Aalto University
School of Electrical Engineering

ELEC-E8740 — Sensors, Models, and Least Squares Criterion

Simo Särkkä

Aalto University

September 20, 2022

Contents

Intended Learning Outcomes

After this lecture, you will be able to:

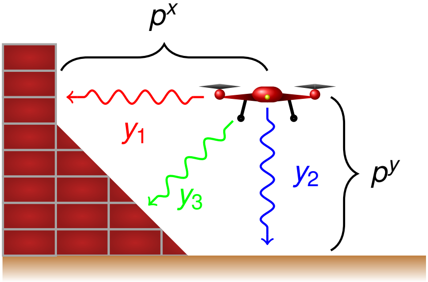
- Recognize and name the components of sensor fusion systems: sensors, models, estimation algorithms;
- Understand the notation used in sensor fusion (on this course);
- Describe and identify the purpose of an optimality criterion and cost functions;
- Understand what are least squares, weighted least squares, and regularized least squares criteria.

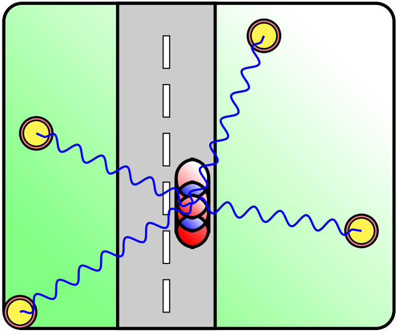
Recap (1)

- Lectures are on Tuesdays in 12:15-14:00
- Exercises on Fridays in 12:15-14:00
- Teaching materials are lecture notes and slides on MyCourses.
- Project work starts later and it is about sensor fusion in a mobile robot.
- There are two mid-term Exams.
- The grade is determined by exams, homeworks, and project work.
- Sensor fusion is methodology for intelligent processing of measurements from multiple sensors.
- In practice, linear/non-linear least squares methods and Kalman filtering methods.

Recap (2)

Typical models that we saw are the following:

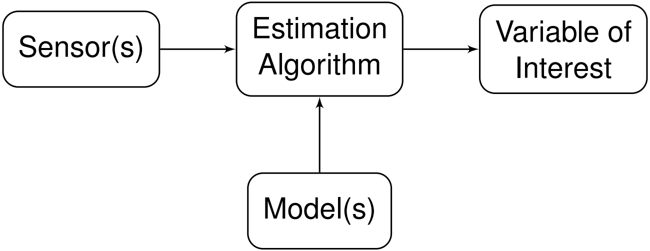




$$\mathbf{y} = \mathbf{G}\,\mathbf{x} + \mathbf{b} + \mathbf{r}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{r}$$

The Components of Sensor Fusion



Variable of Interest

Definition

One or more unknown static quantities of interest, parameters or a time-varying state of a dynamic system of interest that can be measured directly or indirectly.

Notation

- A single (scalar) static parameter is denoted x,
- a vector of K static parameters is denoted as $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_K \end{bmatrix}^\mathsf{T}$,
- a scalar time-varying state is denoted $x_n = x(t_n)$,
- a vector time-varying state is denoted $\mathbf{x}_n = \mathbf{x}(t_n)$.

Sensors (1/3)

Definition

- Sensor is a device that provides a measurement related to the quantity of interest.
- Usually, implemented as a device which converts a physical phenomenon into an electrical signal (Wilson, 2005) which is then further transformed into digital form.

- May measure the variable directly or indirectly
- Measurement range and environmental conditions
- Is affected by noise, biases, and uncertainty
- May give measurements frequently or infrequently
- Scalar or vector measurements

Sensors (2/3)

Notation

- A scalar measurement is denoted y_n
- A vector measurement is denoted $\mathbf{y}_n$
- n is a measurement number, sensor id, time, etc.

