Numerische Mathematik

§4 Überbestimmte lineare Gleichungssysteme

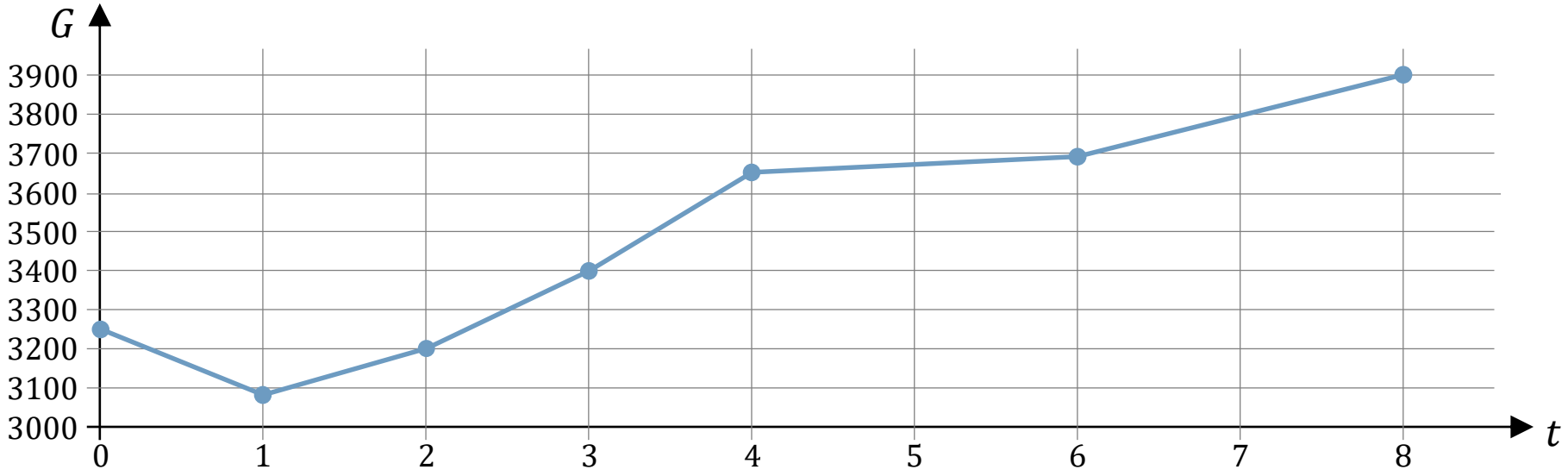
Prof. Dr. Georg Umlauf

Beispiel: Überbestimmtes LGS

Beispiel (Teil 1):

- Das Körpergewicht von Neugeborenen nimmt in den ersten Wochen annähernd linear zu**.**

t in Wochen	0	1	2	3	4	6	8
G in Gramm	3250	3070	3200	3400	3650	3680	3900



Beispiel: Überbestimmtes LGS

- Gesucht ist eine Gerade $g = c_0 + c_1 t$, die diesen Wachstumsprozess möglichst gut wiedergibt.
- $m = 7$ Messungen und $n = 2$ Variablen.

- Annahme: Es gibt eine Gerade, die alle Bedingungen erfüllt:
  - LGS

$$\begin{aligned} c_0 + c_1 \cdot 0 &= 3250 \\ c_0 + c_1 \cdot 1 &= 3070 \\ &\vdots \\ c_0 + c_1 \cdot 8 &= 3900 \end{aligned}$$

Beispiel: Überbestimmtes LGS

- Matrixform:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 6 & 8 \end{pmatrix}^t,$$

$$b = (3250 \quad 3070 \quad \cdots \quad 3900)^t,$$

$$x = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}.$$

- Normalgleichung:
  - minimiere den Fehler $\|Ax - b\|_2^2$
  - $A^t \cdot A \cdot x = A^t \cdot b$

Beispiel: Überbestimmtes LGS

Beispiel (Teil 2):

- Normalgleichung:

$$A^t A = \begin{pmatrix} 7 & 24 \\ 24 & 130 \end{pmatrix}, A^t b = \begin{pmatrix} 24150 \\ 87550 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 24 \\ 24 & 130 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = A^t\, Ax = A^t b = \begin{pmatrix} 24150 \\ 87550 \end{pmatrix}$$

