Probabilistic Robotics Course

Gaussian Distribution

Giorgio Grisetti
grisetti@diag.uniroma1.it

Department of Computer Control and Management Engineering
Sapienza University of Rome

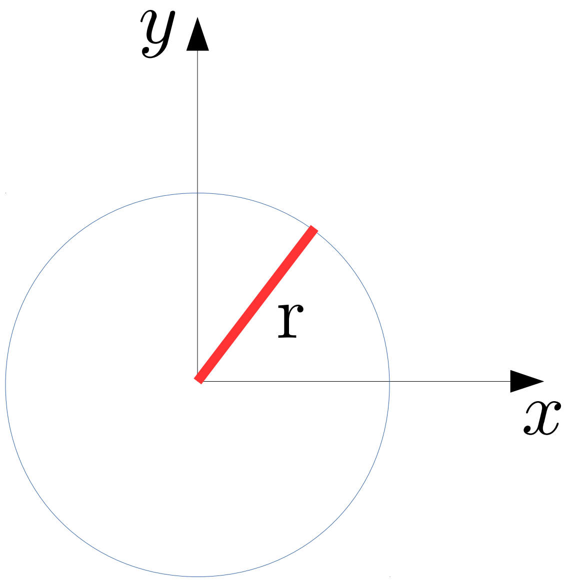
Outline

- Drawing Ellipses
- Parametrizations
- Drawing Gaussians
- Classical Parametrization
  - Marginalization
  - Conditioning
  - Chain Rule
  - Affine Functions
  - Quasi-Affine Functions

Circles

A circle looks like that:

$$\mathrm{x}^2 + y^2 = r^2$$

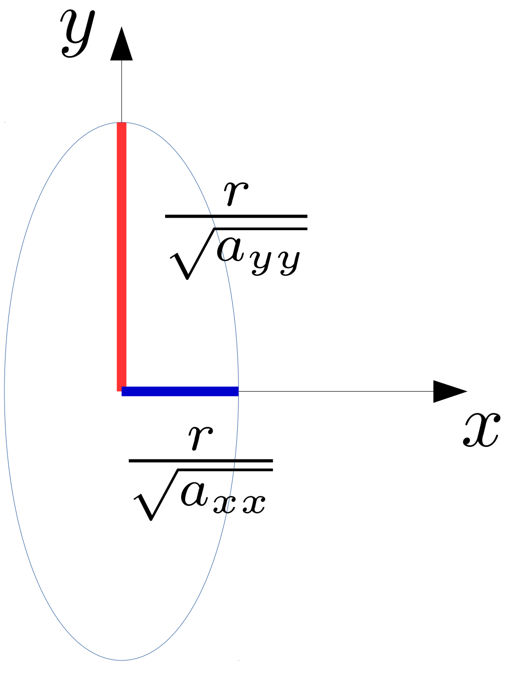


It is a slice of a paraboloid.

Scaled Circles

A *scaled* circle like that:

$$a_{xx}x^2 + a_{yy}y^2 = r^2$$



Slanted Circles

A *slanted scaled* circle like that:

$$\mathrm{a}_{xx}x^2 + a_{xy}xy + a_{yy}y^2 = r^2$$

can be rewritten as:

$$\begin{pmatrix} x & y \end{pmatrix} \underbrace{\begin{pmatrix} a_{xx} & \frac{a_{xy}}{2} \\ \frac{a_{xy}}{2} & a_{yy} \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x \\ y \end{pmatrix} = r^2$$

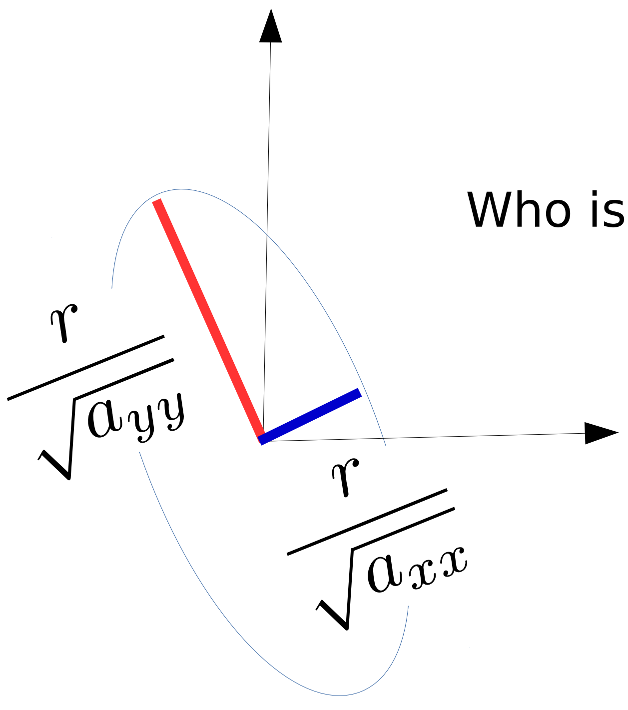
The matrix $\mathbf{A}$ admits an *eigenvalue decomposition:*

