Computer vision: models, learning and inference

Chapter 5

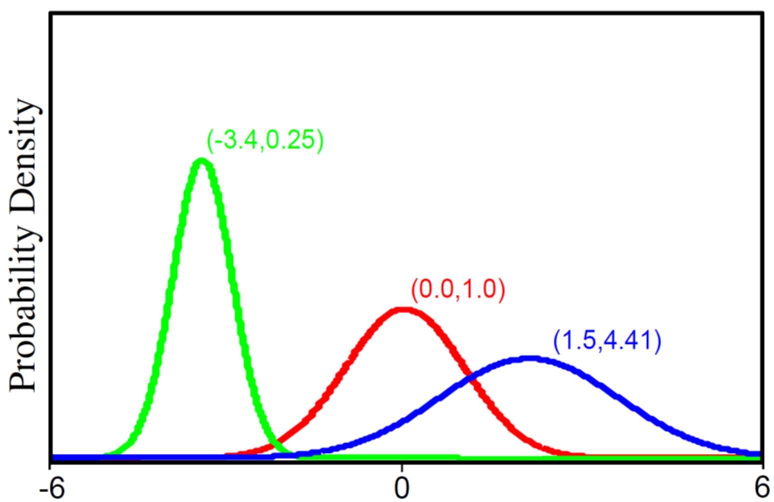
The Normal Distribution

Univariate Normal Distribution

$$Pr(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-0.5(x-\mu)^2/\sigma^2\right]$$

For short we write:

$$Pr(x) = \text{Norm}_x[\mu, \sigma^2]$$



Univariate normal distribution describes single continuous variable.

Takes 2 parameters μ and $\sigma^2 > 0$

Multivariate Normal Distribution

$$Pr(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-0.5(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

For short we write:

$$Pr(\mathbf{x}) = \text{Norm}_{\mathbf{x}}\left[\boldsymbol{\mu}, \boldsymbol{\Sigma}\right]$$

Multivariate normal distribution describes multiple continuous variables. Takes 2 parameters

