Lecture 27 - State Estimation & Kalman Filtering

Last Time: Probability overview

- Multivariate Gaussians
- Conditional Probability

Today:

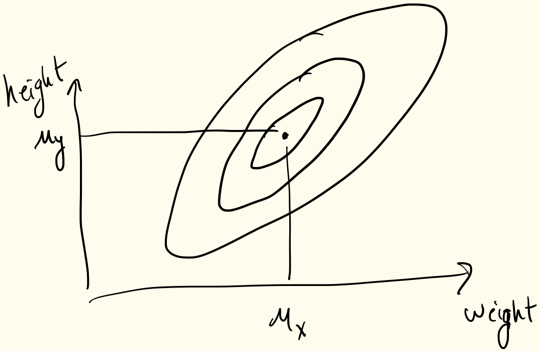
- Optimal Estimation in the LQ case
    - Kalman Filtering
- Estimation for legged robots

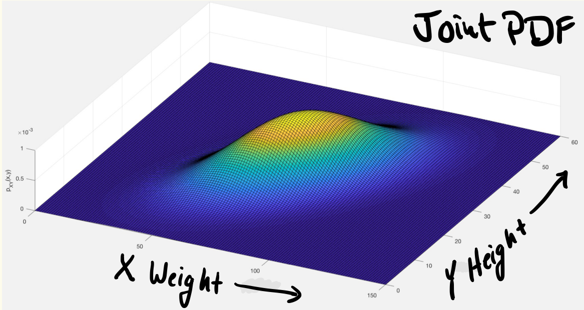
Multi-Dim gaussians - Non independent

REVIEW

Example: Pick a random person @ ND

X = their weight Y = their height





- Let $\underline{z} = \begin{bmatrix} X \\ Y \end{bmatrix}$

- We denote the co-variance of $\underline{z}$ by the 2x2 matrix

$$\Sigma = E\left[(\underline{z}-E[\underline{z}])(\underline{z}-E[\underline{z}])^T\right] = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho \\ \sigma_x\sigma_y\rho & \sigma_y^2 \end{bmatrix}$$

$-1 \le \rho \le 1$

Correlation coefficient

Properties of Multi-D Gaussians

REVIEW

Assume:

$$\underline{X} \sim \mathcal{N}(\mu, \Sigma)$$

$$\underline{Y} \sim \mathcal{N}(r, \Omega)$$

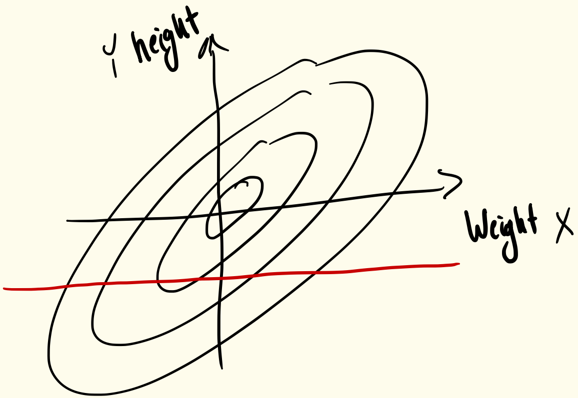
① If X, Y independent and same size

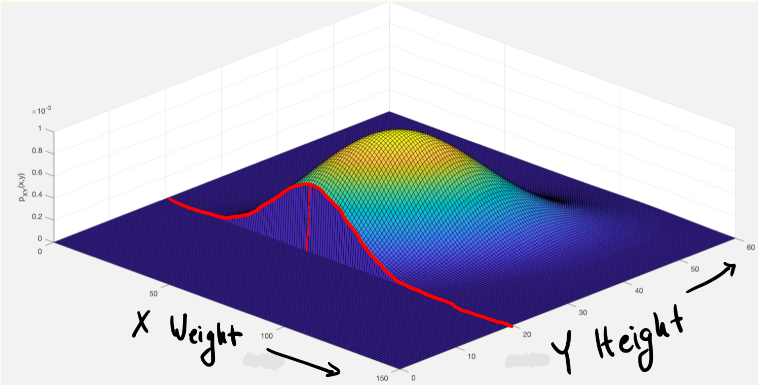
$$\underline{X} + \underline{Y} \sim \mathcal{N}(\mu + r, \Sigma + \Omega)$$

② $$A\underline{X} + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$$

Conditional Distributions: If you gain information, how does it affect uncertainty for other variables **REVIEW**

Example: Suppose height and weight are jointly Normal. If you find out someone's weight, then that affects narrows down your uncertainty in their height...