$$\mathbf{A} = {}^{T} \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$$

Slanted Circles (cont.)

All this because of a_{xy}.

The off diagonal components “rotate” the ellipsoid.

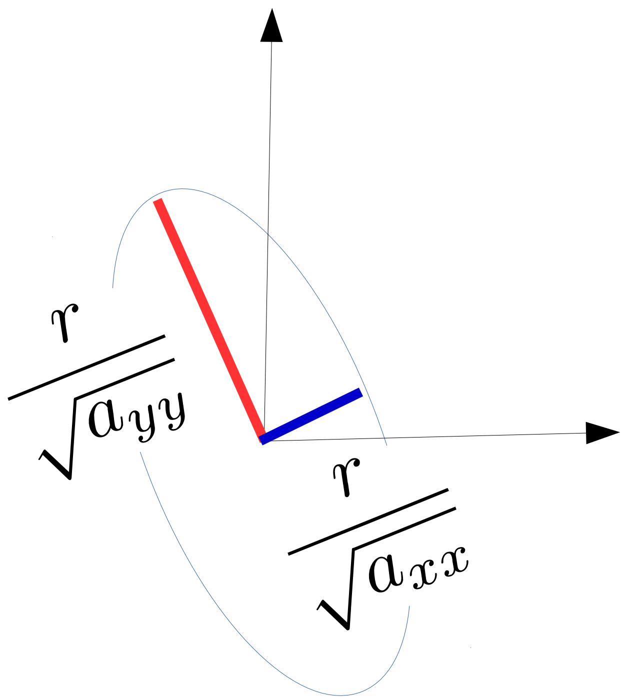


Who is guilty for the rotation?

Breaking News

Ellipses can also be translated:

$$\mathbf{A} = \left[\mathbf{R}\begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix}\right]^T \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \left[\mathbf{R}\begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix}\right] = r^2$$



Gaussian

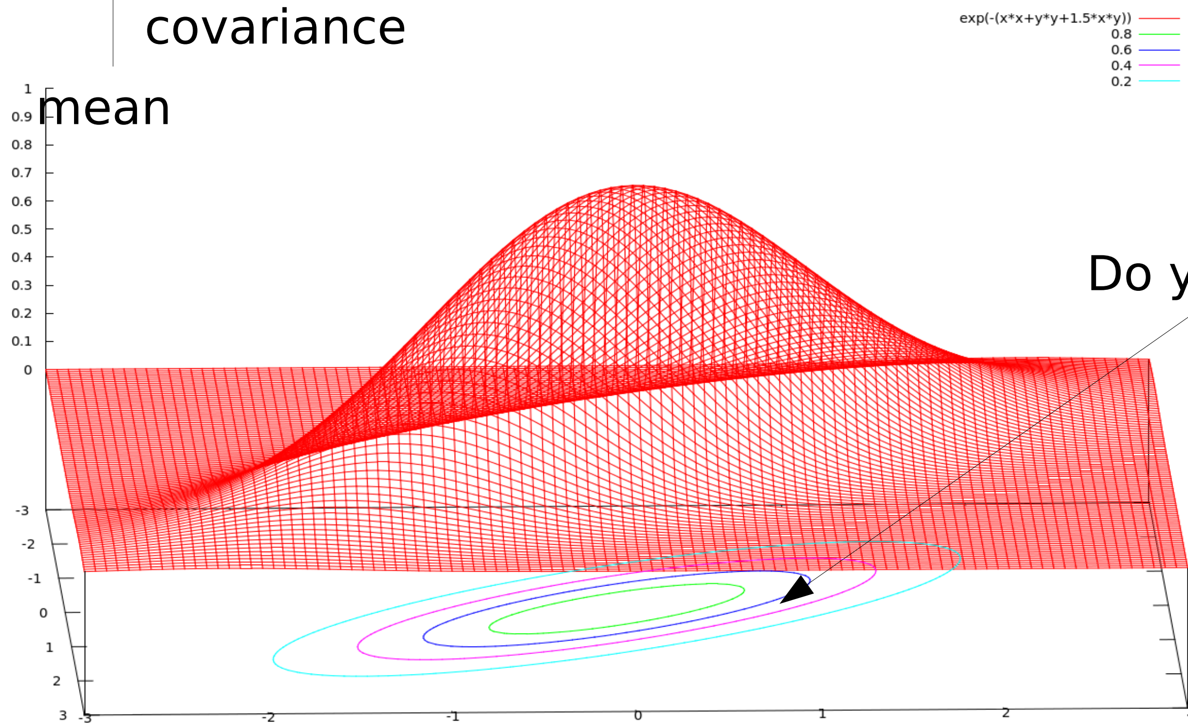
The PDF of a Gaussian distribution has the following form:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{\Sigma}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right)$$

