\[ M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \quad v = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]

\[ M v = \begin{bmatrix} m_{11} x_1 + m_{12} x_2 \\ m_{21} x_1 + m_{22} x_2 \end{bmatrix} \quad m^T = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \]

The inverse \( m^{-1} \) satisfies \( m^{-1} m = m m^{-1} = I \).

\[ \text{determinant = volume of the parallel piped which } M \text{ maps the unit hyper cube on to} \]

\[ \text{area} = |m| \]

If \( M v = \lambda v \) then \( v \) is said to be an eigenvector with eigenvalue \( \lambda \).

\[ M x = \lambda x \Rightarrow (M - \lambda I) v = 0 \]

so \( \det(M - \lambda I) = 0 \)

\( p(\lambda) \) is the characteristic polynomial. The values of \( \lambda \) for which \( p(\lambda) = 0 \) are the eigenvalues of \( M \).

Let \( x_i \in \mathbb{F}^n \) e.g. \( x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \)

\( x_1, x_2, \ldots, x_n \) are linearly dependent if \( \exists \lambda_1, \lambda_2, \ldots, \lambda_n \) such that \( \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n = 0 \).

\( x_1, x_2, \ldots, x_n \) are linearly independent if they are not linearly dependent.

A maximal set of non-zero, linearly independent vectors is called a basis.

A basis \( \{ \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n \} \) is orthonormal if \( \vec{x}_i \cdot \vec{x}_j = 1 \) if \( i = j \) and \( 0 \) otherwise.

\( M^\ast \in \mathbb{F}^{n \times n} \) is the conjugate transpose of \( M \).
If $M = M^*$, $M$ is called Hermitian.
If $x^* M x \geq 0 \quad \forall x \in \mathbb{C}^n$, $M$ is called positive semi-definite.
If $x^* M x > 0 \quad \forall x \in \mathbb{C}^n$, $M$ is called positive definite.

Fact: if $M \in \mathbb{C}^{n \times n}$ is Hermitian, then

\begin{enumerate}
\item[a)] the eigenvalues of $M$ are real.
\item[b)] there exist an orthonormal basis of eigenvectors $\langle v_1, v_2, \ldots, v_n \rangle$ \quad \( (v_i^* v_j = \delta_{i=j}) \), $M v_i = \lambda_i v_i$.
\item[c)] if $P = [v_1 \mid v_2 \mid \ldots \mid v_n]$ then $P^* = P^{-1}$.
\end{enumerate}

\[ P^* M P = \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ 0 & \cdots & \lambda_n \end{bmatrix} \]

\[ M = P \Lambda P^* \]

Proof: see linear algebra notes.
Random Vectors
\(x_1, \ldots, x_n\) jointly defined on \((\mathbb{R}, \tau, P)\)

\[
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\]

\[
E[x(Cx)] = \int_{\mathbb{R}^n} h(x) f_x(x) \, dx
\]

\(\mathbf{x}\) has a joint p.d.f. \(f_x\)

\[
E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} = \mathbf{m}_x
\]

\[
K_x = E[(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T]
\]

\(K_x\) is called the covariance matrix

Properties of \(K_x\)

1) \(K_x\) is symmetric

2) \(K_x\) is pos. sem. def.

\[
\mathbf{v}^T K_x \mathbf{v} = \mathbf{v}^T E[(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T] \mathbf{v} = E[(\mathbf{v}^T \mathbf{x} - \mathbf{v}^T \mathbf{m}_x)(\mathbf{v}^T \mathbf{x} - \mathbf{v}^T \mathbf{m}_x)^T] = \text{Var}(\mathbf{v}^T \mathbf{x}) \geq 0
\]

\[
K_{xy} = E[(x-y)(y-y)^T]
\]

\(K_{xy}\) is called a cross-covariance matrix.

