# Gaussian Random Vectors

For vectors $X, Y$

$$\mathrm{Cov}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \mathbb{E}\left(\begin{bmatrix} X-\mu_x \\ Y-\mu_Y \end{bmatrix}\begin{bmatrix} X-\mu_x \\ Y-\mu_Y \end{bmatrix}^T\right) = \begin{bmatrix} \mathbb{E}[(x-\mu_x)(x-\mu_x)^T] & \mathbb{E}[(x-\mu_x)(y-\mu_y)^T] \\ \mathbb{E}[(y-\mu_y)(x-\mu_x)^T] & \mathbb{E}[(y-\mu_y)(y-\mu_y)^T] \end{bmatrix}$$

One of the most amazing facts about Gaussians:

If $X, Y$ are jointly Gaussian, then we can express:

$$X = \mu_x + \Sigma_{XY}\Sigma_Y^{-1}(Y-\mu_Y) + \underbrace{V}_{\text{independent Gaussian noise}}$$

where $V \sim \mathcal{N}(0, \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX})$ is independent of $Y$

Corollary: If $X, Y$ are jointly Gaussian, then:

$$\mathbb{E}[X \mid Y] = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(Y-\mu_Y)$$
$$= \mathbb{L}[X \mid Y]$$

the best estimator is also the best linear estimator

Pf: Let $\tilde{X} := \mu_x + \Sigma_{xy}\Sigma_y^{-1}(Y-\mu_y) + V$

Claim: $(\tilde{X}, Y)$ are jointly Gaussian

Since $Y, V$ are independent Gaussians, can write:

$Y = \mu_Y + AW_1$, where $W_1 \sim \mathcal{N}(0, I)$ for some matrix $A$

$V = \phantom{x} + BW_2$, where $W_2 \sim \mathcal{N}(0, I)$ for some matrix $B$. $W_1 \perp\!\!\!\perp W_2$

$$\begin{bmatrix} \tilde{X} \\ Y \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} + \begin{bmatrix} \Sigma_{XY}\Sigma_Y^{-1}A & B \\ A & 0 \end{bmatrix}\begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$

$=$ affine transformation of iid standard normals

Goal: Show $(\tilde{X}, Y)$ has correct distribution.

ie, show that $\mathrm{Cov}\left(\begin{bmatrix} \tilde{X} \\ Y \end{bmatrix}\right) = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}$ $\quad \mathbb{E}\left[\begin{matrix} X \\ Y \end{matrix}\right] = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$

$$\mathbb{E}[(\tilde{X}-\mu_x)(Y-\mu_y)^T] = \mathbb{E}[(\Sigma_{xy}\Sigma_y^{-1}(Y-\mu_y)+V)(Y-\mu_y)^T]$$
$$= \Sigma_{xy}\Sigma_y^{-1}$$
$$= \Sigma_{xy}$$

$$\mathbb{E}[(\tilde{X}-\mu_x)(\tilde{X}-\mu_x)^T] = \mathbb{E}[(\Sigma_{xy}\Sigma_y^{-1}(Y-\mu_y)+V)((Y-\mu_y)^T\Sigma_y^{-1}\Sigma_{yx}+V^T)]$$
$$= \Sigma_{xy}\Sigma_y^{-1}\Sigma_y\Sigma_y^{-1}\Sigma_{yx}+\Sigma_v$$
$$= \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}+\Sigma_v$$
$$= \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}-(\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}-\Sigma_x)$$
$$= \Sigma_x$$

Connection to linear regression:

