Aalto University
School of Electrical
Engineering

Statistical linearization and statistical linear regression

Simo Särkkä

Aalto University, Finland

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Linear approximation of a non-linearity

- We wish to approximate the transformation

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- We choose the approximation to be a linear function, say

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}.$$

- One option is to use Taylor series centered at $\mathbf{x} = \mathbf{m}$, which leads to

$$\mathbf{A} = \mathbf{G}_x(\mathbf{m}),$$
$$\mathbf{b} = \mathbf{g}(\mathbf{m}).$$

- Leads to the extended Kalman filter (and smoother).
- But this is not the only option to choose **A** and **b**.

Linear approximation of a non-linearity

- We wish to approximate the transformation

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- We choose the approximation to be a linear function, say

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}.$$

- One option is to use Taylor series centered at $\mathbf{x} = \mathbf{m}$, which leads to

$$\mathbf{A} = \mathbf{G}_x(\mathbf{m}),$$
$$\mathbf{b} = \mathbf{g}(\mathbf{m}).$$

- Leads to the extended Kalman filter (and smoother).
- But this is not the only option to choose $\mathbf{A}$ and $\mathbf{b}$.

Linear approximation of a non-linearity

- We wish to approximate the transformation

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- We choose the approximation to be a linear function, say

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}.$$

- One option is to use Taylor series centered at $\mathbf{x} = \mathbf{m}$, which leads to

$$\mathbf{A} = \mathbf{G}_x(\mathbf{m}),$$
$$\mathbf{b} = \mathbf{g}(\mathbf{m}).$$

- Leads to the extended Kalman filter (and smoother).
- But this is not the only option to choose $\mathbf{A}$ and $\mathbf{b}$.

Linear approximation of a non-linearity

- We wish to approximate the transformation

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- We choose the approximation to be a linear function, say

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}.$$

- One option is to use Taylor series centered at $\mathbf{x} = \mathbf{m}$, which leads to

$$\mathbf{A} = \mathbf{G}_x(\mathbf{m}),$$
$$\mathbf{b} = \mathbf{g}(\mathbf{m}).$$

- Leads to the extended Kalman filter (and smoother).
- But this is not the only option to choose $\mathbf{A}$ and $\mathbf{b}$.

Linear approximation of a non-linearity

- We wish to approximate the transformation

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- We choose the approximation to be a linear function, say

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}.$$

- One option is to use Taylor series centered at $\mathbf{x} = \mathbf{m}$, which leads to

$$\mathbf{A} = \mathbf{G}_x(\mathbf{m}),$$
$$\mathbf{b} = \mathbf{g}(\mathbf{m}).$$

- Leads to the extended Kalman filter (and smoother).
- But this is not the only option to choose $\mathbf{A}$ and $\mathbf{b}$.

Statistical linearization

- Instead, consider minimization of the error

$$\mathrm{MSE}(\mathbf{A},\mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right]$$

w.r.t. some $\mathbf{x} \sim p(\mathbf{x})$ with $\mathrm{E}[\mathbf{x}] = \mathbf{m}$ and $\mathrm{Cov}[\mathbf{x}] = \mathbf{P}$.

- We get

$$\begin{aligned}
&\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x}-\mathbf{m}) - \mathbf{b}\|^2 \\
&= \mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x}) - 2\,\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x}-\mathbf{m}) + \mathrm{tr}\left\{\mathbf{A}\,(\mathbf{x}-\mathbf{m})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}\,\mathbf{A}^\mathsf{T}\right\} \\
&\qquad + 2\,(\mathbf{x}-\mathbf{m})^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{b} - 2\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

- Taking expectation gives

$$\begin{aligned}
&\mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x}-\mathbf{m}) - \mathbf{b}\|^2\right] = \mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x})\right] \\
&- 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x}-\mathbf{m})\right] + \mathrm{tr}\left\{\mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T}\right\} - 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\right]\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