Conditional Probability:

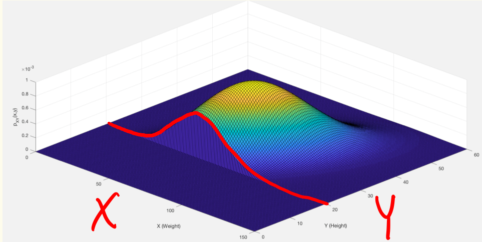
REVIEW

- Let $p_{\underline{X}|\underline{Y}}(x|y)$ denote the PDF of $\underline{X}$ under the restriction that $\underline{Y} = y$.

- Conditional Gaussians

$$\begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$$

$$\Sigma_{xy} = \Sigma_{yx}^T$$

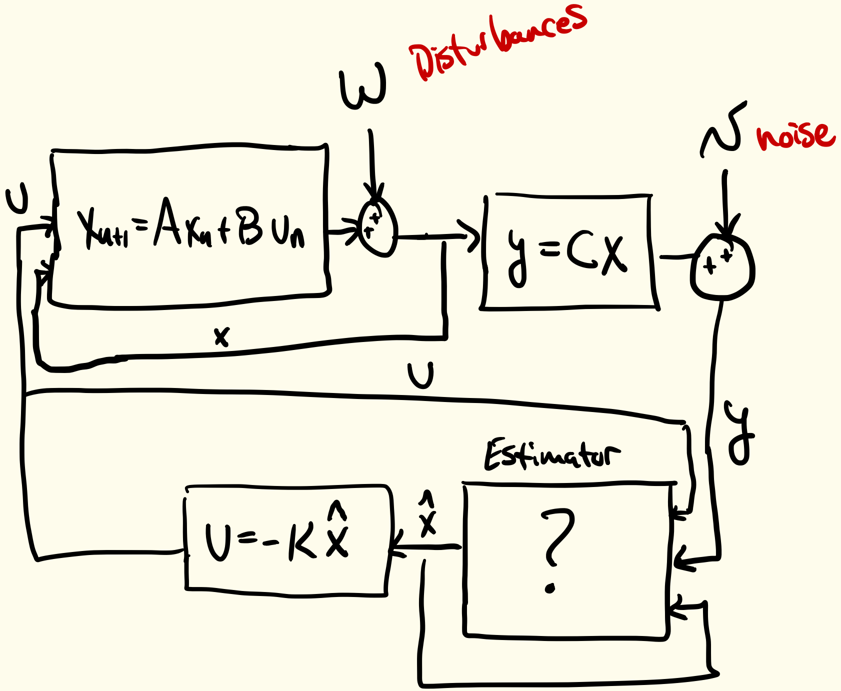


reduction in uncertainty

$$p_{\underline{X}|\underline{Y}}(x|y) \sim \mathcal{N}\left(\mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y), \; \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \right)$$

$$\Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \succeq 0$$

High Level



w
Disturbances
v
noise
u
$x_{n+1} = Ax_n + Bu_n$
$y = Cx$
x
u
y
Estimator
?
$\hat{x}$
$u = -K\hat{x}$

Preliminaries

Process Model

$$\underline{x}_{k+1} = A\underline{x}_k + Bu_k + \underline{w}_k$$

$\underline{w}_k, \underline{v}_{k+1}, \ldots, \underline{w}_0, \underline{v}_1$ mutually independent

$$\underline{w}_k \sim \mathcal{N}(0, Q) \qquad Q \succeq 0$$

Measurement Model

$$y_{k+1} = C x_{k+1} + \underline{v}_{k+1}$$

$$\underline{v}_{k+1} \sim \mathcal{N}(0, R) \qquad R \succ 0$$

Initial State

$$\underline{x}_0 \sim \mathcal{N}(\hat{x}_0, P_{0|0})$$

uncertainty in x_0 given 0 measurements

Definitions:

$$p(x_{k+1|k}) := p(x_{k+1} \mid y_1, \ldots, y_k) \qquad \underline{x}_{k+1|k} \sim \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k})$$

$$p(x_{k+1|k+1}) : p(x_{k+1} \mid y_1, \ldots, y_{k+1}) \qquad \underline{x}_{k+1|k+1} \sim \mathcal{N}(\hat{x}_{k+1|k+1}, P_{k+1|k+1})$$