covariance

mean

Do you see the ellipse???



Why Gaussians are cool

Gaussian distributions are closed under:

- sum
- affine transformation (Ax+b)
- conditioning
- marginalization

This means that in order to implement the above operations, one only needs to compute the **parameters** of the result, from the parameters of the input

Moment Parametrization

The one seen in the previous slide is known as *moment parameterization.*

The parameters can be calculated from a (large) set of samples as:

$$\mu = \frac{1}{N} \sum \mathbf{x}^{(i)}$$

$$\mathbf{\Sigma} = \frac{1}{N} \sum \left(\mathbf{x}^{(i)} - \mu\right)\left(\mathbf{x}^{(i)} - \mu\right)^T$$

Moment Parametrization

The parameters are defined as the 1st and 2nd order moments of the distribution:

$$\mu = \int_{\Omega} \mathbf{x}\, p(\mathbf{x}) d\mathbf{x} = \mathbb{E}[\mathbf{x}]$$

$$\mathbf{\Sigma} = \int_{\Omega} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T\, p(\mathbf{x}) d\mathbf{x} = \mathbb{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$$

Canonical Parametrization

Another parametrization is the so called *canonical*, useful for conditioning:

ν information vector

$\mathbf{\Omega}$ information matrix

$$\mathcal{N}^{-1}(\mathbf{x}; \nu, \mathbf{\Omega}) \quad = \frac{\exp\left(\frac{1}{2}\nu^T \mathbf{\Omega}^{-1} \nu\right)\sqrt{\det \mathbf{\Omega}}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{\Omega} \mathbf{x} + \mathbf{x}^T \nu\right)$$

Partitioned Gaussian Densities

The space can be split in two subspaces.

The density is over a joint distribution:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \nu = \begin{pmatrix} \nu_a \\ \nu_b \end{pmatrix}$$