- Näherung für Gewicht bei der Geburt: $c_0 = \frac{1038300}{334} \approx 3110$

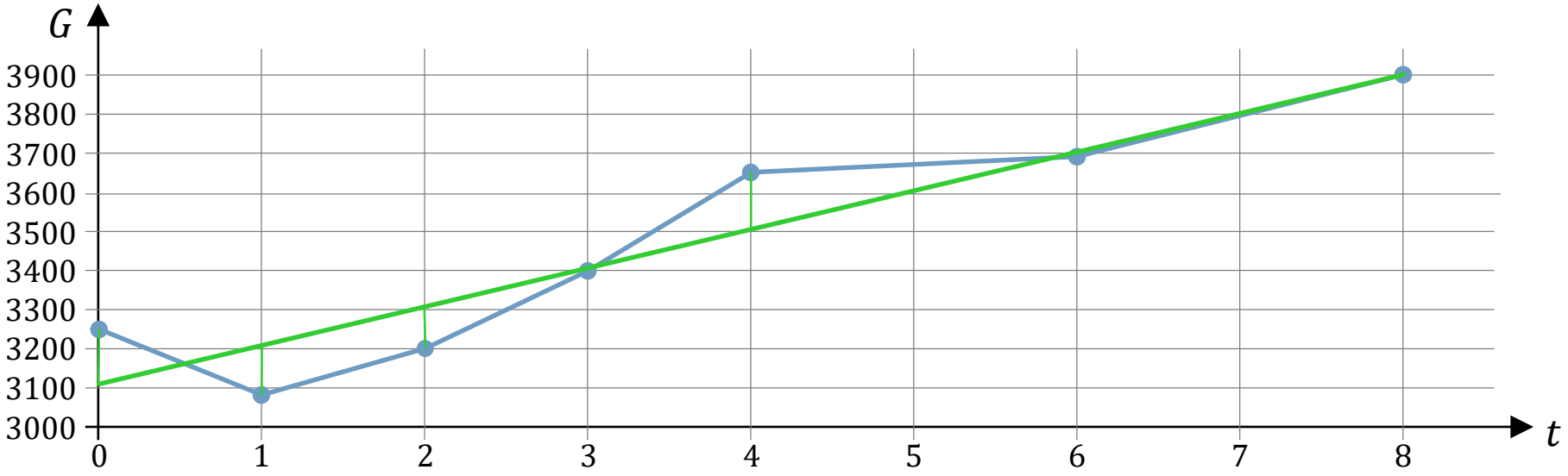
  Wöchentliche Zunahme: $c_1 = \frac{33250}{334} \approx 100$

Beispiel: Überbestimmtes LGS

- Näherung für Gewicht bei der Geburt: $c_0 = \frac{1038300}{334} \approx 3110$

  Wöchentliche Zunahme: $c_1 = \frac{33250}{334} \approx 100$

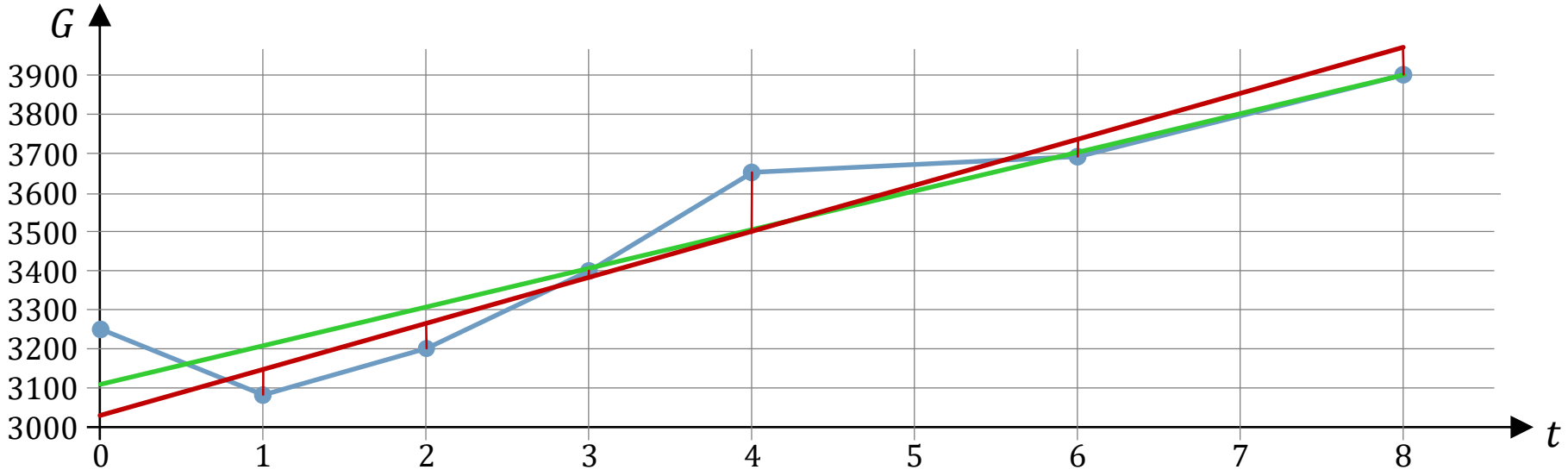
t in Wochen	0	1	2	3	4	6	8
G in Gramm	3250	3070	3200	3400	3650	3680	3900



Beispiel: Überbestimmtes LGS

- Vernachlässigung der Gewichtsabnahme in der ersten Woche, d.h. Verzicht auf die erste Gleichung:
- $c_0 = \frac{615800}{204} \approx 3020, \quad c_1 = \frac{23700}{204} \approx 116.$

t in Wochen	0	1	2	3	4	6	8
G in Gramm	3250	3070	3200	3400	3650	3680	3900



Beispiel: Normalgleichungen sind u.U. instabil

Beispiel:

$$A = \begin{pmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{pmatrix} \text{ mit } \operatorname{rang}(A) = 2.$$

- Normalgleichung:

$$A^t A = \begin{pmatrix} 1 + 10^{-8} & 1 \\ 1 & 1 + 10^{-8} \end{pmatrix}$$

- achtstellige Dezimalarithmetik liefert:

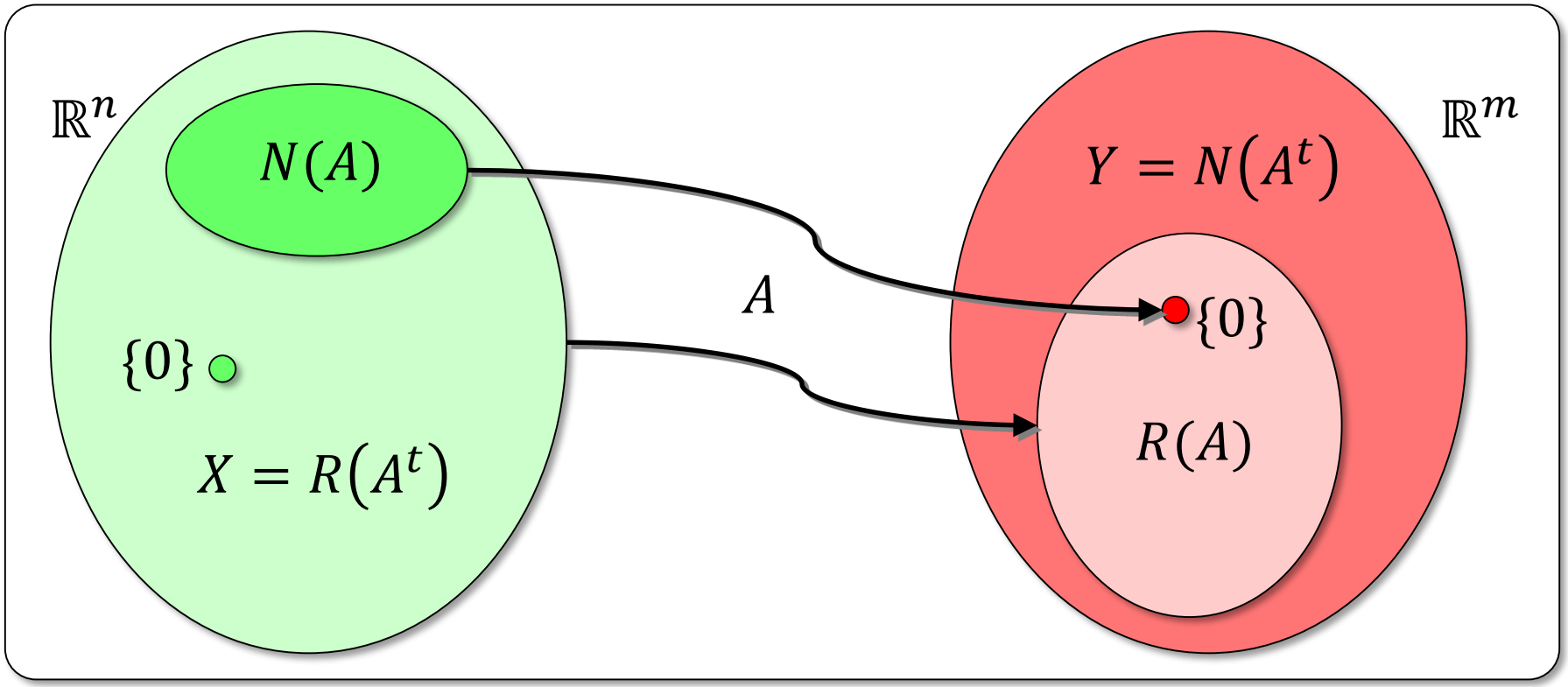
$$\det(\operatorname{rd}(A^t A)) = \det\left(\operatorname{rd}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) = 0.$$

Prinzip der SVD

Lineare Abbildung $A: \mathbb{R}^n \to \mathbb{R}^m$:

- $\mathbb{R}^n = N(A) \oplus X = N(A) \oplus R(A^t)$
- $\mathbb{R}^m = R(A) \oplus Y = R(A) \oplus N(A^t)$

