2 Optimale Schätzer

Dieses Kapitel beschäftigt sich mit der Frage, wie aus einer Eingangsfolge (u_k) und einer gemessenen Ausgangsfolge (y_k) der Zustand $\mathbf{x}_{k+m}$ eines dynamischen Systems geschätzt werden kann. Je nach Wert von m wird der Vorgang des Schätzens als

(1) Glätten (smoothing) für $m < 0$

(2) Filtern (filtering) für $m = 0$ oder

(3) Vorhersagen (prediction) für $m > 0$

bezeichnet. Abbildung 2.1 gibt eine grafische Veranschaulichung dieser drei Fälle.

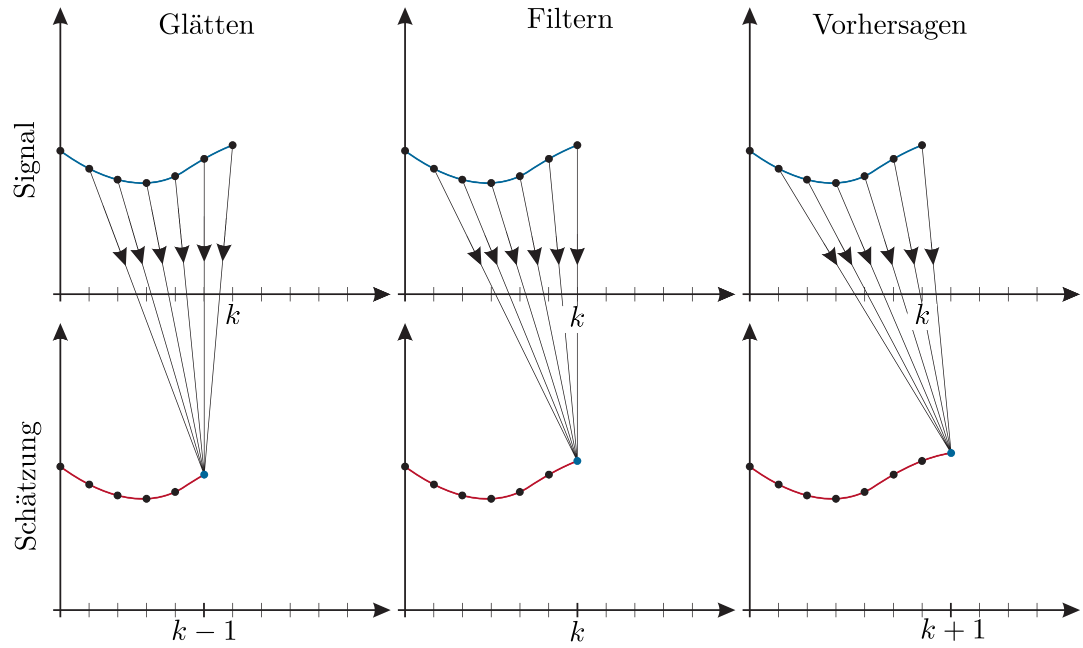

Abbildung 2.1: Zu den Begriffen Glätten, Filtern und Vorhersagen.

Die weiteren Betrachtungen werden sich auf den letzten Fall (3) beschränken, da dieser für regelungstechnische Anwendungen am interessantesten ist. Als Ergebnis der nachfolgenden Betrachtungen wird ein optimaler Zustandsbeobachter, das so genannte *Kalman-Filter*, ermittelt werden, der ein quadratisches Gütekriterium minimiert. Dazu müssen jedoch in einem Zwischenschritt die Ergebnisse der Least-Squares Schätzung des vorigen Kapitels erweitert werden.

Da in diesem Kapitel immer wieder der Erwartungswert und die Kovarianz von Zufallszahlen verwendet werden, sollen diese beiden Begriffe für normalverteilte und gleichverteilte Zufallszahlen erläutert werden.

Bemerkung 2.1 (Normalverteilte Zufallsvariablen). Eine skalare normalverteilte Zufallsvariable x ist durch die Verteilungsdichtefunktion (Gauß'sche Verteilung)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{2.1}$$

mit dem Mittelwert (Erwartungswert) m und der Varianz σ, definiert. Der Mittelwert (Erwartungswert, erstes Moment) und die Varianz (zweites zentrales Moment) sind, wie im Anhang A erläutert, durch

$$\mathrm{E}(x) = m = \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x \tag{2.2a}$$

$$\mathrm{E}\Big((x - \mathrm{E}(x))^2\Big) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mathrm{E}(x))^2 f(x) \, \mathrm{d}x \tag{2.2b}$$

definiert. In Abbildung 2.2 ist die Wahrscheinlichkeitsdichtefunktion für unterschiedliche Parametrierungen einer normalverteilten Zufallszahl dargestellt.

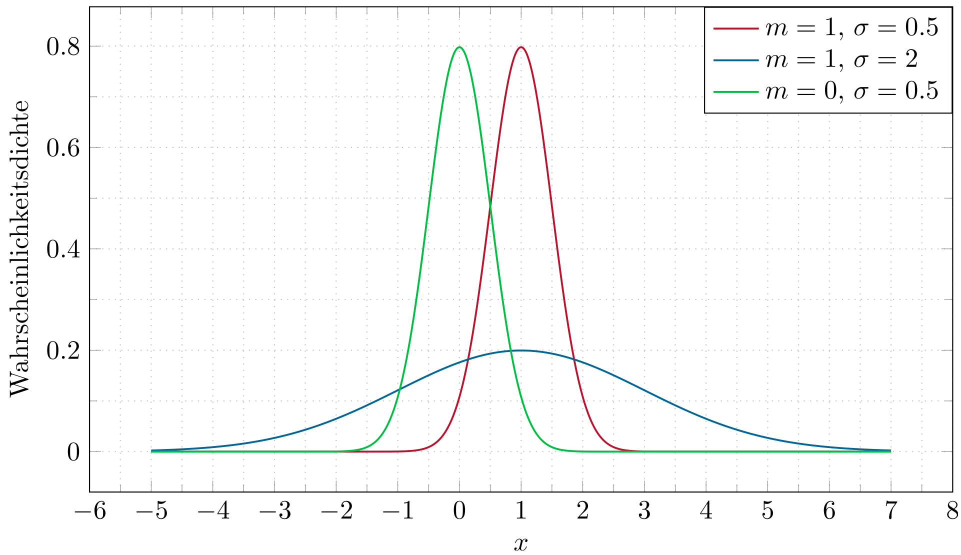

Abbildung 2.2: Wahrscheinlichkeitsdichtefunktion für unterschiedliche normalverteilte Zufallsvariablen.

Die Wahrscheinlichkeit, dass eine Zufallszahl x im Intervall δ um den Erwartungswert $\mathrm{E}(x) = m$ liegt errechnet sich zu

$$\mathrm{P}(m - \delta < x \leq m + \delta) = \int_{m-\delta}^{m+\delta} f(x)\,\mathrm{d}x = \operatorname{erf}\left(\frac{\delta}{\sqrt{2}\sigma}\right), \tag{2.3}$$

vgl. Aufgabe A.1 im Anhang A. Wählt man z.B. $\delta = \sigma$, so liegt eine Zufallszahl x mit einer Wahrscheinlichkeit von 0.68 in diesem Intervall. Weiterhin liegt eine Zufallszahl x mit einer Wahrscheinlichkeit von 0.95 im Intervall $\delta = 2\sigma$. Die Varianz stellt damit ein Maß für die Streuung der Zufallszahlen um den Erwartungswert dar.

