

# Multivariate Gaussians

Sargur Srihari

[srihari@cedar.buffalo.edu](mailto:srihari@cedar.buffalo.edu)

# Topics

1. Multivariate Gaussian: Basic Parameterization
2. Covariance and Information Form
3. Operations on Gaussians
4. Independencies in Gaussians

# Three Representations of Multivariate Gaussian

- Equivalent representations:
  1. Basic parameterization
    1. Covariance form,
    2. Information (or precision ) form
  2. Gaussian Bayesian network
    - Information form captures conditional independencies needed for BN
  3. Gaussian Markov Random Field
    - Easiest conversion

# Parameterizing Multivariate Gaussian

1. Basic parameterization: covariance form
2. Multivariate Gaussian distribution over  $X_1, \dots, X_n$ 
  - is characterized by an  $n$ -dimensional mean vector  $\mu$  and a symmetric  $n \times n$  covariance matrix  $\Sigma$
3. The density function is defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right]$$

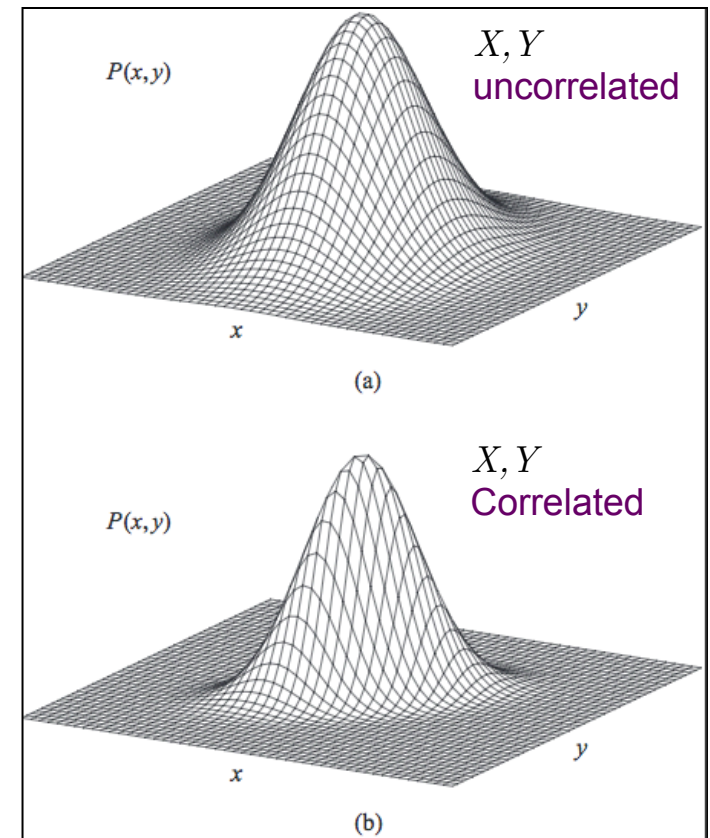
# Covariance Intuition in Gaussian

- The pdf specifies a set of ellipsoidal contours around the mean vector  $\mu$

– Contours are parallel

- Each corresponds to a value of the density function
- Shape of ellipsoid and steepness of contour are determined by  $\Sigma$

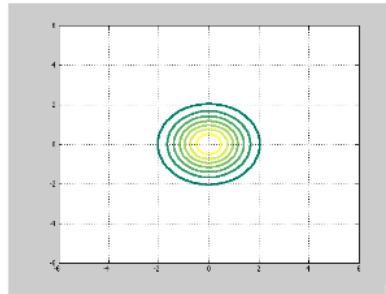
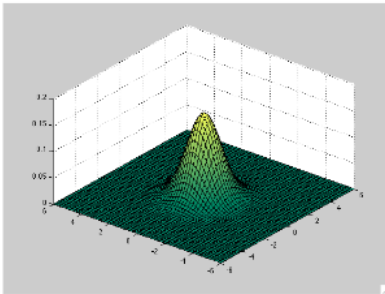
$$\begin{aligned}\mu &= E[\mathbf{x}] \\ \Sigma &= E[\mathbf{x}\mathbf{x}^t] - E[\mathbf{x}]E[\mathbf{x}]^t\end{aligned}$$