Statistical linearization

- Instead, consider minimization of the error
$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right]$$
w.r.t. some $\mathbf{x} \sim p(\mathbf{x})$ with $\mathrm{E}[\mathbf{x}] = \mathbf{m}$ and $\mathrm{Cov}[\mathbf{x}] = \mathbf{P}$.
- We get
$$\begin{aligned}
&\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x} - \mathbf{m}) - \mathbf{b}\|^2 \\
&= \mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x}) - 2\,\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathrm{tr}\left\{\mathbf{A}\,(\mathbf{x} - \mathbf{m})\,(\mathbf{x} - \mathbf{m})^\mathsf{T}\,\mathbf{A}^\mathsf{T}\right\} \\
&\qquad + 2\,(\mathbf{x} - \mathbf{m})^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{b} - 2\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$
- Taking expectation gives
$$\begin{aligned}
&\mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x} - \mathbf{m}) - \mathbf{b}\|^2\right] = \mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x})\right] \\
&- 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x} - \mathbf{m})\right] + \mathrm{tr}\left\{\mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T}\right\} - 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\right]\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

Statistical linearization

- Instead, consider minimization of the error

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right]$$

  w.r.t. some $\mathbf{x} \sim p(\mathbf{x})$ with $\mathrm{E}[\mathbf{x}] = \mathbf{m}$ and $\mathrm{Cov}[\mathbf{x}] = \mathbf{P}$.

- We get

$$\begin{aligned}
&\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x} - \mathbf{m}) - \mathbf{b}\|^2 \\
&= \mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x}) - 2\,\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathrm{tr}\left\{\mathbf{A}\,(\mathbf{x} - \mathbf{m})\,(\mathbf{x} - \mathbf{m})^\mathsf{T}\,\mathbf{A}^\mathsf{T}\right\} \\
&\qquad + 2\,(\mathbf{x} - \mathbf{m})^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{b} - 2\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

- Taking expectation gives

$$\begin{aligned}
&\mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,(\mathbf{x} - \mathbf{m}) - \mathbf{b}\|^2\right] = \mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x})\right] \\
&- 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,(\mathbf{x} - \mathbf{m})\right] + \mathrm{tr}\left\{\mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T}\right\} - 2\,\mathrm{E}\left[\mathbf{g}^\mathsf{T}(\mathbf{x})\right]\,\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

Statistical linearization (cont.)

- Let us set the derivatives w.r.t. $\mathbf{b}$ and $\mathbf{A}$ to zero:

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{b}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right] + 2\,\mathbf{b} = 0,$$
$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{A}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right] + 2\,\mathbf{A}\,\mathbf{P} = 0.$$

- This leads to the optimal linearization parameters

$$\mathbf{b} = \mathrm{E}\left[\mathbf{g}(\mathbf{x})\right],$$
$$\mathbf{A} = \mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1}$$
$$= \mathrm{E}\left[(\mathbf{g}(\mathbf{x})-\mathbf{b})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1}.$$

- Thus the linearization is (with the above $\mathbf{b}$ and $\mathbf{A}$)

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x}-\mathbf{m}) + \mathbf{b}.$$

- This is called classical statistical linearization (SL).

Statistical linearization (cont.)

- Let us set the derivatives w.r.t. $\mathbf{b}$ and $\mathbf{A}$ to zero:

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{b}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right] + 2\,\mathbf{b} = 0,$$

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{A}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right] + 2\,\mathbf{A}\,\mathbf{P} = 0.$$

- This leads to the optimal linearization parameters

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}\left[\mathbf{g}(\mathbf{x})\right], \\
\mathbf{A} &= \mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1} \\
&= \mathrm{E}\left[(\mathbf{g}(\mathbf{x})-\mathbf{b})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1}.
\end{aligned}$$

- Thus the linearization is (with the above $\mathbf{b}$ and $\mathbf{A}$)

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x}-\mathbf{m}) + \mathbf{b}.$$

- This is called classical statistical linearization (SL).

Statistical linearization (cont.)

