Systemidentifikation und Regelung in der Medizin

2. Lineare Regression (1. Teil: Least Squares)

Sommersemester 2020

6. Mai 2020

Thomas Schauer

Technische Universität Berlin
Fachgebiet Regelungssysteme

Literaturempfehlung: E. Ikonen und K. Najim, Advanced Process Identification and Control, Marcel Dekker, Inc., 2002

2. Lineare Regression

Parameterschätzung für Systeme mit linearer Struktur →
analytische Lösung verfügbar

Übersicht

2.1 Lineare Regressionsmodelle

- Parametrisiertes Modell $\mathcal{M}(\boldsymbol{\Theta}, \boldsymbol{\varphi})$ mit dem Regressionsvektor $\boldsymbol{\varphi}$ und dem Parametervektor $\boldsymbol{\Theta}$
- Linearität bezüglich des Parametervektors $\boldsymbol{\Theta}$!
- Der Regressionsvektors $\boldsymbol{\varphi}$ wird anhand der wirklichen Modelleingangsgrößen bestimmt und kann auch nichtlinear von diesen abhängen (siehe Beispiele).
- Lineares Regressionsmodell (k – Index der Messung)

$$\begin{aligned}
y(k) &= \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k) \\
\boldsymbol{\Theta}^T &= \left[\Theta_1, \cdots, \Theta_i, \cdots, \Theta_I\right] \text{ Parametervektor} \\
\boldsymbol{\varphi}^T &= \left[\varphi_1, \cdots, \varphi_i, \cdots, \varphi_I\right] \text{ Regressionsvektor} \\
\zeta(k) &- \text{ Rauschen}
\end{aligned}$$

- Beobachtung von $y(k)$ und $\boldsymbol{\varphi}(k)$ für $k = 1, \ldots, K$ (K – Anzahl der Messungen).
- Modellierung von Offset (Bias): $\varphi_I = 1$

Beispiele für lineare Regressionsmodelle:

(a) Statisches System

$$y(k) = a_1 u_1(k) + a_2 u_2^2(k) + a_3 u_3(k) + a_4 + \zeta(k)$$

a_i $(i = 1, 2, 3, 4)$ sind konstante Parameter. Das lineare Regressionsmodell lautet in diesem Fall

$$\mathbf{\Theta} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \quad \boldsymbol{\varphi} = \begin{bmatrix} u_1 \\ u_2^2 \\ u_3 \\ 1 \end{bmatrix}, \quad y(k) = \mathbf{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k).$$

(b) Dynamisches System (lineare Differenzengleichung)

$$y(k) + a_1 y(k-1) + \cdots + a_{n_A} y(k - n_A) =$$
$$b_0 u(k-d) + \cdots + b_{n_B} u(k - d - n_B) + \zeta(k)$$

Eingangsgrößensequenz $\{u(k)\}$ und Ausgangsgrößensequenz $\{y(k)\}$, abgetastet zu diskreten Abtastzeitpunkten $k = 1, 2, 3, \ldots$, d.h. zu den Zeiten $t = k\Delta$ mit dem Abtastintervall Δ; Konstanten (Parameter) a_i und b_i sowie Totzeit d

$$\begin{aligned} y(k) &= -a_1 y(k-1) - \cdots - a_{n_A} y(k-n_A) + \\ & \quad b_0 u(k-d) + \cdots + b_{n_B} u(k-d-n_B) + \zeta(k) \end{aligned}$$

Das lineare Regressionsmodell lautet in diesem Fall

$$\boldsymbol{\Theta} = \begin{bmatrix} a_1 \\ \vdots \\ a_{n_A} \\ b_0 \\ \vdots \\ b_{n_B} \end{bmatrix}, \quad \boldsymbol{\varphi} = \begin{bmatrix} -y(k-1) \\ \vdots \\ -y(k-n_A) \\ u(k-d) \\ \vdots \\ u(k-d-n_B) \end{bmatrix}, \quad y(k) = \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k).$$

Prädiktor:

$$\hat{y}(k) = \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)$$

$\rightarrow$ Bester Prädiktor im statistischen Sinne, wenn $\{\zeta(k)\}$ eine Sequenz von unabhängigen Zufallsgrößen ist, die unabhängig von den Beobachtungen $\boldsymbol{\varphi}(k)$ ist und einen Mittelwert von Null sowie eine endliche Varianz hat.