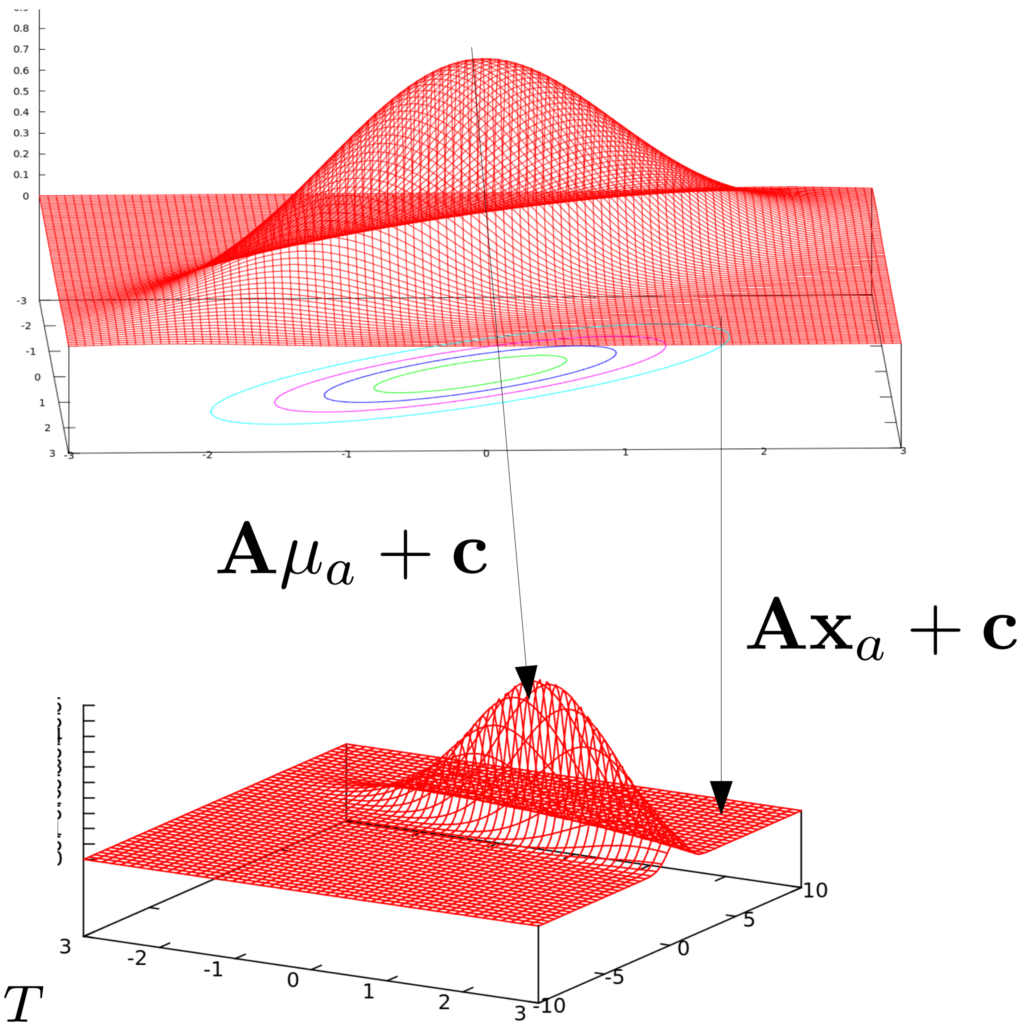
$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix} \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{aa} & \boldsymbol{\Omega}_{ab} \\ \boldsymbol{\Omega}_{ba} & \boldsymbol{\Omega}_{bb} \end{pmatrix}$$

Affine Transformation

Let $\mathbf{x}_a$ be a Gaussian random variable such that:

$$\mathbf{x}_a \sim \mathcal{N}(\mathbf{x}_a, \mu_a, \mathbf{\Sigma}_a).$$

Let $\mathbf{x}_b = \mathbf{f}(\mathbf{x}_a) = \mathbf{A}\mathbf{x}_a + \mathbf{c}$ an affine transformation of $\mathbf{x}_a$.



$\mathbf{x}_b$ is Gaussian:

$$p(\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_b; \mu_b, \mathbf{\Sigma}_b)$$

The parameters are:

$$\mu_b = \mathbf{A}\mu + \mathbf{c} \qquad \mathbf{\Sigma}_b = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$$

Taylor Expansion

For non-linear transformations, we can approximate the function around a *linearization point* $\mathbf{x}_0$:

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}) &\simeq \mathbf{f}(\mathbf{x}_0) + \underbrace{\frac{\partial \mathbf{f}(\mathbf{x},\mathbf{u})}{\partial \mathbf{x}}}_{\mathbf{A}}(\mathbf{x}-\mathbf{x}_0) \\
&= \mathbf{A}\mathbf{x} + \underbrace{\mathbf{f}(\mathbf{x}_0) - \mathbf{A}\mathbf{x}_0}_{\mathbf{b}}
\end{aligned}
$$

- This reduces the transformation to an affine transform
- The approximation holds only around a linearization point.
- The farther ***f*** is from being linear, the worse the approximation

