

Technischer Bericht 79

## Kalibrierung von Quell-Ziel-Matrizen

Lutz Dümbgen, Anja Mühlemann, Johanna F. Ziegel (UniBE)  
Ursina Brunnhofer, Martin Bürki, Livio Kaeslin, Georg Rodriguez (SBB)

Dezember 2018

### 1 Ausgangslage

In der Zeit von September 2017 bis Juni 2018 trafen sich die Autorinnen und Autoren dieses Berichts regelmässig, um diverse Fragen rund um Schätzungen und Hochrechnungen im Personenverkehr zu diskutieren. Insbesondere beschäftigten wir uns intensiv mit der Frage, wie man Schätzungen sogenannter Quell-Ziel-Matrizen, die aufgrund notgedrungen kleiner Stichprobengrössen relativ ungenau sind, anhand umfangreicher und daher recht präziser Daten zu sogenannten Abschnittsbelastungen oder Ein- und Aussteigerzahlen an einzelnen Bahnhöfen kalibrieren, also nachbessern kann. Zahlreiche Ansätze wurden diskutiert, von denen uns letztlich einer am meisten überzeugte. Im vorliegenden Bericht beschreiben und diskutieren wir die bisher verwendete Methode sowie den von uns favorisierten neuen Ansatz.

Gegenüber der ersten Version dieses Berichtes vom Juli 2018 wurden jetzt noch gewichtete Distanzen eingefügt, siehe Abschnitt 7.

### 2 Problemstellung

Ausgangspunkt sind ein oder mehrere Züge im Fern- oder Regionalverkehr und alle sinnvollen Verbindungen, welche von Fahrgästen mit diesen Zügen gefahren werden können. Dabei bezieht sich “Zug” auf einen bestimmten Wochentag oder Tagestyp (Werktag oder Sonn- und Feiertag) und eine bestimmte Uhrzeit.

Wenn wir nur einen einzelnen Zug betrachten, ist die Definition von Verbindungen recht einfach: Der Zug fährt nacheinander Bahnhöfe  $0, 1, \dots, m$  an, und seine Fahrt besteht aus Abschnitten  $1, 2, \dots, m$ . Eine Verbindung entspricht dann einem Paar  $(i, j)$  mit  $0 \leq i < j \leq m$ , bestehend aus einem Einstiegsbahnhof  $i$  und dem Ausstiegsbahnhof  $j$ , beziehungsweise der Menge  $\{i + 1, \dots, j\}$  von aufeinanderfolgenden Abschnitten, welche ein Fahrgast mit diesem Zug fahren kann.

Allgemein bezeichnen wir mit  $\mathcal{V}$  die Menge aller sinnvollen Verbindungen für den oder die

gegebenen Züge. Aus Sicht der Bahnbetriebe sind die Verbindungsbelastungen

$x_v^*$  := mittlere Zahl von Fahrgästen, welche Verbindung  $v$  pro Tag benutzen,

für  $v \in \mathcal{V}$  von besonderem Interesse. Diese werden aufgrund von Fahrgastbefragungen (FQ-Erhebungen) durch Werte

$$\hat{x}_v \geq 0$$

geschätzt. Da diese Befragungen recht aufwändig sind, können solche Befragungen nicht allzu oft durchgeführt werden. Selbst in den Zügen, in welchen Befragungsteams unterwegs sind, wird nur ein Bruchteil aller Fahrgäste befragt. Daher sind diese Schätzwerte  $\hat{x}_v$  relativ ungenau. Im Folgenden schreiben wir  $\mathbf{x}^*$  und  $\hat{\mathbf{x}}$  für das gesamte Tupel  $(x_v^*)_{v \in \mathcal{V}}$  bzw.  $(\hat{x}_v)_{v \in \mathcal{V}}$ .

In der Regel sind die Verbindungen durch Ausgangs- und Zielort eindeutig charakterisiert, und man kann die Komponenten von  $\mathbf{x}^*$  bzw.  $\hat{\mathbf{x}}$  in Form einer oberen Dreiecksmatrix anordnen. Deshalb ist oft die Rede von Quell-Ziel-Matrizen. Im Beispiel eines einzelnen Zuges mit  $m+1$  Bahnhöfen  $0, 1, \dots, m$  sieht die entsprechende Quell-Ziel-Matrix dann wie folgt aus:

$$\begin{bmatrix} x_{(0,1)}^* & x_{(0,2)}^* & \cdots & x_{(0,m)}^* \\ & x_{(1,2)}^* & \cdots & x_{(1,m)}^* \\ & & \ddots & \vdots \\ & & & x_{(m-1,m)}^* \end{bmatrix} \quad \text{bzw.} \quad \begin{bmatrix} \hat{x}_{(0,1)} & \hat{x}_{(0,2)} & \cdots & \hat{x}_{(0,m)} \\ & \hat{x}_{(1,2)} & \cdots & \hat{x}_{(1,m)} \\ & & \ddots & \vdots \\ & & & \hat{x}_{(m-1,m)} \end{bmatrix}$$

Eine weitere umfangreiche Informationsquelle sind Abschnittsbelastungen sowie Ein- und Aussteigerzahlen (FRASY-Erhebungen). Ziemlich regelmässig werden für jeden Zug entweder die mittleren Fahrgastzahlen auf einzelnen Abschnitten durch die Zugbegleiter geschätzt (vor allem im Fernverkehr), oder an allen Bahnhöfen werden mittels Sensoren an den Türen die Ein- und Aussteigerzahlen automatisch gemessen (vor allem im Nah- und Regionalverkehr). In unseren Betrachtungen gehen wir davon aus, dass diese Grössen so genau bestimmt werden, dass die Ungenauigkeit vernachlässigbar ist. In Formeln bedeutet dies: Wir gehen davon aus, dass für vorgegebene Teilmengen

$$\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_J \subset \mathcal{V}$$

die Zahlen

$$c_j^* := \sum_{v \in \mathcal{A}_j} x_v^*, \quad 1 \leq j \leq J,$$

bekannt sind.

Im Beispiel eines einzelnen Zuges mit Bahnhöfen  $0, 1, \dots, m$  bedeutet dies zum Beispiel Folgendes: Bestimmt man die mittleren Ein- und Aussteigerszahlen an allen Bahnhöfen,

dann entspricht dies den Teilsommen von  $\mathbf{x}^*$  über folgende  $J = 2m$  Mengen:

$$\begin{aligned}\mathcal{A}_j &:= \{(j-1, j), (j-1, j+1), \dots, (j-1, m)\} \\ &\quad \text{(Einsteiger an Bahnhof } j-1, 1 \leq j \leq m), \\ \mathcal{A}_{m+k} &:= \{(0, k), (1, k), \dots, (k-1, k)\} \\ &\quad \text{(Aussteiger an Bahnhof } k, 1 \leq k \leq m).\end{aligned}$$

Bestimmt man stattdessen die mittlere Anzahl von Fahrgästen pro Tag auf den  $m$  Abschnitten, dann entspricht dies den Teilsommen über die  $J = m$  Mengen

$$\mathcal{A}_j := \{(i, k) : 0 \leq i < j \leq k \leq m\}, \quad 1 \leq j \leq m.$$

Auch bei netzweiten Betrachtungen mit Verbindungen, welche mehrere verschiedene Züge beinhalten, liefern die FRASY-Daten solche Teilsommen für das unbekannte Tupel  $\mathbf{x}^*$ .

Die Frage ist nun, ob und wie man den Schätzer  $\hat{\mathbf{x}}$  für  $\mathbf{x}^*$  verbessern kann, indem man die Information über die  $J$  Teilsommen  $c_j^*$  verwendet.

### 3 SIMBA-Kalibrierungen

Wir beschreiben und diskutieren nun die bisher verwendete Kalibrierungsmethode. In diesem und den nächsten Abschnitten identifizieren wir die Menge  $\mathcal{V}$  der Verbindungen mit den Zahlen  $1, 2, \dots, V$ , und die Tupel  $\mathbf{x}^*, \hat{\mathbf{x}}$  entsprechen dann Vektoren in  $\mathbb{R}^V$ , genauer gesagt, Vektoren mit nichtnegativen Komponenten, also in  $[0, \infty)^V$ . Für einen beliebigen Vektor  $\mathbf{x} = (x_v)_{v=1}^V \in \mathbb{R}^V$  schreiben wir

$$S_j(\mathbf{x}) := \sum_{v \in \mathcal{A}_j} x_v$$

für  $1 \leq j \leq J$ . Wir gehen davon aus, dass alle Summen  $c_j^* = S_j(\mathbf{x}^*)$  und  $S_j(\hat{\mathbf{x}})$  strikt positiv sind. Für jeden Vektor  $\mathbf{x} \in [0, \infty)^V$  mit  $S_j(\mathbf{x}) > 0$  für  $1 \leq j \leq J$  betrachten wir nun die Kalibrationsabbildung  $\mathbf{x} \mapsto \mathbf{K}(\mathbf{x}) = (K_v(\mathbf{x}))_{v=1}^V$  mit

$$\begin{aligned}K_v(\mathbf{x}) &:= x_v \prod_{j=1}^J \left( \frac{c_j^*}{S_j(\mathbf{x})} \right)^{\gamma_{v,j}} \\ &= x_v \exp \left( - \sum_{j=1}^J \gamma_{v,j} \log(S_j(\mathbf{x})/c_j^*) \right).\end{aligned}$$

