

# 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

2.2 Methode der kleinsten Fehlerquadrate (Least Squares)

2.2.1 Herleitung der Lösung

2.2.2 Matrixdarstellung

2.2.3 Beispiel - Least Squares in Matlab

2.2.4 Eigenschaften des Least-Squares-Schätzers

2.2.5 Interpretation der Varianz und Kovarianzmatrix – Konfidenzintervalle

2.2.6 Konfidenzintervall für den prädizierten Systemausgang

---

## 2.1 Lineare Regressionsmodelle

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

$$y(k) = \Theta^T \varphi(k) + \zeta(k)$$

$$\Theta^T = [\Theta_1, \dots, \Theta_i, \dots, \Theta_I] \quad \text{Parametervektor}$$

$$\varphi^T = [\varphi_1, \dots, \varphi_i, \dots, \varphi_I] \quad \text{Regressionsvektor}$$

$$\zeta(k) - \text{Rauschen}$$

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

## Beispiele für lineare Regressionsmodelle:

## (a) Statisches System

Was ist eine Regressionsmodell?

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

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

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

ist nicht linear, aber die Parameter bleibt linear

## (b) Dynamisches System (lineare Differenzengleichung)

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

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

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

Das lineare Regressionsmodell lautet in diesem Fall

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

### Prädiktor:

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

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

### Beweis:

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

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

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

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

Wenn  $\{\zeta(k)\}$  <sup>unabhängig</sup> von  $\{\varphi(k)\}$  ist, folgt


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

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

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

Dies führt zu

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

und somit folgt  wir suchen die

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

sowie

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



## 2.2 Methode der kleinsten Fehlerquadrate (Least Squares)

- Methode zur Schätzung von linearen Systemparametern
- *Modell:*  $y(k) = \Theta^T \varphi(k) + \zeta(k)$
- *Ziel:*  $\Theta$  soll aus den Beobachtungen  $y(k), \varphi(k), k = 1, 2, 3, \dots, K$  bestimmt werden.
- *Annahme:*  $\Theta$  ist unabhängig von  $\varphi(k)$
- *Vorgehen:* Bestimmung des Parametervektors, der die Summe der Fehlerquadrate minimiert.
- *Gütekriterium:* Was ist Gütekriterium? Ein Maß für das Regelverhalten einer Regelung. Mit ihr kann eine Aussage über die Qualität der Regelung gemacht werden.

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

- *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

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

1. Ableitung

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

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:

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

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

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

Nullsetzen der 1. Ableitung:

$$\frac{\partial J}{\partial \Theta} = 0 \quad \text{Minimieren}$$

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

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

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

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  $\frac{\partial^2 J}{\partial \boldsymbol{\Theta}^2}$  positiv definit sein.

## 2.2.2 Matrixdarstellung

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

Messung x Parameter

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

Beobachtungen des Ausgangs und Messrauschen als Vektoren:

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

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

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

Vektoren sind in VL immer **fett** geschrieben

Gütefunktional in Vektorform für  $\alpha_k = 1$  Wichtungsfaktor

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

Ableitung nach dem Parametervektor und zu Null setzen:

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

$$\hat{\Theta} = \left( \Phi^T \Phi \right)^{-1} \Phi^T \mathbf{y} \quad \leftarrow \text{Lösung} \tag{1}$$

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

Für Bestimmung der Lösung

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

## 2.2.3 Beispiel - Least Squares in Matlab

System: Ist immer gegeben?