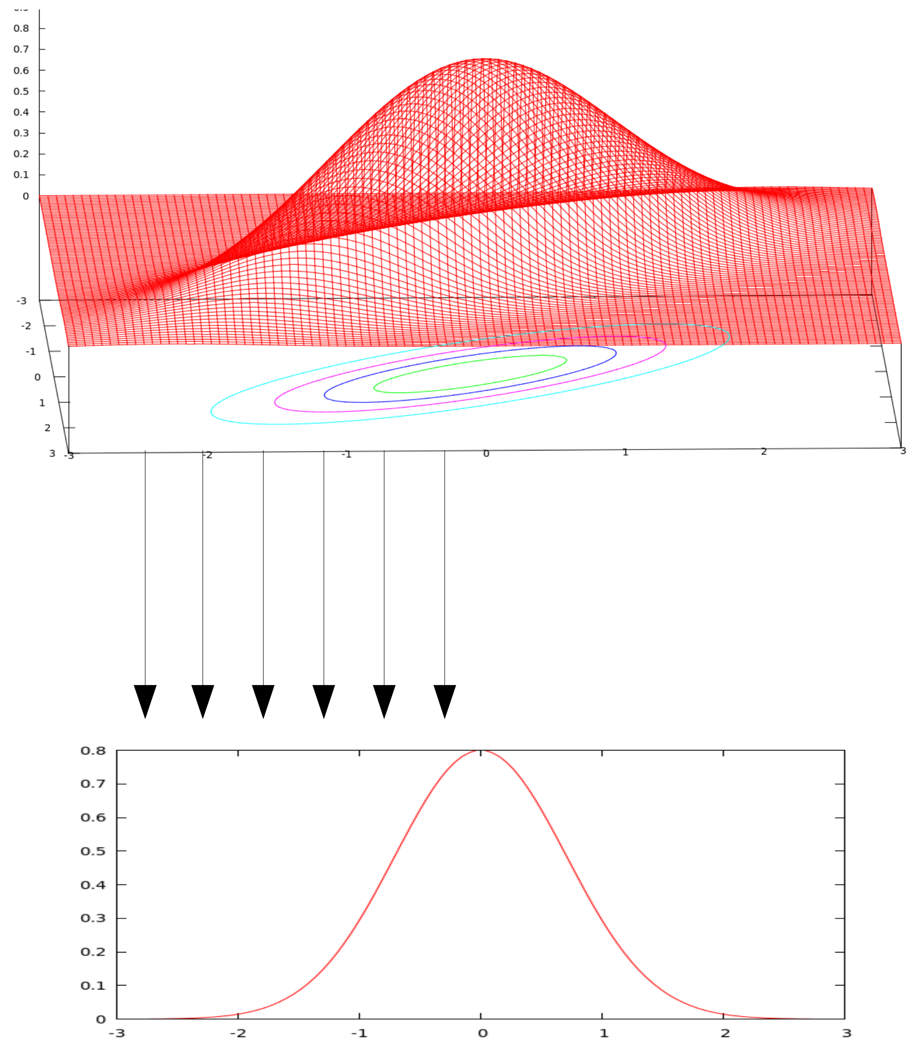
Marginalization

Let $\mathbf{x}^T = (\mathbf{x}_a^T \; \mathbf{x}_b^T)$ be a Gaussian random variable such that $\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mu, \mathbf{\Sigma})$.

The marginal $p(\mathbf{x}_a) = \int_{\mathbf{x}_b} p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b$

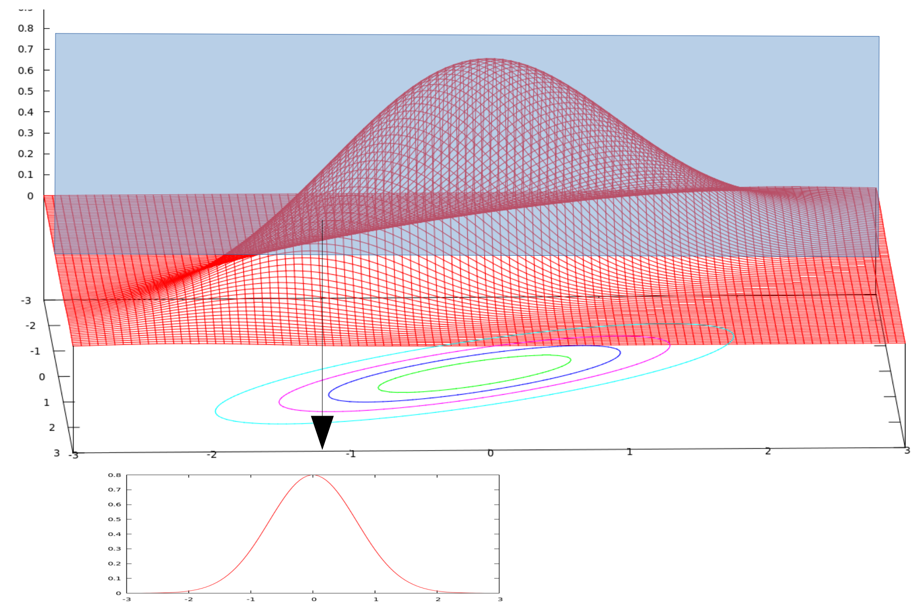
is Gaussian with parameters

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_a, \mathbf{\Sigma}_a)$$



Conditioning (1)

Let $\mathbf{x}^T = (\mathbf{x}_a^T \; \mathbf{x}_b^T)$ be a Gaussian random variable such that $\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mu, \mathbf{\Sigma})$.