MGF for random vectors

\[
g_{\mathbf{x}}(s) = E[e^{s^T \mathbf{x}}]
\]
Last Time:

\[ K_{\hat{x}} = E\left[(\hat{x} - \hat{M}_{\hat{x}})(X - M_x)^T\right] \]

Properties:
1) \( K_{\hat{x}} \) is symmetric
2) \( K_{\hat{x}} \) is pos. sem. definite

(For \( v, \hat{v}^T K_{\hat{x}} v \geq 0 \))

Random vectors \( \hat{x}, \hat{y} \)

\[ K_{xy} = E\left[(\hat{x} - \hat{M}_{\hat{x}}) (\hat{y} - \hat{M}_{\hat{y}})^T\right] \]

This is called a cross-covariance matrix.

---

Moment Generating Function

\[ g_{\hat{x}}(s) = \mathbb{E}\left[e^{s^T \hat{x}}\right] \]

\[ \mathbb{E}\left[e^{s^T \hat{x}}\right] = \int e^{s^T \hat{x}} f_{\hat{x}}(x) \, dx \]

---

Functions of a Random Vector

Suppose \( \hat{x} \) has a density \( f_{\hat{x}} \) and \( h : \mathbb{R}^n \to \mathbb{R}^n \) is one-to-one.

Let \( \hat{z} = h(\hat{x}) \)

What is the density of \( \hat{z} \)?
\[ p(\hat{z} \in B(z)) = \sum_{B(z)} \int_{f(z)} f(z) \, dz \equiv f(z) \circ \text{area}(B(z)) \]

\[ = p(\hat{h}(\hat{z}) \in \hat{h}(B(z))) \]

\[ = p(\hat{x} \in \hat{h}(B(z))) \]

\[ \equiv f(x) (h^{-1}(z)) \circ \text{area}(h^{-1}(B(z))) \]

\[ f_{\hat{z}}(z) \equiv f_{\hat{x}}(h^{-1}(z)) \cdot \frac{\text{area}(h^{-1}(B(z)))}{\text{area}(B(z))} \]

\[ \text{Note: } h^{-1}(\hat{z} + \varepsilon) = \begin{bmatrix} \frac{dh_{1}(\hat{z})}{d\hat{z}} & \cdots & \frac{dh_{n}(\hat{z})}{d\hat{z}} \\ \vdots & \ddots & \vdots \\ \frac{dh_{1}(\hat{z})}{d\varepsilon} & \cdots & \frac{dh_{n}(\hat{z})}{d\varepsilon} \end{bmatrix} \varepsilon + h(\hat{z}) \]
\[
\frac{\mathbf{\hat{y}}}{\mathbf{J}} = J \text{ Jacobian matrix of } \mathbf{h}(\mathbf{y})
\]

\[
f_{\mathbf{\hat{y}}}(\mathbf{z}) = f_{\mathbf{\hat{x}}}(\mathbf{h}(\mathbf{z})) \left| \det (J(\mathbf{\hat{z}})) \right|
\]

\[\uparrow\]
\[J \text{ is Jacobian of } \mathbf{h}(\mathbf{y})\]

Suppose \( J^* \) is Jacobian of \( \mathbf{h}(\mathbf{y}) \)

We can show this:

\[
f_{\mathbf{\hat{y}}}(\mathbf{z}) = f_{\mathbf{\hat{x}}}(\mathbf{h}(\mathbf{z})) \left| \det (J^*(\mathbf{h}(\mathbf{z}))) \right|^{-1}
\]

**Gaussian Random Vectors**

\( \mathbf{W} \sim N(0, \mathbf{I}) \) means that

\[
f_{\mathbf{W}}(\mathbf{w}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \mathbf{w}^2}
\]

Why give attention to Gaussian?