Dabei sind  $\gamma_{v,j} \geq 0$  gewisse Exponenten, so dass

$$\sum_{j=1}^J \gamma_{v,j} = \begin{cases} 1 & \text{falls } v \in \bigcup_{k=1}^J \mathcal{A}_k, \\ 0 & \text{sonst.} \end{cases}$$

Derzeit arbeiten die SBB mit

$$\gamma_{v,j} = 1_{\mathcal{A}_j}(v) / \sum_{k=1}^J 1_{\mathcal{A}_k}(v),$$

es wären aber auch andere Varianten denkbar, beispielsweise

$$\gamma_{v,j} = 1_{\mathcal{A}_j}(v)c_j^* / \sum_{k=1}^J 1_{\mathcal{A}_k}(v)c_k^*.$$

Ausgehend von  $\mathbf{x}^{(0)} = \widehat{\mathbf{x}}$  wird diese Abbildung  $\mathbf{K}$  nun mehrfach iteriert, das heisst, man berechnet  $\mathbf{x}^{(1)} = \mathbf{K}(\mathbf{x}^{(0)})$ ,  $\mathbf{x}^{(2)} = \mathbf{K}(\mathbf{x}^{(1)})$ ,  $\mathbf{x}^{(3)} = \mathbf{K}(\mathbf{x}^{(2)})$  und so weiter. Nach einer nicht allzu grossen Zahl  $n$  von Iterationen sollte  $\mathbf{x}^{(n)} = \mathbf{K}(\mathbf{x}^{(n-1)}) = \mathbf{K}^n(\widehat{\mathbf{x}})$  näherungsweise ein Fixpunkt der Abbildung  $\mathbf{K}$  sein, und man hofft, dass dieser näher an der Wahrheit  $\mathbf{x}^*$  ist als der ursprüngliche Vektor  $\widehat{\mathbf{x}}$ .

Offensichtlich ist  $\mathbf{x}^*$  ein Fixpunkt dieser Kalibrationsabbildung  $\mathbf{K}$ , das heisst,

$$\mathbf{K}(\mathbf{x}^*) = \mathbf{x}^*.$$

Allgemeiner gilt: Wenn  $\bigcup_{k=1}^J \mathcal{A}_k = \mathcal{V}$ , dann ist

$$\mathbf{K}(\lambda \mathbf{x}^*) = \mathbf{x}^* \quad \text{für beliebige } \lambda > 0.$$

Das heisst, wenn  $\mathbf{x}^{(0)} = \widehat{\mathbf{x}}$  bis auf einen positiven Faktor mit  $\mathbf{x}^*$  übereinstimmt, ist bereits  $\mathbf{x}^{(1)} = \mathbf{x}^*$ , und weitere Anwendungen von  $\mathbf{K}$  bewirken keine Veränderung mehr.

Um zu verstehen, inwiefern  $\mathbf{K}^n(\widehat{\mathbf{x}})$  tatsächlich besser ist als der ursprüngliche Schätzer  $\widehat{\mathbf{x}}$ , analysieren wir nun  $\mathbf{K}(\mathbf{x})$  für Vektoren  $\mathbf{x} = \mathbf{x}^* + \mathbf{h}$  mit Störungen  $\mathbf{h} \in \mathbb{R}^V$ , so dass

$$\delta(\mathbf{h}) := \max_{v=1,\dots,V} \frac{|h_v|}{x_v^*} \ll 1.$$

Wegen  $|S_j(\mathbf{h})|/c_j^* = |S_j(\mathbf{h})|/S_j(\mathbf{x}^*) \leq \delta(\mathbf{h})$  ist

$$\begin{aligned} K_v(\mathbf{x}^* + \mathbf{h}) &= (x_v^* + h_v) \exp\left(-\sum_{j=1}^J \frac{\gamma_{j,v} S_j(\mathbf{h})}{c_j^*} + O(\delta(\mathbf{h})^2)\right) \\ &= x_v^* + h_v - \sum_{j=1}^J \frac{x_v^* \gamma_{v,j}}{c_j^*} S_j(\mathbf{h}) + O(\|\mathbf{x}^*\| \delta(\mathbf{h})^2). \end{aligned}$$

Das heisst,  $\mathbf{K}(\mathbf{x}^* + \mathbf{h}) \approx \mathbf{x}^* + \mathbf{L}^*(\mathbf{h})$  mit dem linearen Operator  $\mathbf{L}^* : \mathbb{R}^V \rightarrow \mathbb{R}^V$ ,

$$\mathbf{L}^*(\mathbf{h}) = \mathbf{h} - \left( \sum_{j=1}^J \frac{x_v^* \gamma_{v,j}}{c_j^*} S_j(\mathbf{h}) \right)_{v \in \mathcal{V}}.$$

Nun führen wir folgende spezielle Vektoren ein:

$$\begin{aligned} \mathbf{e}_v &:= (1_{[u=v]})_{u=1}^V, \quad 1 \leq v \leq V, \\ \mathbf{a}_j &:= (1_{\mathcal{A}_j}(v))_{v=1}^V, \quad 1 \leq j \leq J. \end{aligned}$$

Für  $\mathbf{x} \in \mathbb{R}^V$  ist dann  $S_j(\mathbf{x}) = \mathbf{a}_j^\top \mathbf{x}$  und

$$\mathbf{L}^*(\mathbf{h}) = \mathbf{L}\mathbf{h}$$

mit den Matrizen

$$\begin{aligned} \mathbf{L} &:= \mathbf{I} - \mathbf{G}\mathbf{A}^\top, \\ \mathbf{I} &:= \text{Einheitsmatrix im } \mathbb{R}^{V \times V}, \\ \mathbf{A} &:= [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J] \in \mathbb{R}^{V \times J}, \\ \mathbf{G} &:= \left( \frac{x_v^* \gamma_{v,j}}{c_j^*} \right)_{1 \leq v \leq V, 1 \leq j \leq J} \in \mathbb{R}^{V \times J}. \end{aligned}$$

Die Kalibrationsabbildung  $\mathbf{K}$  liefert zumindest in der Nähe von  $\mathbf{x}^*$  eine Verbesserung von  $\hat{\mathbf{x}}$ , wenn diese Matrix  $\mathbf{L}$  folgende Eigenschaften hat:

(K.1) Jeder Eigenwert  $\lambda \in \mathbb{C}$  von  $\mathbf{L}$  erfüllt die (Un-)Gleichung  $|\lambda| < 1$  oder  $\lambda = 1$ .

(K.2) Algebraische und geometrische Vielfachheit des Eigenwertes 1 sind identisch.

Die geometrische Vielfachheit des Eigenwertes 1 ist mindestens gleich  $V - J$ , denn

$$\mathbf{L}\mathbf{h} = \mathbf{h} \quad \text{falls } \mathbf{h} \perp \{\mathbf{a}_1, \dots, \mathbf{a}_J\},$$

was gleichbedeutend ist mit  $\mathbf{A}^\top \mathbf{h} = \mathbf{0}$ . Wenn also  $\mathbf{L}$  genau  $J$  Eigenwerte  $\lambda$  mit  $|\lambda| < 1$  hat, sind die Bedingungen (K.1-2) erfüllt.

Allgemein ergibt sich aus (K.1-2), dass

$$\mathbb{R}^V = \mathbb{V}_1 \oplus \mathbb{V}_2$$

mit zwei Untervektorräumen  $\mathbb{V}_1, \mathbb{V}_2$  derart, dass  $\dim(\mathbb{V}_1) \geq V - J$  und

$$\mathbf{L}\mathbf{h} = \mathbf{h} \quad \text{für alle } \mathbf{h} \in \mathbb{V}_1,$$

$$\mathbf{L}\mathbb{V}_2 \subset \mathbb{V}_2,$$

$$\lim_{n \rightarrow \infty} \mathbf{L}^n \mathbf{h} = \mathbf{0} \quad \text{für alle } \mathbf{h} \in \mathbb{V}_2.$$

Dies lässt sich mit Hilfe der Jordanschen Normalform von  $\mathbf{L}$  leicht erklären. Mit anderen Worten,

$$\lim_{n \rightarrow \infty} \mathbf{L}^n \mathbf{h} = \mathbf{P}\mathbf{h}$$

mit der Projektion(matrix)  $\mathbf{P}$ , so dass

$$\mathbf{P}\mathbf{h} = \begin{cases} \mathbf{h} & \text{falls } \mathbf{h} \in \mathbb{V}_1, \\ \mathbf{0} & \text{falls } \mathbf{h} \in \mathbb{V}_2. \end{cases}$$

In diesem Sinne ist  $\mathbf{L}$  algebraisch schwach kontraktiv. Im Allgemeinen ist allerdings

$$\mathbb{V}_1 \not\perp \mathbb{V}_2$$

und dann gibt es definitiv Vektoren  $\mathbf{h}$  mit der Eigenschaft, dass

$$\|\mathbf{P}\mathbf{h}\| > \|\mathbf{h}\|.$$

Dabei bezeichnen wir mit  $\|\mathbf{h}\|$  die übliche Euklidische Norm  $(\sum_v h_v^2)^{1/2}$ . In der Tat sind folgende Eigenschaften von  $\mathbf{P}$  äquivalent:

$$\begin{aligned} \|\mathbf{P}\mathbf{h}\| &\leq \|\mathbf{h}\| \quad \text{für alle } \mathbf{h} \in \mathbb{R}^V; \\ \mathbb{V}_1 &\perp \mathbb{V}_2. \end{aligned}$$

Die oben genannten Bedingungen (K.1-2) lassen sich in konkreten Fällen numerisch überprüfen und waren in allen bisher betrachteten Beispielen erfüllt; ein allgemeines theoretisches Resultat konnten wir allerdings nicht beweisen.