$$y(k) = \Theta_1 u(k) + \Theta_2 u(k)^{1.4} + \zeta(k)$$

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

\_\_\_\_\_MATLAB-Lösung für *Least Squares* \_\_\_\_\_

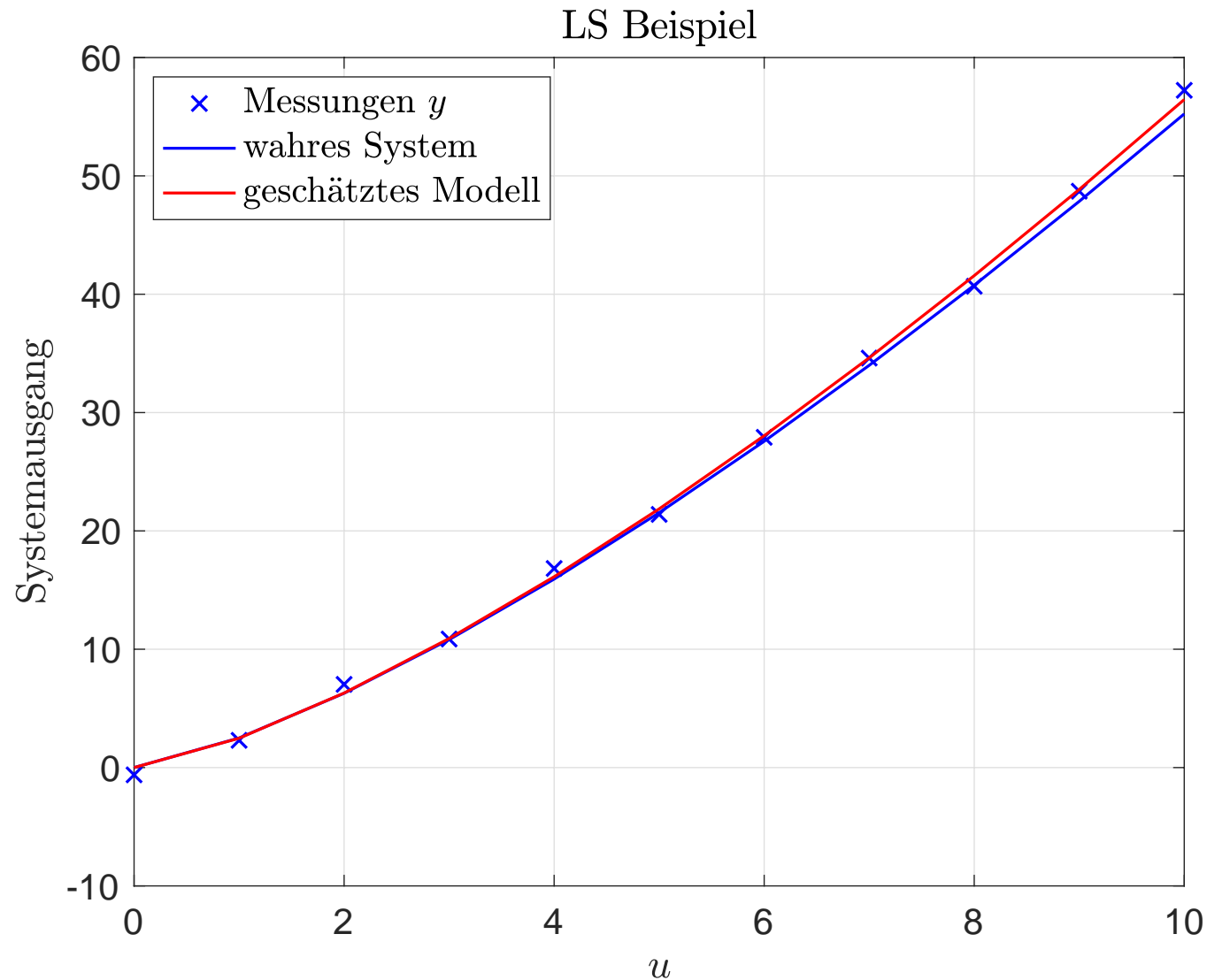
```

1 >> u=(0:10)'; K=length(u);
2 Theta1=0.5; Theta2=2;
3 y=Theta1*u+Theta2*u.^1.4+randn(K,1); %wahres System + Messrauschen
4 Phi=[u u.^1.4];
5 Theta_hat=Phi\y %Least Squares Loesung (vgl. (1))
6 y_hat=Phi*Theta_hat;
7 Theta_hat =
8     0.3866
9     2.0935