- Let us set the derivatives w.r.t. $\mathbf{b}$ and $\mathbf{A}$ to zero:

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{b}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right] + 2\,\mathbf{b} = 0,$$

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{A}} = -2\,\mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right] + 2\,\mathbf{A}\,\mathbf{P} = 0.$$

- This leads to the optimal linearization parameters

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}\left[\mathbf{g}(\mathbf{x})\right], \\
\mathbf{A} &= \mathrm{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1} \\
&= \mathrm{E}\left[(\mathbf{g}(\mathbf{x})-\mathbf{b})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\right]\,\mathbf{P}^{-1}.
\end{aligned}$$

- Thus the linearization is (with the above $\mathbf{b}$ and $\mathbf{A}$)

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x}-\mathbf{m}) + \mathbf{b}.$$

- This is called classical statistical linearization (SL).

Statistical linearization (cont.)

- Let us set the derivatives w.r.t. $\mathbf{b}$ and $\mathbf{A}$ to zero:

$$\frac{\partial \text{MSE}}{\partial \mathbf{b}} = -2\,\text{E}\left[\mathbf{g}(\mathbf{x})\right] + 2\,\mathbf{b} = 0,$$
$$\frac{\partial \text{MSE}}{\partial \mathbf{A}} = -2\,\text{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}\right] + 2\,\mathbf{A}\,\mathbf{P} = 0.$$

- This leads to the optimal linearization parameters

$$\begin{aligned}
\mathbf{b} &= \text{E}\left[\mathbf{g}(\mathbf{x})\right], \\
\mathbf{A} &= \text{E}\left[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}\right]\,\mathbf{P}^{-1} \\
&= \text{E}\left[(\mathbf{g}(\mathbf{x})-\mathbf{b})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}\right]\,\mathbf{P}^{-1}.
\end{aligned}$$

- Thus the linearization is (with the above $\mathbf{b}$ and $\mathbf{A}$)

$$\mathbf{y} = \mathbf{A}\,(\mathbf{x}-\mathbf{m}) + \mathbf{b}.$$

- This is called classical statistical linearization (SL).

Statistical linearization: properties

- The expectation of SL is indeed exact:

$$\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}\right]=\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right].$$

- The covariance is not exact:

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})\,(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})^{\mathsf{T}}\right] \\
&=\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\right] \\
&=\mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}\neq\mathrm{Cov}\left[\mathbf{g}(\mathbf{x})\right] \qquad \{\text{ in general }\}.
\end{aligned}$$

- The corresponding SL filter and smoother have the same limitations.
- But we can fix this!

Statistical linearization: properties

- The expectation of SL is indeed exact:

$$\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}\right]=\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right].$$

- The covariance is not exact:

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})\,(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})^{\mathsf{T}}\right]\\
&=\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\right]\\
&=\mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}\neq\mathrm{Cov}\left[\mathbf{g}(\mathbf{x})\right]\qquad\{\text{ in general }\}.
\end{aligned}$$

- The corresponding SL filter and smoother have the same limitations.
- But we can fix this!

Statistical linearization: properties

- The expectation of SL is indeed exact:

$$\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}\right]=\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right].$$

- The covariance is not exact:

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})\,(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})^{\mathsf{T}}\right]\\
&=\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\right]\\
&=\mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}\neq\mathrm{Cov}\left[\mathbf{g}(\mathbf{x})\right]\qquad\{\text{ in general }\}.
\end{aligned}$$

- The corresponding SL filter and smoother have the same limitations.
- But we can fix this!

Statistical linearization: properties

- The expectation of SL is indeed exact:

$$\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}\right]=\mathrm{E}\left[\mathbf{g}(\mathbf{x})\right].$$

- The covariance is not exact:

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})\,(\mathbf{A}\,(\mathbf{x}-\mathbf{m})+\mathbf{b}-\mathbf{b})^{\mathsf{T}}\right]\\
&=\mathrm{E}\left[\mathbf{A}\,(\mathbf{x}-\mathbf{m})\,(\mathbf{x}-\mathbf{m})^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\right]\\
&=\mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}\neq\mathrm{Cov}\left[\mathbf{g}(\mathbf{x})\right]\qquad\{\text{ in general }\}.
\end{aligned}$$

- The corresponding SL filter and smoother have the same limitations.
- But we can fix this!