Analyzing $\underline{x}_{k+1|k}$, $\underline{y}_{k+1|k}$:

$$\underline{x}_{k+1|k} = A \underbrace{\underline{x}_{k|k}}_{\mathcal{N}(\hat{x}_{k|k},\, P_{k|k})} + B \underbrace{u_k}_{\text{known}} + \underbrace{\underline{w}_k}_{\mathcal{N}(0,\, Q)} \sim \mathcal{N}\left(A\hat{x}_{k|k} + Bu_k,\ \underbrace{AP_{k|k}A^T + Q}_{P_{k+1|k}}\right)$$

$$\underline{y}_{k+1|k} = C\underline{x}_{k+1|k} + v_{k+1} \sim \mathcal{N}\left(C\hat{x}_{k+1|k},\ C P_{k+1|k} C^T + R\right)$$

Cross Variance:

$$\Sigma_{xy} = E\left[\left(\underline{x}_{k+1|k} - \hat{x}_{k+1|k}\right)\left(\underline{y}_{k+1|k} - C\hat{x}_{k+1|k}\right)^T\right]$$

$$= E\left[\left(\underline{x}_{k+1|k} - \hat{x}_{k+1|k}\right)\left(C\underline{x}_{k+1|k} + \cancel{v_{k+1}} - C\hat{x}_{k+1|k}\right)^T\right]$$

$$= E\left[\left(\underline{x}_{k+1|k} - \hat{x}_{k+1|k}\right)\left(\underline{x}_{k+1|k} - \hat{x}_{k+1|k}\right)^T C^T\right]$$

Joint PDF:

$$\begin{bmatrix} \underline{x}_{k+1|k} \\ \underline{y}_{k+1|k} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A\hat{x}_{k|k} + Bu_k \\ C\hat{x}_{k+1|k} \end{bmatrix}, \begin{bmatrix} P_{k+1|k} & P_{k+1|k}C^T \\ C P_{k+1|k} & \underbrace{CP_{k+1|k}C^T + R}_{\Sigma_{YY}} \end{bmatrix} \right)$$

what happens if $\underbrace{\underline{y}_{k+1}}_{RV}$ is sampled as $\underbrace{z_{k+1}}_{\text{Numbers off of sensor}}$

$$\underline{x}_{k+1|k+1} \sim \mathcal{N}\left(\hat{x}_{k+1|k} + \Sigma_{XY}\Sigma_{YY}^{-1}(z_{k+1} - C\hat{x}_{k+1|k}), \underbrace{P_{k+1|k} - P_{k+1|k}C^T\Sigma_{yy}^{-1}CP_{k+1|k}}_{P_{k+1|k+1}} \right)$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + \Sigma_{XY}\Sigma_{YY}^{-1}(z_{k+1} - C\hat{x}_{k+1|k})$$

Algorithm

Given $\hat{x}_0$, $P_{0|0}$

for $i = 1, \dots$

$$\hat{x}_{i|i-1} = A\hat{x}_{i-1|i-1} + Bu_{i-1}$$

$$P_{i|i-1} = A P_{i-1|i-1} A^T + Q$$

z_i = Measurement of y (noisy)

$$\hat{x}_{i|i} = \hat{x}_{i|i-1} + \underbrace{P_{i|i-1} C^T (C P_{i|i-1} C^T + R)^{-1}}_{K_i} (z_i - C\hat{x}_{i|i-1})$$

$$P_{i|i} = P_{i|i-1} - P_{i|i-1} C^T (C P_{i|i-1} C^T + R)^{-1} C P_{i|i-1}$$

end

Evolution of $P_{i+1|i}$ follows a riccati equation!

$$P_{i+1|i} = Q + A\left[P_{i|i-1} - P_{i|i-1} C^T (C P_{i|i-1} C^T + R)^{-1} C P_{i|i-1}\right] A^T$$

Aside: Why does this problem have commonality with LQR?

Consider the problem:

$$\max_{x_1} P_{x_1|y_1}(x_1 | y_1 = z_1)$$

$v_0 = 0$ for simplicity

$$\underline{x}_1 = A\underline{x}_0 + \underline{w}_0$$
$$\underline{y}_1 = C\underline{x}_1 + \underline{v}_1$$

$$\underline{x}_0 \sim \mathcal{N}(\hat{x}_0, \Sigma_0)$$
$$\underline{w}_0 \sim \mathcal{N}(0, Q) \qquad \underline{v}_0 \sim \mathcal{N}(0, R)$$

This problem is the same as:

(Find the maximum likelihood set of events that led to $y_1 = z_1$)

$$\max_{x_1, x_0, v_1, w_0} P_{\underline{x}_0}(x_0)\, P_{\underline{w}_0}(w_0)\, P_{\underline{v}_1}(v_1)$$
$$\text{s.t. } x_1 = A x_0 + w_0$$
$$z_1 = C x_1 + v_1$$