Die Verbundwahrscheinlichkeitsdichtefunktion $f(\mathbf{x})$ eines n-dimensionalen normalverteilten Zufallsvektors $\mathbf{x}$ errechnet sich in der Form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{Q})}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{m})}, \tag{2.4}$$

mit dem Erwartungswert $\mathbf{m}$ und der Kovarianzmatrix $\mathbf{Q}$,

$$\mathbf{m} = \mathrm{E}(\mathbf{x}) \tag{2.5a}$$

$$\mathbf{Q} = \mathrm{E}\Big((\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{x} - \mathrm{E}(\mathbf{x}))^{\mathrm{T}}\Big)\ . \tag{2.5b}$$

Für den Sonderfall $n = 2$ ist die Wahrscheinlichkeit, dass ein Zufallsvektor $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ in einer Ellipse der Form

$$(\mathbf{x} - \mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x} - \mathbf{m}) = C^2 \tag{2.6}$$

liegt, durch

$$\mathrm{P} = 1 - \mathrm{e}^{-\frac{C^2}{2}} \tag{2.7}$$

gegeben. Der Beweis für diese Aussage kann z.B. in [2.1] gefunden werden.

Damit kann mit Hilfe der Kovarianzmatrix $\mathbf{Q}$ das Gebiet (d.h. die Ellipsen für $n = 2$) ermittelt werden, in dem ein Zufallsvektor $\mathbf{x}$ mit einer Wahrscheinlichkeit P liegt. In der Abbildung 2.3 ist die Verteilung von 3000 Zufallsvektoren $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$, mit den normalverteilten, nicht korrelierten Zufallsvariablen x_1 ($\mathrm{E}(x_1) = m_1 = 1$, $\sigma_1 = 2$) und x_2 ($\mathrm{E}(x_2) = m_2 = 2$, $\sigma_2 = 1$) dargestellt. Weiterhin sind die Ellipsen für $\mathrm{P} = 0.5$ und $\mathrm{P} = 0.95$, und der Erwartungswert $\mathrm{E}(\mathbf{x}) = \mathbf{m}$ eingezeichnet.

Diese Betrachtung für $n = 2$ kann für n-dimensionale normalverteilte Zufallsvariablen verallgemeinert werden. In diesem Fall beschreibt das n-dimensionale Ellipsoid $(\mathbf{x} - \mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x} - \mathbf{m}) = C^2$ ein Maß für die Verteilung des Zufallsvektors. Die Wahrscheinlichkeit P dafür, dass ein Zufallsvektor $\mathbf{x}$ in diesem Ellipsoid liegt, ist durch

$$1 - \mathrm{P} = \frac{n}{2^{n/2}\Gamma\left(\frac{n}{2} + 1\right)} \int_C^\infty \xi^{n-1}\mathrm{e}^{-\frac{\xi^2}{2}}\,\mathrm{d}\xi, \tag{2.8}$$

mit der Gamma-Funktion Γ, gegeben [2.1].

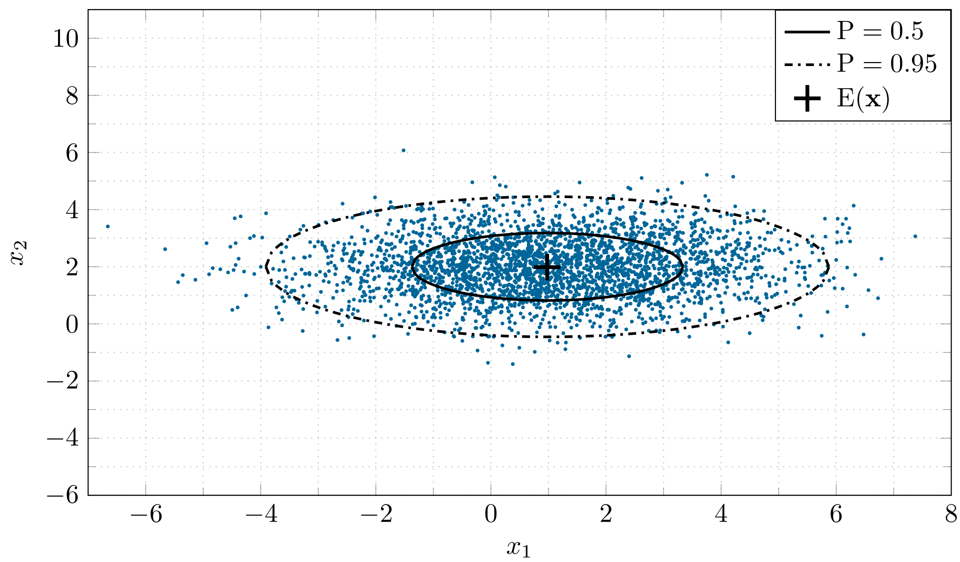

Abbildung 2.3: Grafische Darstellung des Erwartungswertes $\mathrm{E}(\mathbf{x})$ und der Kovarianzmatrix $\mathbf{Q}$ für normalverteilte Zufallsvektoren $\mathbf{x}$.

Ein exakter Zusammenhang zwischen den durch die Kovarianzmatrix $\mathbf{Q}$ definierten Ellipsoiden und der Wahrscheinlichkeit, dass ein Zufallsvektor $\mathbf{x}$ in diesem Gebiet liegt, ist nur für normalverteilte Zufallsvariablen definiert. Für andere Verteilungen stellen diese Ellipsoide nur eine mehr oder weniger genaue Approximation dar. Trotzdem ist die Kovarianzmatrix $\mathbf{Q}$ auch für diese Verteilungen ein sinnvolles Maß zur Abschätzung der Verteilung der Zufallsvektoren, weswegen auch hier häufig die zugehörigen Ellipsen bzw. Ellipsoide dargestellt werden.

Bemerkung 2.2. Eine im Intervall $[a, b]$ gleichverteilte skalare Zufallsvariable x wird durch die Wahrscheinlichkeitsdichtefunktion

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{sonst} \end{cases} \tag{2.9}$$

definiert. Der zugehörige Erwartungwert m und die Varianz σ errechnen sich zu

$$\mathrm{E}(x) = m = \frac{a+b}{2} \tag{2.10a}$$

$$\mathrm{E}\Big((x - \mathrm{E}(x))^2\Big) = \sigma^2 = \frac{1}{12}(b-a)^2 \ . \tag{2.10b}$$

In der Abbildung 2.4 sind die Wahrscheinlichkeitsdichtefunktionen für unterschiedliche gleichverteilte Zufallsvariablen dargestellt.

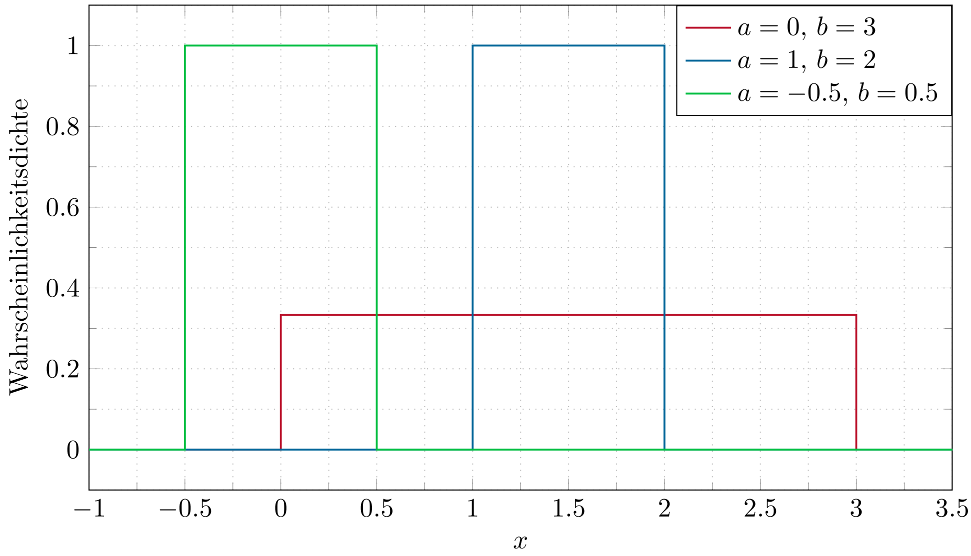

Abbildung 2.4: Wahrscheinlichkeitsdichtefunktion für unterschiedliche gleichverteilte Zufallsvariablen.

