Multivariate Gaussian Random Variables

M. Sami Fadali
Professor of Electrical Engineering
University of Nevada, Reno

Outline

- Multivariate Gaussian random variables.
- Jointly Gaussian random vectors.
- Gaussian density function.
- Minimum variance estimator (another derivation).
- Orthogonality principle.
- Measurement residuals (innovations).

Properties of Multivariate Gaussian Random Variables

1. Completely characterized by the first two moments \((m_x, C_x)\).
2. Independent \(\Leftrightarrow\) uncorrelated.
3. Linear transformation of Gaussian random vector (vector of jointly Gaussian random variables) gives a Gaussian random vector.
Properties

• If $x$ and $z$ are jointly Gaussian then they are marginally Gaussian
  $f_{xy}$ Gaussian $\Rightarrow f_x \& f_y$ Gaussian
• If $x$ and $z$ are marginally Gaussian and mutually uncorrelated (independent) then they are jointly Gaussian.
• If $x$ and $z$ are marginally Gaussian but not mutually uncorrelated then they may or may not be jointly Gaussian.

Jointly Gaussian Random Vectors

$x \in \mathbb{R}^n \sim N(m_x, C_x), \quad z \in \mathbb{R}^m \sim N(m_z, C_z)$

$C_{xz} = E\{(x - m_x)(z - m_z)^T\}$
$C_{zx} = E\{(z - m_z)(x - m_x)^T\} = C_{xz}^T$

$\bar{z} = \text{col}\{x, z\}, \quad m_{\bar{z}} = \text{col}\{m_x, m_z\}$

$C_{\bar{z}} = \begin{bmatrix} C_x & C_{xz} \\ C_{zx} & C_z \end{bmatrix}$

$f_{\bar{z}}(\bar{z}) = \frac{1}{[2\pi]^{(n+m)/2}\sqrt{\det(C_{\bar{z}})}} e^{-\frac{1}{2}(\bar{z} - m_{\bar{z}})^T C_{\bar{z}}^{-1}(\bar{z} - m_{\bar{z}})}$

Covariance $C_{\bar{z}}$ Properties

$C_{\bar{z}}^{-1} = \begin{bmatrix} C_x & C_{xz} \\ C_{zx} & C_z \end{bmatrix}^{-1} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$

• Use inverse of partitioned matrix and the matrix inversion lemma

$A = (C_x - C_{xz}C_{zz}^{-1}C_{zx})^{-1}$
$B = -AC_{xz}C_{zz}^{-1} = -C_{zz}^{-1}C_{xz}C$
$C = (C_z - C_{zx}C_{xx}^{-1}C_{xz})^{-1}$
$= C_z^{-1} + C_{zz}^{-1}C_{zx}AC_{xz}C_{zz}^{-1}$

Conditional Density Function

Theorem 1: For $x, z$ jointly Gaussian

$f_{x|z}(x|z) = f_{xz}(x, z)/f_z(z)$

$f_{x|z}(x|z)$

$= \frac{1}{[2\pi]^{n/2}\sqrt{\det(C_{x|z})}} e^{-\frac{1}{2}(x - m_{x|z})^T C_{x|z}^{-1}(x - m_{x|z})}$

$m_{x|z} = E\{x|z\} = m_x + C_{xz}C_{zz}^{-1}(z - m_z)$

$C_{x|z} = C_x - C_{xz}C_z^{-1}C_{zx}$
Proof
\[
\frac{f_{xz}(x, z)}{f_z(z)} = \frac{f_z(z)}{f_z(z)}, \quad \tilde{z} = \text{col}\{x, z\}
\]
\[
\frac{f_{xz}(x, z)}{f_z(z)} = \frac{1}{[2\pi]^{(n+m)/2}\sqrt{\det(C_z)}} e^{\frac{-1}{2}(z-m_z)^T C_z^{-1}(z-m_z)}
\]
\[
= \frac{1}{[2\pi]^{m/2}\sqrt{\det(C_z)}} e^{\frac{-1}{2}(z-m_z)^T C_z^{-1}(z-m_z)}
\]
\[
= \frac{1}{[2\pi]^{n/2}\sqrt{\det(C_z)}} e^{\frac{-1}{2}(z-m_z)^T C_z^{-1}(z-m_z)}
\]
\[
\times \exp\left\{\frac{1}{2} (\bar{z} - m_z)^T \begin{bmatrix} A & B \\ B^T & C - C_z^{-1} \end{bmatrix} (\bar{z} - m_z) \right\}
\]

Proof: Expand Quadratic
\[
(\bar{z} - m_z)^T \begin{bmatrix} A & B \\ B^T & C - C_z^{-1} \end{bmatrix} (\bar{z} - m_z)
\]
\[
= (x - m_x)^T A (x - m_x) + 2(x - m_x)^T B (z - m_z) + (z - m_z)^T (C - C_z^{-1}) (z - m_z)
\]
\[
A = C_x^{-1} + C_x^{-1} C_{xz} C_{zz} C_x^{-1}
\]
\[
B = -AC_{xz} C_z^{-1}, \quad C = C_z^{-1} + C_z^{-1} C_{xz} A C_{xz} C_z^{-1}
\]
\[\text{Substitute for A, B, C, then messy algebra}
\]
\[
(x - m_x - C_{xz} C_z^{-1} (z - m_z))^T A (x - m_x - C_{xz} C_z^{-1} (z - m_z))
\]
\[
= (x - m_x|z)^T C_{x|z}^{-1} (x - m_x|z)
\]

Properties of Conditional Mean
• If \( z \) is a random vector ⇒ \( E\{x|z\} \) also random.