The conditional

$$p(\mathbf{x}_a \mid \mathbf{x}_b) = \frac{p(\mathbf{x}_a, \mathbf{x}_b)}{\int_{\mathbf{x}_a} p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_a}$$

is Gaussian with parameters

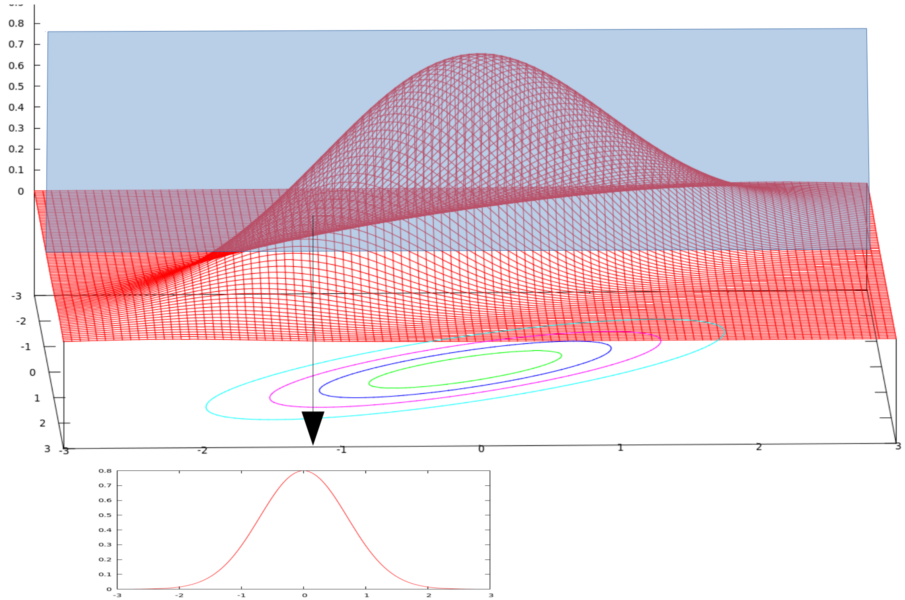
$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_{a|b}, \mathbf{\Sigma}_{a|b})$$

$$\mu_{a|b} = \mu_a + \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mu_b)$$

$$\mathbf{\Sigma}_{a|b} = \mathbf{\Sigma}_{aa} - \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}\mathbf{\Sigma}_{ba}$$

Conditioning (2)

Let $\mathbf{x}^T = (\mathbf{x}_a^T \ \mathbf{x}_b^T)$ be a Gaussian random variable such that $\mathbf{x} \sim \mathcal{N}^{-1}(\mathbf{x}, \nu, \mathbf{\Omega})$.



The conditional

$$p(\mathbf{x}_a \mid \mathbf{x}_b)$$

is Gaussian with parameters

$$p(\mathbf{x}_a) = \mathcal{N}^{-1}(\mathbf{x}_a; \nu_{a|b}, \mathbf{\Omega}_{a|b})$$

$$\nu_{a|b} = \nu_a - \mathbf{\Omega}_{ab}\mathbf{x}_b$$

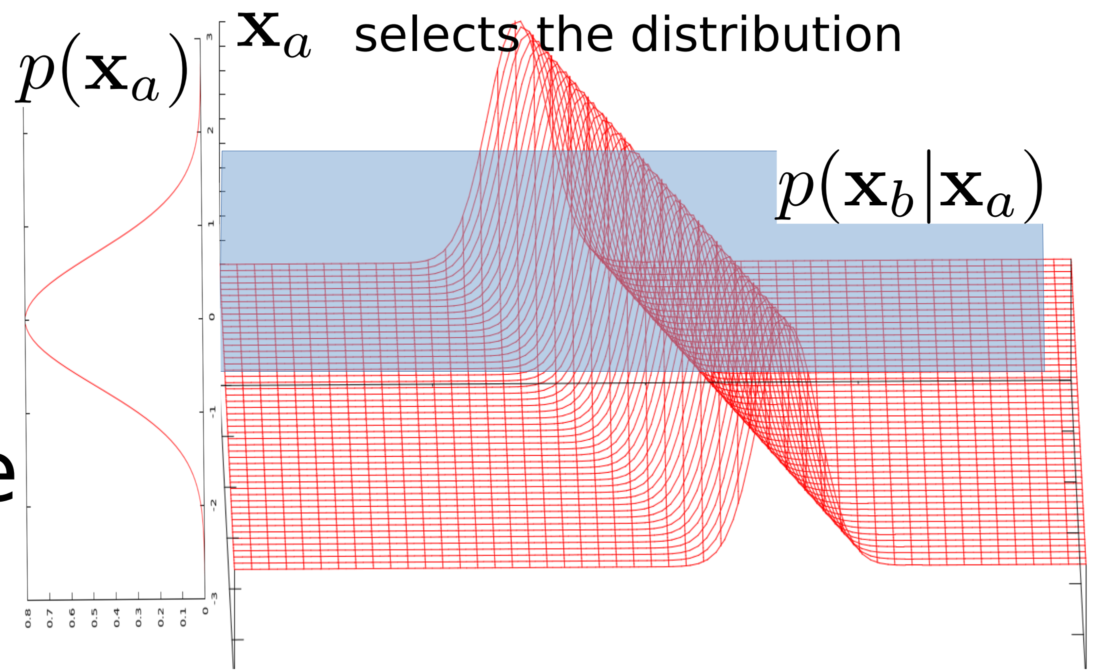
$$\mathbf{\Omega}_{a|b} = \mathbf{\Omega}_{aa}$$