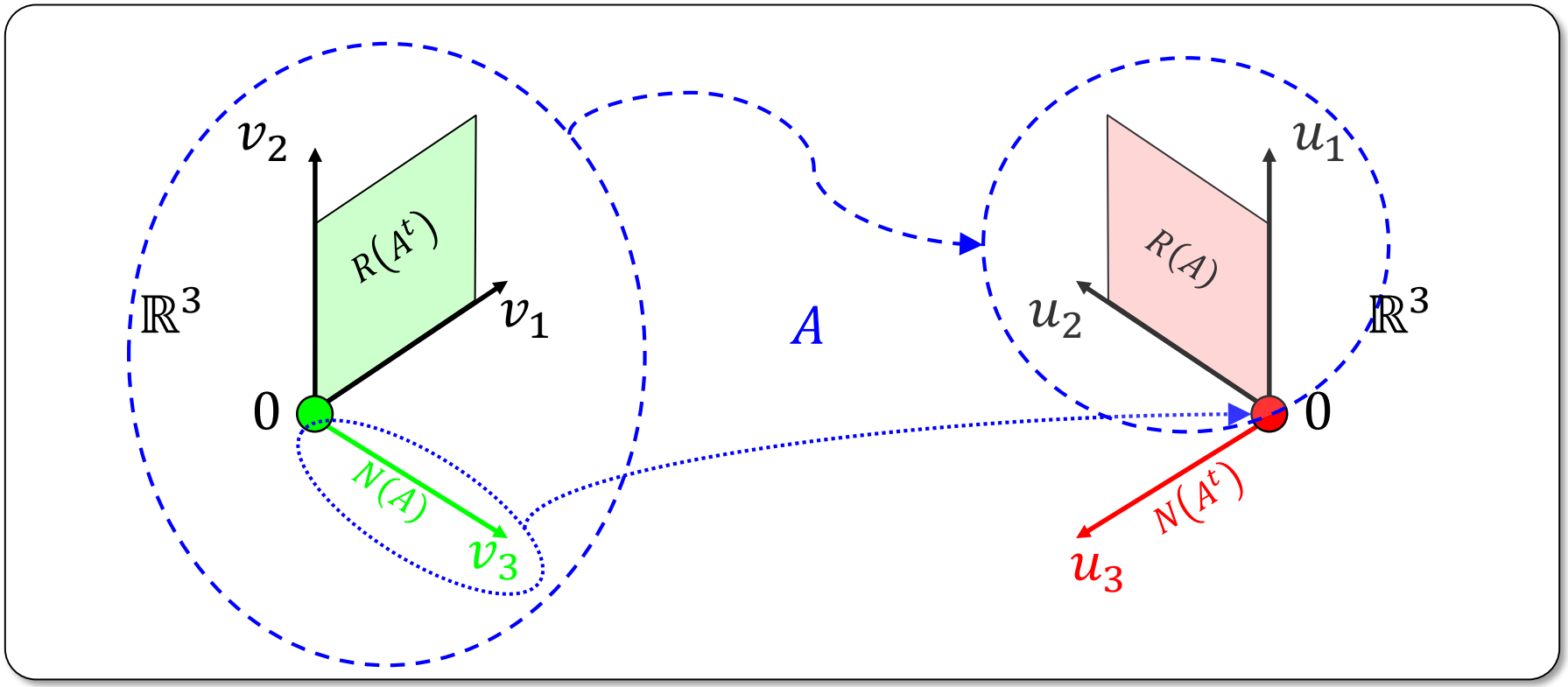
} Wähle geeignete ONBen



Prinzip der SVD

Lineare Abbildung $A: \mathbb{R}^n \to \mathbb{R}^m$:

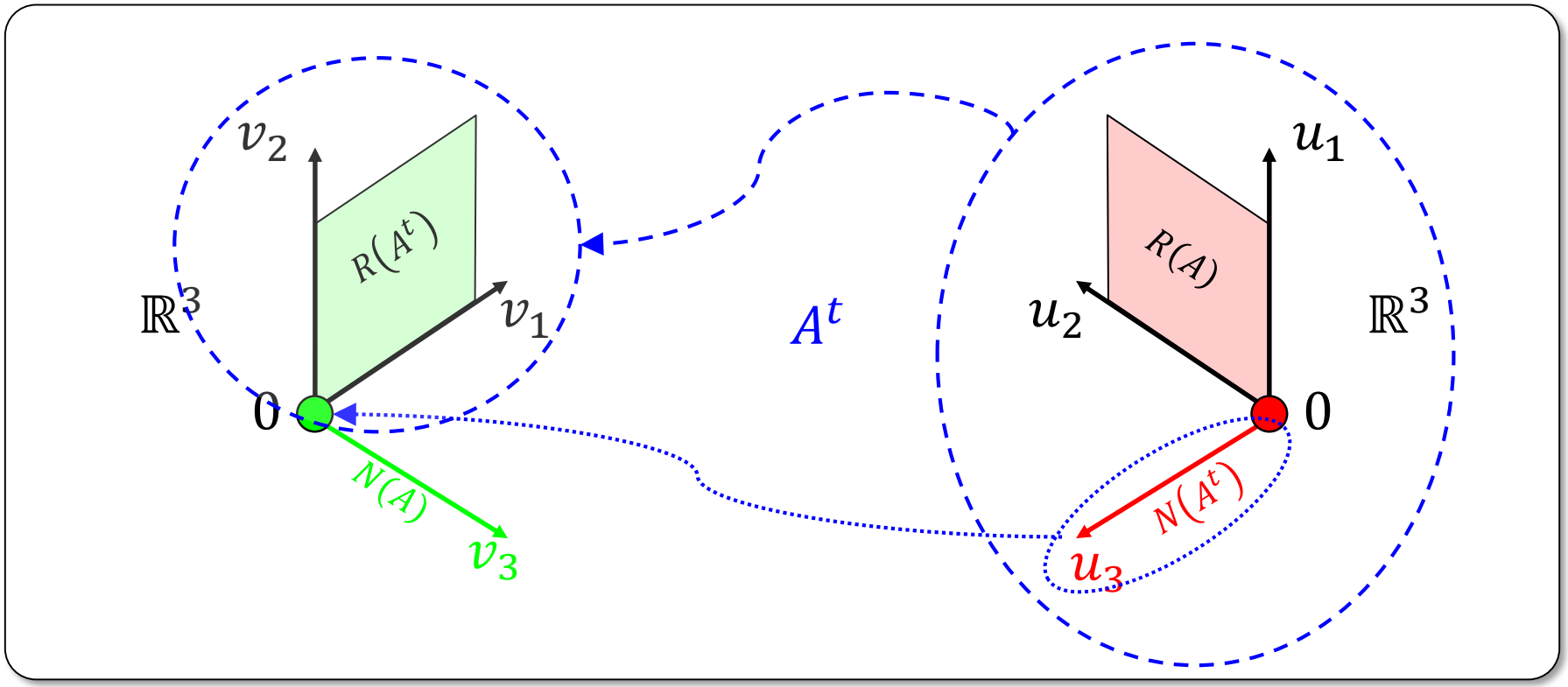
- $\mathbb{R}^n = R(A^t) \oplus N(A) = \langle v_1, \ldots, v_r, v_{r+1}, \ldots, v_n \rangle$
- $\mathbb{R}^m = R(A) \oplus N(A^t) = \langle u_1, \ldots, u_r, u_{r+1}, \ldots, u_m \rangle$



