Gaussians

Pieter Abbeel
UC Berkeley EECS

Many slides adapted from Thrun, Burgard and Fox, Probabilistic Robotics

Outline

- Univariate Gaussian
- Multivariate Gaussian
- Law of Total Probability
- Conditioning (Bayes' rule)

Disclaimer: lots of linear algebra in next few lectures. See course homepage for pointers for brushing up your linear algebra.

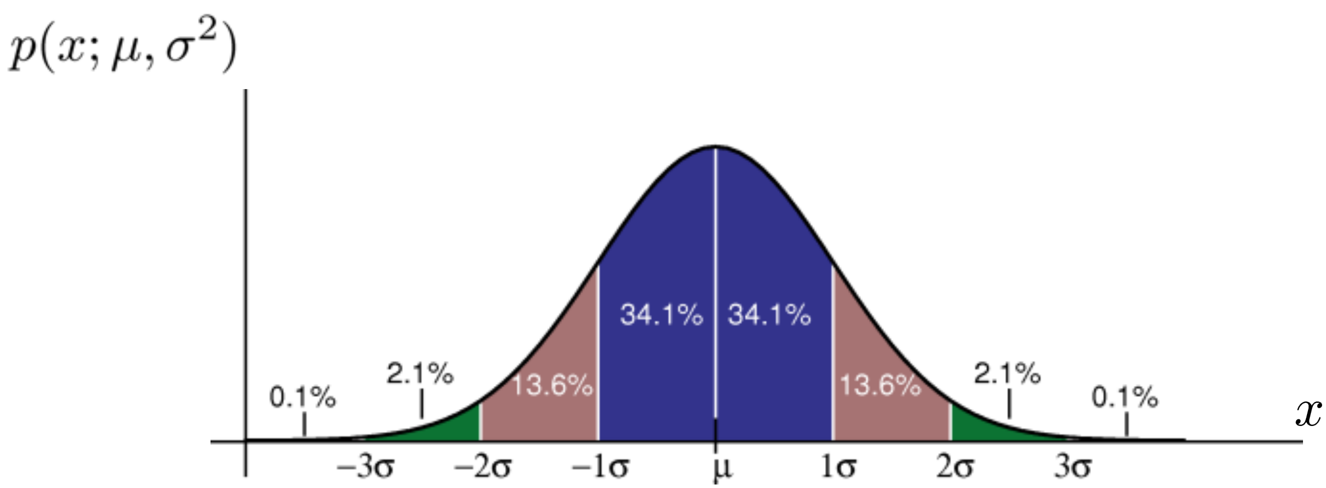
In fact, pretty much all computations with Gaussians will be reduced to linear algebra!

Univariate Gaussian

- Gaussian distribution with mean μ, and standard deviation σ:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$p(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$



Properties of Gaussians

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$p(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

- Densities integrate to one:

$$\int_{-\infty}^{\infty} p(x; \mu, \sigma^2) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx = 1$$

- Mean:

$$\begin{aligned} \mathrm{E}_X[X] &= \int_{-\infty}^{\infty} x p(x; \mu, \sigma^2) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx \\ &= \mu \end{aligned}$$

- Variance:

$$\begin{aligned} \mathrm{E}_X[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x; \mu, \sigma^2) dx \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx \\ &= \sigma^2 \end{aligned}$$

Central limit theorem (CLT)

- Classical CLT:
  - Let $X_1, X_2, \ldots$ be an infinite sequence of *independent* random variables with $E\, X_i = \mu$, $E(X_i - \mu)^2 = \sigma^2$
  - Define $Z_n = ((X_1 + \ldots + X_n) - n\,\mu) / (\sigma\, n^{1/2})$
  - Then for the limit of n going to infinity we have that Z_n is distributed according to $N(0,1)$
- Crude statement: things that are the result of the addition of lots of small effects tend to become Gaussian.