2.1 Gauß-Markov-Schätzung

Man betrachte das um die stochastische Störung $\mathbf{v}$ erweiterte, überbestimmte lineare Gleichungssystem von (1.60) (vergleiche dazu auch (1.87)-(1.90))

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.11}$$

mit der bekannten $(m \times n)$-Matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$, dem n-dimensionalen Vektor der Unbekannten $\mathbf{p} \in \mathbb{R}^n$ sowie dem m-dimensionalen Messvektor $\mathbf{y} \in \mathbb{R}^m$. Man nimmt nun an, dass die stochastische Störung $\mathbf{v}$ folgende stochastische Eigenschaften besitzt

$$\mathrm{E}(\mathbf{v}) = \mathbf{0} \quad \text{und} \quad \mathrm{cov}(\mathbf{v}) = \mathrm{E}\Big(\mathbf{v}\mathbf{v}^{\mathrm{T}}\Big) = \mathbf{Q} \quad \text{mit} \quad \mathbf{Q} > 0\ . \tag{2.12}$$

Es sei an dieser Stelle angemerkt, dass $\mathbf{v}$ auch als Messfehler interpretiert werden kann. Gesucht ist nun ein *linearer Schätzer* der Form

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} \tag{2.13}$$

mit einer konstanten $(n \times m)$-Matrix $\mathbf{K} \in \mathbb{R}^{n\times m}$. Da $\mathbf{y}$ die Summe eines konstanten Vektors $\mathbf{S}\mathbf{p}$ und eines stochastischen Vektors (Vektor mit stochastischen Einträgen) $\mathbf{v}$ ist, sind $\mathbf{y}$, der geschätzte Parameter $\hat{\mathbf{p}}$ nach (2.13) und der Parameterfehler $\mathbf{e} = \mathbf{p} - \hat{\mathbf{p}}$ stochastische Größen. Es macht deshalb keinen Sinn, die Matrix $\mathbf{K}$ so zu bestimmen, dass $\|\mathbf{e}\|_2^2$ minimiert wird, sondern es muss die Aufgabe

$$\min_{\mathbf{K}} \mathrm{E}\Big(\|\mathbf{e}\|_2^2\Big) = \min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{y}]\Big) \tag{2.14}$$

gelöst werden.

Setzt man nun für $\mathbf{y}$ die Beziehung (2.11) in (2.14) ein, dann erhält man unter Berücksichtigung von (2.12) und den Ergebnissen von Anhang A (siehe Aufgabe A.3)

$$
\begin{aligned}
&\min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p}-\mathbf{KSp}-\mathbf{Kv}]^{\mathrm{T}}[\mathbf{p}-\mathbf{KSp}-\mathbf{Kv}]\Big) = \\
&\min_{\mathbf{K}} \left\{ \mathrm{E}\Big([\mathbf{p}-\mathbf{KSp}]^{\mathrm{T}}[\mathbf{p}-\mathbf{KSp}]\Big) - 2\underbrace{\mathrm{E}\Big([\mathbf{p}-\mathbf{KSp}]^{\mathrm{T}}\mathbf{Kv}\Big)}_{=0} + \underbrace{\mathrm{E}\Big(\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{Kv}\Big)}_{=\mathrm{spur}(\mathrm{E}(\mathbf{Kvv}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}))} \right\} = \\
&\min_{\mathbf{K}} \Big\{ \|\mathbf{p}-\mathbf{KSp}\|_2^2 + \mathrm{spur}\Big(\mathbf{KQK}^{\mathrm{T}}\Big) \Big\} \ .
\end{aligned}
\tag{2.15}
$$

Die Matrix $\mathbf{K}$, die den Ausdruck (2.15) minimiert, ist offensichtlich eine Funktion des unbekannten Parametervektors $\mathbf{p}$. Daher ist die Lösung dieser Minimierungsaufgabe auch nicht geeignet, einen Schätzer für $\mathbf{p}$ gemäß (2.13) zu liefern. Um dieses Problem zu umgehen, wird im Folgenden eine Ersatzaufgabe gelöst: Man erkennt, dass mit der Wahl

$$
\mathbf{KS} = \mathbf{E} \tag{2.16}
$$

und $\mathbf{E}$ als Einheitsmatrix die Lösung der Minimierungsaufgabe (2.15) unabhängig vom Parameter $\mathbf{p}$ ist. Diese *Nebenbedingung* (2.16) scheint zwar im ersten Augenblick willkürlich zu sein, doch berechnet man den Erwartungswert von $\hat{\mathbf{p}}$ des linearen Schätzers (2.13), dann folgt

$$
\mathrm{E}(\hat{\mathbf{p}}) = \mathrm{E}(\mathbf{Ky}) = \mathrm{E}(\mathbf{KSp}+\mathbf{Kv}) = \mathbf{KS}\underbrace{\mathrm{E}(\mathbf{p})}_{=\mathbf{p}} + \mathbf{K}\underbrace{\mathrm{E}(\mathbf{v})}_{=\mathbf{0}} \ . \tag{2.17}
$$

D.h. die Nebenbedingung (2.16) bringt mit sich, dass $\mathrm{E}(\hat{\mathbf{p}}) = \mathbf{p}$ ist und somit der Schätzer *erwartungstreu* ist.

Die Aufgabe

$$
\min_{\mathbf{K}} \Big\{ \mathrm{spur}\Big(\mathbf{KQK}^{\mathrm{T}}\Big) \Big\} \quad \text{unter der NB} \quad \mathbf{KS} = \mathbf{E} \tag{2.18}
$$

ist nun äquivalent zur Lösung n separater Optimierungsaufgaben der Form

$$
\min_{\mathbf{k}_j} \mathbf{k}_j^{\mathrm{T}}\mathbf{Q}\mathbf{k}_j \quad \text{unter den NBen} \quad \mathbf{k}_j^{\mathrm{T}}\mathbf{S} = \mathbf{e}_j^{\mathrm{T}} \quad \text{für} \quad j = 1,\ldots,n \ . \tag{2.19}
$$