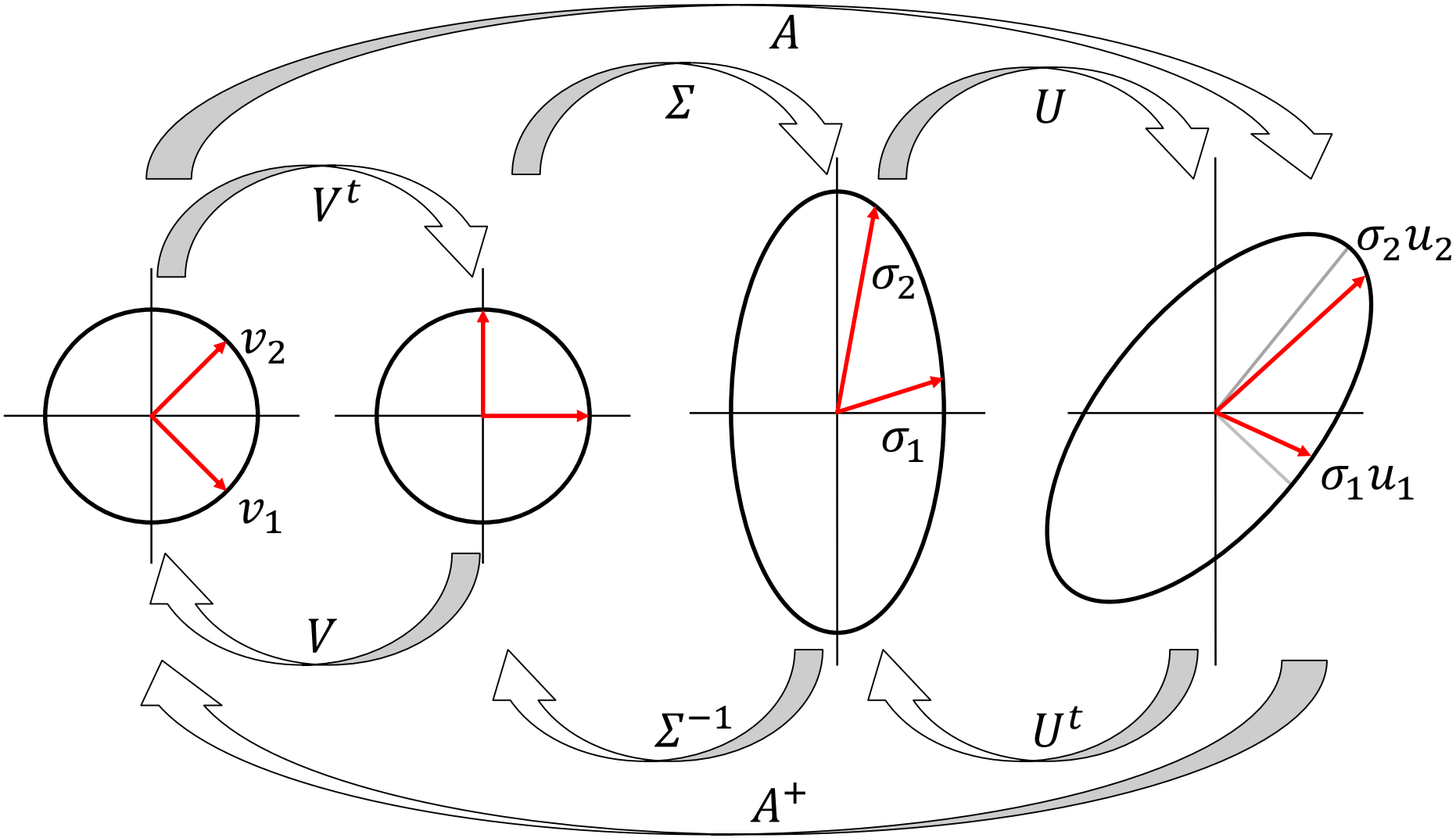
Prinzip der SVD

Lineare Abbildung $A: \mathbb{R}^n \to \mathbb{R}^m$:

- $\mathbb{R}^n = R(A^t) \oplus N(A) = \langle v_1, \ldots, v_r, v_{r+1}, \ldots, v_n \rangle$
- $\mathbb{R}^m = R(A) \oplus N(A^t) = \langle u_1, \ldots, u_r, u_{r+1}, \ldots, u_m \rangle$



