# Notes on the Kalman filter

Geoff Gordon

ggordon@cs.cmu.edu

# 1d Gaussian

Usual form of a 1d Gaussian is

$$\mathbb{P}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

*Expectation parameters* are mean $\mu$, variance $\sigma^2$

$\mathbb{P}($variable; parameters$)$ means a probability density function

# Natural parameters

By defining $p = \frac{1}{\sigma^2}$ and $\theta = p\mu$, we have

$$\mathbb{P}(x;\theta,p) = Z(\theta,p)e^{-\frac{p}{2}x^2+\theta x}$$

where the normalizing constant $Z$ is

$$Z(\theta,p) = \sqrt{\frac{p}{2\pi}}\,e^{-\frac{\theta^2}{2p}}$$

$p$ is the *precision*
$\theta, p$ are *natural parameters*

# Operations on Gaussians

Expectation and natural parameters are useful for different purposes

Adding two Gaussians is easy in expectation parameters

"Intersecting" two Gaussians (i.e., multiplying their PDFs and renormalizing—what we need to do in the observation step of a Bayes filter) is easy in natural parameters

# Adding Gaussians

When adding two Gaussians, we just add their means and their variances

If $Z = X + Y$ then

- $\mu_Z = \mu_X + \mu_Y$
- $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$

# Intersecting Gaussians

When intersecting two Gaussians, we just add their $\theta$s and precisions

If $\mathbb{P}Z = \alpha\mathbb{P}X\mathbb{P}Y$ then

- $\theta_Z = \theta_X + \theta_Y$
- $p_Z = p_X + p_Y$

# An example

Suppose the robot starts at $x = 0 \pm 5cm$ then moves right $2m \pm 20cm$

Then it ends up at $x = 2m \pm \sqrt{425}cm$

The corresponding natural parameters are $p = \frac{1}{425} \approx 0.00235$ and $\theta = \frac{2}{425} \approx 0.00471$

# Example cont'd

Now suppose that we observe $x = 2.5m \pm 20cm$ (i.e., $p = 0.0025$ and $\theta = 0.00675$)

Then the final $p$ is $0.0025 + 0.00235 = 0.00485$ and the final $\theta$ is $0.00675 + 0.00471 = 0.01146$

In expectation parameters, that's $2.36m \pm 14.4cm$

# Multivariate Gaussians

Usual form of a multivariate Gaussian is

$$\mathbb{P}(x; \mu, \sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)}$$

Here the random variable $x$ and its mean $\mu$ are both $d$-dimensional vectors

$\Sigma$ is a $d \times d$ covariance matrix

$|\cdot|$ is the determinant of a matrix (i.e., the product of its eigenvalues)

# Natural parameters

Let $P = \Sigma^{-1}$ and $\theta = P\mu$

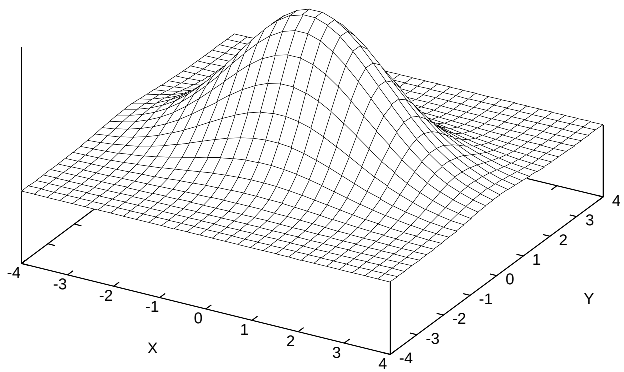
Then we can rewrite

$$\mathbb{P}(x;\theta,P) = Z(\theta,p)e^{-\frac{1}{2}x^{\mathrm{T}}Px+\theta x}$$

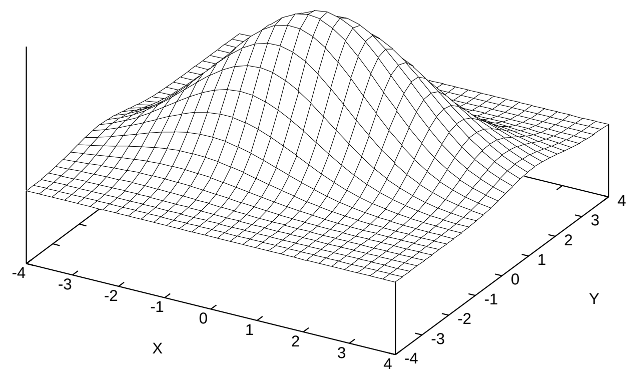
where the normalizing constant $Z$ is

$$Z(\theta,P) = \sqrt{\left|\frac{1}{2\pi}P\right|}\,e^{-\frac{1}{2}\theta^{\mathrm{T}}P^{-1}\theta}$$

# Multivariate Gaussian examples




$$P = \begin{pmatrix} .4 & 0 \\ 0 & 1 \end{pmatrix}$$




$$P = \begin{pmatrix} .4 & -.3 \\ -.3 & 1 \end{pmatrix}$$

Diagonal elements of $\Sigma$ are variances of $x_i$s

$i, j$ element of $\Sigma$ is covariance of $x_i$ and $x_j$

# Operations on multivariate Gaussians

Just as before, when adding two Gaussians, the means add and the variances add

Similarly, when intersecting two Gaussians, the $\theta$s add and the precisions add

# Low-rank updates

Often our observation only gives us information along some directions in state space

We might keep track of many landmarks, but only observe a few on each time step

Similarly, our motion model may only change a few pieces of information

Odometry updates only change our robot's state, not the state of the landmarks

In either case, we save computation

# The Sherman-Morrison formula

$$(A + uv^{\mathrm{T}})^{-1} = A^{-1} - \frac{1}{1 + v^{\mathrm{T}} A^{-1} u} A^{-1} u v^{\mathrm{T}} A^{-1}$$

Rank-1 update to $A$ yields rank-1 update to $A^{-1}$

E.g., if $A$ is precision, $v = u$ and either represents direction along which we measured state

So, with low-rank motions or observations, we can update both precision and variance efficiently

# Sherman-Morrison savings

Inverting is $O(d^3)$ (or slightly less if we get tricky)

SM update is $O(d^2)$ (or less if $u$ or $v$ is sparse)

Preview: in (useful!) special cases, can get down to $O(\ln d)$—allows KFs with 10K–100K dimensions!