78.5

State Estimation for Robotics. Midterm

TU Berlin WS 2021/22.
Chair: Robotic Interactive Perception.
December 16, 2021

Student name: ____________

Student's TU Berlin ID: ____________

"Closed books" (no slides, no cheat-sheet, no computer, no smartphone).

Honor code: I hereby certify that I have not given or received aid in the examination: ________ (Sign your name).

1. (5 pts) Assume a joint Gaussian density over a pair of variables $\mathbf{x}, \mathbf{y}$. If the variables are uncorrelated, which of the following is false?

   (a) $E[\mathbf{x}\mathbf{y}^\top] = E[\mathbf{x}]E[\mathbf{y}^\top]$ (the 2nd moment factorizes as the product of the means)

   (b) $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ (the joint probability density factorizes into conditional · prior)

   (c) $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ (the joint density factorizes into the product of the marginals)

   (d) $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})$ (the conditionals satisfy this cross-term equality)

2. (5 pts) A random variable $\mathbf{x}$ with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance $\boldsymbol{\Sigma}_\mathbf{xx}$ is passed through a system described by the equation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{ww})$. What is the mean of the output?

   (a) $\boldsymbol{\mu}_\mathbf{y} = \mathbf{H}\boldsymbol{\mu}_\mathbf{x}$

   (b) $\boldsymbol{\mu}_\mathbf{y} = \boldsymbol{\mu}_\mathbf{w} + \mathbf{H}\boldsymbol{\mu}_\mathbf{x}$

   (c) $\boldsymbol{\mu}_\mathbf{y} = \mathbf{H}\Sigma_\mathbf{xx}^{-1}\boldsymbol{\mu}_\mathbf{x} + \boldsymbol{\mu}_\mathbf{w}$

   (d) $\boldsymbol{\mu}_\mathbf{y} = \mathbf{H}\Sigma_\mathbf{xx}^{-1}\boldsymbol{\mu}_\mathbf{x} + \Sigma_\mathbf{ww}^{-1}\boldsymbol{\mu}_\mathbf{w}$

3. (5 pts) A random variable $\mathbf{x}$ with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance $\boldsymbol{\Sigma}_\mathbf{xx}$ is passed through a system described by the equation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{ww})$ is uncorrelated with $\mathbf{x}$. What is the covariance of the output?

   (a) $\boldsymbol{\Sigma}_\mathbf{yy} = \mathbf{H}^\top\boldsymbol{\Sigma}_\mathbf{xx}\mathbf{H}$

   (b) $\boldsymbol{\Sigma}_\mathbf{yy} = \mathbf{H}\boldsymbol{\Sigma}_\mathbf{xx}\mathbf{H}^\top$

   (c) $\boldsymbol{\Sigma}_\mathbf{yy} = \mathbf{H}^\top\boldsymbol{\Sigma}_\mathbf{xx}\mathbf{H} + \boldsymbol{\Sigma}_\mathbf{ww}$

   (d) $\boldsymbol{\Sigma}_\mathbf{yy} = \mathbf{H}\boldsymbol{\Sigma}_\mathbf{xx}\mathbf{H}^\top + \boldsymbol{\Sigma}_\mathbf{ww}$

4. (5 pts) Match. Join by lines all that apply. Please be clear or it will not count:

$p(x)$	posterior density
~~$p(y)$~~	sensor model
$p(x\|y)$	prior density
$p(y\|x)$	~~motion model~~

5. (5 pts) Which option of the following ones describes the information fusion of two Gaussian densities $\{\mathcal{N}(\mu_i, \Sigma_i)\}_{i=1}^{2}$ in a Bayesian manner to produce another Gaussian density $\mathcal{N}(\mu, \Sigma)$?

   (a) $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$

   (b) $\sigma^2\mu = \sigma_1^2\mu_1 + \sigma_2^2\mu_2$ and $1/\sigma^2 = 1/\sigma_1^2 + 1/\sigma_2^2$

   (c) $\sigma^2\mu = \sigma_1^2\mu_1 + \sigma_2^2\mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$

   (d) $\mu/\sigma^2 = \mu_1/\sigma_1^2 + \mu_2/\sigma_2^2$ and $1/\sigma^2 = 1/\sigma_1^2 + 1/\sigma_2^2$

6. (5 pts) An estimator is unbiased when ...