Beweis:

Gesucht ist der Prädiktor $\hat{y}(k)$, der den Erwartungswert des quadratischen Fehlers (Engl.: mean square error) minimiert:

$$\hat{y}(k) \quad = \quad \arg\min_{\hat{y}} \mathcal{E}\{(y(k) - \hat{y}(k))^2\}$$

Ersetzen von $y(k)$ durch $\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k)$ ergibt:

$$\begin{aligned}
\mathcal{E}\{(y(k) - \hat{y}(k))^2\} &= \mathcal{E}\{(\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k) - \hat{y}(k))^2\} \\
&= \mathcal{E}\{(\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) - \hat{y}(k))^2 + \zeta(k)^2 + 2\zeta(k)(\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) - \hat{y}(k))\} \\
&= \mathcal{E}\{(\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) - \hat{y}(k))^2\} + \mathcal{E}\{\zeta(k)^2\} + \mathcal{E}\{2\zeta(k)(\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) - \hat{y}(k))\}
\end{aligned}$$

Wenn $\{\zeta(k)\}$ vonabhängig von $\{\boldsymbol{\varphi}(k)\}$ ist, folgt

$$\mathcal{E}\{2\zeta(k)(\boldsymbol{\Theta}^T\boldsymbol{\varphi}(k)-\hat{y}(k))\} = 2\mathcal{E}\{\zeta(k)\}\mathcal{E}\{(\boldsymbol{\Theta}^T\boldsymbol{\varphi}(k)-\hat{y}(k))\}$$

Da $\{\zeta(k)\}$ eine Sequenz unabhängiger Zufallsgrößen mit Erwartungswert Null ist, folgt

$$2\mathcal{E}\{\zeta(k)\}\mathcal{E}\{(\boldsymbol{\Theta}^T\boldsymbol{\varphi}(k)-\hat{y}(k))\} = 0.$$

Dies führt zu

$$\mathcal{E}\{(y(k)-\hat{y}(k))^2\} = \mathcal{E}\{(\boldsymbol{\Theta}^T\boldsymbol{\varphi}(k)-\hat{y}(k))^2\} + \mathcal{E}\{\zeta(k)^2\}$$

und somit folgt

$$\hat{y}(k) \quad = \quad \arg\min_{\hat{y}} \mathcal{E}\{(y(k)-\hat{y}(k))^2\} = \boldsymbol{\Theta}^T\boldsymbol{\varphi}(k)$$

sowie

$$\min \mathcal{E}\{(y(k)-\hat{y}(k))^2\} = \mathcal{E}\{\zeta(k)^2\} \text{ (Varianz des Messrauschens).}$$

$\square$

2.2 Methode der kleinsten Fehlerquadrate (Least Squares)

- Methode zur Schätzung von linearen Systemparametern
- *Modell:* $y(k) = \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k) + \zeta(k)$
- *Ziel:* $\boldsymbol{\Theta}$ soll aus den Beobachtungen $y(k)$,$\boldsymbol{\varphi}(k)$, $k = 1, 2, 3, \ldots, K$ bestimmt werden.
- *Annahme:* $\boldsymbol{\Theta}$ ist unabhängig von $\boldsymbol{\varphi}(k)$
- *Vorgehen:* Bestimmung des Parametervektors, der die Summe der Fehlerquadrate minimiert.
- *Gütekriterium:*

$$J(\boldsymbol{\Theta}) = \frac{1}{K} \sum_{k=1}^{K} \alpha_k \left[y(k) - \underbrace{\boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)}_{\hat{y}(k)} \right]^2, \quad \alpha_k \text{ Wichtungsfaktor}$$

- *Bemerkung:* $J \Rightarrow$ quadratische Funktion $\Rightarrow$ Form einer Hyperparabel $\Rightarrow$ ein optimaler Punkt $\Rightarrow$ analytische Lösung durch Nullsetzen der 1. Ableitung und überprüfen der 2. Ableitung von J

2.2.1 Herleitung der Lösung

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} J(\boldsymbol{\Theta}) \text{ mit } J(\boldsymbol{\Theta}) = \frac{1}{K}\sum_{k=1}^{K} \alpha_k \left[y(k) - \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)\right]^2$$

1. Ableitung

$$\begin{aligned}
\frac{\partial J}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}\frac{1}{K}\sum_{k=1}^{K} \alpha_k \left[y(k) - \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)\right]^2 = \frac{1}{K}\sum_{k=1}^{K} \alpha_k \frac{\partial}{\partial \Theta_i}\left[y(k) - \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)\right]^2 \\
&= \frac{1}{K}\sum_{k=1}^{K} \alpha_k \left\{2\left[y(k) - \boldsymbol{\Theta}^T \boldsymbol{\varphi}(k)\right]\left[-\varphi_i(k)\right]\right\} \\
&= \frac{2}{K}\left[\boldsymbol{\Theta}^T \sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\varphi_i(k) - \sum_{k=1}^{K} \alpha_k y(k)\varphi_i(k)\right]
\end{aligned}$$

