



Aalto University
School of Electrical
Engineering

ELEC-E8740 — Static Linear Models and Linear Least Squares

Simo Särkkä

Aalto University

September 18, 2019

Contents

- 1 Intended Learning Outcomes and Recap
- 2 Linear Scalar and Vector Models
- 3 Least Squares for General Linear Model
- 4 Weighted Least Squares
- 5 Regularized Least Squares
- 6 Sequential Least Squares
- 7 Summary

Intended Learning Outcomes

After this lecture, you will be able to:

- Identify and construct scalar and vector linear models;
- apply and derive (weighted, regularized, and sequential) linear least squares estimators;
- investigate the properties of linear least squares estimators.

Recap

- **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.

Scalar Model: Model & Cost Function

- Many sensors measure (a scaled) version of a **single unknown x**

$$y_n = gx + r_n,$$

with $E\{r_n\} = 0$ and $\text{var}\{r_n\} = \sigma_{r,n}^2$

- The **error** for one measurement is

$$e_n = y_n - gx$$

and the **least squares cost function** is given by

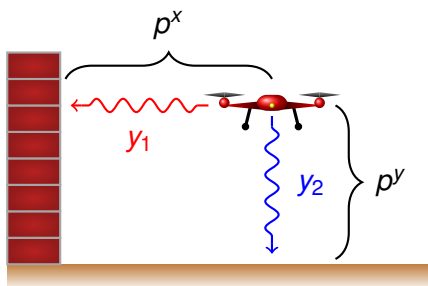
$$J_{\text{LS}}(x) = \sum_{n=1}^N (y_n - gx)^2$$

Scalar Model: Example

Example

- **Radar** measures the time difference between the sent and reflected signals.
- **Double distance** divided by the speed of light
 $\tau = 2p^x / c$, where
 $c = 299792458$ m/s.
- The **measurement model** is (beware of the notation!)

$$y_n = \frac{2}{c} p^x + r_n, \quad n = 1, \dots, N.$$



Scalar Model: Minimizing the Cost

- The derivative is given by

$$\frac{\partial J_{\text{LS}}(x)}{\partial x} = -2g \sum_{n=1} y_n + 2Ng^2x$$

- Setting the derivative to zero and solving for x yields

$$\hat{x}_{\text{LS}} = \frac{1}{Ng} \sum_{n=1}^N y_n.$$

- This is the least squares estimator for the model

$$y_n = gx + r_n.$$

Scalar Model: Estimator Properties

- What are the estimator's **statistical properties**?
- **Expected value:**

$$E\{\hat{x}_{LS}\} = x + \sum_{n=1}^N E\{r_n\} = x$$

- **Variance:**

$$\text{var}\{\hat{x}\} = \frac{1}{N^2 g^2} \sum_{n=1}^N \sigma_{r,n}^2$$

and when $\sigma_{r,n}^2 = \sigma^2$ we get

$$\text{var}\{\hat{x}\} = \frac{\sigma^2}{N g^2} \quad \text{and} \quad \text{std}\{\hat{x}\} = \frac{1}{\sqrt{N}} \frac{\sigma}{|g|}.$$

- The expectation of \hat{x} is $x \Rightarrow$ estimator is **unbiased**.

Scalar Model: Example

Example

Let us consider the wall-distance measurement model and assume that we estimate p^x with N measurements:

$$\hat{x}_1 = \frac{c}{2N} \sum_{n=1}^N y_n.$$

This estimator is unbiased. Further assume that the standard deviation of the measurement is $\sigma = 10^{-9}$ s (1 nanosecond). Then the standard deviation of the estimator is

$$\text{std}\{\hat{x}_1\} = \frac{1}{\sqrt{N}} \frac{c\sigma}{2}.$$

With a single measurement we get the error of 15 cm whereas by averaging 100 measurements the error drops to 1.5 cm.

Vector Models

- Scalar observations, several parameters x_1, x_2, \dots, x_K :

$$\begin{aligned}y_n &= g_1 x_1 + g_2 x_2 + \dots + g_K x_K + r_n \\ &= \mathbf{g}\mathbf{x} + r_n\end{aligned}$$

- Slightly more generally:

$$y_n = \mathbf{g}_n \mathbf{x} + r_n$$

- Stacking measurements together gives:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_N \end{bmatrix} \mathbf{x} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$

- This has the general form

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{r}, \text{ with } \text{Cov}\{\mathbf{r}\} = \mathbf{R} = \text{diag}(\sigma_{r,1}^2, \dots, \sigma_{r,N}^2).$$

Vector Models (cont.)