```

$$\hat{\Theta} = \left( \Phi^T \Phi \right)^{-1} \Phi^T y \quad \leftarrow \text{Lösung} \quad (1)$$





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



```
1 figure(1); %Neues Figure mit Nummer 1
2 clf(1); %Loeschen von bereits vorhandenem Figure-Inhalt
3 plot(u,y,'bx','MarkerSize',9,'linewidth',1.2);% Messungen
4 set(gca,'FontSize',15);%Set Default Fontsize
5         %(gca - get handle to current axis)
6 hold on; %vorhandenen Figure-Inhalt ab hier nicht mehr loeschen
7 plot(u,Theta1*u+Theta2*u.^1.4,'b','Linewidth',1.2); %wahres System M
8 plot(u,y_hat,'r','Linewidth',1.2); %geschaetztes Modell N
9 l=legend('Messungen $y$', 'wahres System',...
10         'gesch\ "atztes Modell',...
11         'location','nw','interpreter','latex');
12 l.FontSize = 15;
13 h=xlabel('$u$', 'interpreter','latex');
14 ylabel('Systemausgang','interpreter','latex');
15 title('LS Beispiel','interpreter','latex'); grid on;
16 %EPS vom Figure generieren
17 print -deps2c ls_example_plot.eps
18 %Aufruf eines Systembefehls zum Umwandeln des EPS in ein PDF
19 !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  $\hat{\Theta}^\varepsilon$  entspricht dem wahren Parametervektor  $\Theta$  (d.h. bei unendlich vielen Messungen).
- Hinreichende Voraussetzung für eine biasfreie LS-Schätzung:
  - (a)  $\mathcal{E}\{\zeta\} = 0$
  - (b)  $\zeta$  und  $\Phi$  sind statistisch unabhängig

#### Herleitung:

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

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

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

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

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

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

$$\begin{aligned}
 \mathbf{P} &= \mathcal{E}\{(\tilde{\Theta} - \mathcal{E}\{\tilde{\Theta}\})(\tilde{\Theta} - \mathcal{E}\{\tilde{\Theta}\})^T\} = \mathcal{E}\{\tilde{\Theta}\tilde{\Theta}^T\} \\
 &= \mathcal{E}\{[\Theta - \hat{\Theta}][\Theta - \hat{\Theta}]^T\} \\