1. They are common

2. They are easy to work with
   - Easy conditional density and expectation calculations
   - Preserved by linear systems

3. Elegant solutions to Kalman and Wiener Filter problems.
why common?

Central Limit Theorem

Roughly, suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d). Then

$$\frac{1}{\sigma \sqrt{n}} \sum_{i=1}^{n} (X_i - N) \sim N(0, 1) \quad N = E[X_i]$$

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - N) \to 0$$

$$\sigma^2 = Var(X_i)$$

In general,

$$X = N(\mu, \sigma^2)$$

Means

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A collection $\langle X_1, \ldots, X_m \rangle$ is said to be Jointly Gaussian (JO) if for all $a \in \mathbb{R}^m$, $\sum_{i=1}^{m} a_i X_i$ is Gaussian

Example of individually but not jointly Gaussian

Let $X$ be $N(0, 1)$

Let $Y$ be $\langle X \text{ if } |X| \leq 1 \rangle$. $Y$ is Gaussian
But \[ \tilde{z} = x + y = \begin{cases} 2x & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases} \]

Thus \( x + y \) is not \( \mathcal{G} \).

Let \( \tilde{x} \in \mathbb{R}^n \) be a vector with iid components \( X_i \sim N(0, 1) \). Then \( \tilde{x} \) is a \( \mathcal{G} \) random vector.

**Proof:** Let \( Y = a^T \tilde{x} \)

\[
\begin{align*}
G_Y(s) &= E \left[ e^{s a^T \tilde{x}} \right] \\
&= \prod_{i=1}^{n} E \left[ e^{sa_i X_i} \right] \\
&= \prod_{i=1}^{n} \left( 1 + \frac{s^2a_i^2}{2} \right)^{-1} \\
&= \frac{1}{\prod_{i=1}^{n} \left( 1 + \frac{s^2a_i^2}{2} \right)}^n = e^{\frac{1}{2}s^2\sigma^2} \\
\end{align*}
\]

\( \sigma^2 = \frac{\sum a_i^2}{m} \)

Yes \( Y \) is Gaussian.

Thus \( \tilde{x} \) is \( \mathcal{G} \).
Definition: We say $\tilde{x} = N(\tilde{\mu}, K_{\tilde{x}})$ if $\tilde{x}$ is i.i.d.

with covariance $K_{\tilde{x}}$, and mean $\tilde{\mu}$.

Fact: $g_{\tilde{x}}(s) = \exp\left(s^T \tilde{\mu} + \frac{1}{2} s^T K_{\tilde{x}} s\right)$

Proof: $g_{\tilde{x}}(s) = E\left[\exp(s^T \tilde{x})\right]$

let $y' = s^T \tilde{x}$

$= E\left[\exp(y')\right]$  

$= g_y(1) = c \left[\frac{1}{\sigma_y^2} \sqrt{\frac{s^T \tilde{\mu}}{\sigma_y^2}} + \frac{1}{2} s^T \tilde{K} s\right]_{s^T=1}$

$\sigma_y^2 := \text{var}(y) = \text{var}(s^T \tilde{x})$

$= \frac{1}{s^T \tilde{K} s}$

$= \frac{1}{s^T \tilde{K} \tilde{x}}$

$= s^T \tilde{K} s$

$\bar{y} = E[s^T \tilde{x}] = s^T \tilde{\mu}$

Substitute

$g_{\tilde{x}}(s) = \exp\left(s^T \tilde{\mu} + \frac{1}{2} s^T K_{\tilde{x}} s\right)$

Corollary: Joint distribution of a $\tilde{\mu} + \tilde{N}(\bar{y})$ is specified completely by $\tilde{\mu}$ and $K_{\tilde{x}}$

Let $\tilde{w} = [w_1, \ldots, w_m]^T$ consist of iid Gaussian r.v.s with distribution $N(0, 1)$.
Then $f_{\mathbf{w}}(\mathbf{w}) = \frac{1}{(2\pi)^{m/2}} \exp\left[-\frac{\mathbf{w}^T \mathbf{w}}{2}\right]$ is joint pdf.