Chain Rule (1)

We know

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_a, \mathbf{\Sigma}_a).$$

$$p(\mathbf{x}_b|\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_b; \underbrace{\mathbf{A}\mathbf{x}_a + \mathbf{c}}_{\mu_{b|a}}, \mathbf{\Sigma}_{b|a})$$



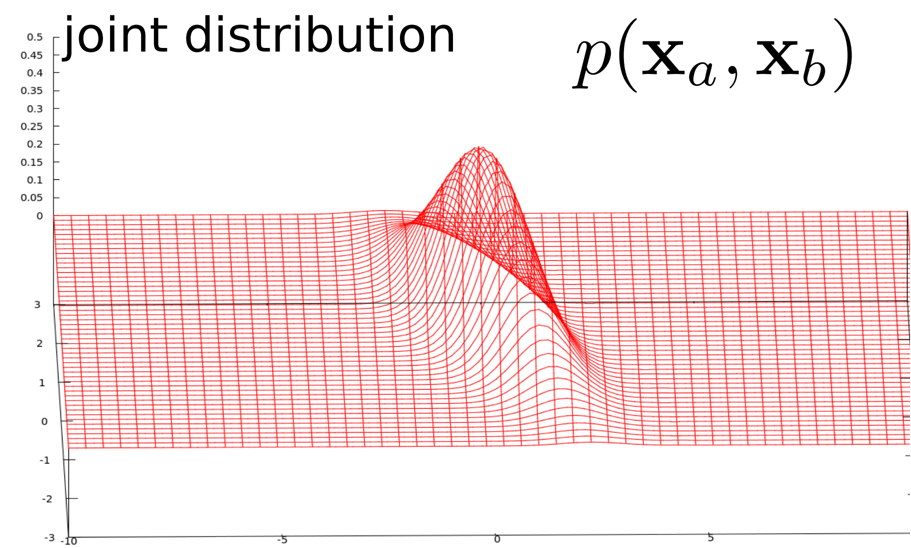
We want to compute

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\mathbf{x}_{a,b}; \mu_{a,b}, \mathbf{\Sigma}_{a,b}\right)$$

The parameters are

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ \mathbf{A}\mu_a + \mathbf{c} \end{pmatrix}$$

$$\mathbf{\Sigma}_{a,b} = \begin{pmatrix} \mathbf{\Sigma}_a & \mathbf{\Sigma}_a \mathbf{A}^T \\ \mathbf{A}\mathbf{\Sigma}_a & \mathbf{\Sigma}_{b|a} + \mathbf{A}\mathbf{\Sigma}_a\mathbf{A}^T \end{pmatrix}$$