Multi-variate Gaussians

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

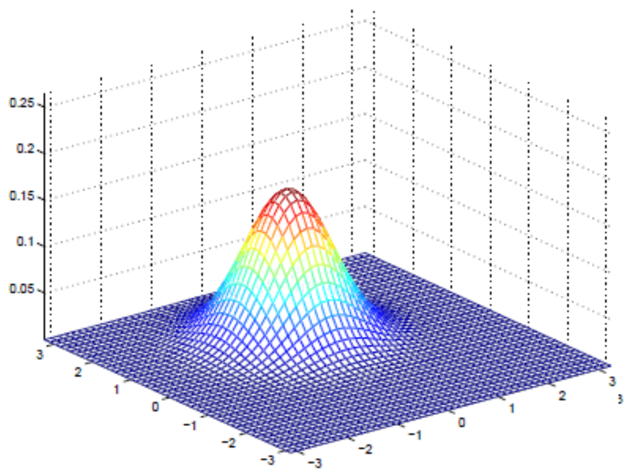
$$\int \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) dx = 1$$

For a matrix $A \in \mathbb{R}^{n\times n}$, $|A|$ denotes the determinant of A.

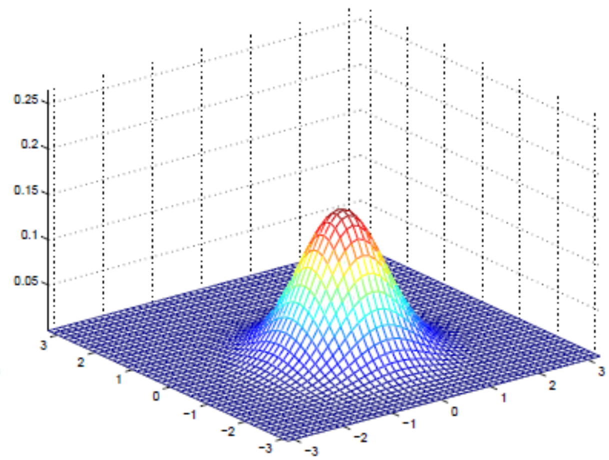
For a matrix $A \in \mathbb{R}^{n\times n}$, A^{-1} denotes the inverse of A, which satisfies $A^{-1}A = I = AA^{-1}$ with $I \in \mathbb{R}^{n\times n}$ the identity matrix with all diagonal entries equal to one, and all off-diagonal entries equal to zero.

Hint: often when trying to understand matrix equations, it's easier to first consider the special case of the dimensions of the matrices being one-by-one. Once parsing them that way makes sense, a good second step can be to parse them assuming all matrices are diagonal matrices. Once parsing them that way makes sense, usually it is only a small step to understand the general case.

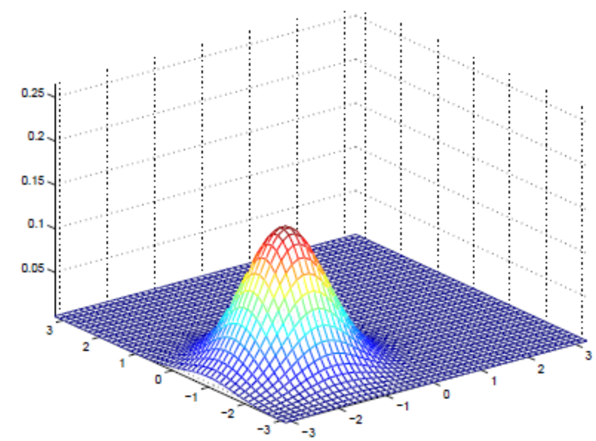
Multi-variate Gaussians: examples



- μ = [1; 0]
- Σ = [1 0; 0 1]