Suppose $\mathbf{z} = A\mathbf{w} + \mathbf{v}$ then $\mathbf{z} \sim N(\mathbf{u}, \mathbf{A} \mathbf{A}^T)$

Proof: $\mathbf{z}$ is JG because its elements are lin. comb. of elements of $\mathbf{w}$

$E[\mathbf{z}] = E[A\mathbf{w} + \mathbf{v}] = A \cdot E[\mathbf{w}] + E[\mathbf{v}] = \mathbf{u}$

$K_{\mathbf{z}} = E[(A\mathbf{w})(A\mathbf{w})^T]$

$= A \cdot E[\mathbf{w} \mathbf{w}^T] A^T$

($\mathbf{w}$ independent)

$E[w_i w_j] = E[w_i] E[w_j]$ so $A^T = 0$

$= A \cdot I \cdot A^T$

$= \mathbf{A} \mathbf{A}^T$
Theorem
Assume a joint distribution with arbitrary covariance matrix \( \Sigma \) and mean \( \mu_r \). A transform can be constructed by taking
\[
\tilde{z} = A\tilde{w} + \nu
\]
for some \( A \) and \( \nu \).

Proof: Let
\[
K_z = P \Lambda P^T
\]
\( P \) = matrix of orthonormal eigenvectors,
\( \Lambda \) = diagonal matrix of eigenvalues

Let \( A = P \Lambda^{1/2} \)

Covariance \( (Aw + \nu) \) = \( E[Aww^TA^T] \)
\[
= A E[ww^T] A^T
\]
\[
= P \Lambda^{1/2} I \Lambda^{1/2} P^T
\]
\[
= K_z
\]

\[
f_{\tilde{z}}(z) = \left| \det(CA^{-1}) \right| f_w(CA^{-1}(z-\mu)) \Lambda^{-1/2} \Lambda^{-1/2}
\]
\[
= \frac{1}{(2\pi)^{n/2} \det(K_z)^{n/2}} \exp \left[ -\frac{1}{2} C(z-\mu)^T K_z^{-1} (z-\mu) \right]
\]
Last Time:

Theorem: An \( \text{AR}(1) \) with an arbitrary covariance matrix \( K_z \) and mean \( \mu_z \) can be constructed by taking

\[ \hat{z} = A \hat{w} + \nu \]

for some \( A \) and \( \nu \).

---

\[ K_z = P \Lambda K_z P^* \quad P = \text{matrix of orthonormal eigenvectors} \]
\[ \Lambda = \text{diagonal matrix of eigenvalues} \]

Let \( A = P \Lambda^{\frac{1}{2}} \)

\[ \text{Covariance} \left( A \hat{w} + \nu \right) = E \left[ A \hat{w} \hat{w}^T A^T \right] \]
\[ = A E[\hat{w} \hat{w}^T] A^T \]
\[ = P \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} P^T \]
\[ = K_z \]

\[ f_z(z) = \det A \right| f_z \left( A^{-1} (z - \mu_z) \right) \]
\[ = \frac{1}{(2\pi)^{n/2} \det(K_z)^{n/2}} \exp \left[ -\frac{1}{2} (z - \mu_z)^T K_z^{-1} (z - \mu_z) \right] \]

---

Contour's of \( r_{z(W)} \) and \( f_z(z) \) contours of \( f_z(z) \)
$V_1 = \text{eigen vector}_1$, eigenvalue $\lambda_1$  \\
$V_2 = \text{eigen vector}_2$, eigenvalue $\lambda_2$

\[
V_i^T K_2^{-1} V_i = \lambda_i^{-1} V_i^T V_i = \lambda_i \| V_i \|^2
\]

i.e. $\| V_i \| = \sqrt{\lambda_i}$, then $V_i^T K_2^{-1} V_i = 1$

\[
V_1^T K_2^{-1} V_2 = \lambda_2^{-1} \| V_2 \|^2
\]

i.e. $\| V_2 \| = \sqrt{\lambda_2}$, then $V_2^T K_2^{-1} V_2 = 1$

Note: RVs $Z_1$ and $Z_j$ are said to be uncorrelated if

$E[(Z_i - \overline{Z}_i)(Z_j - \overline{Z}_j)] = 0$

Note: independence $\iff$ uncorrelated

---

Corollary to Theorem

If RVs $Z_1, \ldots, Z_n$ are uncorrelated if\th e they are independent.