Theorem 2: If \( x \) and \( z \) are jointly Gaussian, then \( E\{x|z\} \) is a Gaussian affine transformation of \( z \).

Proof: Follows from the expression (affine)
\[
m_{x|z} = E\{x|z\} = m_x + C_{xz} C_z^{-1} (z - m_z)
\]
\[x \text{ and } z \text{ jointly Gaussian} \Rightarrow z \text{ Gaussian}
\]
\[\Rightarrow m_{x|z} \text{ Gaussian}
\]
State Estimation

- Unbiased estimator: $m_x = E\{x_k|\hat{x}_k\} = \hat{x}_k$
- Measurements: $z_k = H_k x_k + \nu_k$

$$m_z = E\{z_k|\hat{x}_k\} = H_k \hat{x}_k$$
$$\tilde{z}_k = z_k - H_k \hat{x}_k = H_k e_k^- + \nu_k$$
$$C_z = E\{\tilde{z}_k \tilde{z}_k^T|\hat{x}_k\} = H_k P_k^- H_k^T + R_k$$
$$e_k^- = \hat{x}_k - x_k$$
$$C_x = E\{e_k^- e_k^-^T|\hat{x}_k\} = P_k^-$$

$C_{xz} = E\{\tilde{x}_k \tilde{x}_k^T|\hat{x}_k\} = P_k^- H_k^T$,  
$C_{zx} = H_k P_k^-$

Kalman Filter

$$m_{x|z} = m_x + C_{xz} C_z^{-1} (z - m_z)$$

$$P_{k+}^+ = C_{x|z} = C_x - C_{xz} C_z^{-1} C_{zx}$$

- Substitute $m_x = E\{x_k|\hat{x}_k\} = \hat{x}_k$, $C_x = P_k^-$

$$C_{xz} = P_k^- H_k^T, C_z = H_k P_k^- H_k^T + R_k, \quad m_z = H_k \hat{x}_k^-$$

$$\tilde{x}_k^+ = \hat{x}_k^- + P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} (z_k - H_k \hat{x}_k^-)$$

$$P_k^+ = P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^-$$

- Kalman gain

$$K_k = C_{xz} C_z^{-1} = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}$$

Comments

- Kalman filter equation assuming Gaussian noise using the fundamental theorem of estimation theory.
- Linear estimate for Gaussian case without prior assumption.
- For non-Gaussian case the Kalman filter is the minimum variance linear filter but there may be better nonlinear filters.
- If knowledge of the process is incomplete (only up to 2nd order statistics), the Kalman filter is the MMSE estimator but not the minimum variance estimator.

Orthogonality

- $x$ and $z$ are jointly orthogonal if

$$E\{xz^T\} = [0]$$

- Orthogonality principle

For a MMSE estimator, the estimation error vector is orthogonal to the hyperplane of the measurements.

- All measurements used

$$z = \{z_i, i = 0,1, \ldots \}$$
Theorem: Orthogonality

1. If $x_k$ and $z$ are jointly Gaussian, then the estimation error $\hat{x}_{k|k} = x_k - \hat{x}_{k|k}$ is orthogonal to $z$.

\[
E\{\hat{x}_{k|k} z^T\} = [0]
\]
\[
\hat{x}_{k|k} = K^o z + b^o
\]
\[
K^o = C_{xz}C_z^{-1}, b^o = m_x - C_{xz}C_z^{-1}m_z
\]

2. If $(K, b)$ satisfy $E\{(x_k - Kz - b)z^T\} = [0]$ and $E\{x_k - Kz - b\} = 0$ (unbiased), then $(K, b) = (K^o, b^o)$

Proof of (1)

• The optimum linear estimate is

\[
\hat{x}_{k|k} = m_{x|z} = m_x + C_{xz}C_z^{-1}(z - m_z)
\]

• To show orthogonality (drop subscript)

\[
E\{(x - \hat{x}_{k|k})z^T\} = E\{(x - m_x - C_{xz}c^{-1}(z - m_z))z^T\}
\]
\[
= E\{xz^T\} - m_x E\{z^T\}
- C_{xz}C_z^{-1}E\{(z - m_z)(z - m_z)^T\}
- C_{xz}C_z^{-1}E\{z - m_z\}m_z^T
= C_{xz} - C_{xz}C_z^{-1}C_z - [0] = [0]
\]