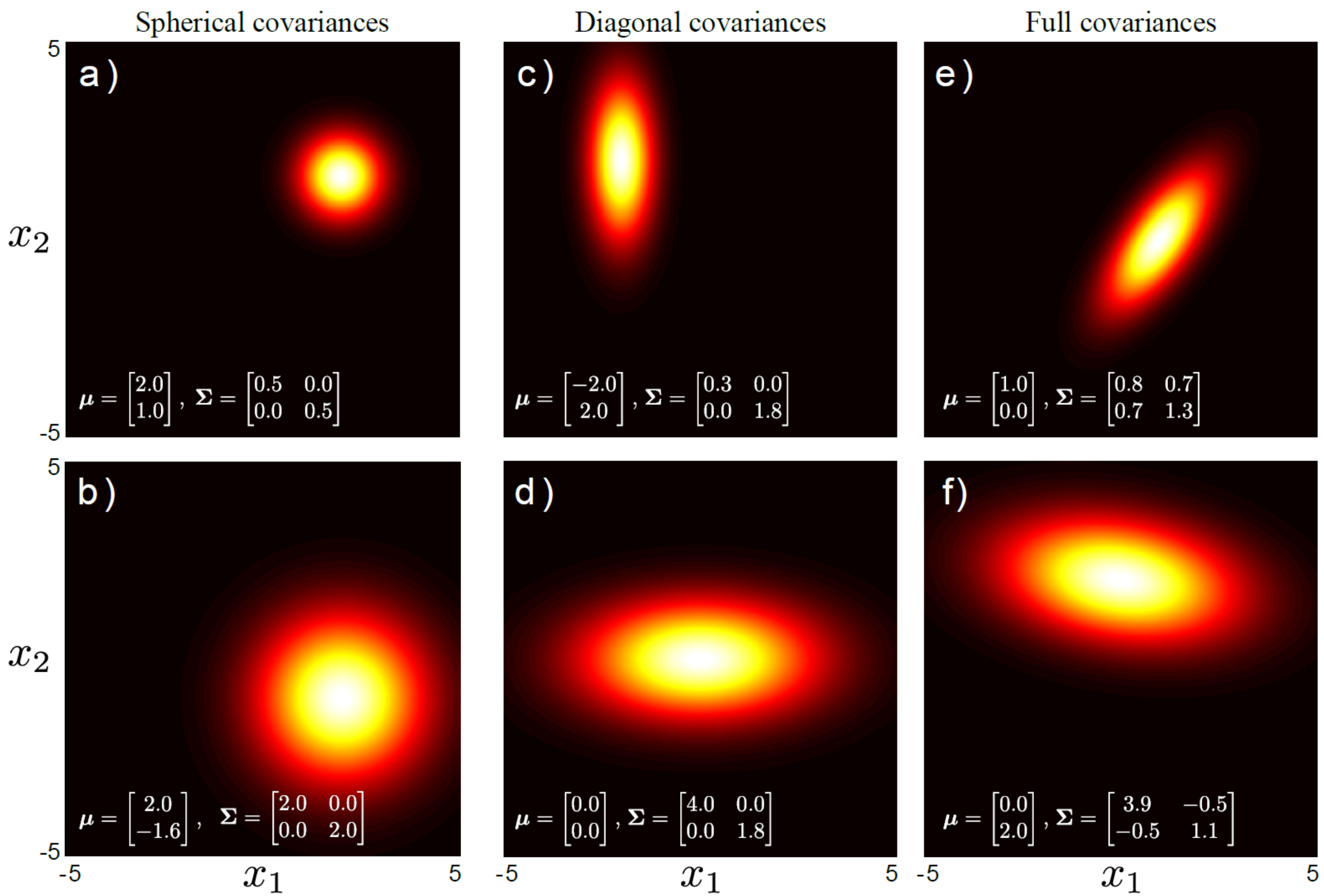
- a vector containing mean position, $\boldsymbol{\mu}$
- a symmetric "positive definite" covariance matrix $\boldsymbol{\Sigma}$

Positive definite: $\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}$ is positive for any real $\mathbf{z}$

Types of covariance

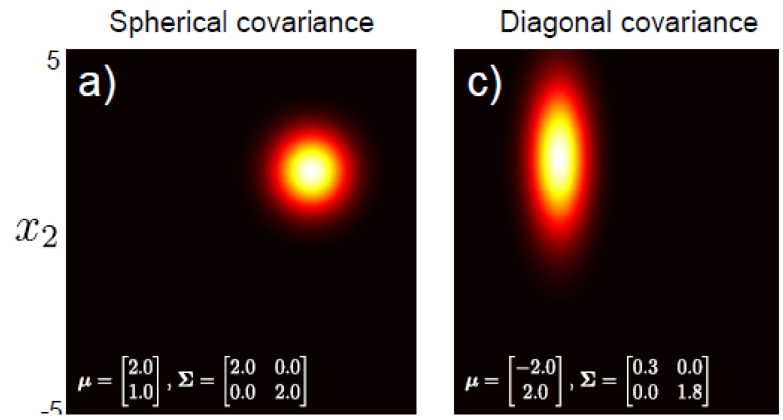
$$\boldsymbol{\Sigma}_{spher} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{diag} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{full} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$$

Symmetric
$\sigma_{12}^2 = \sigma_{21}^2$



Diagonal Covariance = Independence

$$\boldsymbol{\Sigma}_{\text{spher}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{\text{diag}} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$



$$\begin{aligned}
Pr(x_1, x_2) &= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-0.5 \begin{pmatrix} x_1 & x_2 \end{pmatrix} \boldsymbol{\Sigma}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-0.5 \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{x_1^2}{2\sigma_1^2}\right] \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{x_2^2}{2\sigma_2^2}\right] \\
&= Pr(x_1)Pr(x_2)
\end{aligned}$$