$\Rightarrow$ Let $\hat{Z} = [Z_1, \ldots, Z_n]^T$

Then $K_Z$ is diagonal:

$\hat{Z} = A \hat{W} + N$

for some diagonal $A$ ($A = K_Z^{1/2}$)

and $\hat{W} = N(0, I)$

$Z_1 = q_1 W_1 + N_1$, $Z_2 = q_2 W_2 + N_2$, \ldots

$W_1, W_2, \ldots$ are independent

thus $Z_1, Z_2, \ldots$ are independent.

$\iff$ True for any independent
For JG:

Uncorrelated $\iff$ independent

For general RV's

independent $\implies$ uncorrelated

uncorrelated $\not\implies$ independence

Suppose $\tilde{X}$ and $\tilde{Y}$ are JG and zero mean
and also suppose $K_Y^{-1}$ exists.

What is $E[\tilde{X}|\tilde{Y}]$?

Approach: Suppose $A\tilde{Y}$ is our estimate of $\tilde{X}$

Find $A$ such that our error $(\tilde{X} - A\tilde{Y})$ is

uncorrelated with $\tilde{Y}$, and therefore independent

of $\tilde{Y}$.

$E[(\tilde{X} - A\tilde{Y})\tilde{Y}^T] = K_{\tilde{X}\tilde{Y}} - A K_Y = 0$

$\implies A = K_{\tilde{X}\tilde{Y}} K_Y^{-1}$

Thus $(\tilde{X} - K_{\tilde{X}\tilde{Y}} K_Y^{-1} \tilde{Y})$ is independent of $\tilde{Y}$

$E[\tilde{X}|\tilde{Y}] = E[(\tilde{X} - A\tilde{Y}) + A\tilde{Y}|\tilde{Y}]$

$= E[(\tilde{X} - A\tilde{Y})|\tilde{Y}] + E[A\tilde{Y}|\tilde{Y}]$

$= 0 + A\tilde{Y}$

$E[\tilde{X}|\tilde{Y}] = A\tilde{Y}$
In general if \( \tilde{x} \) and \( \tilde{y} \) are JO but not zero mean

\[
E[\tilde{x} | \tilde{y}] = K_{xy} K_{\tilde{y}} (\tilde{y} - \tilde{\mu}_y) + \tilde{\mu}_x
\]

(1)

Facts

(Call \( \hat{x}(\tilde{y}) \) our estimate)

1) If \( \tilde{x} \) and \( \tilde{y} \) are JO, setting \( \hat{x}(\tilde{y}) = 0 \) minimizes \( E[|x - \hat{x}(\tilde{y})|^2] \) amongst all possible functions of \( \tilde{y} \). It is a "minimum mean squared error estimator" (MMSE).
   proof: see prob. note

2) If \( \tilde{x} \) and \( \tilde{y} \) have arbitrary distributions, setting \( \hat{x}(\tilde{y}) = 0 \) minimizes \( E[|x - \hat{x}(\tilde{y})|^2] \) amongst all possible linear functions of \( \tilde{y} \). It is a "Linear Least Squared Estimator" (LLSE).
   proof: see prob. notes.
Example
\[ X \sim N(0, \sigma_x^2) \]
\[ Z \sim N(0, \sigma_z^2) \]
\( X, Z \) are independent
\[ Y = X + Z \]

What is \( E[X \mid Y] \)?

\[ K_y = E\left[ (X+Z)(X+Z)^T \right] \]
\[ = E\left[ (X+Z)^2 \right] = E\left[ X^2 + 2XZ + Z^2 \right] \]
\[ = \sigma_x^2 + \sigma_z^2 \]

\[ K_y^{-1} = \frac{1}{\sigma_x^2 + \sigma_z^2} \]

\[ K_{xy} = E[XY] = E[X(X+Z)] = \sigma_x^2 \]

\[ E[X \mid Y] = K_{xy} K_y^{-1} Y \]

\[ = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} Y \]
A random process \( \{X(t)\} \)