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Statistical linear regression

- We can now replace the deterministic approximation on $\mathbf{g}(\mathbf{x})$ with stochastic approximation

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e},$$

where $\mathbf{e} \sim \mathrm{N}(0, \Sigma)$ is a pseudo-noise.

- The covariance is now

$$\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})\,(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})^\mathsf{T} \right]$$
$$= \mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T} + \Sigma.$$

- Thus, we can force the correct covariance by putting

$$\Sigma = \mathrm{Cov}\left[\mathbf{g}(\mathbf{x})\right] - \mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T}.$$

- This is called statistical linear regression (SLR).

Statistical linear regression

- We can now replace the deterministic approximation on $\mathbf{g}(\mathbf{x})$ with stochastic approximation

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e},$$

where $\mathbf{e} \sim \mathrm{N}(0, \Sigma)$ is a pseudo-noise.

- The covariance is now

$$\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})\,(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})^{\mathsf{T}}\right]$$
$$= \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}} + \Sigma.$$

- Thus, we can force the correct covariance by putting

$$\Sigma = \mathrm{Cov}\,[\mathbf{g}(\mathbf{x})] - \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}.$$

- This is called statistical linear regression (SLR).

Statistical linear regression

- We can now replace the deterministic approximation on $\mathbf{g}(\mathbf{x})$ with stochastic approximation

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e},$$

where $\mathbf{e} \sim \mathrm{N}(0, \Sigma)$ is a pseudo-noise.

- The covariance is now

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})\,(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})^{\mathsf{T}}\right] \\
&= \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}} + \Sigma.
\end{aligned}$$

- Thus, we can force the correct covariance by putting

$$\Sigma = \mathrm{Cov}\,[\mathbf{g}(\mathbf{x})] - \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}.$$

- This is called statistical linear regression (SLR).

Statistical linear regression

- We can now replace the deterministic approximation on $\mathbf{g}(\mathbf{x})$ with stochastic approximation

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e},$$

where $\mathbf{e} \sim \mathrm{N}(0, \Sigma)$ is a pseudo-noise.

- The covariance is now

$$\begin{aligned}
&\mathrm{E}\left[(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})\,(\mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b} + \mathbf{e} - \mathbf{b})^{\mathsf{T}}\right] \\
&= \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}} + \Sigma.
\end{aligned}$$

- Thus, we can force the correct covariance by putting

$$\Sigma = \mathrm{Cov}\,[\mathbf{g}(\mathbf{x})] - \mathbf{A}\,\mathbf{P}\,\mathbf{A}^{\mathsf{T}}.$$

- This is called statistical linear regression (SLR).

Statistical linear regression: prediction step

- Let us use the SLR approximation on the dynamic model:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1},$$

where $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q})$.

- When in SLR we select $\mathbf{m} = \mathbf{m}_{k-1}$ and $\mathbf{P} = \mathbf{P}_{k-1}$ we get

$$\mathbf{x}_k = \mathbf{F}\,(\mathbf{x}_{k-1} - \mathbf{m}_{k-1}) + \mathbf{b} + \mathbf{e} + \mathbf{q}_{k-1},$$

with

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}_{k-1}\left[\mathbf{f}(\mathbf{x}_{k-1})\right], \\
\mathbf{F} &= \mathrm{E}_{k-1}\left[(\mathbf{f}(\mathbf{x}_{k-1}) - \mathbf{b})\,(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})^\mathsf{T}\right] \mathbf{P}_{k-1}^{-1}, \\
\mathrm{Cov}_{k-1}[\mathbf{e}] &= \mathrm{Cov}_{k-1}\left[\mathbf{f}(\mathbf{x}_{k-1})\right] - \mathbf{F}\,\mathbf{P}_{k-1}\,\mathbf{F}^\mathsf{T}.
\end{aligned}$$

Statistical linear regression: prediction step

- Let us use the SLR approximation on the dynamic model:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1},$$

where $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q})$.