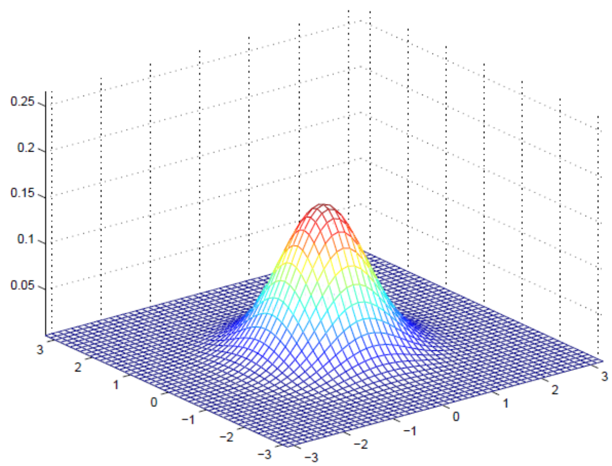


- μ = [-.5; 0]
- Σ = [1 0; 0 1]

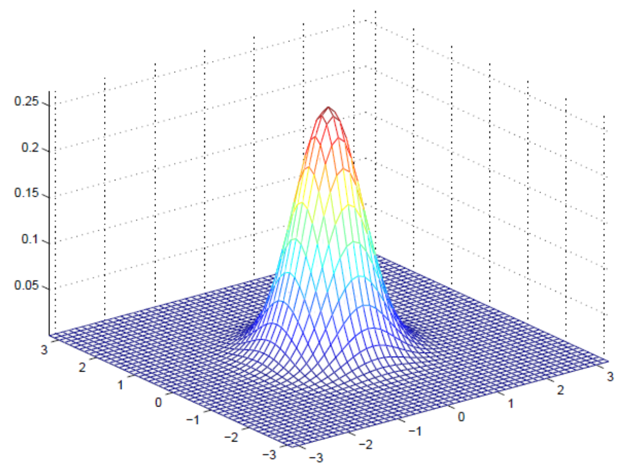


- μ = [-1; -1.5]
- Σ = [1 0; 0 1]