Sensors (3/3)

Examples

Sensor	Measurement	Application Examples
Accelerometer	Gravity, acceleration	Inertial navigation, activity tracking, screen rotation
Gyroscope	Rotational velocity	Inertial navigation, activity tracking
Magnetometer	Magnetic field strength	Inertial navigation, digital compass, object tracking
Radar	Range, bearing, speed	Target tracking, autonomous vehicles
LIDAR	Range, bearing, speed	Target tracking, autonomous vehicles, robotics
Ultrasound	Range	Robotics
Camera	Visual scene	Security systems, autonomous vehicles, robotics
Barometer	Air pressure	Inertial navigation, autonomous vehicles, robotics
GNSS	Position	Autonomous vehicles, aerospace applications
Strain gauge	Strain	Condition monitoring, scales

Model

Definition

Describes how the variable of interest is observed by the sensor in a systematic way.

- May be very simple or very complex
- Takes noise, uncertainty, and other error sources into account
- Formulated using mathematics

Estimation Algorithm

Definition

Combines the measurements from multiple sensors by using the corresponding models to estimate the quantities of interest in some optimal sense.

- Combining multiple measurements increases the precision (on average)
- Measurements from different sensors can be incorporated
- Can account for the uncertainty of different measurements
- In this course the algorithms minimize a least squares cost criterion to achieve these.

A Basic Model

- The general form is:

$$\text{Measurement} = \text{Function of Parameter(s)} + \text{Noise}$$

- Mathematically:

$$y_n = g_n(\mathbf{x}) + r_n$$

- Anatomy:
  - Measurement y_n is on the left hand side, and
  - Function $g_n(\mathbf{x})$ of $\mathbf{x}$ and a *noise* term r_n on the right side.
- This is called *sensor model*, *measurement model*, or *observation model*.

Measurement Noise

- Encodes thermal sensor noise, uncertainty, etc.
- r_n is modeled as a random variable, follows a probability density function (pdf)

$$r_n \sim p(r_n)$$

- For now, we assume zero-mean, independent random variables with variance $\sigma_{r,n}^2$

$$\begin{aligned} \mathrm{E}\{r_n\} &= 0, \\ \mathrm{var}\{r_n\} &= \mathrm{E}\{r_n^2\} - (\mathrm{E}\{r_n\})^2 = \sigma_{r,n}^2, \\ \mathrm{Cov}\{r_m, r_n\} &= \mathrm{E}\{r_m r_n\} - \mathrm{E}\{r_m\}\,\mathrm{E}\{r_n\} = 0 \quad (m \neq n) \end{aligned}$$

Vector Model

- Extending the basic model for *vector-valued* measurements:

$$\mathbf{y}_n = \mathbf{g}_n(\mathbf{x}) + \mathbf{r}_n,$$

- $\mathbf{y}_n$ and $\mathbf{r}_n$ are d_y-dimensional column vectors
- $\mathbf{r}_n$ is a multivariate random variable with pdf

$$\mathbf{r}_n \sim p(\mathbf{r}_n)$$

- Assume zero-mean, independent random variables with covariance $\mathbf{R}_n$

$$\begin{aligned}
\mathrm{E}\{\mathbf{r}_n\} &= 0, \\
\mathrm{Cov}\{\mathbf{r}_n\} &= \mathrm{E}\{\mathbf{r}_n \mathbf{r}_n^\mathsf{T}\} - \mathrm{E}\{\mathbf{r}_n\}\,\mathrm{E}\{\mathbf{r}_n\}^\mathsf{T} = \mathbf{R}_n, \\
\mathrm{Cov}\{\mathbf{r}_m, \mathbf{r}_n\} &= \mathrm{E}\{\mathbf{r}_m \mathbf{r}_n^\mathsf{T}\} - \mathrm{E}\{\mathbf{r}_m\}\,\mathrm{E}\{\mathbf{r}_n\}^\mathsf{T} = 0 \quad (m \neq n)
\end{aligned}$$

Multiple Measurements

- Sensor fusion requires multiple sensors, repeated measurements, or both
- In the terminology of the measurement model, they can be regarded the same: $y_1, y_2, \ldots, y_N$ or $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$
- We denote a set of measurements:
  - $y_{1:N} = \{y_1, y_2, \ldots, y_N\}$ for the scalar case
  - $\mathbf{y}_{1:N} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ for the vector case
- Examples: Sensor networks, sensor arrays, multi-view imaging, etc.