1) An infinite collection of RV described on prob. space \( \Omega \)

2) For each \( \omega \), specify a simple path \( \gamma : \mathbb{R} \rightarrow \mathbb{R} \) or \( \cdot \cdot \cdot \)

Example 1

\( \{w(n), n \in \mathbb{Z}\} \ldots \ w(-1), w(0), w(1), \ldots \)

are iid Gaussian RVs.

This is called iid Gaussian Random Process

Example 2

\( \langle X(n), n \in \mathbb{Z}, n \geq 0 \rangle \)

\[
\begin{align*}
W(n) & \quad \rightarrow \quad \oplus \quad \rightarrow \quad D \quad \rightarrow \quad X(n) \\
\downarrow \quad \quad \quad \quad \quad \downarrow_D \\
X(n+1) & = a \cdot X(n) + W(n) \quad n \geq 0 \\
X(0) & = 0
\end{align*}
\]

Definition: Gaussian Process \( \langle X(t) \rangle \) (discrete or cont. time)

\( \forall \ k, t_1, \ldots, t_k \quad (X(t_1), \ldots, X(t_k)) \) is N.G.

Example 2 continued

\[
E[X(n)] = \bar{x}(n) = 0
\]
Gaussian processes are described by

\[ \tilde{x}(t) = E(x(t)) \text{ at } t \text{ mean at every } t \]

\[ K_x(t, u) = E \left[ (x(t) - \tilde{x}(t))(x(u) - \tilde{x}(u)) \right] \text{ } \forall t, u \]

called the covariance function

(discrete or cont. time)

---

**Example 2:**

\[ E[X(n)] = \tilde{x}(n) = 0 \]

\[ K_x(n, m) = E[X(n)X(m)] \]

\[ = E \left[ (m-n)X(n) + \sum_{k=1}^{m-n} \epsilon_k w(n+k-1) \right] \]

\[ = d^{m-n} E[X^2(n)] + 0 \]

\[ = d^{m-n} \left( \sum_{k=1}^{n} \epsilon_k^2 w(n+k-1) \right) \]

\[ = d^{m-n} \sum_{k=1}^{n} \sigma^2 d^{2(k-1)} \]

\[ = d^{m-n} \sigma^2 \frac{(1 - d^2n)}{1 - d^2} \]
\[
X(n+1) = \alpha X(n) + w(n)
\]

we showed
\[
K_X(n, m) = E[X(n)X(m)] = \alpha^{m-n} \sigma^2 \frac{(1 - \alpha^2)}{1 - \alpha^2}
\]

if \(|\alpha| < 1\), and \(n \text{ large}\)

\[
\approx \alpha^{m-n} \frac{\sigma^2}{1 - \alpha^2}
\]

**Definition: Stationarity**

\(\langle X(t) \rangle\) is stationary \(\forall k, \forall t_1, \ldots, t_k\) and for all \(T\), the RVs \((X(t_1), \ldots, X(t_k))\) and \((X(t_1+T), \ldots, X(t_k+T))\)

have the same distribution.

\[
F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k) = F_{X(t_1+T), \ldots, X(t_k+T)}(x_1, \ldots, x_k)
\]

**Suppose a Gaussian Process is Stationary**

Then

\[
E[X(t)] = \bar{X}(0) \quad \forall t
\]

\[
K_X(t_1, t_2) = E[(X(t_1) - \bar{X}(t_1))(X(t_2) - \bar{X}(t_2))]
\]
\[ K_x(t_1, t_2) = E \left[ (X(t_1 + T) - \bar{X}(t_1 + T))(X(t_2 + T) - \bar{X}(t_2 + T)) \right] \]

\[ = K_x(t_1 + T, t_2 + T) \]