Prinzip der SVD und Pseudoinversen

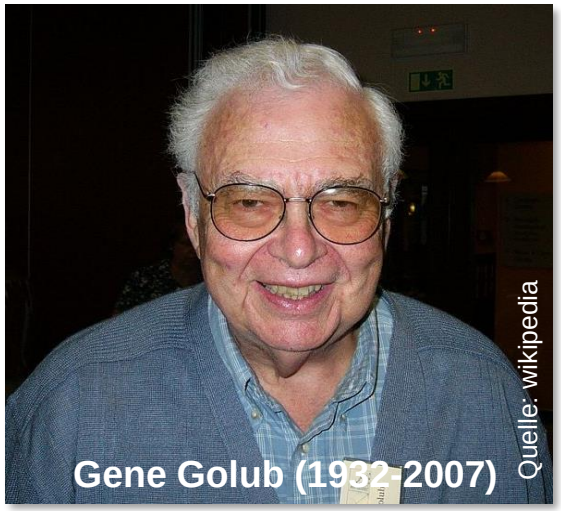


Moore/Penrose





Golub/Kahan/Reinsch







Anwendung: SVD

Anwendungen für die SVD

- Lösen von singulären, schlecht konditionierten, unterbestimmten oder überbestimmten LGSen.
- Matrixapproximation mit niedrigerem Rang
- Bildkompression
- Rauschunterdrückung
- Finden von Konzepten oder Themen, Latent Semantic Indexing (LSI)

Beispiel: SVD

- Beispiel: Was ist die SVD zu

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}?$$

1. Berechne

$$A^t A = \left(V \Sigma^t U^t\right)\left(U \Sigma V^t\right) = V \Sigma^2 V^t = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

mit Eigenvektoren $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ und $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ zu den Eigenwerten 8 und 2.

- Normierung: $v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ und $v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

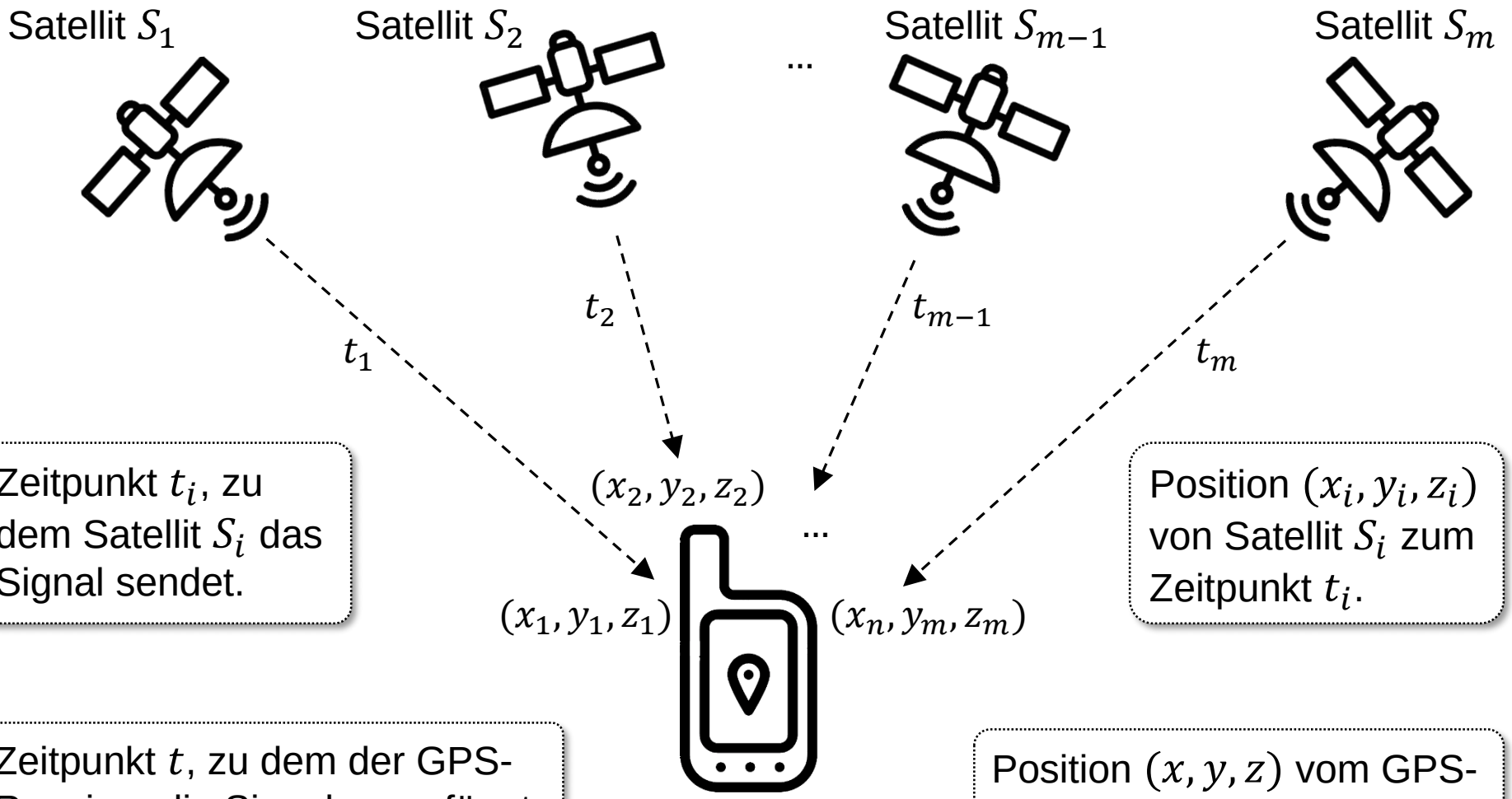
Beispiel: SVD

2. Berechne $Av_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}2\\0\end{bmatrix}$ und $Av_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\end{bmatrix}$, d.h. $u_1 = \begin{bmatrix}1\\0\end{bmatrix}$ und $u_2 = \begin{bmatrix}0\\1\end{bmatrix}$.
3. Da $Av_1 = 2\sqrt{2}u_1$ ist, ist $\sigma_1 = 2\sqrt{2}$ und analog $\sigma_2 = \sqrt{2}$.