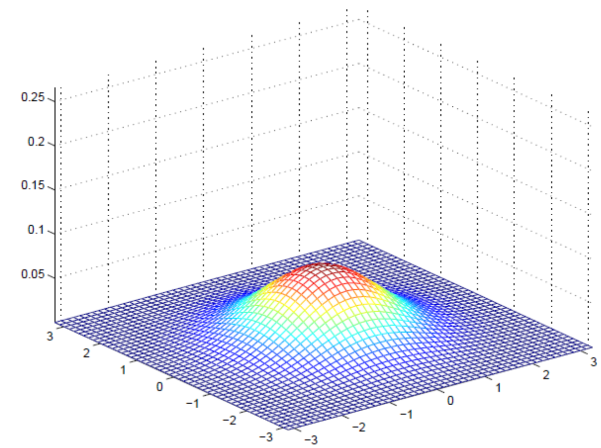
Multi-variate Gaussians: examples



- $\mu = [0; 0]$
- $\Sigma = [1\ 0\ ;\ 0\ 1]$

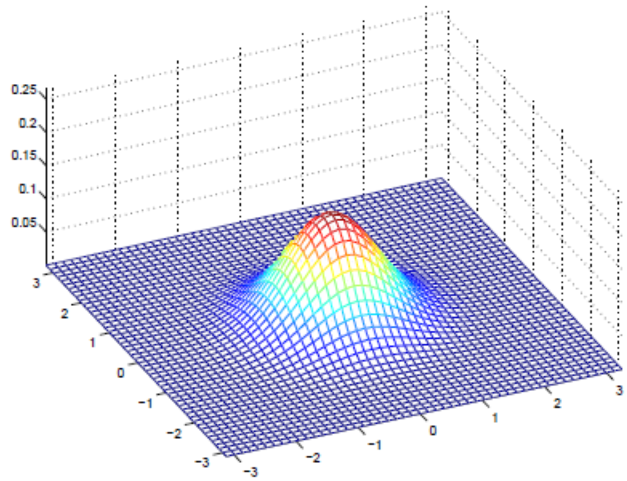


- $\mu = [0; 0]$
- $\Sigma = [.6\ 0\ ;\ 0\ .6]$

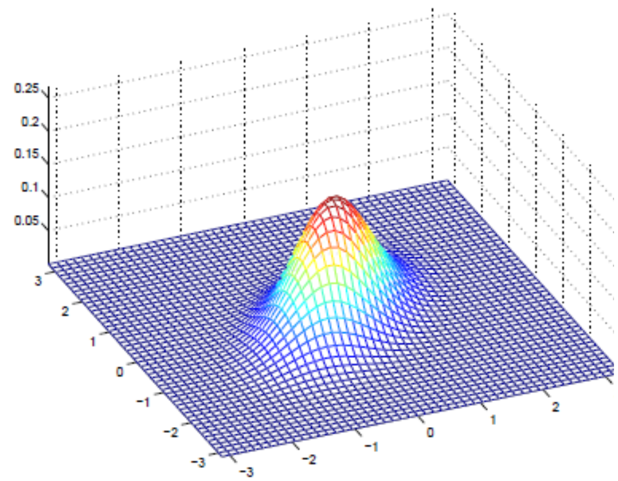


- $\mu = [0; 0]$
- $\Sigma = [2\ 0\ ;\ 0\ 2]$

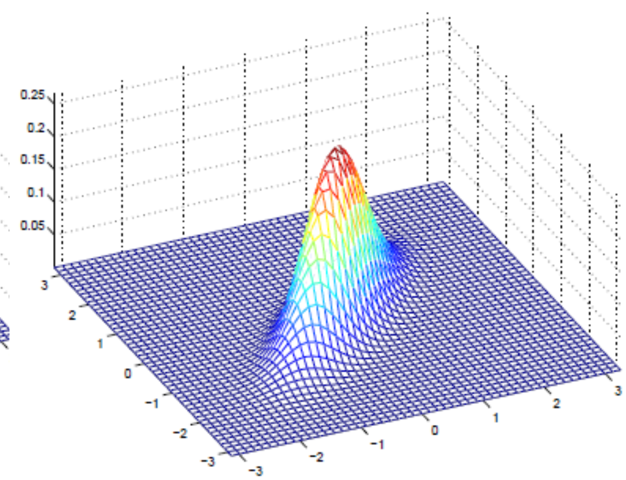
Multi-variate Gaussians: examples



- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0; 0 \;\; 1]$

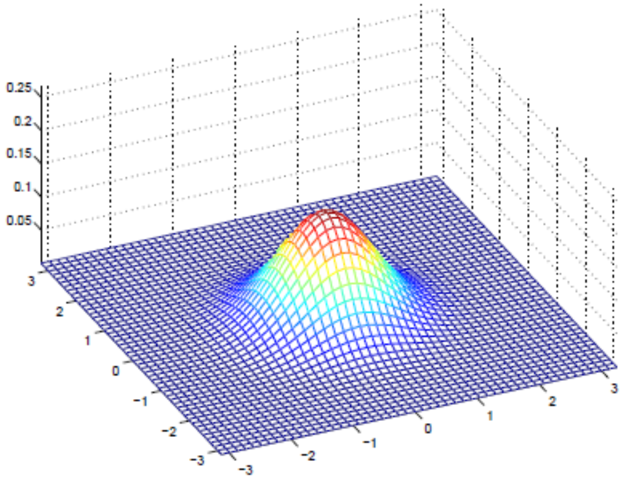


- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.5; 0.5 \; 1]$

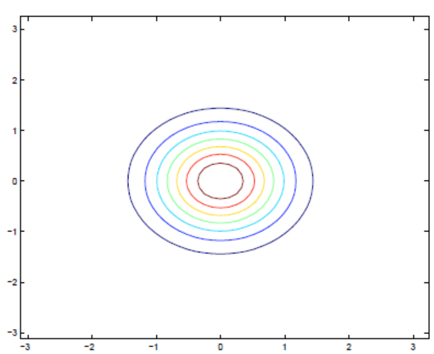


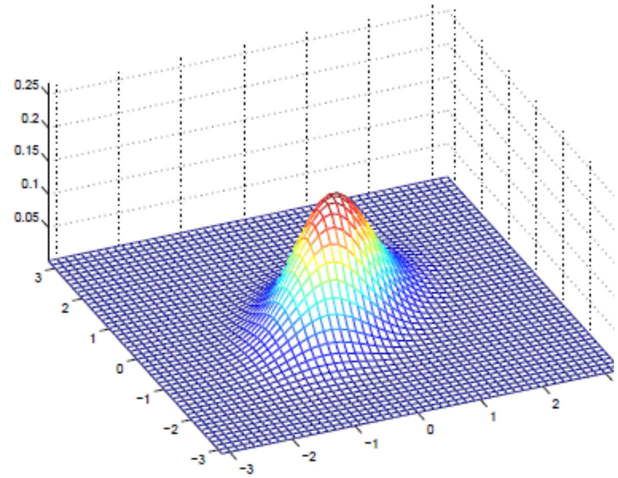
- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.8; 0.8 \;\; 1]$

Multi-variate Gaussians: examples

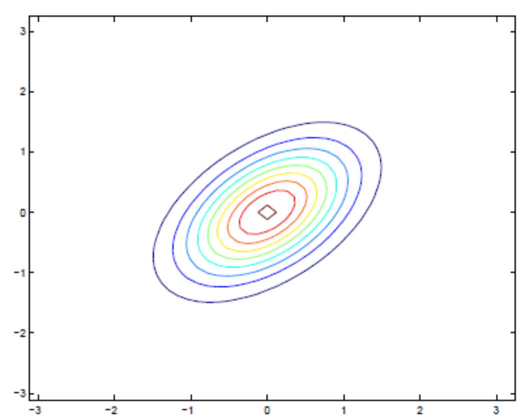


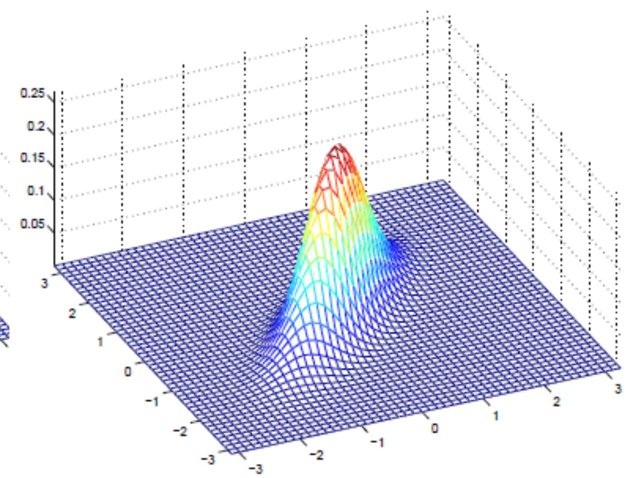
- μ = [0; 0]
- Σ = [1 0; 0 1]



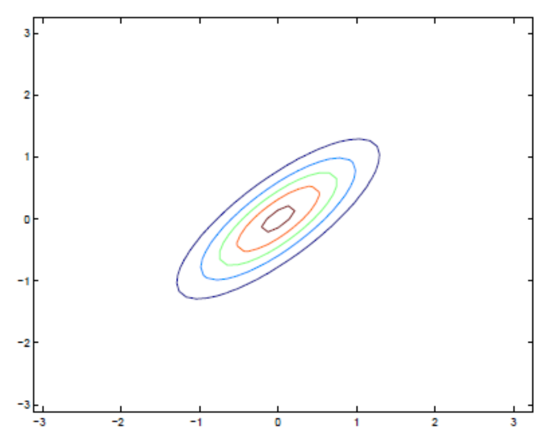


- μ = [0; 0]
- Σ = [1 0.5; 0.5 1]

