

COMPRESSIVE SENSING

Vorlesung im Sommersemester 2025 von Prof. Dr. Ivan Veselić

Vorwort

Dieses Skript ist aus Vorlesungen an der Technischen Universität Dortmund im Wintersemester 2016/17 und im Sommersemester 2025 entstanden. Im zweiten Fall wurden die einzelnen Vorlesungen mit Kamera aufgenommen und bei YouTube veröffentlicht. Die Vorlesung beruht in großen Teilen auf dem Buch *A Mathematical Introduction to Compressive Sensing* von Simon Foucart und Holger Rauhut [FR13]. Desweiteren stand mir 2016 ein handschriftliches Manuskript von Henning Kempka zu einer Vorlesung zum selben Thema, dass er an der Technischen Universität Chemnitz gehalten hat, zu Verfügung. Dafür möchte ich mich herzlich bedanken. Sein Manuskript wiederum beruht auf demselben Buch [FR13]. Einige Abschnitte dieses Skripts sind von den Originalarbeiten [CP88, BDF⁺11, CT05, CT06, CRT06b, CRT06a, FPRU10, LN17] inspiriert. Der größte Teil des Textes wurde von Herr Jens Koniarski in LaTeX gesetzt, andere Teile von Daniel Rosen, Michela Egidi, Janine Textor und von mir. Alexander Dicke, Michela Egidi und Daniel Rosen haben das Skript Korrektur gelesen. Daniel Rosen und Jakob Schütz haben die Kameraaufnahmen und die Nachbearbeitung der Videos übernommen. Daniel Rosen hat mich freundlicherweise mehrmals bei den Vorlesungen vertreten und der Aufnahme zugestimmt. Bei all diesen Personen möchte ich mich bei dieser Gelegenheit ganz herzlich bedanken.

Trotz des Korrekturlesens ist in die Abfassung des Textes wesentlich weniger Arbeit investiert worden, als dies bei einem Buch der Fall ist. Es gibt sicherlich eine Reihe von kleineren Fehlern und ich arbeite — zusammen mit freundlichen Menschen, die mich auf Fehler hinweisen — daran, diese auszumerzen.

Das Skript gibt bewusst recht treu den Vortragsstil an der Tafel wieder und will kein Buch sein. Die URLs der Videos auf YouTube sind an den entsprechenden Stellen im Skript verlinkt. An manchen Stellen gibt es kleinere Abweichungen zwischen der aufgezeichneten Tafelvorlesungen und dem Skript.

Das Skript soll dazu dienen, dem Leser eine Auswahl des Stoffes zu liefern, die in eine vierstündige Vorlesung passt. Als Referenzen sollte man die angegebene Hauptquelle [FR13] stets in der Hinterhand haben.

Ein allgemeiner Hinweis zum Sinn und Zweck von Skripten: Das Lesen von Vorlesungsnotizen und Skripten kann das eigene Denken und Nachvollziehen der Sachverhalte, Schritte und Argumente nicht ersetzen. Das eigentliche Ziel ist die Verinnerlichung der mathematischen Wahrheiten, so dass man sie wie etwas Eigenes beherrscht und anderen erklären kann. Falls Ihnen an einzelnen Stellen ein Skript zu knapp formuliert ist, um die dargestellten Sachverhalte zu verstehen, konsultieren Sie die angegebene Literatur.

Ivan Veselić

Dortmund im Juli 2025

Literaturverzeichnis

- [BDF⁺11] Jean Bourgain, Stephen Dilworth, Kevin Ford, Sergei Konyagin & Denka Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Math. J.*, 159(1):145–185, 2011.
- [CRT06a] Emmanuel J. Candès, Justin Romberg & Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [CRT06b] Emmanuel J. Candès, Justin K. Romberg & Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [CT05] Emmanuel J. Candes and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [CT06] Emmanuel J. Candes and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
- [CP88] Bernd Carl and Alain Pajor. Gel’fand numbers of operators with values in a Hilbert space. *Invent. Math.*, 94(3):479–504, 1988.
- [FPRU10] Simon Foucart, Alain Pajor, Holger Rauhut & Tino Ullrich. The Gelfand widths of ℓ_p -balls for $0 < p \leq 1$. *J. Complexity*, 26(6):629–640, 2010.
- [FR13] Simon Foucart & Holger Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.
- [GG84] A. Yu. Garnaev & E. D. Gluskin. The widths of a Euclidean ball. *Dokl. Akad. Nauk SSSR*, 277(5):1048–1052, 1984.
- [LN17] Kasper Green Larsen & Jelani Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 633–638. IEEE Computer Soc., Los Alamitos, CA, 2017.

Inhaltsverzeichnis

Vorwort	1
Literaturverzeichnis	3
Kapitel 0. Motivation und Begrifflichkeiten	7
Kapitel 1. Dünne Lösungen unter-bestimmter Systeme	11
1. Dünne und komprimierbare Vektoren	11
2. Minimal erforderliche Anzahl von Messungen zur Rekonstruktion s -dünner Signale	19
3. Exkurs/Anhang A: Diskrete Fouriertransformation/Fouriermatrizen	22
4. ℓ_0 -Minimierung ist NP -schwer	27
Kapitel 2. ℓ_1 -Minimierung	31
1. Optimierungsprobleme	31
Kapitel 3. Anhang B: Matrix Analysis	53
1. B.1 Normen	53
2. B.2 Singulärwertzerlegung/ singular value decomposition (SVD)	55
3. B.3 Prinzip der kleinsten Quadrate	59
4. B.4 Rekonstruktion von Matrizen	61
Kapitel 4. Kohärenz	63
1. Definition und Eigenschaften	63
2. Kohärenz und Basisjagd	69
Kapitel 5. Restricted Isometry Property (RIP)	73
1. Definition und Eigenschaften	73
2. Vergleich mit Orthogonalität	74
Kapitel 6. Basisjagd (P_1) mit RIP	83
Kapitel 7. Basisjagd mit zufälligen Matrizen	89
Kapitel 8. Exkurs: Das Lemma von Johnson & Lindenstrauss	97
Kapitel 9. CS mit strukturierten zufälligen Matrizen	101
Kapitel 10. Unschärferelation für inkohärente Basen	103
Kapitel 11. Gelfand-Zahlen	107
1. Definition Gelfand-Zahlen, Gelfand-Weiten, und Rekonstruktionsfehler	107
2. Zusammenhang zwischen Gelfand-Weiten und CS	111

Motivation und Begrifflichkeiten

LINK: VORLESUNGSVIDEO 1 VOM 7.4.25

Ursprüngliches Datenobjekt x soll kodiert/verarbeitet/gemessen und in einem Format y gespeichert/versendet/verarbeitet werden.

In der Praxis bestehen sowohl x , als auch y aus endlich vielen Werten, also können sie als Vektoren in einem endlich dimensionalen Raum betrachten werden.

Beachte, dass $y = y(x)$ eine Funktion von x ist, da bei der Erstellung von y keine anderen Informationen verwendet werden außer denen, die in x enthalten sind. In Anwendungen ist y meistens eine lineare Funktion von x , also existiert eine Matrix A so, dass $y = Ax$.¹ Die Matrix A beschreibt also die Prozedur, wie die ursprüngliche Informationen x kodiert, gespeichert oder gemessen wird.

BEISPIEL 0.1. x sind die Rohdaten eines Fotos (einzelne Pixel als Vektor- oder Matrixarray gespeichert), A ist die JPEG oder PNG Kodierung und y ist das Bild in JPEG Format.

Offensichtlich ist hier $y = y(x)$ eine Funktion der Rohdaten.

Je nach Einstellung des Kodiereres ist

- Kodierung ohne Qualität/Informationsverlust möglich,
- Kodierung mit Kompression/Datenreduktion und damit verbundenen Informationsverlust möglich.

Darstellung der Matrix A mit Zeilenvektoren:

$$A = \begin{pmatrix} a_1 \\ \dots \\ \dots \\ \dots \\ a_m \end{pmatrix} = \begin{pmatrix} a_{11}, \dots, a_{1N} \\ \dots \\ a_{j1}, \dots, a_{jN} \\ \dots \\ a_{m1}, \dots, a_{mN} \end{pmatrix}, \quad a_j = (a_{j1}, \dots, a_{jN}), \text{ mit } m, N \in \mathbb{N}.$$

Dann gilt

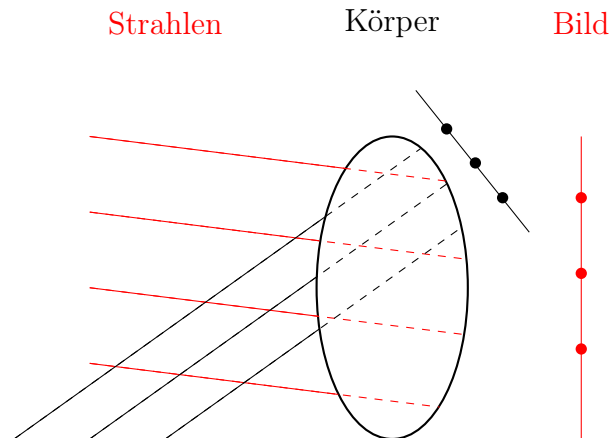
$$y = Ax = \begin{pmatrix} \langle a_1, x \rangle \\ \dots \\ \dots \\ \dots \\ \langle a_m, x \rangle \end{pmatrix}$$

¹In dieser Vorlesung werden wir nur lineare Funktionen $y = y(x) = Ax$ betrachten.

wobei das j te Skalarprodukt gegeben ist durch

$$\langle a_j, x \rangle = \sum_{i=1}^N a_{ji} x_i = (a_{j1}, \dots, a_{jN}) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

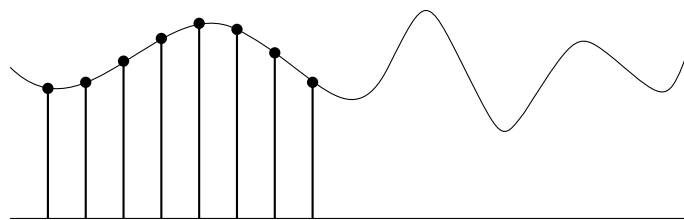
und in den Anwendungen (manchmal) der j -ten Messung entspricht.



BEISPIEL 0.2 (Röntgen-Bilder).

Arzt/ZA schaut sich Aufnahmen aus mehreren Winkeln (Perspektiven) an, um (in seinem Kopf) die tatsächliche dreidimensionale Form des bildgebenden Verfahrens zu rekonstruieren.

BEISPIEL 0.3. Falls man ein analoges Audiosignal digital kodieren will, wird es in diskreten Zeitintervallen, “abgetastet”



und in einem Datenvektor gespeichert. Dieser Vorgang entspricht wieder einer linearen Abbildung A . Daher nennt man auch oft:

- $x \in \mathbb{R}^N$ das Signal
- $y \in \mathbb{R}^m$ die Messung/Messvektor/die Messdaten.

Damit ist $A : \mathbb{R}^N \longrightarrow \mathbb{R}^m$ eine (lineare) Abbildung. In den “interessanten Fällen” ist meistens $m < N$.

Warum?

- JPEG-Bild ist komprimiert und enthält weniger Daten als Originalbild.
- digitaler Mitschnitt einer Konzerts ist weniger präzise als das Klangerlebnis von Ort.

BEMERKUNG 0.4. [*Ausnahme*: Fehler korrigierende Kodierung (error correcting codes)]
 Versende einen Text

ursprüngliche Textnachricht $\xrightarrow{\text{Verfälschung, Korruption}}$ empfangene Textnachricht

Simpleste Idee, um korrektes Lesen des Text zu gewährleisten
 sende Text zweimal.

In der Praxis:

ursprünglicher Text erweiterter, redundanter, kodierter Text

$$\begin{aligned} \mathbb{R}^N \ni x &\longrightarrow y = Ax \in \mathbb{R}^m, m > N \\ &\longrightarrow \tilde{y} = Ax + e \end{aligned}$$

“wegen doppelt versendeter” Information kann man hoffen, dass der ursprüngliche Text/Kode x aus der verzerrten/verrauschten/verfälschten Nachricht $\tilde{y}$ rekonstruiert werden kann.

Resultate des Compressive Sensing sind auch für die Theorie der error-correcting-codes nützlich, auch wenn wir dies in dieser Vorlesung nicht näher untersuchen.

Offensichtliches Problem: $m < N$. Falls $m < N$, wie soll ich dann aus dem Bild/den Messwerten

$$(1) \quad y = Ax \in \mathbb{R}^m$$

das ursprüngliche Signal $x \in \mathbb{R}^N$ rekonstruieren?

Lineare Algebra lehrt uns: Zu jedem Bild y ist Gleichung (1) eindeutig lösbar, falls der Rang von A mindestens N ist. Das ist hier aber unmöglich: $\text{rank}(A) \leq \min(m, N) = m < N$.

Fazit: Im Allgemeinen, ohne zusätzliche Annahme, ist folgende Dilemma nicht aufzulösen:

- ich möchte wenig Daten messen/speichern/verarbeiten, d.h. $m \ll N$,
- ich will aus der Messung/Kodierung das ursprüngliche Bild/Signal rekonstruieren können.

Einziger Ausweg ist: Wir müssen etwas über die Strukturen

- des Signals,
- oder der Kodierungsmatrix,
- oder von beiden

annehmen.

Was für eine Struktur sollen wir wählen? Man könnte sich ja viele Ausdenken.

Die gewählte Struktur sollte

- einen starken, vereinfachenden Effekt auf die Verarbeitung grosser Datenvektoren haben,
- bei praktischen Anwendungen zumindest näherungsweise erfüllt sein,

Sparsity (Dünnheit/Dünne Besetztheit) des Signals x bedeutet, dass die meisten Einträge des Vektors x Null sind. Es ist die zentrale Annahme in der Theorie des Compressive Sensing und in dieser Vorlesung. Es ist erstaunlich welche Effekte diese auf die mathematische Theorie der Datenrekonstruktion hat.

Insbesondere ist bemerkenswert:

Geeignete Annahmen an die Matrix A ermöglichen es, ein sparses Signal $x \in \mathbb{R}^N$ aus $y = Ax \in \mathbb{R}^m$ auch für $m \ll N$ zu rekonstruieren.

Dünne Lösungen unter-bestimmter Systeme

LINK: VORLESUNGSVIDEO 2 VOM 7.4.25

1. Dünne und komprimierbare Vektoren

1.1. Geometrie von B_p^N Kugeln.

Notation:

- $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$
- $\overline{N} := \{1, \dots, N\}$
- $S^c = \overline{N} \setminus S$ ist das Komplement von S im $\overline{N}$
- $x \in \mathbb{K}^N$, $N \in \mathbb{N}$, $\text{supp } x := \{j \in \overline{N} \mid x_j \neq 0\}$

DEFINITION 1.1. $x \in \mathbb{K}^N$ heißt s -sparse (s -dünn) für $s \in \mathbb{N}_0$, falls $\|x\|_0 := \#(\text{supp } x) \leq s$.

Notation: Für $x \in \mathbb{K}^N$ und $0 < p < \infty$ bezeichnen wir

$$\|x\|_p := \left(\sum_{j=1}^N |x_j|^p \right)^{1/p}$$

und für $p = \infty$

$$\|x\|_\infty := \max_{j=1}^N |x_j|.$$

- $1 \leq p \leq \infty$: $\|\cdot\|_p$ ist eine Norm, namens ℓ_p^N -Norm von x .
- $0 < p < 1$: $\|\cdot\|_p$ ist eine Quasinorm, d. h. $\exists C \geq 1$ so dass für alle $x, y \in \mathbb{K}^N$

$$\|x + y\|_p \leq C(\|x\|_p + \|y\|_p)$$

(Vergleiche Bild 1.1 und Lemma 1.10).

- $p \rightarrow 0$:

$$\lim_{p \rightarrow 0} |x_j|^p = \begin{cases} 1 & x_j \neq 0 \\ 0 & \text{sonst} \end{cases}$$

So,

$$\lim_{p \rightarrow 0} \|x\|_p^p = \lim_{p \rightarrow 0} \sum_{j=1}^N |x_j|^p = \#(\{j \mid x_j \neq 0\}) = \|x\|_0,$$

namens " ℓ_0 -Norm" von x .

- $p \rightarrow \infty$: $\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty$.

Es gilt folgende Monotonie:

LEMMA 1.2. Für alle $x \in \mathbb{K}^N$ ist die Funktion

$$(0, \infty) \ni p \mapsto \|x\|_p$$

antiton. Konkret

$$0 < q < p < \infty \implies \|x\|_q \geq \|x\|_p$$

LINK: VORLESUNGSVIDEO 3 VOM 8.4.25

BEWEIS. Für $x = 0$ gilt es offensichtlich. Sei also $x \neq 0$, damit $y_k := |x_k|/\|x\|_q$ wohldefiniert und $y_k \leq 1$ für alle $k \in \overline{N}$. Damit auch $y_k^q \geq y_k^p$. Also

$$\sum_k y_k^q \geq \sum_k y_k^p = \sum_k \left(\frac{x_k}{\|x\|_p} \right)^p = 1$$

insbesondere $\left\| \frac{x}{\|x\|_p} \right\|_q = \|y\|_q = (\sum_k y_k^q)^{1/q} \geq 1$ und damit $\|x\|_q \geq \|x\|_p$. □

BEMERKUNG 1.3 (Geometrie von ℓ_p -Einheitskugeln).

$$B_p^N := \{x \in \mathbb{K}^N \mid \|x\|_p \leq 1\}$$

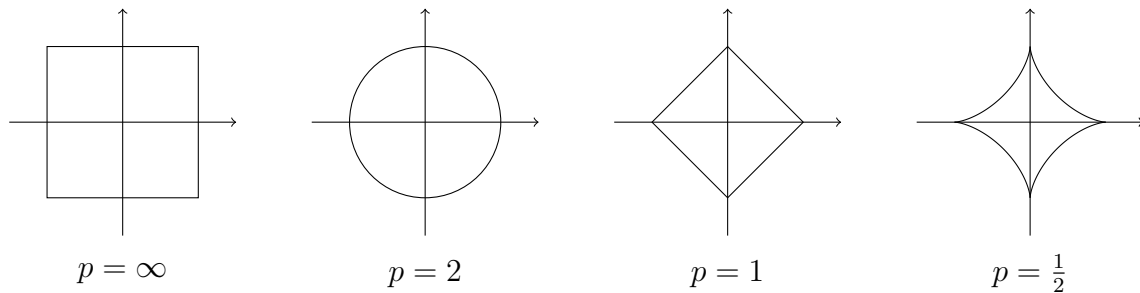


ABBILDUNG 1.1. ℓ_p -Einheitskugeln

B_p^2 ist konvex für $1 \leq p \leq \infty$, da $\|\cdot\|_p$ eine Norm ist.

B_p^2 ist nicht konvex für $0 < p < 1$.

B_p^N mit $p < 1$ ist ein Modell für komprimierbare Vektoren, siehe spaeter.

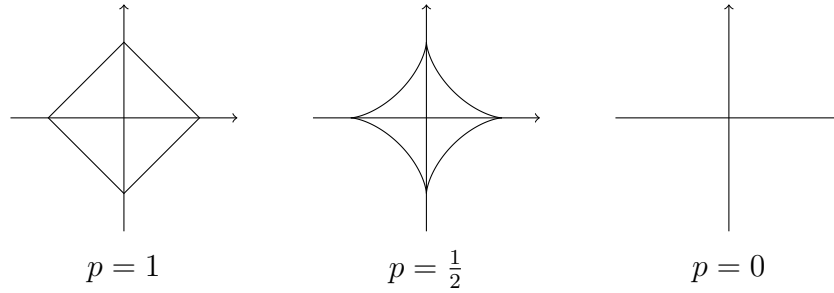
1.2. Strukturen von s -dünnen Vektoren. Sei $S \subset \overline{N}$ mit $\#(S) = s$.

$$\mathbb{K}_S^N := \{x \in \mathbb{K}^N \mid \text{supp } x \subset S\} = \{x \in \mathbb{K}^N \mid x_j = 0 \text{ für } j \notin S\}$$

ist s -dimensionaler Unterraum von $\mathbb{K}^N$ und

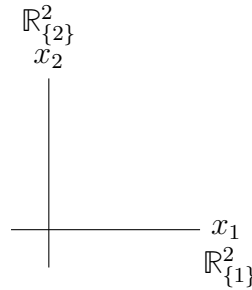
$$\mathbb{K}_s^N = \bigcup_{S \subset \overline{N}, \#(S)=s} \mathbb{K}_S^N$$

ist die Menge der s -dünnen Vektoren.

ABBILDUNG 1.2. ℓ_p -Einheitskugeln

$\mathbb{K}_s^N$ ist kein linearer Unterraum.

Warum? $x, y \in \mathbb{K}_s^N$, impliziert i.a. nicht $x + y \in \mathbb{K}_s^N$, aber schon $x + y \in \mathbb{K}_{2s}^N$, da $\text{supp}(x + y) \subset (\text{supp } x) \cup (\text{supp } y)$.



$$\mathbb{R}_1^2 = \mathbb{R}_{\{1\}}^2 \cup \mathbb{R}_{\{2\}}^2$$

Für viele Zwecke ist die Forderung s -dünn zu sein, zu strikt.

Manchmal sind viele Koeffizienten “fast Null”, aber nicht exakt. Benötigen Begriff, der dies beschreibt: komprimierbare Vektoren.

DEFINITION 1.4. (i) Sei $x \in \mathbb{K}^N$ dann existiert (mindestens) eine Permutation $\pi \in \text{Sym}(N)$ mit

$$|x_{\pi(1)}| \geq |x_{\pi(2)}| \geq \dots \geq |x_{\pi(N)}| \geq 0.$$

Fixiere nun eine solche Permutation π . Man setzt: $x_j^* = |x_{\pi(j)}|$. Dann heißt der Vektor $x^* = (x_j^*) \in \mathbb{R}_+^N = [0, \infty)^N$ *nicht-wachsende Umordnung von x* (non-increasing rearrangement). Er ist eindeutig bestimmt, auch wenn π nicht die einzige Permutation mit den o.g. Eigenschaften ist.

(ii) Sei weiter $s \in \overline{N} := \{1, \dots, N\}$ und $S = \{\pi^{-1}(1), \dots, \pi^{-1}(s)\}$, dann heißt $z \in \mathbb{K}^N$ gegeben durch

$$z_j = \begin{cases} x_j & j \in S \\ 0 & \text{sonst} \end{cases}$$

eine *beste s -Term Approximation* zu x .

BEISPIEL 1.5. Sei $x = (1, -3, 2, 4, -1, -4)$. Dann, $x^* = (4, 4, 3, 2, 1, 1)$.

Für $s = 3$, ist $z = (0, -3, 0, 4, 0, -4)$ eine beste 3-Term Approximation zu x .

Für $s = 5$, sind $z = (0, -3, 2, 4, -1, -4)$ und $z = (1, -3, 2, 4, 0, -4)$ beide eine beste 5-Term Approximation zu x .

Daraus sieht man, dass eine beste s -Term Approximation *nicht* eindeutig bestimmt sein muss.

1.3. Fehler der besten s -Term Approximation.

LINK: VORLESUNGSVIDEO 4 VOM 8.4.25

Wir definieren

$$\begin{aligned}\sigma_s(x)_p &:= \inf\{\|x - y\|_p \mid y \in \mathbb{K}_s^N\} \\ &= \|x - z\|_p\end{aligned}$$

wobei z eine beste s -Term Approximation zu x ist.

Informelle Sprachweise: x ist komprimierbar, wenn $\sigma_s(x)_p$ klein wird für wachsendes s .

LEMMA 1.6. Sei $s \in \{1, \dots, N\} = \overline{N}$, $0 < p < q$ und $x \in \mathbb{K}^N$. Dann gilt, dass

$$\sigma_s(x)_q \leq \frac{\|x\|_p}{s^{1/p-1/q}}.$$

Fakt: $\sigma_s(x)_p$ und $\|x\|_p$ hängen nur von der nicht-wachsenden Umordnung x^* ab.

BEWEIS. Wegen des Fakts, oBdA nehmen wir an, dass $x_1 \geq x_2 \geq \dots \geq x_N \geq 0$. Dann gilt

$$\begin{aligned}\sigma_s(x)_q &= \left(\sum_{j=s+1}^N x_j^q \right)^{1/q} = \left(\sum_{j=s+1}^N x_j^p x_j^{q-p} \right)^{1/q} \leq x_s^{1-p/q} \left(\sum_{j=s+1}^N x_j^p \right)^{1/q} \\ &\leq \left(\frac{s}{s} x_s^p \right)^{1/p-1/q} \|x\|_p^{p/q} \leq \left(\frac{1}{s} \right)^{1/p-1/q} \left(\underbrace{\sum_{j=1}^s x_j^p}_{\leq \|x\|_p^p} \right)^{1/p-1/q} \|x\|_p^{p/q} \\ &= \frac{1}{s^{1/p-1/q}} \|x\|_p.\end{aligned}$$

□

BEMERKUNG 1.7. Es gibt eine scharfe Version mit optimalen Konstante

$$\sigma_s(x)_q \leq c_{p,q} \frac{\|x\|_p}{s^{1/p-1/q}} \text{ mit } c_{p,q} = \left[\left(\frac{p}{q} \right)^{p/q} \left(1 - \frac{p}{q} \right)^{1-p/q} \right] \leq 1.$$

BEISPIEL 1.8. Satz 1.6 mit $p = 1$ und $q = 2$ bedeutet dass, $\sigma_s(x)_2 \leq \frac{\|x\|_1}{\sqrt{s}}$, und die optimale Variante ist $\sigma_s(x)_2 \leq \frac{1}{2} \frac{\|x\|_1}{\sqrt{s}}$.

1.4. Weitere intuitive Modelle für komprimierbare Vektoren.

LINK: VORLESUNGSVIDEO 5 VOM 14.4.25

DEFINITION 1.9. Sei $x \in \mathbb{K}^N$. Wir definieren die *schwache ℓ_p -Norm* auf $\mathbb{K}^N$ für $p > 0$ als

$$(2) \quad \|x\|_{p,+\infty} := \sup_{t>0} |\{j \in \bar{N} \mid |x_j| \geq t\}|^{1/p} \cdot t.$$

Siehe Darstellung in Bild 1.3 im Fall $p = 1$.

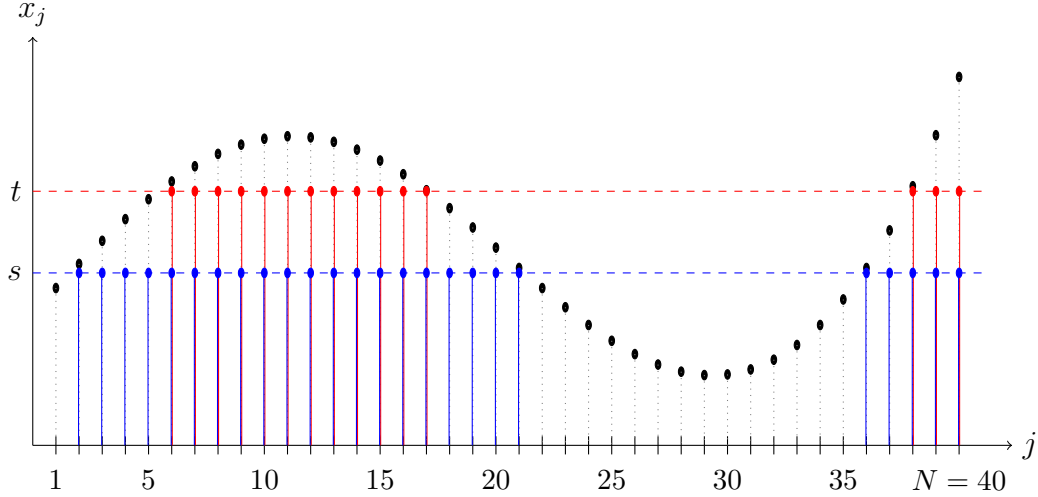


ABBILDUNG 1.3. Die schwache ℓ_1 -Norm

Alternative Darstellung: Wähle $k \in \bar{N}$ beliebig und setze $t = x_k^*$, wobei x_k^* die k -te Komponente der nicht-wachsenden Umordnung x^* von x ist. Dann gilt

$$\|x\|_{p,+\infty} \geq |\{j \in \bar{N} \mid |x_j| \geq x_k^*\}|^{1/p} \cdot x_k^* \geq k^{1/p} \cdot x_k^*,$$

da dies für jedes $k \in \bar{N}$ gilt, folgt

$$(3) \quad \|x\|_{p,+\infty} \geq \max_{1 \leq k \leq N} k^{1/p} x_k^*.$$

Andererseits: für $t > 0$ und $k := |\{j \in \bar{N} \mid |x_j| \geq t\}|$ gilt $t \leq x_k^*$ und damit

$$t |\{j \in \bar{N} \mid |x_j| \geq t\}|^{1/p} \leq x_k^* |\{j \in \bar{N} \mid |x_j| \geq t\}|^{1/p} = x_k^* k^{1/p} \leq \max_{1 \leq j \leq N} x_j^* j^{1/p}$$

Es folgt

$$(4) \quad \sup_{t>0} |\{j \in \bar{N} \mid |x_j| \geq t\}|^{1/p} \cdot t \leq \max_{1 \leq k \leq N} k^{1/p} x_k^*.$$

Insgesamt liefern Ungleichungen (3) und (4)

$$(5) \quad \|x\|_{p,+\infty} = \max_{1 \leq k \leq N} k^{1/p} x_k^*.$$

Wir bemerken dass, $\|\cdot\|_{p,+\infty}$ eine Quasinorm auf $\mathbb{K}^N$ ist, d.h. die Eigenschaften

LEMMA 1.10. Für alle $p > 0$ gilt:

$$(QN1) \quad \|x\|_{p,+\infty} = 0 \Leftrightarrow x = 0;$$

$$(QN2) \quad \|\lambda x\|_{p,+\infty} = |\lambda| \|x\|_{p,+\infty} \text{ für alle } \lambda \in \mathbb{C};$$

$$(QN3) \quad \exists C > 0, \text{ so dass für alle } x, y \in \mathbb{K}^N \text{ gilt } \|x + y\|_{p,+\infty} \leq C(\|x\|_{p,+\infty} + \|y\|_{p,+\infty})$$

erfüllt sind.

BEWEIS VON (QN3). Genauer beweisen wir die folgende Ungleichung für alle $p \in (0, \infty)$

$$(6) \quad \|x^1 + \dots + x^k\|_{p,+\infty} \leq C(\|x^1\|_{p,+\infty} + \dots + \|x^k\|_{p,+\infty}) \quad \text{mit } C = k^{\max(1, 1/p)}.$$

Sei $t > 0$. Falls fuer ein $j \in \overline{N}$ gilt, dass $|x_j^1 + \dots + x_j^k| \geq t$ gilt auch $|x_j^i| \geq t/k$ für mindestens ein $i \in \overline{k}$. Es folgt

$$(7) \quad \{j \in \overline{N} \mid |x_j^1 + \dots + x_j^k| \geq t\} \subset \bigcup_{i=1}^k \left\{ j \in \overline{N} \mid |x_j^i| \geq \frac{t}{k} \right\}.$$

Damit

$$\begin{aligned} \sup_{t>0} |\{j \in \overline{N} \mid |x_j^1 + \dots + x_j^k| \geq t\}| \cdot t^p &\leq \sup_{t>0} \left| \bigcup_{i=1}^k \left\{ j \in \overline{N} \mid |x_j^i| \geq \frac{t}{k} \right\} \right| \cdot t^p \\ &\leq \sum_{i=1}^k \sup_{t/k>0} \left| \left\{ j \in \overline{N} \mid |x_j^i| \geq \frac{t}{k} \right\} \right| \left(\frac{t}{k} \right)^p k^p \\ &\leq \sum_{i=1}^k \|x^i\|_{p,+\infty}^p \cdot k^p \end{aligned}$$

Es folgt,

$$\|x^1 + \dots + x^k\|_{p,+\infty}^p \leq k^p (\|x^1\|_{p,+\infty}^p + \dots + \|x^k\|_{p,+\infty}^p)$$

und dies impliziert

$$\|x^1 + \dots + x^k\|_{p,+\infty} \leq k(\|x^1\|_{p,+\infty}^p + \dots + \|x^k\|_{p,+\infty}^p)^{1/p}$$

Um die rechte Seite abzuschätzen, unterscheiden wir **zwei Fälle**. Falls $\mathbf{p} \geq \mathbf{1}$, folgern wir aus Lemma 1.2 $\|\xi\|_p \leq \|\xi\|_1$ für $\xi = (\|x^1\|_{p,+\infty}, \dots, \|x^k\|_{p,+\infty})$ und erhalten

$$(8) \quad \|x^1 + \dots + x^k\|_{p,+\infty} \leq k(\|x^1\|_{p,+\infty} + \dots + \|x^k\|_{p,+\infty})$$

liefert. Falls $\mathbf{p} < \mathbf{1}$, benutzen wir die Hölder-Ungleichung für $1/p > 1$, d.h.

$$\left(\sum_{i=1}^k \xi_i^p \cdot 1 \right)^{1/p} \leq \left[\left(\sum_{i=1}^k \xi_i^{p/p} \right)^p \left(\sum_{i=1}^k 1 \right)^{1-p} \right]^{1/p} = \sum_{i=1}^k \xi_i \cdot k^{\frac{1}{p}-1},$$

so folgt

$$(9) \quad \|x^1 + \dots + x^k\|_{p,+\infty} \leq k^{1/p} (\|x^1\|_{p,+\infty} + \dots + \|x^k\|_{p,+\infty}).$$

Dann liefern Ungl. (8) und (9) das Ergebnis in Ungl. (6). □

Hilfreich: $j^{-1/p} \|x\|_{p,+\infty} = (\max_k k^{1/p} \cdot j^{-1/p} x_k^*) \geq x_j^*$ für $j \in \overline{N}$.

LEMMA 1.11 (Vergleich zwischen $\|x\|_{p,+\infty}$ und $\|x\|_p$). Sei $x \in \mathbb{K}^N$ und $p > 0$. Für die ℓ_p -Norm und die schwache ℓ_p -Norm gelten

$$(1) \quad \|x\|_{p,+\infty} \leq \|x\|_p,$$

$$(2) \quad \|x\|_p \leq (1 + \log N)^{1/p} \|x\|_{p,+\infty}.$$

BEWEIS. Zu (1): Sei $x_1 \geq x_2 \geq \dots \geq x_n \geq 0$, dann gilt

$$k^{1/p} x_k = (k x_k^p)^{1/p} \leq (x_1^p + \dots + x_k^p)^{1/p} \leq \|x\|_p.$$

Zu (2): Es gilt

$$\|x\|_p^p = \sum_{k=1}^N x_k^p \frac{k}{k} \leq \sum_{k=1}^N \left(\frac{\|x\|_{p,+\infty}}{k^{1/p}} \right)^p = \|x\|_{p,+\infty}^p \sum_{k=1}^N \frac{1}{k} \leq \|x\|_{p,+\infty}^p (1 + \log N).$$

□

BEMERKUNG 1.12. Ist $x \in \mathbb{K}_s^N$ s -dünn, liefern (1) und (2) die Ungleichungskette

$$(10) \quad \|x\|_{p,+\infty} \leq \|x\|_p \leq (1 + \log s)^{1/p} \|x\|_{p,+\infty}.$$

LINK: VORLESUNGSVIDEO 6 VOM 14.4.25

SATZ 1.13. Sei $s \in \overline{N}$, $0 < p < q$, $x \in \mathbb{K}^N$, dann

$$(11) \quad \sigma_s(x)_q \leq \left(\frac{p}{q-p} \right)^{1/q} \frac{\|x\|_{p,+\infty}}{s^{1/p-1/q}}.$$

BEWEIS. Wir nutzen, dass die schwache ℓ_p -Norm $\|\cdot\|_{p,+\infty}$ erfüllt (QN2).

OBdA nehmen wir an, dass $x_1 \geq x_2 \geq \dots \geq x_N \geq 0$ und $\|x\|_{p,+\infty} = \max_{1 \leq k \leq N} k^{1/p} x_k^* = 1$. Es gilt $x_k \leq k^{-1/p}$ für $k \in \overline{N}$ und

$$\begin{aligned} \sigma_s(x)_q &= \left(\sum_{k=s+1}^N x_k^q \right)^{1/q} \leq \left(\sum_{k=s+1}^N k^{-\frac{q}{p}} \right)^{1/q} \\ &\leq \left(\int_s^\infty t^{-q/p} dt \right)^{1/q} \\ &= \left(\frac{1}{\left(\frac{q}{p} - 1 \right) s^{\frac{q}{p}-1}} \right)^{1/q} = \left(\frac{p}{q-p} \right)^{1/q} \cdot \frac{1}{s^{1/p-1/q}} \end{aligned}$$

□

LEMMA 1.14 (Kontraktionseigenschaften). Seien $x, y \in \mathbb{K}^N$ und $1 \leq s < k \leq N$. Dann gelten die folgenden:

(a) Die Umordnung ist eine $\|\cdot\|_\infty$ -Kontraktion, d.h. $\|x^* - y^*\|_\infty \leq \|x - y\|_\infty$;

(b) “Komprimierbarkeit ist stetig”, d.h. $|\sigma_s(x)_1 - \sigma_s(y)_1| \leq \|x - y\|_1$;

(c) $(k-s)x_k^* \leq \|x - y\|_1 + \sigma_s(y)_1$

BEWEIS. Zu (a): Es gilt

$$(12) \quad x_k^* = \min_{T \subset \overline{N}, |T| < k} \max_{j \in T^c} |x_j| \quad \text{und} \quad |x_j| \leq |x_j - y_j| + |y_j| \leq \|x - y\|_\infty + |y_j|.$$

Damit,

$$x_k^* \leq \|x - y\|_\infty + y_k^*$$

und es folgt dass,

$$x_k^* - y_k^* \leq \|x - y\|_\infty.$$

Nun x und y austauschen: $\Rightarrow y_k^* - x_k^* \leq \|x - y\|_\infty$, und so folgt (a).

Zu (b): Sei $z \in \mathbb{K}_s^N$ eine beste s -Term Approximation an y . Dann,

$$(13) \quad \sigma_s(x)_1 = \inf\{\|x - w\|_1 \mid w \in \mathbb{K}_s^N\} \leq \|x - z\|_1 \leq \|x - y\|_1 + \|y - z\|_1 = \|x - y\|_1 + \sigma_s(y)_1.$$

Damit

$$(14) \quad \sigma_s(x)_1 - \sigma_s(y)_1 \leq \|x - y\|_1$$

Rollentausch von x und y ergibt : $\sigma_s(y)_1 - \sigma_s(x)_1 \leq \|x - y\|_1$ und dann

$$|\sigma_s(x)_1 - \sigma_s(y)_1| \leq \|x - y\|_1.$$

Zu (c):

$$(15) \quad (k - s)x_k^* \leq \sum_{j=s+1}^k x_j^* \leq \sum_{j=s+1}^N x_j^* = \sigma_s(x)_1 \leq \|x - y\|_1 + \sigma_s(y)_1.$$

□

2. Minimal erforderliche Anzahl von Messungen zur Rekonstruktion s -dünnere Signale

LINK: VORLESUNGSVIDEO 7 VOM 15.4.25

Compressive sensing beschäftigt sich mit der Aufgabe einen Vektor x unter die Zusatzannahme $x \in \mathbb{K}_s^N$ aus m durch

$$y = Ax$$

kodierte Messungen zu rekonstruieren, mit anderen Worten die eindeutige sparse Lösung des linearen Gleichungssystems $y = Ax$, wobei A eine $(m \times N)$ -Matrix, m die Anzahl der Messungen, und $y \in \mathbb{K}^m$ der Vektor der Messwerte ist.

Nehmen wir für einen Moment Abstand von der Sparsity Bedingung $x \in \mathbb{K}_s^N$ und wollen die Rekonstruktion für sämtliche y gewährleisten, müssen wir offensichtlich mindestens $m \geq N$ Messungen durchführen.

Um wie viel lässt sich diese Mindestanzahl senken, falls wir die sparsity $x \in \mathbb{K}_s^N$ ausnutzen? D.h. wir untersuchen in diesem Abschnitt die Frage: Was ist die minimale Anzahl von Messungen m , welche eine Rekonstruktion von $x \in \mathbb{K}_s^N$ (bei geeignet gewählten A) gewährleistet?

