{\mathbf{X}}$ is real and symmetric (since $\sigma_{ij} = \sigma_{ji}$)
- $\Sigma_{\mathbf{X}}$ is positive semidefinite, i.e., the quadratic form

$$\mathbf{a}^T \Sigma_{\mathbf{X}} \mathbf{a} \ge 0 \quad \text{for every real vector } \mathbf{a}$$

  Equivalently, all the eigenvalues of $\Sigma_{\mathbf{X}}$ are nonnegative, and also all principal minors are nonnegative

- To show that $\Sigma_{\mathbf{X}}$ is positive semidefinite we write

$$\Sigma_{\mathbf{X}} = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}(\mathbf{X}))(\mathbf{X} - \mathrm{E}(\mathbf{X}))^T\right],$$

  i.e., as the expectation of an outer product. Thus

$$\begin{aligned} \mathbf{a}^T \Sigma_{\mathbf{X}} \mathbf{a} &= \mathbf{a}^T \, \mathrm{E}\left[(\mathbf{X} - \mathrm{E}(\mathbf{X}))(\mathbf{X} - \mathrm{E}(\mathbf{X}))^T\right] \mathbf{a} \\ &= \mathrm{E}\left[\mathbf{a}^T (\mathbf{X} - \mathrm{E}(\mathbf{X}))(\mathbf{X} - \mathrm{E}(\mathbf{X}))^T \mathbf{a}\right] \\ &= \mathrm{E}\left[(\mathbf{a}^T (\mathbf{X} - \mathrm{E}(\mathbf{X})))^2\right] \ \ge \ 0 \end{aligned}$$

Which of the Following Can Be a Covariance Matrix ?

1. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$

4. $\begin{bmatrix} -1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$

Coloring and Whitening

- Square root of covariance matrix: Let Σ be a covariance matrix. Then there exists an $n \times n$ matrix $\Sigma^{1/2}$ such that $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T$. The matrix $\Sigma^{1/2}$ is called the square root of Σ

- Coloring: Let $\mathbf{X}$ be white RV, i.e., has zero mean and $\Sigma_{\mathbf{X}} = aI$, $a > 0$. Assume without loss of generality that $a = 1$

  Let Σ be a covariance matrix, then the RV $\mathbf{Y} = \Sigma^{1/2}\mathbf{X}$ has covariance matrix Σ (why?)

  Hence we can generate a RV with any prescribed covariance from a white RV

- Whitening: Given a zero mean RV $\mathbf{Y}$ with nonsingular covariance matrix Σ, then the RV $\mathbf{X} = \Sigma^{-1/2}\mathbf{Y}$ is white

  Hence, we can generate a white RV from any RV with nonsingular covariance matrix

- Coloring and whitening have applications in simulations, detection, and estimation

Finding Square Root of Σ

- For convenience, we assume throughout that Σ is nonsingular
- Since Σ is symmetric, it has n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and n corresponding orthogonal eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$

  Further, since Σ is positive definite, the eigenvalues are all positive
- Thus, we have

$$\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad \lambda_i > 0,\ i = 1, 2, \ldots, n$$

$$\mathbf{u}_i^T \mathbf{u}_j = 0 \quad \text{for every } i \neq j$$

  Without loss of generality assume that the $\mathbf{u}_i$ vectors are unit vectors
- The first set of equations can be rewritten in the matrix form

$$\Sigma U = U \Lambda,$$

  where

$$U = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n]$$

  and Λ is a diagonal matrix with diagonal elements λ_i

- Note that U is a unitary matrix ($U^TU = UU^T = I$), hence

$$\Sigma = U\Lambda U^T$$

and the square root of Σ is

$$\Sigma^{1/2} = U\Lambda^{1/2},$$

where $\Lambda^{1/2}$ is a diagonal matrix with diagonal elements $\lambda_i^{1/2}$