Um dies zu zeigen, schreibe man die Matrix $\mathbf{K}$ und die Einheitsmatrix $\mathbf{E}$ in der Form

$$
\mathbf{K} = \begin{bmatrix} \mathbf{k}_1^{\mathrm{T}} \\ \mathbf{k}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{k}_n^{\mathrm{T}} \end{bmatrix} \quad \text{und} \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}} \\ \mathbf{e}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{e}_n^{\mathrm{T}} \end{bmatrix} \tag{2.20}
$$

an und setzt dies in (2.18) ein

$$
\min_{\mathbf{K}} \Big\{ \mathrm{spur}\Big(\mathbf{KQK}^{\mathrm{T}}\Big) \Big\} = \min_{\mathbf{K}} \left\{ \sum_{j=1}^{n} \mathbf{k}_j^{\mathrm{T}}\mathbf{Q}\mathbf{k}_j \right\} \quad \text{unter den NBen} \quad \mathbf{k}_j^{\mathrm{T}}\mathbf{S} = \mathbf{e}_j^{\mathrm{T}} \tag{2.21}
$$

für $j = 1, \ldots, n$. Da der j-te Summand in (2.21) lediglich von $\mathbf{k}_j$ abhängt, kann die Minimierungsaufgabe von (2.21) durch n Minimierungsprobleme gemäß (2.19) ersetzt werden. Für das Weitere ist also die Lösung einer Aufgabe vom Typ (2.19) für ein festes j von Interesse. Betrachtet man nun den Hilbertraum $\mathcal{H} = \mathbb{R}^m$ mit dem inneren Produkt $\langle \mathbf{x}, \mathbf{z} \rangle = \mathbf{x}^\mathrm{T}\mathbf{z} = \sum_{j=1}^m x_j z_j$, dann erkennt man, dass die Spaltenvektoren $\mathbf{s}_k$, $k = 1, \ldots, n$, der Matrix $\mathbf{S}$, also $\operatorname{span}\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, die die Nebenbedingungen von (2.19) erfüllen, keinen Unterraum des Hilbertraums $\mathcal{H}$ bilden und damit das Projektionstheorem von Satz 1.1 nicht direkt anwendbar ist.

> *Aufgabe* 2.1. Zeigen Sie, dass die Summe $\mathbf{s}_j + \mathbf{s}_k$ die Nebenbedingung von (2.21) nicht erfüllt, selbst wenn $\mathbf{s}_j$ und $\mathbf{s}_k$ jeweils für sich dieser Nebenbedingung genügen.

2.1.1 Quadratische Minimierung mit affinen Nebenbedingungen

Zur Lösung der obigen Aufgabe wird eine Erweiterung des Projektionstheorems von Satz 1.1 benötigt:

Satz 2.1 (Erweiterung des Projektionstheorems)**.** *Es sei $\mathcal{H}$ ein Hilbertraum und $\mathcal{U}$ ein abgeschlossener Unterraum von $\mathcal{H}$. Die translatorische Verschiebung von $\mathcal{U}$ in der Form $\mathcal{A} = \mathbf{x} + \mathcal{U}$ für ein festes $\mathbf{x} \in \mathcal{H}$ wird als* lineare Varietät *oder* affiner Unterraum *bezeichnet. Dann existiert ein eindeutiger Vektor $\mathbf{x}_0 \in \mathcal{A}$ minimaler Norm und dieser ist orthogonal auf $\mathcal{U}$ (siehe Abbildung 2.5).*

> *Beweis von Satz 2.1.* Man verschiebt den affinen Unterraum $\mathcal{A}$ durch $-\mathbf{x}$ so, dass er ein abgeschlossener Unterraum wird und wendet dann das Projektionstheorem von Satz 1.1 an. Man beachte, dass die optimale Lösung $\mathbf{x}_0$ *nicht* orthogonal auf den affinen Unterraum $\mathcal{A}$ sondern orthogonal auf $\mathcal{U}$ ist. □

Bevor nun die Aufgabe von (2.19) gelöst werden kann, sollen noch einige theoretische Grundlagen erläutert werden:

Definition 2.1 (Orthogonales Komplement)**.** Ist $\mathcal{U}$ ein Unterraum eines Hilbertraumes $\mathcal{H}$ mit dem inneren Produkt $\langle \cdot, \cdot \rangle$, dann bezeichnet man die Menge aller Vektoren, die orthogonal zu $\mathcal{U}$ sind, als das *orthogonale Komplement von $\mathcal{U}$* und schreibt dafür $\mathcal{U}^\perp$.

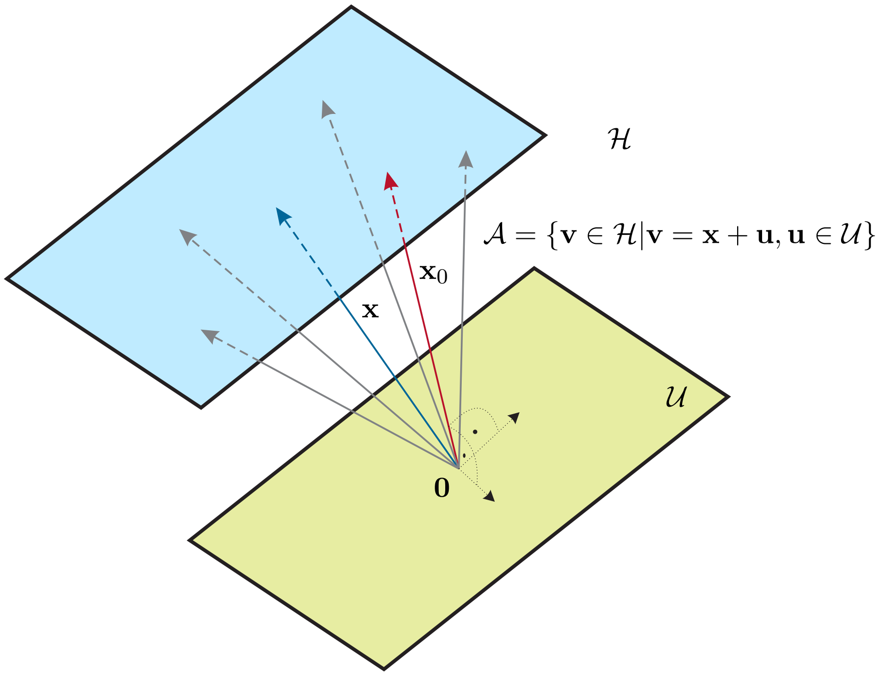

Abbildung 2.5: Zum Projektionstheorem für affine Unterräume.

Für die Unterräume $\mathcal{U}$ und $\mathcal{V}$ eines Hilbertraums gelten nun folgende Eigenschaften:

(1) Das orthogonale Komplement $\mathcal{U}^\perp$ ist ein abgeschlossener Unterraum,

(2) $\mathcal{U} \subseteq \mathcal{U}^{\perp\perp}$,

(3) wenn $\mathcal{U} \subset \mathcal{V}$ ist, dann folgt $\mathcal{V}^\perp \subset \mathcal{U}^\perp$,

(4) $\mathcal{U}^{\perp\perp\perp} = \mathcal{U}^\perp$ und

(5) $\mathcal{U}^{\perp\perp}$ ist der kleinste geschlossene Unterraum der $\mathcal{U}$ beinhaltet.

Beweis zu (1). Da die Linearkombination orthogonaler Vektoren wieder orthogonal ist, folgt unmittelbar, dass $\mathcal{U}^\perp$ ein Unterraum ist. Die Abgeschlossenheit von $\mathcal{U}^\perp$ folgt aus der Tatsache, dass wegen der Stetigkeit des inneren Produktes $\langle \cdot, \cdot \rangle$ für den Grenzwert $\mathbf{x}$ einer konvergenten Folge $(\mathbf{x}_k)$ in $\mathcal{U}^\perp$ gilt

$$\mathbf{0} = \langle \mathbf{y}, \mathbf{x}_k \rangle = \langle \mathbf{y}, \mathbf{x} \rangle \tag{2.22}$$

für alle $\mathbf{y} \in \mathcal{U}$ und damit auch $\mathbf{x} \in \mathcal{U}^\perp$. □

Aufgabe 2.2. Beweisen Sie die obigen Eigenschaften (2) bis (4).