Thus \( K_x(t_1, t_2) \) depends only on \( t_1 - t_2 \)

**Fact:** A Gaussian process is stationary if \( \bar{X}(t) = X(0) \) and

\( K_x(t_1, t_2) \) depends only on \( t_2 - t_1 \)

**Proof:** See above

\[ \iff K_x(t_1, t_2) \text{ depends only on } t_2 - t_1, \]

\[ \bar{X} = (X(t_1), \ldots, X(t_n))^T \quad \bar{X}_s = (X(t_1 + T), \ldots, X(t_n + T))^T \]

Assume zero mean for now

\[ K_x = E[\bar{X}\bar{X}^T] = E \begin{bmatrix} X(t_1)^2 - X(t_1)X(t_1) \\ X(t_1)X(t_2) \\ \vdots \\ X(t_n)X(t_1) - X(t_1)X(t_2) \\ X(t_2)^2 - X(t_2)X(t_1) \\ \vdots \\ X(t_n)^2 - X(t_n)X(t_1) \end{bmatrix} \]

\[ K_{\bar{X}} = E[\bar{X}_s\bar{X}_s^T] = E \begin{bmatrix} X(t_1 + T)^2 - X(t_1 + T)X(t_1 + T) \\ X(t_1 + T)X(t_2 + T) - X(t_1 + T)X(t_1 + T) \\ \vdots \\ X(t_n + T)^2 - X(t_n + T)X(t_1 + T) \end{bmatrix} \]

\[ K_x = K_{\bar{X}} \]

Because \( \bar{X} \) and \( \bar{X}_s \) are jointly Gaussian, we conclude

\[ F_{\bar{X}}(x) = F_{\bar{X}_s}(x) \]

Thus \( \bar{X}(t) \) is stationary.
Example 2 is not stationary, but it is asymptotically stationary.

Define: \( \langle X(t) \rangle \) is Wide-Sense Stationary (WSS) if
\[
\tilde{X}(t) = \tilde{X}(0) \quad \forall t
\]

\( K_X(t, t+v) \) only depends on \( v \)

Note for Gaussian Processes:
WSS \( \iff \) Stationarity

But in General
Stationarity \( \not\iff \) WSS

WSS \( \not\iff \) Stationarity

Linear Systems

\( X(t) = \delta(t) \)

Note \( \sum \delta(t) = 1 \)

\( \delta(t) = \begin{cases} \infty & t = 0 \\ 0 & \text{otherwise} \end{cases} \)

\( X(t) = \delta(t) \)

\( h(t) = h(t) \)

:= impulse response

\( h(t) \)

\( -1 \)

\( h(t) + h(t+1) \)
The document contains a detailed explanation of convolution in both continuous and discrete time domains. Here are the key points:

**Continuous Time**

- **Convolution Integral**:
  \[ y(t) = \int_{-\infty}^{\infty} h(t-\tau) x(\tau) \, d\tau \]

- **Linear System (LTI System)**: Sometimes we denote convolution as \( y(t) = h \ast x(t) \).

- **Causal System**: If causal,
  \[ y(t) = \int_{0}^{t} h(t-\tau) x(\tau) \, d\tau \]

**Discrete Time**

- **Convolution Sum**:
  \[ y[n] = \sum_{m=-\infty}^{\infty} h[n-m] x[m] = h \ast x[n] \]

**Cont Time**

- **Stability**: \( \int_{-\infty}^{\infty} |h(t)| \, dt < \infty \)
  (Bounded Input Bounded Output) \( \text{BIBO} \)

**Fourier Transform**

- **Forward Fourier Transform**:
  \[ X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \, dt \]
  \( j = \sqrt{-1} \)

- **Note**: \( e^{j\omega} = \cos \omega + j \sin \omega \)

- **Inverse Fourier Transform**:
  \[ x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} \, d\omega \]
Fact
\[ y(t) = h(x(t)) \]
\[ y(\omega) = H(\omega) x(\omega) \]