- The inverse of the square root is straightforward to find as

$$\Sigma^{-1/2} = \Lambda^{-1/2}U^T$$

- Example: Let

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

To find the eigenvalues of Σ, we find the roots of the polynomial equation

$$\det(\Sigma - \lambda I) = \lambda^2 - 5\lambda + 5 = 0,$$

which gives $\lambda_1 = 3.62$, $\lambda_2 = 1.38$

To find the eigenvectors, consider

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = 3.62 \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix},$$

and $u_{11}^2 + u_{12}^2 = 1$, which yields

$$\mathbf{u}_1 = \begin{bmatrix} 0.53 \\ 0.85 \end{bmatrix}$$

Similarly, we can find the second eigenvector

$$\mathbf{u}_2 = \begin{bmatrix} -0.85 \\ 0.53 \end{bmatrix}$$

Hence,

$$\Sigma^{1/2} = \begin{bmatrix} 0.53 & -0.85 \\ 0.85 & 0.53 \end{bmatrix} \begin{bmatrix} \sqrt{3.62} & 0 \\ 0 & \sqrt{1.38} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1.62 & 0.62 \end{bmatrix}$$

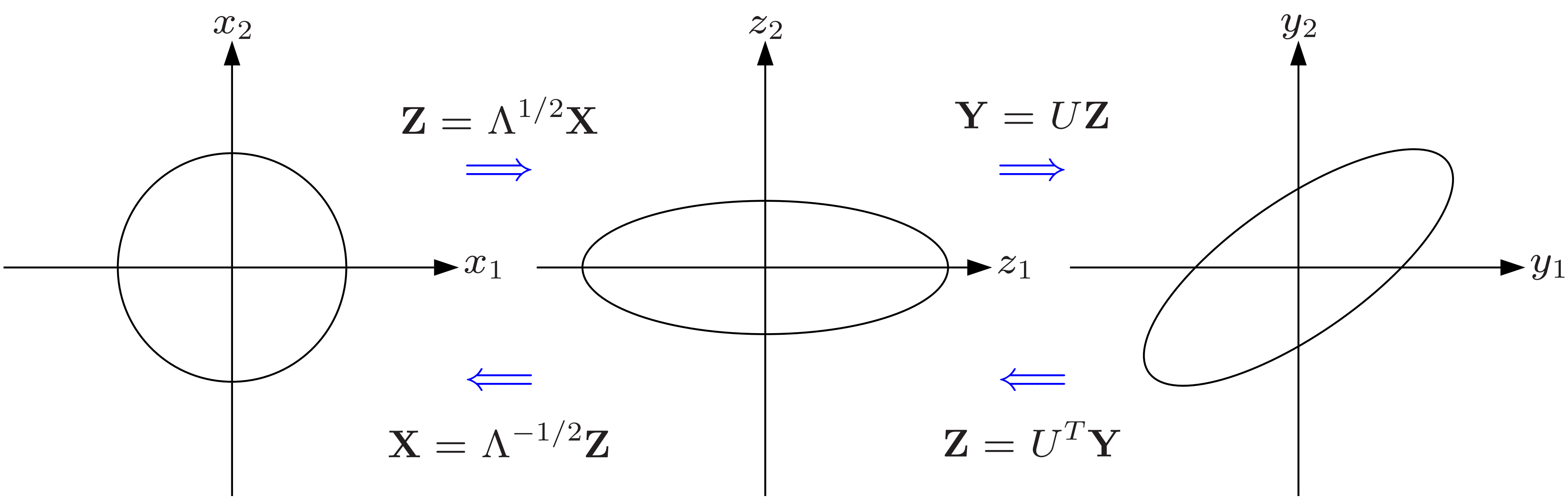
The inverse of the square root is

$$\Sigma^{-1/2} = \begin{bmatrix} 1/\sqrt{3.62} & 0 \\ 0 & 1/\sqrt{1.38} \end{bmatrix} \begin{bmatrix} 0.53 & 0.85 \\ -0.85 & 0.53 \end{bmatrix} = \begin{bmatrix} 0.28 & 0.45 \\ -0.72 & 0.45 \end{bmatrix}$$

Geometric Interpretation

- To generate a RV $\mathbf{Y}$ with covariance matrix Σ from white RV $\mathbf{X}$, we use the transformation $\mathbf{Y} = U\Lambda^{1/2}\mathbf{X}$
- Equivalently, we first scale each component of $\mathbf{X}$ to obtain the RV $\mathbf{Z} = \Lambda^{1/2}\mathbf{X}$