Definition 2.2 (Direkte Summe)**.** Man bezeichnet einen Vektorraum $\mathcal{X}$ als direkte Summe zweier Unterräume $\mathcal{U}$ und $\mathcal{V}$ und schreibt dann $\mathcal{X} = \mathcal{U} \oplus \mathcal{V}$, wenn sich jeder Vektor $\mathbf{x} \in \mathcal{X}$ *eindeutig* als Summe $\mathbf{x} = \mathbf{u} + \mathbf{v}$ mit $\mathbf{u} \in \mathcal{U}$ und $\mathbf{v} \in \mathcal{V}$ darstellen lässt.

Ohne Beweis gilt nun als Folgerung des Projektionstheorems von Satz 1.1 nachfolgender Satz:

Satz 2.2. *Wenn $\mathcal{U}$ ein abgeschlossener linearer Unterraum eines Hilbertraums $\mathcal{H}$ ist, dann gilt $\mathcal{H} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ und $\mathcal{U} = \mathcal{U}^{\perp\perp}$.*

Damit lässt sich nachfolgender Satz zur Lösung der Minimierungsaufgabe mit affinen Nebenbedingungen von (2.19) angeben:

Satz 2.3 (Minimierung mit affinen Nebenbedingung)**.** *Es sei $\mathcal{H}$ ein Hilbertraum mit den linear unabhängigen Vektoren $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$. Unter allen möglichen Vektoren $\mathbf{x} \in \mathcal{H}$, die das affine Gleichungssystem*

$$\begin{aligned} \langle \mathbf{x}, \mathbf{s}_1 \rangle &= c_1 \\ \langle \mathbf{x}, \mathbf{s}_2 \rangle &= c_2 \\ &\vdots \\ \langle \mathbf{x}, \mathbf{s}_n \rangle &= c_n \end{aligned} \tag{2.23}$$

mit den konstanten Koeffizienten $c_1, c_2, \ldots, c_n$ erfüllen, habe der Vektor $\mathbf{x}_0$ die minimale Norm. Dann lässt sich $\mathbf{x}_0$ in der Form

$$\mathbf{x}_0 = \sum_{j=1}^{n} p_{0,j} \mathbf{s}_j = \mathbf{S}\mathbf{p}_0 \tag{2.24}$$

anschreiben, wobei sich $\mathbf{p}_0^T = \begin{bmatrix} p_{0,1} & p_{0,2} & \ldots & p_{0,n} \end{bmatrix}$ aus der Beziehung

$$\underbrace{\begin{bmatrix} \langle \mathbf{s}_1, \mathbf{s}_1 \rangle & \cdots & \langle \mathbf{s}_n, \mathbf{s}_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{s}_1, \mathbf{s}_n \rangle & \cdots & \langle \mathbf{s}_n, \mathbf{s}_n \rangle \end{bmatrix}}_{\mathbf{G}=\mathbf{S}^{\mathrm{T}}\mathbf{S}} \underbrace{\begin{bmatrix} p_{0,1} \\ \vdots \\ p_{0,n} \end{bmatrix}}_{\mathbf{p}_0} = \underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}_{\mathbf{c}} \tag{2.25}$$

mit der Gramschen Matrix $\mathbf{G}$ errechnet.

Beweis zu Satz 2.3. Es sei $\mathcal{U} = \operatorname{span}\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ ein abgeschlossener linearer Unterraum des Hilbertraums $\mathcal{H}$. Die Menge aller möglichen Vektoren $\mathbf{x} \in \mathcal{H}$, die das affine Gleichungssystem (2.23) erfüllen, bilden einen affinen Unterraum von $\mathcal{H}$, nämlich die translatorische Verschiebung von $\mathcal{U}^{\perp}$ um einen Vektor $\boldsymbol{\gamma}$. Da das orthogonale Komplement $\mathcal{U}^{\perp}$ ein abgeschlossener Unterraum ist, kann man Satz 2.1 anwenden und weiß damit, dass die optimale Lösung $\mathbf{x}_0$ existiert und eindeutig ist sowie orthogonal auf $\mathcal{U}^{\perp}$ steht. Damit folgt aber $\mathbf{x}_0 \in \mathcal{U}^{\perp\perp}$ und wegen der Abgeschlossenheit von

> $\mathcal{U}$ und Satz 2.2 erhält man $\mathcal{U}^{\perp\perp} = \mathcal{U}$. Da nun aber $\mathbf{x}_0 \in \mathcal{U}$ gilt, muss sich $\mathbf{x}_0$ als Linearkombination der $\mathbf{s}_j$, $j = 1, \ldots, n$, gemäß (2.24) darstellen lassen. Setzt man nun noch (2.24) in die affine Nebenbedingung (2.23) ein, erhält man das Ergebnis (2.25). □

Wendet man Satz 2.3 auf die Minimierungsaufgabe (2.19) mit $\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_n \end{bmatrix}$ und dem inneren Produkt $\langle \mathbf{x}, \mathbf{z} \rangle_Q = \mathbf{x}^{\mathrm{T}} \mathbf{Q} \mathbf{z}$ im Hilbertraum $\mathcal{H} = \mathbb{R}^m$ an, also

$$\min_{\mathbf{k}_j} \langle \mathbf{k}_j, \mathbf{k}_j \rangle_Q \quad \text{unter den NBen} \quad \left\langle \mathbf{k}_j, \mathbf{Q}^{-1}\mathbf{s}_l \right\rangle_Q = \delta_{jl} = \begin{cases} 1 & \text{für } j = l \\ 0 & \text{sonst} \end{cases} \tag{2.26}$$

für $j = 1, \ldots, n$, dann erhält man für $\mathbf{s}_l$ ersetzt durch $\mathbf{Q}^{-1}\mathbf{s}_l$ in (2.24) und (2.25) das Ergebnis

$$\mathbf{k}_{j,0} = \mathbf{Q}^{-1}\mathbf{S}\mathbf{p}_0 \quad \text{und} \quad \left(\mathbf{Q}^{-1}\mathbf{S}\right)^{\mathrm{T}} \mathbf{Q}\mathbf{Q}^{-1}\mathbf{S}\mathbf{p}_0 = \mathbf{e}_j \tag{2.27}$$

bzw.

$$\mathbf{k}_{j,0} = \mathbf{Q}^{-1}\mathbf{S}\left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} \mathbf{e}_j \; . \tag{2.28}$$

Nach (2.20) errechnet sich die optimale Lösung $\mathbf{K}_0$ für die Matrix $\mathbf{K}$ zu

$$\mathbf{K}_0^{\mathrm{T}} = \mathbf{Q}^{-1}\mathbf{S}\left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} \tag{2.29}$$

und damit lautet der gesuchte lineare *Gauß-Markov-Schätzer* nach (2.13)

$$\hat{\mathbf{p}} = \left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} \mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{y} \; . \tag{2.30}$$

Vergleicht man nun (2.30) mit (1.128), dann erkennt man, dass das Ergebnis identisch zum Ergebnis der Least-Squares Identifikation mit gewichteten kleinsten Quadraten mit der Gewichtungsmatrix $\mathbf{Q}^{-1}$ ist, wobei gilt $\mathbf{Q} = \mathrm{cov}(\mathbf{v})$.