(Während im Rest der Vorlesung die Frage nach der ‘hinreichenden Anzahl’ von Messungen im Vordergrund steht, ist es hier die ‘notwendige Anzahl’.)

Offensichtlich, wie immer, haben wir $s \leq m \leq N$.

Die obige Frage hat zwei Varianten.

- Problem U: Wie groß muss m mindestens sein, damit **eine** $(m \times N)$ -Matrix A existiert, so dass für **jedes** $x \in \mathbb{K}_s^N$ die Gleichung $y = Ax$ eindeutig lösbar ist. Wir werden sehen, dass die Antwort $m = 2s$ ist.
- Problem I: Wie groß muss m mindestens sein, damit zu **jedem** $x \in \mathbb{K}_s^N$ eine $(m \times N)$ -Matrix A (also $A = A(x)$ **abhängig von** x) existiert, so dass $y = Ax$ eindeutig lösbar ist. In diesem Fall wird die Antwort $m = s + 1$ sein.

Zur Vorbereitung überlegen wir:

SATZ 1.15. Seien $A \in \mathbb{K}^{m \times N}$, $x \in \mathbb{K}^N$ mit $\#\text{supp } x = s$, und $y = Ax \in \mathbb{K}^m$. Dann sind folgende Aussagen äquivalent.

(a) x ist die eindeutige Lösung von $Ax = y$ in $\mathbb{K}_s^N$, d.h. $\{z \in \mathbb{K}_s^N \mid Az = y\} = \{x\}$.

(b) x ist die eindeutige Lösung des Minimierungsproblems

$$(16) \quad \min_{z \in \mathbb{K}^N} \|z\|_0 \quad \text{unter die NB} \quad Az = y.$$

Dieses Problem wird mit (P_0) bezeichnet.

BEWEIS. **(b) $\Rightarrow$ (a):** Ist $Az = y$ mit $z \neq x$, dann folgt aus (b)

$$\|z\|_0 > \|x\|_0 = s$$

und damit $z \notin \mathbb{K}_s^N$. Damit ist die Eindeutigkeit gezeigt.

(a)⇒(b): Sei $z_0 \neq x$ eine beliebige Lösung von $Az = y$ (d.h. NB erfüllt). Da nach (a) x die *eindeutige* Lsg. von $Az = y$ in $\mathbb{K}_s^N$ ist, folgt $z_0 \notin \mathbb{K}_s^N$, also $\|z_0\|_0 > s = \|x\|_0$, d.h. z_0 ist keine Minimierer von (P_0) . □

Notation.

(i) Sei $A \in \mathbb{K}^{m \times N}$ und sei $S \subset \bar{N}$ eine Index-Teilmenge. Dann ist A_S die Untermatrix mit genau den Spalten, deren Index in S liegt. Sei z.B. $S = \{1, 2, N\}$ dann

$$A = \begin{pmatrix} \vdots & \vdots & \cdots & \vdots \\ 1-te & 2-te & & N-te \end{pmatrix} \in \mathbb{K}^{m \times N} \quad A_S = \begin{pmatrix} \vdots & \vdots & \vdots \\ 1-te & 2-te & N-te \end{pmatrix} \in \mathbb{K}^{m \times s}.$$

(ii) Seien $x \in \mathbb{K}^N$ und $S \subset \bar{N}$ mit $\#(S) = s$ eine Teil-Indexmenge. Wir bemerken, dass x_S zwei ähnliche aber unterschiedliche Objekten bezeichnen kann, d.h.

- $x_S \in \mathbb{K}^S$ mit $(x_S)_j = x_{k_j}$ wobei $S = \{k_1, \dots, k_s\}$ oder
- $x_S \in \mathbb{K}^N$ mit $x_S = x \cdot \chi_S$, d.h. $(x_S)_j = x_j$ falls $s_j \in S$ und 0 sonst.

BEISPIEL 1.16. Sei $(1, 2, 3, 4, 5, 6) \in \mathbb{R}^6$ und sei $S = \{2, 4, 6\}$. Dann hat man $x_S = (2, 4, 6) \in \mathbb{R}^3$ oder $x_S = (0, 2, 0, 4, 0, 6) \in \mathbb{R}^6$.

SATZ 1.17. Sei $A \in \mathbb{K}^{m \times N}$, $s \in \bar{N}$. Dann sind äquivalent:

(i) Jedes $x \in \mathbb{K}_s^N$ ist die eindeutige Lösung von $Az = Ax$ in $\mathbb{K}_s^N$, d.h.

$$x, y \in \mathbb{K}_s^N, Ax = Ay \Rightarrow x = y,$$

(ii) $\ker(A) \cap \mathbb{K}_{2s}^N = \{0\}$,

(iii) Für jedes $S \subset \bar{N}$ mit $\#S \leq 2s$ gilt: $A_S: \mathbb{K}^{\#S} \rightarrow \mathbb{K}^m$ ist eine injektiv lineare Abbildung,

(iv) Jede Menge von $2s$ Spalten von A ist linear unabhängig.

Unkonventionell: Einzelrichtungen statt Kreisschluss.

BEWEIS. (i) $\Rightarrow$ (ii): Sei $u \in \ker(A) \cap \mathbb{K}_{2s}^N$. Dann u darstellbar als $u = x - z$ mit $x, z \in \mathbb{K}_s^N$. Dann $0 = Au = A(x - z)$, d.h. $Ax = Az \stackrel{(i)}{\Rightarrow} x = z \Rightarrow u = 0$.

(ii) $\Rightarrow$ (i): Sei $Az = Ax$ mit $x, z \in \mathbb{K}_s^N$. Dann $A(z - x) = 0$. Damit ist $z - x \in \ker(A) \cap \mathbb{K}_{2s}^N \stackrel{(ii)}{\Rightarrow} z - x = 0 \Rightarrow x = z$.

(ii) $\Rightarrow$ (iii): Widerspruchs-Annahme: Existiere eine Menge $S \subset \bar{N}$, $|S| \leq 2s$, und ein $z \in \mathbb{K}^{\#S}$, $z \neq 0$, mit $A_S z = 0$. Sei $x \in \mathbb{K}_{\#S}^N$ mit $x_S = z$. Dann $x \in \mathbb{K}_{2s}^N$ und $Ax = A_S x_S = A_S z = 0$. Dann gilt $x \in \ker(A)$ und wegen (ii) $x = 0$: Widerspruch.

(iii) $\Rightarrow$ (ii): Widerspruchs-Annahme: Es existiere ein $x \in \mathbb{K}_{2s}^N$ mit $Ax = 0$ und $x \neq 0$.

Setze $S = \text{supp } x$, dann $\#S \leq 2s$. Wir haben $A_S x_S = Ax = 0 \stackrel{(iii)}{\Rightarrow} x = 0$: Widerspruch.

Äquivalenz (iii) $\Leftrightarrow$ (iv) folgt aus Standardargumenten der Lineare Algebra. □

LINK: VORLESUNGSVIDEO 8 VOM 15.4.25

FOLGERUNG 1.18. Ist A eine $(m \times N)$ -Matrix, die Problem U löst, d.h. (i) im Satz 1.17 erfüllt, dann folgt aus (iv)

$$(17) \quad m \geq \text{rank } A \geq 2s.$$

Man kann also mit weniger als $2s$ Messungen nicht erfolgreich sein. Tatsächlich reichen $2s$ bei geschickt gewähltem A auch aus:

SATZ 1.19. Sei $N \geq m = 2s$. Dann existiert eine $m \times N$ -Matrix A , so dass jedes $x \in \mathbb{K}_s^N$ aus dem Bildvektor $y = Ax \in \mathbb{K}^m$ eindeutig rekonstruierbar ist.

BEWEIS. Wähle $0 < t_1 < t_2 < \dots < t_N$ und betrachte die Matrix $A = (t_j^n)_{n=0, \dots, m-1}^{j=1, \dots, N}$. Sei $S = \{j_1, \dots, j_m\}$ mit $1 \leq j_1 < j_2 < \dots < j_m \leq N$. Dann ergibt $A_S = (t_j^n)_{n=0, \dots, m-1}^{j=j_1, \dots, j_m}$ eine transponierte $(m \times m)$ -Vandermonde-Matrix

$$A_S^T = \begin{pmatrix} 1 & t_{j_1} & t_{j_1}^2 & \dots & t_{j_1}^{m-1} \\ 1 & t_{j_2} & t_{j_2}^2 & \dots & t_{j_2}^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{j_m} & t_{j_m}^2 & \dots & t_{j_m}^{m-1} \end{pmatrix}$$

und

$$\det A_S = \prod_{1 \leq k < l \leq m} (t_{j_l} - t_{j_k}) > 0 \quad \text{Übung}$$

so dass A_S invertierbar ist. Damit gilt (iii) und somit auch (i) von Satz 1.17. Also ist Problem U lösbar. $\square$

Es gibt weitere Klassen von Matrizen, die für die Rekonstruktion in Satz 1.19 benutzt werden können:

DEFINITION 1.20. Eine Matrix $M \in \mathbb{R}^{N \times N}$ heißt *total positiv*, falls für jedes Paar $I, J \subset \overline{N}$ mit $|I| = |J|$ gilt $\det M_{I,J} > 0$, wobei $M_{I,J} = (M_{i,j})_{i \in I, j \in J}$

SATZ 1.21. Die Matrix $M = (t_j^n)_{n=0,1, \dots, n-1}^{j=1, \dots, N}$ mit $0 < t_1 < t_2 < \dots < t_N$ ist total positiv. (Übung)

BEMERKUNG 1.22. Folgende $m \times N$ -Matrizen erfüllen (iii) in Satz 1.17:

- Im Beweis von 1.19 kann man statt konsekutiver natürlicher Zahlen als Exponenten auch beliebige Menge $I \subset \mathbb{N}_0$ mit $|I| = m$ wählen. $\Rightarrow A = (t_j^n)_{n \in I}^{j=1, \dots, N}$ wobei $0 < t_1 < \dots < t_N$.
- Sei $M \in \mathbb{R}^{N \times N}$ total positiv und $I \subset \overline{N}$ mit $|I| = m = 2s$ dann erfüllt $A = M_{I, \overline{N}}$ die Bed (iv). Denn: Für jedes $S \subset \overline{N}$ mit $|S| = m$ ist $A_S = M_{I,S}$ invertierbar.
- Schon $\det(\cdot) \neq 0$ reicht für Invertierbarkeit quadr. Matrix aus $\Rightarrow$ Können auch Vandermonde Matrizen A mit $t_1, \dots, t_N \in \mathbb{C}$ (unterschiedlich) betrachten, solange $\det(A_S) \neq 0$ für alle $S \subset \overline{N}$ mit $|S| = m$ gewährleistet ist.
konkretes Bsp:
- $t_l = e^{2\pi i \frac{l-1}{N}}$, $l \in \overline{m}$

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{2\pi i \frac{1}{N}} & e^{2\pi i \frac{2}{N}} & \dots & e^{2\pi i \frac{N-1}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{2\pi i \frac{m-1}{N}} & e^{2\pi i \frac{m-1}{N} \cdot 2} & \dots & e^{2\pi i \frac{m-1}{N} (N-1)} \end{pmatrix}$$

Nennt man partielle Fouriermatrix (evtl. bis auf Skalierung)

$$\Rightarrow \det A_S = \prod_{1 \leq k < j \leq m} \underbrace{(t_{l_k} - t_{l_j})}_{\neq 0 \text{ da } k \neq j} \neq 0$$

falls $S = \{l_1, \dots, l_m\}$.

e) "Fast alle" $m \times N$ Matrizen erfüllen die Bed (iv)

Denn: Betrachte Matrizen als VR mit $\dim = mN$. Für eine feste Indexmenge $S \subset \overline{N}$, $|S| = m$ ist $\det A_S = 0$ eine polynomielle Gl., die Lebesgue-fast-überall in $\mathbb{R}^{m \times N}$ nicht erfüllbar ist. Dann ist auch endliche Vereinigung

$$\bigcup_{S \subset \overline{N}, |S|=m} \{A \in \mathbb{R}^{m \times N} \mid \det A_S = 0\}$$

Nullmenge (in $\mathbb{R}^{m \times N}$).

LINK: VORLESUNGSVIDEO 9 VOM 22.4.25

Bei den aufgeführten Klassen von Matrizen kann man mittels der Optimierungsproblems (P_0) die sparse Lösung von $y = Ax$ finden. Allerdings ist im allgemeinen die Konstruktion der Lösung NP-schwer, insbesondere zu komplex für praktischen Anwendungen. Für Fouriermatrizen liefert das unten vorgestellte „Prony-Algorithmus“, eine Lösung in polynomialer Zeit. (Wegen anderer Schwächen ist er aber für praktische Anwendungen auch nicht der beste, siehe unten.)

Nichtsdestotrotz sind Fouriermatrizen ein wichtiges Werkzeug für das Thema dieser Vorlesung.

3. Exkurs/Anhang A: Diskrete Fouriertransformation/Fouriermatrizen

Hier: Identifikation $\overline{N} \cong \mathbb{Z}_N$ zyklische Gruppe mit Verknüpfung $k \circ j = (k + j) \bmod N$
Vektor $x \in \mathbb{C}^N \cong \mathbb{C}^{\{0, \dots, N-1\}} \cong \mathbb{C}^{\mathbb{Z}_N}$ als Fkt

$$x : \mathbb{Z}_N \rightarrow \mathbb{C}, \quad x(k) = x_k$$

DEFINITION 1.23. (Diskrete Fouriertransformation) $F_N : \mathbb{C}^N \rightarrow \mathbb{C}^N$ linear, mit $F_N := (e_l(-k))_{l,k=0}^{N-1}$, $e_l(k) := e^{2\pi i \frac{kl}{N}}$. F_N wird Fouriermatrix genannt. Setze $\hat{x} := F_N x$, also

$$\hat{x}(k) = \sum_{l=0}^{N-1} e^{-2\pi i \frac{lk}{N}} x(l) = \sum_{l \in \mathbb{Z}_N} e_l(-k) x(l).$$

Es ist leicht zu zeigen, dass $\frac{1}{\sqrt{N}} F_N$ ist unitär und, insbesondere invertierbar mit

$$F_N^{-1} = \frac{1}{N} F_N^*, \quad x = \frac{1}{N} F_N^* \hat{x}$$

Notation: $\check{y} = F_N^{-1} y$.

Seien $x, y : \mathbb{Z}_N \rightarrow \mathbb{C}$. Die *Faltung* zwischen x und y ist

$$* : \mathbb{C}^N \times \mathbb{C}^N \rightarrow \mathbb{C}^N, \quad (x * y)(k) = \sum_{j \in \mathbb{Z}_N} x(j) y(k - j) = \sum_{j \in \mathbb{Z}_N} x(k - j) y(j), \quad k \in \mathbb{Z}_N.$$

Verträglichkeit mit Fouriertransformation:

$$\widehat{x * y} = \hat{x} \cdot \hat{y}$$

LINK: VORLESUNGSVIDEO 10 VOM 22.4.25

LEMMA 1.24 (Baby-Unschärferelation). Sei $y \in \mathbb{C}^N$, $|S| \leq s$, $S = \text{supp } y$ und $\hat{y}(k) = 0$ für s konsequente $k \in \mathbb{Z}_N \Rightarrow y = 0 \in \mathbb{C}^N$.

BEWEIS. Für $k = t + 1, \dots, t + s \in \mathbb{Z}_N$ gilt $0 = \hat{y}(k) = \sum_{l \in \mathbb{Z}_N} e_l(-k)y(l)$,

$$0 = \left([e^{-2\pi i \frac{l}{N}}]^k \right)_{\substack{l \in S \\ k=t+1, \dots, t+s}} \cdot y_S$$

Multipliziere Matrixgleichung mit Skalar $e^{2\pi i \frac{l}{N}(t+1)}$

$$\Rightarrow 0 = \underbrace{\left([e^{-2\pi i \frac{l}{N}}]^k \right)_{k=0, \dots, s-1}}_{\text{invertierbar}} y_S \Rightarrow y_S = 0$$

$\text{supp } y = S \Rightarrow y = 0 \in \mathbb{C}^N$. □

BEMERKUNG 1.25 (Schnelle Fourier-Transformation). Sei $E_N = \{1, e^{2\pi i \frac{1}{N}}, \dots, e^{2\pi i \frac{N-1}{N}}\}$ die N -ten Einheitswurzeln.

$$\begin{aligned} \text{additive Gruppe } (\mathbb{Z}_N, +_{\text{mod } N}) &\cong (E_N, \cdot) \\ k &\mapsto e^{2\pi i \frac{k}{N}} \end{aligned}$$

Vektorraum $V = \{f : E_N \rightarrow \mathbb{C}\}$
 Skalarprodukt $\langle f, g \rangle_V = \sum_{k \in \mathbb{Z}_N} f(k) \overline{g(k)}$
 $\|f\|_V^2 = \sum_{k \in \mathbb{Z}_N} |f(k)|^2$

LEMMA 1.26. $\langle e_l, e_m \rangle = N \delta_{l,m}$ für $l, m \in \mathbb{Z}_N$.

BEWEIS. Übung □

Insbesondere bilden die Vektoren $E_l = \frac{1}{\sqrt{N}} e_l$, ($l \in \mathbb{Z}_N$) eine ONB.

Zerlegung $V \ni f = \sum_{k \in \mathbb{Z}_N} \langle f, E_k \rangle_V E_k$,

Plancherel $\|f\|_V^2 = \sum_{k \in \mathbb{Z}_N} |\langle f, E_k \rangle_V|^2$. Der n -ter Fourierkoeffizient von f ist als

$$a_n := \hat{f}(n) = \sum_{k \in \mathbb{Z}_N} f(k) e^{-2\pi i \frac{kn}{N}} = \sqrt{N} \langle f, E_n \rangle_V$$

definiert.

SATZ 1.27. Für $f \in V$

$$f(k) = \sum_{n \in \mathbb{Z}_N} \langle f, E_n \rangle_V E_n(k) = \sum_{n \in \mathbb{Z}_N} \frac{1}{\sqrt{N}} a_n E_n(k) = \frac{1}{N} \sum_{n=0}^{N-1} a_n e_n(k)$$

Analog: $\sum_{n \in \mathbb{Z}_N} |a_n|^2 = \dots = N \sum_{k \in \mathbb{Z}_N} |f(k)|^2$.

Wie berechne ich effizient aus $F(0), \dots, F(N-1)$ die Fourierkoeffizienten $\hat{F}(0), \dots, \hat{F}(N-1)$?
 Naiver Weg: Setze

$$w_N = e^{-2\pi i/N}, \quad a_k(F) = \sum_{r \in \mathbb{Z}_N} F(r) w_N^{k-r}$$

Aufwand/Kosten der Berechnung:

- $N - 2$ Multiplikationen um $w_N^2, \dots, w_N^{N-1}$ zu erhalten

- N Multiplikationen und $N - 1$ Additionen um a_k zu berechnen
- N Koeffizienten $a_0, \dots, a_k$

$$\text{Kosten : } N \underbrace{N}_{\text{Mult.}} + N \underbrace{(N-1)}_{\text{Add.}} + \underbrace{N-2}_{\text{Pot. von } w_N} = 2N^2 - 2 < 2N^2$$

Aufwand bei naivem Vorgehen: $\mathcal{O}(N^2)$

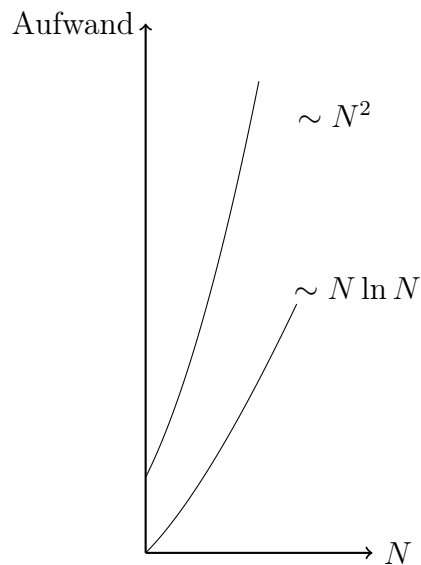
SATZ 1.28. Sei $N = 2^n$ Zweierpotenz und $w_N = e^{-2\pi i/N}$. Dann können alle Fourierkoeffizienten von f in

$$4 \cdot n \cdot 2^n = 4 \cdot \ln_2 N \cdot N = \mathcal{O}(N \log N)$$

Operationen berechnet werden.

BEWEIS. Übung.

□



Vergleich von Rechenzeiten, falls Rechner 10^6 Operationen pro Sekunde schafft.

N	1000	100000	1000000
$\mathcal{O}(N^2)$	15s	10000s $\sim$ 2,8h	10^6 s $\sim$ 278h
$\mathcal{O}(N \ln N)$	10^{-2} s	1,7s	10s

LINK: VORLESUNGSVIDEO 11 VOM 22.4.2025

SATZ 1.29. Sei F die $N \times N$ Fouriermatrix, $m = 2s \leq N$ und A bestehe aus den ersten m Zeilen von F . Dann ex. ein Algorithmus, der jedes $x \in \mathbb{K}_s^N$ aus der Messung $y = Ax \in \mathbb{K}^m$ rekonstruiert und dessen Aufwand höchstens polynomial von N abhängt.

BEWEIS. Sei $x : \mathbb{Z}_N \rightarrow \mathbb{C}$ s -sparse, d.h. $\text{supp } x \subset S \subset \mathbb{Z}_N$, $|S| = s$. $\hat{x}(0), \dots, \hat{x}(2s-1)$ sind gegeben durch

$$\hat{x}(j) = \sum_{k \in \mathbb{Z}_N} x(k) e_j(-k), \quad j \in \mathbb{Z}_N.$$

Wir definieren trigonometrische Polynom vom Grad s :

$$p(t) = \frac{1}{N} \prod_{k \in S} (1 - e^{-2\pi i \frac{k}{N}} e^{2\pi i \frac{t}{N}}) = \frac{1}{N} + \dots$$

Es ist so konstruiert, dass $p(t) = 0 \Leftrightarrow t \in S$. Wir möchten die Menge S finden durch Identifizierung von p oder $\hat{p}$. Wenn S und damit der Träger von x bekannt ist, ist Rekonstruktion fast geschafft

Da $\text{supp } x \subset S$, gilt $p(t) \cdot x(t) = 0$ für alle $t \in \overline{N}$. Dies impliziert: $\hat{p} * \hat{x} = \widehat{p \cdot x} = 0$, genauer

$$(18) \quad (\hat{p} * \hat{x})(j) = \sum_{k=0}^{N-1} \hat{p}(k) \cdot \hat{x}(j-k) = 0 \quad \forall 0 \leq j \leq N-1.$$

Da die zum Monom $e^{2\pi i \frac{kt}{N}}$ gehörigen Koeffizienten von p genau $\frac{1}{N} \hat{p}(k)$ sind und das Polynom p Grad s hat, gilt $\hat{p}(0) = 1$ und $\hat{p}(k) = 0$ für $k > s$. Wir müssen jetzt die Fourier Koeffizienten $\hat{p}(1), \dots, \hat{p}(s)$ finden.

Wir schreiben die Zeilen $s \leq j \leq 2s-1$ von (18) in der Form

$$(19) \quad \begin{cases} \hat{x}(s) + \hat{p}(1)\hat{x}(s-1) + \dots + \hat{p}(s)\hat{x}(0) = 0 \\ \hat{x}(s+1) + \hat{p}(1)\hat{x}(s) + \dots + \hat{p}(s)\hat{x}(1) = 0 \\ \vdots \\ \hat{x}(2s-1) + \hat{p}(1)\hat{x}(2s-2) + \dots + \hat{p}(s)\hat{x}(s-1) = 0 \end{cases}$$

Dies ist äquivalent zur Matrixgleichung mit Toeplitzmatrix T

$$(20) \quad \underbrace{\begin{pmatrix} \hat{x}(s-1) & \dots & \hat{x}(0) \\ \vdots & \ddots & \vdots \\ \hat{x}(2s-2) & \dots & \hat{x}(s-1) \end{pmatrix}}_{:=T} \begin{pmatrix} \hat{p}(1) \\ \vdots \\ \hat{p}(s) \end{pmatrix} = - \begin{pmatrix} \hat{x}(s) \\ \vdots \\ \hat{x}(2s-1) \end{pmatrix}.$$

Da die Fourier-Koeffizienten $\hat{x}(0), \dots, \hat{x}(2s-1)$ ja bekannt sind, wollen wir das Gleichungssystem nach $\hat{p}(1), \dots, \hat{p}(s)$ auflösen.

Daher Frage: Ist die Matrix T invertierbar ?

LINK: VORLESUNGSVIDEO 12 VOM 28.4.25

Die Matrix T ist im Allgemeinen nicht injektiv, z.B. bei

$$x = (1, 0, \dots, 0) \Rightarrow \hat{x} = (1, 1, \dots, 1) \Rightarrow \text{rank } T = 1$$

d.h. die Lösung $\hat{q}(1), \dots, \hat{q}(s)$ von (20) muss nicht notwendigerweise $\hat{p}(1), \dots, \hat{p}(s)$ sein. Allerdings existiert mindestens eine Lösung nach Konstruktion von p und T .

Wir ergänzen nun die s Bedingungen in (20) durch festsetzung der Koeffizienten $\hat{q}(0) = 1$ und $\hat{q}(k) = 0$ für $k > s$. Dann $\hat{q} \in \mathbb{K}_{\{0, \dots, s\}}^N$. Also ist $q \cdot x$ s -dünn und wegen (19) verschwindet $\widehat{q \cdot x} = \hat{q} * \hat{x}$ auf s konsekutiven Elementen von $\mathbb{Z}_N$. Wegen der Baby-Unschärferelation in Lemma 1.24 folgt $q \cdot x = 0 \in \mathbb{K}^N$.

Demnach $\text{supp } x \subset (\text{supp } q)^C$. Da $\text{supp } \hat{q} \subset \{0, \dots, s\}$, ist die Fourierreihe

$$q(t) = \frac{1}{N} \sum_{k=0}^s \hat{q}(k) e^{2\pi i \frac{kt}{N}}$$

ein trigonometrisches Polynom von Grad $\leq s$. Also

$$\#(\text{supp } q)^C = \# \text{ Nullstellen} \leq s = \#(\text{supp } x)$$

Also $S = \text{supp } x = (\text{supp } q)^C$.

Um nun die Koeffizienten von x zu bestimmen, lösen wir das System

$$\hat{x}(j) = \sum_{k \in S} x(k) e^{-2\pi i \frac{jk}{N}} \text{ für } j = 0, \dots, 2s-1.$$

Bemerke, das bereits bekannt:

- Koeffizienten der Messung $\hat{x}(j)$
- Summationsmenge S
- Koeffizienten $e^{-2\pi i \frac{jk}{N}}$

Wir haben $2s$ Gleichungen, die s Unbekannte bestimmen, da wegen der Struktur der Fouriermatrix das System vollen Rang hat. $\square$

Der vorgestellte Algorithmus wird *Prony methode* bezeichnet.

BEMERKUNG 1.30 (Rechenaufwand/Komplexität).

- Löse das $(s \times s)$ -Gleichungssystem (20) um $\hat{q}$ zu erhalten: $O(s^3)$.
(Die Anzahl arithmetischer Operationen für die LR-Zerlegung ist bei einer $s \times s$ -Matrix ca. $\frac{2}{3}s^3$.)
- Berechne q und $\hat{q}$ mittels FFT (Fast Fourier Transform/Schnelle Fouriertransformation): $O(N \log N)$. Damit ist S bekannt.
- Löse das überbestimmte $(s \times 2s)$ -Gleichungssystem um x zu erhalten: $O(s^3)$.

Insgesamt brauchen wir $O(s^3 + N \log N)$ Operationen.

BEMERKUNG 1.31 (Schwäche der vorgestellten Prony Methode).

- Nur für sparse, nicht kompressible Signalen anwendbar.
- Nicht stabil bei fehlerhaften Messungen:

$$\tilde{y} = y + e = Ax + e, \quad e \text{ Fehler mit } \|e\|_2 \text{ klein}$$

Ausblick: Um diese Schwäche zu überwinden, muss man einen Algorithmus mit mindesten $m = O(s \log \frac{N}{s})$ Anzahl von Messungen anwenden $\Rightarrow$ (schwache) N -Abhängigkeit.

[Ende des Exkurses]

LINK: VORLESUNGSVIDEO 13 VOM 28.4.25

Nun zu Problem I.

SATZ 1.32 (Rekonstruktionsmatrix A für vorgegeben sparses x). Sei $N \geq s+1$ und $x \in \mathbb{K}_s^N$. Dann ist die Menge

$W := \{A \in \mathbb{K}^{(s+1) \times N} \mid x \text{ kann nicht als eindeutige Lösung von } (P_0) \text{ (16) rekonstruiert werden}\}$
eine Lebesgue-Nullmenge in $\mathbb{K}^{(s+1) \times N}$.

BEWEIS. Sei A eine solche Matrix, dann existiert ein $z \in \mathbb{K}_s^N$, $z \neq x$ so dass $Ax = Az$. Jetzt, wählen wir $T \subset \bar{N}$, $\text{supp } z \subset T$, $\#T = s$. Wir betrachten zwei Fälle.

Fall 1: $\text{supp } x \subset T$. Es gilt $A(x - z) = 0$, damit auch $(A(x - z))_{\bar{s}} = 0$. Dann ist die quadratische Matrix $A_{\bar{s},T}$ nicht invertierbar.

Setzen nun

$$f_T : \mathbb{K}^{(s+1) \times N} \rightarrow \mathbb{K}, \quad f_T(X) = \det X_{\bar{s},T}$$

Es gilt dann $f_T(A) = 0$.

Fall 2: $\text{supp } x \not\subset T$. Dann betrachten wir den Vektorraum $V := \{u \in \mathbb{K}^N \mid \text{supp } u \subset T\} + \mathbb{K}x$, $\Rightarrow \dim V = s + 1$, und die lineare Abbildung $G : V \rightarrow \mathbb{K}^{s+1}$, $v \mapsto Av$, ist singulär, da $0 \neq z - x \in \ker(G)$. In Basisform bezüglich der Basis $\{e_j\}_{j \in T} + \{x\}$, sieht G wie folgt aus:

$$B_{x,T} := \begin{pmatrix} a_{1,j_1} & \cdots & a_{1,j_s} & \sum_{k \in S} x_k a_{1,k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{s+1,j_1} & \cdots & a_{s+1,j_s} & \sum_{k \in S} x_k a_{s+1,k} \end{pmatrix},$$

wobei $T = \{j_1, \dots, j_s\}$ und $S = \text{supp } x$. Also für $g_T : \mathbb{K}^{(s+1) \times N} \rightarrow \mathbb{K}$ gegeben durch $g_T(A) = \det B_{x,T}$, gilt also $g_T(B_{x,T}) = 0$, wenn man eine Matrix A mit oben geforderten Eigenschaften einsetzt.

Demnach

$$W \subset \underbrace{\bigcup_{\#T=s} f_T^{-1}(\{0\})}_{=(1)} \cup \underbrace{\bigcup_{\#T=s} g_T^{-1}(\{0\})}_{=(2)}$$

Nun, sind sowohl (1) als auch (2) Lebesgue-Nullmengen, da sie Vereinigung von endlichen vielen Nullstellengebilden von Polynomen (die nicht identisch gleich Null sind) Nullmengen ist. Damit ist W selbst eine Lebesgue-Nullmenge. $\square$

4. ℓ_0 -Minimierung ist NP-schwer

LINK: VORLESUNGSVIDEO 14 VOM 28.4.25

Sei $x \in \mathbb{K}_s^N$ mit $Ax = y$. Das (P_0) -Problem ist

$$\min_{z \in \mathbb{K}^N} \|z\|_0 \quad \text{mit NB} \quad y = Az.$$

Naiver intuitiver Ansatz: Zu vorgegebenen (möglichst kleinem Träger) suche Vektor, der das Gleichungssystem löst, also: Zu jedem $t \in \{0, 1, \dots, N\}$ und jedem $T \subset \bar{N}$, $\#T = t$ suche die Lösung $A_T z = y$ (beziehungsweise statt dessen für jedes quadratische System $A_T^* A_T z = A_T^* y$).

Falls schon bekannt ist, dass die gesuchte Lösung $\# \text{supp } x = s$ erfüllt, reduziert sich die Suche auf den Fall $t = s$. Selbst dann sind $\binom{N}{s}$ lineare Gleichungssysteme zu lösen, (z.B. falls $N = 1000$ und $s = 10$, hat man $\binom{N}{s} \geq \left(\frac{1000}{10}\right)^{10} = 10^{20}$.)

Dieser Ansatz für (P_0) ist zu teuer/lang. Tatsächlich ist nicht nur dieser zu teuer, sondern jeder mögliche.

4.1. Rahmen der Komplexitätstheorie.

DEFINITION 1.33. Ein Entscheidungsproblem ist eine Frage mit Ja/Nein Antwort, formal eine Funktion auf einer (abzählbaren) Menge Ω

$$f: \Omega \rightarrow \{0, 1\}$$

Betrachten wir nun den Spezialfall, dass Ω aus endlichen Worten über einem endlichen Alphabet $\mathcal{A}$ besteht, d.h.

$$\Omega = \bigcup_{n \in \mathbb{N}} \{a_1 a_2 \dots a_n \mid a_j \in \mathcal{A} \text{ für alle } j \in \bar{n}\}$$

In diesem Kontext ist ein polynomial-Zeit Algorithmus ein Algorithmus, der eine Aufgabe (z.B. die Berechnung des Funktionswerts $f(a_1 a_2 \dots a_n) \in \{0, 1\}$) in einer Anzahl von Schritten durchführt, die durch eine polynomiale Funktion in der Wortlänge n der Eingabe $a_1 a_2 \dots a_n$ beschränkt ist. Die Zahl n wird Eingabelänge genannt.

Man sagt, dass ein Entscheidungsproblem A auf ein Entscheidungsproblem B reduziert werden kann **genau dann, wenn** die Frage A (mit Hilfe eines polynomial-Zeit Algorithmus') zu Frage B umformuliert werden kann **und** die Antwort zu Frage B (mit Hilfe eines polynomial-Zeit Algorithmus') zu einer Antwort auf A übersetzt werden kann.

$$\Omega_A \xrightarrow{\text{Algorithmus}} \Omega_B \xrightarrow{f_B} \{0, 1\} \xrightarrow{\text{Algorithmus}} \{0, 1\}$$

Notation: In diesem Fall schreiben wir $A \preceq_P B$ und sprechen " B ist mindestens so schwer wie A ".

Bei vielen Entscheidungsproblemen ist die gestellte Frage, ob ein Objekt mit bestimmten Eigenschaften existiert, z.B. eine Lösung eines Gleichungssystems oder ein Punkt an dem ein Funktional sein Infimum annimmt, oder aber ob solch ein Objekt eindeutig ist. Wird nach der Existenz eines Objektes gefragt, so sind zwei Arten der Antwort 'Ja' denkbar: Eine Methode, das gesuchte Objekt zu konstruieren, oder eine Methode, um für ein 'geratenes' Objekt zu prüfen, ob es die gewünschten Eigenschaften besitzt.

Die Klassen der Komplexitätstheorie sind:

- **P:** Entscheidungsprobleme, für die es einen polynomial-Zeit Algorithmus gibt, der eine Lösung findet.
- **NP:** Entscheidungsprobleme, für die es einen polynomial-Zeit Algorithmus gibt, der eine potenzielle Lösung verifiziert.
- **NP-schwer:** Entscheidungsprobleme B , so dass $A \preceq_P B$ für jedes NP-Problem A gilt.
- **NP-vollständig:** $\text{NP} \cap \text{NP-schwer}$.

Letzte beiden Eigenschaften machen Sinn für allgemeine Probleme, wie z. B. unser (P_0) , da sie ja über Transformation von Algorithmus definiert sind. Wir wollen zeigen dass (P_0) NP-schwer ist. Dazu, finden wir NP-schweres Problem B , so dass $B \preceq_P (P_0)$. Dann,

$$\forall \text{ NP-Problem } A: \quad A \preceq_P B \preceq_P (P_0) \implies A \preceq_P (P_0)$$

(Das ist die Transitivität der Relation $\preceq_P$.)

BEISPIEL 1.34 (3-Cover Problem). Gegeben das System $\{\mathcal{C}_1, \dots, \mathcal{C}_N\}$ von Teilmengen $\mathcal{C}_i \subset \overline{m}$ mit $\#\mathcal{C}_i = 3$ für jedes $i \in \overline{N}$, formuliert man die Ja/Nein-Frage: Gibt es eine Indexmenge $J \subset \overline{N}$, so dass

$$\bigcup_{j \in J} \mathcal{C}_j = \overline{m} \quad \text{und} \quad (\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset \implies i = j) \quad ?$$

Diese Problem ist NP-vollständig (ohne Beweis).

Offensichtlich kann man eine Lösung nur dann finden, wenn $m = 3\#J \in 3 \cdot \mathbb{N}$.

LINK: VORLESUNGSVIDEON 15 VOM 29.4.25

SATZ 1.35. Das (P_0) Minimierungsproblem ist für allgemeines $A \in \mathbb{K}^{m \times N}$ und $y \in \mathbb{K}^m$ NP-schwer.

BEWEISIDEE: Suche Transformation, die das 3-Cover Problem in Polynomialzeit auf (P_0) reduziert und die Antwort in Polynomialzeit zurücktransformiert.

Sei $\{\mathcal{C}_1, \dots, \mathcal{C}_N\}$ wie oben. Baue Matrix $A = (a_1 | a_2 | \dots | a_N)$ durch

$$(a_i)_j = \begin{cases} 1, & j \in \mathcal{C}_i \\ 0, & j \notin \mathcal{C}_i \end{cases}$$

Also ist $a_i = \mathbb{1}_{\mathcal{C}_i}$ und A ist Inzidenzmatrix. Setze $y = (1, \dots, 1)^\top \in \mathbb{K}^m$. Konstruktion kann in Polynomialzeit erfolgen wegen $N \leq \binom{m}{3} = \frac{m(m-1)(m-2)}{6}$.

Es gilt: $|\mathcal{C}_j| = 3 \implies \|a_j\|_0 = 3$ für $j \in \overline{N}$.

$\implies$ Jeder Vektor $z \in \mathbb{K}^N$ mit $y = Az \in \mathbb{K}^m$ erfüllt

$$(21) \quad m = \|y\|_0 = \|Az\|_0 = \left\| \sum_{j \in \overline{N}} z_j a_j \right\|_0 \leq 3\|z\|_0.$$

Daher erfüllt jedes NB -zulässige $z \in \mathbb{K}^N$: $\|z\|_0 \geq \frac{m}{3}$.

Intuitiv klar, denn so groß muss die Indexmenge J sein, damit $\overline{m}$ von $\{\mathcal{C}_j | j \in \text{supp } z\}$ überdeckt wird.

Behauptung: $\|x\|_0 = \frac{m}{3} \iff J := \text{supp } x$ löst 3-cover Problem.

Richtung $\implies$:

Vorüberlegung:

$$x \text{ löst } Ax = y = (1, \dots, 1)^\top \in \mathbb{K}^m \xLeftrightarrow{\text{supp } x = S} A_S x_S = y$$

A_S hat $m/3 = \|x\|_0$ Spalten und $m = \dim y$ Zeilen, jede Spalte hat genau $3 = \#\mathcal{C}_j$ Einsen. Also hat A_S genau m Einsen (andere Koeffizienten = 0).

Zuerst zeigen wir Disjunktheit.

Widerspruchsannahme: Seine die $\mathcal{C}_i$ nicht disjunkt. Dann existieren $i, j \in \text{supp } x =: S \subset \overline{N}$ mit $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$. $\implies$ Es existiert Zeile $k \in \overline{m}$ mit $a_{ik} = a_{jk}$, insbesondere existiert eine Zeile (nämlich k te) von A_S mit 2 Einsen

$\implies$ Es existiert Zeile l von A_S mit lauter Nullen.

$$\implies 1 = y_l = \sum_{j \in S} a_{lj} x_j \stackrel{\text{alle } a_{lj}=0}{=} 0$$

$\implies$ Widerspruch.

Jetzt zeigen wir Überdeckungseigenschaft:

Von oben wissen wir: A_S hat genau m Einsen, in jeder Zeile genau eine. Wegen $S = \text{supp } x$ gilt $x_S \equiv 1$ und es folgt

$$A_S x_S = \sum_{j \in S} a_j = (1, \dots, 1)^T = \mathbf{1}_{\bar{m}}$$

Richtung $\Leftarrow$:

Ist $(\mathcal{C}_j : j \in J)$ 3-cover von $\bar{m}$, so setze $z := \mathbf{1}_J \in \mathbb{R}^N$. Dann folgt $Az = y = (1, 1, \dots, 1)^T$ sowie $\|z\|_0 = m/3$. $\square$

Man kann dieses Argument noch erweitern.