$$\mu = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{bmatrix}$$

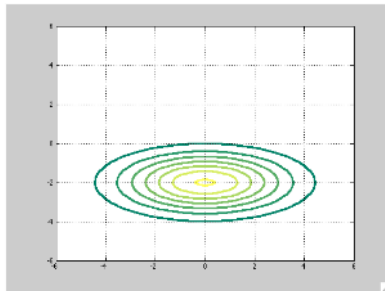
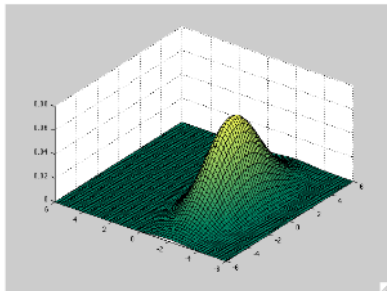
$X_1$  is negatively correlated (-2) with  $X_3$   
 $X_2$  is negatively correlated (-5) with  $X_3$

# Multivariate Gaussian examples



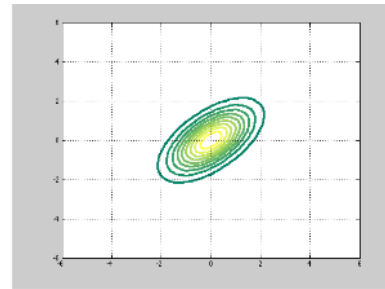
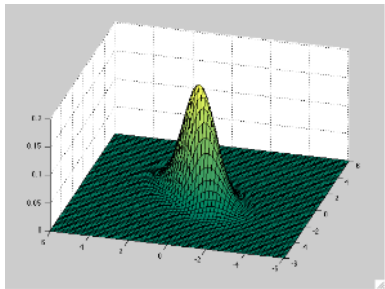
$$\mu = [0, 0]$$

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



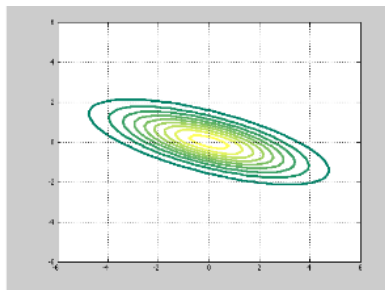
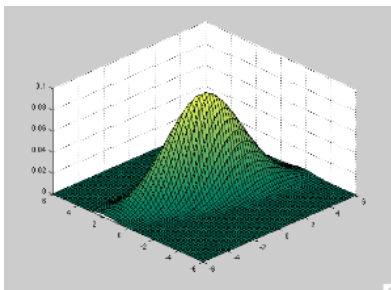
$$\mu = [0, -2]$$

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = [0, 0]$$

$$\Sigma = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$



$$\mu = [0, 0]$$

$$\Sigma = \begin{bmatrix} 5 & -1.5 \\ -1.5 & 1 \end{bmatrix}$$

# Deriving the Information Form

- Covariance Form of multivariate Gaussian is:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

- Using inverse covariance matrix Precision  $J = \Sigma^{-1}$

$$\begin{aligned} -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t J (\mathbf{x} - \boldsymbol{\mu}) \\ &= -\frac{1}{2} \left[ \mathbf{x}^t J \mathbf{x} - 2\mathbf{x}^t J \boldsymbol{\mu} + \boldsymbol{\mu}^t J \boldsymbol{\mu} \right] \end{aligned}$$

– Since last term is constant, we get information form

$$p(\mathbf{x}) \propto \exp\left[-\frac{1}{2} \mathbf{x}^t J \mathbf{x} + (J\boldsymbol{\mu})^t \mathbf{x}\right]$$

- $J\boldsymbol{\mu}$  is called the potential vector

# Information form of Gaussian

- Since  $\Sigma$  is invertible we can also define the Gaussian in terms of the inverse covariance matrix  $J = \Sigma^{-1}$
- Called the *information matrix* or *precision matrix*
- It induces an alternate form of the Gaussian density

$$p(\mathbf{x}) \propto \exp \left[ -\frac{1}{2} \mathbf{x}^t J \mathbf{x} + (J \boldsymbol{\mu})^t \mathbf{x} \right]$$

- $J \boldsymbol{\mu}$  is called the potential vector



# Operations on Gaussians

- There are two main operations we wish to perform on a distribution:
  1. Compute marginal distribution over some subset  $Y$
  2. Conditioning the distribution on some  $Z=z$
- Each operation is very easy in one of the two ways of encoding a Gaussian
  1. Marginalization is trivial in the covariance form
  2. Conditioning a Gaussian is very easy to perform in the information form