- When in SLR we select $\mathbf{m} = \mathbf{m}_{k-1}$ and $\mathbf{P} = \mathbf{P}_{k-1}$ we get

$$\mathbf{x}_k = \mathbf{F}\,(\mathbf{x}_{k-1} - \mathbf{m}_{k-1}) + \mathbf{b} + \mathbf{e} + \mathbf{q}_{k-1},$$

with

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}_{k-1}\left[\mathbf{f}(\mathbf{x}_{k-1})\right], \\
\mathbf{F} &= \mathrm{E}_{k-1}\left[(\mathbf{f}(\mathbf{x}_{k-1}) - \mathbf{b})\,(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})^{\mathsf{T}}\right]\,\mathbf{P}_{k-1}^{-1}, \\
\mathrm{Cov}_{k-1}[\mathbf{e}] &= \mathrm{Cov}_{k-1}\left[\mathbf{f}(\mathbf{x}_{k-1})\right] - \mathbf{F}\,\mathbf{P}_{k-1}\,\mathbf{F}^{\mathsf{T}}.
\end{aligned}$$

Statistical linear regression: prediction step (cont.)

- This now gives the prediction step

$$\mathbf{m}_k^- = \mathrm{E}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})],$$
$$\mathbf{P}_k^- = \mathrm{Cov}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})] + \mathbf{Q}.$$

- The mean and covariance are exact by construction.
- This is indeed the Gaussian filter prediction – leading to UKFs and other sigma-point filters.

Statistical linear regression: prediction step (cont.)

- This now gives the prediction step

$$\begin{aligned} \mathbf{m}_k^- &= \mathrm{E}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})], \\ \mathbf{P}_k^- &= \mathrm{Cov}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})] + \mathbf{Q}. \end{aligned}$$

- The mean and covariance are exact by construction.
- This is indeed the Gaussian filter prediction – leading to UKFs and other sigma-point filters.

Statistical linear regression: prediction step (cont.)

- This now gives the prediction step

$$
\begin{aligned}
\mathbf{m}_k^- &= \mathrm{E}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})], \\
\mathbf{P}_k^- &= \mathrm{Cov}_{k-1}[\mathbf{f}(\mathbf{x}_{k-1})] + \mathbf{Q}.
\end{aligned}
$$

- The mean and covariance are exact by construction.
- This is indeed the Gaussian filter prediction – leading to UKFs and other sigma-point filters.

Statistical linear regression: update step

- For the measurement update we consider

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k,$$

where $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R})$.

- Using SLR with $\mathbf{m} = \mathbf{m}_k^-$ and $\mathbf{P} = \mathbf{P}_k^-$ gives

$$\mathbf{y}_k = \mathbf{H}\,(\mathbf{x}_k - \mathbf{m}_k^-) + \mu_k + \mathbf{e}'_k + \mathbf{r}_k,$$

with

$$\begin{aligned}
\mu_k &= \mathrm{E}_k^-[\mathbf{h}(\mathbf{x}_k)], \\
\mathbf{H} &= \mathrm{E}_k^-\left[(\mathbf{h}(\mathbf{x}_k) - \mu_k)\,(\mathbf{x}_k - \mathbf{m}_k^-)^\mathsf{T}\right]\,\left[\mathbf{P}_k^-\right]^{-1}, \\
\mathrm{Cov}_k^-[\mathbf{e}'] &= \mathrm{Cov}_k^-[\mathbf{h}(\mathbf{x}_k)] - \mathbf{H}\,\mathbf{P}_k^-\,\mathbf{H}^\mathsf{T}.
\end{aligned}$$

Statistical linear regression: update step

- For the measurement update we consider

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k,$$

  where $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R})$.