2. Ableitung

$$\frac{\partial^2 J}{\partial \Theta_{i,j}^2} = \frac{2}{K}\sum_{k=1}^{K} \alpha_k \varphi_j(k)\varphi_i(k)$$

Darstellung der ersten Ableitung in Vektorform:

$$\left(\frac{\partial J}{\partial \boldsymbol{\Theta}}\right)^T = \frac{2}{K}\left[\boldsymbol{\Theta}^T \sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k) - \sum_{k=1}^{K} \alpha_k y(k)\boldsymbol{\varphi}^T(k)\right]$$

Transponieren um Spaltenvektor zu bekommen: $((\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T)$

$$\frac{\partial J}{\partial \boldsymbol{\Theta}} = \frac{2}{K}\left[\left[\sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)\right]\boldsymbol{\Theta} - \sum_{k=1}^{K} \alpha_k y(k)\boldsymbol{\varphi}(k)\right]$$

Nullsetzen der 1. Ableitung:

$$\frac{\partial J}{\partial \boldsymbol{\Theta}} = 0$$

$$\frac{2}{K}\left[\left[\sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)\right]\hat{\boldsymbol{\Theta}} - \sum_{k=1}^{K} \alpha_k y(k)\boldsymbol{\varphi}(k)\right] = 0$$

$$\hat{\boldsymbol{\Theta}} = \left[\sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)\right]^{-1} \sum_{k=1}^{K} \alpha_k y(k)\boldsymbol{\varphi}(k)$$

Die Inverse der Matrix $\left[\sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)\right]$ muss existieren!

Die 2. Ableitung in Vektorform:

$$\frac{\partial^2 J}{\partial \boldsymbol{\Theta}^2} = \left[\frac{\partial^2 J}{\partial \Theta_i \partial \Theta_j}\right]_{i,j} = \left[\frac{2}{K}\sum_{k=1}^{K} \alpha_k \varphi_j(k)\varphi_i(k)\right]_{i,j} = \frac{2}{K}\sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)$$

Damit das Optimum $\hat{\boldsymbol{\Theta}}$ ein Minimum ist, muss die Matrix $\dfrac{\partial^2 J}{\partial \boldsymbol{\Theta}^2}$ positiv definit sein.

2.2.2 Matrixdarstellung

Zusammenfassung aller Beobachtungen des Eingangs des linearen Regressionsmodells in einer $K \times I$ Matrix

$$\boldsymbol{\Phi} = \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(K) \end{bmatrix} = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \cdots & \varphi_I(1) \\ \varphi_1(2) & \varphi_2(2) & \cdots & \varphi_I(2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(K) & \varphi_2(K) & \cdots & \varphi_I(K) \end{bmatrix}$$

Beobachtungen des Ausgangs und Messrauschen als Vektoren:

$$\begin{aligned} \boldsymbol{y} &= \begin{bmatrix} y(1) & y(2) & \cdots & y(K) \end{bmatrix}^T \\ \boldsymbol{\zeta} &= \begin{bmatrix} \zeta(1) & \zeta(2) & \cdots & \zeta(K) \end{bmatrix}^T \end{aligned}$$

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Theta} + \boldsymbol{\zeta}$$

Gütefunktional in Vektorform für $\alpha_k = 1$

$$\begin{aligned} J(\boldsymbol{\Theta}) &= \frac{1}{K}(\boldsymbol{y} - \boldsymbol{\Phi\Theta})^T(\boldsymbol{y} - \boldsymbol{\Phi\Theta}) \\ &= \frac{1}{K}\left(\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Phi\Theta} - (\boldsymbol{\Phi\Theta})^T\boldsymbol{y} + \boldsymbol{\Theta}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi\Theta}\right) \\ &= \frac{1}{K}\left(\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\Theta}^T\boldsymbol{\Phi}^T\boldsymbol{y} + \boldsymbol{\Theta}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi\Theta}\right) \end{aligned}$$

Ableitung nach dem Parametervektor und zu Null setzen:

$$\frac{\partial J(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \frac{1}{K}\left(-2\boldsymbol{\Phi}^T\boldsymbol{y} + 2\boldsymbol{\Phi}^T\boldsymbol{\Phi\Theta}\right) = 0$$

$$\hat{\boldsymbol{\Theta}} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{y} \tag{1}$$

Die Matrix $\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)$ entspricht der 2. Ableitung des Gütefunktionals (mit Faktor $2/K$) und wird Hessian-Matrix genannt. Die Matrix muss invertierbar sein!