Consider an overdetermined least squares problem:

$$x_{LS} = \underset{x}{\operatorname{argmin}} \|Ax-y\|^2$$ A has full column rank

$$x_{LS} = (A^TA)^{-1}A^Ty$$

Back to probability: Consider $Y = AX + Z$ $\quad X \sim \mathcal{N}(0, \sigma^2 I)$, $Z \sim \mathcal{N}(0, I)$

$$\mathbb{L}[X|Y] = \Sigma_{xy}\Sigma_y^{-1}Y$$
$$\Sigma_{xy} = \mathbb{E}[XY^T] = \mathbb{E}[X(X^TA^T+Z^T)] = \Sigma_x A^T = \sigma^2 A^T$$
$$\Sigma_y = \mathbb{E}[YY^T] = \mathbb{E}[(AX+Z)(X^TA^T+Z^T)]$$
$$= \sigma^2 AA^T + I$$

$$\Rightarrow \mathbb{L}[X|Y] = A^T(AA^T+\tfrac{1}{\sigma^2}I)^{-1}Y \xrightarrow[\sigma^2\to\infty]{} (A^TA)^{-1}A^TY$$

Fact: $A^+ = \lim_{\varepsilon\downarrow 0} A^T(AA^T+\varepsilon I)$

$$= (A^TA)^{-1}A^T$$

when A full column rank

**Kalman Filter**: practical algorithm for doing online prediction filtering/smoothing of a stochastic process that evolves based on a state space model, given noisy observations

## State space model

Let $X_0, V_0, V_1, V_2, \dots, W_0, W_1, W_2 \dots$ be uncorrelated zero-mean (WLOG) random vectors

State space model is evolution of the form:

~~$X_{n+1} = A_n X_n + V_n$ $n \geq 0$ $(A_n)_{n \geq 0}$ known seq. of matrices~~

$X_{n+1} = A X_n + V_n$ $n \geq 0$ $A$ known matrix.

Observations:

$Y_n = C X_n + W_n$, $n \geq 1$ $C$ known matrix

Assume we know covariances: $\Sigma_x = \mathrm{Cov}(X_0)$, $\Sigma_v = \mathrm{Cov}(V_n)$, $\Sigma_w = \mathrm{Cov}(W_n)$

Example: Let $X_n$ = position at time instant $n$.

Physics: $x_{t+dt} = x_t + dt\, x_t'$

Model: $x_{n+1} = \underbrace{x_n}_{\text{old posn}} + \underbrace{\Delta(x_n - x_{n-1})}_{\text{velocity}} + \underbrace{z_n}_{\text{indep. noise}}$

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 1+\Delta & -\Delta \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix} + \begin{bmatrix} z_n \\ 0 \end{bmatrix}$$

$$Y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix} + W_n$$

Many variations of Kalman filters possible

1. Prediction Problem: Estimate $X_{n+k}$ from $Y_1, \dots, Y_n$
2. Filtering: Estimate $X_n$ from $Y_1, \dots, Y_n$
3. Smoothing: Estimate $X_{n-k}$ from $Y_1, \dots, Y_n$

Kalman Filter Statement :

$$\hat{X}_{n|m} = \mathbb{L}[X_n \mid Y_1 \cdots Y_m]$$

$$\Sigma_{n|m} = \mathrm{Cov}(X_n - \hat{X}_{n|m})$$

$$\Sigma_v = \mathrm{Cov}(V_i) \qquad \Sigma_w = \mathrm{Cov}(W_i) \qquad \Sigma_x = \mathrm{Cov}(X_0)$$

Algorithm :

① Initialize :

$$\hat{X}_{0|0} = 0$$

$$\Sigma_{0|0} = \Sigma_x$$

② For $n \geq 1$ do:

update the estimate $\left\{ \hat{X}_{n|n} = A\hat{X}_{n-1|n-1} + \underbrace{K_n}_{\text{Kalman gain}} (Y_n - CA\hat{X}_{n-1|n-1}) \right.$

$$K_n = \Sigma_{n|n-1} C^T (C\Sigma_{n|n-1} C^T + \Sigma_w)^{-1}$$

$$\Sigma_{n|n-1} = A\Sigma_{n-1|n-1} A^T + \Sigma_v$$

new error $\left\{ \Sigma_{n|n} = (I - K_n C)\Sigma_{n|n-1} \right.$