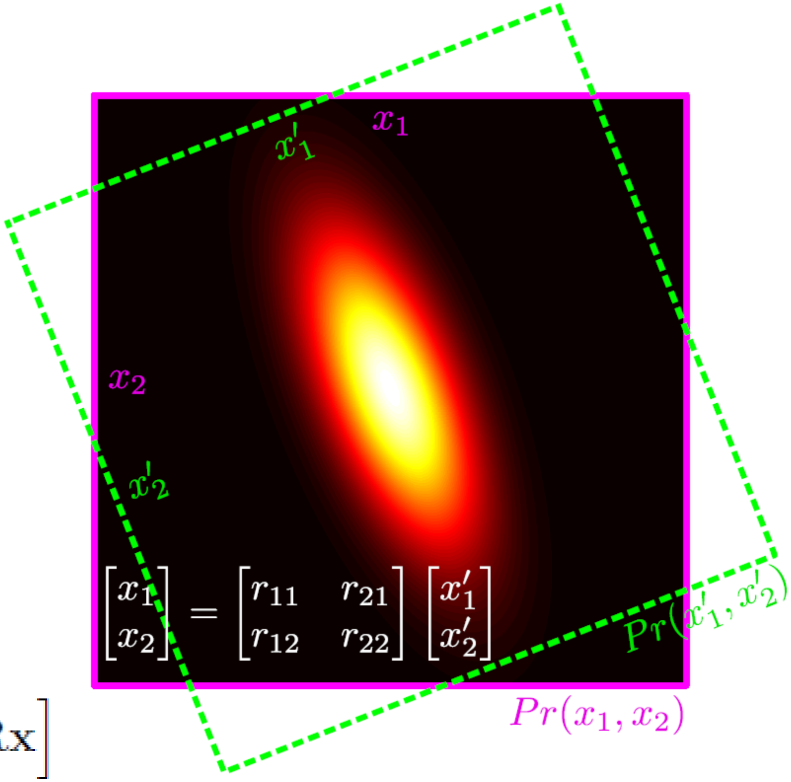
Decomposition of Covariance

Consider green frame of reference:

$$Pr(\mathbf{x}') = \frac{1}{(2\pi)^{K/2}|\Sigma'_{diag}|^{1/2}} \exp\left[-0.5\mathbf{x}'^T\Sigma'^{-1}_{diag}\mathbf{x}'\right]$$

Relationship between pink and green frames of reference:

$$\mathbf{x}' = \mathbf{R}\mathbf{x}$$



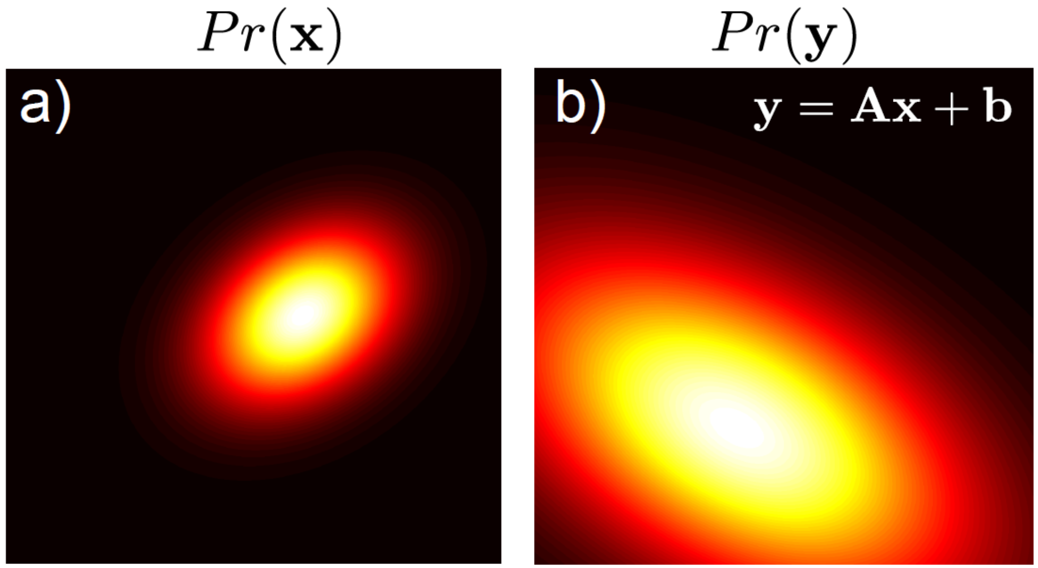
Substituting in:

$$\begin{aligned} Pr(\mathbf{x}) &= \frac{1}{(2\pi)^{K/2}|\Sigma'_{diag}|^{1/2}} \exp\left[-0.5(\mathbf{R}\mathbf{x})^T\Sigma'^{-1}_{diag}\mathbf{R}\mathbf{x}\right] \\ &= \frac{1}{(2\pi)^{K/2}|\mathbf{R}^T\Sigma'_{diag}\mathbf{R}|^{1/2}} \exp\left[-0.5\mathbf{x}^T\mathbf{R}^T\Sigma'^{-1}_{diag}\mathbf{R}\mathbf{x}\right] \end{aligned}$$

Conclusion: $\mathbf{\Sigma}_{full} = \mathbf{R}^T\mathbf{\Sigma}'_{diag}\mathbf{R}$

Full covariance can be decomposed into rotation matrix and diagonal

Transformation of Variables



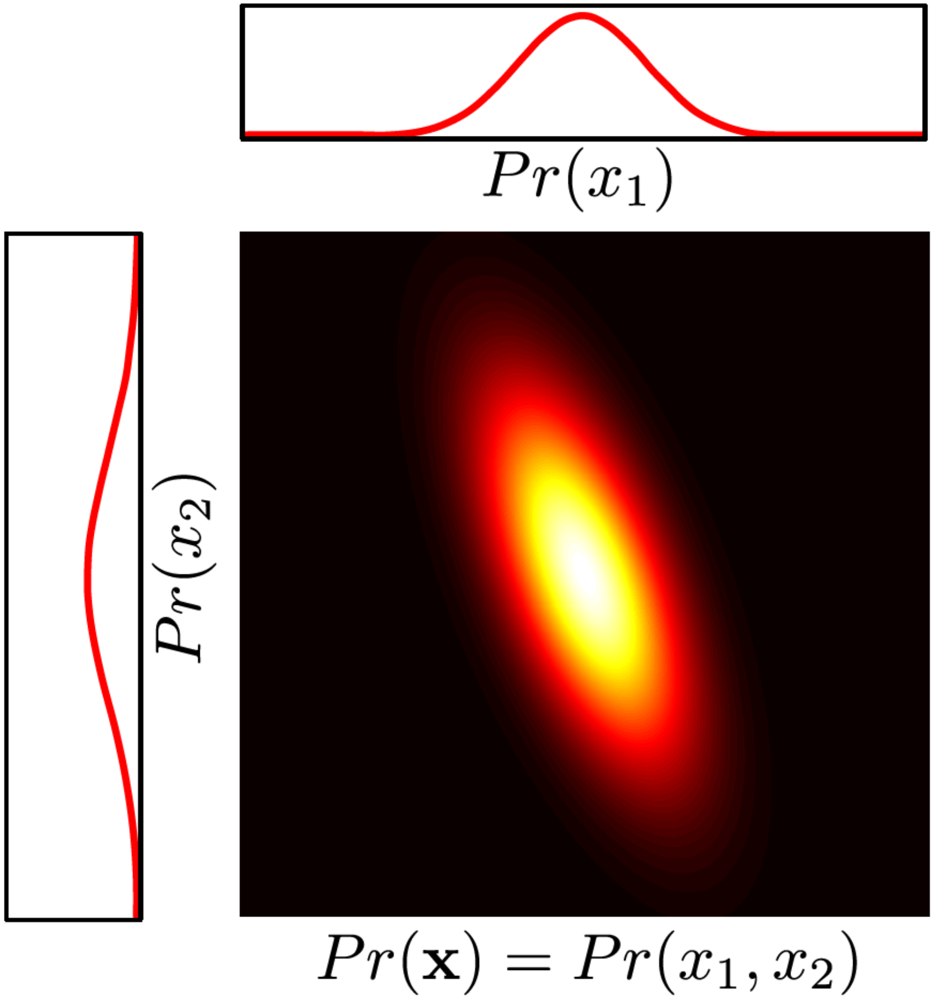
If $Pr(\mathbf{x}) = \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$ and we transform the variable as