Measurement Stacking: Scalar Case

- Remember: $y_n = g_n(\mathbf{x}) + r_n$
- Given the measurements $y_{1:N}$, we can write:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_N(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$

- *Compact* notation for *all measurements*:

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

- Covariance for $\mathbf{r}$: $\mathrm{Cov}\{\mathbf{r}\} = \mathbf{R} = \begin{bmatrix} \sigma_{r,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{r,2}^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{r,N}^2 \end{bmatrix}$

Measurement Stacking: Vector Case

- Vector case: $\mathbf{y}_n = \mathbf{g}_n(\mathbf{x}) + \mathbf{r}_n$
- Stacked notation:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1(\mathbf{x}) \\ \mathbf{g}_2(\mathbf{x}) \\ \vdots \\ \mathbf{g}_N(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_N \end{bmatrix}$$

- Hence,

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

where $\mathrm{Cov}\{\mathbf{r}\} = \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{R}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{R}_N \end{bmatrix}.$

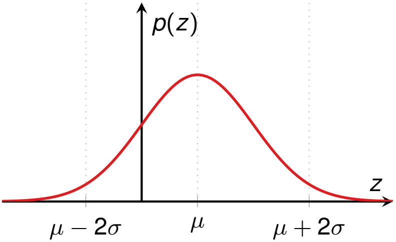
Gaussian Measurement Noise

- Noise is often assumed to be Gaussian, i.e.

$$p(\mathbf{r}) = \frac{1}{(2\pi)^{M/2}|\mathbf{R}|^{1/2}} e^{-\frac{1}{2}\mathbf{r}^\mathsf{T}\mathbf{R}^{-1}\mathbf{r}}$$

- Compact notation $p(\mathbf{r}) = \mathcal{N}(\mathbf{r}; 0, \mathbf{R})$, where

$$\mathcal{N}(\mathbf{z}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\mathsf{T}\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})}$$



Cost Functions (1/2)

- An optimality criterion is required to develop an estimation algorithm
- We focus on algorithms that minimize a cost function of the error

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} J(\mathbf{x})$$

  where
  - $\hat{\mathbf{x}}$ denotes the estimate of $\mathbf{x}$
  - $J(\mathbf{x})$ is the cost function
  - $\operatorname{argmin}_{\mathbf{x}} J(\mathbf{x})$ denotes "the argument $\mathbf{x}$ that minimizes $J(\mathbf{x})$"
- The error is given by the difference between the measurement and the output predicted by $\mathbf{x}$

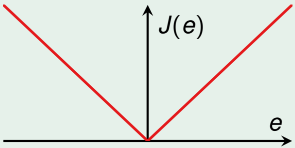
$$e_n = y_n - g_n(\mathbf{x})$$

Cost Functions (2/2)

Absolute Error

Penalizes all errors equally:

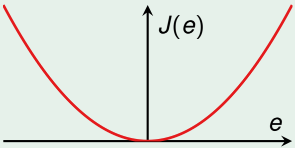
$$|e_n| = |y_n - g_n(\mathbf{x})|,$$



Quadratic Error

Penalizes large errors more than small ones:

$$e_n^2 = (y_n - g_n(\mathbf{x}))^2.$$



Least Squares (1/2)