- Es muss mindestens I Messungen geben!
- Die Elemente in $\mathbf{\Phi}$ entsprechen den gewählten Eingangssignalen.
- Ein konstantes Eingangssignal führt oft zu einer singulären Matrix $\mathbf{\Phi}^T\mathbf{\Phi}$.
- Eingangssignale müssen das zu untersuchende System ausreichend anregen, damit $\mathbf{\Phi}^T\mathbf{\Phi}$ nicht singulär wird.
- Ein gutes Maß für den "Abstand" einer Matrix zur Singularität derselbigen ist die *Konditionierungszahl* der Matrix $\mathbf{\Phi}^T\mathbf{\Phi}$.
- Die Konditionierungszahl ist das Verhältnis des größten zum kleinsten Singulärwert der Matrix $\mathbf{\Phi}^T\mathbf{\Phi}$ (Singulärwerte – Wurzeln der Eigenwerte der Matrix $(\mathbf{\Phi}^T\mathbf{\Phi})^T(\mathbf{\Phi}^T\mathbf{\Phi})$).
- Größere Konditionierungszahlen bedeuten einen kleineren "Abstand" zur Singularität der Matrix.

2.2.3 Beispiel - Least Squares in Matlab

System:

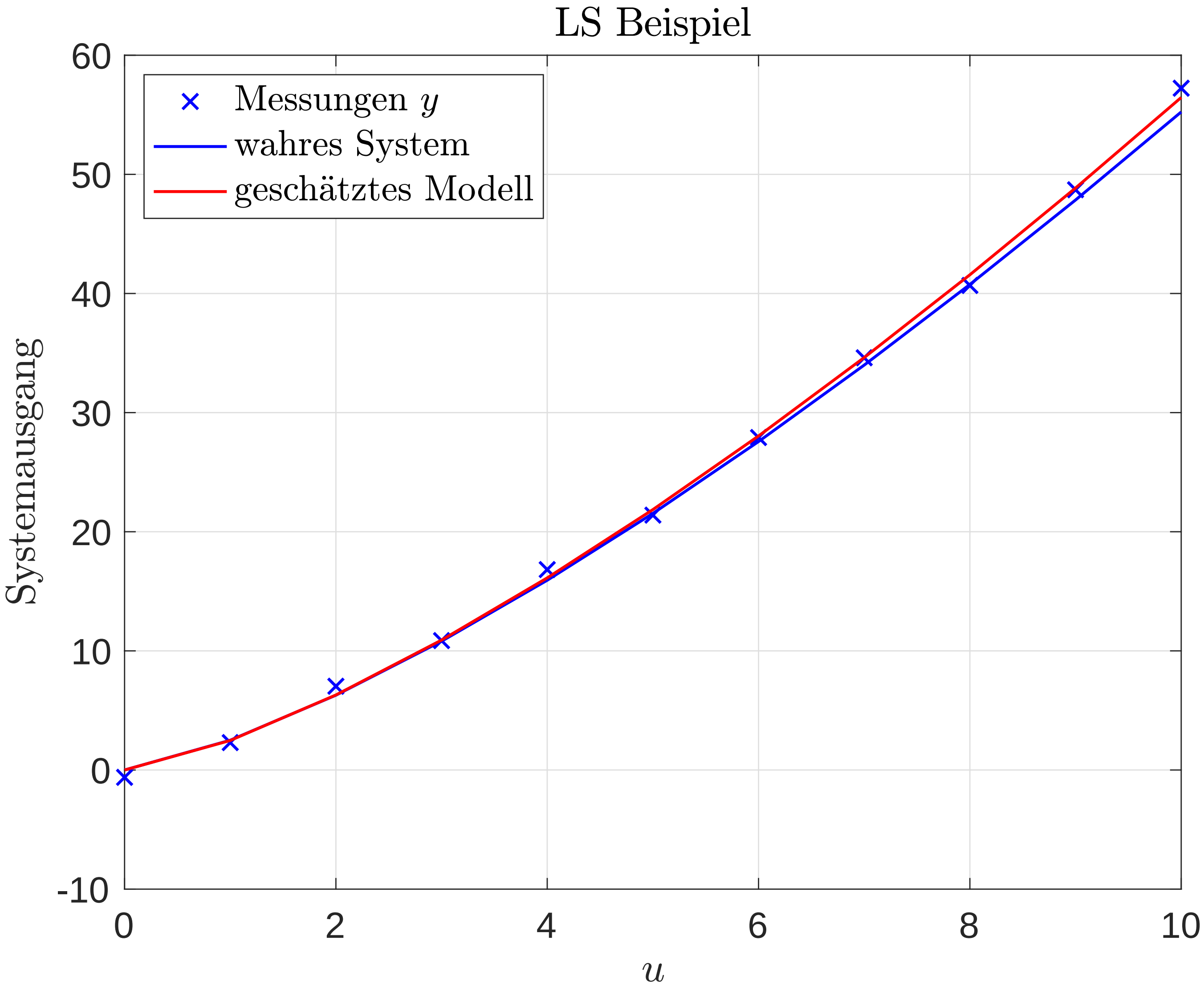
$$y(k) = \Theta_1 u(k) + \Theta_2 u(k)^{1,4} + \zeta(k)$$