The result is also a normal distribution:

$$Pr(\mathbf{y}) = \text{Norm}_{\mathbf{y}}\left[\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A}\right]$$

Can be used to generate data from arbitrary Gaussians from standard one

Marginal Distributions



Marginal distributions of a multivariate normal are also normal

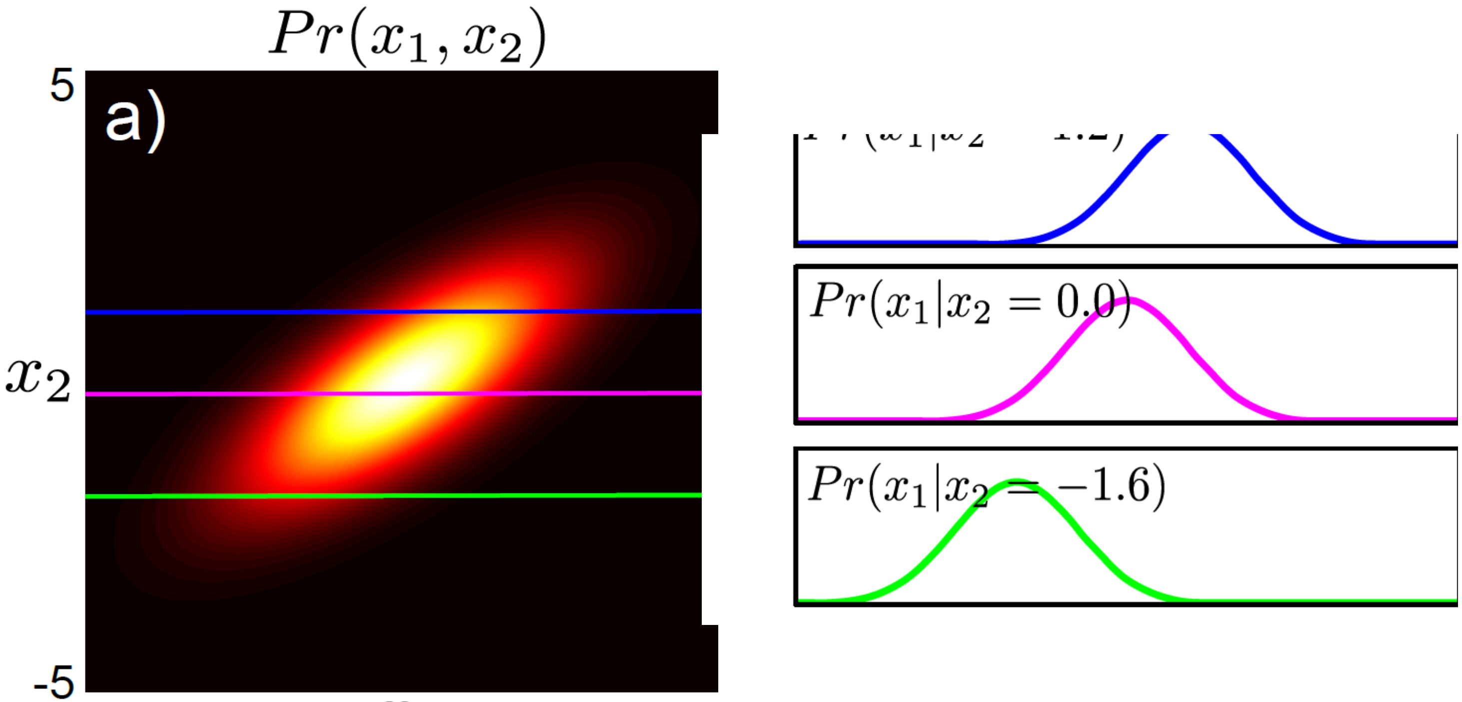
$$Pr(\mathbf{x}) = Pr\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right)$$