 &= \mathcal{E}\left\{\left[\Theta - (\Phi^T \Phi)^{-1} \Phi^T [\Phi \Theta + \zeta]\right] \times \left[\Theta - (\Phi^T \Phi)^{-1} \Phi^T [\Phi \Theta + \zeta]\right]^T\right\} \\
 &= \mathcal{E}\left\{\left(\Phi^T \Phi\right)^{-1} \Phi^T \zeta \zeta^T \Phi \left(\Phi^T \Phi\right)^{-1}\right\} \\
 &= \mathcal{E}\left\{\left(\Phi^T \Phi\right)^{-1} \underbrace{\Phi^T \Phi}_{\mathbf{I}} \left(\Phi^T \Phi\right)^{-1}\right\} \underbrace{\mathcal{E}\{\zeta \zeta^T\}}_{\text{da statistisch unabhängig}} = \mathcal{E}\left\{\left(\Phi^T \Phi\right)^{-1}\right\} \sigma_{\zeta}^2 \\
 &= \left(\Phi^T \Phi\right)^{-1} \sigma_{\zeta}^2
 \end{aligned} \tag{2}$$

–  $\sigma_{\zeta}^2$  - Varianz von  $\zeta$  (Messrauschen)

–  $\left(\Phi^T \Phi\right)^{-1}$  ist bekannt!

- Die Kovarianzmatrix  $\mathbf{P}$  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  $\hat{\Theta}$  darstellt.

- Schätzung der Varianz von  $\zeta$ :

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

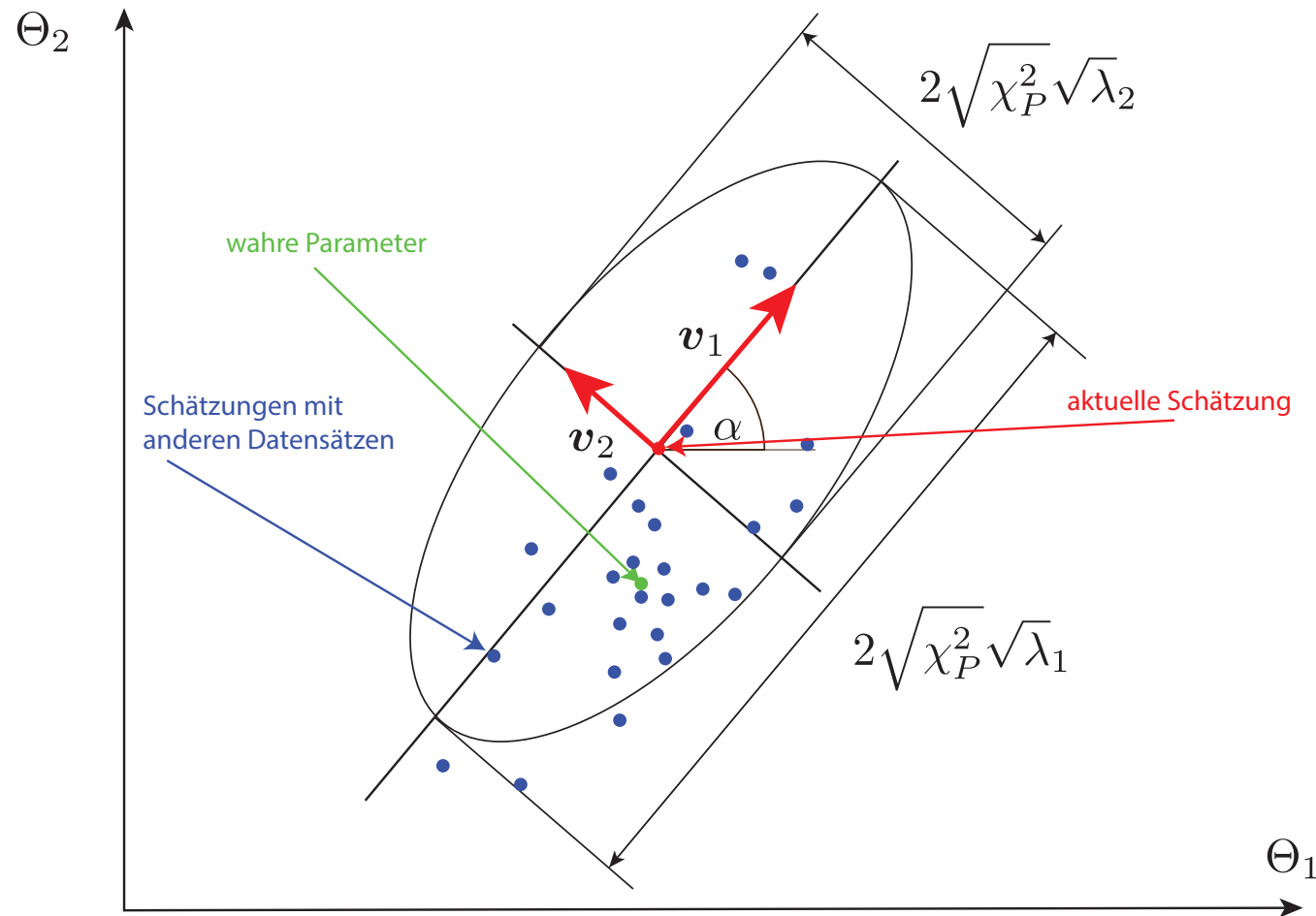
Anzahl der Messung  $K$  - Parameter  $I$

- 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  $\Phi^T \Phi$  verkleinern durch mehr anregende Testsignale.

## 2.2.5 Interpretation der Varianz und Kovarianzmatrix – Konfidenzintervalle und -bereiche

- Normalverteilte univariate Zufallsgröße mit der Varianz  $\sigma^2$ 
  - 68% Konfidenzintervall  $\rightarrow \pm 1\sigma$
  - 95 % Konfidenzintervall  $\rightarrow \pm 2\sigma$
  - 99,7 % Konfidenzintervall  $\rightarrow \pm 3\sigma$
- Kovarianzmatrix (Beispiel 2x2)  $P = \begin{bmatrix} \sigma_{\Theta_1}^2 & \sigma_{\Theta_1 \Theta_2}^2 \\ \sigma_{\Theta_1 \Theta_2}^2 & \sigma_{\Theta_2}^2 \end{bmatrix}$ 
  - Konfidenzbereich ist eine Ellipse.
  - $\lambda_1, \lambda_2$  und  $\mathbf{v}_1, \mathbf{v}_2$  sind die Eigenwerte und entsprechenden Eigenvektoren.
  - Konfidenzbereich ergibt sich aus der Chi-Quadrat-Verteilung (für zwei Freiheitsgrade), in dem man zunächst  $\chi_P^2$  bestimmt:

$P[\%]$	99,5	99	95,7	95	90