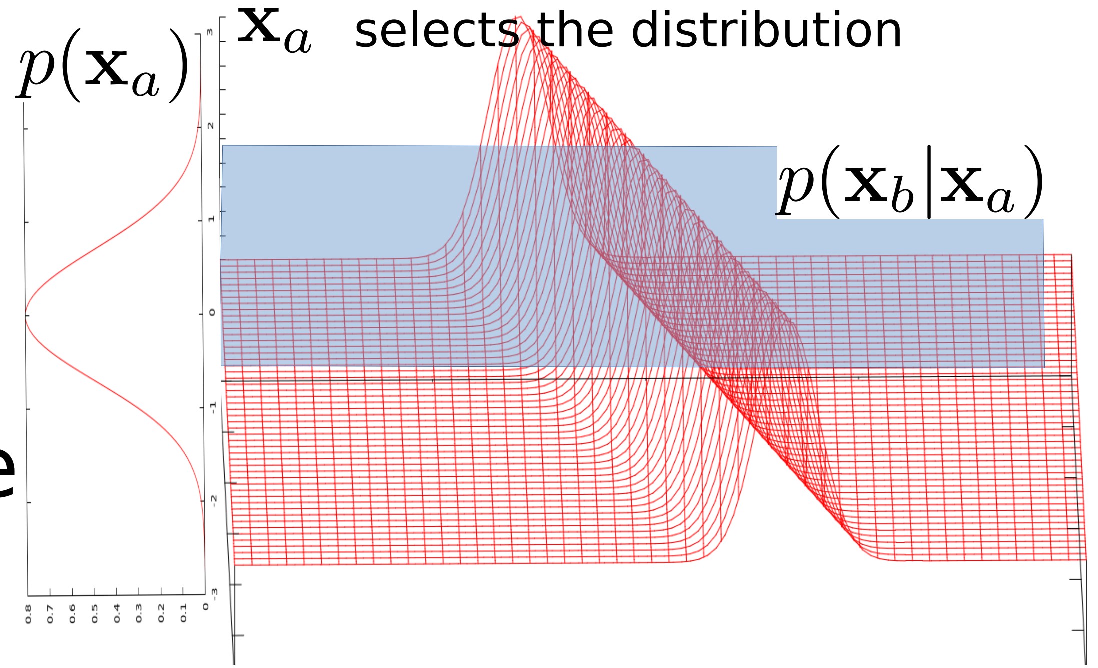


Chain Rule (2)

We know

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_a, \mathbf{\Sigma}_a).$$

$$p(\mathbf{x}_b|\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_b; \mathbf{A}\mathbf{x}_a + \mathbf{c}, \mathbf{\Sigma}_{b|a})$$



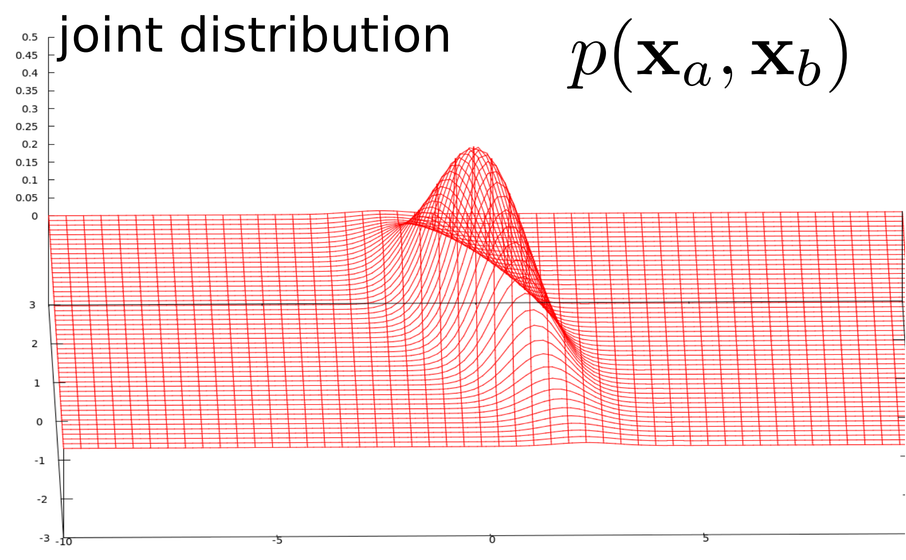
We want to compute

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\mathbf{x}_{a,b}; \mu_{a,b}, \mathbf{\Sigma}_{a,b}\right)$$

The parameters are

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ \mathbf{A}\mu_a + \mathbf{c} \end{pmatrix}$$

$$\mathbf{\Omega}_{a,b} = \begin{pmatrix} \mathbf{A}^T\mathbf{\Omega}_{b|a}\mathbf{A} + \mathbf{\Omega}_a & -\mathbf{A}^T\mathbf{\Omega}_{b|a} \\ -\mathbf{\Omega}_{b|a}\mathbf{A}^T & \mathbf{\Omega}_{b|a} \end{pmatrix}$$



References

Further details are here (warmly recommended):

- *Thomas Schoen,* On Manipulating the Multivariate Gaussian Density