$\Rightarrow$ Since log is monotonic

$$\max_{x_1, x_0, v_1, w_0} \log(P_{\underline{x}_0}(x_0)) + \log(P_{\underline{w}_0}(w_0)) + \log(P_{\underline{v}_1}(v_1))$$
$$\text{s.t. } x_1 = A x_0 + w_0$$
$$z_1 = C x_1 + v_1$$

Using the PDF of a guassian:

$$\min_{x_0, x_1, w_0, v_1} (x_0 - \hat{x}_0)^T \Sigma_0^{-1} (x_0 - \hat{x}_0) + w_0^T Q^{-1} w_0 + v_1^T R^{-1} v_1 + \text{const.}$$
$$\text{s.t. } x_1 = A x_0 + w_0$$
$$z_1 = C x_1 + v_0$$

Measures how unlikely the outcome $\underline{w}_0 = w_0$ is

Appying FONC for optimality Results in optimal x_1^* the same as $\hat{x}_{1|1}$ from Kalman filter!
(Lots of algebra needed)

Separation Principle:

Consider the linear system

$$x_{n+1} = A x_n + B u_n + w$$

$$y_{n+1} = C x_{n+1} + v$$

- Design an estimator & controller such that

$$E\left[\sum_{k=0}^{N-1} \left(x_k^T Q x_k + u_k^T R_k u_k \right) + x_N^T P_N x_n \right]$$

is minimal

- Remarkable solution
  - Optimal controller is LQR
  - Optimal observer is Kalman Filter

What about nonlinear systems?

$$x_{n+1} = f(x_n, u_n) + w_n$$

$$y_{n+1} = h(x_{n+1}) + v_{n+1}$$

Caveats

1. Separation principle doesn't hold :(
2. If x_0 is gaussian, x_1 need not be
3. Practical solution: Use linearizations to approximate distributions
   $\Rightarrow$ Extended Kalman Filter (EKF)

Algorithm: Extended Kalman Filter

Given $\hat{x}_0$, $P_{0|0}$

for $i = 1, \dots$

Use nonlinear model to push forward $\hat{x}$

$$\hat{x}_{i|i-1} = f(x_{i-1|i-1}, u_{i-1}) \qquad A_{i-1} := \left.\frac{\partial f}{\partial x}\right|_{x_{i-1|i-1}, u_{i-1}} \qquad C_i := \left.\frac{\partial h}{\partial x}\right|_{x_{i|i-1}}$$

$$P_{i|i-1} = A_{i-1} P_{i-1|i-1} A_{i-1}^T + Q$$

Linearized model to push forward the covarince

z_i = Measurement of y (noisy)

$$\hat{x}_{i|i} = \hat{x}_{i|i-1} + \underbrace{P_{i|i-1} C_i^T (C_i P_{i|i-1} C_i^T + R)^{-1}}_{K_i} (z_i - h(\hat{x}_{i|i-1}))$$

$$P_{i|i} = P_{i|i-1} - P_{i|i-1} C_i^T (C_i P_{i|i-1} C_i^T + R)^{-1} C_i P_{i|i-1}$$

end

See notes by Burgard & Thrun for further detail

State Estimation For Legged Systems

$$q = \begin{bmatrix} q_B \\ q_J \end{bmatrix}$$

q_B ← Difficult to sense

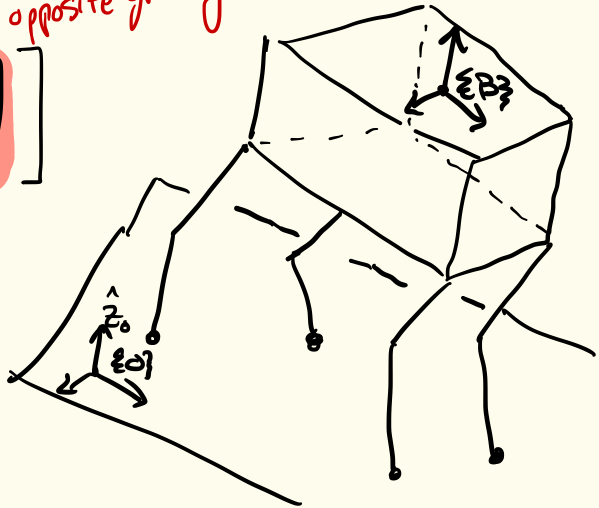
q_J ← Easy to sense

Sensors

1. Encoders for Legs
2. Gyro for body ${}^{B}\omega_B$
3. Accelerometer $a = {}^{B}R_0 \left[{}^{0}\ddot{p}_B + \begin{bmatrix} 0 \\ 0 \\ 9.81\, m/s^2 \end{bmatrix} \right]$