SATZ 1.36. Für jedes $\eta > 0$ ist das Minimierungsproblem

Gegeben $A \in \mathbb{K}^{m \times N}, y \in \mathbb{K}^m$ finde

$$\min_{z \in \mathbb{K}^N} \|z\|_0 \quad \text{mit} \quad \|Az - y\|_2 \leq \eta \quad (P_{0,\eta})$$

NP-schwer.

BEWEIS. Reskalierung erlaubt anzunehmen: $\eta < 1$. Wählen $A \in \{0, 1\}^{m \times N}$ und $y = (1, \dots, 1)$ wie oben. Erfüllt $z \in \mathbb{K}^N$ die NB, so gilt für $k \in \bar{m}$:

$$1 > \eta \geq \|(Az)_k - y_k\|_2 \geq |(Az)_k - y_k| = |(Az)_k - 1| \implies (Az)_k > 0 \implies k \in \text{supp } Az$$

Also $\|Az\|_0 = m$ und damit wie in (22)

$$(22) \quad m = \|Az\|_0 = \left\| \sum_{j \in \bar{N}} z_j a_j \right\|_0 \leq 3\|z\|_0.$$

$$\text{d.h. } \|z\|_0 \geq \frac{m}{3}.$$

Sei nun x ein Minimierer von $(P_{0,\eta})$.

Falls $\|x\|_0 = \frac{m}{3}$, dann ist $\{\mathcal{C}_j \mid j \in \text{supp } x\}$ Zerlegung von $\bar{m}$ vgl. Richtung $\implies$ im vorherigem Beweis (in polynomialer Zeit berechenbar).

Falls $\|x\|_0 > \frac{m}{3}$, finden wir keine disjunkte Zerlegung von $\bar{m}$ in $\{\mathcal{C}_j\}_{j \in \bar{N}}$, exakt wie oben.

In der Tat: Nehmen wir an dass, eine disjunkte Zerlegung mit Indexmenge J existiert. Setze $z = \mathbf{1}_J \in \mathbb{K}^N \implies z_1 a_1 + z_2 a_2 + \dots + z_N a_N = (1, \dots, 1)^T$, also $Az = y$. Und analog zu Richtung $\Leftarrow$ oben folgt $\|z\|_0 = \frac{m}{3} \implies$ Widerspruch zu x Minimierer. $\square$

KAPITEL 2

ℓ_1 -Minimierung

LINK: VORLESUNGSVIDEO 16 VOM 29.4.25

Sind mit (P_0) gescheitert, da NP-schwer. Suchen nach alternativen/modifizierten Zugängen.

1. Optimierungsprobleme

Optimierungsprobleme haben die Form

$$\textcircled{\text{P}} \quad \min_{x \in \mathbb{R}^N} F_0(x) \quad \text{unter NB} \quad F_i(x) \leq b_i, i \in \bar{n},$$

wobei $b_1, \dots, b_n \in \mathbb{R}$ Konstanten, $F : \mathbb{R}^N \rightarrow \mathbb{R}$ Zielfunktion, $F_1, \dots, F_n : \mathbb{R}^N \rightarrow \mathbb{R}$ Nebenbedingungen. Für die Lösung von $\textcircled{\text{P}}$ schreibt man $x^\# := \arg \min_{x \in \mathbb{R}^N} F_0(x)$ unter NB $F_i \leq b_i, i \in \bar{n}$.

- Kann man auch für $F_0, \dots, F_n : \mathbb{C}^N \rightarrow \mathbb{R}$ formulieren. Hier ist dennoch wichtig $\mathbb{C}/\mathbb{R}$ zu unterscheiden.
- Falls in einer NB Gleichheit auftaucht, z.B. $F_i(x) = b_i$, kann sie durch zwei $\leq$ -Bedingungen implementiert werden: $F_i(x) \leq b_i$ und $-F_i(x) \leq -b_i$.
- Sind $F_0, \dots, F_n$ konvexe Funktionen, so heißt $\textcircled{\text{P}}$ *konvexes Problem*.
- Sind $F_0, \dots, F_n$ lineare Funktionen, so heißt $\textcircled{\text{P}}$ *lineares Problem*, welches auf Simplexalgorithmus führt.

$\{\text{lineare Probleme}\} \subset \{\text{konvexe Probleme}\} \leftarrow$ gut lösbar, enthält nicht (P_0) , da NP-schwer.

BEISPIEL 2.1. Die Funktion $F : \mathbb{R}^N \rightarrow \mathbb{R} : F(x) = \|x\|_0$ ist nicht konvex:

$$\text{supp } x \neq \text{supp } y, F(x) = 1 = F(y) \implies F(x+y) = 2 = F\left(\frac{x+y}{2}\right) \not\leq 1 = \frac{F(x) + F(y)}{2}$$

Erinnerung: Für $q \rightarrow 0$ gilt $\|x\|_q^q \rightarrow \|x\|_0$. Vielleicht kann man (P_0) durch ein

$$(P_q) \quad \min_{z \in \mathbb{R}^N} \|z\|_q \quad \text{unter NB} \quad Az = y$$

ersetzen?

Für $q > 1$:

- ist das Funktional $F(x) = \|x\|_q$ konvex $\textcircled{+}$
- Minimierer von (P_q) sind nicht dünn $\textcircled{-}$

$\implies (P_q)$ ist ein konvexes Optimierungsproblem, aber das falsche.

Für $0 < q < 1$:

- Minimierungsproblem (P_q) bevorzugt sparse Vektoren
- P_q nicht konvex
- NP-schwer

Für $q = 1$:

- (P_1) ist konvex, $F(x) = \sum_{i=1}^N |x_i|$, insbesondere (P_1) nicht NP-schwer.
- Minimierer sind sparse (unter geeigneten Bedingungen)

SATZ 2.2 (ℓ_1 -Minimierer sind dünn). Sei $y \in \mathbb{R}^m$ und $A = (a_1, \dots, a_N) \in \mathbb{R}^{m \times N}$. Ist $x \in \mathbb{R}^N$ einziger Minimierer von

$$(P_1) \quad \min_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{unter } NB \quad Az = y,$$

so sind die Spalten von $A_S \sim \{a_j\}_{j \in S}$ mit $S = \text{supp } x$ linear unabhängig. Insbesondere gilt $\|x\|_0 = |S| \leq m$, da es höchstens m linear unabhängige Spalten in A geben kann.

BEWEIS. Sind die Spalten A_S nicht linear unabhängig, so existiert ein $v \in \mathbb{R}^N$, $v \neq 0$, $\text{supp } v \subset S$, mit

$$Av = \sum_{j=1}^N v_j a_j = \sum_{j \in S} v_j a_j = 0$$

Für alle $t \neq 0$ gilt: $A(x + tv) = Ax + tAv = y$ erfüllt die NB. Sei jetzt $t \neq 0$ so dass $|t| < \min_{j \in S} \frac{|x_j|}{\|v\|_\infty}$, dann

$$\begin{aligned} \|x\|_1 &\stackrel{\text{Eindeutigkeit}}{<} \|x + tv\|_1 = \sum_{j \in S} |x_j + tv_j| \\ &= \sum_{j \in S} \text{sgn}(x_j + tv_j)(x_j + tv_j) \stackrel{\text{Wahl von } t}{\leq} \sum_{j \in S} \text{sgn}(x_j)(x_j + tv_j) \\ &= \sum_{j \in S} |x_j| + t \sum_{j \in S} \text{sgn}(x_j)v_j. \end{aligned}$$

Das liefert einen Widerspruch, da man immer $t \neq 0$ wählen kann, so dass $t \sum_{j \in S} \text{sgn}(x_j)v_j \leq 0$. □

LINK: VORLESUNGSVIDEO 17 VOM 5.5.25

Was passiert, falls

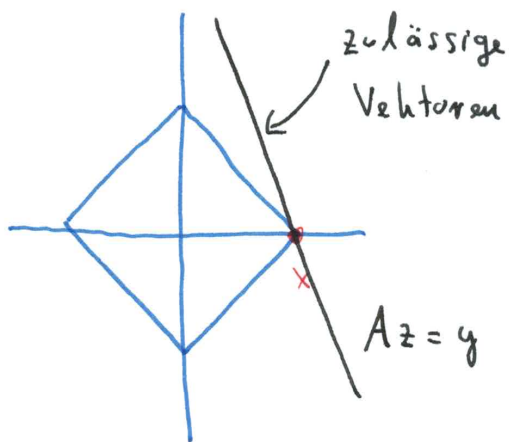
$$\sum_{j \in S} \text{sgn}(x_j) v_j = 0 ?$$

Auch in diesem Fall liefert die Argumentation vom letzten Beweis einen Widerspruch, denn wegen der angenommenen Eindeutigkeit des Minimums durften wir die Ungleichungskette mit der strikten Ungleichheit

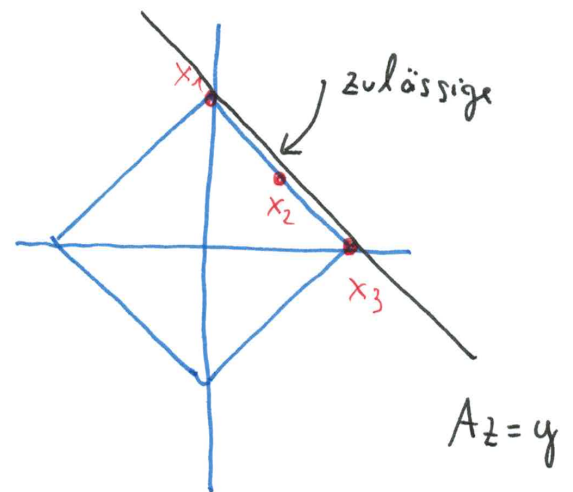
$$\|x\|_1 < \|x + \underbrace{tv}_{\neq 0}\|_1$$

beginnen. In der Tat, ohne diese Annahme ist Satz 2.2 falsch. Sie schliesst eine geometrische Sondersituation aus:

generisch



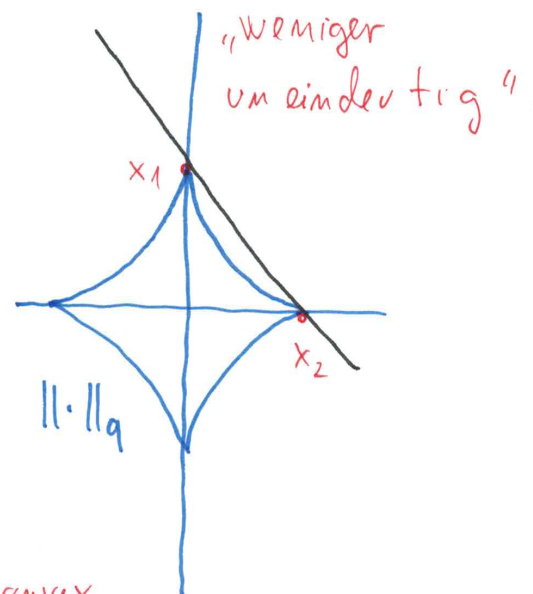
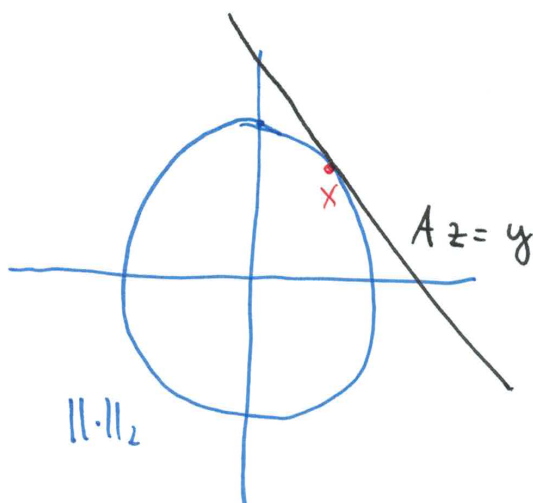
Sondersituation



$$x = \arg \min \|z\|_1 \quad \text{unter NB} \quad Az = y$$

 $\{z | Az = y\}$ Seitenfläche
 $\{z | Az = y\}$ Seitenf.

Dieser Sonderfall ist typisch für $\|\cdot\|_1$ Norm. Vgl. mit $\|\cdot\|_2$ und $\|\cdot\|_q$ -Norm mit $q \in (0, 1)$



x eindeutig. da $\|\cdot\|_2$ strikt konvex

Wir benutzen Eindeutigkeit von x im Beweis. In der Tat, ohne diese Annahme ist der Satz falsch.

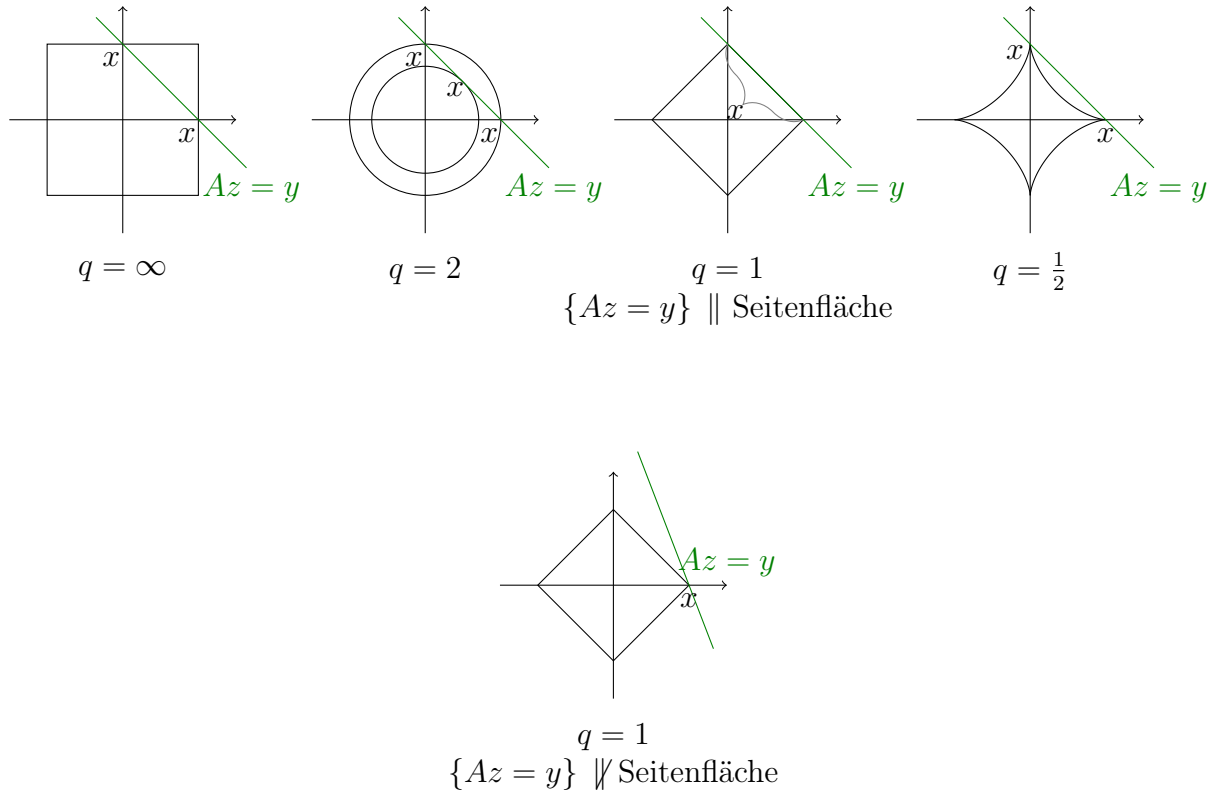


ABBILDUNG 1.1. ℓ_q -Einheitsballe: Vektoren $\{Az = y\}$, Minimierer $x = \arg \min \|z\|_q$ NB $Az = y$.

Minimierungsproblem (P_1) und zugehorige Algorithmen werden ℓ_1 -Minimierung oder *Basissjagd* (basic pursuit) genannt. Das konvexe Problem (P_1) kann man im reellen Fall durch Einfuhrung von ‘‘Schlupfvariablen’’ in ein lineares uberfuhren und Simplexalgorithmus anwenden.

Dazu: Setze $z^+, z^- \in \mathbb{R}_+^N$ mit

$$z_j^+ = \begin{cases} 0 & , z_j < 0 \\ z_j & , \text{sonst} \end{cases} \quad \text{und} \quad z_j^- = \begin{cases} 0 & , z_j > 0 \\ -z_j & , \text{sonst} \end{cases}$$

$$\implies z = z^+ - z^-, |z_j| = z_j^+ + z_j^-.$$

Problem (P_1) ist aquivalent zu $(\tilde{P}_1)$:

$$(\tilde{P}_1) \quad \min_{z^+, z^- \in \mathbb{R}_+^N} \sum_{j=1}^N (z_j^+ + z_j^-) \quad \text{unter NB} \quad (A| - A) \begin{pmatrix} z^+ \\ z^- \end{pmatrix} = y; \begin{pmatrix} z^+ \\ z^- \end{pmatrix} \geq 0$$

$\Rightarrow$ sei $\xi = \begin{pmatrix} \xi^+ \\ \xi^- \end{pmatrix} \in \mathbb{R}^{2N}$ Minimierer von $(\tilde{P}_1)$

$\Rightarrow x = \xi^+ - \xi^-$ ist Minimierer von (P_1) .

[Bisher nur $\mathbb{R}$ -Fall. Für $\mathbb{C}$ -Fall wählen wir andere Transformation, die auch fehlerbehaftete Messungen $(P_{1,\eta})$ abdeckt.]

1.1. Basisjagd unter quadratischen Bedingungen

LINK: VORLESUNGSVIDEO 18 VOM 5.5.25

Satz 2.2 und Reduktion auf ein lineares Problem funktioniert nur für $\mathbb{K} = \mathbb{R}$.

Nun: $\mathbb{K} = \mathbb{C}$ und lassen (quantifiziert) fehlerbehaftete Messungen zu.

Messmatrix $A \in \mathbb{K}^{m \times N}$ und $y \in \mathbb{K}^m$ gegeben. Vorgegeben ist auch ein noise level $\eta \geq 0$ (maximal zulässige Verzerrung der Messdaten). Führt auf Optimierungsproblem:

$$x^\# := \arg \min_{z \in \mathbb{K}^N} \|z\|_1 \quad \text{mit NB} \quad \|Az - y\|_2 \leq \eta \quad (P_{1,\eta})$$

genannt *Basisjagd mit quadratischen Nebenbedingungen*. Ist konvexes Optimierungsproblem.

Zunächst, falls $\mathbb{K} = \mathbb{C}$ benutzen wir die

Transformation in reellen Variablen. : Seien $z \in \mathbb{C}^N$ und $u, v \in \mathbb{R}^N$ mit $z = u + iv$. Dann

$$\|z\|_1 = \sum_{j=1}^N |z_j| = \sum_{j=1}^N \sqrt{u_j^2 + v_j^2} \quad \text{und} \quad Az = (\Re(A) + i\Im(A))(u + iv).$$

Seien weiter $c_1, \dots, c_N$ zusätzliche reelle Variablen mit $c_j \geq |z_j| = \sqrt{u_j^2 + v_j^2}$. Schreiben dann $(P_{1,\eta})$ in neuen Koordinaten (äquivalent) hin:

$$\min_{u,v,c \in \mathbb{R}^N} \sum_{j=1}^N c_j \quad \text{unter NB} \quad \sqrt{u_j^2 + v_j^2} \leq c_j, \quad j = 1, \dots, N \quad (P'_{1,\eta})$$

und NB

$$\left\| \begin{pmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} \Re(y) \\ \Im(y) \end{pmatrix} \right\|_2 \leq \eta. \quad (P'_{1,\eta})$$

Ist ein “Kegelloptimierungsproblem 2. Ordnung” (second order cone program) (“2. Ordnung” wegen quadratischer NB), denn die Menge

$$K := \left\{ (c, u, v) \in \mathbb{R}^{3N} : \sqrt{u_j^2 + v_j^2} \leq c_j, \quad (j = 1, \dots, N) \right\}$$

ist ein konvexer Kegel im $\mathbb{R}^{3N}$. Für solche Optimierungsprobleme gibt es effiziente Algorithmen implementiert in Standardsoftware.

LINK: VORLESUNGSVIDEO 19 VOM 5.5.25

$(P_{1,\eta})$ ist in enger Beziehungen zu den folgenden beiden:

(LASSO): gegeben $\tau > 0$, suche

$$(L_\tau) \quad \min_{z \in \mathbb{C}^N} \|Az - y\|_2 \quad \text{unter NB} \quad \|z\|_1 \leq \tau$$

Dies ist ein Beispiel von Dualität: Vertauschen von NB und Zielfunktion.

Sowie Tikhonov Regularisierung: Für $\lambda > 0$, suche

$$(T_\lambda) \quad \min_{z \in \mathbb{C}^N} \lambda \|z\|_1 + \|Az - y\|_2^2$$

SATZ 2.3. Sei $A \in \mathbb{K}^{m \times N}$, $y \in \mathbb{K}^m$. Dann:

- a) Ist x eindeutiger Minimierer von $(P_{1,\eta})$, dann gibt es ein $\tau = \tau_x$, sodass x eindeutiger Minimierer von (L_τ) ist.
- b) Ist x Minimierer von (L_τ) , so gibt es ein $\lambda = \lambda_x$, sodass x Minimierer von (T_λ) ist.
- c) Ist x (eindeutiger) Minimierer von (T_λ) , so existiert ein $\eta = \eta_x$, sodass x (eindeutiger) Minimierer von $(P'_{1,\eta})$ ist.

BEWEIS. Zu (a): x sei eind. Minimierer von $\|z\|_1$ unter NB $\|Az - y\|_2 \leq \eta$. Setze $\tau = \tau_x = \|x\|_1$. Sei $z \in \mathbb{K}^N$, $z \neq x$ mit $\|z\|_1 \leq \tau$, d.h. zulässig bzgl. (L_τ) . Da x eindeutig, erfüllt z nicht die NB des Problems $(P_{1,\eta})$. $\Rightarrow \|Az - y\|_2 > \eta \geq \|Ax - y\|_2$, d.h. x eindeutiger Minimierer von (L) .

Zu (b): Erfordert etwas mehr konvexe Analysis.

Zu (c): x sei Minimierer von $\lambda \|z\|_1 + \|Az - y\|_2^2$. Setze $\eta = \|Ax - y\|_2^2 = \eta_x$. Wähle $z \in \mathbb{K}^N$, $z \neq x$, $\|Az - y\|_2 \leq \eta$, d.h. zulässig bzgl. $(P_{1,\eta})$.

Da x Minimierer, folgt

$$\lambda \|x\|_1 + \cancel{\|Ax - y\|_2^2} \leq \lambda \|z\|_1 + \underbrace{\|Az - y\|_2^2}_{\leq \eta^2} \leq \lambda \|z\|_1 + \cancel{\|Ax - y\|_2^2}$$

$$\Rightarrow \|x\|_1 \leq \|z\|_1 \Rightarrow x \text{ Minimierer von } (P_{1,\eta}). \quad \square$$

1.2. Die Nullraumeigenschaft. Voraussetzung: $x^\#$ ist sparse.

Frage: Unter welchen Bedingungen an die Matrix A sind

$$(P_0) \quad x^\# = \arg \min_{z \in \mathbb{K}^N} \|z\|_0 \text{ unter NB } Az = y$$

$$(P_1) \quad x^\# = \arg \min_{z \in \mathbb{K}^N} \|z\|_1 \text{ unter NB } Az = y$$

so, dass die Lösg. übereinstimmen?

DEFINITION 2.4. Sei $A \in \mathbb{K}^{n \times N}$.

(1) Sei $S \subset \overline{N}$. Dann hat A die *Nullraumeigenschaft* bzgl. S (NR_S), falls gilt

$$(23) \quad \|v_S\|_1 < \|v_{S^c}\|_1 \quad \forall v \in \ker A \setminus \{0\}$$

(2) A hat eine Nullraumeigenschaft der Ordnung $s \in \overline{N}$, (NR_s), falls A für jedes $S \subset \overline{N}$ mit $|S| = s$ (NR_S) erfüllt.

BEMERKUNG 2.5. (a) Addiere auf beiden Seiten von (23) $\|v_S\|_1$ und erhalte die äqui. Bed. $2\|v_S\|_1 < \|v\|_1$.

(b) Genauso: Add. $\|v_{S^c}\|_1$ zu (23): $\|v\|_1 < 2\|v_{S^c}\|_1$

(c) Ist $v \in \ker A \setminus \{0\}$ und gilt (23) für die Menge $S := \text{supp } z$, wobei z eine Beste s -Term-Approx. von v ist, so gilt (23) sicher für alle $S \subset \overline{N}$, $|S| = s$.

- (d) (a)-(c) implizieren: A hat (NR_S) genau dann, wenn $\|v\|_1 < 2\sigma_s(v)_1$ für alle $v \in \ker A \setminus \{0\}$.

LINK: VORLESUNGSVIDEO 20 VOM 6.5.25

SATZ 2.6. Sei $A \in \mathbb{K}^{m \times N}$ und $S \subset \overline{N}$. Dann sind äquiv.:

- (i) Jeder Vektor $x \in \mathbb{K}^N$ mit $\text{supp } x \subset S$ ist der eind. Minimierer von (P_1) mit $y = Ax$.
- (ii) A hat (NR_S) .

Insbes.: Jeder Vektor $x \in \mathbb{K}_s^N$ ist eind. Min. von (P_1) mit $y = Ax \Leftrightarrow A$ hat Nullraumeigenschaft der Ordnung s .

BEWEIS. (i) $\Rightarrow$ (ii): Für jedes $x \in \mathbb{K}^N$ mit $\text{supp } x \subset S$ sei x eind. Lös. von

$$x = \arg \min_{z \in \mathbb{K}^N} \|z\|_1 \quad \text{mit NB } Az = Ax$$

Sei $v \in \ker A \setminus \{0\}$. Dann ist $\text{supp } v_S \subset S$, also ist v_S eind. best. Minimum von $\|z\|_1$ unter NB $Az = Av_S$.

$0 = Av = Av_S + Av_{S^c} \Rightarrow Av_S = A(-v_{S^c})$ (damit $-v_{S^c}$ zulässig) und $v_S \neq -v_{S^c}$ (da $v \neq 0$)
 $\Rightarrow \|v_{S^c}\|_1 > \|v_S\|_1$, (denn v_S ist ja eindeutiger Minimierer), somit gilt (NR_S) für A .

(ii) $\Rightarrow$ (i): A habe (NR_S) . Sei $x \in \mathbb{K}^N$ mit $\text{supp } x \subset S$ bel. Sei $z \in \mathbb{K}^N$ mit $Az = Ax$ und $z \neq x$. Es ist zu zeigen: $\|z\|_1 > \|x\|_1$.

Dazu sei $v := x - z$, dann gilt $A(x - z) = 0$ und so $v \in \ker A \setminus \{0\}$. Weiter

$$\|x\|_1 \leq \|x - z_S\|_1 + \|z_S\|_1 = \|v_S\|_1 + \|z_S\|_1 \stackrel{NR_S}{<} \|v_{S^c}\|_1 + \|z_S\|_1 = \|z_{S^c}\|_1 + \|z_S\|_1 = \|z\|_1.$$

□

BEMERKUNG 2.7. a) ℓ_0 -Minimierung ist NP-schwer für bel./allgemeine Daten $A \in \mathbb{K}^{m \times N}$, $y \in \mathbb{K}^m$. Ist aber x s -sparse und hat A die NR_s , so kann das ℓ_0 -Min.problem

$$x_0^\# = \arg \min \|z\|_0 \quad \text{mit NB } Az = y$$

effizient durch Basisjagd gelöst werden:

$$x_1^\# = \arg \min \|z\|_1 \quad \text{mit NB } Az = y$$

(rekonstruiert x durch $x_1^\# = x$.)

b) Hat A die (NR_S) , so auch die Matrizen

(i) $\hat{A} = GA$ für jedes reguläre $G \in \mathbb{K}^{m \times m}$ (insb. Permut.)

(ii) $\tilde{A} = \begin{bmatrix} A \\ B \end{bmatrix}$ mit bel. $\tilde{m} \times N$ -Matrix B .

Dagegen hat $A \cdot \tilde{G}$ für reg. $\tilde{G} \in \mathbb{K}^{N \times N}$ nicht automatisch die (NR_S) .

LINK: VORLESUNGSVIDEO 21 VOM 6.5.25

SATZ 2.8 (Unabhängigkeit der NR_S von $\mathbb{K}$). Sei A eine reelle $m \times N$ Matrix und $S \subset \overline{N}$. Es gilt: A hat reelle $(NR_S) \Leftrightarrow A$ hat die komplexe (NR_S) .

BEMERKUNG 2.9. Aus vorherigem Satz folgt unmittelbar, dass $\forall s \in \overline{N}$ gilt: A hat reelle $(NR_s) \Leftrightarrow A$ hat komplexe (NR_s) .

BEWEIS. Es ist zu zeigen, dass

$$\left\{ \begin{array}{l} \forall w \in \mathbb{R}^N \setminus \{0\} \quad \text{mit} \quad Aw = 0 \\ \sum_{j \in S} |w_j| < \sum_{j \in S^c} |w_j| \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \forall u + iv \in \mathbb{C}^N \setminus \{0\} \quad \text{mit} \quad \underbrace{A(u + iv) = 0}_{Au + iAv = 0 + i0} \\ \sum_{j \in S} \sqrt{u_j^2 + v_j^2} < \sum_{j \in S^c} \sqrt{u_j^2 + v_j^2} \end{array} \right.$$

Zu $\Leftarrow$: Betrachte Spezialfall $v \equiv 0$.

Zu $\Rightarrow$: Falls u und v lin. abh., so ist entweder $v \equiv 0$ oder es existiert ein $t \in \mathbb{R}$ mit $u = tv$.

Im letzteren Fall folgt

$$\begin{aligned} \sum_{j \in S} \sqrt{u_j^2 + v_j^2} &= \sum_{j \in S} \sqrt{t^2 v_j^2 + v_j^2} \\ &= \sqrt{t^2 + 1} \sum_{j \in S} \sqrt{v_j^2} \stackrel{\text{reelle NRE}}{<} \sqrt{t^2 + 1} \sum_{j \in S^c} \sqrt{v_j^2} \\ &= \sum_{j \in S^c} \sqrt{t^2 v_j^2 + v_j^2} = \sum_{j \in S^c} \sqrt{u_j^2 + v_j^2} \end{aligned}$$

was gewünscht war.

Falls u und v sind unabhängig, dann ist

$$w = \cos \theta \cdot u + \sin \theta \cdot v \neq 0, \quad \text{für bel. } \theta \in \mathbb{R}.$$

Linearität gibt: $Aw = \cos \theta Au + \sin \theta Av = 0$. Reelle (NR_S) liefert

$$(24) \quad \sum_{j \in S} |(\cos \theta)u_j + (\sin \theta)v_j| = \sum_{j \in S} |w_j| < \sum_{j \in S^c} |w_j| = \sum_{j \in S^c} |(\cos \theta)u_j + (\sin \theta)v_j|$$

Polarkoordinaten für $(u_j, v_j) \in \mathbb{R}^2$ ergeben $u_j = \sqrt{v_j^2 + u_j^2} \cos \theta_j$, $v_j = \sqrt{v_j^2 + u_j^2} \sin \theta_j$ mit $\theta_j \in [-\pi, \pi)$. Dann

$$\text{LHS von (24)} = \sum_{j \in S} \underbrace{|(\cos \theta)(\cos \theta_j) + (\sin \theta)(\sin \theta_j)|}_{=|\cos(\theta - \theta_j)|} \sqrt{v_j^2 + u_j^2},$$

und ähnlich für RHS von (24).

Integriere (24) über $\theta \in [-\pi, \pi)$: Wegen Periodizität $\int_{-\pi}^{\pi} |\cos(\theta - t)| d\theta = \int_{-\pi}^{\pi} |\cos(\theta)| d\theta = 4$, so dass

$$4 \sum_{j \in S} \sqrt{u_j^2 + v_j^2} \leq 4 \sum_{j \in S^c} \sqrt{u_j^2 + v_j^2},$$

daher folgt die Beobachtung. □

1.3. Stabilität. ($Ax = y$ lösen für x komprimierbar)

LINK: VORLESUNGSVIDEO 22 VOM 6.5.25

Die Annahme $x \in \mathbb{K}_s^N$ ist oft zu restriktiv.

Frage: Gelten die obigen Aussagen auch, falls x nur komprimierbar, d.h. nahe an einem Vektor z in $\mathbb{K}_s^N$?

Wir (können/erzielen) nicht identische Resultate, aber analog mit zus. Fehlerterm, der durch $\|x - z\|_1$ bestimmt ist.

Dazu verschärfen wir die Definition von Nullraumeigenschaft:

DEFINITION 2.10. (i) Sei $S \subset \overline{N}$. A hat die *stabile NRE* bzgl. S mit Konstante $\rho \in (0, 1)$ falls gilt

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 \quad \forall v \in \ker A. \quad (\text{SNR}_S^\rho)$$

(ii) Sei $s \in \overline{N}$. A hat die *stabile NRE der Ordnung s* (SNR_s^ρ) falls $\forall S \subset \overline{N}$ Matrix A die (SNR_S^ρ) hat.

BEMERKUNG 2.11 (Bez. zwischen NRE, SNR). Die Implikation $(\text{SNR}_S^\rho) \Rightarrow (\text{NR}_S)$ ist trivial.

Um die andere Richtung besser zu verstehen, sei $A \in \mathbb{K}^{m \times N}$, $S \subset \overline{N}$, $|S| \leq s$. Sei weiter

$$R_S : \ker A \rightarrow \mathbb{K}^N, \quad R_S v = \mathbb{1}_S v = v_S$$

der Restriktionsoperator. Dann

$$\begin{aligned} \text{NR}_S &\Leftrightarrow 2\|v_S\|_1 < \|v\|_1 \quad \forall v \in \ker A \setminus \{0\} \\ &\Leftrightarrow \|R_S v\|_1 < \frac{1}{2}\|v\|_1 \quad \forall v \in \ker A \setminus \{0\} \\ &\Leftrightarrow \|R_S\| < \frac{1}{2}. \end{aligned}$$

wobei $\|R_S\| = \|R_S\|_{1,1}$ die Operatornorm von $R_S : (\ker A, \|\cdot\|_1) \rightarrow (\mathbb{K}^N, \|\cdot\|_1)$ ist, d.h.

$$\|R_S\| = \sup_{\substack{v \in \ker A \\ \|v\|_1=1}} \|R_S v\|_1$$

Damit bewiesen:

$$A \text{ hat } (\text{NR}_s) \Leftrightarrow \mu := \max_{\substack{S \subset \overline{N} \\ |S| \leq s}} \|R_S\| < \frac{1}{2}$$

Setze nun $\rho := \frac{\mu}{1-\mu} < 1$ (da $1 - \mu > \frac{1}{2}$). Nun äquivalente Umformungen

$$A \text{ hat } \text{NR}_S \Leftrightarrow \text{für ein:} \quad \mu \in [0, 1/2) : \|R_S v\|_1 \leq \mu \|v\|_1 \quad \forall v \in \ker A$$

und

$$\begin{aligned}
& \|R_S v\|_1 \leq \mu \|v\|_1 \quad \forall v \in \ker A \\
& \Leftrightarrow \|v_S\|_1 \leq \frac{\mu}{1-\mu} \Big/ \frac{1}{1-\mu} \cdot \|v_1\| \quad \forall v \in \ker A \\
& \Leftrightarrow \|v_S\|_1 \leq \frac{\rho}{1+\rho} \|v\|_1 \quad \forall v \in \ker A \\
& \Leftrightarrow (1+\rho) \|v_S\|_1 \leq \rho \|v\|_1 \quad \forall v \in \ker A \\
& \Leftrightarrow \|v_S\|_1 \leq \rho \|v_{S^c}\|_1 \quad \forall v \in \ker A \\
& \Leftrightarrow \text{SNR}_S^\rho.
\end{aligned}$$

Wir haben oben die folgende Aussage gezeigt: A die $\text{NRE}_S \Rightarrow \exists \rho \in (0, 1)$ s.d. A die SNR_S^ρ hat.

Nützlich hier und sonst:

LEMMA 2.12. Sei $S \subset \overline{N}$, $z, x \in \mathbb{K}^N$. Dann

$$\|(x - z)_{S^c}\|_1 \leq \|z\|_1 - \|x\|_1 + \|(x - z)_S\|_1 + 2\|x_{S^c}\|_1.$$

BEWEIS.

$$\begin{aligned}
\|x\|_1 &= \|x_{S^c}\|_1 + \|x_S\|_1 \quad (\text{gilt nur für } \|\cdot\|_1) \\
&\leq \|x_{S^c}\|_1 + \|(x - z)_S\|_1 + \|z_S\|_1.
\end{aligned}$$

Außerdem gilt:

$$\|(x - z)_{S^c}\|_1 \leq \|x_{S^c}\|_1 + \|z_{S^c}\|_1,$$

Addieren impliziert

$$\|x\|_1 + \|(x - z)_{S^c}\|_1 \leq \|z\|_1 + \|(x - z)_S\|_1 + 2\|x_{S^c}\|_1.$$

□

LINK: VORLESUNGSVIDEO23 VOM 12.5.25

SATZ 2.13. a) Seien $A \in \mathbb{K}^{m \times N}$, $\rho \in (0, 1)$, $s \in \overline{N}$, $x \in \mathbb{K}^N$. Hat A die (SNR_s^ρ) und ist $x^\#$ eine Lösung von (P_1) mit $y = Ax$, so gilt:

$$\|x^\# - x\|_1 \leq 2 \frac{1+\rho}{1-\rho} \sigma_s(x)_1.$$

Teil (a) ist Spezialfall von

b) Seien $A \in \mathbb{K}^{m \times N}$, $\rho \in (0, 1)$, $S \subset \overline{N}$. Dann: A erfüllt (SNR_S^ρ) genau dann, wenn

$$(25) \quad \|z - x\|_1 \leq \frac{1+\rho}{1-\rho} \left(\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1 \right) \quad \forall x, z \in \mathbb{K}^N \text{ mit } Az = y := Ax.$$

Insb. falls $\text{supp } x \subset S$ $\|z - x\|_1 \leq \frac{1+\rho}{1-\rho} (\|z\|_1 - \|x\|_1)$.

BEWEIS. Zunächst beweisen wir, dass (a) Spezialfall von (b) ist. Setze $z = x^\#$. Zulässiger Vektor, da $Ax = Az$. $\|z\|_1 = \|x^\#\| \leq \|x\|_1$, da Optimum. Wähle $S \subset \bar{N}$ s.d. $\sigma_S(x)_1 = \|x_{S^c}\|_1$ (d.h. S enthält Indiz. der größten Komp. von x). Dann

$$\|x^\# - x\|_1 \stackrel{(b)}{\leq} \left(\underbrace{\|x^\#\|_1 - \|x\|_1}_{\leq 0} + 2\|x_{S^c}\|_1 \right) \left(\frac{1+\rho}{1-\rho} \right) \leq \left(\frac{2(1+\rho)}{1-\rho} \right) \sigma_S(x)_1.$$

Jetzt beweisen wir Teil (b).

Zu (25) $\Rightarrow$: A hat SNR_S^ρ : Sei $v \in \ker A$, z.z. $\|v_S\|_1 \leq \rho\|v_{S^c}\|_1$.