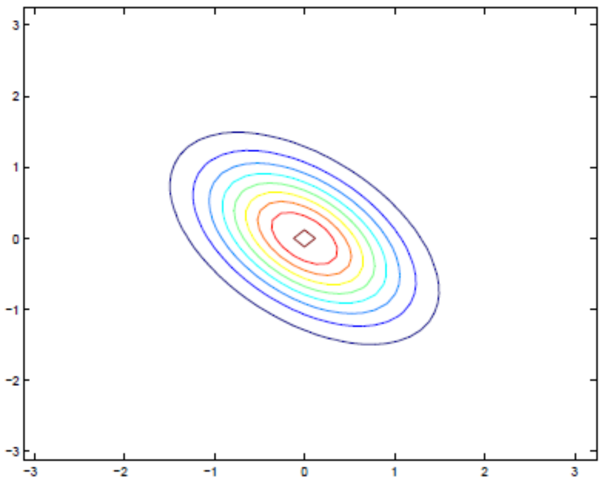




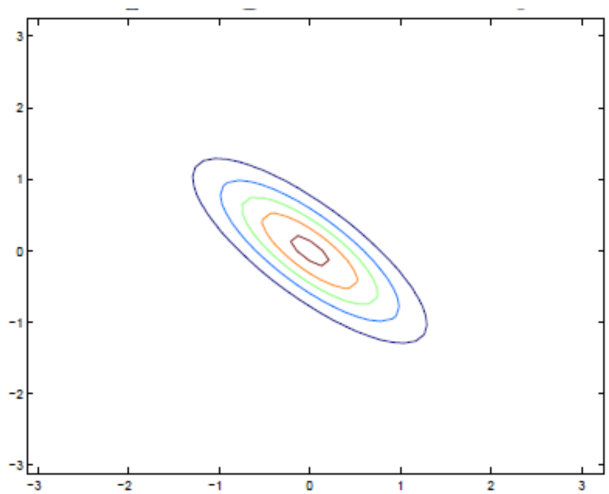
- μ = [0; 0]
- Σ = [1 0.8; 0.8 1]



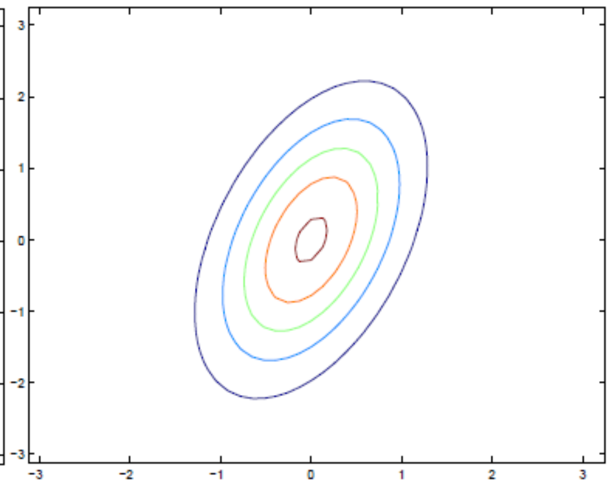
Multi-variate Gaussians: examples



- $\mu = [0; 0]$
- $\Sigma = [1\ \ -0.5\ ;\ -0.5\ \ 1]$



- $\mu = [0; 0]$
- $\Sigma = [1\ \ -0.8\ ;\ -0.8\ \ 1]$



- $\mu = [0; 0]$
- $\Sigma = [3\ \ 0.8\ ;\ 0.8\ \ 1]$

Partitioned Multivariate Gaussian

- Consider a multi-variate Gaussian and partition random vector into (X,Y).

$$\mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

$$p(\begin{bmatrix}x \\ y\end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{(n/2)}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)^\top \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}^{-1} \left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)\right)$$

$$\begin{aligned}
\mu_X &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[X] \\
\mu_Y &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[Y] \\
\Sigma_{XX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(X-\mu_X)^\top] \\
\Sigma_{YY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(Y-\mu_Y)^\top] \\
\Sigma_{XY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(Y-\mu_Y)^\top] = \Sigma_{YX}^\top \\
\Sigma_{YX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(X-\mu_X)^\top] = \Sigma_{XY}^\top
\end{aligned}$$