   (a) the true value of the unknown variable that is being estimated coincides with the expected value of the estimator.

   (b) the mean square value of the estimator coincides with the true mean square value of the unknown variable that is being estimated.

   (c) the expected value of the estimator and the true value of the unknown variable that is being estimated are less than 3σ away.

   (d) the expected value of the estimator is zero.

7. (5 pts) An estimator is consistent when...

   (a) the true value of the unknown variable that is being estimated coincides with the expected value of the estimator.

   (b) the mean square value of the estimator coincides with the true mean square value of the unknown variable that is being estimated.

   (c) the expected value of the estimator and the true value of the unknown variable that is being estimated are less than 1σ away.

   (d) the true uncertainty in the system is perfectly modelled by the estimated covariance.

8. (5 pts) Which of the following filters is a MAP estimator?

   (a) The EKF (Extended Kalman Filter)

   (b) The IEKF (Iterated EKF)

   (c) The SPKF (sigmapoints Kalman Filter)

   (d) The ISPKF (Iterated SPKF)

9. (5 pts) What assumptions does the particle filter make on the non-linear models of the system dynamics, the observation equation and the state?

   (a) The state distribution is Gaussian and the models are differentiable.

   (b) The state distribution can adopt any shape but the models need be differentiable.

   (c) The state distribution is Gaussian and the models need not be differentiable.

   (d) The state distribution can adopt any shape and the models need not be differentiable.

10. (5 pts) What assumptions does the generalized Gaussian filter make on the non-linear models of the system dynamics, the observation equation and the state?

   (a) The state distribution is Gaussian and the models are differentiable.

   (b) The state distribution can adopt any shape but the models need be differentiable.

   (c) The state distribution is Gaussian and the models need not be differentiable.

   (d) The state distribution can adopt any shape and the models need not be differentiable.

11. (10 pts) One way to obtain the Kalman gain is via gain optimization. If the error in the state estimate is $\hat{\mathbf{e}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$, then the covariance of $\hat{\mathbf{e}}_k$ is $E[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^\top] = (1 - \mathbf{K}_k \mathbf{C}_k)\check{\mathbf{P}}_k(1 - \mathbf{K}_k \mathbf{C}_k)^\top + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^\top$, and a cost function to quantify the uncertainty of $\hat{\mathbf{e}}_k$ is defined:

$$J(\mathbf{K}_k) \doteq \text{trace } E[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^\top].$$

Show how J is related to the Mean Square value of $\hat{\mathbf{e}}_k$ (MSE). Please specify and justify (with text) all steps carried out. Example: in this step, I apply property X of $E[\cdot]$.

$\text{MSE}(\hat{e}_k) = E[\hat{e}_k^2] = E[\hat{e}_k^T \cdot \hat{e}_k]$

a few steps missing here. -3

- from matrix cookbook apply trace:

$E[\text{trace}(\hat{e}_k \cdot \hat{e}_k^T)] \xrightarrow{\text{linearity}} \text{trace}(E[\hat{e}_k \cdot \hat{e}_k^T])$

↳ the last term is like the equation above (in problem statement)

$J(K_k) = \text{trace } E[\hat{e}_k \hat{e}_k^T]$

12. (10 pts) The correction step of the generalized Gaussian filter uses "Gaussian inference": it assumes a joint Gaussian density $p(\mathbf{x}_k, \mathbf{y}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1})$ for the state and the measurement, and then computes the posterior

$$p(\mathbf{x}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k}) = \mathcal{N}(\underbrace{\boldsymbol{\mu}_{x,k} + \boldsymbol{\Sigma}_{xy,k}\boldsymbol{\Sigma}_{yy,k}^{-1}\overbrace{(\mathbf{y}_k - \boldsymbol{\mu}_{y,k})}^{\text{innovation}}}_{\hat{\mathbf{x}}_k}, \underbrace{\boldsymbol{\Sigma}_{xx,k} - \boldsymbol{\Sigma}_{xy,k}\boldsymbol{\Sigma}_{yy,k}^{-1}\boldsymbol{\Sigma}_{yx,k}}_{\hat{\mathbf{P}}_k}).$$

(a) Which term is the "Kalman gain"?

(b) Which term is the innovation and what does it represent?