Gesucht ist die Schätzung $\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\Theta}_1 & \hat{\Theta}_2 \end{bmatrix}^T$ für den Parametervektor.

MATLAB-Lösung für *Least Squares*

```
>> u=(0:10)'; K=length(u);
Theta1=0.5;Theta2=2;
y=Theta1*u+Theta2*u.^1.4+randn(K,1);%wahres System + Messrauschen
Phi=[u u.^1.4];
Theta_hat=Phi\y %Least Squares Loesung (vgl. (1))
y_hat=Phi*Theta_hat;
Theta_hat =
    0.3866
    2.0935
```

△



Achtung: Mit dem backslash (left devision) bekommt man auch im Fall $K < I$ eine Lösung. Bei einem unterbestimmten LS-Problem ist die Lösung jedoch nicht eindeutig.

```
figure(1); %Neues Figure mit Nummer 1
clf(1); %Loeschen von bereits vorhandenem Figure-Inhalt
plot(u,y,'bx','MarkerSize',9,'linewidth',1.2);% Messungen
set(gca,'FontSize',15);%Set Default Fontsize
                       %(gca - get handle to current axis)
hold on; %vorhandenen Figure-Inhalt ab hier nicht mehr loeschen
plot(u,Theta1*u+Theta2*u.^1.4,'b','Linewidth',1.2); %wahres System M
plot(u,y_hat,'r','Linewidth',1.2); %geschaetztes Modell N
l=legend('Messungen $y$','wahres System',...
    'gesch\"atztes Modell',...
    'location','nw','interpreter','latex');
l.FontSize = 15;
h=xlabel('$u$','interpreter','latex');
ylabel('Systemausgang','interpreter','latex');
title('LS Beispiel','interpreter','latex'); grid on;
%EPS vom Figure generieren
print -deps2c ls_example_plot.eps
%Aufruf eines Systembefehls zum Umwandeln des EPS in ein PDF
!pstopdf ls_example_plot.eps  %epstopdf on Linux systems
```

2.2.4 Eigenschaften des Least-Squares-Schätzers

- Die Messgrößen sind gestört durch Zufallsgrößen, daher wird auch die Schätzung eine Zufallsgröße sein.
- *Biasfreie Schätzung:* Erwartungswert von $\hat{\mathbf{\Theta}}$ entspricht dem wahren Parametervektor $\mathbf{\Theta}$ (d.h. bei unendlich vielen Messungen).
- Hinreichende Voraussetzung für eine biasfreie LS-Schätzung:

  (a) $\mathcal{E}\{\boldsymbol{\zeta}\} = \mathbf{0}$

  (b) $\boldsymbol{\zeta}$ und $\mathbf{\Phi}$ sind statistisch unabhängig

Herleitung:

- Schätzfehler $\tilde{\mathbf{\Theta}} = \mathbf{\Theta} - \hat{\mathbf{\Theta}}$
- Erwartungswert des Schätzfehlers:

$$\mathcal{E}\{\tilde{\mathbf{\Theta}}\} \quad = \quad \mathcal{E}\{\mathbf{\Theta} - \left(\mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \boldsymbol{y}\}$$

$$
\begin{aligned}
\mathcal{E}\{\tilde{\mathbf{\Theta}}\} &= \mathcal{E}\{\mathbf{\Theta} - \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\left[\mathbf{\Phi}\mathbf{\Theta} + \boldsymbol{\zeta}\right]\} \\
&= -\mathcal{E}\{\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\}\mathcal{E}\{\boldsymbol{\zeta}\},
\end{aligned}
$$

da $\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{\Phi} = \boldsymbol{I}$ und $\boldsymbol{\zeta}$, $\mathbf{\Phi}$ statistisch unabhängig.