$\chi_P^2$	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 Messrauschs 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

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

wobei  $\chi_P^2$  das P-Perzentil einer  $\chi^2$ -Verteilung mit  $I$  Freiheitsgraden ist.



## 2.2.6 Konfidenzintervall für den prädizierten Systemausgang

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

$$\begin{aligned}
 P_{\hat{y}} &= \mathcal{E} \{ (\hat{y} - \mathcal{E}\{\hat{y}\})(\hat{y} - \mathcal{E}\{\hat{y}\})^T \} = \mathcal{E} \{ (\Phi_{\text{neu}}(\hat{\Theta} - \mathcal{E}\{\hat{\Theta}\}))(\Phi_{\text{neu}}(\hat{\Theta} - \mathcal{E}\{\hat{\Theta}\}))^T \} \\
 &= \Phi_{\text{neu}} \mathcal{E} \{ (\hat{\Theta} - \mathcal{E}\{\hat{\Theta}\})(\hat{\Theta} - \mathcal{E}\{\hat{\Theta}\})^T \} \Phi_{\text{neu}}^T \\
 &= \Phi_{\text{neu}} \mathcal{E} \{ (\hat{\Theta} - \hat{\Theta})(\hat{\Theta} - \hat{\Theta})^T \} \Phi_{\text{neu}}^T \\
 &= \underbrace{\Phi_{\text{neu}} \mathcal{E} \{ (\hat{\Theta} - \hat{\Theta})(\hat{\Theta} - \hat{\Theta})^T \}}_{P_{\text{ident}}} \Phi_{\text{neu}}^T
 \end{aligned}$$

↑ ↓ Ist vergleichbar, da es eine Quadrate bilden

Was ist hier ident? Unsere gemessene P

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

Anzahl der Messung K - Parameter I

mit  $P_{\text{ident}} = (\Phi_{\text{ident}}^T \Phi_{\text{ident}})^{-1} \sigma_{\zeta}^2$  (vgl. Gleichung (2)).  $\sigma_{\zeta}^2$  kann mittels (3) geschätzt werden.

Die **Diagonalelemente** von  $\mathbf{P}_{\hat{\mathbf{y}}} = \mathcal{E} \{ (\hat{\mathbf{y}} - \mathcal{E}\{\hat{\mathbf{y}}\})(\hat{\mathbf{y}} - \mathcal{E}\{\hat{\mathbf{y}}\})^T \}$  **repräsentieren** die **Varianzen** des geschätzten Modellausganges  $\hat{\mathbf{y}}(i)$  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:

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

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

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

Warum ist es nicht 2-Sigma?