$$= \text{Norm}_{\mathbf{x}}\left[\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}^T\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right]$$

then

$$Pr(\mathbf{x}_1) = \text{Norm}_{\mathbf{x}_1}[\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}]$$
$$Pr(\mathbf{x}_2) = \text{Norm}_{\mathbf{x}_2}[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}]$$

Conditional Distributions



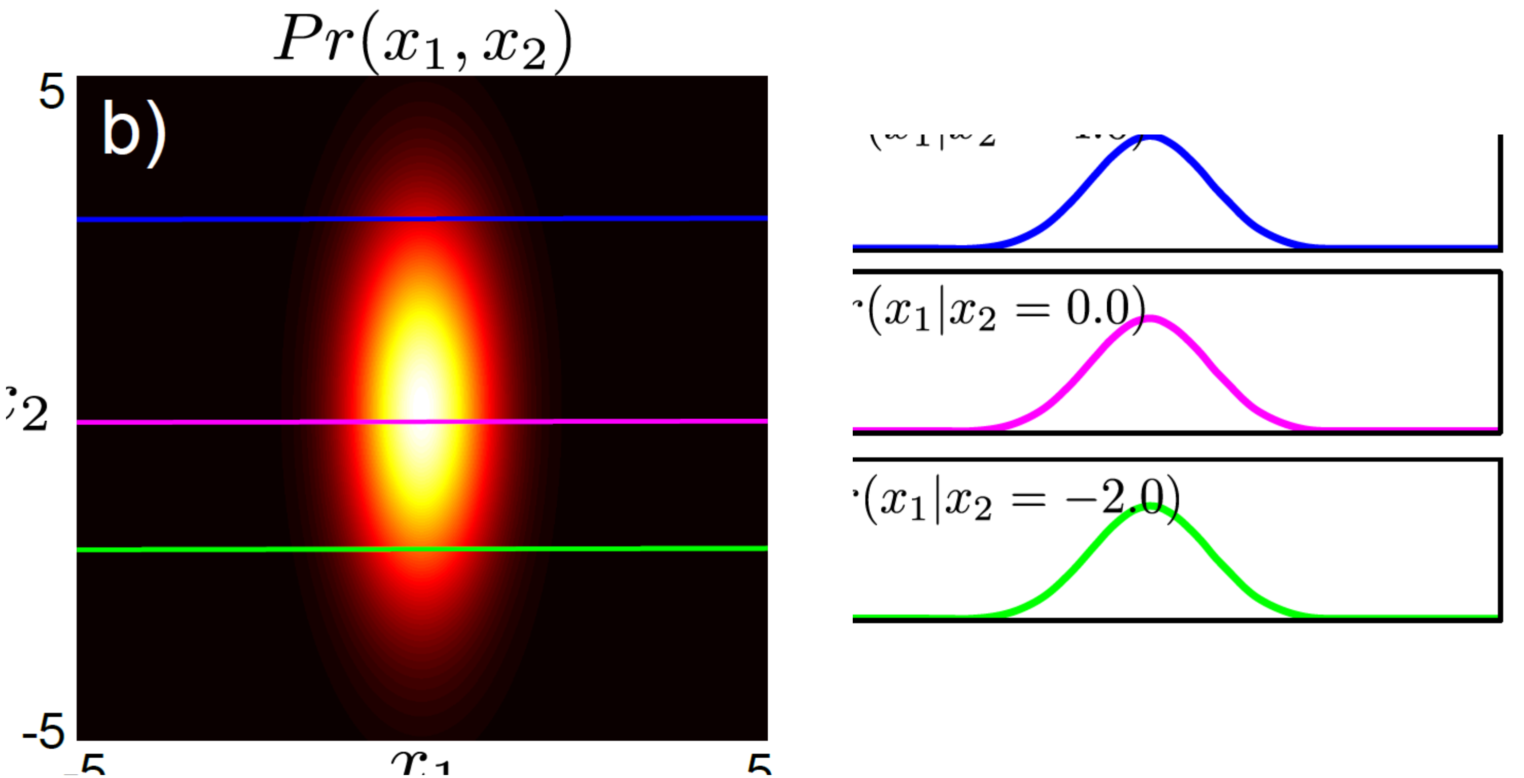
If

$$Pr(\mathbf{x}) = Pr\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right) = \text{Norm}_{\mathbf{x}}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}^T\\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

then