Wenn $\mathcal{E}\{\boldsymbol{\zeta}\} = \mathbf{0}$, dann gilt $\mathcal{E}\{\tilde{\mathbf{\Theta}}\} = \mathbf{0}$ und somit $\mathcal{E}\{\hat{\mathbf{\Theta}}\} = \mathbf{\Theta}$.

- Kovarianz des Schätzfehlers (Streuung von $\tilde{\mathbf{\Theta}}$ um den Erwartungswert):

$$\begin{aligned}
\boldsymbol{P} &= \mathcal{E}\{(\tilde{\mathbf{\Theta}} - \mathcal{E}\{\tilde{\mathbf{\Theta}}\})(\tilde{\mathbf{\Theta}} - \mathcal{E}\{\tilde{\mathbf{\Theta}}\}^T)\} = \mathcal{E}\{\tilde{\mathbf{\Theta}}\tilde{\mathbf{\Theta}}^T\} \\
&= \mathcal{E}\{[\mathbf{\Theta} - \hat{\mathbf{\Theta}}][\mathbf{\Theta} - \hat{\mathbf{\Theta}}]^T\} \\
&= \mathcal{E}\left\{\left[\mathbf{\Theta} - \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\left[\mathbf{\Phi}\mathbf{\Theta} + \boldsymbol{\zeta}\right]\right] \times \left[\mathbf{\Theta} - \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\left[\mathbf{\Phi}\mathbf{\Theta} + \boldsymbol{\zeta}\right]\right]^T\right\} \\
&= \mathcal{E}\left\{\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\boldsymbol{\zeta}\boldsymbol{\zeta}^T\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\right\} \\
&= \mathcal{E}\left\{\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\right\}\mathcal{E}\{\boldsymbol{\zeta}\boldsymbol{\zeta}^T\} = \mathcal{E}\left\{\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\right\}\sigma_\zeta^2 \\
&= \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\sigma_\zeta^2 \qquad (2)
\end{aligned}$$

  - σ_ζ^2 - Varianz von $\boldsymbol{\zeta}$ (Messrauschen)
  - $\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}$ ist bekannt!

- Die Kovarianzmatrix $\boldsymbol{P}$ ist ein Maß für die Genauigkeit der Schätzung, wobei $\sqrt{P_{i,i}}$ den Standardfehler (-abweichung) für jedes Element $\hat{\Theta}_i$ des geschätzten Parametervektors $\hat{\mathbf{\Theta}}$ darstellt.

- Schätzung der Varianz von $\boldsymbol{\zeta}$:

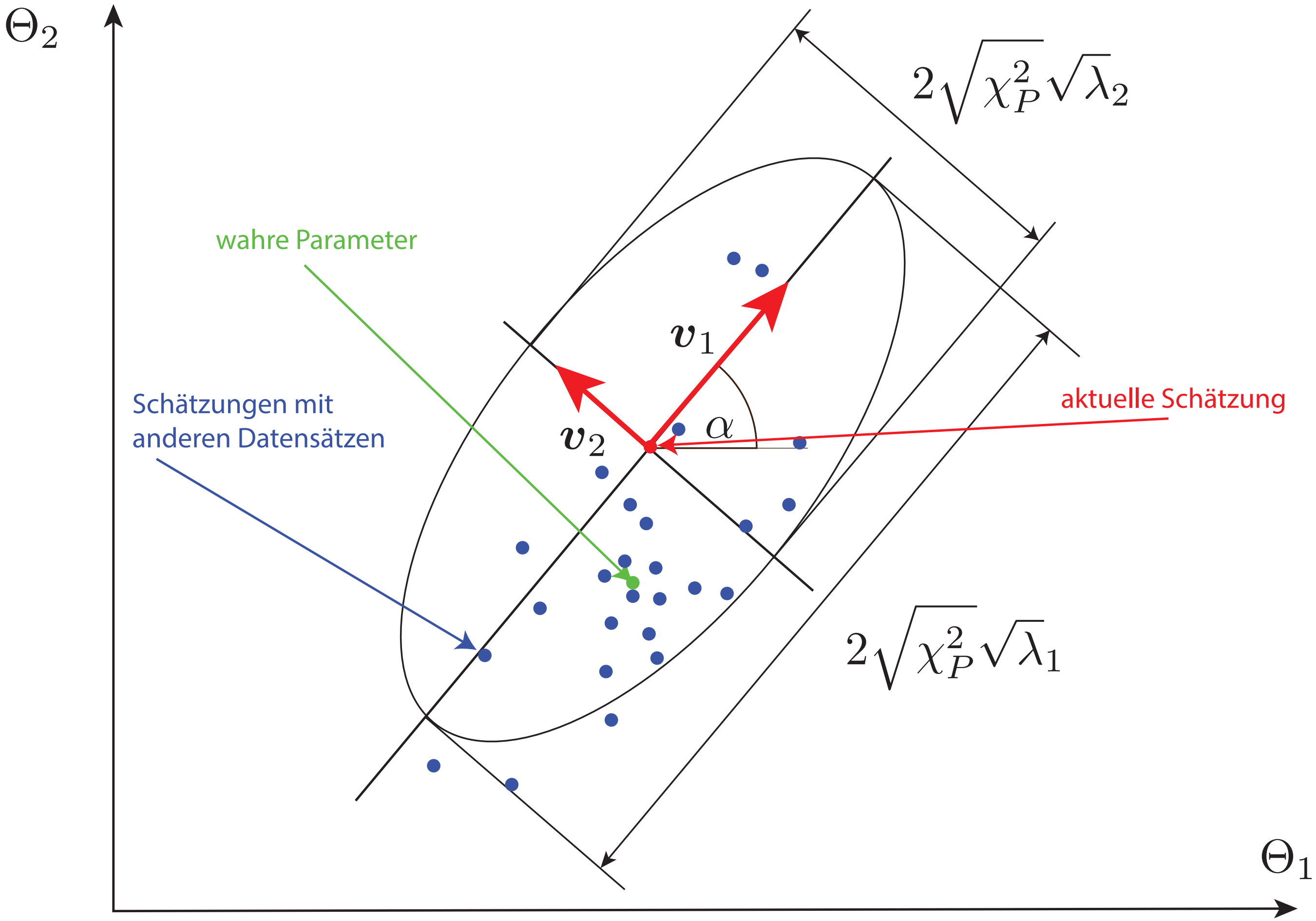
$$\hat{\sigma}_{\zeta}^{2} = \frac{\left[y - \mathbf{\Phi}\hat{\mathbf{\Theta}}\right]^{T}\left[y - \mathbf{\Phi}\hat{\mathbf{\Theta}}\right]}{K - I} \tag{3}$$

- Ansätze zur Verkleinerung der Kovarianz des Schätzfehlers:
  - Mehr Messungen verwenden, da die Kovarianz proportional zu $1/K$ ist. Durch genügend Daten kann jedes Rauschen kompensiert werden.
  - Konditionierungszahl der Matrix $\mathbf{\Phi}^{T}\mathbf{\Phi}$ verkleinern durch mehr anregende Testsignale.

2.2.5 Interpretation der Varianz und Kovarianzmatrix – Konfidenzintervalle und -bereiche

- Normalverteilte univariate Zufallsgröße mit der Varianz σ^2
  - 68% Konfidenzintervall $\rightarrow \pm 1\sigma$
  - 95 % Konfidenzintervall $\rightarrow \pm 2\sigma$
  - 99,7 % Konfidenzintervall $\rightarrow \pm 3\sigma$
- Kovarianzmatrix (Beispiel 2x2) $P = \begin{bmatrix} \sigma^2_{\Theta_1} & \sigma^2_{\Theta_1\Theta_2} \\ \sigma^2_{\Theta_1\Theta_2} & \sigma^2_{\Theta_2} \end{bmatrix}$
  - Konfidenzbereich ist eine Ellipse.
  - λ_1, λ_2 und $\boldsymbol{v}_1, \boldsymbol{v}_2$ sind die Eigenwerte und entsprechenden Eigenvektoren.
  - Konfidenzbereich ergibt sich aus der Chi-Quadrat-Verteilung (für zwei Freiheitsgrade), in dem man zunächst χ^2_P bestimmt:

$P[\%]$	$99,5$	99	$95,7$	95	90
χ^2_P	$10,6$	$9,21$	$7,38$	$5,99$	$4,66$



Der gezeigte Konfidenzbereich gehört zur aktuellen (roten) Schätzung. Der grüne Punkt stellt die wahren Parameter dar. Die anderen blauen Punkte zeigen beispielhaft Schätzergebnisse mit anderen gleich großen Datensätzen, die jedoch immer andere Realisierungen des Messrauschen beinhalten. Die wahren Parameterwerte liegen zu P [%] im dargestellten Konfidenzbereich.