  And then rotate $\mathbf{Z}$ using U to obtain $\mathbf{Y} = U\mathbf{Z}$
- We can visualize this by plotting $\mathbf{x}^T I \mathbf{x} = c$, $\mathbf{z}^T \Lambda \mathbf{z} = c$, and $\mathbf{y}^T \Sigma \mathbf{y} = c$



Cholesky Decomposition

- Σ has many square roots:

  If $\Sigma^{1/2}$ is a square root, then for any unitary matrix V, $\Sigma^{1/2}V$ is also a square root since $\Sigma^{1/2}VV^T(\Sigma^{1/2})^T = \Sigma$

- The Cholesky decomposition is an efficient algorithm for computing lower triangle square root that can be used to perform coloring causally (sequentially)

- For $n = 3$, we want to find a lower triangle matrix (square root) A such that

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ 0 & a_{22} & a_{32} \\ 0 & 0 & a_{33} \end{bmatrix}$$

  The elements of A are computed in a raster scan manner:

  a_{11}: $\sigma_{11} = a_{11}^2 \Rightarrow a_{11} = \sqrt{\sigma_{11}}$

  a_{21}: $\sigma_{21} = a_{21}a_{11} \Rightarrow a_{21} = \sigma_{21}/a_{11}$

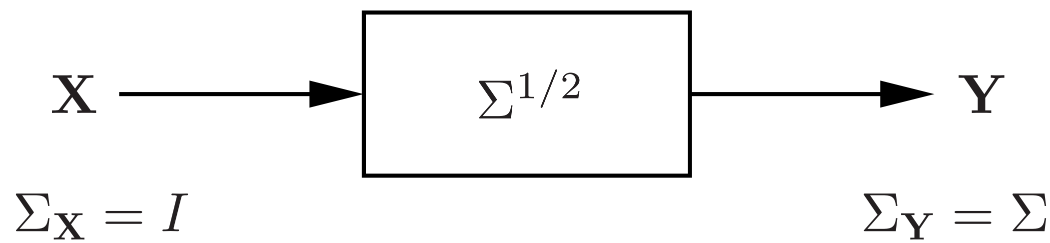
  a_{22}: $\sigma_{22} = a_{21}^2 + a_{22}^2 \Rightarrow a_{22} = \sqrt{\sigma_{22} - a_{21}^2}$

  a_{31}: $\sigma_{31} = a_{11}a_{31} \Rightarrow a_{31} = \sigma_{31}/a_{11}$

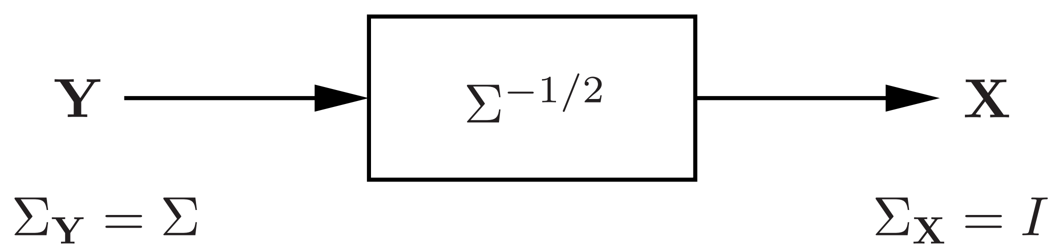
a_{32}: $\sigma_{32} = a_{21}a_{31} + a_{22}a_{32} \Rightarrow a_{32} = (\sigma_{32} - a_{21}a_{31})/a_{22}$

a_{33}: $\sigma_{33} = a_{31}^2 + a_{32}^2 + a_{33}^2 \Rightarrow a_{33} = \sqrt{\sigma_{33} - a_{31}^2 - a_{32}^2}$

- The inverse of a lower triangle square root is also lower triangular
- Coloring and whitening summary:
  - Coloring:



  - Whitening:



  - Lower triangle square root and its inverse can be efficiently computed using Cholesky decomposition