# Marginalization in Covariance Form

- Marginal distribution of any subset of variables
  - Trivially read from the mean and covariance matrix

$$\begin{array}{|l}
 p(X_1, X_2, X_3) = N(\mu, \Sigma) \\
 p(X_2, X_3) = \sum_{X_1} p(X_1, X_2, X_3) \\
 \quad = N(\mu', \Sigma')
 \end{array}
 \quad
 \mu = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix} \quad
 \Sigma = \begin{bmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{bmatrix}
 \text{ then } \mu' = \begin{bmatrix} -3 \\ 4 \end{bmatrix} \quad
 \Sigma' = \begin{bmatrix} 5 & -5 \\ -5 & 8 \end{bmatrix}$$

- More generally, if we have a joint distribution over  $\{\mathbf{X}, \mathbf{Y}\}$ ,  $\mathbf{X} \in \mathcal{R}^n$ ,  $\mathbf{Y} \in \mathcal{R}^m$  then we can decompose mean and covariance of joint as:

$$p(\mathbf{X}, \mathbf{Y}) = N \left( \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}; \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$

where  $\mu_X \in \mathcal{R}^n$ ,  $\mu_Y \in \mathcal{R}^m$ ,  $\Sigma_{XX}$  is  $n \times n$ ,  $\Sigma_{XY}$  is  $n \times m$ ,  $\Sigma_{YX} = \Sigma_{XY}^T$  is  $m \times n$ ,  $\Sigma_{YY}$  is  $m \times m$

- **Lemma:** If  $\{\mathbf{X}, \mathbf{Y}\}$  have a joint normal distribution then marginal of  $\mathbf{Y}$  is a normal distribution  $N(\mu_Y; \Sigma_{YY})$

# Conditioning in Information Form

- Conditioning on an observation  $Z=z$  is very easy to perform in the information form
- Simply assign  $Z=z$  in the equation

$$p(\mathbf{x}) \propto \exp\left[-\frac{1}{2}\mathbf{x}^t J \mathbf{x} + (J\boldsymbol{\mu})^t \mathbf{x}\right]$$

- It turns some of the quadratic terms into linear terms or even constant terms
- Expression has same form with fewer terms
- Not so straightforward in covariance form

# Summary of Marginalization & Conditioning

- To marginalize a Gaussian over a subset of variables:
  - Compute pairwise covariances
  - Which is precisely generating the distribution in its covariance form
- To condition a Gaussian on an observation:
  - Invert covariance matrix to obtain information form
    - Inversion feasible for small matrices
    - Too costly for high-dimensional space

# Independencies in Gaussians

- *Theorem:* If  $\mathbf{X} = X_1, \dots, X_n$  have a joint normal distribution  $N(\mu, \Sigma)$  then  $X_i$  and  $X_j$  are independent if and only if  $\Sigma_{i,j} = 0$
- This property does not hold in general
  - If  $p(X, Y)$  is not Gaussian then it is possible that  $\text{Cov}[X; Y] = 0$  while  $X$  and  $Y$  are still dependent in  $p$

# Independencies from Information Matrix

- Theorem:*

Consider a Gaussian distribution

$$p(X_1, \dots, X_n) = N(\mu, \Sigma)$$

and let  $J = \Sigma^{-1}$  be the information matrix.

Then  $J_{ij} = 0$  if and only if  $p \Rightarrow (X_i \perp X_j \mid \mathcal{X} - \{X_i, X_j\})$

- Example:*

$$\mu = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{bmatrix} \quad \text{then } J = \begin{bmatrix} 0.3125 & -0.125 & 0 \\ -0.125 & 0.5833 & 0.333 \\ 0 & 0.333 & 0.333 \end{bmatrix}$$

– Then  $X_1, X_3$  are conditionally independent given  $X_2$

# Information matrix defines an I-map

- Information matrix  $J$  captures independencies between pairs of variables conditioned on all the remaining variables

– Can use  $J$  to construct unique minimal I-map for  $p$

## Definition of I-map

- $I$  independence assertion set of form  $(X \perp Y / Z)$ , which are true of distribution  $P$
- $G$  is an I-map of  $P$  implies:  $I_G \subseteq I$ 
  - » Minimal I-map: removal of a single edge renders the graph not an I-map
- Introduce an edge between  $X_i$  and  $X_j$  whenever  $(X_i \perp X_j \mid \mathcal{X} - \{X_i, X_j\})$  does not hold in  $p$ 
  - This is precisely when  $J_{ij} \neq 0$