Biased opposite gravity

4. Vision (sometimes)



State Estimation For Legged Systems

$$q_B = \begin{bmatrix} {}^oP_B \\ {}^oR_B \end{bmatrix} \qquad \dot{q}_B = \begin{bmatrix} {}^o\dot{p}_B \\ {}^B\omega_B \end{bmatrix}$$

Assume known

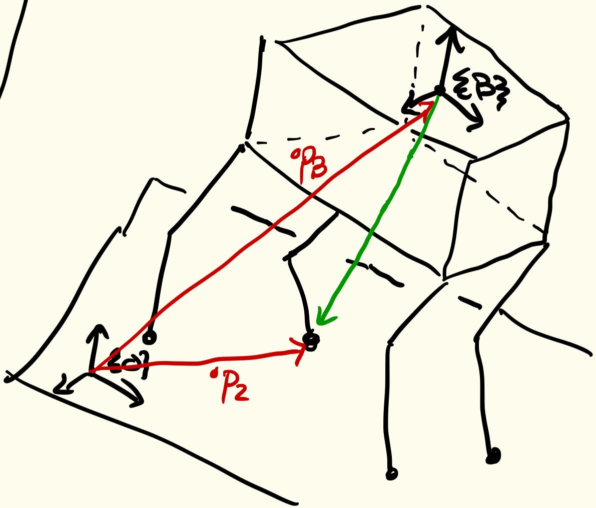
State:

$$X = \begin{bmatrix} {}^oP_B \\ {}^o\dot{P}_B \\ {}^oP_1 \\ {}^oP_2 \\ {}^oP_3 \\ {}^oP_4 \end{bmatrix}$$

Dynamics

Treat acceleration as input u_n

$$X_{n+1} = \begin{bmatrix} {}^oP_B + \Delta t\, {}^o\dot{p}_B \\ {}^o\dot{p}_B + \Delta t \left[{}^oR_b\, a - \begin{bmatrix} 0 \\ 0 \\ 9.81\,m/s^2 \end{bmatrix} \right] \\ {}^oP_1 \\ {}^oP_2 \\ {}^oP_3 \\ {}^oP_4 \end{bmatrix}$$



Dynamics can be written in the form:

$$X_{n+1} = A x_n + B u_n + w_n$$

User tuned process noise

Measurement Model

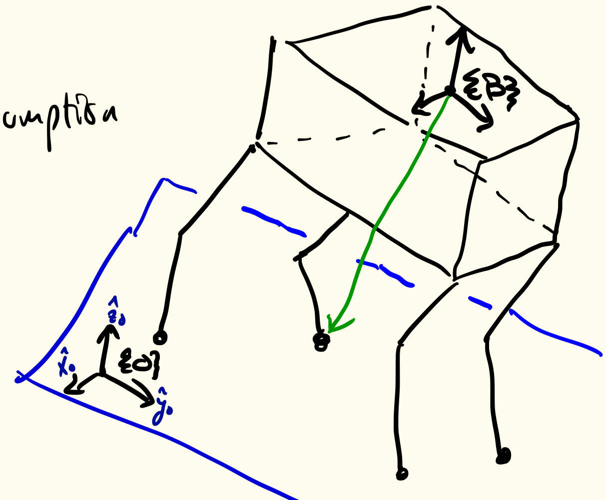
$$
y = \begin{bmatrix} {}^{0}p_1 - {}^{0}p_B \\ {}^{0}p_2 - {}^{0}p_B \\ \vdots \\ {}^{0}p_4 - {}^{0}p_B \\ \hline [0\;0\;1]\,{}^{0}p_1 \\ \vdots \\ [0\;0\;1]\,{}^{0}p_4 \end{bmatrix}
$$

Position of feet relative to body ("measured" by computing forward kinematics from encoders)

Height of feet (Not really "measured" but assumed at $z=0$ or other value if vision is available)

$$
= Cx + v
$$

user-tuned uncertainty in measurements & height assumption



Overall

- This setup enables Kalman Machinery to estimate ${}^{0}p_B$, ${}^{0}\dot{p}_B$
- Extended Kalman Filter can be used to estimate ${}^{0}R_B$ and ${}^{B}\omega_B$

Overall:

Modeling - Continuous Dynamics & Impacts

Optimal Control - LQR, HJB, Pontryagin

Trajectory optimization

Locomotion - Templates, MPC, Limit cycles

Task-Space control

State Estimation - Kalman Filtering