(c) Where and how is the *nonlinear* observation model used? (not the linearized one)

(d) Where and how is the *nonlinear* motion model used? (not the linearized one)

✓ a) $K = \Sigma_{xy,k} \Sigma_{yy,k}^{-1}$ in front of $(y_k - \mu_{y,k})$

✗ b) Innovation is the whole calc. for $\hat{x}_k$ (mean of posterior) :

↳ it corrects the predicted mean with the measurement of current timestep -2.5

✓ c) nonlinear observ. model $g(\cdot)$ is used in $(y_k - \mu_{y,k})$ of $\hat{x}_k$ calc.

↳ it transforms the predicted mean $\mu_{\check{x},k}$ into measurement ~~spacbe~~ space :

↳ ~~yk~~ $\mu_{y,k} = g(\mu_{\check{x},k})$

✓ d) nonlin. motion model $f(\cdot)$ is used for the calculation of the predicted mean $\mu_{\check{x},k}$

↳ $\mu_{\check{x},k} = f(\mu_{\hat{x},k-1}; v_k)$

↳ this step happened before the calc of the posterior

13. (10 pts) The Bayes filter adopts the expression

$$\underbrace{p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}\mathbf{y}_{0:k})}_{r} = \eta\, \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{a} \int \underbrace{p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{v}_k)}_{b}\, \underbrace{p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1})}_{c}\, d\mathbf{x}_{k-1}$$

(a) Please provide a name (from the state estimation terminology) for each of the terms a, b, c and r.

(b) Pleas explain how a and b are related to the state equations described by functions **f** and **g**.

(c) Why is it said that the Bayes filter has a predictor-corrector form? Please explain the steps of the method. You may draw a block diagram to help you explain.

a) r: posterior probability distribution for $\underbrace{\text{state}}_{\text{estimate}}$ x at time k given all past inputs and measurements.

↳ also called: belief in the state ~~x~~ x_k

a: measurement model

b: system dynamics

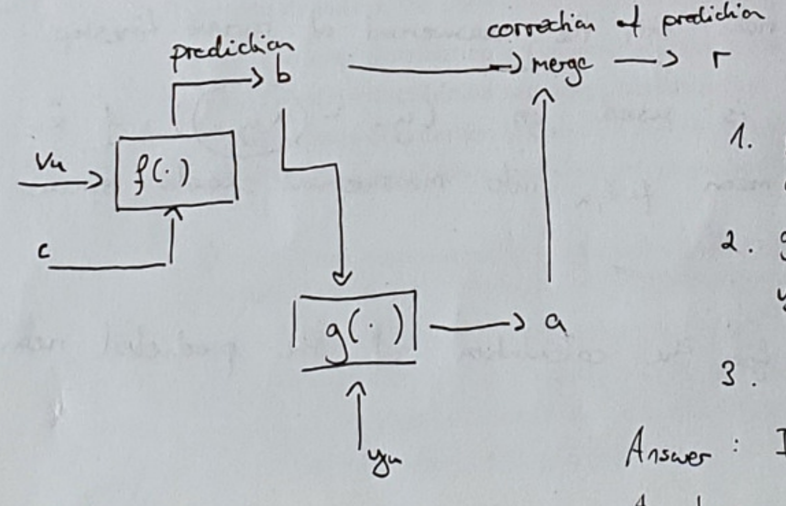
c: prior belief (posterior of previous estimate x_{k-1})

b) f(·) is the system dynamics function / motion model in general state space.

b represents this/applies this dynamics by forwarding the previous estimate with help of the input through f(·) to generate the ~~prior~~ prediction.

g(·) is the obsefation function that relates the prediction with the measurement space. It is used in ⓐ to have a direct measure on how likely the actual measurement is, given the prediction

c)



Steps:

1. get b as prediction with f(·), v_k and ~~x~~ x_{k-1}
2. get a through prediction, g(·) and y_k
3. correct prediction with a

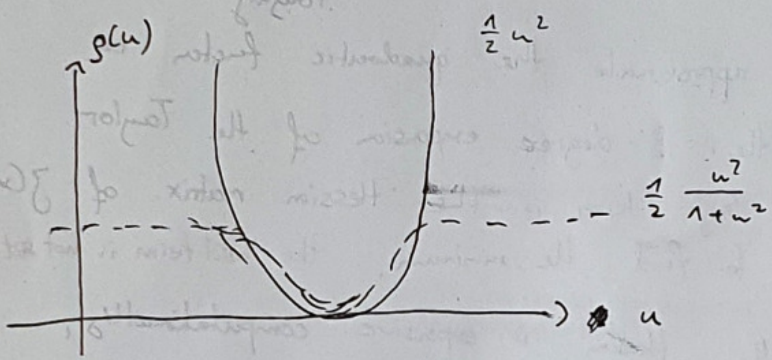
Answer: It is called predictor-corrector because we first predict and then correct with measurement