Da $0 = Av = A(v_S + v_{S^c})$ gilt $Av_{S^c} = -Av_S$. Einsetzen in (25) mit $x = v_S$ und $z = -v_{S^c}$:

$$\begin{aligned} \|v\|_1 &\leq \left(\frac{1+\rho}{1-\rho} \right) (\|v_{S^c}\|_1 - \|v_S\|_1) \\ &\Leftrightarrow (1-\rho)\|v\|_1 + (1+\rho)\|v_S\|_1 \leq (1+\rho)\|v_{S^c}\|_1 \\ &\Leftrightarrow \|v\|_1 - \rho\|v_S\|_1 - \rho\|v_{S^c}\|_1 + \|v_S\|_1 + \rho\|v_S\|_1 \leq \|v_{S^c}\|_1 + \rho\|v_{S^c}\|_1 \\ &\Leftrightarrow 2\|v_S\|_1 \leq 2\rho\|v_{S^c}\|_1. \end{aligned}$$

Zu $\text{SNR}_S^\rho \Rightarrow$ (25): Seien $x, z \in \mathbb{K}^N$ mit $Ax = Az \Rightarrow v := z - x \in \ker A$.

$$\text{SNR}_S^\rho \Rightarrow \|v_S\|_1 \leq \rho\|v_{S^c}\|_1.$$

$$\stackrel{\text{Lemma 2.12}}{\Rightarrow} \|v_{S^c}\|_1 \leq \|z\|_1 - \|x\|_1 + \underbrace{\|v_S\|_1}_{\leq \rho\|v_{S^c}\|_1} + 2\|x_{S^c}\|_1$$

$$\Leftrightarrow \|v_{S^c}\|_1 \leq \frac{1}{1-\rho} (\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1).$$

Dann folgt

$$\|z - x\|_1 = \|v\|_1 = \|v_{S^c}\|_1 + \|v_S\|_1 \stackrel{\text{SNR}}{\leq} (1+\rho)\|v_{S^c}\|_1 \leq \frac{1+\rho}{1-\rho} (\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1).$$

□

LINK: VORLESUNGSVIDEO 24 VOM 12.5.25

1.4. Robustheit (bzgl. Messfehler). Bisher: $x \in \mathbb{K}_S^N$ "leicht verletzt".

Jetzt: $y = Ax$ "leicht verletzt".

Mit Fehler $e \in \mathbb{K}^m$ behaftete Messung:

$$y = Ax + e \Leftrightarrow Ax - y = -e \Rightarrow \|Ax - y\|_2 = \|e\|_2$$

Terminologie: Basisjagd heißt *robust*, wenn sie stabil bzgl. Messfehler ist.

Erinnerung: Basisjagd mit Rauschen:

$$(P_{1,\eta}) \quad x^\# := \arg \min_{z \in \mathbb{K}^N} \|z\|_1 \quad \text{mit NB} \quad \|Az - y\|_2 \leq \eta$$

Noch eine Variante der NRE:

DEFINITION 2.14. Seien $A \in \mathbb{K}^{m \times N}$, $\rho \in (0, 1)$, $\tau > 0$.

(i) Sei $S \subset \bar{N}$. A hat die robuste NRE bzgl. S mit Konstanten ρ, τ falls gilt

$$\forall v \in \mathbb{K}^N : \|v_S\|_1 \leq \rho\|v_{S^c}\|_1 + \tau\|Av\|_2. \quad (\text{RNR}_S^{\rho,\tau})$$

- (ii) Sei $s \in \overline{N}$. A hat die *robuste NRE der Ordnung s mit Konst. ρ, τ* falls gilt hat $(\text{RNR}_S^{\rho, \tau})$ für alle $S \subset \overline{N}$ mit $|S| \leq s$.

BEMERKUNG 2.15. (a) $(\text{RNR} \implies \text{SNR})$ A habe $\text{RNR}_S^{\rho, \tau}$ und $v \in \ker A \implies Av = 0 \implies \|v_S\|_1 \leq \rho \|v_{S^c}\|_1$

- (b) Der Term $\|Av\|_2$ könnte statt der euklidischen Norm eine beliebige Norm $\|\cdot\|$ auf $\mathbb{K}^m$ beinhalten (Wir konzentrieren uns vor allem auf $\|\cdot\| := \|\cdot\|_2$)

Wieder haben wir eine Äquivalenz:

SATZ 2.16. Seien $A \in \mathbb{K}^{m \times N}$, $\rho \in (0, 1)$, $\tau > 0$, $T \subset \overline{N}$. Dann: A hat $\text{RNR}_T^{\rho, \tau}$ genau dann, wenn

$$(26) \quad \|z - x\|_1 \leq \frac{1 + \rho}{1 - \rho} (\|z\|_1 - \|x\|_1 + 2\|x_{T^c}\|_1) + \frac{2\tau}{1 - \rho} \|A(z - x)\| \quad \forall x, z \in \mathbb{K}^N.$$

BEWEIS. Zu (26) $\implies$ RNR: Es ist zu zeigen: $\forall v \in \mathbb{K}^N$ gilt $\|v_T\|_1 \leq \rho \|v_{T^c}\|_1 + \tau \|Av\|$. Setze $x := -v_T$, $z = v_{T^c}$, so dass $v = z - x$.

$$\begin{aligned} \stackrel{(26)}{\implies} \|v\|_1 &\leq \frac{1 + \rho}{1 - \rho} (\|v_{T^c}\|_1 - \|v_T\|_1) + \frac{2\tau}{1 - \rho} \|Av\| \\ &\stackrel{\rho \leq 1}{\Leftrightarrow} (1 - \rho) (\|v_T\|_1 + \|v_{T^c}\|_1) + \|v_T\|_1 + \rho \|v_T\|_1 \leq (1 + \rho) \|v_{T^c}\|_1 + 2\tau \|Av\| \\ &\Leftrightarrow 2\|v_T\|_1 \leq 2\rho \|v_{T^c}\|_1 + 2\tau \|Av\|. \end{aligned}$$

"RNR $\implies$ (26)". Seien $x, z \in \mathbb{K}^N$ und setze $v := z - x$.

RNR $\implies \|v_T\|_1 \leq \rho \|v_{T^c}\|_1 + \tau \|Av\|$.

Lemma 2.12 $\implies$

$$\|v_{T^c}\|_1 \leq \|z\|_1 - \|x\|_1 + \underbrace{\|v_T\|_1}_{\leq \rho \|v_{T^c}\|_1} + 2\|x_{T^c}\|_1$$

Also:

$$\begin{aligned} \|v_{T^c}\|_1 &\leq \|z\|_1 - \|x\|_1 + \rho \|v_{T^c}\|_1 + \tau \|Av\| + 2\|x_{T^c}\|_1 \\ &\stackrel{\rho \leq 1}{\Leftrightarrow} \|v_{T^c}\|_1 \leq \frac{1}{1 - \rho} (\|z\|_1 - \|x\|_1 + 2\|x_{T^c}\|_1 + \tau \|Av\|) \end{aligned}$$

Es folgt:

$$\begin{aligned} \|v\|_1 &= \|v_{T^c}\|_1 + \|v_T\|_1 \leq (1 + \rho) \|v_{T^c}\|_1 + \tau \|Av\| \\ &\leq \frac{1 + \rho}{1 - \rho} (\|z\|_1 - \|x\|_1 + 2\|x_{T^c}\|_1) + \tau \left(1 + \frac{1 + \rho}{1 - \rho}\right) \|Av\|. \end{aligned}$$

□

LINK: VORLESUNGSVIDEO 25 VOM 12.5.25

Das gilt für allgemeine $x, z \in \mathbb{K}^N$. Nun optimieren:

KOROLLAR 2.17. Sei $s \in \overline{N}$, $\eta \geq 0$, $A \in \mathbb{K}^{m \times N}$ mit $\text{RNR}_s^{\rho, \tau}$, $x \in \mathbb{K}^N$, $y := Ax \in \mathbb{K}^m$ und $x^\#$ sei Lösung von:

$$\min_{z \in \mathbb{K}^N} \|z\|_1 \quad \text{mit NB} \quad \|y - Az\| \leq \eta.$$

Dann:

$$\|x - x^\#\|_1 \leq 2 \frac{1 + \rho}{1 - \rho} \sigma_s(x)_1 + \frac{2\tau}{1 - \rho} \eta.$$

BEWEIS. Wähle $T \subset \overline{N}$, so dass $\|x_{T^c}\|_1 = \sigma_s(x)_1$, $z = x^\#$, $|T| = s$. Dann $\|z\|_1 = \|x^\#\|_1 \leq \|x\|_1$, da minimal, und

$$\|x - x^\#\|_1 \leq \frac{1+\rho}{1-\rho} \left[\underbrace{\|x^\#\|_1 - \|x\|_1}_{\leq 0} + 2 \underbrace{\|x_{T^c}\|_1}_{=\sigma_s(x)_1} \right] + \frac{2\tau}{1-\rho} \underbrace{\|A(x^\# - x)\|}_{\leq \eta}.$$

□

BEMERKUNG 2.18. (a) Für $\eta = 0$ erhalten wir Satz 2.13 (a)

(b) Satz 2.16 gilt für allgemeine Normen auf $\mathbb{K}^m$, nicht nur $\|\cdot\| = \|\cdot\|_2$.

Auch im folgenden Kapitel.

1.5. Erweiterung auf ℓ_p -Normen auf $\mathbb{K}^N$. Sei nun $p \geq 1$. Wollen Rekonstruktionsfehler in $\|\cdot\|_p$ auf $\mathbb{K}^N$ messen.

DEFINITION 2.19. Seien $q \geq 1$, $s \in \overline{N}$, $A \in \mathbb{R}^{m \times N}$, $\rho \in (0, 1)$, $\tau > 0$. A hat ℓ_q -robuste NRE der Ordnung s mit Konstanten ρ, τ , falls: $\forall T \subset \overline{N}$ mit $|T| \leq s$ und $\forall v \in \mathbb{K}^N$ gilt

$$\|v_T\|_q \leq \frac{\rho}{s^{1-\frac{1}{q}}} \|v_{T^c}\|_1 + \tau \|Av\|. \quad (\text{RNR})_{s,q}^{\rho,\tau}$$

LINK: VORLESUNGSVIDEO 26 VOM 13.5.25

BEMERKUNG 2.20. Für $|T| \leq s$, $1 \leq p \leq q$ gilt:

$$\|v_T\|_p^p = \sum_{j \in T} |v_j|^p \cdot 1 \stackrel{\text{Hölder}}{\leq} \left(\sum_{j \in T} \left(|v_j|^p \right)^{\frac{q}{p}} \right)^{\frac{p}{q}} \cdot \left(\sum_{j \in T} 1 \right)^{1-\frac{p}{q}} = \|v_T\|_q^p s^{1-\frac{p}{q}}.$$

$$\Rightarrow \|v_T\|_p \leq s^{\frac{1}{p}-\frac{1}{q}} \|v_T\|_q$$

Falls $\text{RNR}_{s,q}^{\rho,\tau}$ gilt, dann gilt: $\forall T \subset \overline{N}$ mit $|T| \leq s$, $\forall v \in \mathbb{K}^N$

$$(27) \quad \|v_T\|_p \leq s^{\frac{1}{p}-\frac{1}{q}} \|v_T\|_q \leq \frac{\rho}{s^{1-1/p}} \|v_{T^c}\|_1 + \tau s^{\frac{1}{p}-\frac{1}{q}} \|Av\|$$

Demnach ergibt sich für $1 \leq p \leq q$:

Gilt $\text{RNR}_{s,q}^{\rho,\tau}$ mit Norm $\|\cdot\|$, so gilt auch $\text{RNR}_{s,p}^{\rho,\tau}$ mit Norm $\|\cdot\| = s^{\frac{1}{p}-\frac{1}{q}} \|\cdot\|$.

Insbesondere $\text{RNR}_{s,1}^{\rho,\tau}$ mit $\|\cdot\| = s^{1-\frac{1}{q}} \|\cdot\|$.

Also kann man $\text{RNR}_{s,q}^{\rho,\tau}$ als Verschärfung von $\text{RNR}_s^{\rho,\tau}$ (genauer $\text{RNR}_s^{\rho,s^{1-1/q}\tau}$) sehen. Uns wird später insbesondere der Fall $q = 2$ interessieren.

SATZ 2.21 (Robustheit der Basisjagd mit quadratischen NB). Seien $\eta \geq 0$, $s \in (0, 1)$, $\tau > 0$, $s \in \overline{N}$ und $A \in \mathbb{K}^{m \times N}$ mit $\text{RNR}_{s,2}^{\rho,\tau}$. Dann existieren Konstanten c, D abhängig nur von ρ und τ , s. d.

$$\forall x \in \mathbb{K}^N \text{ Lösung von } x^\# = \arg \min_{z \in \mathbb{K}^N} \|z\|_1 \quad \text{mit NB} \quad \|Ax - Az\|_2 \leq \eta$$

erfüllt

$$(28) \quad \|x - x^\#\|_p \leq \frac{c}{s^{1-1/p}} \sigma_s(x)_1 + D s^{\frac{1}{p}-\frac{1}{2}} \eta \quad p \in [1, 2].$$

BEMERKUNG 2.22. Insbesondere:

$$\begin{aligned}\|x - x^\# \|_1 &\leq c\sigma_s(x)_1 + D\sqrt{s}\eta, \\ \|x - x^\# \|_2 &\leq \frac{c}{\sqrt{s}}\sigma_s(x)_1 + D\eta.\end{aligned}$$

Die allgemeine Aussage für $p \in [1, 2]$ kann man dann durch Interpolation bekommen.

Frage: Geht auch

$$(29) \quad \|x - x^\# \|_p \leq c\sigma_s(x)_p + Ds^{\frac{1}{p}-\frac{1}{2}}\eta ?$$

(Insbesondere für $p = 2$?)

Antwort: Nein! Zumindest nicht für interessante Dimensionen m, N .

Frage: Haben wir dadurch viel verloren?

Antwort: Nein (zumindest in repräsentativen Fällen)

Zur Begründung: Erinnern wir uns, dass x aus dem $\|\cdot\|_{\ell_q}$ -Ball, $q < 1$, Modelle für komprimierbare Vektoren sind. Seien $q < 1$, $\|x\|_q \leq 1$, $p \geq 1$, und sei $\eta = 0$ (also perfekte Messungen). Lemma 1.6 besagt in diesem Fall

$$\sigma_s(x)_p \leq s^{\frac{1}{p}-\frac{1}{q}}$$

Nun folgt einerseits aus (28) mit $p \in [1, 2]$

$$\begin{aligned}\Rightarrow \|x - x^\# \|_p &\leq \frac{c}{s^{1-\frac{1}{p}}}\sigma_s(x)_1 + 0 \\ &\stackrel{\text{Lemma 1.6}}{\leq} \frac{c}{s^{1-1/p}} \frac{1}{s^{1/q-1}} \leq \frac{c}{s^{1/q-1/p}} = c s^{1/p-1/q}\end{aligned}$$

Andererseits impliziert die hypothetische Abschätzung (29) zusammen mit Lemma 1.6

$$(30) \quad \|x - x^\# \|_p \leq c\sigma_s(x)_p + 0 \leq cs^{\frac{1}{p}-\frac{1}{q}}$$

Das heißt: Zumindest für die betrachtete Klasse von Vektoren und $\eta = 0$ erhalten wir auf diese Art zwei obere Schranken von derselben Größenordnung.

LINK: VORLESUNGSVIDEO 27 VOM 13.5.25

Satz 2.21 ergibt sich als Spezialfall von:

SATZ 2.23. Seien $1 \leq p \leq q$, $s \in \overline{N}$, $\rho \in (0, 1)$, $\tau > 0$, $A \in \mathbb{K}^{m \times N}$ mit $RNR_{s,q}^{\rho,\tau}$. Dann gilt:

$$\forall x, z \in \mathbb{K}^N : \|z - x\|_p \leq \frac{c}{s^{1-1/p}} \left[\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1 \right] + Ds^{1/p-1/q} \|A(z - x)\|$$

mit $c = \frac{(1+\rho)^2}{1-\rho}$, $D = \frac{(3+\rho)}{1-\rho} \tau$.

BEWEIS. $RNR_{s,q}^{\rho,\tau}$ impliziert durch (27)

$$\|v_T\|_1 \leq \rho \|v_{T^c}\|_1 + \tau s^{1-1/p} \|Av\| \quad (RNR_{s,1}^{\rho,\tau s^{1-1/p}})$$

und

$$\|v_T\|_p \leq \frac{\rho}{s^{1-1/p}} \|v_{T^c}\|_1 + \tau s^{1/p-1/q} \|Av\| \quad (RNR_{s,p}^{\rho s^{1/p-1}, \tau s^{1/p-1/q}})$$

jeweils für alle $v \in \mathbb{K}^N$ und $T \subset \bar{N}$ mit $|T| \leq s$. Wähle erst $T \subset \bar{N}$ s.d. $|T| \leq s$ und $\|x_{T^c}\|_1 = \sigma_s(x)_1$. Satz 2.16 liefert

$$\|z - x\|_1 \leq \frac{1 + \rho}{1 - \rho} \left[\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1 \right] + \frac{2\tau}{1 - \rho} s^{1-1/q} \|A(z - x)\|.$$

Wähle jetzt $T \subset \bar{N}$ s.d. $|T| \leq s$ und $\|(z - x)_{T^c}\|_p = \sigma_s(z - x)_p$. Dann

$$\begin{aligned} \|z - x\|_p &\leq \|(z - x)_T\|_p + \|(z - x)_{T^c}\|_p \\ &= \sigma_s(z - x)_p + \|(z - x)_T\|_p \\ &\stackrel{\text{Lem 1.6}}{\leq} s^{1/p-1} \|z - x\|_1 + \|(z - x)_T\|_p \\ &\leq s^{1/p-1} \|z - x\|_1 + \rho s^{1/p-1} \|(z - x)_{T^c}\|_1 + \tau s^{1/p-1/q} \|A(z - x)\| \\ &\leq \frac{(1 + \rho)^2}{1 - \rho} s^{1/p-1} \left[\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1 \right] \\ &\quad + 2\tau \left(\frac{1 + \rho}{1 - \rho} \right) \|A(z - x)\| s^{1/p-1/q} + \tau s^{1/p-1/q} \|A(z - x)\|. \end{aligned}$$

□

1.6. Rekonstruktion “individueller Vektoren”.

LINK: VORLESUNGSVIDEO 28 VOM 13.5.25

Lösen von (Fehlerbehafteten) Gleichungen in ∞ -dim. Räumen.

Typisches Setting: $A : \ell_2(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$ linear und injektiv, d.h. $\forall x, y \in \ell_2$ gilt: $Ax = Ay \Rightarrow x = y$. Insbesondere ex. der inverse Operator $A^{-1} : \text{Bild } A \rightarrow \ell_2(\mathbb{N})$, d.h. $A^{-1}Ax = x$ für alle $x \in \ell_2(\mathbb{N})$.

Wo ist also das Problem ?

In endlich dim. VR ist jede lineare Abbildung (dann genannt: Operator) stetig, ja sogar Lipschitz-stetig mit Lipschitz-Konstante gegeben durch

$$\|A\|_{\text{Operatornorm von } A} := \sup_{\|x\|=1} \|Ax\| < \infty.$$

Solche linearen Operatoren nennt man auch "beschränkt" (Vorsicht vorn Missverständnissen).

Damit gilt für fehlerhafte Messung:

$$y = Ax + e, \quad A^{-1}y = A^{-1}Ax + A^{-1}e = x + A^{-1}e$$

mit Fehlergröße $\|A^{-1}e\| \leq \|A^{-1}\| \underbrace{\|e\|}_{\leq \eta} \leq \|A^{-1}\| \eta$. D.h. falls ich η verkleinere (z.B. durch präzisere Messungen), so wird der Fehler $A^{-1}e$ *proportional* dazu kleiner.

Bei $A : \ell_2(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$ linear & injektiv muss A^{-1} nicht Lipschitz-stetig sein. In diesem Fall ex. normierte Folge $(e_n)_n$, mit $\|A^{-1}e_n\| \xrightarrow{n \rightarrow \infty} \infty$. Dann folgt für $y_n = Ax + \eta e_n, \eta > 0$ dass,

$$\|\eta e_n\| \leq \eta, \quad \text{aber} \quad \|A^{-1}y_n - x\| = \|x + \eta A^{-1}e_n - x\| = \eta \|A^{-1}e_n\| \xrightarrow{n \rightarrow \infty} \infty.$$

D.h. selbst kleine Messfehler können gravierende Folgen für die Rekonstruktion des Signals haben.

Hier: Nur endlich dimensionaler Fall. Da wir uns aber für große Dimensionen interessieren, werden ähnliche Effekte auftreten, wie bei unendlich-dimensionalen normierten Räumen. Viele Resultate gelten auch im ∞ -dimensionalen Fall und entsprechen recht frischen Ergebnissen.

LINK: VORLESUNGSVIDEO 29 VOM 19.5.25

Hier unterscheiden wir $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$. Für $z \in \mathbb{C}$ sei

$$\operatorname{sgn} z = \begin{cases} \frac{z}{|z|}, & \text{für } z \neq 0 \\ 0, & z = 0 \end{cases}$$

$$\Rightarrow z = |z| \operatorname{sgn} z, \quad |z| = \frac{z}{\operatorname{sgn} z} = z \overline{\operatorname{sgn} z}$$

sowie: Kontinuum von Werten im Gegensatz zu $\operatorname{sgn} x \in \{-1, 0, 1\}$ bei $x \in \mathbb{R}$.

Für $z \in \mathbb{C}^N$: setze $\operatorname{sgn} z = (\operatorname{sgn} z_j)_{j=1}^N$.

SATZ 2.24. Sei $A \in \mathbb{C}^{m \times N}$, $S \subset \overline{N}$, $x \in \mathbb{C}^N$, $\operatorname{supp} x = S$. Dann ist x eind. Minimierer von $\min_{z \in \mathbb{C}^N} \|z\|_1$ unter NB $Az = Ax$, falls eine der Bed (a), (b) erfüllt ist:

(a)

$$\underbrace{\left| \sum_{j \in S} v_j \overline{\operatorname{sgn} x_j} \right|}_{|\langle v_S, \operatorname{sgn} x \rangle|} < \|v_{S^c}\|_1 \text{ für alle } v \in \ker A \setminus \{0\}$$

(b) A_S ist injektiv und es gibt ein $h \in \mathbb{C}^m$, s.d. $(A^*h)_j = \operatorname{sgn} x_j \forall j \in S$, und $|(A^*h)_l| < 1 \forall l \in S^c$.

BEWEIS. “(a) $\Rightarrow x$ eind. Minimierer von (P_1) mit $y = Ax$ ”:

Sei $z \neq x \in \mathbb{C}^N$ mit $Ax = Az$ und damit $v := x - z \in \ker A \setminus \{0\}$. Dann

$$\|z\|_1 = \|z_S\|_1 + \|z_{S^c}\|_1 = \|(x - v)_S\|_1 + \|v_{S^c}\|_1 \quad \text{no cancelations}$$

$$\stackrel{(a)}{>} |\langle x - v, (\operatorname{sgn} x)_S \rangle| + |\langle v, (\operatorname{sgn} x)_S \rangle| \quad \text{cancelations possible}$$

$$\stackrel{\Delta\text{-Ungl.}}{>} |\langle x, (\operatorname{sgn} x)_S \rangle| \stackrel{S=\operatorname{supp} x}{=} \|x\|_1$$

“(b) $\Rightarrow$ (a)”: Sei $v \in \ker A \setminus \{0\} \Rightarrow Av_S = -Av_{S^c}$. Dann:

$$\left| \sum_{j \in S} v_j \overline{\operatorname{sgn} x_j} \right| \stackrel{(b)}{=} |\langle v_S, A^*h \rangle_{\mathbb{C}^N}| = |\langle Av_S, h \rangle_{\mathbb{C}^m}|$$

$$= |\langle Av_{S^c}, h \rangle| = |\langle v_{S^c}, A^*h \rangle| \leq \max_{l \in S^c} |(A^*h)_l| \|v_{S^c}\|_1 \stackrel{(b)}{<} \|v_{S^c}\|_1$$

da $v_{S^c} \neq 0$. Dies folgt aus $v \neq 0$. Denn wäre $v_{S^c} = 0$, so $v = v_S \Rightarrow A_S v_S = Av = 0$. Widerspruch zu A_S injektiv. $\square$

In der Tat sind (a) $\Leftrightarrow$ (b). Um das zu zeigen zitieren wir einen Trennungssatz:

SATZ 2.25. Seien $A, B \in \mathbb{R}^n$ konvex, $A^\circ \cap B^\circ = \emptyset$. Dann ex. $\omega \in \mathbb{R}^n, \lambda \in \mathbb{R}$: mit $A \subset \{x \in \mathbb{R}^n | \langle x, \omega \rangle \leq \lambda\}$, $B \subset \{x \in \mathbb{R}^n | \langle x, \omega \rangle \geq \lambda\}$

BEMERKUNG 2.26.

- Falls $A \cap B$ einelementig $= \{x_0\}$ so $A^\circ \cap B^\circ = \emptyset$ und $\lambda = \langle x_0, \omega \rangle$ gute Wahl.
- Falls $B \subset$ Hyperebene, so ex $\omega \in \mathbb{R}^n, \lambda \in \mathbb{R}$ mit $B \subset \{x \mid \langle x, \omega \rangle = \lambda\}$.

LINK: VORLESUNGSVIDEO 30 VOM 19.5.25

SATZ 2.27. (a) $\Rightarrow$ (b) in Satz 2.24.

BEWEIS. (a) $\Rightarrow \|v_{S^c}\|_1 > 0 \ \forall v \in \ker A \setminus \{0\}$.

Zeigen zuerst: A_S injektiv. Denn: Falls $A_S v_S = 0$ für $v_S \neq 0$, so setze

$$\mathbb{C}^N \ni v = \begin{cases} v_S & \text{auf } S \\ 0 & \text{sonst} \end{cases}$$

$\Rightarrow Av = A_S v_S = 0 \Rightarrow v \in \ker A \setminus \{0\}$ und $\|v_{S^c}\|_1 = 0$, Widerspruch zu Bedingung (a).

Die Fkt. $v \mapsto \frac{|\langle v, \text{sgn } x_S \rangle|}{\|v_{S^c}\|_1} \stackrel{(a)}{<} 1$ auf ℓ_2 -Einheitssphäre von $\ker A$ ist stetig.

Kompaktheit

$$\Rightarrow \max_{v \in \ker A, \|v\|_2=1} \frac{|\langle v, \text{sgn } x_S \rangle|}{\|v_{S^c}\|_1} = \mu < 1$$

Skalarmultiplikation mit $\|v_{S^c}\|_1$ ergibt

$$|\langle v, \text{sgn } x_S \rangle| \leq \mu \|v_{S^c}\|_1, \quad \forall v \in \ker A, \|v\|_2 = 1$$

Homogenität ergibt:

$$|\langle v, \text{sgn } x_S \rangle| \leq \mu \|v_{S^c}\|_1, \quad \forall v \in \ker A,$$

Wähle $\nu \in (\mu, 1)$ und definiere

$$C := \{z \in \mathbb{C}^n \mid \|z_S\|_1 + \nu \|z_{S^c}\|_1 \leq \|x\|_1\} \quad (\text{konvexe Menge in } \mathbb{R}^{2N})$$

$$D := \{z \in \mathbb{C}^N \mid Az = Ax\} \quad (\text{affine Menge}).$$

Es gilt: $C \cap D = \{x\}$. Denn: Sei $x \neq z \in C \cap D$ und $v := x - z \in \ker A \setminus \{0\}$. Dann

$$\begin{aligned} \|x\|_1 &\stackrel{\text{Def. von } C}{\geq} \|z_S\|_1 + \nu \|z_{S^c}\|_1 \stackrel{\text{supp } x \in S}{=} \|(x-v)_S\|_1 + \nu \|v_{S^c}\|_1 \\ &\stackrel{\mu < \nu}{>} \|(x-v)_S\|_1 + \mu \|v_{S^c}\|_1 \\ &\stackrel{(a)}{>} |\langle x-v, \text{sgn } x_S \rangle| + |\langle v, \text{sgn } x_S \rangle| \\ &\geq |\langle x, \text{sgn } x_S \rangle| = \|x\|_1 \Rightarrow \text{Widerspruch} \end{aligned}$$

Satz 2.25 für $\mathbb{C}^N \cong \mathbb{R}^{2N}$. Es gibt ein $\omega \in \mathbb{C}^N$, s.d.

$$(31) \quad C \subset \{z \in \mathbb{C}^N \mid \Re(\langle z, \omega \rangle) \leq \|x\|_1\} := C^*,$$

und

$$(32) \quad D \subset \{z \in \mathbb{C}^N \mid \Re(\langle z, \omega \rangle) = \|x\|_1\}. \quad (\text{Details als HA})$$

Ungl. (31) impliziert

$$\begin{aligned}
 \|x\|_1 &\geq \sup_{z \in C} \Re(\langle z, \omega \rangle) \geq \sup_{z \in C} \Re \left[\sum_{j \in S} z_j \overline{\omega_j} + \sum_{j \in S^c} \nu z_j \frac{1}{\nu} \overline{\omega_j} \right] \\
 &= \sup_{z \in C} \Re(\langle z_S + \nu z_{S^c}, \omega_S + \frac{1}{\nu} \omega_{S^c} \rangle) \\
 &= \|x\|_1 \cdot \underbrace{\left\| \omega_S + \frac{1}{\nu} \omega_{S^c} \right\|_\infty}_{=\max(\|\omega_S\|_\infty, \frac{1}{\nu} \|\omega_{S^c}\|_\infty)} \quad \text{durch Optimierung von } z
 \end{aligned}$$

Auf beiden Seiten steht Faktor $\|x\|_1$.

Falls $x = 0$, dann wähle $h \in \mathbb{C}^m$, $h = 0$, $\Rightarrow A^*h = 0 = \operatorname{sgn} x$ und $|(A^*h)_j| = 0 < 1$.

Im Fall $x \neq 0$, können wir $\|x\|_1 > 0$ kürzen

$$\begin{aligned}
 &\Rightarrow \begin{cases} 1 \geq \|\omega_S\|_\infty \\ 1 > \nu \geq \|\omega_{S^c}\|_\infty \end{cases} \\
 \sum_{j \in S} |x_j| \cdot 1 &\geq \underbrace{\Re(\sum_{j \in S} x_j \overline{\omega_j})}_{\text{wegen } x \in D} = \|x\|_1 = \sum_{j \in S} x_j \overline{\operatorname{sgn} x_j}
 \end{aligned}$$

$\Re(\langle x, \omega \rangle) = \|x\|_1$ wegen $x \in D$

$\Rightarrow$ Für Gleichheit ist $\omega_S = \operatorname{sgn} x_S$ notwendig.

Sei $v \in \ker A$, dann erfüllt $z = v + x \in D \subset \mathbb{C}^N$, denn $Az = Ax$, Mit (32) schliessen wir:

$$\begin{aligned}
 &\Rightarrow \Re(\langle v, \omega \rangle) = \Re(\langle z, \omega \rangle) - \Re(\langle x, \omega \rangle) = \|x\|_1 - \|x\|_1 = 0 \\
 &\Rightarrow \omega \in (\ker A)^\perp = \operatorname{Bild} A^*, \quad A^* : \mathbb{C}^m \rightarrow \mathbb{C}^N \\
 &\Rightarrow \text{ex. } h \in \mathbb{C}^m \text{ mit } \omega = A^*h
 \end{aligned}$$

□

LINK: VORLESUNGSVIDEO 31 VOM 19.5.25

BEMERKUNG 2.28. (a) Wir betrachten die Rekonstruktion von individuellen Vektoren. Trotzdem bekommen wir automatisch uniforme Rekonstruktionskriterien in Hyper-Oktanten, die durch Vorzeichen $\operatorname{sgn} x$ bestimmt sind:

Genauer: Erfüllt ein Vektor $x \in \mathbb{C}^N$ Bed. (a), dann können alle $\tilde{x} \in \mathbb{C}^N$ mit $\tilde{S} = \operatorname{supp} \tilde{x} \subset \operatorname{supp} x = S$ und $(\operatorname{sgn} \tilde{x})_{\tilde{S}} = (\operatorname{sgn} x)_{\tilde{S}}$ exakt rekonstruiert werden mit (P_1) , denn:

$$\begin{aligned}
 \forall v \in \ker A \setminus \{0\} : \left| \sum_{j \in \tilde{S}} v_j \operatorname{sgn} \tilde{x}_j \right| &= \left| \sum_{j \in S} v_j \operatorname{sgn} x_j - \sum_{j \in S \setminus \tilde{S}} v_j \operatorname{sgn} x_j \right| \\
 &\leq \left| \sum_{j \in S} v_j \operatorname{sgn} x_j \right| + \sum_{j \in S \setminus \tilde{S}} |v_j| \\
 &\stackrel{(a) \text{ für } x}{<} \|v_{S^c}\|_1 + \|v_{S \setminus \tilde{S}}\|_1 = \|v_{\tilde{S}^c}\|_1
 \end{aligned}$$

LINK: VORLESUNGSVIDEO 32 VOM 20.5.25

- (b) Satz 2.24 und 2.27 können modifiziert werden, sodass sie robust (bzgl. y fehlerhaft) und stabil (bzgl. x dünn, x komprimierbar) sind. (ÜA)
- (c) In Bed. (b) ist A_S injektiv vorausgesetzt und Bed. $\forall j \in S : (A^*h)_j = \text{sgn } x_j \Leftrightarrow A_S^*h = \text{sgn } x_S$,

Da A_S inj. existiert ist *Moore-Penrose-Pseudoinverse* von A_S wohldefiniert

$$A_S^\dagger := (A_S^* A_S)^{-1} A_S^* : \mathbb{C}^m \rightarrow \mathbb{C}^S, \quad (A_S^* : \mathbb{C}^m \rightarrow \mathbb{C}^S, \quad A_S^* A_S : \mathbb{C}^S \rightarrow \mathbb{C}^S),$$

$\Rightarrow$ Natürliche Wahl: $h = (A_S^\dagger)^* \text{sgn } x_S$, denn dann

$$\begin{aligned} A_S^* h &= A_S^* ((A_S^* A_S)^{-1} A_S^*)^* \text{sgn } x_S \\ &= A_S^* A_S (A_S^* A_S)^{-1} \text{sgn } x_S. \end{aligned}$$

$$\text{Bed: } \forall l \in S^c : |(A^*h)_l| < 1 \Leftrightarrow \forall l \in S^c : |\langle a_l, h \rangle| < 1.$$

KOROLLAR 2.29 (zu 2.27 / 2.24). Sei $A = (a_1, \dots, a_N) \in \mathbb{C}^{m \times N}$, $x \in \mathbb{C}^N$, $S = \text{supp } x$. Ist A_S injektiv und $|\langle A_S^\dagger a_j, \text{sgn } x_S \rangle| < 1$ für alle $j \in S^c$, so ist x eind. Lsg. von (P_1) mit $y = Ax$.

Frage: Gilt in Satz 2.27 / 2.24 sogar 3er-Äquivalenz? Hängt von Skalarkörper ab.

SATZ 2.30. Seien $A \in \mathbb{R}^{m \times N}$, $x \in \mathbb{R}^N$, $S = \text{supp } x \subset \overline{N}$. Dann sind äquivalent:

- (c) x ist eindeutiger Minimierer von $\min_{z \in \mathbb{R}^N} \|z\|_1$ unter $Az = Ax$
- (a) $\forall v \in \ker A \setminus \{0\} : |\sum_{j \in S} v_j \text{sgn } x_j| < \|v_{S^c}\|_1$
- (b) A_S inj. und es ex. $h \in \mathbb{R}^m$ s.d. $(A^T h)_j = \text{sgn } x_j$ für alle $j \in S$ und $|(A^T h)_l| < 1$ für alle $l \in S^c$

BEWEIS. Der Beweis der Implikationen “(a) $\Rightarrow$ (b) $\Rightarrow$ (c)” folgt genauso wie bei $\mathbb{K} = \mathbb{C}$. Dafür zeigen wir nur “(c) $\Rightarrow$ (a)”. Sei $x \in \mathbb{R}^N$ Originalsignal und $v \in \ker A \setminus \{0\}$. Sei $z := x - v \Rightarrow Az = Ax - Av = Ax$, $z \neq x$.

$$\|x\|_1 < \|z\|_1 = \|z_S\|_1 + \|z_{S^c}\|_1 = \langle z_S, \text{sgn } z_S \rangle + \|z_{S^c}\|_1$$

Da $\|x\|_1 = \langle x, \text{sgn } x_S \rangle \geq \langle x, \text{sgn } z_S \rangle$, gilt $\langle x - z, \text{sgn } z_S \rangle \leq \|x\|_1 - (\|z\|_1 - \|z_{S^c}\|_1) < \|z_{S^c}\|_1$. Substituiere $v = x - z$

$$\langle v, \text{sgn } (x - v)_S \rangle < \|(x - v)_{S^c}\|_1 = \|v_{S^c}\|_1$$

Ersetze $v \rightarrow tv$, $t > 0$.

$$\langle tv, \text{sgn } (x - tv)_S \rangle < \|(x - tv)_{S^c}\|_1 = t\|v_{S^c}\|_1$$

Für t klein genug: $\forall j \in S : \text{sgn } (x - tv)_j = \text{sgn } x_j$ ($S = \text{supp } x$) und damit

$$\langle v, \text{sgn } x_S \rangle < \|v_{S^c}\|_1 \quad (v \in \ker A \setminus \{0\})$$

gilt auch für $-v \Rightarrow$ (a) gilt. □

Falls es nicht einfach ist, den exakten dualen Vektor h zu konstruieren, kann man sich mit einer approximativen Variante behelfen:

LINK: VORLESUNGSVIDEO 33 VOM 20.5.25

SATZ 2.31. Seien $A = (a_1, \dots, a_N) \in \mathbb{C}^{m \times N}$, $x \in \mathbb{C}^N$, $S = \text{supp } x$, $\alpha, \beta, \gamma, \theta \geq 0$ mit $\theta + \alpha\beta\gamma < 1$. Falls

$$\|(A_S^* A_S)^{-1}\|_2 \leq \alpha, \quad \max_{l \in S^c} \|A_S^* a_l\|_2 \leq \beta,$$

und ein Vektor $h \in \mathbb{C}^m$ ex. s.d. für $v = A^* h \in \mathbb{C}^N$: $\|v_S - \text{sgn } x_S\|_2 \leq \gamma$ und $\|v_{S^c}\|_\infty \leq \theta$, so ist x eind. Minim. von (P_1) mit $y = Ax$.

BEWEIS. Wg. Satz 2.24 reicht es Bed. (a) zz:

$$\forall v \in \ker A \setminus \{0\} : |\langle v, \text{sgn } x_S \rangle| < \|v_{S^c}\|_1$$

sei $v \in \ker A$, $u \in \text{Bild } A^* = (\ker A)^\perp$. Dann $\langle u, v \rangle = 0$ und

$$|\langle v, \text{sgn } x_S \rangle| = |\langle v, \text{sgn } x_S - u \rangle| \leq |\langle v_{S^c}, u_{S^c} \rangle| + |\langle v_S, \text{sgn } x_S - u_S \rangle| \leq \theta \|v_{S^c}\|_1 + \gamma \|v_S\|_2$$

$$\begin{aligned} \|v_S\|_2 &= \|(A_S^* A_S)^{-1} (A_S^* A_S) v_S\|_2 \leq \alpha \|A_S^* A_S v_S\|_2 = \alpha \|A_S^* A_{S^c} v_{S^c}\|_2 \\ &= \alpha \left\| \sum_{l \in S^c} v_l A_S^* a_l \right\|_2 \leq \alpha \sum_{l \in S^c} |v_l| \|A_S^* a_l\|_2 \stackrel{\text{Ann.}}{\leq} \alpha \beta \|v_{S^c}\|_1 \end{aligned}$$

$$\Rightarrow |\langle v, \text{sgn } x_S \rangle| \leq \underbrace{(\theta + \alpha\beta\gamma)}_{<1} \|v_{S^c}\|_1 \quad \square$$

Auch hiervon gibt es eine robuste und stabile Variante.

SATZ 2.32. Seien $A = (a_1, \dots, a_N) \in \mathbb{C}^{m \times N}$, $x \in \mathbb{C}^N$, $s \in \overline{N}$, $z \in \mathbb{C}^N$ eine beste s -Term Approximation an x , $T = \text{supp } z$, $y = Ax + e$ mit $\|e\|_2 \leq \eta$. Es seien $\delta \in [0, 1)$, $\beta, \gamma, \theta, \tau \geq 0$, $\rho := \theta + \frac{\beta\gamma}{1-\delta}$. Falls :

- (i) $\rho < 1$,
- (ii) $\|A_T^* A_T - I\|_{2 \rightarrow 2} \leq \delta$ und $\max_{l \in T^c} \|A_T^* a_l\|_2 \leq \beta$,
- (iii) und ein $h \in \mathbb{C}^m$, ex. , so dass mit $u = A^* h \in \mathbb{C}^N$ gilt $\|u\|_2 \leq \tau \sqrt{s}$, $\|u_T - \text{sgn } x_T\| \leq \gamma$, $\|u_{T^c}\|_\infty \leq \theta$,

dann erfüllt jeder Minimierer $x^\#$ von $\|z\|_1$ mit NB $\|Az - y\|_2 \leq \eta$

$$\|x - x^\#\|_2 \leq C_1 \sigma_s(x)_1 + (C_2 + C_3 \sqrt{s}) \eta$$

mit $C_1 = \frac{2}{1-\rho}(1 + \frac{\beta}{1-\rho})$, $C_2 = \frac{2\sqrt{1+\delta}}{1-\delta}(\frac{\gamma}{1-\rho}(1 + \frac{\beta}{1-\delta}) + 1)$, $C_3 = 2\frac{\tau}{1-\rho}(1 + \frac{\beta}{1-\delta})$.