Gaussian Random Vectors

- A random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is a Gaussian random vector (GRV) (or $X_1, X_2, \ldots, X_n$ are jointly Gaussian r.v.s) if the joint pdf is of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

  where $\boldsymbol{\mu}$ is the mean and Σ is the covariance matrix of $\mathbf{X}$, and $\left|\Sigma\right| > 0$, i.e., Σ is positive definite

- Verify that this joint pdf is the same as the case $n = 2$ from Lecture Notes 2

- Notation: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ denotes a GRV with given mean and covariance matrix

- Since Σ is positive definite, Σ^{-1} is positive definite. Thus if $\mathbf{x} - \boldsymbol{\mu} \neq \mathbf{0}$,

$$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) > 0,$$

  which means that the contours of equal pdf are ellipsoids

- The GRV $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, aI)$, where I is the identity matrix and $a > 0$, is called white; its contours of equal joint pdf are spheres centered at the origin

Properties of GRVs

- Property 1: For a GRV, uncorrelation implies independence

  This can be verified by substituting $\sigma_{ij} = 0$ for all $i \neq j$ in the joint pdf.

  Then Σ becomes diagonal and so does Σ^{-1}, and the joint pdf reduces to the product of the marginals $X_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$

  For the white GRV $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, aI)$, the r.v.s are i.i.d. $\mathcal{N}(0, a)$

- Property 2: Linear transformation of a GRV yields a GRV, i.e., given any $m \times n$ matrix A, where $m \leq n$ and A has full rank m, then

$$\mathbf{Y} = A\mathbf{X} \sim \mathcal{N}(A\boldsymbol{\mu},\, A\Sigma A^T)$$

- Example: Let

$$\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\right)$$

  Find the joint pdf of

$$\mathbf{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{X}$$

Solution: From Property 2, we conclude that

$$\mathbf{Y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 3\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\right) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}7 & 3\\3 & 2\end{bmatrix}\right)$$

Before we prove Property 2, let us show that

$$\mathrm{E}(\mathbf{Y}) = A\boldsymbol{\mu} \quad \text{and} \quad \Sigma_{\mathbf{Y}} = A\Sigma A^T$$

These results follow from linearity of expectation. First, expectation:

$$\mathrm{E}(\mathbf{Y}) = \mathrm{E}(A\mathbf{X}) = A\,\mathrm{E}(\mathbf{X}) = A\boldsymbol{\mu}$$

Next consider the covariance matrix:

$$\begin{aligned}\Sigma_{\mathbf{Y}} &= \mathrm{E}\left[(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))^T\right]\\ &= \mathrm{E}\left[(A\mathbf{X} - A\boldsymbol{\mu})(A\mathbf{X} - A\boldsymbol{\mu})^T\right]\\ &= A\,\mathrm{E}\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\right]A^T = A\Sigma A^T\end{aligned}$$

Of course this is not sufficient to show that $\mathbf{Y}$ is a GRV — we must also show that the joint pdf has the right form

We do so using the characteristic function for a random vector

- Definition: If $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x})$, the characteristic function of $\mathbf{X}$ is

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = \mathrm{E}\left(e^{i\boldsymbol{\omega}^T\mathbf{X}}\right),$$

where $\boldsymbol{\omega}$ is an n-dimensional real valued vector and $i = \sqrt{-1}$

Thus

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})e^{i\boldsymbol{\omega}^T\mathbf{x}}\,d\mathbf{x}$$

This is the inverse of the multi-dimensional Fourier transform of $f_{\mathbf{X}}(\mathbf{x})$, which implies that there is a one-to-one correspondence between $\Phi_{\mathbf{X}}(\boldsymbol{\omega})$ and $f_{\mathbf{X}}(\mathbf{x})$. The joint pdf can be found by taking the Fourier transform of $\Phi_{\mathbf{X}}(\boldsymbol{\omega})$, i.e.,

$$f_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^n}\Phi_{\mathbf{X}}(\boldsymbol{\omega})e^{-i\boldsymbol{\omega}^T\mathbf{x}}\,d\boldsymbol{\omega}$$