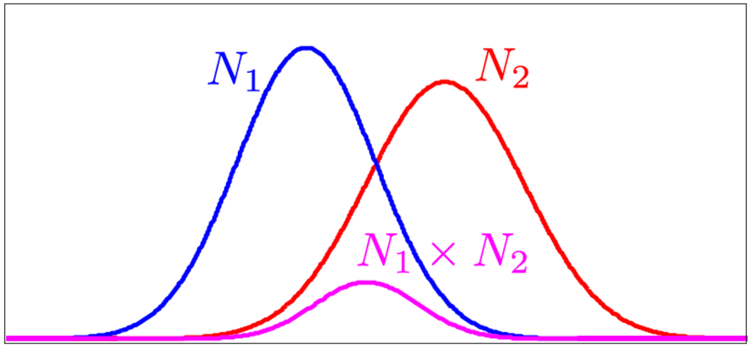
$$\begin{aligned} Pr(\mathbf{x}_1|\mathbf{x}_2) &= \text{Norm}_{\mathbf{x}_1}\left(\boldsymbol{\mu_1} + \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}\right)\\ Pr(\mathbf{x}_2|\mathbf{x}_1) &= \text{Norm}_{\mathbf{x}_2}\left(\boldsymbol{\mu_2} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}^T\right) \end{aligned}$$

Conditional Distributions



For spherical / diagonal case, x_1 and x_2 are independent so all of the conditional distributions are the same.

Product of two normals
(self-conjugacy w.r.t mean)



The product of any two normal distributions in the same variable is proportional to a third normal distribution