Proof
\[
y(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-r) x(r) \, dr \, dt \quad e^{-j\omega t} \, dt
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-r) x(r) \, dr \, dt \quad e^{-j\omega(t-r)} \, e^{-j\omega t} \, dt \, dv
\]
\[
= \int_{-\infty}^{\infty} h(\tau) x(\tau) \, d\tau \quad e^{-j\omega \tau} \, \frac{dv}{dt}
\]
\[
= \left( \int_{-\infty}^{\infty} h(\tau) e^{-j\omega \tau} \, d\tau \right) \left( \int_{-\infty}^{\infty} x(\tau) e^{-j\omega \tau} \, d\tau \right)
\]
\[
y(\omega) = H(\omega) X(\omega)
\]

Example
\[
\begin{array}{c}
\begin{tikzpicture}
\node (in) at (0,0) [input] {$X(t)$};
\node (operation) at (2,0) [operation] {$y(t)$};
\node (out) at (4,0) [output] {$Y(t)$};
\draw [->] (in) -- node [midway, above] {} (operation);
\draw [->] (operation) -- node [midway, above] {} (out);
\end{tikzpicture}
\end{array}
\]
\[
y(t) = -a \dot{y}(t) + x(t)
\]
\[
y(0^-) = 0
\]

Impulse Response
Let \[ x(t) = \delta(t) \]
\[
y(0+) = 1 \quad \dot{y}(t) = -ay(t)
\]
\[
x(t) = e^{-at} 1(t > 0)
\]
thus \[ h(t) = e^{-at} 1(t > 0) \]

\[
\begin{array}{c}
\begin{tikzpicture}
\node (in) at (0,0) [input] {$x(t)$};
\node (operation) at (2,0) [operation] {$\frac{1}{a}$};
\node (out) at (4,0) [output] {$\dot{y}(t)$};
\draw [->] (in) -- node [midway, above] {} (operation);
\draw [->] (operation) -- node [midway, above] {} (out);
\end{tikzpicture}
\end{array}
\]
\[
F(t) = ay(t) + b\dot{y}(t)
\]
\[
\dot{y}(t) = -ay(t) + F(t)
\]
\[ H(\omega) = \sum_{0}^{\infty} e^{-at} e^{-j\omega t} 1(t > 0) \, dt \]

\[ = \int_{0}^{\infty} e^{-at} e^{-j\omega t} \, dt \]

\[ = \left[ \frac{-1}{j\omega + a} e^{-j\omega t} \right]_{0}^{\infty} \quad \text{assume } a > 0 \]

\[ = \frac{1}{j\omega + a} \]

Now suppose \( X(\omega) = e^{-bt} 1(t > 0) \)

What's \( Y(\omega) \)?

\( X(\omega) = \frac{1}{j\omega + b} \)

\( Y(\omega) = \frac{1}{j\omega + b} \cdot \frac{1}{j\omega + a} = \frac{Y_b}{j\omega + a} + \frac{Y_a}{j\omega + b} \)

\text{Inverse Fourier transform}

\[ y(t) = \left[ \frac{1}{b-a} \cdot e^{-at} + \frac{1}{a-b} \cdot e^{-bt} \right] 1(t > 0) \]
Suppose $X(t)$ is wide sense stationary.

\[ Y(t) = \int_{-\infty}^{\infty} X(t-\tau) h(\tau) \, d\tau \]

\[
E[Y(t)] = \overline{Y}(t) = E\left[ \int_{-\infty}^{\infty} X(t-\tau) h(\tau) \, d\tau \right]
\]

\[
= \int_{-\infty}^{\infty} E[X(t-\tau)] h(\tau) \, d\tau \quad \text{(w.s.s.)}
\]

\[
= \int_{-\infty}^{\infty} \overline{X}(\tau) h(\tau) \, d\tau
\]

\[
= \overline{X}(\tau) \int_{-\infty}^{\infty} h(\tau) \, d\tau
\]