Für mehr als zwei Parameter ($I > 2$) ist der P [%]-Konfidenzbereich ein I-dimensionales Ellipsoid

$$\left(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}\right)^T \boldsymbol{P}^{-1} \left(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}\right) \leq \chi_P^2, \tag{4}$$

wobei χ_P^2 das P-Perzentil einer χ^2-Verteilung mit I Freiheitsgraden ist.

2.2.6 Konfidenzinterval für den prädizierten Systemausgang

Die Kovarianzmatrix des Prädiktionsfehlers für einen prädizierten Ausganges $\hat{\boldsymbol{y}} = \boldsymbol{\Phi}_{\text{neu}}\hat{\boldsymbol{\Theta}}$ mit neuer Regressionsmatrix $\boldsymbol{\Phi}_{\text{neu}}$ (andere Eingangsdaten als bei der Schätzung) ergibt sich für einen mit der Regressionsmatrix $\boldsymbol{\Phi}_{\text{ident}}$ geschätzten Parametervektor $\hat{\boldsymbol{\Theta}}$ aus

$$\begin{aligned}
\boldsymbol{P}_{\hat{\boldsymbol{y}}} = \mathcal{E}\left\{(\hat{\boldsymbol{y}} - \mathcal{E}\{\hat{\boldsymbol{y}}\})(\hat{\boldsymbol{y}} - \mathcal{E}\{\hat{\boldsymbol{y}}\})^T\right\} &= \mathcal{E}\left\{(\boldsymbol{\Phi}_{\text{neu}}(\hat{\boldsymbol{\Theta}} - \mathcal{E}\{\hat{\boldsymbol{\Theta}}\}))(\boldsymbol{\Phi}_{\text{neu}}(\hat{\boldsymbol{\Theta}} - \mathcal{E}\{\hat{\boldsymbol{\Theta}}\}))^T\right\} \\
&= \boldsymbol{\Phi}_{\text{neu}}\mathcal{E}\left\{(\hat{\boldsymbol{\Theta}} - \mathcal{E}\{\hat{\boldsymbol{\Theta}}\})(\hat{\boldsymbol{\Theta}} - \mathcal{E}\{\hat{\boldsymbol{\Theta}}\})^T\right\}\boldsymbol{\Phi}_{\text{neu}}^T \\
&= \boldsymbol{\Phi}_{\text{neu}}\mathcal{E}\left\{(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^T\right\}\boldsymbol{\Phi}_{\text{neu}}^T \\
&= \boldsymbol{\Phi}_{\text{neu}}\underbrace{\mathcal{E}\left\{(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})^T\right\}}_{\boldsymbol{P}_{\text{ident}}}\boldsymbol{\Phi}_{\text{neu}}^T
\end{aligned}$$

mit $\boldsymbol{P}_{\text{ident}} = \left(\boldsymbol{\Phi}_{\text{ident}}^T\boldsymbol{\Phi}_{\text{ident}}\right)^{-1}\sigma_\zeta^2$ (vgl. Gleichung (2)). σ_ζ^2 kann mittels (3) geschätzt werden.

Die Diagonalelemente von $\boldsymbol{P}_{\hat{\boldsymbol{y}}} = \mathcal{E}\left\{(\hat{\boldsymbol{y}} - \mathcal{E}\{\hat{\boldsymbol{y}}\})(\hat{\boldsymbol{y}} - \mathcal{E}\{\hat{\boldsymbol{y}}\})^T\right\}$ repräsentieren die Varianzen des geschätzten Modellausganges $\hat{\boldsymbol{y}}(i)$ für jeden Teilregressionsvektor $\boldsymbol{\varphi}_{\text{neu}}(i)$ der neuen Regressionsmatrix $\boldsymbol{\Phi}_{\text{neu}}$.

Das 68%-Konfidenzintervall (1-sigma) ergibt sich somit wie folgt:

$$\hat{\boldsymbol{y}} \pm \sqrt{\text{diag}(\boldsymbol{P}_{\hat{\boldsymbol{y}}})}$$

Üblich ist auch die Angabe des 95%- Konfidenzintervalls (1,96-sigma):

$$\hat{\boldsymbol{y}} \pm 1,96\sqrt{\text{diag}(\boldsymbol{P}_{\hat{\boldsymbol{y}}})}$$