Für den Erwartungswert des Schätzfehlers $\mathbf{e} = \mathbf{p} - \hat{\mathbf{p}}$ erhält man gemäß (2.17) $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ und die Kovarianzmatrix des Schätzfehlers lautet mit (2.16)

$$\begin{aligned} \mathrm{E}\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right) &= \mathrm{E}\left([\mathbf{p} - \mathbf{K}\mathbf{y}][\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}\right) = \mathrm{E}\left([\mathbf{p} - \mathbf{K}(\mathbf{S}\mathbf{p} + \mathbf{v})][\mathbf{p} - \mathbf{K}(\mathbf{S}\mathbf{p} + \mathbf{v})]^{\mathrm{T}}\right) = \\ &= \mathrm{E}\left([\mathbf{K}\mathbf{v}][\mathbf{K}\mathbf{v}]^{\mathrm{T}}\right) = \mathbf{K}\mathbf{Q}\mathbf{K}^{\mathrm{T}} \end{aligned} \tag{2.31}$$

bzw. mit (2.29)

$$\mathrm{E}\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right) = \left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} \mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{S}\left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} = \left(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} . \tag{2.32}$$

In der Literatur wird der lineare Schätzer (2.30) oft auch als *BLUE (best linear unbiased estimate)* bezeichnet.

Ist die Kovarianzmatrix $\mathrm{cov}(\mathbf{v}) = \mathrm{E}\left(\mathbf{v}\mathbf{v}^{\mathrm{T}}\right) = \mathbf{Q}$ eine Diagonalmatrix der Form $\mathbf{Q} = \mathrm{diag}(q_0, q_1, \ldots)$ mit $q_j > 0$ für alle $j \geq 0$, dann kann für den Gauß-Markov-Schätzer (2.30) gemäß der rekursiven Methode der gewichteten kleinsten Quadrate (1.121) unmittelbar mit $\alpha_j = 1/q_j$ eine rekursive Version angegeben werden. Dies klärt auch die Frage, wie die optimale Wahl der Folge positiver Gewichtungskoeffizienten α_j ($\alpha_j > 0$ für alle j) in (1.121) aussieht.

2.2 Minimum-Varianz-Schätzung

Bisher wurde angenommen, dass über den n-dimensionalen Vektor der Unbekannten $\mathbf{p}$ in (2.11) keine Information vorhanden ist. Nun ist aber in manchen Fällen sehr wohl a priori Information über $\mathbf{p}$ in Form von stochastischen Kenngrößen (Erwartungswert, Kovarianzmatrix) bekannt. Daher wird angenommen, dass für das Gleichungssystem (vergleiche dazu auch (2.11))

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.33}$$

mit der stochastischen Störung $\mathbf{v}$, der bekannten $(m \times n)$-Matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$, dem n-dimensionalen Zufallsvektor $\mathbf{p} \in \mathbb{R}^n$ sowie dem m-dimensionalen Messvektor $\mathbf{y} \in \mathbb{R}^m$ gilt

$$\begin{aligned}
\mathrm{E}(\mathbf{v}) &= \mathbf{0}, & \operatorname{cov}(\mathbf{v}) = \mathrm{E}\Big(\mathbf{v}\mathbf{v}^{\mathrm{T}}\Big) &= \mathbf{Q} & \text{mit} \quad \mathbf{Q} \geq 0 \\
\mathrm{E}(\mathbf{p}) &= \mathbf{0}, & \operatorname{cov}(\mathbf{p}) = \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) &= \mathbf{R} & \text{mit} \quad \mathbf{R} \geq 0 \\
& & \mathrm{E}\Big(\mathbf{p}\mathbf{v}^{\mathrm{T}}\Big) &= \mathbf{N}\ .
\end{aligned} \tag{2.34}$$

Im Weiteren sei vorausgesetzt, dass die Matrix $\Big(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}} + \mathbf{Q} + \mathbf{S}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big)$ regulär ist. Gesucht ist nun wiederum ein linearer Schätzer

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} \tag{2.35}$$

mit einer konstanten $(n \times m)$-Matrix $\mathbf{K} \in \mathbb{R}^{n\times m}$ so, dass nachfolgende Minimierungsaufgabe

$$\min_{\mathbf{K}} \mathrm{E}\Big(\|\mathbf{e}\|_2^2\Big) = \min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{y}]\Big) \tag{2.36}$$

gelöst wird. Durch Ausmultiplizieren von (2.36) und mit Hilfe der Beziehung $\operatorname{spur}(\mathbf{K}\mathbf{S}\mathbf{R}) = \operatorname{spur}\Big(\mathbf{R}(\mathbf{K}\mathbf{S})^{\mathrm{T}}\Big)$ erhält man

$$\begin{aligned}
&\min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{y}]\Big) = \\
&\min_{\mathbf{K}} \left\{ \underbrace{\mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]\Big)}_{\operatorname{spur}\left(\mathrm{E}\left([\mathbf{E}-\mathbf{K}\mathbf{S}]\mathbf{p}\mathbf{p}^{\mathrm{T}}[\mathbf{E}-\mathbf{K}\mathbf{S}]^{\mathrm{T}}\right)\right)} - 2\,\underbrace{\mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]^{\mathrm{T}}\mathbf{K}\mathbf{v}\Big)}_{\operatorname{spur}(\mathrm{E}([\mathbf{E}-\mathbf{K}\mathbf{S}]\mathbf{p}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}))} + \underbrace{\mathrm{E}\Big(\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{K}\mathbf{v}\Big)}_{\operatorname{spur}(\mathrm{E}(\mathbf{K}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}))} \right\} = \\
&\min_{\mathbf{K}} \Big\{ \operatorname{spur}\Big([\mathbf{E} - \mathbf{K}\mathbf{S}]\mathbf{R}[\mathbf{E} - \mathbf{K}\mathbf{S}]^{\mathrm{T}} - 2[\mathbf{E} - \mathbf{K}\mathbf{S}]\mathbf{N}\mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{Q}\mathbf{K}^{\mathrm{T}}\Big) \Big\} = \\
&\min_{\mathbf{K}} \Big\{ \operatorname{spur}\Big(\mathbf{K}\Big(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}} + \mathbf{Q} + \mathbf{S}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big)\mathbf{K}^{\mathrm{T}} - 2\mathbf{K}\Big(\mathbf{S}\mathbf{R} + \mathbf{N}^{\mathrm{T}}\Big)\Big) \Big\}\ .
\end{aligned} \tag{2.37}$$

Schreibt man die Matrix $\mathbf{K}$ und die Einheitsmatrix $\mathbf{E}$ wie in (2.20)

$$\mathbf{K} = \begin{bmatrix} \mathbf{k}_1^{\mathrm{T}} \\ \mathbf{k}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{k}_n^{\mathrm{T}} \end{bmatrix} \quad \text{und} \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} \tag{2.38}$$

an, dann folgt (2.37) zu

$$\min_{\mathbf{K}}\left\{\sum_{j=1}^{n}\left(\mathbf{k}_j^{\mathrm{T}}\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\mathbf{k}_j-2\mathbf{k}_j^{\mathrm{T}}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)\mathbf{e}_j\right)\right\}. \tag{2.39}$$