Diese Betrachtungen bestätigen, dass die bisher verwendete Kalibrierungsmethode durchaus sinnvoll ist. Allerdings stellte sich auch in allen von uns untersuchten numerischen Beispielen heraus, dass  $\mathbb{V}_1 \not\perp \mathbb{V}_2$ , was bedeutet, dass  $\|\mathbf{K}^n(\hat{\mathbf{x}}) - \mathbf{x}^*\|$  mitunter strikt grösser ist als  $\|\hat{\mathbf{x}} - \mathbf{x}^*\|$ . Diese Problematik, auf die bereits in der Masterarbeit von Lukas Bossart [1] hingewiesen wurde, und auch die schwierige Interpretation von  $\mathbf{K}$  veranlassten uns, nach einer neuen Kalibrierungsmethode zu suchen.

**SIMBA-Kalibrierung logarithmisch betrachtet.** Zur Ergänzung noch eine weitere mathematische Betrachtung: Unter der stärkeren Annahme, dass alle Komponenten von  $\mathbf{x}^*$  und  $\hat{\mathbf{x}}$  strikt positiv sind, kann man die Kalibrationsabbildung auch auf der logarithmischen Skala beschreiben: Wir betrachten nun also Vektoren  $\mathbf{x}^*, \hat{\mathbf{x}} \in \mathbb{R}^V$  mit der Bedeutung, dass  $\exp(\mathbf{x}^*), \exp(\hat{\mathbf{x}})$  die tatsächlichen und die geschätzten mittleren Verbindungsbelastungen sind; dabei ist  $\exp(\cdot)$  komponentenweise zu interpretieren. Nun betrachten wir die Kalibrationsabbildung  $\mathbf{K} : \mathbb{R}^V \rightarrow \mathbb{R}^V$  mit Komponenten

$$K_v(\mathbf{x}) := x_v - \sum_{j=1}^J \gamma_{v,j} \log(S_j(\exp(\mathbf{x}))/c_j^*)$$

wobei nun  $c_j^* = S_j(\exp(\mathbf{x}^*))$ . Nun gilt für beliebige  $\mathbf{x}, \mathbf{h} \in \mathbb{R}^V$  mit  $\|\mathbf{h}\| \ll 1$ , dass

$$K_v(\mathbf{x} + \mathbf{h}) - K_v(\mathbf{x}) = h_v - \sum_{j=1}^J \gamma_{v,j} \frac{S_j(\exp(\mathbf{x})\mathbf{h})}{S_j(\exp(\mathbf{x}))} + O(\|\mathbf{h}\|^2)$$

mit  $\exp(\mathbf{x})\mathbf{h} := (\exp(x_v)h_v)_{v=1}^V$ . Demnach ist

$$\mathbf{K}(\mathbf{x} + \mathbf{h}) = \mathbf{K}(\mathbf{x}) + \mathbf{L}\mathbf{h} + O(\|\mathbf{h}\|^2)$$

mit den Matrizen

$$\mathbf{L} := \left( 1_{[v=w]} - \sum_{j=1}^J \frac{\gamma_{v,j} 1_{\mathcal{A}_j}(w) \exp(x_w)}{\sum_{z \in \mathcal{A}_j} \exp(x_z)} \right)_{v,w=1}^V = \mathbf{I} - \mathbf{G}\mathbf{B}^\top$$

und

$$\begin{aligned} \mathbf{B} &:= \left[ \frac{\exp(\mathbf{x})\mathbf{a}_1}{\mathbf{a}_1^\top \exp(\mathbf{x})}, \frac{\exp(\mathbf{x})\mathbf{a}_2}{\mathbf{a}_2^\top \exp(\mathbf{x})}, \dots, \frac{\exp(\mathbf{x})\mathbf{a}_J}{\mathbf{a}_J^\top \exp(\mathbf{x})} \right] \in \mathbb{R}^{V \times J}, \\ \mathbf{G} &:= (\gamma_{v,j})_{1 \leq v \leq V, 1 \leq j \leq J} \in \mathbb{R}^{V \times J}. \end{aligned}$$

Sowohl  $\mathbf{G}$  als auch  $\mathbf{B}$  haben nichtnegative Einträge, bei  $\mathbf{B}$  sind die Spalten- und bei  $\mathbf{G}$  die Zeilensummen gleich 1.

Auch hier zeigte sich in zahlreichen numerischen Beispielen, dass die zu  $\mathbf{L}$  gehörende lineare Abbildung algebraisch schwach kontraktiv ist, aber die entsprechende Projektionsmatrix  $\mathbf{P} = \lim_{n \rightarrow \infty} \mathbf{L}^n$  hat in der Regel einen oder mehrere Singulärwerte, welche deutlich grösser als 1 sind.

## 4 Distanzminimierung

### 4.1 Zielfunktionen $F$

Einen beliebigen Kandidaten  $\mathbf{x} \in [0, \infty)^V$  für  $\mathbf{x}^*$  kann man unter zwei Aspekten beurteilen:

- (a) Wie nahe ist er am Vektor  $\hat{\mathbf{x}}$  der aus den FQ-Erhebungen geschätzten Verbindungsbelastungen?
- (b) Wie nahe ist der entsprechende Vektor  $\mathbf{A}^\top \mathbf{x}$  am Vektor  $\mathbf{c}^*$  der tatsächlichen Abschnittsbelastungen bzw. Ein- und Aussteigerzahlen?

Beide Aspekte lassen sich kombinieren, indem man für einen gegebenen Parameter  $\lambda$  mit  $0 \ll \lambda < 1$  einen Vektor  $\mathbf{x} \in \mathbb{R}^V$  bestimmt, so dass

$$F(\mathbf{x}, \hat{\mathbf{x}}) := (1 - \lambda) \frac{d_V(\mathbf{x}, \hat{\mathbf{x}})^2}{2} + \lambda \frac{d_J(\mathbf{A}^\top \mathbf{x}, \mathbf{c}^*)^2}{2}$$

unter der Nebenbedingung

$$\mathbf{x} \geq \mathbf{x}^{\min}$$

(komponentenweise gemeint) minimal wird. Dabei sind  $d_V(\cdot, \cdot)$ ,  $d_J(\cdot, \cdot)$  geeignete Abstandsmasse auf  $\mathbb{R}^V$  bzw.  $\mathbb{R}^J$ . Der Parameter  $\lambda$  steuert also, wie sehr man Abweichungen von bekannten mittleren Abschnittsbelastungen oder Ein- und Aussteigerzahlen bestraft. Im Grenzfall  $\lambda \rightarrow 0$  ergäbe sich einfach  $\mathbf{x} = \hat{\mathbf{x}}$ , doch wir sind eher an Werten  $\lambda \approx 1$  interessiert. Der Vektor  $\mathbf{x}^{\min} \geq \mathbf{0}$  ist ein vorgegebener Vektor von unteren Schranken für die Verbindungsbelastungen. Im einfachsten Falle ist  $\mathbf{x}^{\min} = \mathbf{0}$ , denkbar wäre beispielsweise auch  $\mathbf{x}^{\min} = \epsilon \hat{\mathbf{x}}$  mit einem kleinen Faktor  $\epsilon \in (0, 1)$ .

Als Abstandsmass  $d_J$  wählen wir die übliche euklidische Distanz auf  $\mathbb{R}^J$ , also

$$d_J(\mathbf{A}^\top \mathbf{x}, \mathbf{c}^*)^2 = \|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2 = \sum_{j=1}^J (\mathbf{a}_j^\top \mathbf{x} - c_j^*)^2.$$

Auf Ebene der Verbindungsbelastungen könnte man analog mit der euklidischen Distanz auf  $\mathbb{R}^V$  arbeiten. Dies liefert dann die Zielfunktion

$$F_1(\mathbf{x}, \hat{\mathbf{x}}) := (1 - \lambda) \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{2} + \lambda \frac{\|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2}{2}.$$

In Abschnitt 7 wird kurz auf andere Distanzen  $d_V(\cdot, \cdot)$  eingegangen.