- Vector observations, several parameters:

$$\mathbf{y}_n = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1K} \\ g_{21} & g_{22} & \cdots & g_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{d_y 1} & g_{d_y 2} & \cdots & g_{d_y K} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} + \mathbf{r}_n$$
$$= \mathbf{G}_n \mathbf{x} + \mathbf{r}_n,$$

- Batch notation:

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

- In compact notation we again get the same **general form**:

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{r}, \text{ with } \text{Cov}\{\mathbf{r}\} = \mathbf{R} = \text{diag}(\mathbf{R}_1, \dots, \mathbf{R}_N).$$

General Linear Model: Definition

- General form of a linear models:

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

with $E\{\mathbf{r}\} = 0$ and $\text{Cov}\{\mathbf{r}\} = \mathbf{R}$.

- This is the general linear model, both the scalar and vector cases can be expressed in this way.

Affine Models

- We might also have a **constant bias** \mathbf{b} in the model:

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

- We can now compute a **modified measurement** $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{b}$ and rewrite this as

$$\tilde{\mathbf{y}} = \mathbf{G}\mathbf{x} + \mathbf{r}.$$

- Thus is again a **general linear model**.

Example: Localizing a Drone

- Recall the drone model:

$$y_1 = p^x + r_1,$$

$$y_2 = p^y + r_2,$$

$$y_3 = \frac{1}{\sqrt{2}}(p^x - x_0) + \frac{1}{\sqrt{2}}p^y + r_3.$$

- It has the affine form:

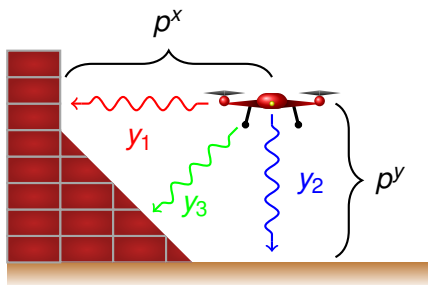
$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{b} + \mathbf{r}.$$

- Can be reduced to linear model by defining

$$\tilde{y}_1 = y_1,$$

$$\tilde{y}_2 = y_2,$$

$$\tilde{y}_3 = y_3 + \frac{1}{\sqrt{2}}x_0.$$



Example: Localizing a Car

- We have:

$$y_1 = s_1^x - p^x + r_1,$$

$$y_2 = s_1^y - p^y + r_2,$$

⋮

$$y_{2M} = s_M^y - p^y + r_{2M}.$$

- Again leads to form

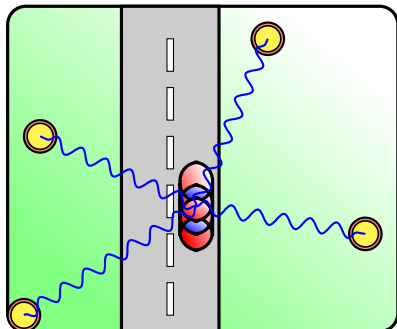
$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{b} + \mathbf{r}.$$

- We can now define

$$\tilde{y}_1 = y_1 - s_1^x,$$

⋮

$$\tilde{y}_{2M} = y_{2M} - s_M^y.$$



General Linear Model: Least Squares (1)

- The **least squares cost function** to minimize:

$$\begin{aligned} J_{LS}(\mathbf{x}) &= (\mathbf{y} - \mathbf{G}\mathbf{x})^T (\mathbf{y} - \mathbf{G}\mathbf{x}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{G}\mathbf{x} - \mathbf{x}^T \mathbf{G}^T \mathbf{y} + \mathbf{x}^T \mathbf{G}^T \mathbf{G}\mathbf{x} \end{aligned}$$

- Some **vector calculus identities** (when \mathbf{A} is symmetric):

$$\begin{aligned} \frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \\ \frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x} \end{aligned}$$

General Linear Model: Least Squares (2)