- Using SLR with $\mathbf{m} = \mathbf{m}_k^-$ and $\mathbf{P} = \mathbf{P}_k^-$ gives

$$\mathbf{y}_k = \mathbf{H}\,(\mathbf{x}_k - \mathbf{m}_k^-) + \boldsymbol{\mu}_k + \mathbf{e}'_k + \mathbf{r}_k,$$

  with

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}_k^-[\mathbf{h}(\mathbf{x}_k)], \\
\mathbf{H} &= \mathrm{E}_k^-\left[(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)\,(\mathbf{x}_k - \mathbf{m}_k^-)^\mathsf{T}\right]\,\left[\mathbf{P}_k^-\right]^{-1}, \\
\mathrm{Cov}_k^-[\mathbf{e}'] &= \mathrm{Cov}_k^-[\mathbf{h}(\mathbf{x}_k)] - \mathbf{H}\,\mathbf{P}_k^-\,\mathbf{H}^\mathsf{T}.
\end{aligned}$$

Statistical linear regression: update step (cont.)

- The joint mean and covariance of $\mathbf{x}_k$ and $\mathbf{y}_k$ are given as

$$\mathrm{E}_k^-\left[\begin{pmatrix}\mathbf{x}_k \\ \mathbf{y}_k\end{pmatrix}\right] = \begin{pmatrix}\mathbf{m}_k^- \\ \boldsymbol{\mu}_k\end{pmatrix}$$

$$\mathrm{Cov}_k^-\left[\begin{pmatrix}\mathbf{x}_k \\ \mathbf{y}_k\end{pmatrix}\right] = \begin{pmatrix}\mathbf{P}_k^- & \mathrm{E}_k^-\left[(\mathbf{x}_k - \mathbf{m}_k^-)\,(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T}\right] \\ (\bullet)^\mathsf{T} & \mathrm{Cov}_k^-[\mathbf{h}(\mathbf{x}_k)] + \mathbf{R}\end{pmatrix}$$

- Gaussian conditioning leads to Gaussian filter update step

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}_k^-[\mathbf{h}(\mathbf{x}_k)], \\
\mathbf{S}_k &= \mathrm{Cov}_k^-[\mathbf{h}(\mathbf{x}_k)] + \mathbf{R}, \\
\mathbf{C}_k &= \mathrm{E}_k^-\left[(\mathbf{x}_k - \mathbf{m}_k^-)(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T}\right], \\
\mathbf{K}_k &= \mathbf{C}_k\,\mathbf{S}_k^{-1}, \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\,(\mathbf{y}_k - \boldsymbol{\mu}_k), \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^\mathsf{T}.
\end{aligned}$$

Statistical linear regression: update step (cont.)

- The joint mean and covariance of $\mathbf{x}_k$ and $\mathbf{y}_k$ are given as

$$\mathrm{E}_k^- \left[\begin{pmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{pmatrix} \right] = \begin{pmatrix} \mathbf{m}_k^- \\ \boldsymbol{\mu}_k \end{pmatrix}$$

$$\mathrm{Cov}_k^- \left[\begin{pmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{pmatrix} \right] = \begin{pmatrix} \mathbf{P}_k^- & \mathrm{E}_k^- \left[(\mathbf{x}_k - \mathbf{m}_k^-) \, (\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T} \right] \\ (\bullet)^\mathsf{T} & \mathrm{Cov}_k^- [\mathbf{h}(\mathbf{x}_k)] + \mathbf{R} \end{pmatrix}$$

- Gaussian conditioning leads to Gaussian filter update step

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}_k^- [\mathbf{h}(\mathbf{x}_k)], \\
\mathbf{S}_k &= \mathrm{Cov}_k^- [\mathbf{h}(\mathbf{x}_k)] + \mathbf{R}, \\
\mathbf{C}_k &= \mathrm{E}_k^- \left[(\mathbf{x}_k - \mathbf{m}_k^-)(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T} \right], \\
\mathbf{K}_k &= \mathbf{C}_k \, \mathbf{S}_k^{-1}, \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k \, (\mathbf{y}_k - \boldsymbol{\mu}_k), \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k \, \mathbf{S}_k \, \mathbf{K}_k^\mathsf{T}.
\end{aligned}$$