**(Partielle) Skaleninvarianz.** Obwohl die FQ-Schätzungen so konzipiert sind, dass  $\hat{\mathbf{x}}$  ein unverzerrter Schätzer für  $\mathbf{x}^*$  sein sollte, wollten wir uns vor allem darauf konzentrieren, wie sehr sich die entsprechenden Richtungsvektoren  $\mathbf{u}(\mathbf{x})$  und  $\mathbf{u}(\hat{\mathbf{x}})$  unterscheiden, wobei

$$\mathbf{u}(\mathbf{y}) := \begin{cases} \|\mathbf{y}\|^{-1}\mathbf{y} & \text{falls } \mathbf{y} \neq \mathbf{0}, \\ \mathbf{0} & \text{falls } \mathbf{y} = \mathbf{0}. \end{cases}$$

Die Skalierung von  $\mathbf{x}$  wird ja bereits durch den Abstand zwischen  $\mathbf{A}^\top \mathbf{x}$  und  $\mathbf{c}^*$  berücksichtigt. Von daher wäre es naheliegend, mit  $d_V(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\hat{\mathbf{x}})\|$  zu arbeiten. Rein algorithmisch wäre jedoch eine in  $\mathbf{x}$  quadratische Zielfunktion angenehmer. Daher verwenden wir

$$\begin{aligned} d_V(\mathbf{x}, \hat{\mathbf{x}})^2 &= \min_{s \in \mathbb{R}} \|\mathbf{x} - s\hat{\mathbf{x}}\|^2 \\ &= \min_{s \in \mathbb{R}} \|\mathbf{x} - s\mathbf{u}(\hat{\mathbf{x}})\|^2 \\ &= \|\mathbf{x} - (\mathbf{x}^\top \mathbf{u}(\hat{\mathbf{x}}))\mathbf{u}(\hat{\mathbf{x}})\|^2 \\ &= \|\mathbf{x}\|^2 - (\mathbf{x}^\top \mathbf{u}(\hat{\mathbf{x}}))^2. \end{aligned}$$

Diese Grösse misst, wie sehr sich  $\mathbf{x}$  von irgendeinem skalaren Vielfachen von  $\hat{\mathbf{x}}$  unterscheidet. Sie ist also skaleninvariant bezüglich  $\hat{\mathbf{x}}$ . Da wir nur Vektoren  $\mathbf{x} \geq \mathbf{0}$  betrachten und auch  $\hat{\mathbf{x}} \geq \mathbf{0}$ , ist automatisch  $\mathbf{x}^\top \mathbf{u}(\hat{\mathbf{x}}) \geq 0$ .

Diese Betrachtungen führen zu der Zielfunktion

$$\begin{aligned} F_2(\mathbf{x}, \hat{\mathbf{x}}) &:= (1 - \lambda) \min_{s \in \mathbb{R}} \frac{\|\mathbf{x} - s\hat{\mathbf{x}}\|^2}{2} + \lambda \frac{\|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2}{2} \\ &= (1 - \lambda) \frac{\|\mathbf{x}\|^2 - (\mathbf{x}^\top \hat{\mathbf{u}})^2}{2} + \lambda \frac{\|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2}{2}. \end{aligned}$$

Mit der Matrix

$$\Pi(\hat{\mathbf{x}}) := \mathbf{I} - \mathbf{u}(\hat{\mathbf{x}})\mathbf{u}(\hat{\mathbf{x}})^\top,$$

welche die orthogonale Projektion auf  $\mathbf{u}(\hat{\mathbf{x}})^\perp = \hat{\mathbf{x}}^\perp$  beschreibt, kann man auch schreiben:

$$\begin{aligned} F_2(\mathbf{x}, \hat{\mathbf{x}}) &= (1 - \lambda) \frac{\|\Pi(\hat{\mathbf{x}})\mathbf{x}\|^2}{2} + \lambda \frac{\|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2}{2} \\ &= (1 - \lambda) \frac{\mathbf{x}^\top \Pi(\hat{\mathbf{x}})\mathbf{x}}{2} + \lambda \frac{\|\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*\|^2}{2}. \end{aligned}$$

**Anmerkung.** Zusätzlich zu den Nebenbedingungen, dass  $\mathbf{x} \geq \mathbf{x}^{\min}$ , könnte man noch Nebenbedingungen der Form  $\mathbf{x} \leq \mathbf{x}^{\max}$  einführen, wobei  $\mathbf{x}^{\max} > \mathbf{x}^{\min}$  ein Vektor von oberen Schranken für die Komponenten von  $\mathbf{x}$  ist. Die nachfolgenden Betrachtungen lassen sich recht einfach auf diese Variante anpassen.



## 4.2 Minimierung von $F$

Bei festem  $\hat{\mathbf{x}}$  stellen  $F_1(\cdot, \hat{\mathbf{x}})$  und  $F_2(\cdot, \hat{\mathbf{x}})$  quadratische Funktionen dar. Die entsprechenden Gradienten sind

$$\begin{aligned}\nabla F_1(\mathbf{x}, \hat{\mathbf{x}}) &= (1 - \lambda)(\mathbf{x} - \hat{\mathbf{x}}) + \lambda \mathbf{A}(\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*) \\ &= \mathbf{H}_1 \mathbf{x} - ((1 - \lambda)\hat{\mathbf{x}} + \lambda \mathbf{A} \mathbf{c}^*), \\ \nabla F_2(\mathbf{x}, \hat{\mathbf{x}}) &= (1 - \lambda)\Pi(\hat{\mathbf{x}})\mathbf{x} + \lambda \mathbf{A}(\mathbf{A}^\top \mathbf{x} - \mathbf{c}^*) \\ &= \mathbf{H}_2(\hat{\mathbf{x}})\mathbf{x} - \lambda \mathbf{A} \mathbf{c}^*\end{aligned}$$

mit den Hesse-Matrizen

$$\begin{aligned}D^2 F_1(\cdot, \hat{\mathbf{x}}) &\equiv \mathbf{H}_1 := (1 - \lambda)\mathbf{I} + \lambda \mathbf{A} \mathbf{A}^\top, \\ D^2 F_2(\cdot, \hat{\mathbf{x}}) &\equiv \mathbf{H}_2(\hat{\mathbf{x}}) := (1 - \lambda)\Pi(\hat{\mathbf{x}}) + \lambda \mathbf{A} \mathbf{A}^\top.\end{aligned}$$

Offensichtlich ist  $\mathbf{H}_1$  positiv definit. Dies gilt auch für  $\mathbf{H}_2(\hat{\mathbf{x}})$  sofern  $\mathbf{A}^\top \hat{\mathbf{x}} \neq \mathbf{0}$ . Denn für einen beliebigen Vektor  $\mathbf{y} \in \mathbb{R}^V$  folgt aus

$$\mathbf{y}^\top \mathbf{H}_2(\hat{\mathbf{x}})\mathbf{y} = (1 - \lambda)\|\Pi(\hat{\mathbf{x}})\mathbf{y}\|^2 + \lambda\|\mathbf{A}^\top \mathbf{y}\|^2 = 0,$$

dass  $\Pi(\hat{\mathbf{x}})\mathbf{y} = \mathbf{0}$ , also  $\mathbf{y} = s\hat{\mathbf{x}}$  für ein  $s \in \mathbb{R}$ , und dann ist  $\mathbf{A}^\top \mathbf{y} = s\mathbf{A}^\top \hat{\mathbf{x}} = \mathbf{0}$  genau dann, wenn  $s = 0$ .

Für  $F = F_1, F_2$  zeigen Standardargumente aus der konvexen Optimierung (siehe z.B. [2]), dass ein Vektor  $\mathbf{x} \geq \mathbf{x}^{\min}$  das Optimierungsproblem genau dann löst, wenn er folgende Bedingungen erfüllt:

$$\nabla F(\mathbf{x}, \hat{\mathbf{x}})_v = 0 \quad \text{falls } x_v > x_v^{\min}, \quad (1)$$

$$\nabla F(\mathbf{x}, \hat{\mathbf{x}})_v \geq 0 \quad \text{falls } x_v = x_v^{\min}. \quad (2)$$

Daher bietet sich eine Active-Set-Methode an. Für einen Vektor  $\mathbf{x} \in \mathbb{R}^V$  definieren wir

$$\mathcal{A}(\mathbf{x}) := \{v \in \{1, \dots, V\} : x_v = x_v^{\min}\},$$

die Menge seiner ‘aktiven Nebenbedingungen’. Ausserdem betrachten wir für eine beliebige Indexmenge  $\mathcal{A} \subset \{1, \dots, V\}$  die Menge

$$\mathbb{X}(\mathcal{A}) := \{\mathbf{x} \in \mathbb{R}^V : x_v = x_v^{\min} \text{ für alle } v \in \mathcal{A}\} = \{\mathbf{x} \in \mathbb{R}^V : \mathcal{A} \subset \mathcal{A}(\mathbf{x})\}.$$