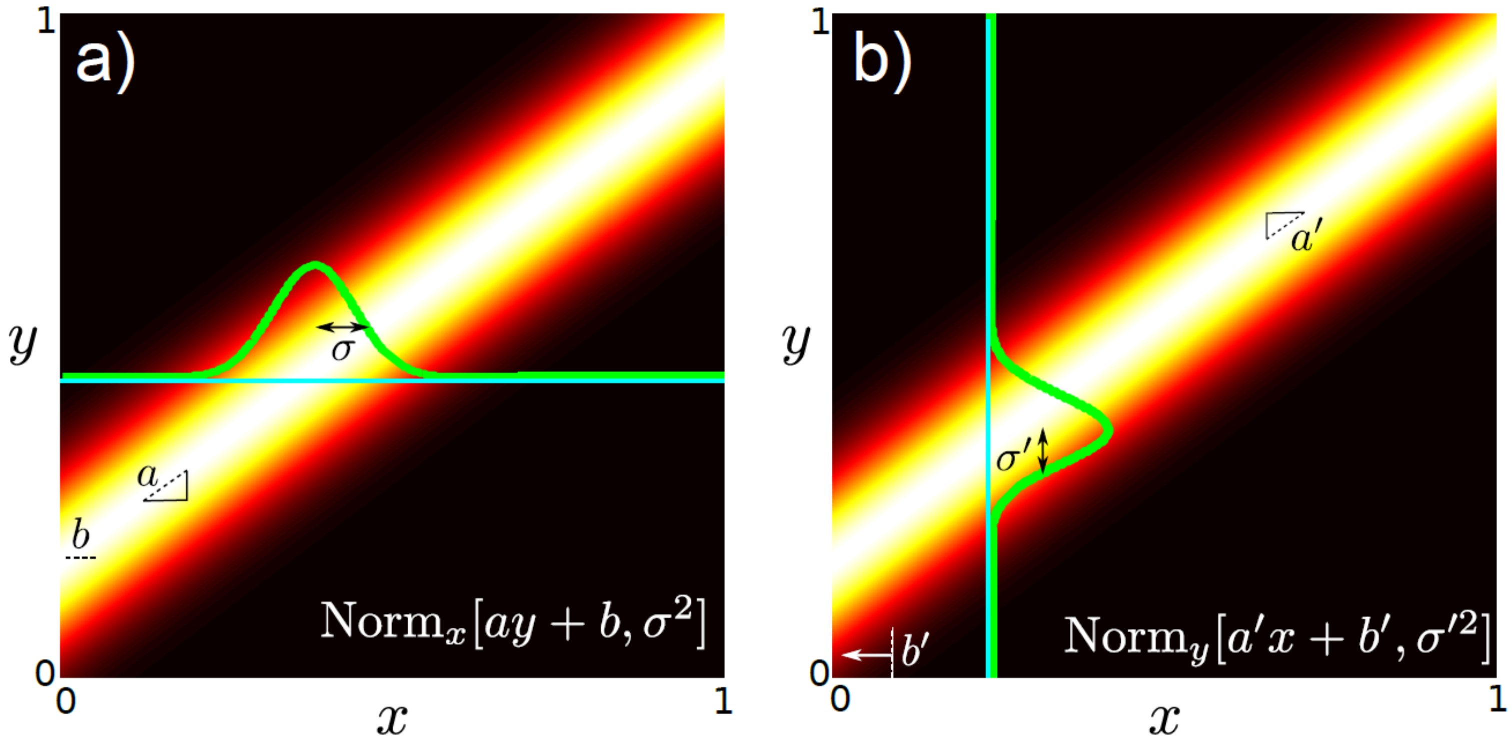
$$\text{Norm}_{\mathbf{x}}[\mathbf{a}, \mathbf{A}]\text{Norm}_{\mathbf{x}}[\mathbf{b}, \mathbf{B}] \quad = $$
$$\kappa \cdot \text{Norm}_{\mathbf{x}}\left[\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}(\mathbf{A}^{-1}\mathbf{a}+\mathbf{B}^{-1}\mathbf{b}), \left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}\right]$$

Amazingly, the constant also has the form of a normal!

$$\kappa = \text{Norm}_{\mathbf{a}}[\mathbf{b}, \mathbf{A}+\mathbf{B}]$$

Change of Variables

If the mean of a normal in **x** is proportional to **y** then this can be re-expressed as a normal in **y** that is proportional to **x**



$$\text{Norm}_{\mathbf{x}}[\mathbf{A}\mathbf{y}+\mathbf{b},\boldsymbol{\Sigma}] = \kappa.\text{Norm}_{\mathbf{y}}[\mathbf{A}'\mathbf{x}+\mathbf{b}',\boldsymbol{\Sigma}']$$

where $\mathbf{A}' = \boldsymbol{\Sigma}'\mathbf{A}^T\boldsymbol{\Sigma}^{-1}$ $\quad \mathbf{b}' = -\boldsymbol{\Sigma}'\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{b}$ $\quad \boldsymbol{\Sigma}' = (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$

Conclusion

- Normal distribution is used ubiquitously in computer vision

- Important properties:

  - Marginal dist. of normal is normal
  - Conditional dist. of normal is normal
  - Product of normals prop. to normal
  - Normal under linear change of variables