- Die vollständige SVD ist

$$\begin{bmatrix}2 & 2\\-1 & 1\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2\sqrt{2} & 0\\0 & \sqrt{2}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\0 & \sqrt{2}\end{bmatrix}.$$

Beispiel: GPS



Zeitpunkt t_i, zu dem Satellit S_i das Signal sendet.

Position (x_i, y_i, z_i) von Satellit S_i zum Zeitpunkt t_i.

Zeitpunkt t, zu dem der GPS-Receiver die Signale empfängt.

Position (x, y, z) vom GPS-Receiver zum Zeitpunkt t.

Beispiel: GPS

- Der GPS-Receiver empfängt Signale von m Satelliten S_i, $i = 1, \ldots, m$.
- Zeitpunkt t_i, zu dem Satellit S_i das Signal sendet, ist **hochgenau bekannt**.
- Position $p_i = (x_i, y_i, z_i)$ von Satellit S_i zum Zeitpunkt t_i ist dem GPS-Receiver **bekannt**.

} $4m$ bekannte Parameter

- Zeitpunkt t, zu dem die Satelliten-Signale bei GPS-Receiver empfangen werden, ist nur **relativ ungenau bekannt**.
- Position $p = (x, y, z)$ vom GPS-Receiver zum Zeitpunkt t ist **unbekannt**.

} 4 unbekannte Parameter

Beispiel: GPS

- Beziehung zwischen den Größen

$$\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} = c(t-t_i), i = 1, \dots, m,$$

  mit der Lichtgeschwindigkeit c.

- Da die Empfangszeit t sehr ungenau ist, wird ein *clock-bias* b verwendet, d.h.

$$t = \tilde{t} - b,$$

  wobei $\tilde{t}$ die gemessene Zeit ist.

- n nichtlineare Gleichungen in den vier Unbekannten x, y, z, b

$$\|p - p_i\|_2 = c(\tilde{t} - t_i) - cb, \qquad i = 1, \dots, m.$$

Beispiel: GPS

- Approximiere diese Gleichungen durch Minimierung von

$$Q(x,y,z,b) = \sum_{i=1}^{m} (\|p - p_i\|_2 - c(\tilde{t} - t_i) + cb)^2$$

$$= \sum_{i=1}^{m} (r_i(x,y,z,b))^2 = \sum_{i=1}^{m} (r_i(w))^2$$

1. Idee: Berechne Nullstelle der ersten Ableitung von Q mit dem Newton-Verfahren.

- Problem: Berechnung der zweiten Ableitungen!

2. Idee: Linearisiere die Residuen $r_i(w)$…

Beispiel: GPS

2. Idee: Linearisiere die Residuen $r_i(w) = r_i(x, y, z, b)$ an einem Startwert $\left(x^0, y^0, z^0, b^0\right) =: w^0$

$$r_i(w) \approx r_i\left(w^0\right) + \nabla r_i\left(w^0\right)^T \cdot \left(w - w^0\right)$$

mit dem Gradienten

$$\nabla r_i(w) = \frac{1}{\|p - p_i\|_2} \begin{pmatrix} x - x_i \\ y - y_i \\ z - z_i, \\ c\|p - p_i\|_2 \end{pmatrix}.$$

- Minimiere die linearisierte Zielfunktion $\bar{Q}$ mit $\Delta w := w - w^0$

$$\bar{Q}(x, y, z, b) = \sum_{i=1}^{m} \left(r_i\left(w^0\right) + \nabla r_i\left(w^0\right)^T \cdot \Delta w\right)^2.$$

Beispiel: GPS

- Minimierung von $\bar{Q}(x, y, z, b)$ ist dann äquivalent zur kleinsten-Quadrate Lösung des LGS

$$\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_1 & b_2 & c_2 & d_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_m & b_m & c_m & d_m \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta b \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

mit den Unbekannten x, y, z, b bzw.

$$\Delta x = x - x^0, \Delta y = y - y^0, \Delta z = z - z^0, \Delta b = b - b^0$$

und den bekannten Parameter $f_i := r_i\left(w^0\right)$ und

$$(a_i, b_i, c_i, d_i) := \nabla r_i\left(w^0\right)^T, i = 1, \dots, m.$$

Beispiel: GPS

- Dies liefert eine verbesserte Lösung $\left(x^1, y^1, z^1, b^1\right)$

$$x^1 = x^0 + \Delta x,$$
$$y^1 = y^0 + \Delta y,$$
$$z^1 = z^0 + \Delta z,$$
$$b^1 = b^0 + \Delta b$$

  … und iteriere dieses Vorgehen.