BEMERKUNG 2.33. (a) Bsp-Werte: $\delta = \beta = \gamma = \frac{1}{2}$, $\theta = \frac{1}{4} \Rightarrow \rho = \frac{3}{4}$ und $C_1 \approx 16$, $C_2 = 10 \cdot \sqrt{6} \approx 24,5$, $C_3 = 32$.

- (b) Vgl. mit $\|x^\# - x\|_2 \leq C \frac{1}{\sqrt{s}} \sigma_s(x)_1 + D\eta$ welches unter Bed. $\text{RNR}_{S,2}^{\delta,\tau}$ gilt. Also ist Abschätzung von Satz 2.32 von Faktor $\sqrt{s}$ schlechter. Eher angewandt, wenn RNR schwer zu zeigen ist.
- (c) Vergleich mit vorherigen Resultat zeigt Analogie der Werte:

$$\alpha \rightsquigarrow \frac{1}{1-\delta}$$

Dies liegt an der Neumannschen Reihe, siehe Anhang B: Matrix Analysis

LINK: VORLESUNGSVIDEO 34 VOM 20.5.25

Weiteres geometrische Kriterium für Erfolg der Basisjagd:

DEFINITION 2.34. (a) $K \subset \mathbb{R}^n$ heißt *Kegel* (cone) falls gilt

$$\forall x \in K, t \geq 0 : tx \in K$$

(b) $K \subset \mathbb{R}^n$ heißt *konvexer Kegel* falls K ein konvexer Kegel ist, d.h. $\forall z, x \in K$ und $\forall s, t \geq 0 : sz + tx \in K$.

(c) Zu einem $B \subset \mathbb{R}^n$ heißt $\text{cone}(B) := \{\sum_{j=1}^k t_j x_j \mid k \in \mathbb{N}, t_1, \dots, t_k \geq 0, x_1, \dots, x_k \in B\}$ *konisch konvexe Hülle* von B .

SATZ 2.35. Seien $A \in \mathbb{R}^{m \times N}$, $x \in \mathbb{R}^N$. Dann ist x eind. Minimum von $\|z\|_1$ mit $Az = Ax$

$$\Leftrightarrow \ker A \cap \underbrace{\text{cone}\{z - x \mid z \in \mathbb{R}^N, \|z\|_1 \leq \|x\|_1\}}_{:=\kappa(x)} = \{0\}.$$

LINK: VORLESUNGSVIDEO 35 VOM 26.5.25

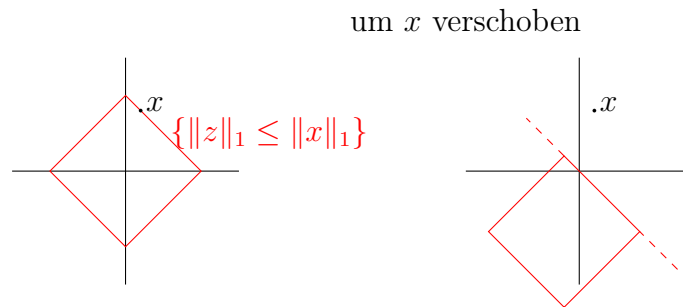


ABBILDUNG 1.2. Hier: $\ker A \cap \kappa(x) = \{0\} \Leftrightarrow \ker A = \{0\}$, da $\kappa(x)$ ein Halbraum ist.

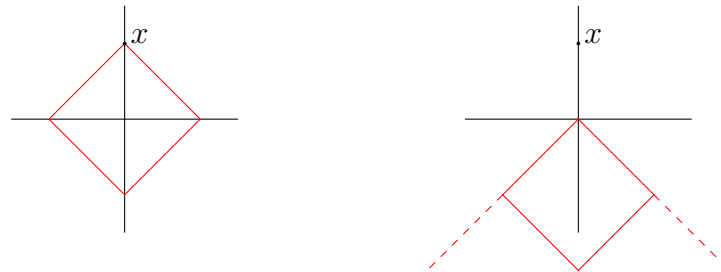


ABBILDUNG 1.3. Hier $\ker A \cap \kappa(x) = \{0\}$. Dann ist $\ker A$ eine nicht zu steile Gerade oder Punkt.

Eine Möglichkeit die Aussage des Satzes zu “glauben” ist:

(P₁) klappt $\Rightarrow A$ hat $NR_S \Rightarrow \|(x - x^\#)_S\|_1 < \|(x - x^\#)_{S^c}\|_1$, falls $x \neq x^\#$,
 $x^\#$ Minim. von (P₁) $\Rightarrow \|x^\#\|_1 \leq \|x\|_1 \Rightarrow \|(x - x^\#)_{S^c}\|_1 \leq \|(x - x^\#)_S\|_1$. Also $x = x^\#$ muss gelten.

BEWEIS. Sei $\ker A \cap \kappa(x) = \{0\}$ und $x^\#$ ℓ_1 -Minimierer
 $\Rightarrow \|x^\#\|_1 \leq \|x\|_1$ und $Ax = Ax^\#$,
 $v := x^\# - x \Rightarrow v \in \ker A, v \in \kappa(x)$.

Nun sei bekannt, dass x eind. Minim. von (P₁), insbesondere $\|z_j\|_1 \leq \|x\|_1$ sobald $Az_j = Ax$.

Jedes $v \in \kappa(x) \setminus \{0\}$ kann nach Def. des konvexen Kegels geschrieben werden als:

$$v = \sum_{j=1}^K t_j(z_j - x), K \in \mathbb{N}, t_j > 0$$

Falls

$$v \in \ker A \Rightarrow A\left(\sum_j t_j z_j\right) = A\left(\sum_j t_j x\right) = Ax \sum_j t_j$$

Dreieckungleichung und x Minimierter implizieren

$$\left\| \sum_j t_j z_j \right\|_1 \leq \sum_j t_j \|z_j\| \leq \|x\|_1 \sum_j t_j \quad \text{neuer Kandidat für Minimierer}$$

Setze $s_j = t_j / \sum_k t_k \Rightarrow$

$$\left\| \sum_j s_j z_j \right\|_1 \leq \|x\|_1$$

Eindeutigkeit impliziert

$$\sum_j s_j z_j = x \text{ [somit } v = \sum_j t_j(z_j - x) = \left(\sum_j t_j\right) \cdot \left(\sum_j s_j(z_j - x)\right) = 0 \Rightarrow \checkmark]$$

□

Kriterium vom Satz umschreiben:

$$\inf_{\substack{v \in \kappa(x) \\ \|v\|_2=1}} \|Av\|_2 > 0$$

Es gibt auch eine robuste Variante der Aussage. Bei den Voraussetzungen spielt eine implizite, lokalisierte Invertierbarkeitseigenschaft eine Rolle, wie bei den früheren Sätzen in diesem Abschnitt.

SATZ 2.36. Seien $A \in \mathbb{R}^{m \times N}$, $x \in \mathbb{R}^N$, $y = Ax + e \in \mathbb{R}^m$, $\|e\|_2 \leq \eta$. Falls ein $\tau > 0$ ex. mit

$$\inf_{\substack{v \in \kappa(x) \\ \|v\|_2=1}} \|Av\|_2 \geq \tau$$

dann erfüllt jedes Minimum $x^\#$ von $\|z\|_1$ mit NB $\|Az - y\|_2 \leq \eta$ die Schranke

$$\|x - x^\#\| \leq \frac{2\eta}{\tau}$$

BEWEIS. $x \neq x^\#$. Dann $\|x^\#\|_1 \leq \|x\|_1$ und $v := \frac{x^\# - x}{\|x^\# - x\|_2} \in \kappa(x)$

$$\|v\|_2 = 1 \Rightarrow \|Av\|_2 \geq \tau \Rightarrow \|x^\# - x\|_2 \tau \leq \|A(x^\# - x)\|_2 \leq \|Ax^\# - y\|_2 + \|Ax - y\|_2 \leq 2\eta.$$

□

KAPITEL 3

Anhang B: Matrix Analysis

1. B.1 Normen

LINK: VORLESUNGSVIDEO 36 VOM 26.5.25

DEFINITION 3.1. Sei $\|\cdot\|$ Norm auf $\mathbb{K}^N$. Dann ist die *Dualnorm* def. durch

$$\|x\|_* := \sup_{\|y\| \leq 1} |\langle y, x \rangle| \quad (x \in \mathbb{K}^N).$$

Die Dualnorm von $\|\cdot\|_p$ ist $\|\cdot\|_q$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Insb. $\|\cdot\|_2$ selbstdual:

$$\|x\|_2 = \sup_{\|y\|_2 \leq 1} |\langle y, x \rangle|.$$

DEFINITION 3.2. Seien $(X, \|\cdot\|)$, $(Y, \|\cdot\|)$ norm. lin. Räume und $A : X \rightarrow Y$ lineare Abbildung. Die *Operatornorm* ist definiert als

$$\|A\| := \sup_{\|x\| \leq 1} \|Ax\| = \sup_{\|x\|=1} \|Ax\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|},$$

falls das sup endlich ist.

Insbesondere falls $\|x\| = (\sum_j |x_j|^p)^{1/p}$ und $\|y\| = (\sum_k |y_k|^q)^{1/q}$ ist

$$\|A\|_{p \rightarrow q} := \sup_{\|x\|_p = 1} \|Ax\|_q.$$

Offensichtlich

$$\|Ax\| \leq \|A\| \|x\|,$$

insbesondere falls $A \in \mathbb{K}^{m \times n}$ und $B \in \mathbb{K}^{n \times k}$,

$$\|AB\|_{p \rightarrow r} \leq \|A\|_{q \rightarrow r} \|B\|_{p \rightarrow q}.$$

Insb. ist $\|\cdot\|_2$ auf $\mathbb{K}^{n \times n}$ submultiplikativ.

LEMMA 3.3. $A \in \mathbb{C}^{m \times n}$. Dann

$$\|A\|_{2 \rightarrow 2} = \sqrt{\lambda_{\max}(A^*A)}.$$

BEWEIS. A^*A ist hermitesch und positiv semidefinit $\Rightarrow$ diagonalisierbar.

$$A^*A = U^*DU, \quad U \text{ unitär}, \quad D = \text{diag}(\lambda_1, \dots, \lambda_n)$$

$$x \in \mathbb{C}^n, \|x\|_2 = 1 \Rightarrow \|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A^*Ax, x \rangle = \langle U^*DUx, x \rangle = \langle D Ux, \underbrace{Ux}_z \rangle.$$

Unitär $\Rightarrow \|Ux\|_2 = \|x\|_2 = 1$. Außerdem

$$\forall z \in \mathbb{C}^n : \langle Dz, z \rangle = \sum_{j=1}^n \lambda_j |z_j|^2 \leq (\max_{j \in \overline{n}} \lambda_j) \sum_j |z_j|^2 = \lambda_{\max}(D) \|z\|_2^2.$$

Es folgt

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{\langle DUx, Ux \rangle}{\|Ux\|_2^2} \leq \lambda_{\max}(D) = \lambda_{\max}(A^*A)$$

da Eigenwerte invariant unter unitären Transformationen. $\square$

LEMMA 3.4. Für $A \in \mathbb{C}^{m \times n}$ gilt:

$$\|A\|_{2 \rightarrow 2} = \sup_{\|y\|_2 \leq 1} \sup_{\|x\|_2 \leq 1} |\langle Ax, y \rangle| = \sup_{\|y\|_2 \leq 1} \sup_{\|x\|_2 \leq 1} \Re(\langle Ax, y \rangle)$$

Insb. für $B \in \mathbb{C}^{n \times n}$ hermitesch $\|B\|_{2 \rightarrow 2} = \sup_{\|x\|_2=1} |\langle Bx, x \rangle|$.

BEWEIS. Übung. $\square$

KOROLLAR 3.5. Für $A \in \mathbb{C}^{m \times n}$ und unitäre $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ gilt:

$$\|UAV\|_{2 \rightarrow 2} = \|A\|_{2 \rightarrow 2}$$

LINK: VORLESUNGSVIDEO 37 VOM 26.5.25

LEMMA 3.6. $B \in \mathbb{C}^{n \times n}$ erfülle $\|B - I\|_{2 \rightarrow 2} \leq \eta$, $\eta \in [0, 1)$

$\Rightarrow B$ invertierbar mit $\|B^{-1}\|_{2 \rightarrow 2} \leq \frac{1}{1-\eta}$.

BEWEIS. $H := I - B$,

$$\underbrace{\left\| \sum_{k=0}^N H^k \right\|}_{\text{Cauchyfolge in } \mathbb{C}^{m \times k}} \leq \sum_{k=0}^N \|H^k\| \stackrel{\text{subm.}}{\leq} \sum_{k=0}^N \|H\|^k \leq \sum_{k=0}^{\infty} \eta^k = \frac{1}{1-\eta} \quad \Rightarrow \quad \sum_{k=0}^{\infty} H^k \text{ existiert.}$$

Also

$$(I - H) \sum_{k=0}^{\infty} H^k = \sum_{k=0}^{\infty} H^k - \sum_{k=1}^{\infty} H^k = H^0 = I.$$

$\square$

SATZ 3.7. Sei λ Eigw. von $A \in \mathbb{C}^{n \times n}$. Dann ex, $j \in \bar{n}$, s.d. $|\lambda - A_{j,j}| \leq \sum_{l \in \bar{n} \setminus \{j\}} |A_{j,l}|$.

BEWEIS. Sei $u \neq 0$ EV zu λ , $Au = \lambda u$, $j \in \bar{n}$ mit $\|u\|_{\infty} = |u_j|$. Dann

$$\sum_{l \in \bar{n}} A_{j,l} u_l = \lambda u_j \Leftrightarrow \sum_{l \in \bar{n} \setminus \{j\}} A_{j,l} u_l = \lambda u_j - A_{j,j} u_j$$

und so

$$|\lambda - A_{j,j}| |u_j| \leq \sum_{l \in \bar{n} \setminus \{j\}} |A_{j,l}| |u_l| \leq \|u\|_{\infty} \sum_{l \neq j} |A_{j,l}| \leq \|u\|_{\infty} \sum_{l \neq j} |A_{j,l}|.$$

$\square$

LINK: VORLESUNGSVIDEO 38 VOM 27.5.25

DEFINITION 3.8. Für $B \in \mathbb{C}^{n \times n}$ ist die *Spur* von B definiert durch

$$\text{tr}(B) := \sum_{j=1}^n B_{j,j}.$$

BEMERKUNG 3.9. Die Spur einer Matrix hat die folgende Eigenschaften.

- kommutativ, d.h. $\text{tr}(AB) = \text{tr}(BA)$, $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times m}$.
- zyklisch, d.h. $\text{tr}(ABC) = \text{tr}(CAB)$, $A, B, C \in \mathbb{C}^{n \times n}$.
- $\text{tr}(A) = \text{tr}(U^*AU)$, falls $U \in \mathbb{C}^{n \times n}$ unitär.
- $\text{tr}(A) = \text{tr}(B^{-1}AB)$, falls $B \in \mathbb{C}^{n \times n}$ invertierbar.
- $(A, B) \mapsto \langle A, B \rangle_F := \text{tr}(AB^*)$ ist Skalarprodukt auf $\mathbb{C}^{m \times n}$.
- Induzierte Norm: *Frobenius- oder Hilbert-Schmidt-Norm*:

$$\|A\|_{\text{HS}} = \|A\|_F := \sqrt{\text{tr}(AA^*)} = \sqrt{\sum_{j \in \bar{m}, k \in \bar{n}} |A_{j,k}|^2}$$

Diese Norm ist also grade die ℓ_2 -Norm auf $\mathbb{C}^{m \cdot n}$.

- Vergleich mit Operatornorm:

$$\|A\|_{2 \rightarrow 2} \leq \|A\|_F$$

denn

$$\|Ax\|_2^2 = \sum_{j=1}^m \left(\sum_{l=1}^m A_{j,l} x_l \right)^2 \leq \sum_{j=1}^m \left(\sum_{l=1}^m |x_l|^2 \right) \left(\sum_{l=1}^m |A_{j,l}|^2 \right) = \|A\|_F^2 \|x\|_2^2$$

2. B.2 Singulärwertzerlegung/ singular value decomposition (SVD)

PROPOSITION 3.10. Sei $A \in \mathbb{C}^{m \times n}$. Dann ex. unitäre Matrizen $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ und eind. bestimmte nicht-negative Zahlen $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(m,n)} \geq 0$, s.d. $A = U \Sigma V^*$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(m,n)}) \in \mathbb{C}^{m \times n}$.

BEMERKUNG 3.11. Die σ_j heißen *Singulärwerte* von A . Schreiben wir $U := (u_1, \dots, u_m)$, $V := (v_1, \dots, v_n)$, wobei u_j, v_j die Spalten von U und V sind, so heißen die Spaltenvektoren u_j *linke singuläre Vektoren* von A , v_j *rechte singuläre Vektoren* von A .

BEWEIS. Da $\mathbb{S}^{n-1} = \{x \in \mathbb{C}^n \mid \|x\|_2 = 1\}$ kompakt, ex. Vektor $v_1 \in \mathbb{S}^{n-1}$, so dass:

$$\|A\|_{2 \rightarrow 2} = \sup_{x \in \mathbb{S}^{n-1}} \|Ax\|_2 = \|Av_1\|_2$$

Setze $\sigma_1 = \|Av_1\|_2$.

Falls $\sigma_1 = 0 \Leftrightarrow A = 0$, dann setze $\forall j : \sigma_j = 0$ und U, V beliebig unitär.

Falls $\sigma_1 > 0$, setze $u_1 := \sigma_1^{-1} Av_1$. Ergänze u_1 und v_1 zu orthonormalbasen

$\Rightarrow (u_1 | \tilde{U}_1) =: U_1 \in \mathbb{C}^{m \times m}$, $(v_1 | \tilde{V}_1) =: V_1 \in \mathbb{C}^{n \times n}$ unitäre Matrizen.

Beachte: Aus $\tilde{U}_1^* Av_1 = \sigma_1 \tilde{U}_1^* u_1 = 0$ folgt

$$A_1 = U_1^* A V_1 = \begin{pmatrix} u_1^* \\ \tilde{U}_1^* \end{pmatrix} A (\tilde{v}_1 | V_1) = \begin{pmatrix} \sigma_1 & b^* \\ 0 & B \end{pmatrix}$$

mit

$$b^* = u_1^* A \tilde{V}_1$$

$$B = \tilde{U}_1^* A \tilde{V}_1 \in \mathbb{C}^{(m-1) \times (n-1)}$$

$$\Rightarrow \|A_1\|_{2 \rightarrow 2} \sqrt{\sigma_1^2 + \|b\|_2^2} \geq \left\| A_1 \begin{pmatrix} \sigma_1 \\ b \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} \sigma_2 + \|b\|_2^2 \\ Bb \end{pmatrix} \right\|_2 \geq \sigma_1^2 + \|b\|_2^2$$

Aus unitärer Invarianz der $2 \rightarrow 2$ -Operatornorm folgt:

$$\Rightarrow \sigma_1 = \|A\|_{2 \rightarrow 2} = \|A_1\|_{2 \rightarrow 2} \geq \sqrt{\sigma_1^2 + \|b\|_2^2} \Rightarrow b = 0$$

$$\Rightarrow A_1 = U_1^* A V_1 = \begin{pmatrix} \sigma_1 & 0 \\ 0 & B \end{pmatrix}$$

Wende dieselbe Prozedur auf $B \in \mathbb{C}^{(m-1) \times (n-1)}$ an. $\square$

LINK: VORLESUNGSVIDEO 39 VOM 27.5.25

BEMERKUNG 3.12. (a) Für größten und kleinsten Singulärwert von A gilt:

$$\begin{aligned} \sigma_{\max}(A) &:= \sigma_1(A) = \|A\|_{2 \rightarrow 2} = \max_{x \in S^{n-1}} \|Ax\|_2 \\ \sigma_{\min}(A) &:= \sigma_l(A) = \min_{x \in S^{n-1}} \|Ax\|_2, \quad l := \min(m, n) \end{aligned}$$

Hat A Rang $= r$, so gilt $\sigma_1 \geq \dots \geq \sigma_r > 0$ und $\sigma_{r+1} = \dots = \sigma_l = 0$.

(b) *Reduzierte/kompakte Singulärwertzerlegung:*

Sei $A = U \Sigma V^*$ die SVD von A mit $\text{rank } A = r$. Setze

$$\begin{aligned} \tilde{\Sigma} &:= \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{C}^{r \times r} \\ \tilde{U} &:= (u_1, \dots, u_r) \in \mathbb{C}^{m \times r} \\ \tilde{V} &:= (v_1, \dots, v_r) \in \mathbb{C}^{n \times r} \\ A &= \tilde{U} \tilde{\Sigma} \tilde{V}^* = \sum_{j=1}^r \sigma_j u_j v_j^* \end{aligned}$$

Dann heißt $A = \tilde{U} \tilde{\Sigma} \tilde{V}^* \in \mathbb{C}^{m \times n}$ reduzierte SVD. Es folgt:

$$\begin{aligned} \underbrace{A^* A}_{\text{hermitesche Matrix}} &= \tilde{V} \tilde{\Sigma} \underbrace{\tilde{U}^* \tilde{U}}_{=I} \tilde{\Sigma} \tilde{V}^* = \underbrace{\tilde{V} \tilde{\Sigma}^2 \tilde{V}^*}_{\text{reduzierte Eigenwertzerlegung}} \in \mathbb{C}^{n \times n} \\ \underbrace{A A^*}_{\text{nicht negativ def.}} &= \tilde{U} \tilde{\Sigma} \underbrace{\tilde{V}^* \tilde{V}}_{=I} \tilde{\Sigma} \tilde{U}^* = \tilde{U} \tilde{\Sigma}^2 \tilde{U}^* \in \mathbb{C}^{m \times m} \end{aligned}$$

$$(c) \sigma_j(A) = \sqrt{\lambda_j(A^* A)} = \sqrt{\lambda_j(A A^*)} \text{ mit } j \in \bar{l},$$

wobei $\lambda_1(A^* A) \geq \lambda_2(A^* A) \geq \dots$

$$\text{Insb: } \|A\|_{2 \rightarrow 2} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^* A)}$$

PROPOSITION 3.13. $A \in \mathbb{C}^{m \times n}$, $m \geq n$ und $\delta \in [0, 1]$. Dann sind äquivalent:

- (a) $\|A^* A - I\|_{2 \rightarrow 2} \leq \delta$
- (b) $\sigma_{\max}(A) \leq \sqrt{1 + \delta}$, $\sigma_{\min}(A) \geq \sqrt{1 - \delta}$

BEWEIS. $\forall j \in \bar{n}$,

$$\begin{aligned} \lambda_j(A^* A) &= \sigma_j^2(A) \\ \lambda_j(A^* A - I) &= \lambda_j(A^* A) - 1 = \sigma_j^2(A) - 1 \end{aligned}$$

$$\begin{aligned} \|A^* A - I\|_{2 \rightarrow 2} &= \sigma_{\max}(A^* A - I) = \max(\lambda_{\max}(A^* A - I), |\lambda_{\min}(A^* A - I)|) \\ &= \max(\sigma_{\max}^2(A) - 1, |\sigma_{\min}^2(A) - 1|) \\ &= \max(\sigma_{\max}^2(A) - 1, 1 - \sigma_{\min}^2(A)) \end{aligned}$$

$\square$

PROPOSITION 3.14. $A, B \in \mathbb{C}^{m \times n}$. Dann:

$$\begin{aligned} |\sigma_{\max}(A) - \sigma_{\max}(B)| &\leq \|A - B\|_{2 \rightarrow 2} \leq \|A - B\|_{HS} \\ |\sigma_{\min}(A) - \sigma_{\min}(B)| &\leq \|A - B\|_{2 \rightarrow 2} \leq \|A - B\|_{HS} \end{aligned}$$

BEWEIS.

$$\begin{aligned} |\sigma_{\max}(A) - \sigma_{\max}(B)| &\leq \left| \|A\|_{2 \rightarrow 2} - \|B\|_{2 \rightarrow 2} \right| \leq \|A - B\|_{2 \rightarrow 2} \\ \sigma_{\min}(A) &= \inf_{\|x\|_2=1} \|Ax\|_2 \leq \inf_{\|x\|_2=1} \left(\|Bx\|_2 + \|(A - B)x\|_2 \right) \leq \sigma_{\min}(B) + \|A - B\|_{2 \rightarrow 2} \end{aligned}$$

□

LINK: VORLESUNGSVIDEO 40 VOM 27.5.25

Dies betraf extremale Singulärwerte. Um “innere” Singulärwerte zu untersuchen, ist variationelle Darstellung nützlich:

$$\sigma_k(A) = \max_{\substack{M \subset \mathbb{C}^m \\ \dim M = k}} \min_{\substack{x \in M \\ \|x\|_2=1}} \|Ax\|_2.$$

Diese Darstellung folgt aus dem Courant-Fischer-Min-Max Theorem für Eigenwerte $\lambda_1(B) \geq \lambda_2(B) \geq \dots \geq \lambda_n(B)$ einer herm. Matrix $B \in \mathbb{C}^{n \times n}$, d.h.

$$\lambda_k(B) = \max_{\substack{M \subset \mathbb{C}^m \\ \dim M = k}} \min_{\substack{x \in M \\ \|x\|_2=1}} \langle Bx, x \rangle$$

Hier eine Art “Dreiecksungleichung für Eigenwerte”.

SATZ 3.15 (Ungleichung Lidskii). Seien $A, B, A + B \in \mathbb{C}^{n \times n}$ hermitesch mit Eigenwerten

$$\begin{aligned} \lambda_1(A) &\geq \dots \geq \lambda_n(A) \\ \lambda_1(B) &\geq \dots \geq \lambda_n(B) \\ \lambda_1(A + B) &\geq \dots \geq \lambda_n(A + B) \end{aligned}$$

Dann gilt für jedes $S \subset \bar{n}$:

$$\sum_{j \in S} \lambda_j(A + B) \leq \sum_{j \in S} \lambda_j(A) + \sum_{k=1}^{|S|} \lambda_k(B).$$

LINK: VORLESUNGSVIDEO 41 VOM 2.6.25

KOROLLAR 3.16. $A, B \in \mathbb{C}^{m \times n}$ mit Singulärwerten $\sigma_1(A) \geq \dots \geq \sigma_l(A)$, $\sigma_1(B) \geq \dots \geq \sigma_l(B)$, $l = \min(m, n)$. Dann:

$$k \in \bar{l}: \quad \sum_{j=1}^k |\sigma_j(A) - \sigma_j(B)| \leq \sum_{j=1}^k \sigma_j(A - B).$$

BEWEIS. Übung. Benutze $S(A), S(B) \in \mathbb{C}^{(m+n) \times (m+n)}$, wobei

$$S(A) = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}, S(B) = \begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix}$$

□

FOLGERUNG 3.17.

- a) $\sum_{j=1}^l \sigma_j(A+B) \leq \sum_{j=1}^l \sigma_j(A) + \sum_{j=1}^l \sigma_j(B), \quad A, B \in \mathbb{C}^{m \times n}, \quad l = \min(m, n).$
- b) $\sum_{j=1}^l \sigma_j(A) = 0 \Leftrightarrow \sigma_{\max}(A) = 0 \Leftrightarrow \|A\|_{2 \rightarrow 2} = 0 \Leftrightarrow A = 0.$
- c) $\sum_{j=1}^l \sigma_j(\lambda A) = |\lambda| \sum_{j=1}^l \sigma_j(A), \quad \lambda \in \mathbb{R}$

Deminach definiert der Ausdruck

$$\|A\|_* := \sum_{j=1}^l \sigma_j(A)$$

eine Norm auf den $m \times n$ -Matrizen, nämlich die *Nuklearnorm*, *Spur-Norm*, oder *Schatten-1-Norm*.

Dies ist ein Spezialfall der Schatten- p -Norm, $p \in [1, \infty]$, welche definiert ist durch:

$$\|A\|_{S_p} := \left[\sum_{j=1}^l \sigma_j(A)^p \right]^{1/p}, \quad A \in \mathbb{C}^{m \times n}$$

Für $p = 2 \Rightarrow$ Frobenius-Norm.

Für $p = \infty \Rightarrow$ Grenzfall

$$\|A\|_{S_\infty} := \max\{\sigma_1(A), \dots, \sigma_l(A)\} = \sigma_{\max}$$

also grade die Operatornorm $\|A\|_{2 \rightarrow 2}$.

DEFINITION 3.18. Habe $A \in \mathbb{C}^{m \times n}$, $\text{rank } A = r$ mit reduzierten SVD

$$A = \tilde{U} \tilde{\Sigma} \tilde{V}^* = \sum_{j=1}^r \sigma_j(A) u_j v_j^*.$$

Dann ist die *Moore-Penrose Pseudoinverse* def. als:

$$A^\dagger := \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^* = \sum_{j=1}^r \sigma_j^{-1}(A) v_j u_j^* = \sum_{j=1}^l \sigma_j^{-1}(A) v_j u_j^* = \cancel{V \Sigma^{-1} U^*}$$

Falls A quadr. und invertierbar, dann $A^\dagger = A^{-1}$ und $\sigma_{\max}(A^\dagger) = \|A^\dagger\|_{2 \rightarrow 2} = \sigma_r^{-1}(A)$.

Ist $A^* A \in \mathbb{C}^{n \times n}$ invert., so $m \geq n$ und :

$$(A^* A)^{-1} A^* = \left(\tilde{V} \tilde{\Sigma}^2 \tilde{V}^* \right)^{-1} \tilde{V} \tilde{\Sigma} \tilde{U}^* = \tilde{V} \tilde{\Sigma}^{-2} \tilde{V}^* \tilde{V} \tilde{\Sigma} \tilde{U}^* = \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^* = A^\dagger$$

Genauso, falls $AA^* \in \mathbb{C}^{m \times m}$ invert. $\Rightarrow A^*(AA^*)^{-1} = A^\dagger$.

Warum interessiert uns dieser Ausdruck? Was ist $A^\dagger A$?

$$A^\dagger A = \left(\sum_{k=1}^r \sigma_k^{-1} v_k u_k^* \right) \left(\sum_{j=1}^r \sigma_j u_j v_j^* \right) = \sum_{k,j=1}^r \sigma_k^{-1} \sigma_j v_k \underbrace{u_k^* u_j}_{=\delta_{k,j}} v_j^* = \sum_{j=1}^r v_j v_j^*$$

also der orthogonale Projektor auf $\text{span}(v_1, \dots, v_n)$. Denn für $x = \sum_{k=1}^n \alpha_k v_k \in \mathbb{C}^n$

$$A^\dagger Ax = \sum_{k=1}^n \alpha_k \left(\sum_{j=1}^r v_j v_j^* \right) v_k = \sum_{j=1}^r \alpha_k v_k$$

HA: Was ist $AA^\dagger$?

3. B.3 Prinzip der kleinsten Quadrate

LINK: VORLESUNGSVIDEO 42 VOM 2.6.25

Relevanz der Moore-Penrose-Pseudoinversen.

PROPOSITION 3.19. Sei $A \in \mathbb{C}^{m \times n}$, $y \in \mathbb{C}^m$, und $M \subset \mathbb{C}^n$ die Menge der Minimierer des Funktionals $\mathbb{C}^n \ni x \mapsto \|Ax - y\|_2$. Das Optimierungsproblem

$$(33) \quad \min_{x \in M} \|x\|_2$$

hat $x^\# = A^\dagger y$ als einzige Lsg.

BEWEIS. SVD:

$$A = U\Sigma V^*, \Sigma = \begin{pmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}, \tilde{\Sigma} = \begin{pmatrix} \sigma_1(A) & & 0 \\ & \ddots & \\ 0 & & \sigma_r(A) \end{pmatrix} \in \mathbb{R}^{r \times r}, r = \text{rank } A$$

Setze

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = V^* x \in \mathbb{C}^n, z_1 \in \mathbb{C}^r, \\ b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = U^* y \in \mathbb{C}^m, b_1 \in \mathbb{C}^r,$$

Dann:

$$\|Ax - y\|_2 = \|U^*(Ax - y)\|_2 = \|\Sigma V^* x - b\|_2 = \left\| \begin{pmatrix} \tilde{\Sigma} z_1 - b_1 \\ -b_2 \end{pmatrix} \right\| =: f(z_1, z_2).$$

Minimierer von $f(z_1, z_2)$ hat die Form $z_1 = \tilde{\Sigma}^{-1} b_1$, z_2 beliebig.

$$M = \left\{ \begin{pmatrix} \tilde{\Sigma}^{-1} b_1 \\ z_2 \end{pmatrix} \mid z_2 \in \mathbb{C}^{n-r} \right\}$$

ist also affiner Teilraum. Betrachte darauf Funktional

$$\|x\|_2^2 = \|V^* x\|_2^2 = \|z\|_2^2 = \underbrace{\|z_1\|_2^2}_{=\tilde{\Sigma}^{-1} b_1 \text{ fixiert}} + \|z_2\|_2^2 =: g(z_2)$$

$z_2 \mapsto g(z_2)$ wird minimal für $z_2 = 0 \in \mathbb{C}^{n-r}$. Also ist das Minimum von (33)

$$x = Vz = V \begin{pmatrix} z_1 \\ 0 \end{pmatrix} = V \begin{pmatrix} \tilde{\Sigma}^{-1} b_1 \\ 0 \end{pmatrix} = V \begin{pmatrix} \tilde{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix} b = V \begin{pmatrix} \tilde{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^* y$$

$$\stackrel{\text{reduzierte SVD}}{=} \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^* y = A^\dagger y$$

□

KOROLLAR 3.20. Sei $A \in \mathbb{C}^{m \times n}$, $m \geq n = \text{rank } A$, $y \in \mathbb{C}^m$. Dann:

$$\min_{x \in \mathbb{C}^m} \|Ax - y\|_2$$

hat $x^\# = A^\dagger y$ als einzige Lsg.

KOROLLAR 3.21. Sei $A \in \mathbb{C}^{m \times n}$, $n \geq m = \text{rank } A$, $y \in \mathbb{C}^m$. Dann hat

$$\min_{x \in \mathbb{K}^N} \|x\|_2 \quad \text{mit NB} \quad Ax = y$$

$x^\# = A^\dagger y$ als einzige Lsg.

BEMERKUNG 3.22 (Verallg. von Korollar 3.21). Betrachte Folge $w_1, \dots, w_n > 0$ von Gewichten. Sei

$$\|x\|_{2,w} := \left(\sum_{j=1}^n |x_j|^2 w_j \right)^{1/2}, \quad \langle x, y \rangle_w := \sum_{j=1}^n x_j \overline{y_j} w_j, \quad x, y \in \mathbb{C}^m$$

eine Norm bzw. ein Skalarprodukt. Betrachte das Optimierungsproblem: Für $A \in \mathbb{C}^{m \times n}$, $y \in \mathbb{C}^m$

$$(34) \quad \min_{z \in \mathbb{C}} \|z\|_{2,w} \quad \text{mit NB} \quad Az = y$$

Setzt man $D = D_w = \text{diag}(w_1, \dots, w_n) \in \mathbb{R}^{n \times n}$, $\xi = D^{1/2}z = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_n})z \in \mathbb{C}^n$, so liefert jeder Minimierer $\xi^\#$ von

$$(35) \quad \min_{\xi \in \mathbb{C}^n} \|\xi\|_2 \quad \text{unter NB} \quad AD^{-1/2}\xi = y$$

eine Lsg. von (34).

Denn: Setze $z^\# = D^{-1/2}\xi^\# = D^{-1/2}(AD^{-1/2})^\dagger y$. Dann gilt: $Az^\# = AD^{-1/2}(AD^{-1/2})^\dagger y \stackrel{\text{HA}}{=} y$ und $\|z^\#\|_{2,w}^2 = \|D^{-1/2}\xi^\#\|_{2,w}^2 = \sum_{j=1}^n |w_j^{-1/2}\xi_j^\#|^2 w_j = \|\xi^\#\|_2^2$. Gäbe es $\zeta \in \mathbb{C}^n$ mit $A\zeta = y$ und $\|\zeta\|_{2,w} < \|z^\#\|_{2,w}$, so folgt: $\|D_w^{1/2}\zeta\|_2 < \|\xi^\#\|_2$ und $AD^{-1/2}(D_w^{1/2}\zeta) = y \Rightarrow \xi^\#$ ist kein Minimierer $\Rightarrow$ Widerspruch.

Insb.: Gilt $n \geq m = r = \text{rank } A$, so

$$z^\# = D^{-1/2}(AD^{-1/2})^*(AD^{-1/2}D^{-1/2}A^*)^{-1}y = D^{-1}A^*(AD^{-1}A^*)^{-1}y = Ay$$

Alternativzugang:

PROPOSITION 3.23. Ein $z^\#$ minimiert $\mathbb{C}^n \ni z \mapsto \|z\|_{2,w}$, mit NB $Az = y \Leftrightarrow Az^\# = y$ und $\forall v \in \ker A: \Re(\langle z^\#, v \rangle_w) = 0$.

BEWEIS. “ $\Leftarrow$ ”. Sei $z^\#$ mit $Az^\# = y$. Ein $z \in \mathbb{C}^n$ erfüllt die NB $Az = y$ genau dann, wenn gilt: $z = z^\# + tv$ mit $t \in \mathbb{R}$, $v \in \ker A$.

$$\Rightarrow \|z^\# + tv\|_{2,w}^2 = \|z^\#\|_{2,w}^2 + t^2\|v\|_{2,w}^2 + 2t\Re(\langle z^\#, v \rangle_w)$$

Also ist $t = 0$ Minimierer von

$$t \mapsto \|z^\# + tv\|_{2,w}$$

Da $v \in \ker A$ bel. ist $z^\#$ Minimierer.

“ $\Rightarrow$ ”. Sei nun $z^\#$ Minimierer. Dann ist für jedes $v \in \ker A$, $t =$ Minimierer von $t \mapsto \|z^\# + tv\|_{2,w}$. Gäbe es ein $v \in \ker A$ mit $\Re(\langle z^\#, v \rangle_w) \neq 0$, so könnte man ein kleines $t \in \mathbb{R}$ mit $t = \Re(\langle z^\#, v \rangle_w) < 0$. Wählen, so dass $\|z^\# + tv\|_{2,w} < \|z^\#\|_{2,w}$, da für kleine t der lineare Term den quadratischen dominiert. $\square$

4. B.4 Rekonstruktion von Matrizen

LINK: VORLESUNGSVIDEO 43 VOM 3.6.25

Bisher haben wir uns mit der Lösung von G-Systemen $Ax = y$, also der Rekonstruktion des Vektors x beschäftigt. In vielen Fällen ist das gesuchte Objekt eine Matrix, z.B. ein mit Pixeln kodierte Bild, die aus möglichst wenigen Messungen rekonstruiert werden soll. Eine mögl. Formulierung: $X \in \mathbb{C}^{n_1 \times n_2}$ hat $\text{Rang} \leq R$ (entspricht Sparsity). Messungen ergeben Vektor $y = \mathcal{A}(X) \in \mathbb{C}^m$, wobei $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^m$ linear. Um zulässige Matrix mit kleinstem Rang zu erhalten, wäre

$$\min_{Z \in \mathbb{C}^{n_1 \times n_2}} \text{rank } Z \quad \text{mit NB } \mathcal{A}(Z) = y$$

natürlichste Aufgabenstellung, ist jedoch NP-schwer. Analog zu (P_0) für Vektoren. Ergibt sich über

$$\text{rank } X = \max\{j | \sigma_j(X) > 0\} = \|(\sigma_1(X), \dots, \sigma_n(X))\|_0, \quad n = \min(n_1, n_2)$$

Wieder in Analogie zur Rekonstruktion von Vektoren schlagen wir stattdessen die Aufgabe

$$(36) \quad \min_{Z \in \mathbb{C}^{n_1 \times n_2}} \|Z\|_* \quad \text{unter } \mathcal{A}(Z) = y \in \mathbb{C}^m$$

vor, wobei:

$$\|Z\|_* = \sum_{j=1}^n \sigma_j(Z) = \text{Spurnorm} = \ell_1\text{-Norm der Singulärwerte}$$

Da $\|\cdot\|_*$ tatsächlich eine Norm ist, ist (36) ein konvexes Optimierungsproblem und äquivalent zu einem semi-definiten Programm.