Vergleicht man die Minimierungsaufgabe (2.39) mit der von (1.61), also

$$\min_{\mathbf{p}}(\mathbf{y}-\mathbf{Sp})^{\mathrm{T}}(\mathbf{y}-\mathbf{Sp})=\min_{\mathbf{p}}\left(\mathbf{y}^{\mathrm{T}}\mathbf{y}-2\mathbf{p}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{y}+\mathbf{p}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{Sp}\right), \tag{2.40}$$

dann erkennt man, dass die beiden Aufgaben äquivalent sind. Es lässt sich also die Lösung von (2.39) direkt durch die Lösung von (2.40) (vergleiche dazu (1.63))

$$\mathbf{p}_0=\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{y} \tag{2.41}$$

angeben, indem man in (2.41) $\mathbf{S}^{\mathrm{T}}\mathbf{S}=\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}$ und $\mathbf{S}^{\mathrm{T}}\mathbf{y}=\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)\mathbf{e}_j$ setzt. Damit ergibt sich die optimale Lösung $\mathbf{K}_0$ für die Matrix $\mathbf{K}$ von (2.35) zu

$$\mathbf{K}_0^{\mathrm{T}}=\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right) \tag{2.42}$$

und der gesuchte lineare *Minimum-Varianz-Schätzer* nach (2.35) lautet

$$\hat{\mathbf{p}}=\left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\mathbf{y}\,. \tag{2.43}$$

Man beachte, dass am Beginn dieses Abschnittes vorausgesetzt wurde, dass die Matrix $\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)$ regulär ist. Die Kovarianzmatrix des Schätzfehlers errechnet sich zu

$$\begin{aligned}
\mathrm{cov}(\mathbf{e}) &= \mathrm{E}\left([\mathbf{p}-\mathbf{K}(\mathbf{Sp}+\mathbf{v})][\mathbf{p}-\mathbf{K}(\mathbf{Sp}+\mathbf{v})]^{\mathrm{T}}\right)=\mathrm{E}\left([\mathbf{E}-\mathbf{KS}]\mathbf{pp}^{\mathrm{T}}[\mathbf{E}-\mathbf{KS}]^{\mathrm{T}}\right)-\\
&\quad \mathrm{E}\left(\mathbf{Kvp}^{\mathrm{T}}\left(\mathbf{E}-\mathbf{S}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)+(\mathbf{E}-\mathbf{KS})\mathbf{pv}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)+\mathrm{E}\left(\mathbf{Kvv}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)\\
&= (\mathbf{E}-\mathbf{KS})\mathbf{R}(\mathbf{E}-\mathbf{KS})^{\mathrm{T}}-\mathbf{KN}^{\mathrm{T}}\left(\mathbf{E}-\mathbf{S}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)-(\mathbf{E}-\mathbf{KS})\mathbf{NK}^{\mathrm{T}}+\mathbf{KQK}^{\mathrm{T}}\\
&= \mathbf{R}-\mathbf{K}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)-\left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\mathbf{K}^{\mathrm{T}}+\mathbf{K}\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\mathbf{K}^{\mathrm{T}}\\
&= \mathbf{R}-\left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)-\\
&\quad \left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)+\left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\\
&\quad \left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\\
&\quad \left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right)\\
&= \mathbf{R}-\left(\mathbf{RS}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{SRS}^{\mathrm{T}}+\mathbf{Q}+\mathbf{SN}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR}+\mathbf{N}^{\mathrm{T}}\right).
\end{aligned} \tag{2.44}$$

Aufgabe 2.3. Wenn Sie einmal Zeit und Muße haben, rechnen Sie (2.44) nach.

Im Gegensatz zum Gauß-Markov-Schätzer (2.30) liefert der Minimum-Varianz-Schätzer (2.43) auch dann bereits sinnvolle Ergebnisse, wenn weniger Messungen m als Unbekannte n zur Verfügung stehen, sofern die Matrix $\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}$ regulär ist. Der Grund dieser Eigenschaft liegt offensichtlich darin, dass beim Minimum-Varianz-Schätzer stochastische Informationen über den Parametervektor $\mathbf{p}$ vorliegen und somit eine sinnvolle Schätzung auch mit weniger Messungen, ja sogar mit keiner Messung für $\mathbf{Q} > 0$, möglich ist.

Wendet man nun das Matrizeninversionslemma Satz 1.3

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1} \tag{2.45}$$

auf die Kovarianzmatrix des Schätzfehlers $\mathrm{cov}(\mathbf{e})$ von (2.44) mit $\mathbf{A} = \mathbf{R}^{-1}$, $\mathbf{B} = \mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}$, $\mathbf{C} = \left(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\right)^{-1}$ und $\mathbf{D} = \mathbf{S} + \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}$ an, dann erhält man

$$\begin{aligned}\mathrm{cov}(\mathbf{e}) &= \mathbf{R} - \left(\mathbf{RS}^{\mathrm{T}} + \mathbf{N}\right)\left(\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{SR} + \mathbf{N}^{\mathrm{T}}\right)\\ &= \left(\mathbf{R}^{-1} + \left(\mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}\right)\left(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\right)^{-1}\left(\mathbf{S} + \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\right)\right)^{-1} .\end{aligned} \tag{2.46}$$

Im Weiteren kann man sich einfach davon überzeugen, dass sich der Minimum-Varianz-Schätzer (2.43) in der Form

$$\hat{\mathbf{p}} = \mathrm{cov}(\mathbf{e})\left(\mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}\right)\left(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\right)^{-1}\mathbf{y} \tag{2.47}$$

schreiben lässt.

Aufgabe 2.4. Zeigen Sie die Gültigkeit von (2.47).

Beziehung (2.47) zeigt nun, dass für $\mathbf{R}^{-1} = \mathbf{0}$, d. h. unendlich hohe Varianz des Parametervektors $\mathbf{p}$ – es ist also keine sinnvolle a priori Information für $\mathbf{p}$ vorhanden – und $\mathbf{Q} > 0$ der Minimum-Varianz-Schätzer (2.43) bzw. (2.47) in den Gauß-Markov-Schätzer (2.30) übergeht.

Das bisher Gesagte, insbesondere die Ergebnisse (2.43) und (2.44), in Kombination mit den Beziehungen

$$\mathrm{E}\left(\mathbf{py}^{\mathrm{T}}\right) = \mathrm{E}\left(\mathbf{pp}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}} + \mathbf{pv}^{\mathrm{T}}\right) = \left(\mathbf{RS}^{\mathrm{T}} + \mathbf{N}\right) \tag{2.48}$$

und

$$\mathrm{E}\left(\mathbf{yy}^{\mathrm{T}}\right) = \mathrm{E}\left([\mathbf{Sp} + \mathbf{v}][\mathbf{Sp} + \mathbf{v}]^{\mathrm{T}}\right) = \left(\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right) \tag{2.49}$$

lässt sich im nachfolgenden Satz zusammenfassen:

Satz 2.4 (Minimum-Varianz-Schätzer)**.** *Für das Gleichungssystem* (2.33)

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.50}$$

mit den stochastischen Größen $\mathbf{p}$, $\mathbf{v}$ *und* $\mathbf{y}$ *wird angenommen, dass* $\mathrm{E}\Big(\mathbf{y}\mathbf{y}^{\mathrm{T}}\Big)$ *invertierbar ist. Die optimale lineare Schätzung* $\hat{\mathbf{p}}$ *von* $\mathbf{p}$*, die den Erwartungswert des quadratischen Fehlers* $\mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}[\mathbf{p}-\hat{\mathbf{p}}]\Big)$ *minimiert, ist durch*

$$\hat{\mathbf{p}} = \mathrm{E}\Big(\mathbf{p}\mathbf{y}^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\mathbf{y}\mathbf{y}^{\mathrm{T}}\Big)\Big]^{-1}\mathbf{y} \tag{2.51}$$

mit der zugehörigen Fehlerkovarianzmatrix

$$\begin{aligned} \mathrm{cov}(\mathbf{e}) &= \mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}][\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) - \mathrm{E}\Big(\hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}\Big) \\ &= \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) - \mathrm{E}\Big(\mathbf{p}\mathbf{y}^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\mathbf{y}\mathbf{y}^{\mathrm{T}}\Big)\Big]^{-1}\mathrm{E}\Big(\mathbf{y}\mathbf{p}^{\mathrm{T}}\Big) \end{aligned} \tag{2.52}$$

gegeben.