- Dieses Vorgehen zum Minimieren der Summe von Quadraten n-variater, nicht-linearer Funktionen heißt **Gauß-Newton-Verfahren**.

Beispiel: Gauß-Newton-Verfahren

- Dieses Vorgehen zum Minimieren der Summe von Quadraten n-variater, nicht-linearer Funktionen $f_i \in \mathrm{C}^1, f_i\colon \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m,$

$$\min_{x\in\mathbb{R}^n}\left(\frac{1}{2}\sum_{i=1}^{m}\big(f_i(x)\big)^2\right) = \min_{x\in\mathbb{R}^n}\left(\frac{1}{2}\|F(x)\|_2^2\right), \qquad F\colon \mathbb{R}^n \to \mathbb{R}^m,$$

  heißt **Gauß-Newton-Verfahren:**

1. Linearisiere die zu minimierende nicht-lineare Funktion $F = (f_1, \dots, f_m)^T$ an einem Startwert $x^0 \in \mathbb{R}^n$

$$\bar{F}(x) = F\big(x^0\big) + J_F\big(x^0\big)\big(x - x^0\big), \qquad \underbrace{J_F(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{i,j} \in \mathbb{R}^{m\times n}.}_{\text{aka Jacobi-Matrix}}$$

2. …

Beispiel: Gauß-Newton-Verfahren

- … heißt **Gauß-Newton-Verfahren:**

1. Linearisiere die zu minimierende nicht-lineare Funktion $F = (f_1, \dots, f_m)^T$ an einem Startwert $x^0 \in \mathbb{R}^n$ mit $\Delta x = x - x^0$

$$\bar{F}(x) = F\left(x^0\right) + J_F\left(x^0\right) \cdot \Delta x, \qquad J_F(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{i,j} \in \mathbb{R}^{m \times n}.$$

2. Löse das so erhaltene überbestimmte lineare Gleichungssystem mit der Methode der kleinsten Quadrate, d.h.

$$\min_{x \in \mathbb{R}^n} \left(\frac{1}{2}\left\|\bar{F}(x)\right\|_2^2\right) = \min_{x \in \mathbb{R}^n} \left(\frac{1}{2}\sum_{i=1}^{m}\left(F\left(x^0\right) + J_F\left(x^0\right) \cdot \Delta x\right)^2\right)$$

wird gelöst mit den Normalgleichungen

$$x = x^0 - \left(J_F^T\left(x^0\right) \cdot J_F\left(x^0\right)\right)^{-1} \cdot J_F^T\left(x^0\right) \cdot F\left(x^0\right).$$

Beispiel: Gauß-Newton-Verfahren

- … heißt **Gauß-Newton-Verfahren:**

1. …
2. Löse das so erhaltene überbestimmte LGS mit der Methode der kleinsten Quadrate, d.h.

$$\min_{x\in\mathbb{R}^n}\left(\frac{1}{2}\left\|\bar{F}(x)\right\|_2^2\right)=\min_{x\in\mathbb{R}^n}\left(\frac{1}{2}\sum_{i=1}^{m}\left(F\left(x^0\right)+J_F\left(x^0\right)\cdot\Delta x\right)^2\right)$$

   wird gelöst mit den Normalgleichungen

$$x=x^0-\left(J_F^T\left(x^0\right)\cdot J_F\left(x^0\right)\right)^{-1}\cdot J_F^T\left(x^0\right)\cdot F\left(x^0\right).$$

3. Damit ergibt sich der **Gauß-Newton-Schritt** zur Schrittweite $\alpha_k>0$

$$x^{k+1}=x^k-\alpha_k\left(J_F^T\left(x^k\right)\cdot J_F\left(x^k\right)\right)^{-1}\cdot J_F^T\left(x^k\right)\cdot F\left(x^k\right),k=0,1,\ldots$$

Beispiel: Gauß-Newton-Verfahren

- … heißt **Gauß-Newton-** bzw. **Levenberg-Marquardt-Verfahren:**

1. …
2. …
3. Damit ergibt sich der **Gauß-Newton-Schritt (GN-Schritt)**

$$x^{k+1} = x^k - \alpha_k \left(J_F^T\left(x^k\right) \cdot J_F\left(x^k\right) \right)^{-1} \cdot J_F^T\left(x^k\right) \cdot F\left(x^k\right), k = 0,1, \ldots$$

4. Falls die Matrix $J_F^T\left(x^k\right) \cdot J_F\left(x^k\right) \in \mathbb{R}^{n \times n}$ singulär oder schlecht konditioniert ist, wähle den **Levenberg-Marquardt-Schritt (LM-Schritt)**

$$x^{k+1} = x^k - \alpha_k \left(J_F^T\left(x^k\right) \cdot J_F\left(x^k\right) + \beta_k I_n \right)^{-1} \cdot J_F^T\left(x^k\right) \cdot F\left(x^k\right)$$

mit der $n \times n$ Einheitsmatrix I_n und $\beta_k \geq 0$.

- Für $\beta_k = 0$ reduziert sich der LM-Schritt zu einem GN-Schritt.
- Für $\beta_k \gg 0$ reduziert sich der LM-Schritt zu einem **Gradienten-Schritt**.

Beispiel: Levenberg-Marquardt