Partitioned Multivariate Gaussian: Dual Representation

- Precision matrix $\Gamma = \Sigma^{-1} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1} = \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix}$ (1)

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^(n/2)|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right) \right)$$

- Straightforward to verify from (1) that:

$$\begin{aligned}
\Sigma_{XX} &= \left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} \\
\Sigma_{YY} &= \left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} \\
\Sigma_{XY} &= -\Gamma_{XX}^{-1}\Gamma_{XY}\left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} = \Sigma_{YX}^\top \\
\Sigma_{YX} &= -\Gamma_{YY}^{-1}\Gamma_{YX}\left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} = \Sigma_{XY}^\top
\end{aligned}$$

- And swapping the roles of Γ and Σ:

$$\begin{aligned}
\Gamma_{XX} &= \left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} \\
\Gamma_{YY} &= \left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} \\
\Gamma_{XY} &= -\Sigma_{XX}^{-1}\Sigma_{XY}\left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} = \Gamma_{YX}^\top \\
\Gamma_{YX} &= -\Sigma_{YY}^{-1}\Sigma_{YX}\left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} = \Gamma_{XY}^\top
\end{aligned}$$

Marginalization Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X &\sim \mathcal{N}(\mu_X, \Sigma_{XX}) \\ Y &\sim \mathcal{N}(\mu_Y, \Sigma_{YY}) \end{aligned}$$

Conditioning Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X|Y = y_0 &\sim \mathcal{N}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y_0 - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}) \\ Y|X = x_0 &\sim \mathcal{N}(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x_0 - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}) \end{aligned}$$

Optimal estimation in linear-Gaussian systems

Consider the partially-observed system

$$\begin{aligned} \mathbf{x}_{k+1} &= A\mathbf{x}_k + C\boldsymbol{\omega}_k \\ \mathbf{y}_k &= H\mathbf{x}_k + D\boldsymbol{\varepsilon}_k \end{aligned}$$

with hidden state $\mathbf{x}_k$, measurement $\mathbf{y}_k$, and noise $\boldsymbol{\varepsilon}_k, \boldsymbol{\omega}_k \sim N(0, I)$.

Given a Gaussian prior $\mathbf{x}_0 \sim N(\widehat{\mathbf{x}}_0, \Sigma_0)$ and a sequence of measurements $\mathbf{y}_0, \mathbf{y}_1, \cdots \mathbf{y}_k$, we want to compute the posterior $p_{k+1}(\mathbf{x}_{k+1})$.

We can show by induction that the posterior is Gaussian at all times. Let $p_k(\mathbf{x}_k)$ be $N(\widehat{\mathbf{x}}_k, \Sigma_k)$. This will act as a prior for estimating $\mathbf{x}_{k+1}$. Now $\mathbf{x}_{k+1}$ and $\mathbf{y}_k$ are jointly Gaussian, with mean and covariance

$$\begin{aligned} E\begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{y}_k \end{bmatrix} &= \begin{bmatrix} A\widehat{\mathbf{x}}_k \\ H\widehat{\mathbf{x}}_k \end{bmatrix} \\ Cov\begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{y}_k \end{bmatrix} &= \begin{bmatrix} CC^\mathsf{T} + A\Sigma_k A^\mathsf{T} & A\Sigma_k H^\mathsf{T} \\ H\Sigma_k A^\mathsf{T} & DD^\mathsf{T} + H\Sigma_k H^\mathsf{T} \end{bmatrix} \end{aligned}$$