Proof of (2)

\[
E\{(x - Kz - b)z^T\} = [0], \quad E\{x - Kz - b\} = 0
\]

• The optimum linear estimate satisfies

\[
E\{(x - K^o z - b^o)z^T\} = [0], \quad E\{x - K^o z - b^o\} = 0
\]

• Subtract

\[
(K - K^o)E\{zz^T\} + (b - b^o)E\{z^T\}
\]
\[
= (K - K^o)(C_z + m_z m_z^T) + (b - b^o)m_z^T = [0]
\]
\[
[K - K^o] \begin{bmatrix} C_z & m_z m_z^T \end{bmatrix} = [0]
\]
\[
K - K^o = [0] \Rightarrow b - b^o = 0
\]

Example

• First-order Gauss Markov $x$

\[
R_{xx}(\tau) = \sigma_x^2 e^{-\beta|\tau|}
\]

• Measurement

\[
z_k = x_k + v_k, \quad R = E\{v_k^2\} = \sigma_v^2
\]

• Initialization

\[
\hat{x}_0^- = 0, \quad P_0^- = \sigma_x^2
\]

• Initial estimation error

\[
e_0^- = x_0 - \hat{x}_0^- = x_0
\]
Gain

- Gain $H_k = 1, R = \sigma_v^2$

\[ K_k = \frac{P_k^{-}H_k^T}{H_kP_k^{-} H_k^T + R} = \frac{P_k^{-}}{P_k^{-} + \sigma_v^2} \]

- Initial Gain

\[ K_0 = \frac{P_0^{-}}{P_0^{-} + \sigma_v^2} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2} \]

\[ 1 - K_0 = \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2} \]

Example (a posteriori error)

- Update estimate $\hat{x}_0^- = 0$

\[ \hat{x}_0^+ = \hat{x}_0^- + K_0(z_0 - \hat{x}_0^-) = K_0z_0 \]

\[ e_0^+ = x_0 - \hat{x}_0^+ = x_0 - K_0z_0 \]

\[ = x_0 - K_0(x_0 + \nu_0) \]

\[ e_0^+ = (1 - K_0)x_0 - K_0\nu_0 \]

\[ E\{e_0^+z_0\} = (1 - K_0)E\{x_0^2\} - K_0E\{\nu_0^2\} \]

\[ = \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2} \cdot \sigma_x^2 - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2} \cdot \sigma_v^2 = 0 \]

Example (a priori error)

- Predict using the estimate $\hat{x}_0^+ = K_0z_0$

\[ \hat{x}_1^- = \phi \hat{x}_0^+ \]

\[ e_1^- = x_1 - \hat{x}_1^- = \phi x_0 + w_0 - \phi \hat{x}_0^+ \]

\[ e_1^- = \phi e_0^+ + w_0 \]

- Orthogonality

\[ E\{e_1^-z_0\} = \phi E\{e_0^+z_0\} + E\{w_0z_0\} \]

\[ = 0 + 0 = 0 \]

- Similarly show (by induction)

\[ E\{e_{k+1}^-z_k\} = 0, E\{e_k^+z_k\} = 0, k = 0,1,2, \ldots \]

Measurement Residuals (Innovations)

Residuals: in statistics $z - \hat{z}$

Residuals: one-step ahead prediction errors

\[ \tilde{z}_k = z_k - E\{z_k| z_{k-1}^*, z_{k-1}^* = \{z_i, i \leq k-1\} \]

\[ = z_k - H_k\hat{x}_k = H_k e_k + \nu_k \]

Innovations sequence is

(i) zero-mean (unbiased estimate)

\[ E\{\tilde{z}_k\} = H_k E\{e_k\} + E\{\nu_k\} = 0 \]

(ii) Gaussian, linear combination of Gaussian

(iii) White: proof requires the Law of Iterated Expectations
Law of Iterated Expectations

\[ E\{E\{x|y}\} = E\{x\} \] (Outer expectation over y)

**Proof**

\[
E\{E\{x|y\}\} = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} xf_{x|y}(x|y)dx \right] f_y(y)dy \\
= \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f_{xy}(x,y)dy \right] dx \\
= \int_{-\infty}^{\infty} xf_x(x)dx = E\{x\}
\]

Proof: innovations white

Consider \( j \leq k - 1 \) with measurements \( z^*_k = \{z_i, i \leq k - 1\} \)

Smoothing Property

\[
E\left\{\hat{z}_k \hat{z}_j^T\right\} = E\left\{E\left\{\hat{z}_k \hat{z}_j^T | z^*_k\right\}\right\}
\]

Given \( z^*_k \) the term \( \hat{z}_j^T \) can be moved

\[
E\left\{\hat{z}_k \hat{z}_j^T\right\} = E\left\{E\{\hat{z}_k | z^*_k\} \hat{z}_j^T\right\}
\]

\[
E\{\hat{z}_k | z^*_k\} = E\{H_k e_k^- + v_k | z^*_k\} = 0 \\
E\left\{\hat{z}_k \hat{z}_j^T\right\} = [0]
\]

References