Diese ist ein affiner Teilraum vom  $\mathbb{R}^V$  mit Dimension  $V - \#\mathcal{A}$ . Man kann daher die Funktion  $F(\cdot, \hat{\mathbf{x}})$  auf diesem Raum problemlos minimieren, indem man das Gleichungssystem

$$\nabla F(\mathbf{x}, \hat{\mathbf{x}})_v \stackrel{!}{=} 0 \quad \text{für } v \in \mathcal{V} \setminus \mathcal{A}$$

unter der Nebenbedingung, dass

$$x_v = x_v^{\min} \quad \text{für } v \in \mathcal{A}$$

löst. Nun geht man wie folgt vor:

---

### Active-Set-Algorithmus

**Schritt 1 (Grobe erste Suche):** Setze

$$\begin{aligned}\mathbf{x} &\leftarrow \max(\widehat{\mathbf{x}}, \mathbf{x}^{\min}), \\ \mathbf{x} &\leftarrow \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}))} F(\mathbf{y}, \widehat{\mathbf{x}}).\end{aligned}$$

Solange  $\mathbf{x} \not\geq \mathbf{x}^{\min}$ , setze

$$\begin{aligned}\mathbf{x} &\leftarrow \max(\mathbf{x}, \mathbf{x}^{\min}), \\ \mathbf{x} &\leftarrow \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}))} F(\mathbf{y}, \widehat{\mathbf{x}}).\end{aligned}$$

Dabei ist  $\max(\cdot, \cdot)$  komponentenweise gemeint. Bei dieser Suche wird die Menge  $\mathcal{A}(\mathbf{x})$  stets grösser. Spätestens wenn  $\mathcal{A}(\mathbf{x}) = \{1, \dots, V\}$ , werden die Bedingungen  $\mathbf{x} \geq \mathbf{x}^{\min}$  und (1) erfüllt sein.

**Schritt 2 (Deaktivieren einer Nebenbedingung):** Gegeben sei ein Vektor  $\mathbf{x} \geq \mathbf{x}^{\min}$  mit der Eigenschaft, dass

$$\mathbf{x} = \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}))} F(\mathbf{y}, \widehat{\mathbf{x}}),$$

was gleichbedeutend mit Eigenschaft (1) ist. Nun bestimme

$$\delta \leftarrow \min_{v=1, \dots, V} \nabla F(\mathbf{x}, \widehat{\mathbf{x}})_v$$

Falls  $\delta = 0$ , ist  $\mathbf{x}$  das gesuchte Optimum gemäss (1) und (2). Falls nicht, wähle einen Index  $v_o \in \mathcal{A}(\mathbf{x})$  mit  $\nabla F(\mathbf{x})_{v_o} = \delta < 0$  und bestimme

$$\mathbf{x}^{\text{neu}} \leftarrow \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}) \setminus \{v_o\})} F(\mathbf{y}, \widehat{\mathbf{x}}).$$

Dieser Vektor  $\mathbf{x}^{\text{neu}}$  hat garantiert die Eigenschaft, dass  $x_{v_o}^{\text{neu}} > x_{v_o}^{\min}$ . Nun setze

$$\begin{aligned}t_o &\leftarrow \max\{t \in (0, 1] : (1-t)\mathbf{x} + t\mathbf{x}^{\text{neu}} \geq \mathbf{x}^{\min}\}, \\ \mathbf{x} &\leftarrow (1-t_o)\mathbf{x} + t_o\mathbf{x}^{\text{neu}}\end{aligned}$$

und gehe zum nächsten Schritt.

**Schritt 3 (Lokale Suche):** Für einen gegebenen Vektor  $\mathbf{x} \geq \mathbf{x}^{\min}$  setze

$$\mathbf{x}^{\text{neu}} \leftarrow \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}))} F(\mathbf{y}, \widehat{\mathbf{x}}).$$

Solange  $\mathbf{x}^{\text{neu}} \not\geq \mathbf{x}^{\min}$ , setze

$$\begin{aligned}t_o &\leftarrow \max\{t \in (0, 1] : (1-t)\mathbf{x} + t\mathbf{x}^{\text{neu}} \geq \mathbf{x}^{\min}\}, \\ \mathbf{x} &\leftarrow (1-t_o)\mathbf{x} + t_o\mathbf{x}^{\text{neu}}, \\ \mathbf{x}^{\text{neu}} &\leftarrow \arg \min_{\mathbf{y} \in \mathbb{X}(\mathcal{A}(\mathbf{x}))} F(\mathbf{y}, \widehat{\mathbf{x}}).\end{aligned}$$

Schliesslich setze

$$\mathbf{x} \leftarrow \mathbf{x}^{\text{neu}}$$

und wiederhole Schritt 2.

---

Sowohl in Schritt 2 als auch in Schritt 3 nimmt der Wert  $F(\mathbf{x}, \hat{\mathbf{x}})$  bei jeder Veränderung von  $\mathbf{x}$  strikt ab. Wichtig ist, dass man bei der Durchführung von

$$\mathbf{x} \leftarrow (1 - t_o)\mathbf{x} + t_o\mathbf{x}^{\text{neu}}$$

erzwingt, dass tatsächlich neue Nebenbedingungen aktiviert werden. Ansonsten läuft man Gefahr, in Endlosschleifen zu landen. Bei geeigneter Implementierung erhält man nach endlich vielen Schritten eine Lösung  $\mathbf{x}$  des besagten Optimierungsproblems.

In konkreten Beispielen zeigte sich, dass bereits Schritt 1 in der Regel schon einen nahezu optimalen Punkt  $\mathbf{x}$  liefert, und nach nur wenigen Durchführungen der Schritte 2 und 3 ergab sich eine Lösung.

## 5 Eine einfache (Nach-)Kalibrierung

Auch nach Durchführung der vorangehend beschriebenen Kalibrierungsmethoden bietet es sich an, einen Kandidaten  $\mathbf{x}$  für  $\mathbf{x}^*$  noch durch

$$\mathbf{S}(\mathbf{x}) = s(\mathbf{x})\mathbf{x}$$

mit dem Skalenfaktor

$$s(\mathbf{x}) = \frac{\sum_{j=1}^J c_j^*}{\sum_{j=1}^J \mathbf{a}_j^\top \mathbf{x}}$$

zu ersetzen. Dies führt dazu, dass die Summe der  $J$  vorgegebenen Abschnittsbelastungen oder Ein- und Aussteigerzahlen mit der entsprechenden Summe für  $\mathbf{S}(\mathbf{x})$  übereinstimmt.

## 6 Numerisches Beispiel

Wir illustrieren die Kalibrierung mittels SIMBA und Distanzminimierung mit  $F = F_2$  am Beispiel eines einzelnen Zuges, nämlich des Intercity 710 von St. Gallen nach Genf Flughafen. Dieser Zug verbindet 13 Bahnhöfe, doch von den  $78 = \binom{13}{2}$  möglichen Verbindungen hatten nur 69 einen strikt positiven Wert aus den FQ-Erhebungen. Daher schränkten wir uns auf diese  $V = 69$  Verbindungen ein. Für diesen Zug sind keine Ein- oder Aussteigerzahlen bekannt, nur Abschnittsbelastungen von  $J = 10$  der 12 Abschnitte.

Die Zahlenwerte in den nachfolgenden Tabellen sind stets auf eine bis drei Nachkommastellen gerundet. Tabelle 1 zeigt die zu  $\hat{\mathbf{x}}$  gehörende Quell-Ziel-Matrix.

Tabelle 2 zeigt die für diesen Zug bekannten Abschnittsbelastungen  $c_j^*$  sowie die sich aus  $\hat{\mathbf{x}}$  ergebenden geschätzten Abschnittsbelastungen  $\mathbf{a}_j^\top \hat{\mathbf{x}}$  und  $\mathbf{a}_j^\top \mathbf{S}(\hat{\mathbf{x}})$ .

SG	13.2	11.9	8.9	21.7	43.5	22.8	74.0	8.8	1.2	2.1	2.1	0.3
	GSS	18.7	5.7	11.4	22.4	11.4	34.1	2.6	0	0.4	0.1	0
		FLA	6.8	11.9	6.6	1.1	12.4	0.6	0	0.1	0.5	0
			UZW	18.7	17.9	6.3	18.2	0.3	0	0	0.2	0
				WIL	73.2	25.3	89.0	7.9	0.2	0.7	0.2	0
					W	56.4	64.4	31.5	1.1	3.0	1.0	0
						ZFH	83.7	18.7	5.3	3.2	1.5	0.1
							ZUE	329.0	17.5	35.2	24.5	1.6
								BN	117.0	61.2	23.8	4.2
									FRI	43.2	10.2	2.2
										LS	81.0	11.1
											GE	14.7
												GEAP

Tabelle 1: Geschätzte Quell-Ziel-Matrix  $\hat{\mathbf{x}}$ .

$j$	$c_j^*$	$\mathbf{a}_j^\top \hat{\mathbf{x}}$	$\mathbf{a}_j^\top \mathbf{S}(\hat{\mathbf{x}})$
SG-GSS	246.2	210.7	243.8
UZW-WIL	394.8	353.8	409.5
WIL-W	521.2	486.3	562.8
W-ZFH	562.5	480.1	555.6
ZFH-ZUE	534.1	469.2	543.0
ZUE-BN	562.8	501.2	580.1
BN-FRI	344.5	308.1	356.6
FRI-LS	260.2	221.5	256.4
LS-GE	203.2	164.5	190.4

Tabelle 2: Bekannte und aus  $\hat{\mathbf{x}}$  geschätzte Abschnittsbelastungen.