Kalman filter

Lemma

If $\mathbf{u}, \mathbf{v}$ *are jointly Gaussian with means* $\widehat{\mathbf{u}}, \widehat{\mathbf{v}}$ *and covariances* $\Sigma_{\mathbf{uu}}$, $\Sigma_{\mathbf{vv}}$, $\Sigma_{\mathbf{uv}} = \Sigma_{\mathbf{vu}}^{\mathsf{T}}$, *then* $\mathbf{u}$ *given* $\mathbf{v}$ *is Gaussian with mean and covariance*

$$\begin{aligned} E\left[\mathbf{u}|\mathbf{v}\right] &= \widehat{\mathbf{u}} + \Sigma_{\mathbf{uv}}\Sigma_{\mathbf{vv}}^{-1}\left(\mathbf{v} - \widehat{\mathbf{v}}\right) \\ Cov\left[\mathbf{u}|\mathbf{v}\right] &= \Sigma_{\mathbf{uu}} - \Sigma_{\mathbf{uv}}\Sigma_{\mathbf{vv}}^{-1}\Sigma_{\mathbf{vu}} \end{aligned}$$

Applying this to our problem with $\mathbf{u} = \mathbf{x}_{k+1}$ and $\mathbf{v} = \mathbf{y}_k$ yields

Theorem (Kalman filter)

The mean $\widehat{\mathbf{x}}$ *and covariance* Σ *of the Gaussian posterior satisfy*

$$\begin{aligned} \widehat{\mathbf{x}}_{k+1} &= A\widehat{\mathbf{x}}_k + K_k\left(\mathbf{y}_k - H\widehat{\mathbf{x}}_k\right) \\ \Sigma_{k+1} &= CC^{\mathsf{T}} + \left(A - K_kH\right)\Sigma_kA^{\mathsf{T}} \\ K_k &\triangleq A\Sigma_kH^{\mathsf{T}}\left(DD^{\mathsf{T}} + H\Sigma_kH^{\mathsf{T}}\right)^{-1} \end{aligned}$$

Duality of LQG control and Kalman filtering

LQG controller

State dynamics:

$$\mathbf{x}_{k+1} = (A - BL_k)\,\mathbf{x}_k + C\boldsymbol{\varepsilon}_k$$

Gain matrix:

$$L_k = \left(R + B^\mathsf{T} V_{k+1} B\right)^{-1} B^\mathsf{T} V_{k+1} A$$

Backward Riccati equation:

$$V_k = Q + A^\mathsf{T} V_{k+1}\,(A - BL_k)$$

Kalman filter

Estimated state dynamics:

$$\widehat{\mathbf{x}}_{k+1} = (A - K_k H)\,\widehat{\mathbf{x}}_k + K_k \mathbf{y}_k$$

Gain matrix:

$$K_k = A\Sigma_k H^\mathsf{T}\left(DD^\mathsf{T} + H\Sigma_k H^\mathsf{T}\right)^{-1}$$

Forward Riccati equation:

$$\Sigma_{k+1} = CC^\mathsf{T} + (A - K_k H)\,\Sigma_k A^\mathsf{T}$$

Duality of LQG control and Kalman filtering

LQG controller

State dynamics:

$$\mathbf{x}_{k+1} = (A - BL_k)\,\mathbf{x}_k + C\boldsymbol{\varepsilon}_k$$

Gain matrix:

$$L_k = \left(R + B^\mathsf{T} V_{k+1} B\right)^{-1} B^\mathsf{T} V_{k+1} A$$

Backward Riccati equation:

$$V_k = Q + A^\mathsf{T} V_{k+1} (A - BL_k)$$

Kalman filter

Estimated state dynamics:

$$\widehat{\mathbf{x}}_{k+1} = (A - K_k H)\,\widehat{\mathbf{x}}_k + K_k \mathbf{y}_k$$

Gain matrix:

$$K_k = A\Sigma_k H^\mathsf{T} \left(DD^\mathsf{T} + H\Sigma_k H^\mathsf{T}\right)^{-1}$$

Forward Riccati equation:

$$\Sigma_{k+1} = CC^\mathsf{T} + (A - K_k H)\,\Sigma_k A^\mathsf{T}$$

This form of duality does not generalize to non-LQG systems. However there is a different duality which does generalize (see later). It involves an information filter, computing Σ^{-1} instead of Σ.