- The **least squares estimator** for the general linear model is

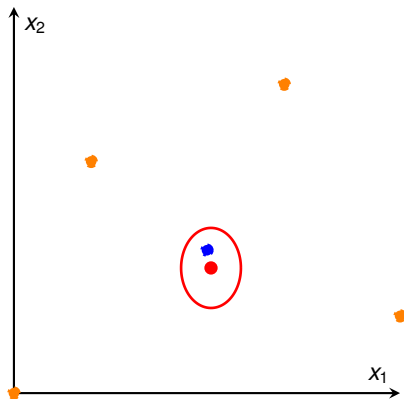
$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y}$$

- Its **statistical properties** are

$$E\{\hat{\mathbf{x}}_{\text{LS}}\} = \mathbf{x}$$

$$\text{Cov}\{\hat{\mathbf{x}}_{\text{LS}}\} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{R} \mathbf{G} ((\mathbf{G}^T \mathbf{G})^{-1})^T.$$

Example: Localizing a Car (1)



Weighted Linear Least Squares (1)

- Recall the **general linear model**:

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{r}.$$

with $E\{\mathbf{r}\} = \mathbf{0}$ and $\text{Cov}\{\mathbf{r}\} = \mathbf{R}$.

- Weighted least squares** cost function:

$$J_{\text{WLS}}(\mathbf{x}) = (\mathbf{y} - \mathbf{G}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{G}\mathbf{x})$$

- We can now derive the estimator **in the same way** as for (plain) least squares.

Weighted Linear Least Squares (2)

- Weighted linear least squares estimator:

$$\hat{\mathbf{x}}_{\text{WLS}} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{y}.$$

- Properties:

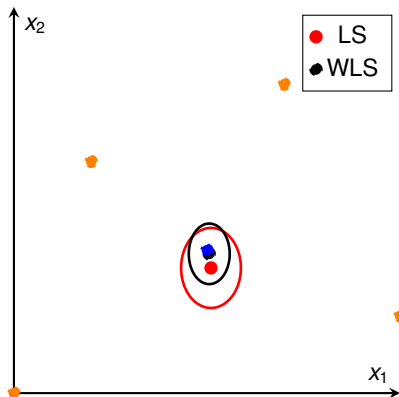
$$E\{\hat{\mathbf{x}}_{\text{WLS}}\} = \mathbf{x}$$

$$\text{Cov}\{\hat{\mathbf{x}}_{\text{WLS}}\} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1}$$

- It can be shown that $\mathbf{W} = \mathbf{R}^{-1}$ minimizes $\text{Cov}\{\hat{\mathbf{x}}_{\text{WLS}}\}$ over all choices for \mathbf{W} and in this case

$$\text{Cov}\{\hat{\mathbf{x}}_{\text{WLS}}\} \leq \text{Cov}\{\hat{\mathbf{x}}_{\text{LS}}\}$$

Example: Localizing a Car (2)



Regularized Linear Least Squares (1/2)

- The **regularized least squares** criterion

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

- The **regularized linear least squares estimator**

$$\hat{\mathbf{x}}_{\text{ReLS}} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} + \mathbf{P}^{-1})^{-1} (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{y} + \mathbf{P}^{-1} \mathbf{m}).$$

- The expectation is **not \mathbf{x}** !
- The **covariance** of the estimator is

$$\text{Cov}\{\hat{\mathbf{x}}_{\text{ReLS}}\} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} + \mathbf{P}^{-1})^{-1}.$$

- The covariance is **always smaller** (or equal) to the WLS estimator.

Regularized Linear Least Squares (2/2)

- By using the **matrix inversion formula** we can write

$$(\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} + \mathbf{P}^{-1})^{-1} = \mathbf{P} - \mathbf{P} \mathbf{G}^T (\mathbf{G} \mathbf{P} \mathbf{G}^T + \mathbf{R})^{-1} \mathbf{G} \mathbf{P}.$$

- This gives

$$\mathbf{K} = \mathbf{P} \mathbf{G}^T (\mathbf{G} \mathbf{P} \mathbf{G}^T + \mathbf{R})^{-1},$$