$v$	1, ..., 59	60	61	62	63	64	65	66	67	68	69
$\lambda_v$	1	0.84	0.81	0.74	0.70	0.55	0.46	0.32	0.20	0.07	0.00

Tabelle 3: Eigenwerte  $\lambda_v$  von  $D\mathbf{K}(\mathbf{x}^{\text{SIMBA}})$ .

$v$	1	2	3	4	5	6	7	8	9	10	11, ..., 59	60, ..., 69
$\sigma_v$	4.62	2.60	1.98	1.81	1.81	1.59	1.48	1.44	1.28	1.26	1	0

Tabelle 4: Singulärwerte  $\sigma_v$  von  $\lim_{n \rightarrow \infty} D\mathbf{K}(\mathbf{x}^{\text{SIMBA}})^n$ .

Nach 20 Iterationen der SIMBA-Kalibrationsabbildung  $\mathbf{K}$  ergab sich ein Vektor  $\mathbf{x}^{\text{SIMBA}} \in \mathbb{R}^{69}$ , der im Wesentlichen ein Fixpunkt von  $\mathbf{K}$  ist. Die absteigend sortierten Eigenwerte  $\lambda_v$  der entsprechenden Ableitungsmatrix  $\mathbf{L} = D\mathbf{K}(\mathbf{x}^{\text{SIMBA}})$  sieht man in Tabelle 3, die Singulärwerte  $\sigma_v$  der Matrix  $\lim_{n \rightarrow \infty} \mathbf{L}^n$  sieht man in Tabelle 4. Offensichtlich sind die Bedingungen (K.1-2) erfüllt, denn  $0 \leq \lambda_v < 1$  für  $J = 10$  Verbindungen  $v \in \mathcal{V}$ . Andererseits sind zehn Singulärwerte grösser als 1, einige recht deutlich. Es lässt sich also nicht ausschliessen, dass  $\|\mathbf{x}^{\text{Simba}} - \mathbf{x}^*\|$  grösser ist als  $\|\hat{\mathbf{x}} - \mathbf{x}^*\|$ .

Der neue Ansatz der Distanzminimierung lieferte für  $\lambda = 0.999$  und  $\mathbf{x}^{\text{min}} = 0.01 \cdot \hat{\mathbf{x}}$  bereits nach Schritt 1 des Active-Set-Algorithmus in zwei Iterationen die Minimalstelle  $\mathbf{x}^{\text{DIST}}$ . Tabelle 5 zeigt die tatsächlichen und geschätzten Abschnittsbelastungen für  $\mathbf{x} = \mathbf{S}(\hat{\mathbf{x}}), \mathbf{S}(\mathbf{x}^{\text{SIMBA}}), \mathbf{S}(\mathbf{x}^{\text{DIST}})$  sowie in Klammern die relativen Fehler

$$\frac{\mathbf{a}_j^\top \mathbf{x}}{c_j^*} - 1$$

in Prozent. Eine graphische Darstellung der relativen Fehler sieht man in Abbildung 1.

Schliesslich sieht man noch in Tabelle 6 die Quell-Zielmatrizen zu  $\mathbf{S}(\hat{\mathbf{x}}), \mathbf{S}(\mathbf{x}^{\text{SIMBA}})$  und  $\mathbf{S}(\mathbf{x}^{\text{DIST}})$ , und in Tabelle 7 werden die Faktoren

$$\frac{\mathbf{S}(\mathbf{x})}{\mathbf{S}(\hat{\mathbf{x}})}$$

für  $\mathbf{x} = \mathbf{x}^{\text{SIMBA}}, \mathbf{x}^{\text{DIST}}$  angegeben.

$j$	$c_j^*$	Rohdaten		SIMBA		Distanzmin.	
SG-GSS	246.2	243.8	(-0.99)	246.3	(+0.007)	246.2	(±0.000)
UZW-WIL	394.8	409.5	(+3.72)	394.6	(-0.050)	394.8	(±0.000)
WIL-W	521.2	562.8	(+7.99)	521.6	(+0.075)	521.2	(+0.001)
W-ZFH	562.5	555.6	(-1.22)	562.1	(-0.074)	562.5	(-0.001)
ZFH-ZUE	534.1	543.0	(+1.68)	534.3	(+0.039)	534.1	(+0.001)
ZUE-BN	562.8	580.1	(+3.07)	562.8	(+0.000)	562.8	(±0.000)
BN-FRI	344.5	356.6	(+3.51)	344.5	(+0.001)	344.5	(+0.001)
FRI-LS	260.2	256.4	(-1.47)	260.2	(-0.001)	260.2	(±0.000)
LS-GE	203.2	190.4	(-6.29)	203.2	(+0.000)	203.2	(+0.003)
GE-GEAP	108.4	39.7	(-63.40)	108.4	(+0.000)	108.3	(-0.014)

Tabelle 5: Bekannte und geschätzte Abschnittsbelastungen (samt relativen Fehlern in Prozent).

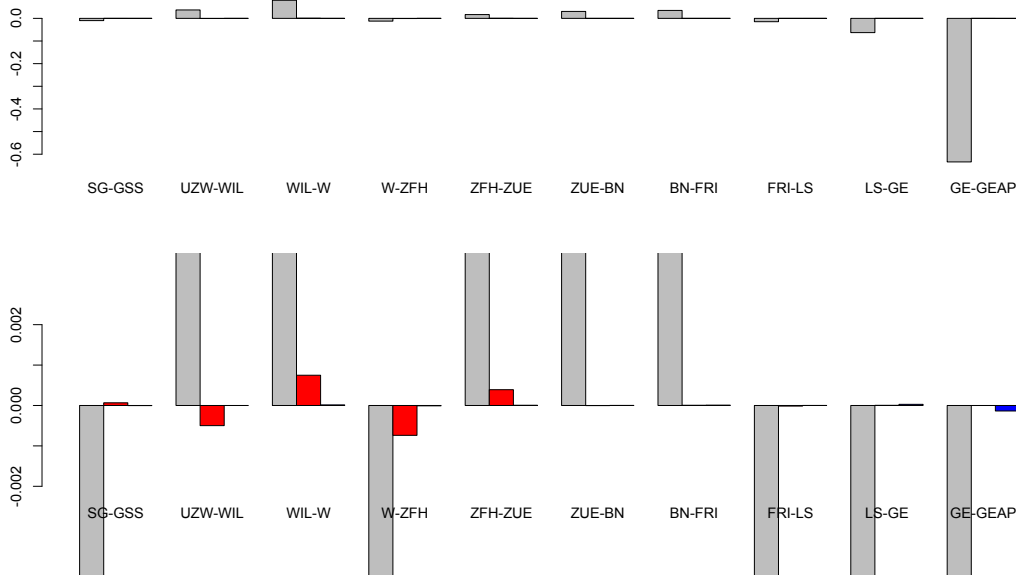


Abbildung 1: Relative Fehler der geschätzten Abschnittsbelastungen für  $\mathbf{S}(\hat{\mathbf{x}})$  (grau),  $\mathbf{S}(\mathbf{x}^{\text{SIMBA}})$  (rot) und  $\mathbf{S}(\mathbf{x}^{\text{DIST}})$  (blau).

SG	15.3	13.8	10.3	25.2	50.4	26.4	85.7	10.2	1.4	2.4	2.4	0.3
	17.2	15.5	11.6	26.4	47.6	26.4	84.9	10.1	1.3	2.4	2.4	0.4
	15.4	14.0	10.8	25.7	42.9	27.6	80.3	12.2	2.2	3.6	0.0	11.6
GSS	21.7	6.6	13.2	25.9	13.2	39.4	3.0	0	0.5	0.1	0	
	18.7	5.7	12.9	22.4	12.7	37.9	2.9	0	0.5	0.1	0	
	19.8	6.0	13.4	19.1	14.1	36.6	4.1	0	0.4	0.0	0	
FLA	7.9	13.8	7.6	1.3	14.4	0.6	0	0	0.1	0.5	0	
	6.8	13.5	6.6	1.2	13.8	0.6	0	0	0.1	0.5	0	
	7.2	13.9	2.5	3.2	13.8	2.0	0	0	0.1	0.0	0	
UZW	21.7	20.7	7.3	21.0	0.3	0	0	0	0	0.2	0	
	21.2	17.9	7.0	20.2	0.3	0	0	0	0	0.2	0	
	21.1	14.4	8.7	19.8	1.6	0	0	0	0	0.0	0	
WIL	84.7	29.3	103.0	9.1	0.2	0.8	0.2	0				
	65.0	27.9	98.3	8.7	0.2	0.8	0.2	0				
	71.5	27.4	93.3	8.4	0.0	0.0	0.0	0				
W	65.3	74.5	36.5	1.3	3.4	1.2	0					
	77.3	79.3	37.5	1.3	3.5	1.2	0					
	66.1	73.1	39.1	5.1	7.6	0.4	0					
ZFH	96.8	21.6	6.1	3.7	1.7	0.2						
	92.7	20.8	5.8	3.6	1.7	0.2						
	87.0	19.1	3.0	1.4	0.0	8.0						
ZUE	380.7	20.2	40.7	28.4	1.9							
	366.6	19.1	39.9	27.7	2.5							
	348.1	17.2	36.4	20.0	10.9							
BN	135.5	70.8	27.5	4.9								
	125.4	69.9	27.0	6.8								
	121.7	63.2	18.5	12.9								
FRI	50.0	11.8	2.5									
	52.7	11.9	4.1									
	46.2	6.1	12.7									
LS	93.8	12.8										
	91.0	25.4										
	80.4	21.6										
GE	17.0											
	69.0											
	30.6											
GEAP												

Tabelle 6: Geschätzte Quell-Ziel-Matrizen  $\mathbf{S}(\hat{\mathbf{x}})$ ,  $\mathbf{S}(\mathbf{x}^{\text{SIMBA}})$  und  $\mathbf{S}(\mathbf{x}^{\text{DIST}})$ .