SATZ 3.24. Sei $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^m$ linear, $n = \min(n_1, n_2) \geq R \geq 0$. Dann gilt: Jedes $X \in \mathbb{C}^{n_1 \times n_2}$ mit $\text{rank } X \leq R$ ist eind. Lsg. von (36) mit $y = \mathcal{A}(X)$ genau dann, wenn $\forall M \in \ker \mathcal{A} \setminus \{0\}$ gilt

$$\sum_{j=1}^R \sigma_j(M) < \sum_{j=R+1}^n \sigma_j(M). \quad \text{cf. NRE}$$

BEWEIS. “ $\Rightarrow$ ” Gegeben: Für beliebiges $X \in \mathbb{C}^{n_1 \times n_2}$ mit $\text{rank } X \leq R$ und

$$X^\# = \arg \min \|Z\|_* \quad \text{mit NB } \mathcal{A}(Z) = y := \mathcal{A}(X)$$

gilt $X^\# = X$.

Sei $M \in \ker \mathcal{A} \setminus \{0\}$ mit SVD:

$$M = \underbrace{U}_{\in \mathbb{C}^{n_1 \times n_1}} \text{diag}(\sigma_1, \dots, \sigma_n) \underbrace{V^*}_{\in \mathbb{C}^{n_2 \times n_2}}$$

Zerlege $M = M_1 - M_2$:

$$M_1 = U \text{diag}(\sigma_1, \dots, \sigma_R, 0, \dots, 0) V^*$$

$$M_2 = U \text{diag}(0, \dots, 0, -\sigma_{R+1}, \dots, -\sigma_n) V^*$$

NB erfüllt wegen $\mathcal{A}(M) = 0 \Leftrightarrow \mathcal{A}(M_1) = \mathcal{A}(M_2)$ und $\text{rank } M_1 \leq R$. Voraussetzung angewandt auf M_1 besagt

$$\sum_{j=1}^R \sigma_j = \|M_1\|_* < \|M_2\|_* = \sum_{j=R+1}^n \sigma_j.$$

also die Behauptung.

“ $\Leftarrow$ ” $\forall M \in \ker \mathcal{A} \setminus \{0\}$ sei bekannt:

$$\sum_{j=1}^R \sigma_j(M) < \sum_{j=R+1}^n \sigma_j(M)$$

Sei $X, Z \in \mathbb{C}^{n_1 \times n_2}$, $\mathcal{A}(X) = \mathcal{A}(Z)$, $Z \neq X$, $\text{rank } X \leq R$. Setze $M := X - Z \in \ker \mathcal{A} \setminus \{0\}$. Wir wissen:

$$\|Z\|_* = \sum_{j=1}^n \sigma_j(X - M) \stackrel{\text{Kor. 3.16}}{\geq} \sum_{j=1}^n \underbrace{|\sigma_j(X) - \sigma_j(M)|}_{\geq \sigma_j(X) - \sigma_j(M)}$$

Jetzt gilt:

$$|\sigma_j(X) - \sigma_j(M)| \geq \sigma_j(X) - \sigma_j(M) \text{ für alle } j = 1, \dots, R$$

und

$$|\sigma_j(X) - \sigma_j(M)| = \sigma_j(M) \text{ für alle } j = R+1, \dots, n$$

Dann liefert die Voraussetzung

$$\|Z\|_* \geq \sum_{j=1}^R (\sigma_j(X) - \sigma_j(M)) + \sum_{j=R+1}^n \sigma_j(M) > \sum_{j=1}^R \sigma_j(X) = \|X\|_*$$

Und damit, wie behauptet, $X = X^\sharp$. □

KAPITEL 4

Kohärenz

LINK: VORLESUNGSVIDEO 44 VOM 3.6.25

1. Definition und Eigenschaften

In diesem Kapitel: $A \in \mathbb{K}^{m \times N}$, $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ mit ℓ_2 -normierten Spalten.

DEFINITION 4.1. (a) Sei A wie oben $A = (a_1, \dots, a_N)$.

$$\mu := \max_{\substack{i,j=1,\dots,n \\ i \neq j}} |\langle a_i, a_j \rangle|$$

heißt *Kohärenz* von A .

(b) Die Funktion

$$\mu_1 : \overline{N-1} \rightarrow \mathbb{R}, \quad \mu_1(s) := \max_{i \in \overline{N}} \max \left\{ \sum_{j \in T} |\langle a_i, a_j \rangle| : T \subset \overline{N}, \#T \leq s, i \notin T \right\}$$

heißt ℓ_1 -Kohärenzfunktion von A .

BEMERKUNG 4.2. (a) $\mu_1(1) = \mu$, $\mu \leq \mu_1(s) \leq s\mu$, $s \in \overline{N-1}$

(b) $\max(\mu_1(s), \mu_1(t)) \leq \mu_1(s+t) \leq \mu_1(s) + \mu_1(t)$ für $1 \leq s, t < t+s \leq N-1$

(c) $U \in \mathbb{K}^{m \times m}$ unitär/orthogonal: $\mu(UA) = \mu(A)$ und μ_1 von Matrix A und UA ist identisch

(d) $|\langle a_i, a_j \rangle| \leq \|a_i\|_2 \cdot \|a_j\|_2 = 1 \Rightarrow 0 \leq \mu(A) \leq 1$

(e) $\mu = 0 \Leftrightarrow a_1, \dots, a_N \in \mathbb{K}^m$ bilden ONS.

Motivation für den Begriff der Kohärenz bzw μ :

Finde quantitative Version:

“ μ fast 0” $\Leftrightarrow$ “ $a_1, \dots, a_N$ fast ONS”

bzw. definiere letzteres damit.

SATZ 4.3. Sei $s \in \overline{N}$. Dann:

$$(37) \quad \forall x \in \mathbb{K}_s^N : [1 - \mu_1(s-1)] \|x\|_2^2 \leq \|Ax\|_2^2 \leq [1 + \mu_1(s-1)] \|x\|_2^2$$

BEMERKUNG 4.4. (37) ist äquivalent zu: alle Eigenw. von $A_T^* A_T$ liegen in

$$(38) \quad [1 - \mu_1(s-1), 1 + \mu_1(s-1)]$$

für jedes $T \subset \overline{N}$, $|T| \leq s$.

Insbesondere:

$$\mu_1(s-1) < 1 \Rightarrow A_T^* A_T \text{ invertierbar für alle } T \subset \bar{N}, \#T \leq s$$

Dabei folgt (37) $\Leftrightarrow$ (38) aus:

$T \subset \bar{N} \Rightarrow A_T^* A_T$ ist pos. semid.

$\Rightarrow$ Eig.w. ≥ 0 , Eig.w. orthogonal.

Für $x \in \mathbb{K}^N$ mit $\text{supp } x \subset T$ ist $Ax = A_T x_T$ insbesondere

$$\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A_T x_T, A_T x_T \rangle = \langle A_T^* A_T x_T, x_T \rangle$$

$$\begin{aligned} \Rightarrow \max_{\|x\|=1, \text{supp } x \subset T} \|Ax\|_2^2 &= \max_{\|x\|=1, \text{supp } x \subset T} \langle A_T^* A_T x_T, x_T \rangle = \lambda_{\max}(A_T^* A_T) \\ \Rightarrow \max_{x \in \mathbb{K}_s^N, \|x\|=1} \|Ax\|_2^2 &= \max_{\substack{T \subset \bar{N} \\ |T|=s}} \lambda_{\max}(A_T^* A_T). \end{aligned}$$

analog:

$$\min_{x \in \mathbb{K}_s^N, \|x\|=1} \|Ax\|_2^2 = \min_{\substack{T \subset \bar{N} \\ |T|=s}} \lambda_{\min}(A_T^* A_T),$$

und so folgt die Äquivalenz.

SATZ 4.5 (Satz von Gershgorin). *Ist $A = (a_{ij})_{ij} \in \mathbb{R}^{n \times n}$ und λ ein Eigenwert von A , so gibt ein $i \in \bar{n}$ so dass*

$$|\lambda - a_{ii}| \leq \sum_{j \in \bar{n} \setminus \{i\}} |a_{ij}|.$$

BEWEIS VON SATZ 4.3. Sei $T \subset \bar{N}$ bel. mit $|T| = s$. Vektoren $a_{i_1}, \dots, a_{i_s}$ sind normalisiert nach Ann.

$$A_T^* A_T = \begin{pmatrix} \overline{a_{i_1}} \\ \vdots \\ \overline{a_{i_s}} \end{pmatrix} \begin{pmatrix} a_{i_1} & \dots & a_{i_s} \end{pmatrix}$$

$\Rightarrow$ Diagonalelemente von $A_T^* A_T$ sind jeweils $= 1$.

Satz von Gershgorin liefert: Für λ Eig.w. von $A_T^* A_T$ gilt:

$$|\lambda - 1| \leq \max_{i \in T} \sum_{j \in T \setminus \{i\}} \underbrace{|(A_T^* A_T)_{i,j}|}_{\leq |a_{i,j}|} \leq \mu_1(s-1).$$

Beobachtung folgt, weil die EW von $A_T^* A_T$ alle reell sind, also im Intervall $[1 - \mu_1(s-1), 1 + \mu_1(s-1)]$ angeordnet. $\square$

LINK: VORLESUNGSVIDEO 45 VOM 10.6.25

KOROLLAR 4.6. *Für $s \in \overline{N-1}$ gelte $\mu_1(s) + \mu_1(s-1) < 1$. Dann:*

$\forall T \subset \bar{N}$ mit $|T| \leq 2s$ ist $A_T^ A_T$ regulär und A_T injektiv.*

BEMERKUNG 4.7. Voraussetzung erfüllt, wenn $\mu < \frac{1}{2s-1}$, denn

$$\mu_1(s) + \mu_1(s-1) \leq s\mu + (s-1)\mu < 1.$$

BEWEIS. $\mu_1(2s-1) < \mu_1(s) + \mu_1(s-1) < 1$.

Sei $T \subset \overline{N}$, $|T| \leq 2s$, dann $\lambda_{\min}(A_T^* A_T) \geq 1 - \mu_1(2s-1) > 0$ wegen (38)

$\Rightarrow A_T^* A_T$ strikt pos. def., insb. regulär.

Sei nun $A_T z = 0 \Rightarrow A_T^* A_T = 0 \xrightarrow{\text{invertierbar}} z = 0 \Rightarrow A_T$ injektiv. $\square$

1.1. Matrizen mit kleiner Kohärenz. Identifiziere sie, da sie leicht lösbare Gleichungen liefern.

DEFINITION 4.8. Seien $a_1, \dots, a_N \in \mathbb{K}^m$ mit $\|a_i\|_2 = 1$. $\{a_1, \dots, a_N\}$ heißt *gleichwinkliges System* (equiangular system) falls gilt

$$\exists c > 0 \quad \forall 1 \leq i < j \leq N \quad : \quad |\langle a_i, a_j \rangle| = c.$$

Für $\lambda > 0$ und $a_1, \dots, a_N \in \mathbb{K}^m$ sind äquivalent:

- (1) $\forall x \in \mathbb{K}^m : \|x\|_2^2 = \lambda \sum_{j=1}^N |\langle x, a_j \rangle|^2$;
- (2) $\forall x \in \mathbb{K}^m : x = \lambda \sum_{j=1}^N \langle x, a_j \rangle a_j$;
- (3) $\underbrace{A}_{=(a_1, \dots, a_N)} A^* = \frac{1}{\lambda} I \in \mathbb{K}^{m \times m}$.

DEFINITION 4.9. (a) Falls ein solches $\lambda > 0$ ex, heißt System $\{a_1, \dots, a_N\}$ *tight frame*.

(b) Erfüllt ein gleichwinkliges System $\{a_1, \dots, a_N\}$ auch (a), so heißt es *equiangular tight frame* (ETF).

Offensichtlich ist eine ONB $a_1, \dots, a_N \in \mathbb{K}^m$ ein tight frame mit $\lambda = 1$. Insofern haben wie eine Abschwächung des Konzeptes ONB gefunden.

BEMERKUNG 4.10 (Beweis der Äquivalenzen (1)–(3) oben). Zu (3) $\Rightarrow$ (2): Es gilt

$$\mathbb{K}^m \ni x = Ix = \lambda A A^* x = \lambda A (\langle x, a_j \rangle)_{j=1}^N = \lambda \sum_{j=1}^N \langle x, a_j \rangle a_j$$

Zu (2) $\Rightarrow$ (1): Es gilt

$$\langle x, x \rangle = \langle \lambda \sum_j \langle x, a_j \rangle a_j, x \rangle = \lambda \sum_j \underbrace{\langle x, a_j \rangle \langle a_j, x \rangle}_{=|\langle x, a_j \rangle|^2}$$

Zu (1) $\Rightarrow$ (3): Seien

$$\left. \begin{array}{l} (x, y) \mapsto \langle x, y \rangle \\ (x, y) \mapsto [x, y] := \lambda \sum_{j=1}^N \langle x, a_j \rangle \langle a_j, y \rangle \end{array} \right\} \text{Bilinearform auf } \mathbb{K}^m$$

(1) $\Rightarrow x \mapsto [x, x]$ pos. definit. $\Rightarrow$ beides sind Skalarprodukte. Da induzierte Normen übereinstimmen $\Rightarrow$ Skalarprodukte auch (Polarisationsformel).

$$\delta_{i,k} = \langle e_i, e_k \rangle = [e_i, e_k] = \lambda \sum_{j=1}^N \langle e_i, a_j \rangle \langle a_j, e_k \rangle = \lambda \sum_{j=1}^N \overline{a_{i,j}} a_{k,j}.$$

SATZ 4.11. Es gilt

(a)

$$\mu \geq \sqrt{\frac{N-m}{(N-1)m}} \quad (\text{Welch-Schranke})$$

(b) $\mu = \sqrt{\frac{N-m}{(N-1)m}} \Leftrightarrow \text{Spalten } a_1, \dots, a_N \text{ von } A \text{ bilden ETF.}$

BEWEIS. Zu (a): $G := A^*A \in \mathbb{K}^{N \times N}$, $G_{i,j} := \overline{\langle a_i, a_j \rangle} = \langle a_j, a_i \rangle$

$$\Rightarrow \operatorname{tr}(G) = \sum_{i=1}^N G_{i,i} = \sum_{i=1}^N \underbrace{\langle a_i, a_i \rangle}_{=1} = N.$$

$$H := AA^* \in \mathbb{K}^{m \times m}$$

$$\Rightarrow \operatorname{tr}(G) = \operatorname{tr}(A^*A) = \operatorname{tr}(AA^*) = \operatorname{tr}(H) = \operatorname{tr}(HI_m) = \langle H, I_m \rangle_F \leq \|H\|_F \|I_m\|_F = \sqrt{m} \sqrt{\operatorname{tr}(HH^*)}$$

$$\begin{aligned} \operatorname{tr}(HH^*) &= \operatorname{tr}(AA^*AA^*) \stackrel{\text{zyklisch}}{=} \operatorname{tr}(A^*AA^*A) = \operatorname{tr}(GG^*) \\ &= \sum_{i,j}^N |G_{i,j}|^2 = \sum_{i,j} |\langle a_i, a_j \rangle|^2 = N + \sum_{i \neq j} |\langle a_i, a_j \rangle|^2. \end{aligned}$$

Kombinieren:

$$\begin{aligned} N = \operatorname{tr}(G) &= \operatorname{tr}(H) \leq \sqrt{m} \sqrt{N + \sum_{\substack{i,j=1 \\ i \neq j}}^N |\langle a_i, a_j \rangle|^2} \\ \Rightarrow N^2 &\leq m(N + \sum_{i \neq j}^N \mu^2) = m(N + N(N-1)\mu^2) \\ \Rightarrow N^2 - mN &= N(N-m) \leq mN(N-1)\mu^2 \\ \Rightarrow \frac{N-m}{m(N-1)} &\leq \mu^2 \stackrel{\mu \geq 0}{\Rightarrow} \sqrt{\frac{N-m}{m(N-1)}} \leq \mu \end{aligned}$$

Zu (b): Gleichheit gilt genau dann wenn

(i) $\langle H, I_m \rangle_F = \|H\|_F \|I_m\|_F \Leftrightarrow$ es ex. $\lambda \in \mathbb{K}$ mit $H = \lambda I_m$
und

(ii) $|\langle a_i, a_j \rangle|^2 = \mu^2$.

Dann: (i) und (ii) $\Rightarrow a_1, \dots, a_N$ bilden ETF. □

LINK: VORLESUNGSVIDEO 46 VOM 10.6.25

SATZ 4.12. $A \in \mathbb{K}^{m \times N}$, ℓ_2 -normierte Spalten, $s < \sqrt{N-1}$. Dann:

$$\mu_1(s) \geq s \sqrt{\frac{N-m}{m(N-1)}}$$

Gleichheit gilt genau dann, wenn $\{a_1, \dots, a_N\}$ ETF.

Frage: Wie groß kann N sein, wenn $\{a_1, \dots, a_N\}$ ETF bilden?

SATZ 4.13. Sei $m > 1, \{a_1, \dots, a_N\} \subset \mathbb{K}^m$ gleichwinkliges System mit $c < 1$. Dann:

$$N \leq \begin{cases} \frac{m(m+1)}{2} & \mathbb{K} = \mathbb{R} \\ m^2 & \mathbb{K} = \mathbb{C} \end{cases}$$

Falls Gleichheit gilt, dann ETF.

LEMMA 4.14. $z \in \mathbb{C}$, $B \in \mathbb{C}^{n \times n}$, $B := zE_n + (1-z)I_n = \begin{pmatrix} 1 & & z \\ & \ddots & \\ z & & 1 \end{pmatrix}$

Dann: B hat $1+(n-1)z$ als einfachen Eigenwert und $(1-z)$ als $(n-1)$ -fachen Eigenwert.

BEWEIS VON LEMMA 4.14.

$$B \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = (1+(n-1)z) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$e_1 - e_2, \dots, e_1 - e_n$ ist EV zu $1-z$, wobei $e_1, \dots, e_n$ Standard-Einheitsvektoren in $\mathbb{R}^n$. $\square$

DEFINITION 4.15. Sei V ein Vektorraum mit Skalarprodukt über $\mathbb{K}$. Wir definieren die Gram-Matrix von $v_1, \dots, v_n \in V$ durch $G_{i,j} = \langle v_i, v_j \rangle$. Ist $X = (v_1, \dots, v_n)$, definieren wir entsprechend $G = \overline{X^* X}$.

LEMMA 4.16. Für $v_1, \dots, v_n \in V$ gilt:

$v_1, \dots, v_n \in V$ linear unabhängig $\Leftrightarrow \det G = 0$.

BEWEIS VON SATZ 4.13. Idee: Wir zeigen

$$N \leq \dim X = \begin{cases} \frac{m(m+1)}{2} & , \mathbb{K} = \mathbb{R} \\ m^2 & , \mathbb{K} = \mathbb{C} \end{cases}$$

wobei

$$X = \begin{cases} \text{Menge aller symm. Matr. aus } \mathbb{R}^{m \times m} & , \mathbb{K} = \mathbb{R} \\ \text{ganz } \mathbb{C}^{m \times m} & , \mathbb{K} = \mathbb{C} \end{cases}$$

$P_i \in X, i \in \overline{N}$ sei orthogonaler Projektor auf $\text{span}\{a_i\}$, d.h.

$$\mathbb{K}^m \ni v \mapsto P_i v = \langle v, a_i \rangle a_i \quad \text{insbesondere } P_i e_k = \langle e_k, a_i \rangle a_i$$

Also orthogonaler Projektor gilt: $P_i^* = P_i = P_i^2$ insbesondere

$$\langle P_i, P_i \rangle_F = \text{tr}(P_i P_i^*) = \text{tr}(P_i) = 1 = \sum_{k=1}^m \langle P_i e_k, e_k \rangle = \sum_{k=1}^m \langle e_k, a_i \rangle \langle a_i, e_k \rangle \sum_{k=1}^m |\langle e_k, a_i \rangle|^2 = \|a_i\|^2 = 1$$

Und für $i \neq j$:

$$\begin{aligned} \langle P_i, P_j \rangle_F &= \text{tr}(P_i P_j^*) = \text{tr}(P_i^* P_j) = \sum_{k=1}^m \langle P_j e_k, P_i e_k \rangle \\ &= \sum_{k=1}^m \langle e_k, a_j \rangle \langle a_j, a_i \rangle \overline{\langle e_k, a_i \rangle} = \overline{\langle a_i, a_j \rangle} \langle a_i, a_j \rangle = |\langle a_i, a_j \rangle|^2 = c^2. \end{aligned}$$

Dann ist X der kleinste UVR von $\mathbb{K}^{m \times m}$, der alle Matrizen vom Typ P_i enthält.

Statten Vektorraum X mit Skalarprodukt $\langle A, B \rangle_F = \text{tr}(AB^*)$ aus.

Betrachte Gram-Matrix der Elemente $P_1, \dots, P_N$ des Vektorraums X

$$B := (\langle P_i, P_j \rangle_F)_{i,j}^N = \begin{pmatrix} 1 & & c^2 \\ & \ddots & \\ c^2 & & 1 \end{pmatrix}$$

Nach obigen Lemma sind $1 + (N - 1)c^2$ und $1 - c^2$ die EW von B , beide ungleich Null wegen $c \in [0, 1)$

$\Rightarrow B$ invertierbar $\Rightarrow (P_1, \dots, P_N)$ lin. unabhängig in X .

Somit $N \leq \dim X$.

Wir betrachten jetzt den Fall der Gleichheit $N = \dim X$. Dann sind $(I, P_1, \dots, P_N)$ linear abhängig im VR X . Setze $A_0 = I, A_1 = P_1, \dots, A_N = P_N$, dann ist entsprechende Gram-Matrix

$$C = (\langle A_i, A_j \rangle_F)_{i,j} = \begin{pmatrix} m & 1 & \dots & 1 \\ 1 & 1 & & c^2 \\ \vdots & & \ddots & \\ 1 & c^2 & & 1 \end{pmatrix}$$

singulär. Also

$$\begin{aligned} 0 = \det C &= \det \begin{pmatrix} m & 1 & \dots & 1 \\ 0 & 1 - 1/m & & c^2 - 1/m \\ \vdots & & \ddots & \\ 0 & c^2 - 1/m & & 1 - 1/m \end{pmatrix} = m \cdot \det \begin{pmatrix} 1 - 1/m & & & c^2 - 1/m \\ & \ddots & & \\ c^2 - 1/m & & & 1 - 1/m \end{pmatrix} \\ &= m \cdot \left(1 - \frac{1}{m}\right)^N \det \underbrace{\begin{pmatrix} 1 & & b \\ & \ddots & \\ b & & 1 \end{pmatrix}}_{\text{hat EW } 1-b \text{ und } 1+(N-1)b}, \quad b = \frac{mc^2 - 1}{m - 1} \end{aligned}$$

Wegen $\det C = 0$ muss (mind) einer der beiden EW verschwinden. Welcher?

$$0 = 1 - b = \frac{m - 1 + 1 - mc^2}{m - 1} \Leftrightarrow m = mc^2 \quad \text{Der nicht!}$$

$$\begin{aligned} 0 = 1 + (N - 1)b &= \frac{m - 1 + (N - 1)(mc^2 - 1)}{m - 1} \Leftrightarrow 0 = m - 1 + (N - 1)(mc^2 - 1) \\ &\Leftrightarrow -(N - 1)mc^2 = m - N \Leftrightarrow c^2 = \frac{N - m}{(N - 1)m} \end{aligned}$$

Also $\max_{\substack{i,j=1,\dots,n \\ i \neq j}} |\langle a_i, a_j \rangle| = \mu = \sqrt{\frac{N-m}{(N-1)m}}$. Damit impliziert Satz 4.11: $\{a_1, \dots, a_N\}$ ist ETF. □

LINK: VORLESUNGSVIDEO 47 VOM 16.6.25

BEISPIEL 4.17. Ist die obere Schranke $N \leq \dim X$ scharf, oder kann sie verbessert werden? Beispiele, wo in der Schranke Gleichheit eintritt.

(a) $m = 3, N = 6 = \frac{3 \cdot 4}{2}, c := \frac{\sqrt{5}-1}{2}$

Dann bilde normierte Spalten von

$$\begin{pmatrix} 1 & 0 & c & 1 & 0 & -c \\ c & 1 & 0 & -c & 1 & 0 \\ 0 & c & 1 & 0 & -c & 1 \end{pmatrix}$$

gleichwinkliges System in $\mathbb{R}^3$. (Beachte: Spalten ergeben sich durch zyklische Verschiebung.)

(b) $m = 7, N = 28 = \frac{7 \cdot 8}{2}$

Vorzeichenflips ergeben vier Vektoren

$$(1, \pm 1, 0, \pm 1, 0, 0, 0)^T \in \mathbb{R}^7$$

zyklische Verschiebung ergeben 28 Vektoren. Normierung ergibt gleichwinkliges System in $\mathbb{R}^7$.

(c) $m = 2, N = 4 = 2^2, c := e^{i\pi/4} \sqrt{2 - \sqrt{3}}$

$$\begin{pmatrix} 1 & c & 1 & -c \\ c & 1 & -c & 1 \end{pmatrix}$$

bilden gleichwinkliges System in $\mathbb{C}^2$.

Falls $\mathbb{K} = \mathbb{R}$ wird aber die Schranke $N \leq \frac{m(m+1)}{2}$ nicht in allen Dimensionen angenommen. Denn es gilt folgende zahlentheoretische Bedingung:

SATZ 4.18. *Sei $m \geq 3$. Gibt es ein gleichwinkliges System von $N = \frac{m(m+1)}{2}$ Vektoren im $\mathbb{R}^m$, so gilt*

$$m + 2 = (2k + 1)^2$$

für ein $k \in \mathbb{N}$.

Dagegen wird vermutet, dass für jedes $m \in \mathbb{N}$ ein gleichwinkliges System aus $N = m^2$ Vektoren in $\mathbb{C}^m$ existiert, s.d. die erzeugte Matrix $A \in \mathbb{C}^{m \times m^2}$ Kohärenz

$$\mu = \frac{1}{\sqrt{m+1}}$$

hat.

SATZ 4.19. *Sei $m \geq 5$ Primzahl. Dann kann man ein $A \in \mathbb{C}^{m \times m^2}$ konstruieren mit*

$$\mu = \frac{1}{\sqrt{m}}$$

2. Kohärenz und Basisjagd

SATZ 4.20. *$A \in \mathbb{K}^{m \times N}$ mit ℓ_2 -normierte Spalten. Angenommen*

$$\mu_1(s) + \mu_1(s-1) < 1$$

Dann kann jeder Vektor $x \in \mathbb{K}_s^N$ via Basisjagd rekonstruiert werden.

Erinnerung: Basisjagd hisst ℓ_1 -Minimierung (P_1).

BEWEIS. Wir wollen zeigen, dass A die NR_s hat.

(Also $S \in \overline{N}, |S| = s, \|v_S\|_1 < \|v_{S^c}\|_1 \forall v \in \ker A \setminus \{0\}$)

$a_1, \dots, a_N$ ℓ_2 -normiert. , $v \in \ker A \setminus \{0\}, i \in S$

$$v_i = v_i \langle a_i, a_i \rangle \stackrel{Av=0}{=} - \sum_{\substack{j=1 \\ j \neq i}}^N v_j \langle a_j, a_i \rangle = - \sum_{j \in S^c} v_j \langle a_j, a_i \rangle - \sum_{\substack{j \in S \\ j \neq i}} v_j \langle a_j, a_i \rangle$$

Dreieckungleichung:

$$|v_i| \leq \sum_{j \in S^c} |v_j| |\langle a_j, a_i \rangle| + \sum_{\substack{j \in S \\ j \neq i}} |v_j| |\langle a_j, a_i \rangle|$$

Aufsummieren:

$$\begin{aligned} \|v_S\|_1 &= \sum_{i \in S} |v_i| \leq \sum_{i \in S} \sum_{j \in S^c} |v_j| |\langle a_j, a_i \rangle| + \sum_{i \in S} \sum_{\substack{j \in S \\ j \neq i}} |v_j| |\langle a_j, a_i \rangle| \\ &= \underbrace{\sum_{j \in S^c} |v_j|}_{=\|v_{S^c}\|_1} \underbrace{\left(\sum_{i \in S} |\langle a_j, a_i \rangle| \right)}_{\leq \mu_1(s)} + \underbrace{\sum_{j \in S} |v_j|}_{=\|v_S\|_1} \underbrace{\left(\sum_{\substack{i \in S \\ i \neq j}} |\langle a_j, a_i \rangle| \right)}_{\leq \mu_1(s-1)} \\ \Rightarrow \|v_S\|_1 &\leq \mu_1(s-1) \|v_S\|_1 + \mu_1(s) \|v_{S^c}\|_1 \end{aligned}$$

Falls $\mu_1(s) = 0$, so $\|v_S\|_1 = 0$ aber $\|v_{S^c}\|_1 \neq 0$, für $v \neq 0$.

Umstellen und Annahme vom Satz ergeben:

$$\mu_1(s) \|v\|_1 < (1 - \mu_1(s-1)) \|v_S\|_1 \leq \mu_1(s) \|v_{S^c}\|_1$$

Also $\|v_S\|_1 < \|v_{S^c}\|_1 \forall v \in \ker A \setminus \{0\}$ wie erwünscht. $\square$

BEMERKUNG 4.21. (1) Wegen $\mu_1(s) + \mu_1(s-1) \leq s\mu + (s-1)\mu = (2s-1)\mu$, ist Satz 4.20 anwendbar, falls

$$(39) \quad (2s-1)\mu < 1$$

(2) Bei einer Messmatrix $A \in \mathbb{C}^{m \times N}$, die $\mu \leq \frac{c}{\sqrt{m}}$ erfüllt (wie z.B. solchen in Satz 4.19) so reichen

$$m \geq C \cdot s^2, \quad C \geq 4c^2$$

Messungen, um

$$(2s-1)\mu \leq \frac{2s-1}{\sqrt{m}} c \leq \frac{2s-1}{s} \frac{c}{\sqrt{C}} \leq \frac{2s-1}{s} < 1$$

zu gewährleisten und damit die Anwendbarkeit von Satz 4.20.

(3) Der bisherige Stand der Forschung zu diese Methode erlaubt es nicht, allgemeine Rekonstruktionskriterien zu etablieren bei denen die Anzahl von Messungen m in s sub-quadratisch wächst.

Illustration dazu Seien folgende Relationen erfüllt:

$$\mu_1(s) + \mu_1(s-1) < 1, \quad m \leq \frac{(2s-1)^2}{2}, \quad 2m \leq N,$$

Dann folgt mit verbesserter Welch-Schranke 4.12

$$\begin{aligned} 1 > \mu_1(s) + \mu_1(s-1) &\geq (2s-1) \sqrt{\frac{N-m}{(N-1)m}} = \sqrt{(2s-1)^2} \sqrt{\frac{N-m}{(N-1)m}} \\ &\geq \sqrt{2 \frac{N-m}{N-1}} \geq \sqrt{\frac{N}{N-1}} > 1 \end{aligned}$$

also Widerspruch.

Da wir bereits früher die Intuition entwickelt haben, dass eine linear in s wachsende Anzahl m von Messungen ausreichend für die Rekonstruktion mittels Basisjagd sein sollte, ist die vorliegende Methode unbefriedigend.

Die nun folgende Verallgemeinerung, soll dazu dienen die Anzahl m nur linear wachsend in s wählen zu können.

Restricted Isometry Property (RIP)

LINK: VORLESUNGSVIDEO 48 VOM 16.6.25

1. Definition und Eigenschaften

DEFINITION 5.1. $A \in \mathbb{K}^{m \times N}$, $s \in \overline{N}$, $\delta_s(A)$ ist die *RIP-Konstante der Ordnung s* , wenn δ die kleinste Konstante ist, so dass:

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2 \quad \forall x \in \mathbb{K}_s^N$$

Ähnlich wie in Bemerkung 4.4 zu Satz 4.3 sehen wir, dass die obige Definition äquivalent zu

$$\delta_s(A) := \max_{\substack{S \subset \overline{N} \\ |S| \leq s}} \|A_S^* A_S - I_S\|_{2 \rightarrow 2}.$$

Denn: Wir haben $|\|Ax\|_2^2 - \|x\|_2^2| \leq \delta\|x\|_2^2$. Es gilt

$$\begin{aligned} \|Ax\|_2^2 - \|x\|_2^2 &= \langle Ax, Ax \rangle - \langle x, x \rangle \stackrel{x \text{ s-sparse}}{=} \langle A_S x, A_S x \rangle - \langle x, x \rangle \\ &= \langle A_S^* A_S x, x \rangle - \langle x, x \rangle = \langle (A_S^* A_S - I_S)x, x \rangle \end{aligned}$$

$$\Rightarrow \|A_S^* A_S - I_S\|_{2 \rightarrow 2} = \max_{x \neq 0} \frac{|\langle (A_S^* A_S - I_S)x, x \rangle|}{\|x\|_2^2} \leq \delta$$

PROPOSITION 5.2. (1) $\delta_1 \leq \delta_2 \leq \dots \leq \delta_s \leq \dots \leq \delta_N$

(2) *RIP bedeutet, dass für $|S| \leq s$ alle Eigenwerte von $A_S^* A_S$ in $[1 - \delta_s, 1 + \delta_s]$ liegen. Insbesondere, falls $\delta_s < 1$, dann ist A_S injektiv.*

BEWEIS. Zu (1): Folgt aus: $\mathbb{K}_s^N \subset \mathbb{K}_{s+1}^N$.

Zu (2): Ahme Beweis von Bemerkung 4.4 nach.

Falls $\delta_s < 1$, dann ist $A_S^* A_S$ positiv definit und A_S injektiv. □

SATZ 5.3. $A \in \mathbb{C}^{m \times N}$ mit ℓ_2 - normiert Spalten $a_1, \dots, a_N$. Dann:

- (a) $\delta_1 = 0$
- (b) $\delta_2 = \mu$
- (c) $\delta_s \leq \mu_1(s-1) \leq (s-1)\mu$, $s \geq 2$

BEWEIS. (a) $x \in \mathbb{C}^N$ 1-sparse.

$$|\|Ax\|_2^2 - \|x\|_2^2| \stackrel{?}{=}$$

Ohne Einschränkung betrachte $e_i \in \mathbb{C}^N$ i -te Einheitsvektor: $\|Ae_i\|_2^2 = \|a_i\|_2^2 = 1 = \|e_i\|_2^2$. Damit: $\delta_1 = 0$.

(b) Sei $A_{\{i,j\}} := (a_i, a_j)$.

$$(40) \quad \delta_2 = \max \|A_{\{i,j\}}^* A_{\{i,j\}} - I_2\|_{2 \rightarrow 2}$$

$$A_{\{i,j\}}^* A_{\{i,j\}} - I_2 = \begin{pmatrix} \overline{a_i} \\ \overline{a_j} \end{pmatrix} (a_i, a_j) - I_2 = \begin{pmatrix} 1 & \langle a_j, a_i \rangle \\ \langle a_i, a_j \rangle & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & \langle a_j, a_i \rangle \\ \langle a_i, a_j \rangle & 0 \end{pmatrix}}_{\text{EW} \pm |\langle a_j, a_i \rangle|}$$

$$\Rightarrow (40) = \max_{1 \leq i \neq j \leq N} |\langle a_i, a_j \rangle| = \mu(A)$$

(c) Nach Satz 4.3 und Bem. 4.4 liegen alle Eigenwerte von $A_S^* A_S$ in

$$[1 - \mu_1(s-1), 1 + \mu_1(s-1)]$$

Also:

$$(41) \quad [1 - \delta_s, 1 + \delta_s] \subset [1 - \mu_1(s-1), 1 + \mu_1(s-1)]$$

damit $\delta_s \leq \mu_1(s-1) \leq (s-1)\mu$.

□

2. Vergleich mit Orthogonalität

LEMMA 5.4. $u, v \in \mathbb{C}^N$, $\|u\|_0 \leq s, \|v\|_0 \leq t$, $\text{supp } u \cap \text{supp } v = \emptyset$. Dann:

$$|\langle Au, Av \rangle| \leq \delta_{s+t} \|u\|_2 \|v\|_2$$

BEWEIS. $\text{supp } u \cap \text{supp } v = \emptyset \Rightarrow \langle u_S, v_S \rangle = 0$, wobei $S := (\text{supp } u) \cup (\text{supp } v)$. Damit $\|u + v\|_0 \leq s + t$ und $|S| \leq s + t$.

$$\begin{aligned} |\langle Au, Av \rangle| &= |\langle A_S u_S, A_S v_S \rangle| \\ &= |\langle A_S u_S, A_S v_S \rangle - \langle u_S, v_S \rangle| \\ &= |\langle (A_S^* A_S - I_S) u_S, v_S \rangle| \\ &\leq \|(A_S^* A_S - I_S) u_S\|_2 \|v_S\|_2 \\ &\leq \|A_S^* A_S - I_S\|_{2 \rightarrow 2} \|u_S\|_2 \|v_S\|_2 \\ &\leq \delta_{s+t} \underbrace{\|u_S\|_2}_{=\|u\|_2} \underbrace{\|v_S\|_2}_{=\|v\|_2} \end{aligned}$$

□

DEFINITION 5.5. $A \in \mathbb{C}^{m \times N}$, $\theta_{s,t}(A)$ heißt (s, t) -restricted orthogonality constant und ist das kleinste θ , sodass

$\forall u \in \mathbb{C}^N$ s -sparse, $\forall v \in \mathbb{C}^N$ t -sparse mit $\text{supp } u \cap \text{supp } v = \emptyset$:

$$|\langle Au, Av \rangle| \leq \theta \|u\|_2 \|v\|_2$$

Äquivalent: $\theta_{s,t} = \max\{\|A_T^* A_S\|_{2 \rightarrow 2} \mid S, T \subset \overline{N}, |S| \leq s, |T| \leq t, S \cap T = \emptyset\}$

SATZ 5.6 (Vergleich zwischen $\theta_{s,t}$ und $\delta_{s,t}$). Es gilt

$$\theta_{s,t} \leq \delta_{s+t} \leq \frac{1}{s+t} (s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t}).$$

Falls $t = s$, dann $\theta_{s,s} \leq \delta_{2s} \leq \delta_s + \theta_{s,s}$.