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Discussion and connections

- Statistical linear regression leads to the Gaussian filter – and to UKF, CKF, GHKF, etc.
- The corresponding smoother can be derived similarly.
- We have arbitrarily chosen to do the SLR w.r.t. priors $\mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$ and $\mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$ –
  - We don't need to do that, we can use any $\mathbf{x} \sim p(\mathbf{x})$.
  - Leads to e.g. posterior-linearization filters and smoothers that Ángel will talk about next.
- Statistical linearization corresponds to putting $\mathbf{e} = 0$ in

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}.$$

- We recover the statistically linearized filter.
- If we do the statistical linearization with prior means and take $\mathbf{P}_k \to 0$, $\mathbf{P}_k^- \to 0$ we obtain the EKF.

Learning Outcomes

1. Statistical linearization
2. Statistical linear regression
3. Discussion and connections
4. Summary

Summary

- In statistical linearization (SL) we approximate $\mathbf{y} = \mathbf{g}(\mathbf{x})$ as $\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are selected to minimize

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right].$$

- In statistical linear regression (SLR) we use $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}$, where $\mathbf{e}$ is a Gaussian pseudo-noise that makes the covariance exact.
- By forming a filter with SLR we get Gaussian filter, UKFs, and other related filters.
- By using SL we get the classical statistically linearized filter and also EKF as a special case.
- We can also do SL and SLR w.r.t. other than the prior – leading to posterior linearization filters (and smoothers).

Summary

- In statistical linearization (SL) we approximate $\mathbf{y} = \mathbf{g}(\mathbf{x})$ as $\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are selected to minimize

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right].$$

- In statistical linear regression (SLR) we use $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}$, where $\mathbf{e}$ is a Gaussian pseudo-noise that makes the covariance exact.
- By forming a filter with SLR we get Gaussian filter, UKFs, and other related filters.
- By using SL we get the classical statistically linearized filter and also EKF as a special case.
- We can also do SL and SLR w.r.t. other than the prior – leading to posterior linearization filters (and smoothers).

Summary

- In statistical linearization (SL) we approximate $\mathbf{y} = \mathbf{g}(\mathbf{x})$ as $\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are selected to minimize

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right].$$

- In statistical linear regression (SLR) we use $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}$, where $\mathbf{e}$ is a Gaussian pseudo-noise that makes the covariance exact.
- By forming a filter with SLR we get Gaussian filter, UKFs, and other related filters.
- By using SL we get the classical statistically linearized filter and also EKF as a special case.
- We can also do SL and SLR w.r.t. other than the prior – leading to posterior linearization filters (and smoothers).

Summary

- In statistical linearization (SL) we approximate $\mathbf{y} = \mathbf{g}(\mathbf{x})$ as $\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are selected to minimize

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right].$$

- In statistical linear regression (SLR) we use $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}$, where $\mathbf{e}$ is a Gaussian pseudo-noise that makes the covariance exact.
- By forming a filter with SLR we get Gaussian filter, UKFs, and other related filters.
- By using SL we get the classical statistically linearized filter and also EKF as a special case.
- We can also do SL and SLR w.r.t. other than the prior – leading to posterior linearization filters (and smoothers).

Summary

- In statistical linearization (SL) we approximate $\mathbf{y} = \mathbf{g}(\mathbf{x})$ as $\mathbf{y} = \mathbf{A}\,(\mathbf{x} - \mathbf{m}) + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are selected to minimize

$$\mathrm{MSE}(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\|\mathbf{g}(\mathbf{x}) - \mathbf{A}\,\mathbf{x} - \mathbf{b}\|^2\right].$$

- In statistical linear regression (SLR) we use $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} + \mathbf{e}$, where $\mathbf{e}$ is a Gaussian pseudo-noise that makes the covariance exact.
- By forming a filter with SLR we get Gaussian filter, UKFs, and other related filters.
- By using SL we get the classical statistically linearized filter and also EKF as a special case.
- We can also do SL and SLR w.r.t. other than the prior – leading to posterior linearization filters (and smoothers).