SG	1.1	1.1	1.1	1.0	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.1
	1.0	1.0	1.1	1.0	0.9	1.0	0.9	1.2	1.6	1.5	0.0	34.5
	GSS	0.9	0.9	1.0	0.9	1.0	1.0	1.0		1.0	1.0	
		0.9	0.9	1.0	0.7	1.1	0.9	1.4		0.9	0.0	
		FLA	0.9	1.0	0.9	1.0	1.0	1.0		1.0	1.0	
			0.9	1.0	0.3	2.5	1.0	3.0		0.7	0.0	
			UZW	1.0	0.9	1.0	1.0	1.0			1.0	
				1.0	0.7	1.2	0.9	5.6			0.0	
				WIL	0.8	1.0	1.0	1.0	1.0	1.0	1.0	
					0.8	0.9	0.9	0.9	0.0	0.0	0.0	
					W	1.2	1.1	1.0	1.0	1.0	1.0	
						1.0	1.0	1.1	4.1	2.2	0.3	
						ZFH	1.0	1.0	0.9	1.0	1.0	1.2
							0.9	0.9	0.5	0.4	0.0	48.8
							ZUE	1.0	0.9	1.0	1.0	1.3
								0.9	0.9	0.9	0.7	5.8
								BN	0.9	1.0	1.0	1.4
									0.9	0.9	0.7	2.6
									FRI	1.1	1.0	1.6
										0.9	0.5	5.0
										LS	1.0	2.0
											0.9	1.7
											GE	4.9
												1.8
												GEAP

Tabelle 7: Faktoren  $\mathbf{S}(\mathbf{x})/\mathbf{S}(\hat{\mathbf{x}})$  für  $\mathbf{x} = \mathbf{x}^{\text{SIMBA}}, \mathbf{x}^{\text{DIST}}$  (in Prozent).



## 7 Gewichtung der Verbindungsbelastungen

Es gibt gute Gründe, die Abweichungen  $|x_v - \hat{x}_v|$  je nach Verbindung  $v \in \mathcal{V}$  unterschiedlich zu gewichten. Zum einen liefern unterschiedliche Verbindungen auch unterschiedliche mittlere Einnahmen, so dass die Auswirkungen allfälliger Fehler je nach Verbindung unterschiedlich gross sind. Ausserdem hängen auch die Ungenauigkeiten der geschätzten Werte  $\hat{x}_v$  stark von der Verbindung ab. Zum Beispiel zeigen einfache Modellrechnungen, dass bei FQ-Erhebungen für einen einzelnen Zug die Standardabweichung von  $\hat{x}_v$  in etwa proportional zu  $1/\sqrt{\ell(v)}$  ist, wobei  $\ell(v)$  die Fahrzeit von Verbindung  $v$  mit dem betrachteten Zug ist.

Solche Betrachtungen legen gewichtete euklidische Abstände nahe. Das heisst, anstelle von  $\|\mathbf{x} - \hat{\mathbf{x}}\|$  verwenden wir

$$d_V(\mathbf{x}, \hat{\mathbf{x}}) := \left( \sum_{v=1}^V w_v^2 (x_v - \hat{x}_v)^2 \right)^{1/2}$$

mit gewissen Gewichten  $w_v > 0$ , zum Beispiel  $w_v = 1/\sqrt{\ell(v)}$ . Die zuvor beschriebenen Methoden bleiben aber gültig, wenn man einfach wie folgt vorgeht:

Zu Beginn der Kalibration werden die Vektoren  $\hat{\mathbf{x}}$ ,  $\mathbf{x}^{\min}$  und  $\mathbf{x}^{\max}$  durch  $(w_v \hat{x}_v)_{v=1}^V$ ,  $(w_v x_v^{\min})_{v=1}^V$  bzw.  $(w_v x_v^{\max})_{v=1}^V$  ersetzt. Gleichzeitig werden die Spaltenvektoren  $\mathbf{a}_j = (a_{vj})_{v=1}^V$  der Matrix  $\mathbf{A}$  jeweils durch  $(a_{vj}/w_v)_{v=1}^V$  ersetzt, aber  $\mathbf{c}^*$  bleibt unverändert. Mit anderen Worten, Zeile Nr.  $v$  von  $\mathbf{A}$  wird durch  $w_v$  dividiert. Dann führt man die Kalibration wie gehabt durch und erhält einen Vektor  $\mathbf{x}$ . Letztlich ersetzt man diesen Vektor durch  $(x_v/w_v)_{v=1}^V$ . Auf diese Weise erhält man eine Minimalstelle der Zielfunktion  $F(\mathbf{x}, \hat{\mathbf{x}})$  mit dem gewichteten euklidischen Abstand  $d_V(\cdot, \cdot)$  anstelle des üblichen euklidischen Abstands.

## 8 Schwache Kontraktivität

Ein Kritikpunkt an der SIMBA-Kalibrierung war, dass oftmals

$$\limsup_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|\mathbf{K}(\mathbf{x}^* + \mathbf{h}) - \mathbf{x}^*\|}{\|\mathbf{h}\|} > 1.$$

Die linke Seite der vorangehenden Ungleichung ist der grösste Singulärwert der Matrix  $D\mathbf{K}(\mathbf{x}^*) \in \mathbb{R}^{V \times V}$ , also die Zahl

$$\max_{\mathbf{h} \neq \mathbf{0}} \frac{\|D\mathbf{K}(\mathbf{x}^*)\mathbf{h}\|}{\|\mathbf{h}\|}.$$

Eine naheliegende Frage ist, ob die zuvor beschriebenen Varianten der Distanzminimierung in dieser Hinsicht besser abschneiden. Die entsprechenden Kalibrationsabbildungen sind

$$\mathbf{0} \leq \mathbf{y} \mapsto \psi(\mathbf{y}) := \arg \min_{\mathbf{x} \in \mathbb{R}^V : \mathbf{x} \geq \mathbf{x}^{\min}} F(\mathbf{x}, \mathbf{y})$$

mit  $F = F_1$  oder  $F = F_2$ . Im Falle von  $\mathbf{x}^* > \mathbf{x}^{\min}$  gelten für die Abbildung  $\psi(\cdot)$  folgende Taylor-Entwicklungen: Wenn  $\mathbb{R}^V \ni \mathbf{h} \rightarrow \mathbf{0}$ , dann ist

$$\psi(\mathbf{x}^* + \mathbf{h}) = \psi(\mathbf{x}^*) - \left( \mathbf{I} + \frac{\lambda}{1-\lambda} \mathbf{A} \mathbf{A}^\top \right)^{-1} \mathbf{h} + O(\|\mathbf{h}\|^2) \quad \text{falls } F = F_1, \quad (3)$$

$$\psi(\mathbf{x}^* + \mathbf{h}) = \psi(\mathbf{x}^*) - \left( \Pi(\mathbf{x}^*) + \frac{\lambda}{1-\lambda} \mathbf{A} \mathbf{A}^\top \right)^{-1} \Pi(\mathbf{x}^*) \mathbf{h} + O(\|\mathbf{h}\|^2) \quad \text{falls } F = F_2. \quad (4)$$

Offensichtlich ist stets

$$\left\| \left( \mathbf{I} + \frac{\lambda}{1-\lambda} \mathbf{A} \mathbf{A}^\top \right)^{-1} \mathbf{h} \right\| \leq \|\mathbf{h}\|.$$

Im Falle von  $F = F_1$  ist also schwache Kontraktivität gegeben. Bei Verwendung von  $F = F_2$  geht die schwache Kontraktivität leider verloren. Numerische Beispiele zeigen, dass oftmals der erste Singulärwert der Matrix

$$\left( \Pi(\mathbf{x}^*) + \frac{\lambda}{1-\lambda} \mathbf{A} \mathbf{A}^\top \right)^{-1} \Pi(\mathbf{x}^*)$$

strikt grösser als 1 ist (wenn auch kleiner als der entsprechende Singulärwert für die SIMBA-Methode), und alle übrigen sind kleiner oder gleich 1.