- The quadratic cost is much more common
- Closely related to Gaussian measurement noise
- Minimizing the quadratic cost function is the least squares method
- Cost function for N scalar measurements $y_{1:N} = \{y_1, y_2, \ldots, y_N\}$

$$J_{\text{LS}}(\mathbf{x}) = \sum_{n=1}^{N} e_n^2 = \sum_{n=1}^{N} (y_n - g_n(\mathbf{x}))^2$$

Least Squares (2/2)

- Quadratic error for vector measurements

$$e_n^2 = (\mathbf{y}_n - \mathbf{g}_n(\mathbf{x}))^\mathsf{T}(\mathbf{y}_n - \mathbf{g}_n(\mathbf{x}))$$

- Cost function for N vector measurements $\mathbf{y}_{1:N} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$

$$J_\mathrm{LS}(\mathbf{x}) = \sum_{n=1}^{N} (\mathbf{y}_n - \mathbf{g}_n(\mathbf{x}))^\mathsf{T}(\mathbf{y}_n - \mathbf{g}_n)$$

- Quadratic error and cost function for stacked (batch) notation

$$J_\mathrm{LS}(\mathbf{x}) = (\mathbf{y} - \mathbf{g}(\mathbf{x}))^\mathsf{T}(\mathbf{y} - \mathbf{g}(\mathbf{x}))$$

Weighted Least Squares (1/2)

- How to include confidence in sensor readings?
- Weighted least squares (WLS) cost function:

$$J_{\mathrm{WLS}}(\mathbf{x}) = \sum_{n=1}^{N} w_n (y_n - g_n(\mathbf{x}))^2,$$

where $w_n > 0$ is a weighing factor for the nth measurement

- WLS vector cost function:

$$J_{\mathrm{WLS}}(\mathbf{x}) = \sum_{n=1}^{N} (\mathbf{y}_n - \mathbf{g}_n(\mathbf{x}))^{\mathsf{T}} \mathbf{W}_n (\mathbf{y}_n - \mathbf{g}_n(\mathbf{x})),$$

where $\mathbf{W}_n$ is a positive-definite weighing matrix

Weighted Least Squares (2/2)

- WLS stacked cost function:

$$J_{\text{WLS}}(\mathbf{x}) = (\mathbf{y} - \mathbf{g}(\mathbf{x}))^{\mathsf{T}} \mathbf{W} (\mathbf{y} - \mathbf{g}(\mathbf{x}))$$

where $\mathbf{W}$ is the positive-definite weighing matrix

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & w_N \end{bmatrix} \text{ or } \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 & \dots & 0 \\ 0 & \mathbf{W}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{W}_N \end{bmatrix}$$

- Choice of w_n or $\mathbf{W}_n$ is in principle arbitrary
- In practice, good choices are

$$w_n = 1/\sigma_{r,n}^2 \text{ and } \mathbf{W}_n = \mathbf{R}_n^{-1}$$

Regularized Least Squares

- In plain (weighted) least squares we find an estimate that best explains the measurements.
- In regularized least squares we add a penalty term in the estimate.
- The penalty term can force the estimate to be "small" or close to certain "a priori" know value.
- The general form of regularized least squares (ReLS) that we use is

$$J_{\text{ReLS}}(\mathbf{x}) = (\mathbf{y} - \mathbf{g}(\mathbf{x}))^{\mathsf{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{g}(\mathbf{x})) + (\mathbf{x} - \mathbf{m})^{\mathsf{T}} \mathbf{P}^{-1} (\mathbf{x} - \mathbf{m}).$$

- Regularized least squares can also be used to formulate dynamic estimation methods.

Summary

- Sensor fusion involves three components:
  1. Sensor: Measures a variable of interest, directly or indirectly
  2. Model: A mathematical formulation that relates the variables of interest to the measurements
  3. Estimation Algorithm: Combines the measurements and models to estimate the variables of interest
- Multiple measurements and multidimensional measurements can be written in the same vector notation.
- The least squares method is a good way for deriving estimators.
- (Plain) least squares, weighted least squares, and regularized least squares are useful criteria for estimators.