BEWEIS. Erste Ungleichung folgt aus Lemma 5.4. Für die zweite Ungleichung müssen wir zeigen:

$$\left| \|Ax\|_2^2 - \|x\|_2^2 \right| \leq \frac{1}{s+t} (s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t}) \quad \forall x \in \mathbb{K}^N, \text{ sparse und mit } \|x\|_2 = 1.$$

LINK: VORLESUNGSVIDEO 49 VOM 17.6.25

Zerlege normiertes $x = u + v$ in $u \in \mathbb{K}_s^N, v \in \mathbb{K}_t^N$ mit $\text{supp } v \cap \text{supp } u = \emptyset$. Insbesondere $u \perp v$ und

$$1 = \|x\|_2^2 = \|u\|_2^2 + \|v\|_2^2$$

$$\begin{aligned} \|Ax\|_2^2 &= \langle A(u+v), A(u+v) \rangle \\ &= \|Au\|_2^2 + \|Av\|_2^2 + 2\Re(\langle Au, Av \rangle) \end{aligned}$$

Damit

$$\begin{aligned} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| &\leq \left| \|Au\|_2^2 - \|u\|_2^2 \right| + \left| \|Av\|_2^2 - \|v\|_2^2 \right| + 2|\langle Au, Av \rangle| \\ &\leq \delta_s \|u\|_2^2 + \delta_t \|v\|_2^2 + 2\theta_{s,t} \|u\|_2 \|v\|_2 \\ &= \delta_s \alpha + \delta_t (1 - \alpha) + 2\theta_{s,t} \sqrt{\alpha(1 - \alpha)} := f(\alpha), \quad \alpha := \|u\|_2^2 \in [0, 1] \end{aligned}$$

Kurvendiskussion ergibt: f ist unimodal:

Es existiert Maximalstelle α^* von f in $[0, 1]$ mit:

f ist isoton/nichtfallend auf $[0, \alpha^*]$ und f ist antiton/nichtwachsend auf $[\alpha^*, 1]$

Nun machen wir eine Fallunterscheidung

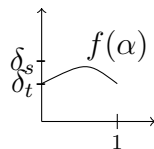


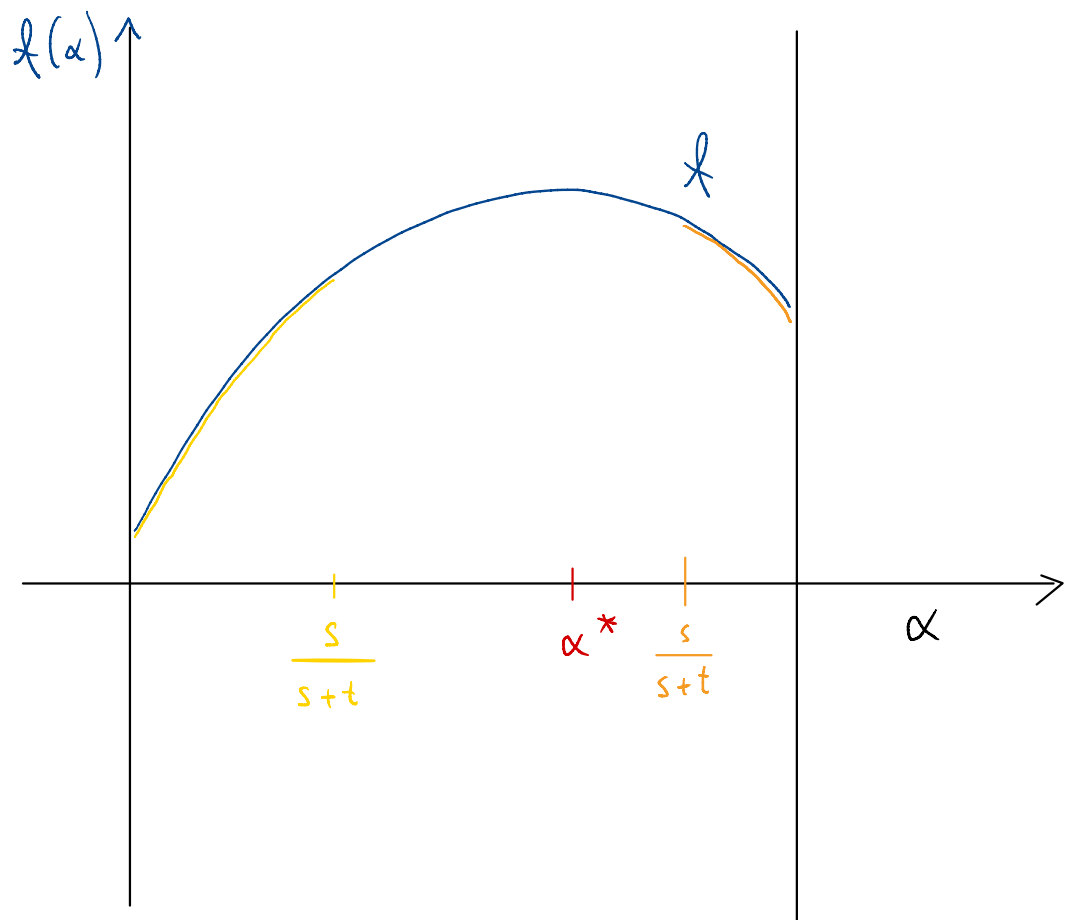
ABBILDUNG 2.1. Funktion $f(\alpha)$

Kurvendiskussion ergibt:

$f: [0, 1] \rightarrow \mathbb{R}$ ist unimodal,

d.h. es ex. ein $\alpha^* \in [0, 1]$ so dass

- f isotom / nicht fallend auf $[0, \alpha^*]$
- f antitom / nicht wachsend auf $[\alpha^*, 1]$



① Fall: $\alpha^* \leq \frac{s}{s+t} \Rightarrow f$ antitom auf $[\frac{s}{s+t}, 1]$

② Fall: $\frac{s}{s+t} \leq \alpha^* \Rightarrow f$ isotom auf $[0, \frac{s}{s+t}]$

- (1) Falls $\alpha^* \leq \frac{s}{s+t}$, insbesondere f antiton auf $[\frac{s}{s+t}, 1]$

Wähle u und v so, sodass

- u die s größten Einträge von x enthält,
- v die t kleinsten Einträge $\neq 0$ enthält.

Dann gilt:

$$\frac{\|u\|_2^2}{s} = \frac{|u_1|^2 + \dots + |u_s|^2}{s} \geq |u_s|^2 \geq \frac{t}{t} |v_{\max}|^2 \geq \frac{\|v\|_2^2}{t} = \frac{1 - \|u\|_2^2}{t}$$

$$\Rightarrow \frac{\|u\|_2^2}{s} + \frac{\|u\|_2^2}{t} \geq 1/t \Rightarrow \|u\|_2^2 \geq \frac{s}{s+t} \text{ insbesondere}$$

$$f(\|u\|_2^2) \leq f\left(\frac{s}{s+t}\right)$$

- (2) Falls $\frac{s}{s+t} \leq \alpha^*$, insbesondere f isoton auf $[0, \frac{s}{s+t}]$

Wähle u und v so, sodass

- u die s kleinsten Einträge $\neq 0$ von x enthält,
- v die t größten Einträge enthält.

$$\Rightarrow \|u\|_2^2 \leq \frac{s}{s+t} \text{ insbesondere}$$

$$f(\|u\|_2^2) \leq f\left(\frac{s}{s+t}\right)$$

In beiden Fällen können wir also $f(\alpha)$ durch $f(\frac{s}{s+t})$ abschätzen. Damit

$$\Rightarrow \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \leq f(\alpha) \leq f\left(\frac{s}{s+t}\right) = \frac{1}{s+t} (s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t})$$

□

Wie ist das Verhältnis von RIP-Konstanten δ_s und der restringierte Orthogonalitätskonstanten $\Theta_{s,t}$ bei unterschiedlichen Dimensionen s zueinander?

PROPOSITION 5.7. Für $r, s, t \in \mathbb{N}$ mit $t \geq s$ gilt

- (a) $\theta_{t,r} \leq \sqrt{\frac{t}{s}} \theta_{s,r}$
 (b) $\delta \leq \frac{t-d}{s} \delta_{2s} + \frac{d}{s} \delta_s$, $d = ggT(s, t)$.

Spezialfall: $t = cs$, $\delta_{cs} \leq c\delta_{2s}$

BEWEIS. Seien $u \in \mathbb{C}_t^N, v \in \mathbb{C}_r^N$, $\text{supp } u \cap \text{supp } v = \emptyset$.

Z.z:

- (a) $|\langle Au, Av \rangle| \leq \sqrt{\frac{t}{s}} \theta_{s,r} \|u\|_2 \|v\|_2$
 (b) $\left| \|Au\|_2^2 - \|u\|_2^2 \right| \leq \left[\frac{t-d}{s} \delta_{2s} + \frac{d}{s} \delta_s \right] \|u\|_2^2$

Sei d ein gem. Teiler von s und t . Dann ex. $n, k \in \mathbb{N}$, sd $s = kd$, $t = nd$. Bezeichne $T = \{j_1, \dots, j_t\} \supset \text{supp } u$. Definiere $S_1, \dots, S_n \subset T$ durch:

$$S_i = \{j_{(i-1)d+1}, \dots, j_{(i-1)d+s} \pmod{t}\}$$

$$|T| = t, |S_i| = s, \sum_{i=1}^n |S_i| = ns = nkd = kt$$

$\Rightarrow$ gleichmäßige k -Überlagerung.

$\Rightarrow$ jedes $j \in T$ gehört zu genau k Mengen $S_i \Rightarrow$

$$(42) \quad u = \frac{1}{k} \sum_{i=1}^n u_{S_i} \text{ und } \|u\|_2^2 = \frac{1}{k} \sum_{i=1}^n \|u_{S_i}\|_2^2$$

Daher:

$$\begin{aligned} |\langle Au, Av \rangle| &\leq \frac{1}{k} \sum_{i=1}^n |\langle \underbrace{A}_{\in \mathbb{C}_s^N} u_{S_i}, \underbrace{A}_{\in \mathbb{C}_r^N} v \rangle| \leq \frac{1}{k} \sum_{i=1}^n \theta_{s,r} \|u_{S_i}\|_2 \cdot 1 \cdot \|v\|_2 \\ \text{Cauchy-Schwarz} \quad &\leq \frac{\theta_{s,r}}{k} \|v\|_2 \sqrt{n} \left(\sum_{i=1}^n \|u_{S_i}\|_2^2 \right)^{1/2} \stackrel{(42)}{\leq} \theta_{s,r} \sqrt{\frac{n}{k}} \|v\|_2 \|u\|_2 \end{aligned}$$

Also folgt Teil (a)

Nun zu (b):

$$\begin{aligned} |\|Au\|_2^2 - \|u\|_2^2| &= |\langle (A^*A - I)u, u \rangle| \\ &\leq \frac{1}{k^2} \sum_{i,j=1}^n |\langle (A^*A - I)u_{S_i}, u_{S_j} \rangle| \\ &= \frac{1}{k^2} \left[\sum_{\substack{i,j=1 \\ i \neq j}}^n |\langle (A_{S_i \cup S_j}^* A_{S_i \cup S_j} - I)u_{S_i}, u_{S_j} \rangle| + \sum_{i=1}^n |\langle (A_{S_i}^* A_{S_i} - I)u_{S_i}, u_{S_i} \rangle| \right] \\ |\langle Bx, y \rangle| &\leq \|B\| \cdot \|x\| \cdot \|y\| \leq \frac{1}{k^2} \left[\sum_{\substack{i,j=1 \\ i \neq j}}^n \delta_{2s} \|u_{S_i}\|_2 \|u_{S_j}\|_2 + \sum_{i=1}^n \delta_s \|u_{S_i}\|_2^2 \right] \\ &\quad + \frac{1}{k^2} \left[\sum_{i=1}^n \delta_{2s} \|u_{S_i}\|_2 \|u_{S_i}\|_2 - \sum_{i=1}^n \delta_{2s} \|u_{S_i}\|_2 \|u_{S_i}\|_2 \right] \\ &= \frac{1}{k^2} \left[\delta_{2s} \left(\sum_{i=1}^n 1 \|u_{S_i}\|_2 \right)^2 - (\delta_{2s} - \delta_s) \sum_{i=1}^n \|u_{S_i}\|_2^2 \right] \\ \text{Cauchy-Schwarz} \quad &\leq \frac{1}{k^2} \left[\delta_{2s} n \sum_{i=1}^n \|u_{S_i}\|_2^2 - (\delta_{2s} - \delta_s) \sum_{i=1}^n \|u_{S_i}\|_2^2 \right] \\ &= (n\delta_{2s} - (\delta_{2s} - \delta_s)) \underbrace{\frac{1}{k^2} \sum_{i=1}^n \|u_{S_i}\|_2^2}_{\frac{1}{k} \cdot \|u\|_2^2} \\ &\stackrel{(42)}{=} \left(\frac{n}{k} \delta_{2s} - \frac{1}{k} (\delta_{2s} - \delta_s) \right) \|u\|_2 \\ &= \left(\delta_{2s} \frac{t}{s} - \frac{d}{s} (\delta_{2s} - \delta_s) \right) \|u\|_2^2 \end{aligned}$$

wegen

$$\frac{n}{k} = \frac{n}{k} \frac{d}{d} = \frac{t}{s}, \quad \frac{1}{k} = \frac{d}{k} \frac{d}{d} = \frac{d}{s}$$

Je kleiner k desto besser die Ungleichung. Wegen Bedingung $s = kd$: Je größer d desto besser die Ungleichung. Also wähle $d = \text{ggT}(s, t)$.

$$|\|Au\|_2^2 - \|u\|_2^2| \leq \left(\delta_{2s} \frac{t-d}{s} + \frac{d}{s} \delta_s \right) \|u\|_2^2$$

□

LINK: VORLESUNGSVIDEO 50 VOM 17.6.25

Je kleiner δ_s ist, desto stärkere Implikationen ergeben sich.

Daher Frage: Wie klein kann man δ_s machen (bei geeignet gewähltem $A \in \mathbb{C}^{m \times N}$)?

Kann δ_s bel. nahe an 0 sein? Oder gibt es untere Schranken an δ_s ?

Dies hängt natürlich von den Dimensionen $m, N, s \in \mathbb{N}$ ab.

Es gilt ähnlich wie beim Welch-Schranke für Kohärenz:

SATZ 5.8. Seien $c = \frac{1}{162}$, $C = 30$, $D = \frac{2}{3}$. Wähle $s, m, N \in \mathbb{N}$ mit $N \geq Cm$ und $s \in [2, N]$. Ist $\delta_s \in [0, D)$ und die RIP Konst. eines $A \in \mathbb{C}^{m \times N}$, so:

$$(43) \quad m\delta_s^2 \geq cs \quad \Leftrightarrow \delta_s \geq \sqrt{\frac{cs}{m}}$$

BEMERKUNG 5.9. $s = 1$ ausgeschlossen. Falls A normierte Spalten hat, dann $\delta_1 = 0$.

BEWEISANLEITUNG VON SATZ 5.8. DIESER BEWEIS IST UNREDIGIERT:

(1) Wähle $t = \lfloor \frac{s}{2} \rfloor \geq 1$. Zerlege

$$A = \left[\underbrace{A_1}_{t\text{-Spalten}} \mid \dots \mid \underbrace{A_{n-1}}_{t\text{-Spalten}} \mid \underbrace{A_n}_{\text{Rest}} \right]$$

$\Rightarrow N \leq nt$, $t_1, \dots, t_{n-1} \in \mathbb{C}^{m \times t}$. Für $i, j \in \bar{n}$, $i \neq j$ gilt:

$$\|A_i^* A_j - I\|_{2 \rightarrow 2} \leq \delta_t \leq \delta_s$$

$$\|A_i^* A_j\|_{2 \rightarrow 2} \leq \theta_{t,t} \leq \delta_{2t} \leq \delta_s$$

Damit gilt für EW:

$$1 - \delta_s \leq \lambda_k(A_i^* A_i) \leq 1 + \delta_s$$

$$\sigma_k(A_i^* A_j) \leq \delta_s$$

(2) $H = AA^* \in \mathbb{C}^{m \times m}$, $G = A^* A = [A_i^* A_j]_{i,j=1}^n \in \mathbb{C}^{N \times N}$. Zeige mit (1):

$$\text{tr}(H) \geq nN(1 - \delta_s)$$

$$\text{tr}(H)^2 \leq mnt[(n-1)\delta_s^2 + (1 + \delta_s)^2]$$

(3) Kombiniere zu:

$$m \geq \frac{nt(1 - \delta_s)^2}{(n-1)\delta_s^2 + (1 + \delta_s)^2}$$

$\Rightarrow$ Beh.

□

BEMERKUNG 5.10 (Stellen nun bisherige Ungleichungen für δ_s zusammen). Bekannte obere Schranke: Für $A \in \mathbb{C}^{m \times N}$ mit optimaler Kohärenz $\mu \leq \frac{c'}{\sqrt{m}}$ folgt:

$$\delta_s \leq (s-1)\mu \leq \frac{c's}{\sqrt{m}} \sim \frac{s}{\sqrt{m}}$$

Kann man diese (für gewisse Klassen von Matrizen A) verbessern?

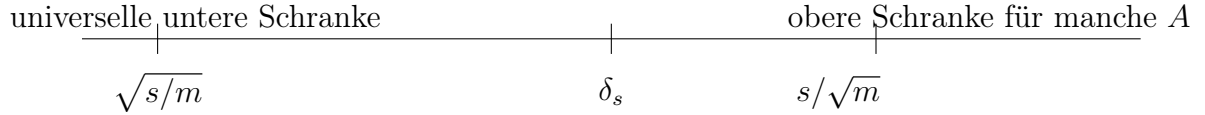
Gerade gezeigt in (43): Universelle untere Schranke

$$\delta_s \geq \sqrt{\frac{cs}{m}} \sim \sqrt{\frac{s}{m}}$$

Ziemlich große Diskrepanz zwischen bewiesener unterer und oberer Schranke. Faktor $\sqrt{s}$: Welchen Einfluss hat das?

$$s = 100 \Rightarrow \sqrt{s} = 10, \quad s = 25 \Rightarrow \sqrt{s} = 5$$

Daher wäre es wünschenswert, (für gewisse Matrizenklassen) eine obere Schranke wie die untere nachzuweisen.



BEMERKUNG 5.11 (Welche Bedeutung hat dies für die notwendige Anzahl m von Messungen, damit sich eine RIP-Konstante δ_s ergibt, die (hinreichend) kleiner als Eins ist?). Falls wir (nur) eine Schranke $\delta_s \leq c' \frac{s}{\sqrt{m}}$ zu Verfügung haben, liefert eine lineare Anzahl von Messungen $m \geq c^* s$

$$\delta_s \leq c' \frac{s}{\sqrt{m}} \leq \frac{c'}{c^*} \frac{s}{\sqrt{s}} \leq \frac{c'}{c^*} \sqrt{s}$$

also eine nutzlose Schranke, die nicht einmal beschränkt bleibt. Demnach gibt erst eine quadratische Anzahl von Messungen $m \geq c^* s^2$

$$\delta_s \leq c' \frac{s}{\sqrt{m}} \leq \frac{c'}{c^*} \frac{s}{s} \leq \frac{c'}{c^*}$$

ein nützliche Schranke, sofern der Quotient $\frac{c'}{c^*}$ kleiner als Eins ist. (1/3 ist schon eine gut brauchbare Schranke, siehe später.)

Wären wir imstande, Matrizen A mit RIP-Konstanten der Qualität

$$\delta_s \leq c' \sqrt{\frac{s}{m}}$$

(wie sie potentiell möglich sind, da sie die untere Schranke erlaubt) nachzuweisen oder sogar zu konstruieren, würde auch eine lineare Anzahl von Messungen $m \geq c^* s$ zu

$$\delta_s \leq c' \frac{\sqrt{s}}{\sqrt{m}} \leq \frac{c'}{c^*} \frac{s}{\sqrt{s}} \leq \frac{c'}{c^*}$$

also einer für Basisjagd effektiven Schranke führen, sofern z.B. $c'/c^* < 1/3$.

LINK: VORLESUNGSVIDEO 51 VOM 23.6.25

Zum jetzigen Zeitpunkt kennt man keine Matrizen $A \in \mathbb{C}^{m \times N}$ für die $\delta_s \leq c\sqrt{\frac{s}{m}}$ gilt. (Idealfall und ermöglicht effektive Basisjagd mit linearer Anzahl von Messungen). Dagegen ist die Existenz von Matrizen mit etwas schlechterem δ_s bekannt. Für geg. $\delta \in (0, 1)$, $m, N, s \in \mathbb{N}$ gibt es zufällige Matrizen $A \in \mathbb{C}^{m \times N}$ mit

$$\delta_s \leq \delta \quad \text{sofern } m \geq c \frac{s}{\delta^2} \ln\left(\frac{e \cdot N}{s}\right)$$

Bis auf den Wurzel-log-Faktor stimmt dies mit der bestmöglichen Schranke aus dem Satz 5.8 überein.

KAPITEL 6

Basisjagd (P₁) mit RIP

LINK: VORLESUNGSVIDEO 52 VOM 23.6.25

SATZ 6.1. Seien $s, m, N \in \mathbb{N}, A \in \mathbb{C}^{m \times N}$. Falls für A gilt

$$\delta_{2s} < \frac{1}{3}$$

so ist jedes $x \in \mathbb{C}_s^N$ die eindeutige Lösung von

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ mit NB } Az = Ax$$

LEMMA 6.2. Seien $q > p > 0, s, t \in \mathbb{N}, u \in \mathbb{C}^s, v \in \mathbb{C}^t$ mit

$$\max_{i \in \bar{s}} |u_i| \leq \min_{j \in \bar{t}} |v_j|$$

Dann $t^{1/p} \|u\|_q \leq s^{1/q} \|v\|_p$.

BEISPIEL 6.3. $s = t, p = 1, q = 2 \Rightarrow \|z\|_2 \leq \frac{1}{\sqrt{s}} \|v\|_1$

BEWEIS VON LEMMA 6.2.

$$\min_{i \in \bar{s}} |u_i| \leq \frac{\|u\|_q}{s^{1/q}} = \left(\frac{1}{s} \sum_{i=1}^s |u_i|^q \right)^{1/q} \leq \max_{i \in \bar{s}} |u_i|$$

Analog für $\|v\|_p t^{-1/p}$. Beobachtung folgt bei Anwendung der Voraussetzung. □

BEWEIS VON SATZ 6.1. Es ist ausreichend die NRE zz., dh.

$$\|v_T\|_1 < \frac{1}{2} \|v\|_1 \quad \forall v \in \ker A \setminus \{0\} \quad \forall T \subset \bar{N}, \#T = s.$$

Wir werden sogar

$$\|v_T\|_2 \leq \frac{\rho}{2\sqrt{s}} \|v\|_1 \quad \forall v \in \ker A, T \subset \bar{N}, |T| = s$$

zeigen, wobei $\rho = \frac{2\delta_{s2}}{1-\delta_{2s}} < 1$, da $\delta_{2s} < \frac{1}{3}$. C-S-Ungl. sagt nämlich $\|v_T\|_1 \leq \sqrt{s} \|v_T\|_2 \leq \frac{\rho}{2} \|v\|_1$.

zz: Falls wir die Ungleichung für $T = T_0 := \{\text{Indizes mit } s \text{ größten Werten von } |v_j|\}$ beweisen, gilt sie auch für alle anderen T . Wir unterteilen $T_0^c = \bar{N} \setminus T_0$ in

$T_1 = \{\text{Indizes mit } s \text{ größten Einträgen von } |v_j| \text{ in } T_0^c\},$

$T_2 = \{\text{Indizes mit } s \text{ größten Einträgen von } |v_j| \text{ in } (T_0 \cup T_1)^c\},$

usw.

Dann für $v \in \ker A$ gilt $Av_{T_0} = A(-v_{T_1} - v_{T_2} - \dots)$ und so

$$\|v_{T_0}\|_2^2 \leq \frac{1}{1 - \delta_{2s}} \|Av_{T_0}\|_2^2 = \frac{1}{1 - \delta_{2s}} \sum_{k \geq 1} \underbrace{\langle Av_{T_0}, A(-v_{T_k}) \rangle}_{\leq \delta_{2s} \|v_{T_0}\|_2 \|v_{T_k}\|_2}.$$

Dividieren mit $\|v_{T_0}\|_2$ liefert

$$\|v_{T_0}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \sum_{k \geq 1} \|v_{T_k}\|_2 \leq \frac{\rho}{2} \sum_{k \geq 1} \frac{1}{\sqrt{s}} \|v_{T_{k-1}}\|_1 = \frac{\rho}{2} \frac{1}{\sqrt{s}} \|v\|_1$$

□

BEMERKUNG 6.4. Es reicht $2\theta_{s,s} + \delta_s < 1$ statt $\delta_{2s} < \frac{1}{3}$ anzunehmen.

LINK: VORLESUNGSVIDEO 53 VOM 23.6.25

Satz sagt nichts über Stabilität & Robustheit von (P₁) aus.

SATZ 6.5. Seien $s, m, N \in \mathbb{N}$, $A \in \mathbb{C}^{m \times N}$. Erfüllt A

$$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0,6246,$$

so erfüllt für jedes $x \in \mathbb{C}^N$ und $y \in \mathbb{C}^m$ mit $\|Ax - y\|_2 \leq \eta$ die Lösung $x^\#$ von

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ unter } \|Az - y\|_2 \leq \eta$$

die Fehlerabschätzungen:

$$\|x - x^\#\|_2 \leq \frac{c}{\sqrt{s}} \sigma_s(x)_1 + D\eta,$$

$$\|x - x^\#\|_1 \leq c\sigma_s(x)_1 + D\sqrt{s}\eta,$$

wobei C, D nur von δ_{2s} abhängen.

Sowohl die ℓ_1 - als auch die ℓ_2 -Schranke, als auch ℓ_p -Schranke für $p \in [1, 2]$ folgen aus:

SATZ 6.6. $\delta_{2s}(A) \leq \frac{4}{\sqrt{41}} \Rightarrow A$ erfüllt $RNR_{s,2}^{\rho,\tau}$ mit $\rho \in (0, 1)$ und $\tau \in (0, \infty)$, die nur von $\delta_{2s}(A)$ abhängen.

LEMMA 6.7 (“Umkehrung” C-S-Ungl.). Seien $a_1 \geq a_2 \geq \dots \geq a_s \geq 0$. Dann

$$\sqrt{a_1^2 + \dots + a_s^2} \leq \frac{1}{\sqrt{s}} \sum_{j=1}^s a_j + \frac{\sqrt{s}}{4} (a_1 - a_s)$$

BEWEIS VON LEMMA 6.7. Übung

□

BEWEIS VON SATZ 6.6. Wir suchen $\rho \in (0, 1), \tau > 0$, sd.

$$\forall v \in \mathbb{C}^N, T \subset \overline{N} \text{ mit } |T| = s : \|v_T\|_2 \leq \frac{\rho}{\sqrt{s}} \|v_{T^c}\|_1 + \tau \|Av\|_2$$

vergleiche $(\text{RNR})_{s,2}^{\rho,\tau}$ in § 1.5 in Kapitel 2

Wie vorher reicht $T = T_0 := \{\text{größte } s \text{ Werte von Komp. von } v\}$. Definieren wir

$$T_1 = \{\text{größte } s \text{ Werte von Komp. von } v \text{ in } T_0^c\}$$

$$T_2 = \{\text{größte } s \text{ Werte von Komp. von } v \text{ in } (T_0 \cup T_1)^c\}$$

usw.

Da v_{T_0} s -dünn ist, ex. ein $|t| \leq \delta_s$, so dass

$$(44) \quad \|Av_{T_0}\|^2 = (1+t)\|v_{T_0}\|_2^2$$

Immer gilt $t^2 \leq \delta_s^2 \leq \delta_{2s}^2$. Wir nehmen zuerst an $|t| < \delta_{2s}$, und betrachten Spezialfall $|t| = \delta_{2s}$ am Ende des Beweises.

Wir zeigen $\forall k \geq 1$

$$(45) \quad |\langle Av_{T_0}, Av_{T_k} \rangle| \leq \sqrt{\delta_{2s}^2 - t^2} \|v_{T_0}\|_2 \|v_{T_k}\|_2$$

Dazu:

$$u = \frac{v_{T_0}}{\|v_{T_0}\|_2}, w = e^{i\theta} \frac{v_{T_k}}{\|v_{T_k}\|_2}$$

wobei $\theta \in [0, 2\pi)$, sodass

$$|\langle Au, Aw \rangle| = \Re(\langle Au, Aw \rangle)$$

Für bel. $\alpha, \beta \geq 0$ liefert Binom. Formel:

$$\begin{aligned} 2(\alpha + \beta) |\langle Au, Aw \rangle| &= 2(\alpha + \beta) \Re(\langle Au, Aw \rangle) \\ &= \|A(\alpha u + w)\|_2^2 - \|A(\beta u - w)\|_2^2 - (\alpha^2 - \beta^2) \|Au\|_2^2 \\ &\stackrel{\text{RIP und (44)}}{\leq} (1 + \delta_{2s}) \|\alpha u + w\|_2^2 - (1 - \delta_{2s}) \|\beta u - w\|_2^2 - (\alpha^2 - \beta^2) (1 + t) \|u\|_2^2 \\ &\stackrel{u \perp v, \text{normiert}}{=} \alpha^2 (\delta_{2s} - t) + \beta^2 (\delta_{2s} + t) + 2\delta_{2s} \end{aligned}$$

Wähle $\alpha = \frac{\delta_{2s} + t}{\sqrt{\delta_{2s}^2 - t^2}}, \beta = \frac{\delta_{2s} - t}{\sqrt{\delta_{2s}^2 - t^2}}$

$$\Rightarrow 2 \frac{2\delta_{2s}}{\sqrt{\delta_{2s}^2 - t^2}} |\langle Au, Aw \rangle| \leq \frac{(\delta_{2s} + t)^2}{\delta_{2s}^2 - t^2} (\delta_{2s} - t) + \frac{(\delta_{2s} - t)^2}{\delta_{2s}^2 - t^2} (\delta_{2s} + t) + 2\delta_{2s}$$

$$\stackrel{3.\text{Binom.}}{=} \delta_{2s} + t + \delta_{2s} - t + 2\delta_{2s} = 4\delta_{2s}$$

$$\Rightarrow |\langle Au, Aw \rangle| \leq \sqrt{\delta_{2s}^2 - t^2} \Leftrightarrow (45)$$

Als nächstes:

$$\|Av_{T_0}\|_2^2 = \langle Av_{T_0}, Av_{T_0} \rangle = \langle Av_{T_0}, Av \rangle - \sum_{k \geq 1} \langle Av_{T_0}, Av_{T_k} \rangle$$

$$\text{CS-U \& (45)} \quad \leq \|Av_{T_0}\|_2 \|Av\|_2 + \sum_{k \geq 1} \sqrt{\delta_{2s}^2 - t^2} \|v_{T_0}\|_2 \|v_{T_k}\|_2$$

$$(46) \quad (44) \quad = \|v_{T_0}\|_2 \left(\sqrt{1+t} \|Av\|_2 + \sqrt{\delta_{2s}^2 - t^2} \sum_{k \geq 1} \|v_{T_k}\|_2 \right).$$

Nach Lemma 6.7 gilt

$$\|v_{T_k}\|_2 \leq \frac{1}{\sqrt{s}} \|v_{T_k}\|_1 + \frac{\sqrt{s}}{4} (v_k^+ - v_k^-),$$

wobei

$$\left. \begin{array}{l} v_k^+ \text{ gr   ter} \\ v_k^- \text{ gr   ter} \end{array} \right\} \text{Wert von } |v_j|, j \in T_k \text{ ist.}$$

$$\begin{aligned} \Rightarrow \sum_{k \geq 1} \|v_{T_k}\|_2 &\leq \sum_{k \geq 1} \left(\frac{\|v_{T_k}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4} (v_k^+ - v_k^-) \right) \\ &\leq \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4} v_1^+ && \text{(Teleskop Argument)} \\ &\leq \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4} \frac{1|v_1| + \dots + 1|v_s|}{s} && \text{(Anordnung)} \\ &\leq \frac{1}{\sqrt{s}} \|v_{T_0^c}\|_1 + \frac{1}{4} \|v_{T_0}\|_2 && \text{(Cauchy-Schwarz)} \end{aligned}$$

Einsetzen in (46)

$$\Rightarrow \underbrace{\|Av_{T_0}\|_2^2}_{=(1+t)\|v_{T_0}\|_2^2} \leq \|v_{T_0}\|_2 \left(\sqrt{1+t} \|Av\|_2 + \sqrt{\delta_{2s}^2 - t^2} \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \sqrt{\delta_{2s}^2 - t^2} \frac{\|v_{T_0}\|_2}{4} \right).$$

Da $|t| \leq \delta_s$ gilt $\frac{\sqrt{\delta_{2s}^2 - t^2}}{1+t} \leq \frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2}}$, und damit

$$\|v_{T_0}\|_2 \leq \left[\frac{\|Av\|_2}{\sqrt{1+t}} + \frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2}} \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \frac{\delta_{2s}}{4\sqrt{1-\delta_{2s}^2}} \|v_{T_0}\|_2 \right].$$

Nutzen $\frac{1}{\sqrt{1+t}} \leq \frac{1}{\sqrt{1-\delta_{2s}^2}}$ f  r

$$\leq \frac{\|Av\|_2}{\sqrt{1-\delta_{2s}^2}} + \frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2}} \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \frac{\delta_{2s}}{4\sqrt{1-\delta_{2s}^2}} \|v_{T_0}\|_2$$

Subtraktion:

$$\Rightarrow \|v_{T_0}\|_2 \leq \underbrace{\frac{\delta_{2s}}{\sqrt{1-\delta_{2s}^2} - \frac{\delta_{2s}}{4}}}_{=: \rho < 1} \frac{\|v_{T_0^c}\|_1}{\sqrt{s}} + \underbrace{\frac{\sqrt{1+\delta_{2s}^2}}{\sqrt{1-\delta_{2s}^2} - \frac{\delta_{2s}}{4}}}_{=: \tau} \|Av\|_2$$

Jetzt:

$$\rho < 1 \Leftrightarrow \frac{5}{4} \delta_{2s} < \sqrt{1-\delta_{2s}^2} \Leftrightarrow \frac{41}{16} \delta_{2s}^2 < 1.$$

Dann gilt $\text{RNR}_{s,2}^{\rho,\tau}$ mit ρ und τ wie oben definiert.

Nun zum Spezialfall $|t| = \delta_{2s}$: Da $\delta_{2s} < \frac{4}{\sqrt{41}}$ gibt es

$$\delta \in \left(\delta_{2s}, \frac{4}{\sqrt{41}} \right)$$

Dieses erfüllt dann insbesondere

$$\delta^2 - t^2 > 0 \text{ und } 1 - \delta < 1 - \delta_{2s} \leq \frac{\|Ax\|_2^2}{\|x\|_2^2} \leq 1 + \delta_{2s} < 1 + \delta \text{ für } x \neq 0$$

Also kann man den Beweis oben durchführen, indem man überall δ_{2s} durch δ ersetzt. Wegen $\delta < \frac{4}{\sqrt{41}}$ ergeben sich dieselben Abschätzungen. □

LINK: VORLESUNGSVIDEO 54 VOM 24.6.25

BEMERKUNG 6.8. (a) Aus Kapitel 3 wissen wir, dass falls NRE erfüllt für A (also klappt Basisjagd für A) so auch für Matrizen GA (G invertierbar) und

$$\tilde{A} = \begin{bmatrix} A \\ B \end{bmatrix}$$

denn zusätzliche Messungen erhalten die NR-Eigenschaft. Wie ist das bei RIP?

- $\delta_s(A) = \delta_s(UA)$ für $U \in \mathbb{C}^{m \times m}$ unitär
- Für $\tilde{A} = \begin{bmatrix} A \\ B \end{bmatrix}$, $B \in \mathbb{C}^{m \times N}$ kann $\delta_s(\tilde{A}) > \delta_s(A)$ gelten, obwohl mehr Messungen zur Verfügung stehen
- beispielsweise für

$$A = [a_1, \dots, a_n] \in \mathbb{C}^{m \times N}, B = [0, \dots, 0, \sqrt{2+\delta}] \in \mathbb{C}^{1 \times N} \quad \text{und} \quad x = (0, \dots, 0, 1)^T$$

wobei $\delta = \delta_s(A)$, ergibt sich

$$\begin{aligned} \|\tilde{A}x\|_2^2 &= \left\| \begin{bmatrix} A \\ 0, \dots, 0, \sqrt{2+\delta} \end{bmatrix} (0, \dots, 0, 1)^T \right\|_2^2 \\ &= \|a_n\|_2^2 + |\sqrt{2+\delta}|^2 \geq 2 + \delta > (1 + \delta)\|x\|_2^2 \end{aligned}$$

- Auch Skalieren kann RIP verschlechtern, zB.

$$\exists A \text{ mit } \delta_s(2A) > \delta_s(A)$$

(b) Was ist die größte (schlechtes erlaubte) Konstante δ_{2s}^{opt} , sodass Basisjagd klappt für s -sparse Vektoren, sobald $\rightarrow \delta_{2s} \leq \delta_{2s}^{\text{opt}}$?

- Für jedes $\epsilon > 0$ kann man Matrizen A konstruieren, sd. $|\delta_{2s}(A) - \frac{1}{\sqrt{2}}| < \epsilon$, aber Basisjagd funktioniert nicht für Signal $x \in \mathbb{C}_s^N$.
- Für jedes $n \in \mathbb{N}$ kann man Matrizen A konstruieren, sd. $\delta_{2s}(A) > 1 - \frac{1}{n}$ hat und Basisjagd funktioniert für Signal $x \in \mathbb{C}_s^N$

(c) Was passiert, wenn wir die anderen Ordnungen der RIP betrachten?

$$A \in \mathbb{C}^{m \times N}$$

- $\delta_s(A) < \frac{1}{3} \Rightarrow$ Basisjagd funktioniert immer, d.h. Optimierer $x^\#$ von (P_1) ergibt Originalsignal $x \in \mathbb{C}_s^N$.
- $\delta_s(A) = \frac{1}{3} \Rightarrow$ im Allgemeinen funktioniert Basisjagd nicht, d.h. es gibt Matrizen A mit $\delta_s(A) = 1/3$ und $x^\# \neq x \in \mathbb{C}_s^N$ bei Basisjagd.

(d) Wir wissen, dass $\forall A \in \mathbb{C}^{m \times N}$ gilt:

$$(47) \quad m\delta_s^2 \geq \text{const. } s \quad \text{bzw.} \quad \delta_s \geq \sqrt{\frac{\text{const. } s}{m}}$$

Die Konstruktion von Matrizen A mit einer zu (47) ähnlichen, aber oberen Schranke an δ_s , vgl. Bemerkung 5.11, ist bisher nur mithilfe von Zufallsvariablen erfolgreich implementiert (d.h. A ist Zufallsmatrix) und erreicht zumindest

$$\delta_s(A) \leq \sqrt{\frac{\text{const. } s}{m} \ln\left(\frac{eN}{s}\right)}$$

Basisjagd mit zufälligen Matrizen

LINK: VORLESUNGSVIDEO 55 VOM 24.6.25

BEMERKUNG 7.1 (Idee/ Beispiel). Seien $X = (X_1, \dots, X_m), Y = (Y_1, \dots, Y_m) \in \mathbb{R}^m$, sd. $X_1, Y_1, \dots, X_m, Y_m \in \{-1, 1\}$ jeweils mit Wahrscheinlichkeit $\frac{1}{2}$, unabhängige Familie.

Normalisierung $\tilde{X} = \frac{X}{\|X\|_2}, \tilde{Y} = \frac{Y}{\|Y\|_2}$ mit $\|X\|_2^2 = 1 + \dots + 1 = m$.

$$\langle X, Y \rangle_2 = \sum_{j=1}^m X_j Y_j$$

$$\mathbb{E}(\langle X, Y \rangle_2) = \mathbb{E}\left(\sum_{i=1}^m X_i \cdot Y_i\right) = \sum_{i=1}^m \mathbb{E}(X_i) \mathbb{E}(Y_i) = 0$$

Das bedeutet, $\tilde{X}, \tilde{Y}$ sind im „Schnitt“ orthogonal. Wie weit “ streut ” die Zufallsvariable $\langle X, Y \rangle$ um ihrem Mittelwert Null? Erster Indikator: Varianz:

$$\begin{aligned} \text{Var}(\langle X, Y \rangle) &= \mathbb{E}(\langle X, Y \rangle^2) - 0 = \mathbb{E}\left(\sum_{i=1}^m X_i \cdot Y_i\right)^2 \\ &= \mathbb{E}\left(\sum_{i,j=1}^m X_i X_j Y_i Y_j\right) = \sum_{i,j=1}^m \mathbb{E}(X_i X_j Y_i Y_j) \\ &= \sum_{i,j=1}^m \mathbb{E}(X_i X_j) \mathbb{E}(Y_i Y_j) = \sum_{i,j=1}^m \delta_{i,j} \delta_{i,j} = m, \\ \mathbb{E}(\tilde{X}, \tilde{Y})^2 &= \frac{1}{m} \end{aligned}$$

Intuition: Für „die meisten Realisierungen ” sollte $\langle \tilde{X}, \tilde{Y} \rangle$ nahe Null sein.

Um dies zu präzisieren benutzen wir Konzentrationsungleichungen.

SATZ 7.2 (Markov-Čebyšev-Ungleichung). Sei $h: [0, \infty) \rightarrow [0, \infty)$ monoton wachsend. Dann gilt

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}(h(|X|) \mathbb{1}_{\{|X| \geq c\}})}{h(c)} \leq \frac{\mathbb{E}(h(|X|))}{h(c)}$$

für alle $c > 0$ mit $h(c) > 0$.