**Beweis von (3) und (4).** Wir ignorieren vorübergehend die Nebenbedingung  $\mathbf{x} \geq \mathbf{x}^{\min}$  und setzen

$$\tilde{\psi}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathbb{R}^V} F(\mathbf{x}, \mathbf{y})$$

für  $\mathbf{y} \geq \mathbf{0}$ . Angenommen, wir können zeigen, dass

$$\tilde{\psi}(\mathbf{x}^* + \mathbf{h}) = \mathbf{x}^* + \mathbf{M}(\mathbf{x}^*) \mathbf{h} + O(\|\mathbf{h}\|^2) \quad (5)$$

für  $\mathbf{h} \rightarrow \mathbf{0}$  mit einer gewissen Matrix  $\mathbf{M}(\mathbf{x}^*)$ . Dann folgt aus  $\mathbf{x}^* > \mathbf{x}^{\min}$ , dass auch  $\tilde{\psi}(\mathbf{x}^* + \mathbf{h}) > \mathbf{x}^{\min}$  für hinreichend kleine Werte von  $\|\mathbf{h}\|$ , und dies impliziert die Gleichung  $\psi(\mathbf{x}^* + \mathbf{h}) = \tilde{\psi}(\mathbf{x}^* + \mathbf{h})$ .

Im Falle von  $F = F_1$  folgt aus

$$\nabla F_1(\mathbf{x}, \mathbf{x}^* + \mathbf{h}) = \mathbf{H}_1 \mathbf{x} - ((1-\lambda)(\mathbf{x}^* + \mathbf{h}) + \lambda \mathbf{A} \mathbf{c}^*) = \mathbf{H}_1 \mathbf{x} - \mathbf{H}_1 \mathbf{x}^* - (1-\lambda) \mathbf{h},$$

dass

$$\tilde{\psi}(\mathbf{x}^* + \mathbf{h}) = \mathbf{x}^* + \mathbf{H}_1^{-1} (1-\lambda) \mathbf{h} = (\mathbf{I} + \mu \mathbf{A} \mathbf{A}^\top)^{-1} \mathbf{h}$$

mit  $\mu := \lambda/(1-\lambda)$ . Also ist (5) sogar eine Gleichung ohne Restterm  $O(\|\mathbf{h}\|^2)$  mit der Matrix  $\mathbf{M}(\mathbf{x}^*) = (\mathbf{I} + \mu \mathbf{A} \mathbf{A}^\top)^{-1}$ .

Im Falle von  $F = F_2$  folgt aus

$$\nabla F_2(\mathbf{x}, \mathbf{x}^* + \mathbf{h}) = \mathbf{H}_2(\mathbf{x}^* + \mathbf{h}) \mathbf{x} - \lambda \mathbf{A} \mathbf{c}^*,$$

dass

$$\tilde{\psi}(\mathbf{x}^* + \mathbf{h}) = \mathbf{H}_2(\mathbf{x}^* + \mathbf{h})^{-1} \lambda \mathbf{A} \mathbf{c}^*.$$

Speziell für  $\mathbf{h} = \mathbf{0}$  ergibt sich die Formel

$$\mathbf{H}_2(\mathbf{x}^*)^{-1} \lambda \mathbf{A} \mathbf{c}^* = \mathbf{x}^*,$$

denn mit  $\Pi^* := \Pi(\mathbf{x}^*)$  folgt aus  $\Pi^* \mathbf{x}^* = \mathbf{0}$  und  $\mathbf{c}^* = \mathbf{A}^\top \mathbf{x}^*$ , dass

$$\mathbf{H}_2(\mathbf{x}^*) \mathbf{x}^* = ((1 - \lambda) \Pi^* + \lambda \mathbf{A} \mathbf{A}^\top) \mathbf{x}^* = \lambda \mathbf{A} \mathbf{A}^\top \mathbf{x}^* = \lambda \mathbf{A} \mathbf{c}^*.$$

Mit  $\mathbf{u}^* := \mathbf{u}(\mathbf{x}^*)$  gilt die Entwicklung

$$\begin{aligned} \mathbf{u}(\mathbf{x}^* + \mathbf{h}) &= (\|\mathbf{x}^*\|^2 + 2\mathbf{h}^\top \mathbf{x}^* + \|\mathbf{h}\|^2)^{-1/2} (\mathbf{x}^* + \mathbf{h}) \\ &= \|\mathbf{x}^*\|^{-1} (1 - \|\mathbf{x}^*\|^{-2} \mathbf{h}^\top \mathbf{x}^* + O(\|\mathbf{h}\|^2)) (\mathbf{x}^* + \mathbf{h}) \\ &= \|\mathbf{x}^*\|^{-1} (1 - \|\mathbf{x}^*\|^{-1} \mathbf{h}^\top \mathbf{u}^* + O(\|\mathbf{h}\|^2)) (\mathbf{x}^* + \mathbf{h}) \\ &= \mathbf{u}^* + \|\mathbf{x}^*\|^{-1} (\mathbf{h} - \mathbf{u}^* \mathbf{u}^{*\top} \mathbf{h}) + O(\|\mathbf{h}\|^2) \\ &= \mathbf{u}^* + \|\mathbf{x}^*\|^{-1} \Pi^* \mathbf{h} + O(\|\mathbf{h}\|^2). \end{aligned}$$

Wegen  $\Pi^{*\top} = \Pi^*$  impliziert dies, dass

$$\Pi(\mathbf{x}^* + \mathbf{h}) = \Pi^* + \|\mathbf{x}^*\|^{-1} \mathbf{u}^* \mathbf{h}^\top \Pi^* + \|\mathbf{x}^*\|^{-1} \Pi^* \mathbf{h} \mathbf{u}^{*\top} + O(\|\mathbf{h}\|^2),$$

und wir können schreiben:

$$\begin{aligned} \mathbf{H}_2(\mathbf{x}^* + \mathbf{h}) &= \mathbf{H}_2(\mathbf{x}^*) + \mathbf{D}^*(\mathbf{h}) + O(\|\mathbf{h}\|^2), \\ \mathbf{D}^*(\mathbf{h}) &= (1 - \lambda) \|\mathbf{x}^*\|^{-1} (\mathbf{u}^* \mathbf{h}^\top \Pi^* + \Pi^* \mathbf{h} \mathbf{u}^{*\top}). \end{aligned}$$

Hieraus folgt, dass

$$\begin{aligned} \tilde{\psi}(\mathbf{x}^* + \mathbf{h}) &= (\mathbf{H}_2(\mathbf{x}^*) + \mathbf{D}^*(\mathbf{h}) + O(\|\mathbf{h}\|^2))^{-1} \lambda \mathbf{A} \mathbf{c}^* \\ &= (\mathbf{I} + \mathbf{H}_2(\mathbf{x}^*)^{-1} \mathbf{D}^*(\mathbf{h}) + O(\|\mathbf{h}\|^2))^{-1} \mathbf{H}_2(\mathbf{x}^*)^{-1} \lambda \mathbf{A} \mathbf{c}^* \\ &= (\mathbf{I} - \mathbf{H}_2(\mathbf{x}^*)^{-1} \mathbf{D}^*(\mathbf{h}) + O(\|\mathbf{h}\|^2)) \mathbf{x}^* \\ &= \mathbf{x}^* - \mathbf{H}_2(\mathbf{x}^*)^{-1} \mathbf{D}^*(\mathbf{h}) \mathbf{x}^* + O(\|\mathbf{h}\|^2) \\ &= \mathbf{x}^* - \mathbf{H}_2(\mathbf{x}^*)^{-1} (1 - \lambda) (\mathbf{u}^* \mathbf{h}^\top \Pi^* + \Pi^* \mathbf{h} \mathbf{u}^{*\top}) \mathbf{u}^* + O(\|\mathbf{h}\|^2) \\ &= \mathbf{x}^* - \mathbf{H}_2(\mathbf{x}^*)^{-1} (1 - \lambda) \Pi^* \mathbf{h} + O(\|\mathbf{h}\|^2) \\ &= \mathbf{x}^* - (\Pi^* + \mu \mathbf{A} \mathbf{A}^\top)^{-1} \Pi^* \mathbf{h} + O(\|\mathbf{h}\|^2), \end{aligned}$$

denn  $\Pi^* \mathbf{u}^* = \mathbf{0}$  und  $\mathbf{u}^{*\top} \mathbf{u}^* = 1$ . Dies liefert (5) mit  $\mathbf{M}(\mathbf{x}_*) = (\Pi^* + \mu \mathbf{A} \mathbf{A}^\top)^{-1} \Pi^*$ .  $\square$

## Literatur

- [1] L. BOSSART (2016). *Schätzung von wochentags-scharfen Quell-Ziel-Matrizen*. Masterarbeit, IMSV, Universität Bern.
- [2] L. DÜMBGEN (2017). *Optimization Methods*. Lecture Notes, IMSV, University of Bern.