Man beachte die Ähnlichkeit von (2.51) mit der optimalen Lösung im Sinne der kleinsten Fehlerquadrate von (1.63).

Aufgabe 2.5. Zeigen Sie die Gültigkeit der Beziehung (2.52). Zeigen Sie weiters, dass gilt

$$\mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}][\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}[\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) \quad \text{bzw.} \quad \mathrm{E}\Big(\hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}\hat{\mathbf{p}}^{\mathrm{T}}\Big)\ . \tag{2.53}$$

Hinweis: (zu Aufgabe 2.5) Setzen Sie einfach für $\hat{\mathbf{p}}$ den Ausdruck von (2.51) ein.

Aufgabe 2.6. Zeigen Sie, dass sich die Beziehungen (2.43) und (2.44) direkt mit Hilfe von Satz 2.4 herleiten lassen.

Aufgabe 2.7. Angenommen, die Erwartungswerte $\mathrm{E}(\mathbf{y})$ und $\mathrm{E}(\mathbf{p})$ sind nicht wie in (2.34) Null, sondern $\mathrm{E}(\mathbf{y}) = \mathbf{y}_0 \neq \mathbf{0}$ und $\mathrm{E}(\mathbf{p}) = \mathbf{p}_0 \neq \mathbf{0}$. Zeigen Sie, dass die Minimum-Varianz-Schätzung der Form

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} + \mathbf{b}$$

mit dem konstanten Vektor $\mathbf{b}$ durch

$$\hat{\mathbf{p}} = \mathbf{p}_0 + \mathrm{E}\Big([\mathbf{p}-\mathbf{p}_0][\mathbf{y}-\mathbf{y}_0]^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big([\mathbf{y}-\mathbf{y}_0][\mathbf{y}-\mathbf{y}_0]^{\mathrm{T}}\Big)\Big]^{-1}(\mathbf{y}-\mathbf{y}_0)$$

gegeben ist.

Unter der *Minimum-Varianz-Schätzung einer linearen Funktion*

$$\mathbf{z} = \mathbf{C}\mathbf{p} \tag{2.54}$$

mit dem optimalen Schätzer

$$\hat{\mathbf{z}} = \mathbf{K_z}\mathbf{y} \tag{2.55}$$

basierend auf den Messungen

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.56}$$

versteht man die Lösung der Minimierungsaufgabe

$$\min_{\mathbf{K_z}} \mathrm{E}\Big([\mathbf{z} - \hat{\mathbf{z}}]^{\mathrm{T}}[\mathbf{z} - \hat{\mathbf{z}}]\Big) \ . \tag{2.57}$$

Es gilt nun folgender Satz:

Satz 2.5 (Minimum-Varianz-Schätzer einer linearen Funktion)**.** *Die lineare Minimum-Varianz-Schätzung* (2.55) *einer linearen Funktion* $\mathbf{C}\mathbf{p}$ *basierend auf den Messungen* (2.56) *ist äquivalent der linearen Funktion der Minimum-Varianz-Schätzung* $\hat{\mathbf{p}}$ *selbst, d. h. es gilt, die beste Schätzung von* $\mathbf{C}\mathbf{p}$ *ist* $\mathbf{C}\hat{\mathbf{p}}$*.*

Aufgabe 2.8. Beweisen Sie Satz 2.5.

2.2.1 Rekursive Minimum-Varianz-Schätzung

Im nächsten Schritt soll untersucht werden, wie sich der optimale Schätzwert $\hat{\mathbf{p}}$ von (2.47) durch Hinzunahme neuer Messungen verbessert. Dies ist insbesondere für On-line-Anwendungen von essentieller Bedeutung. Das Verfahren beruht nun wiederum auf den Eigenschaften des Projektionstheorems in einem Hilbertraum. Wenn $\mathcal{U}_1$ und $\mathcal{U}_2$ zwei Unterräume eines Hilbertraums bezeichnen, dann ist die Projektion eines Vektors $\mathbf{p}$ auf den Unterraum $\mathcal{U}_1 + \mathcal{U}_2$ identisch der Projektion von $\mathbf{p}$ auf $\mathcal{U}_1$ plus der Projektion auf $\mathcal{U}_2^*$, wobei $\mathcal{U}_2^*$ orthogonal zu $\mathcal{U}_1$ ist und die Beziehung $\mathcal{U}_1 \oplus \mathcal{U}_2^* = \mathcal{U}_1 + \mathcal{U}_2$ erfüllt. Wenn $\mathcal{U}_2$ durch eine endliche Anzahl von Vektoren aufgebaut ist, dann spannen die Differenzen dieser Vektoren mit ihren Projektionen auf $\mathcal{U}_1$ den Unterraum $\mathcal{U}_2^*$ auf. Die Abbildung 2.6 veranschaulicht diesen Sachverhalt.

Damit kann man folgenden Satz angeben:

Satz 2.6 (Rekursive Minimum-Varianz-Schätzung)**.** *Es sei* $\mathbf{p}$ *ein Zufallsvektor eines Hilbertraumes* $\mathcal{H}$ *und* $\hat{\mathbf{p}}_1$ *bezeichne die orthogonale Projektion von* $\mathbf{p}$ *auf einen geschlossenen Unterraum* $\mathcal{U}_1$ *von* $\mathcal{H}$*. Nach dem Projektionstheorem ist* $\hat{\mathbf{p}}_1$ *also die beste Schätzung von* $\mathbf{p}$ *in* $\mathcal{U}_1$*. Weiters beschreibe* $\mathbf{y}_2$ *alle jene Zufallsvektoren, die den Unterraum* $\mathcal{U}_2$ *von* $\mathcal{H}$ *aufspannen, und* $\hat{\mathbf{y}}_2$ *sei die orthogonale Projektion von* $\mathbf{y}_2$ *auf* $\mathcal{U}_1$*. Nach dem Projektionstheorem ist* $\hat{\mathbf{y}}_2$ *damit die beste Schätzung von* $\mathbf{y}_2$ *in* $\mathcal{U}_1$*. Mit* $\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - \hat{\mathbf{y}}_2$ *lautet die Projektion* $\hat{\mathbf{p}}$ *von* $\mathbf{p}$ *auf* $\mathcal{U}_1 + \mathcal{U}_2$

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}_1 + \mathrm{E}\Big(\mathbf{p}\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\tilde{\mathbf{y}}_2\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big]^{-1}\tilde{\mathbf{y}}_2 \ . \tag{2.58}$$

Damit setzt sich also die beste Schätzung $\hat{\mathbf{p}}$ *auf* $\mathcal{U}_1 + \mathcal{U}_2$ *aus der Summe der besten Schätzung von* $\mathbf{p}$ *auf* $\mathcal{U}_1$ ($\hat{\mathbf{p}}_1$) *und der besten Schätzung von* $\mathbf{p}$ *auf* $\mathcal{U}_2^*$ *(jener Unterraum, der durch* $\tilde{\mathbf{y}}_2$ *generiert wird) zusammen.*