- Example: The characteristic function for $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\Phi_X(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2 + i\mu\omega},$$

and for a GRV $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$,

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = e^{-\frac{1}{2}\boldsymbol{\omega}^T\Sigma\boldsymbol{\omega} + i\boldsymbol{\omega}^T\boldsymbol{\mu}}$$

- Now let's go back to proving Property 2

  Since A is an $m \times n$ matrix, $\mathbf{Y} = A\mathbf{X}$ and $\boldsymbol{\omega}$ are m-dimensional. Therefore the characteristic function of $\mathbf{Y}$ is

$$\begin{aligned}\Phi_{\mathbf{Y}}(\boldsymbol{\omega}) &= \mathrm{E}\left(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\right) \\ &= \mathrm{E}\left(e^{i\boldsymbol{\omega}^T A\mathbf{X}}\right) \\ &= \Phi_{\mathbf{X}}(A^T\boldsymbol{\omega}) \\ &= e^{-\frac{1}{2}(A^T\boldsymbol{\omega})^T\Sigma(A^T\boldsymbol{\omega}) + i\boldsymbol{\omega}^T A\boldsymbol{\mu}} \\ &= e^{-\frac{1}{2}\boldsymbol{\omega}^T(A\Sigma A^T)\boldsymbol{\omega} + i\boldsymbol{\omega}^T A\boldsymbol{\mu}}\end{aligned}$$

  Thus $\mathbf{Y} = A\mathbf{X} \sim \mathcal{N}(A\boldsymbol{\mu}, A\Sigma A^T)$

- An equivalent definition of GRV: $\mathbf{X}$ is a GRV iff for every real vector $\mathbf{a} \neq 0$, the r.v. $Y = \mathbf{a}^T\mathbf{X}$ is Gaussian (see HW for proof)

- Whitening transforms a GRV to a white GRV; conversely, coloring transforms a white GRV to a GRV with prescribed covariance matrix

- Property 3: Marginals of a GRV are Gaussian, i.e., if $\mathbf{X}$ is GRV then for any subset $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$ of indexes, the RV

$$\mathbf{Y} = \begin{bmatrix} X_{i_1} \\ X_{i_2} \\ \vdots \\ X_{i_k} \end{bmatrix}$$

is a GRV

- To show this we use Property 2. For example, let $n = 3$ and $\mathbf{Y} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$

We can express $\mathbf{Y}$ as a linear transformation of $\mathbf{X}$:

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$$

Therefore

$$\mathbf{Y} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} \right)$$

- As we have seen in Lecture Notes 2, the converse of Property 3 does not hold in general, i.e., Gaussian marginals do not necessarily mean that the r.v.s are jointly Gaussian

- Property 4: Conditionals of a GRV are Gaussian, more specifically, if

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ -- \\ \mathbf{X}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ -- \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & | & \Sigma_{12} \\ -- & | & -- \\ \Sigma_{21} & | & \Sigma_{22} \end{bmatrix} \right),$$

where $\mathbf{X}_1$ is a k-dim RV and $\mathbf{X}_2$ is an $n-k$-dim RV, then

$$\mathbf{X}_2 \,|\, \{\mathbf{X}_1 = \mathbf{x}\} \sim \mathcal{N}\left(\Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2\,,\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right)$$

Compare this to the case of $n = 2$ and $k = 1$:

$$X_2 \,|\, \{X_1 = x\} \sim \mathcal{N}\left(\frac{\sigma_{21}}{\sigma_{11}}(x - \mu_1) + \mu_2\,,\ \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}} \right)$$

- Example:

$$\begin{bmatrix} X_1 \\ -- \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ -- \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & | & 2 & 1 \\ -- & | & -- & -- \\ 2 & | & 5 & 2 \\ 1 & | & 2 & 9 \end{bmatrix} \right)$$

From Property 4, it follows that

$$\mathrm{E}(\mathbf{X}_2 \mid X_1 = x) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} (x-1) + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2x \\ x+1 \end{bmatrix}$$

$$\Sigma_{\{\mathbf{X}_2 \mid X_1 = x\}} = \begin{bmatrix} 5 & 2 \\ 2 & 9 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$$

- The proof of Property 4 follows from properties 1 and 2 and the orthogonality principle (HW exercise)