14. (10 pts) M-estimators work by minimizing the error function $J'(\mathbf{x}) = \sum_i \alpha_i \rho(u_i(\mathbf{x}))$, where $\alpha_i > 0$ is a scalar weight, $u_i(\mathbf{x}) = \sqrt{\mathbf{e}_i^\top(\mathbf{x})\mathbf{W}_i^{-1}\mathbf{e}_i(\mathbf{x})}$, $\mathbf{e}_i$ are the error terms and $\rho(\cdot)$ is some function.

(a) Plot the Geman-McClure function $\rho(u) = \frac{1}{2}\frac{u^2}{1+u^2}$ and the quadratic function $\frac{1}{2}u^2$ (using the same axes, for comparison).

(b) Please explain why the Geman-McClure function is better than the quadratic function in case of noisy observations comprising outliers (Note: outliers are measurements that do not conform to our "Gaussian" noise model). (That is, what characteristics or properties of the function ρ (derivative at zero, shape, differentiability, etc.) confer the robustness with respect to (non-linear) least squares?

a)



Values? -3

b) In the normal quadratic function, outliers are given a lot of weight (with quadratic growth) in the cost function, shifting the optimal estimate into the wrong direction.
The ~~German~~ Geman - Mc Clure function constraints itself in penalting the outliers that deviates from our measurement distribution

?? Not clear what you mean

-2

15. (10 pts) Please explain Gauss-Newton's optimization method for a non-linear least squares (NLLS) objective function $J(\mathbf{x}) = \frac{1}{2}\|\mathbf{u}(\mathbf{x})\|^2$. We have seen two ways to explain it; only one is needed to answer, but please be clear about the approximation/derivation process and about what the steps of the method are to minimize $J(\mathbf{x})$.

- Gauss - Newton: Approximate a cost-function $J(x)$ by a quadratic function on the current operating point x_{op} and directly jump to the minimum. Repeat these steps until the operating point coincides with the found minimum.
  roughly

  In normal Newton we approximate the quadratic function on x_{op} by taking ~~3 Taylor~~ the 3-degree expansion of the Taylor Series of $J(x)$. The third Taylor term is the Hessian matrix of $J(x)$ which will be set to zero to find the minimum. the third term is not set to zero

  However, since calculating the Hessian is expensive computationally, Gauss - Newton only approximates the Hessian. how?

  Need to be able to write it with math −6
  and /or plots.

(2) $y = Hx + w$

$$E[y] = H\,E[x] + E[w] = H\mu_x + \mu_w$$

(3) $E[(\cancel{\mu_y}\ (y-\mu_y)(y-\mu_y)^T = E(Hx + w - H\mu_x - \mu_w)(Hx + w - \cancel{H\mu}\ H\mu_x - \mu_w)^T)$

$$\Leftrightarrow E(H(x + \frac{w}{H} - \mu_x - \frac{\mu_w}{H})(x + \frac{w}{H^T} - \mu_x - \frac{\mu_w}{H^T})^T H^T)$$

$$\Leftrightarrow H\,\cancel{E}\,E[((x-\mu_x) + (\frac{w}{H} - \frac{\mu_w}{H}))((x-\mu_x) + (\frac{w}{H^T} - \frac{\mu_w}{H^T})^T)$$

$$\Leftrightarrow H\,E[\underbrace{(x-\mu_x)^{\cancel{T}2}}_{\Sigma_{xx}} + \underbrace{2(\frac{w}{H} - \frac{\mu_w}{H})(x-\mu_x)}_{\text{statist. independent} \; = 0} + \underbrace{(\frac{w}{H^T} - \frac{\mu_w}{H^T})^2}_{\frac{\Sigma_{ww}}{H^T}}])H^T$$

$$\Leftrightarrow H\Sigma_{xx}H^T + \Sigma_{ww}$$

(11) MSE = ~~uncentralized 2nd moment~~ $E[\hat{e}_n^2]$

$= E[\hat{e}_n^T \cdot \hat{e}_n]$