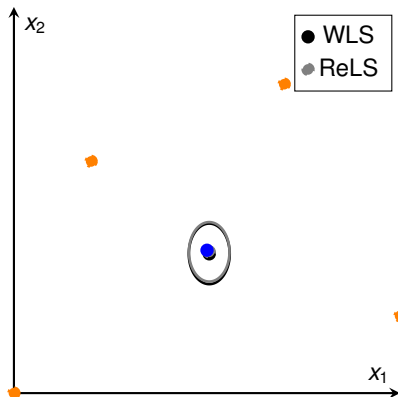
$$\hat{\mathbf{x}}_{\text{ReLS}} = \mathbf{m} + \mathbf{K}(\mathbf{y} - \mathbf{G} \mathbf{m}),$$

$$\text{Cov}\{\hat{\mathbf{x}}_{\text{ReLS}}\} = \mathbf{P} - \mathbf{K}(\mathbf{G} \mathbf{P} \mathbf{G}^T + \mathbf{R}) \mathbf{K}^T.$$

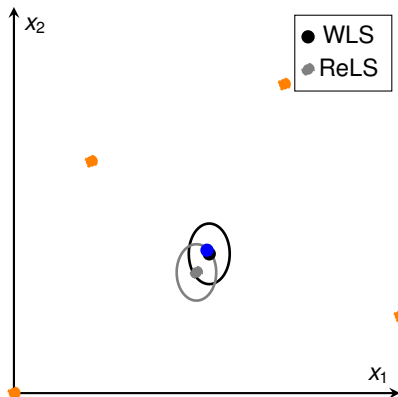
- Finally, we can always **rewrite** regularized least squares as **weighted least squares**:

$$\begin{aligned} J_{\text{ReLS}}(\mathbf{x}) &= (\mathbf{y} - \mathbf{G} \mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{G} \mathbf{x}) + (\mathbf{m} - \mathbf{x})^T \mathbf{P}^{-1} (\mathbf{m} - \mathbf{x}) \\ &= \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} - \begin{bmatrix} \mathbf{G} \\ \mathbf{I} \end{bmatrix} \mathbf{x} \right)^T \begin{bmatrix} \mathbf{R}^{-1} & 0 \\ 0 & \mathbf{P}^{-1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} - \begin{bmatrix} \mathbf{G} \\ \mathbf{I} \end{bmatrix} \mathbf{x} \right). \end{aligned}$$

Example: Localizing a Car (3)



Example: Localizing a Car (4)



Sequential Linear Least Squares (1/2)

- In many cases, the sensor data arrives **sequentially** at the estimator.
- Assume that we have calculated the **weighted least squares (WLS) estimate** using $\mathbf{y}_{1:n-1} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n-1}\}$:

$$\hat{\mathbf{x}}_{n-1} = (\mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{G}_{1:n-1})^{-1} \mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{y}_{1:n-1},$$

$$\text{Cov}\{\hat{\mathbf{x}}_{k-1}\} = (\mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{G}_{1:n-1})^{-1} = \mathbf{P}_{n-1}.$$

- We can now **rewrite** the WLS the cost function in **sequential form**

$$\begin{aligned} J_{\text{SLS}}(\mathbf{x}) &= (\mathbf{y}_{1:n-1} - \mathbf{G}_{1:n-1} \mathbf{x})^T \mathbf{R}_{1:n-1}^{-1} (\mathbf{y}_{1:n-1} - \mathbf{G}_{1:n-1} \mathbf{x}) \\ &\quad + (\mathbf{y}_n - \mathbf{G}_n \mathbf{x})^T \mathbf{R}_n^{-1} (\mathbf{y}_n - \mathbf{G}_n \mathbf{x}). \end{aligned}$$

Sequential Linear Least Squares (2/2)

- Setting gradient to zero and substituting the already computed result gives:

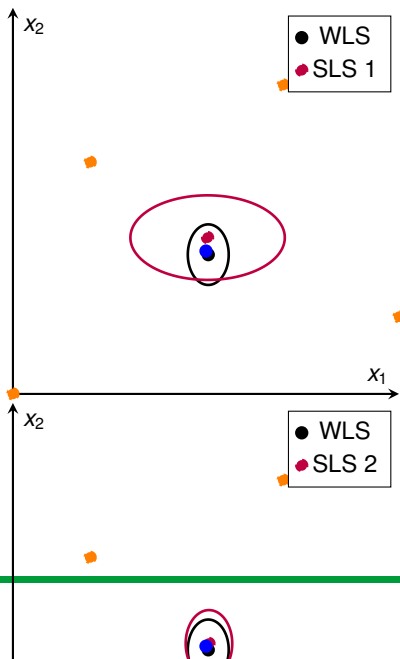
$$\begin{aligned}\hat{\mathbf{x}}_n &= (\mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{G}_{1:n-1} + \mathbf{G}_n^T \mathbf{R}_n^{-1} \mathbf{G}_n)^{-1} \\ &\quad \times (\mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{y}_{1:n-1} + \mathbf{G}_n^T \mathbf{R}_n^{-1} \mathbf{y}_n) \\ &= (\mathbf{P}_{n-1}^{-1} + \mathbf{G}_n^T \mathbf{R}_n^{-1} \mathbf{G}_n)^{-1} (\mathbf{G}_{1:n-1}^T \mathbf{R}_{1:n-1}^{-1} \mathbf{y}_{1:n-1} + \mathbf{G}_n^T \mathbf{R}_n^{-1} \mathbf{y}_n).\end{aligned}$$

- Using matrix inversion formula gives

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n-1} \mathbf{G}_n^T (\mathbf{G}_n \mathbf{P}_{n-1} \mathbf{G}_n^T + \mathbf{R}_n)^{-1}, \\ \hat{\mathbf{x}}_n &= \hat{\mathbf{x}}_{n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n-1}), \\ \text{Cov}\{\hat{\mathbf{x}}_n\} &= \mathbf{P}_{n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n-1} \mathbf{G}_n^T + \mathbf{R}_n) \mathbf{K}_n^T = \mathbf{P}_n.\end{aligned}$$

- This is now a recursion for the estimates.
- Regularized least squares results from $\hat{\mathbf{x}}_0 = \mathbf{m}$ and $\mathbf{P}_0 = \mathbf{P}$.

Example: Localizing a Car (5)



Summary (1)

- The **general linear model** is given by

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{r}, \quad E\{\mathbf{r}\} = 0, \quad \text{Cov}\{\mathbf{r}\} = \mathbf{R}$$

- Affine models** can be tackled by rewriting

$$\begin{aligned}\mathbf{y} &= \mathbf{G}\mathbf{x} + \mathbf{b} + \mathbf{r}, \\ \underbrace{\mathbf{y} - \mathbf{b}}_{\tilde{\mathbf{y}}} &= \mathbf{G}\mathbf{x} + \mathbf{r}.\end{aligned}$$

- Different **least squares estimators**:

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y},$$

$$\hat{\mathbf{x}}_{\text{WLS}} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{y},$$

$$\hat{\mathbf{x}}_{\text{ReLS}} = (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} + \mathbf{P}^{-1})^{-1} (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{y} + \mathbf{P}^{-1} \mathbf{m}).$$

- We also computed their **expectations and covariances**.

Summary (2)

- Alternative form of regularized least squares estimator:

$$\begin{aligned}\mathbf{K} &= \mathbf{P}\mathbf{G}^T(\mathbf{G}\mathbf{P}\mathbf{G}^T + \mathbf{R})^{-1}, \\ \hat{\mathbf{x}}_{\text{ReLS}} &= \mathbf{m} + \mathbf{K}(\mathbf{y} - \mathbf{G}\mathbf{m}), \\ \text{Cov}\{\hat{\mathbf{x}}_{\text{ReLS}}\} &= \mathbf{P} - \mathbf{K}(\mathbf{G}\mathbf{P}\mathbf{G}^T + \mathbf{R})\mathbf{K}^T.\end{aligned}$$

- Sequential (weighted/regularized) least squares estimator:

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n-1}\mathbf{G}_n^T(\mathbf{G}_n\mathbf{P}_{n-1}\mathbf{G}_n^T + \mathbf{R}_n)^{-1}, \\ \hat{\mathbf{x}}_n &= \hat{\mathbf{x}}_{n-1} + \mathbf{K}_n(\mathbf{y}_n - \mathbf{G}_n\hat{\mathbf{x}}_{n-1}), \\ \mathbf{P}_n &= \mathbf{P}_{n-1} - \mathbf{K}_n(\mathbf{G}_n\mathbf{P}_{n-1}\mathbf{G}_n^T + \mathbf{R}_n)\mathbf{K}_n^T.\end{aligned}$$