BEWEIS. Auf $\{|X| \geq c\}$ gilt $h(|X|) \geq h(c)$

$$\Rightarrow h(c) \mathbb{1}_{\{|X| \geq c\}} \leq h(|X|) \mathbb{1}_{\{|X| \geq c\}} \leq h(|X|)$$

Integration ergibt

$$h(c) \mathbb{P}(|X| \geq c) \leq \mathbb{E}(h(|X|); |X| \geq c) \leq \mathbb{E}(h(|X|)).$$

Insbesondere kann man die letzte Ungleichung auch für $h(c) = 0$ formulieren. □

Ein besonders einfacher Fall ist die:

BEMERKUNG 7.3 (Chebyshev-Ungleichung). für ZV $Z = \langle \tilde{X}, \tilde{Y} \rangle$:

$$\mathbb{P}(|Z| \geq \epsilon) = \mathbb{P}(|Z|^2 \geq \epsilon^2) \leq \frac{\mathbb{E}(Z^2)}{\epsilon^2} = \frac{1}{m \epsilon^2}$$

zB. für $\epsilon = m^{-1/4}$ erhalten wir:

$$\mathbb{P}(|Z| \geq m^{-1/4}) \leq \frac{1}{m m^{-1/2}} = \frac{1}{\sqrt{m}}$$

Kontraposition & Interpretation

$$1 - \mathbb{P}(|Z| < m^{-1/4}) = 1 - \mathbb{P}(\tilde{X}, \tilde{Y} \text{ fast } \perp) \text{ ist klein für große } m$$

Um diese Strategie effizient zu implementieren, benötigen wir:

SATZ 7.4 (Bernstein Ungleichung). $X_1, \dots, X_m : \Omega \rightarrow \mathbb{R}$ unabh. ZV mit $\mathbb{E}(X_i) = 0$ und

$$(48) \quad \mathbb{E}|X_i|^n \leq n! R^{n-2} \frac{\sigma_i^2}{2}$$

für $n \geq 2$, $i \in \overline{m}$ und gewisse Konst. $R, \sigma_1, \dots, \sigma_m \in (0, \infty)$. Dann

$$\forall t > 0 : \quad \mathbb{P} \left(\left| \sum_{i=1}^m X_i \right| \geq t \right) \leq 2e^{-\frac{1}{\sigma^2 + Rt} \frac{t^2}{2}}$$

wobei $\sigma^2 = \sum_{i=1}^m \sigma_i^2$

BEMERKUNG 7.5. (i) Bed. (48) erinnert an Taylorentw.

(ii) $n = 2$: $\mathbb{E}X_i^2 \leq \sigma_i^2$ ist Varianzschranke

(iii) (48) $\leftrightarrow$ höhere Momente durch Varianz kontrollierbar.

(iv) Abfallverhalten von $|\sum X_i|$ ist irgendwo zwischen Gaußsch. und exponentiell.

LINK: VORLESUNGSVIDEO 56 VOM 24.6.25

BEISPIEL 7.6. (a) Beschränkte ZV $|X_i| \leq k$ fast sicher mit $\mathbb{E}X_i^2 \leq \sigma_i^2$

$$\Rightarrow \mathbb{P} \left(\left| \sum_{i=1}^m X_i \right| \geq t \right) \leq 2 \exp \left(- \frac{1}{\sigma^2 + \frac{k}{3}t} \frac{t^2}{2} \right)$$

(b) $X_1, \dots, X_m$ subexponentielle ZV, dh. es exist. $\beta, \kappa \in (0, \infty)$ sd.

$$\mathbb{P}(|X_i| \geq t) \leq \beta e^{-\kappa t}, \quad (t \geq 0)$$

$$\Rightarrow \mathbb{P} \left(\left| \sum_{i=1}^m X_i \right| \geq t \right) \leq 2 \exp \left(- \frac{1}{2\beta m + \kappa t} \frac{(\kappa t)^2}{2} \right)$$

Beweis: Übung

BEWEIS VON SATZ 7.4. Wir haben

$$\mathbb{P} \left(\sum_{i=1}^m X_i \geq t \right) \stackrel{\theta > 0 \text{ bel.}}{=} \mathbb{P}(e^{\theta \sum X_i} \geq e^{\theta t}) \stackrel{\text{Markov-Ungl.}}{\leq} e^{-\theta t} \mathbb{E}(e^{\theta \sum X_i}) \stackrel{\text{unabhängig}}{=} e^{-\theta t} \prod_i \mathbb{E}(e^{\theta X_i})$$

Es gilt

$$\begin{aligned}
 \mathbb{E} e^{\theta X_i} &= \sum_{n=0, \neq 1}^{\infty} \frac{\theta^n \mathbb{E}(X_i^n)}{n!} && \text{(Taylor-Entwicklung und Fubini)} \\
 &\stackrel{\Delta\text{-Ungl.}}{\leq} 1 + \sum_{n=2}^{\infty} \frac{\theta^n \mathbb{E}|X_i^n|}{n!} \\
 &\stackrel{\text{Ann.}}{\leq} 1 + \theta^2 \sum_{n=2}^{\infty} \theta^{n-2} R^{n-2} \frac{\sigma_i^2}{2} \\
 &\stackrel{\text{geom. Reihe}}{=} 1 + \theta^2 \frac{\sigma_i^2}{2} \frac{1}{1 - \theta R}
 \end{aligned}$$

Einsetzen:

$$\begin{aligned}
 \mathbb{P}\left(\sum_{i=1}^m X_i \geq t\right) &\leq e^{-\theta t} \prod_i \left(1 + \frac{1}{2} \theta^2 \sigma_i^2 \frac{1}{1 - \theta R}\right) \\
 &\stackrel{1+x \leq e^x}{\leq} e^{-\theta t} \prod_{i=1}^m \exp\left(\frac{\theta \sigma_i^2}{2(1 - \theta R)}\right) \\
 &\stackrel{\text{Def. } \sigma^2}{=} e^{-\theta t} \exp\left(\frac{\theta \sigma^2}{2(1 - \theta R)}\right) \\
 &= \exp\left(-\theta t + \frac{\sigma^2}{2} \frac{\theta}{(1 - \theta R)}\right) \\
 &\stackrel{\text{Wähle } \theta = \frac{t}{\sigma^2 + Rt}}{=} \exp\left(-\frac{t^2}{\sigma^2 + Rt} + \frac{\sigma^2}{2} \frac{t}{\sigma^2 + Rt} \frac{1}{1 - \frac{tR}{\sigma^2 + Rt}}\right) \\
 &= \dots = \exp\left(-\frac{t^2/2}{\sigma^2 + Rt}\right).
 \end{aligned}$$

Ähnlich folgt $\mathbb{P}(-\sum_{i=1}^m X_i \geq t) \leq \exp\left(-\frac{t^2/2}{\sigma^2 + Rt}\right)$ und daher die Beh. □

LINK: VORLESUNGSVIDEO 57 VOM 24.6.25

Nun wollen wir eine Konzentrationsungl. für $\|Ax\|_2^2$ wobei $\|x\|_2 = 1$ und A „Gaußsch“ ist

DEFINITION 7.7. Seien $m, N \in \mathbb{N}$. Seien $g_{ij} \sim N(0, 1)$ für $i \in \overline{m}, j \in \overline{N}$ unabhängige ZV. Dann heißt

$$A = \frac{1}{\sqrt{m}} (g_{i,j})_{\substack{i \in \overline{m} \\ j \in \overline{N}}}$$

(standardisierte) *Gaußsche Zufallsmatrix*.

Konzentrationsungl. für solche A :

SATZ 7.8. Sei $x \in \mathbb{R}^N$ mit $\|x\|_2 = 1$. Dann ex. eine absol. Konstante $C = \frac{1}{36} \in (0, \infty)$ sd.:

$$\forall t \in (0, 1] : \quad \mathbb{P}(|\|Ax\|_2^2 - 1| \geq t) \leq 2e^{-Cmt^2}$$

BEWEIS. Idee: Betrachte Ax , berechne $\mathbb{E}(\|Ax\|_2^2) = \text{Var}(Y_i)$, $\frac{1}{\sqrt{m}}(Y_1, \dots, Y_m)$, $Y_i \sim N(0, 1)$ und wende Bernstein-Ungl. an. Vektor

$$Ax = \frac{1}{\sqrt{m}} \left(\sum_{j=1}^N x_j g_{i,j} \right)_{i \in \bar{m}}$$

hat Gauß ZV als Komponenten, da Summe unabh. von Gauß ZV.

mit Mittelwert Null, da $\mathbb{E}(g_{i,j}) = 0$

$$\text{Var} \left(\sum_{j=1}^N x_j g_{i,j} \right) = \mathbb{E} \left(\sum_{j=1}^N x_j g_{i,j} \right)^2 = \mathbb{E} \left(\sum_{j,k=1}^N x_j x_k g_{i,j} g_{i,k} \right) = \sum_{j,k=1}^N x_j x_k \mathbb{E}(g_{i,j} g_{i,k}) = 1$$

Also

$$Ax \sim \frac{1}{\sqrt{m}} (Y_1, \dots, Y_m)$$

mit unabh. ZV $Y_1, \dots, Y_m \sim \mathcal{N}(0, 1)$

$$\|Ax\|_2^2 \sim \frac{1}{m} (Y_1^2 + \dots + Y_m^2) \Rightarrow \mathbb{E}\|Ax\|_2^2 = \mathbb{E}Y_1^2 = \text{Var}(Y_1) = 1$$

Mit $X_i = Y_i^2 - 1$

$$\mathbb{P}(|\|Ax\|_2^2 - 1| \geq t) = \mathbb{P}\left(\left|\frac{1}{m} \sum_i Y_i^2 - 1\right| \geq t\right) = \mathbb{P}\left(\frac{1}{m} \sum_i X_i \geq t\right)$$

Wollen Bernstein anwenden. Voraussetzungen prüfen!

$\mathbb{E}X_i = 0$ da $\text{Var}(Y_i) = 1$

Y_i hat Gauß-Abfall $\Rightarrow X_i$ ist sub-exponentiell (Variablentransformation)

$Y_i \sim \mathcal{N}(0, 1) \Rightarrow \beta = 2, \kappa = 1/2$. Für $0 < t \leq 1$ folgt:

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{i \in \bar{m}} X_i\right| \geq t m\right) &\leq 2 \exp\left(-\frac{1}{2} \frac{(\kappa t m)^2}{2\beta m + \kappa t m}\right) \\ &= 2 \exp\left(-\frac{1}{2} \frac{\kappa^2 t^2 m}{2\beta + \kappa t}\right) \\ &= 2 \exp\left(-\frac{1}{8} \frac{t^2}{4 + t/2} m\right) \\ &\leq 2 \exp\left(-\frac{1}{8} \frac{t^2}{4 + 1/2} m\right) = 2 \exp\left(-\frac{t^2}{36} m\right) \end{aligned}$$

□

LINK: VORLESUNGSVIDEO 58 VOM 30.6.25

BEMERKUNG. In diesem Abschnitt weicht das Skript leicht vom Vorlesungsvideo ab.

Unser Ziel: Gaußsche Zufallsmatrix $A = \frac{1}{\sqrt{m}}(g_{ij})_{\substack{i \in \bar{m} \\ j \in \bar{N}}}$ wie oben hat mit hoher Wahrscheinlichkeit kleine RIP-Konstante.

Genauer:

SATZ 7.9. Es ex. universelle Konstante $C = 144(1 + \ln 9) > 0$, sodass für bel. $N, m \in \mathbb{N}$, $N \geq m$, $s, \epsilon, \delta \in (0, 1)$ mit $m \geq C\delta^{-2}(s(\ln \frac{eN}{s}) + \ln \frac{2}{\epsilon})$ gilt

$$\mathbb{P}(\delta_s(A) \leq \delta) \geq 1 - \epsilon$$

Beweis-Strategie:

- Brauchen:

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| > \delta \text{ für alle } x \in \mathbb{R}_s^N, \|x\|_2 = 1\right) < \epsilon$$

- Haben: Für festes $x \in \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$ gilt:

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| > \delta\right) \leq 2 \exp\left(-\frac{1}{36}m\delta^2\right)$$

- Idee zur Umsetzung:

Finde für jede Teilmenge $T \subset \overline{N}$, $\#T \leq s$ ein genügend feinmaschiges diskretes Netz von $\mathbb{R}_T^N \cap \mathbb{S}^{d-1}$ und verwende Subadditivität für $\mathbb{P}$.

Dazu benötigen wir:

SATZ 7.10. Sei $t > 0$, $d \in \mathbb{N}$. Es ex. eine Menge $\mathcal{N} \subset \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$ mit $\#\mathcal{N} \leq (1 + 2/t)^d$ und

$$(49) \quad \forall y \in \mathbb{S}^{d-1} \text{ ex. ein } x \in \mathcal{N} \text{ mit } \|y - x\|_2 \leq t$$

BEMERKUNG 7.11 (Interpretation). Satz 7.10 $\Leftrightarrow$ sphärische Kappen mit Radius t und Mittelpunkten $x \in \mathcal{N}$ bilden Überdeckung von $\mathbb{S}^{d-1}$.

Der Beweis von Satz 7.10 wird später geführt. Zunächst verwenden wir ihn, um Satz 7.9 zu beweisen.

BEWEIS VON SATZ 7.9. Union bound / Subadditivität liefern:

$$\mathbb{P}(\delta_s(A) > \delta) \leq \sum_{T \subset \overline{N}, |T|=s} \mathbb{P}(\exists y \in \mathbb{R}_T^N : \|y\|_2 = 1 \text{ und } \left|\|Ay\|_2^2 - 1\right| > \delta)$$

$$\text{Invarianz der Verteilung} = \binom{N}{s} \mathbb{P}(\exists y \in \mathbb{R}_{1,\dots,s}^N : \|y\|_2 = 1 \text{ und } \left|\|Ay\|_2^2 - 1\right| > \delta)$$

Ab jetzt bez. $T = \{1, \dots, s\}$. Satz 7.10 liefert für jedes $\rho > 0$ eine Punktemenge:

$\mathcal{N} \subset \{y \in \mathbb{R}_T^N : \|y\|_2 = 1\}$ mit $\#\mathcal{N} \leq (1+2/\rho)^s$ und $\min_{x \in \mathcal{N}} \|x - y\|_2 \leq \rho$ für alle $y \in \mathbb{S}^{d-1} \cap \mathbb{R}_T^N$

Sei jetzt die Realisierung der ZM A so, dass gilt:

$$\left|\|Ax\|_2^2 - 1\right| \leq t \text{ für alle } x \in \mathcal{N}.$$

Benutze:

$$\begin{aligned} \left|\|Ay\|_2^2 - 1\right| &\leq \delta \quad \text{für alle } y \in \mathbb{R}_T^N \text{ mit } \|y\|_2 = 1 \\ \Leftrightarrow \|B\|_{2 \rightarrow 2} &\leq \delta \quad \text{für } B = A_T^* A_T - I \end{aligned}$$

Nun, wähle t richtig:

$$\begin{aligned} \max_{y \in \mathbb{S}^{s-1}} |\langle By, y \rangle| &= \max_{y \in \mathbb{S}^{s-1}} |\langle Bx, x \rangle + \langle B(y+x), (y-x) \rangle| \\ &\leq \max_{y \in \mathbb{S}^{s-1}} |\langle Bx, x \rangle| + \|B\|_{2 \rightarrow 2} \underbrace{\|y+x\|_2}_{\leq 2} \|y-x\|_2 \end{aligned}$$

für geeignet gewähltes $x = x(y) \in \mathcal{N}$

$$\leq \max_{y \in \mathbb{S}^{s-1}} |\langle Bx, x \rangle| + 2\rho \|B\|_{2 \rightarrow 2}$$

Äquivalenz oben

$$\leq t + 2\rho\|B\|_{2 \rightarrow 2}$$

Charakterisierung der Norm von B :

$$\begin{aligned} \Rightarrow \|B\|_{2 \rightarrow 2} &\leq t + 2\rho\|B\|_{2 \rightarrow 2} \\ \Rightarrow \|B\|_{2 \rightarrow 2} &\leq \frac{t}{1 - 2\rho} \text{ für } \rho \in (0, 1/2). \end{aligned}$$

Setze $t := \delta(1 - 2\rho)$. Komplementrechnung ergibt: ($T \subset \bar{N}$ immer noch fix.)

$$\begin{aligned} \mathbb{P}(\exists y \in \mathbb{R}_T^N \text{ mit } \|y\|_2 = 1 \text{ und } |\|Ay\|_2^2 - 1| > \delta) &\leq \mathbb{P}(\exists x \in \mathcal{N} : |\|Ax\|_2^2 - 1| > t) \\ &\leq \sum_{x \in \mathcal{N}} \mathbb{P}(|\|Ax\|_2^2 - 1| > t) \end{aligned}$$

Satz 7.8 und Satz 7.10 mit $\mathbb{S}^{s-1}$

$$\leq 2 \left(1 + \frac{2}{\rho}\right)^s e^{-mt^2/36}$$

Anzahl der Teilmengen und einsetzen:

$$(50) \quad \mathbb{P}(\delta_s(A) \geq \delta) \leq 2 \left(1 + \frac{2}{\rho}\right)^s \binom{N}{s} e^{-mt^2/36} = 2 \left(1 + \frac{2}{\rho}\right)^s \binom{N}{s} e^{-m\delta^2(1-2\rho)^2/36}$$

Es bleibt zu zeigen, dass (50) $\leq \epsilon$.

Für $\rho \in (0, 1/2)$ ist $t = \delta(1 - 2\rho) \in (0, \delta) \subset (0, 1)$.

$$\begin{aligned} (50) \leq \epsilon &\Leftrightarrow e^{m\delta^2(1-2\rho)^2/36} \geq \frac{2}{\epsilon} \left(1 + \frac{2}{\rho}\right)^s \binom{N}{s} \\ &\Leftrightarrow m \geq \frac{36}{(1-2\rho)^2} \delta^{-2} \left[s \ln \left(1 + \frac{2}{\rho}\right) + \ln \binom{N}{s} + \ln \frac{2}{\epsilon} \right] \end{aligned}$$

Wähle bspweise. $\rho = \frac{1}{4}$ (für andere $\rho \in (0, \frac{1}{2})$ geht es ähnlich). Dann ist $s \ln(1 + \frac{2}{\rho}) = s \ln 9$, und es ist noch (im 2. Summand) zu zeigen

$$\ln \binom{N}{s} \leq s \ln \left(\frac{eN}{s}\right) \Leftrightarrow \binom{N}{s} \leq \left(\frac{eN}{s}\right)^s.$$

Das folgt aus: $\forall n, k \in \mathbb{N}, n \geq k > 0 : \left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ (HA)

Ergibt hinreichende Bedingung:

$$m \geq 144 \delta^{-2} \left[s \left(\ln \left(\frac{eN}{s}\right) + \ln 9 \right) + \ln \frac{2}{\epsilon} \right],$$

und die Beobachtung folgt mit der Wahl $C = 144(1 + \ln 9)$. □

LINK: VORLESUNGSVIDEO 59 VOM 30.6.25

BEWEIS VON SATZ 7.10. Wähle $x_1 \in \mathbb{S}^{d-1}$ bel. und induktiv $x_i \in \mathbb{S}^{d-1}$, $i = 1, 2, \dots$, sodass

$$(51) \quad \|x_{i+1} - x_j\|_2 > t, \quad j = 1, 2, \dots, i$$

so lange es geht. Sobald der Wert i erreicht ist, so dass (51) nicht erfüllt werden kann, setzen wir $N = i$. Dann gilt für jedes $y \in \mathbb{S}^{d-1}$:

$$(52) \quad \min_{i=1}^N \|y - x_i\| \leq t$$

ansonsten ergibt sich ein $x_{N+1} = y$, das (51) erfüllt, im Widerspruch zur Abbruchbedingung.

$\Rightarrow$ Die Menge $\mathcal{N} = \{x_1, \dots, x_N\}$ erfüllt Bed. (49). Wegen (51) sind die Kugeln $B_{t/2}(x_1), \dots, B_{t/2}(x_N)$ disjunkt. Alle Kugeln sind in $B_{1+t/2}(0) \subset \mathbb{R}^d$ enthalten, insbesondere ist die Summe ihrer Volumina abgeschätzt durch

$$N(t/2)^d |B_1(0)| \leq |B_{1+t/2}| = (1 + t/2)^d |B_1(0)|$$

$$\Rightarrow N \leq \frac{(1+t/2)^d}{(t/2)^d} = \left(\frac{2}{t} + 1\right)^d$$

□

LINK: VORLESUNGSVIDEO 60 VOM 30.6.25

KOROLLAR 7.12. Es ex. eine abs. Konst. $C > 0$ so, dass für $\epsilon \in (0, 1)$ und $N, m, s \in \mathbb{N}$ mit $s < N$ und

$$m \geq C \left(s \ln \frac{eN}{s} + \ln \frac{2}{\epsilon} \right)$$

und

$$A = \frac{1}{\sqrt{m}} (g_{ij})_{\substack{i \in \overline{m} \\ j \in \overline{N}}} \text{ mit } g_{ij} \sim \mathcal{N}(0, 1) \text{ unabh.}$$

mit Wahrscheinlichkeit $\geq 1 - \epsilon$ jeder s -dünne Vektor x durch Basisjagd aus $y = Ax$ rekonstruiert werden kann.

BEWEIS. Satz 7.9 liefert $\mathbb{P}(\delta_{2s}(A) < \frac{1}{3}) \geq 1 - \epsilon$ bei geeigneter Anzahl m .

Satz 6.1 bzw. Satz 6.5 garantieren $\delta_{2s} < \frac{1}{3} \Rightarrow$ Basisjagd rekonstruiert x .

□

LINK: VORLESUNGSVIDEO 61 VOM 1.7.25

Auch robuste & stabile Basisjagd ist gewährleistet.

Benutze:

- Satz 7.9: $\mathbb{P}(\delta_{2s}(A) < \frac{4}{\sqrt{41}}) \geq 1 - \epsilon$ bei geeigneter Anzahl m .
- Satz 6.1: $\delta_{2s}(A) \leq \frac{4}{\sqrt{41}} \Rightarrow$ robuste & stabile Basisjagd.

BEMERKUNG 7.13. Wahrscheinlichkeitstheoretische Komponente im Beweis von Satz 7.9 war die Konzentrationsungleichung von Satz 7.8, dh. für festes $x \in \mathbb{S}^{N-1}$ gilt:

$$(53) \quad \mathbb{P}(|\|Ax\|_2^2 - 1| \geq t) \leq 2e^{-\frac{mt^2}{36}}, \quad t \in (0, 1)$$

Für beliebiges orthogonales /unitäres $U: \mathbb{R}^N \rightarrow \mathbb{R}^N$ erfüllt $\tilde{A} = AU$:

$$\mathbb{P}(|\|\tilde{A}x\|_2^2 - 1| \geq t) = \mathbb{P}(|\|AUx\|_2^2 - 1| \geq t) = \mathbb{P}(|\|Az\|_2^2 - 1| \geq t)$$

für $z = Ux \in \mathbb{S}^{N-1}$. Ist also A eine zufällige Matrix, welche die Konzentrationsung. (53) erfüllt für jedes fixe $x \in \mathbb{S}^{N-1}$, so gilt ebendies für AU .

(a) Universalität/Rotationsinvarianz

Was passiert, falls Signal x nicht in der kanonischen Basis sparse ist, sondern in einer anderen (angepassten), z.B. Wavelet Basis? In dieser Situation ex. orth. $N \times N$ basistransformations-Matrix U , sd. $z = Ux$ s -sparse ist.

Da $Az = AUx = \tilde{A}x$ ist Basisjagd auch erfolgreich für Matrix AU .

Fazit: Sparsity darf in diesem Kontext in jeder beliebigen Basis formuliert werden.

(a) Andere zufällige Matrizen:

Das heißt alle Ensembles von Zufallsmatrizen, die diese Eig. erfüllen, können anstelle von Gauß-Matrizen in den letzten 3 Sätzen benutzt werden, insbesondere

$$A = \frac{1}{\sqrt{m}}(X_{ij})_{\substack{i \in \overline{m} \\ j \in \overline{N}}}, X_{ij} \text{ unabh. ident. vert. subgaußsche ZV}$$

konkret z.B. die beschränkten

(A.) Rademacher/Bernoulli ZV $\mathbb{P}(X_{ij} = 1) = \frac{1}{2} = \mathbb{P}(X_{ij} = -1)$

(B.) uniform verteilte ZV $X_{i,j} \in [-a, a]$

DEFINITION 7.14. ZV $X_1, \dots, X_n : (\Omega, \mathbb{P}) \rightarrow \mathbb{R}$ heißen *Rademacher-ZV*, falls sie unabh. sind und

$$\forall i \in \overline{n} : \mathbb{P}(X_i = -1) = \frac{1}{2} = \mathbb{P}(X_i = +1)$$

$X : (\Omega, \mathbb{P}) \rightarrow \mathbb{R}$ heißt *sub-Gaußsche ZV*, falls exist. $\beta, \kappa \in (0, \infty)$ sd.

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\kappa t^2}, \quad (t \geq 0)$$

Exkurs: Das Lemma von Johnson & Lindenstrauss

LINK: VORLESUNGSVIDEO 62 VOM 1.7.25

- eng verw. mit der RIP-Eigenschaft
- Motivation unabh. von CS
- Bew. benutzt Konzentrationsung.

Idee/Aufgabenstellung:

Projiziere M gegebene Punkte $\{x_1, \dots, x_M\} \subset \mathbb{R}^N$ in einem Raum $\mathbb{R}^m$ möglichst kleiner Dimension $m \in \mathbb{N}$, ohne die Abstände zu stark zu verändern.

Anwendung:

Effektives Speichern in Datenbanken. Oftmals ist die Dimension N des Datenraumes hoch, es ex. aber nur wenige Datenpunkte. Dann speichert man die Daten besser nach der Projektion in dem $\mathbb{R}^m$ ab und braucht nur $M \cdot m$ statt $M \cdot N$ reelle Zahlen zu speichern.

SATZ 8.1 (Johnson & Lindenstrauss-Lemma). Seien $x_1, \dots, x_M \in \mathbb{R}^N$, $t \in (0, 1)$ und $m \in \mathbb{N}$, $m > ct^{-2} \ln M = 80t^{-2} \ln M$. Dann ex. eine lin. Abb. $A : \mathbb{R}^N \rightarrow \mathbb{R}^m$ abhängig von $x_1, \dots, x_M$ mit

$$(54) \quad (1-t)\|x_i - x_j\|_2^2 \leq \|Ax_i - Ax_j\|_2^2(1+t)\|x_i - x_j\|_2^2 \quad (i, j \in \overline{M})$$

BEWEIS. Sei $A = \frac{1}{\sqrt{m}}(g_{ij})_{\substack{i \in \overline{m}, \\ j \in \overline{N}}}$, $g_{ij} \sim \mathcal{N}(0, 1)$ unabh.

Kor. 7.8 der Bernstein Ungl. $\Rightarrow$ für festes $x \in \mathbb{R}^N$ und normiertes $y = x/\|x\|_2$ gilt:

$$\mathbb{P}(|\|Ax\|_2^2 - \|x\|_2^2| > t\|x\|_2^2) = \mathbb{P}(|\|Ay\|_2^2 - 1| > t) \leq 2\exp(-mt^2/36)$$

Anwendbar auf jede Differenz $x = x_i - x_j$, $i, j \in \overline{M}$. Nur $i \neq j$ interessant. Es gibt $\binom{M}{2}$ solcher Paare. Subadditivität:

$$\begin{aligned} & \mathbb{P}\left(\exists i \neq j \in \overline{M} \text{ mit } |\|A(x_i - x_j)\|^2 - \|x_i - x_j\|^2| > t\|x_i - x_j\|_2^2\right) \\ & \leq 2 \binom{M}{2} e^{-mt^2/36} \leq M^2 e^{-mt^2/36} \end{aligned}$$

Komplement:

$$\mathbb{P}(|\|Ax_i - Ax_j\|_2^2 - \|x_i - x_j\|_2^2| \leq t\|x_i - x_j\|_2^2 \text{ für alle } i, j \in \overline{M}) \geq 1 - M^2 e^{-mt^2/36}$$

Also ist die Ex. eines A mit (54) gewährleistet, sofern $M^2 e^{-mt^2/36} < 1 \Leftrightarrow 2 \ln M < mt^2/36 \Leftrightarrow m > 72t^2 \ln M$ $\square$

Vergleich mit probabilistische Methode von P. Erdős in der Graphentheorie.

LINK: VORLESUNGSVIDEO 63 VOM 1.7.25

Kann man die Bed (untere Schranke) an m abschwächen und trotzdem (54) erhalten? Dh. wir suchen nach komplementäre Schranke. Sei

$$x_i = \frac{1}{\sqrt{2}} e_i \in \mathbb{R}^N, \quad i = 1, \dots, N, \quad e_1, \dots, e_N \text{ kanonische Basis.}$$

Pythagoras $\xrightarrow{i \neq j} \|x_i - x_j\|_2 = 1$

Sei $A = A(0, x_1, \dots, x_N)$ wie in JLL, dh.

$$\forall i \in \overline{M} : (1 - \delta) \frac{1}{2} = (1 - \delta) \|x_i\|_2^2 \leq \|Ax_i\|_2^2 \leq (1 + \delta) \|x_i\|_2^2 = (1 + \delta) \frac{1}{2}$$

und

$$\forall i, j \in \overline{M} : (1 - \delta) \leq \|Ax_i - Ax_j\|_2^2 \leq (1 + \delta)$$

Die erste Ungleichung impliziert, dass alle Mittelpunkte Ax_i im Ball um $0 \in \mathbb{R}^m$ mit Radius $\sqrt{(1 + \delta)/2}$ liegen. (Tatsächlich liegen sie in der Kugelschale

$$B_{\sqrt{(1+\delta)/2}}(0) \setminus B_{\sqrt{(1-\delta)/2}}(0)$$

aber diese Verbesserung werden wir i.w. nicht verfolgen.) Die zweite Ungleichung impliziert, dass die Kugel um $Ax_1, \dots, Ax_N$ mit Radius $\frac{\sqrt{1-\delta}}{2}$ disjunkt sind. Zusammengenommen sind N disjunkte Bälle mit Radius $\frac{\sqrt{1-\delta}}{2}$ im Ball mit Radius $\sqrt{(1 + \delta)/2} + \frac{\sqrt{1-\delta}}{2}$ enthalten.

Volumenvergleich: mit $B_1 \subset \mathbb{R}^m$ ergibt

$$\begin{aligned} N \left(\frac{\sqrt{1-\delta}}{2} \right)^m |B_1| &\leq \left(\sqrt{(1+\delta)/2} + \frac{\sqrt{1-\delta}}{2} \right)^m |B_1| \\ \Rightarrow N &\leq \left[1 + \sqrt{2} \frac{\sqrt{1-\delta}}{\sqrt{1-\delta}} \right]^m \\ \Leftrightarrow m &\geq \frac{\ln N}{\ln(1 + \sqrt{2})} = c_\delta \ln N \end{aligned}$$

Das ist eine notwendige Bed. (untere Schranke) an die Dim. m damit eine lin. Abb $A: \mathbb{R}^N \rightarrow \mathbb{R}^m$ mit (54) ex. kann. (Hier ist $M = N + 1$.)

Vergl. mit hinreichender Bed. aus JLL: $m \geq 72t^{-2} \ln M = C_t \ln M$

$\Rightarrow$ Was die Abh. von $M = N + 1$ angeht, ist die untere Schr. an m in JLL optimal.

Allgemeines Resultat (ohne Beweis):

SATZ 8.2 ([LN17] Larsen & Nelson, 2017). Für alle $2 \leq N, M \in \mathbb{N}$ und

$$\left(\frac{1}{\min\{N, M\}} \right)^{0,4999} < t < 1$$

existierten $\{x_1, \dots, x_M\} =: X \subset \mathbb{R}^N$, s.d. **jede** Abbildung $f: X \rightarrow \mathbb{R}^m$, die

$$(55) \quad (1 - t) \|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2^2 \leq (1 + t) \|x_i - x_j\|_2^2 \quad (i, j \in \overline{M})$$

erfüllt, auch

$$m \geq C_N \frac{\ln M}{t^2}$$

erfüllen muss.

Beachte: Eine Einbettung

$$f: X \rightarrow \mathbb{R}^{\min\{N,M\}}$$

ist trivial zu realisieren.

BEMERKUNG 8.3. An probal. Werkzeug benutzte das JLL die Konzentrationsungl. in S. 7.8

$$\mathbb{P}(|\|Ax\|_2^2 - 1| > t) \leq 2e^{-mt^2/40} \text{ für } t \in (0, 1), \|x\| = 1$$

Also sind sowohl JLL als auch RIP Kinder dieser Ungleichung. Auch Bewise sind ähnlich. RIP braucht aber Zusatzargument: Netz $N \subset \mathbb{S} \cap \mathbb{R}_s^N$. Insofern ist RIP eine Verfeinerung von JLL.

Dies gilt auch konkret im folgenden Sinne:

SATZ 8.4 (RIP $\Rightarrow$ JLL). Seien $x_1, \dots, x_M \in \mathbb{R}^N, \eta, \epsilon \in (0, 1), A \in \mathbb{R}^{m \times N}$ haben die RIP mit $\delta_{2s} \leq \frac{\eta}{4}$ für ein $s \geq 16 \ln(4 \frac{M}{\epsilon})$. Seien $a = (a_1, \dots, a_n)$ ein Vektor von Rademacher-ZV, dann gilt mit W-keit $> 1 - \epsilon$.

$$(1 - \eta)\|x\|_2^2 \leq \|AD_a x\|_2^2 \leq (1 + \eta)\|x\|_2^2 \quad (i \in \overline{M})$$

mit Diag.matrix $Da = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$

Ohne die Randomisierung mit D_A ist die Aussage des Satzes falsch. Wähle dazu $x_1, \dots, x_M \in \ker A$.

KAPITEL 9

CS mit strukturierten zufälligen Matrizen

LINK: VORLESUNGSVIDEO 64 VOM 7.7.25

Bisher: Zufällige Matrizen mit $m \cdot N$ unabhängigen, (sub-)Gaußschen ZV sind (fast) optimale Meßmatrizen für CS.

(Erklärt in Zusammenhang mit hinreichenden und notwendige Schranken an die Anzahl m von Messungen. Abstrakte, theoretische Begründung liefern Abschnitte ?? und 11.)

Nachteil: Diese Matrizen haben keinerlei Struktur.

Struktur ist wichtig bei Anwendungen, wegen:

- In manchen Situationen sind nicht beliebige Matrizen als Meßmatrix realisierbar (zB wenn nur Daten über Fourier-Koeffizienten zugänglich sind.)
- Struktur der Matrix A erlaubt schnelle Algorithmen, z.B. zur Multiplikation Matrix $\times$ Vektor (z.B. FFT). Solche schnellen Multip. sind notwendig, um CS-Problem hoher Dimension effizient lösen zu können.
- Große umstrukturierte Matrizen erfordern viel Speicherplatz ($d = m \cdot N$), stukturierte Matrizen deutlich weniger, denn man muss weniger "Parameter" speichern.

Allerdings: Keine effiziente Methode bekannt, um deterministische Matrizen zu konstruieren. $\rightarrow$ Brauchen "etwas" Zufall.

Wahrscheinlich bestes deterministisches Resultat (von Bourgain, Dilworth, Ford, Konyagin & Kutzaro) bisher hat starke Einschränkungen an Dimensionen.

SATZ 9.1 ([BDF⁺11] Bourgain et al. 2011). *Es gibt absolute Konstanten $\varepsilon > 0$ und $m_0 \in \mathbb{N}$, s.d. es für alle $\mathbb{N} \ni m \geq \varepsilon m_0$ und $m \leq N \leq m^{1+\varepsilon}$, eine explizite Matrix $A \in \mathbb{C}^{m \times N}$ mit RIP-Konstante $\delta_s = m^{-\varepsilon}$ der Ordnung $s = \left\lceil m^{\frac{1}{2}+\varepsilon} \right\rceil$ gibt.*

$$\varepsilon = \epsilon/5$$

$$c_0 \sim 1/10.430 < 1/10.000$$

$$\alpha \sim 1/200$$

$$m > 86.000/c_0 > 860.000.000$$

$$(\gamma = \beta/50)$$

$$\gamma \leq \alpha \sim 1/200 < 1/100$$

$$\gamma \leq 1/(3m) < 1/30.000$$

$$\epsilon = c_0 \gamma - 43\alpha/m \leq 1/300.000.000$$

BEISPIEL 9.2 (Zufällige partielle Fouriertransformation). Fouriermatrix:

$$F_N = \frac{1}{\sqrt{N}} \left(e^{-2\pi i \frac{kl}{N}} \right)_{l,k=0}^{N-1}$$

andere Normierung als in Definition 1.23 in Kapitel 1. $\Rightarrow$ Unitarität: $F_N^{-1} = F_N^*$

Vorteil: $F_N \cdot x$ lässt sich mittels FFT schnell implementieren.

Konstruieren darauf basierend strukturiertes zufälliges $A \in \mathbb{C}^{m \times N}$:

- 1.) Wähle m Zeilen von F_N **zufällig** aus und bilde $(m \times N)$ -Matrix.
- 2.) Normiere die Spalten auf 1.
- 3.) Erhalte Messmatrix $A = \frac{1}{\sqrt{m}} R_T F_N \in \mathbb{C}^{m \times N}$, $T \subset \overline{N}$, $\#T = m$.

Aufgabe: Rekonstruiere sparses $x \in \mathbb{C}_s^N$ aus Ax , d.h. anhand m zufällig ausgewählter Fourierkoeff. $\hat{x}(n_1), \dots, \hat{x}(n_m)$.

Um m Zeilen **zufällig auszuwählen** gibt es 2 Methoden:

- (I) (Gleichverteilung auf allen m -elementigen Teilmengen $\subset \overline{N}$.)
- (II) Wähle $k_1, \dots, k_m$ unabh. und jeweils gleich-verteilt in $\overline{N} \rightarrow$ (eine Zeile kann mehrfach gewählt werden). Im folgenden wird diese Variante benutzt.

In beiden Fällen: Ax kann schnell berechnet werden: $\mathcal{O}(N \ln N)$. A^*y ebenso.

Wie groß muss $m = \#$ Messungen sein?

Notwendige Bedingung:

- Gelfandweiten
 $\Rightarrow m \geq cs \ln \frac{eN}{s}$ für sämtliche Messmatrizen A
- Unschäreferelationen
 $\Rightarrow m \geq cs \ln N$ für (zufällige) partielle Fouriertransformation

Hinreichende Bedingung:

Ist

$$m \geq c \cdot s (\ln s)^3 \ln N$$

so existiert (zufällige) partielle Fouriermatrix, welche s -dünne Signale rekonstruiert (mit hoher Wahrscheinlichkeit)

Unterschied zwischen notwendiger und hinreichender Bed an m ist nur $(\ln s)^3$.

Offene Frage: Eliminierbar in hinreichender Bedingung?

SATZ 9.3 (RIP von zufälliger FFT). Sei $A = \frac{1}{\sqrt{m}} R_T F_N \in \mathbb{C}^{m \times N}$ die zufällige FFT von oben. Es ex. absolute Konstante $C > 0$ mit: Sind

$$(56) \quad \delta \in (0, 1) \text{ und } m \geq C \delta^{-2} s \ln^4(N) \geq C \delta^{-2} s \ln^3(s) \ln(N)$$

erfüllt, dann gilt

$$\mathbb{P} \left(\delta_s \left(\frac{1}{\sqrt{m}} A \right) \leq \delta \right) \geq 1 - N^{-\ln^3(N)}$$

Unschärferelation für inkohärente Basen

LINK: VORLESUNGSVIDEO 65 VOM 7.7.25

DEFINITION 10.1 (Beschränkte Orthonormalsysteme). • Ω Maßraum, $L^2(\Omega)$ Hilbertraum mit Skalarprodukt

- $(\phi_j)_j$ ONS $\Leftrightarrow \langle \phi_j, \phi_k \rangle = \delta_{k,j}$
- Beschränkt: Ex. Konstante K , sodass $\text{esssup}_{x \in \Omega} |\phi_j(x)| \leq K$ für alle j .

Wir beweisen eine Unschärferelation für die diskrete FFT F_N wie oben:

$$\hat{x} = F_N x, x = F_N^* \hat{x}, F_N^{-1} = F_N^*$$

in der Sprache der “inkohärenten Basen”.

DEFINITION 10.2. Seien $U = (u_0|u_1|\dots|u_{N-1})$, $W = (w_0|\dots|w_{N-1}) \in \mathbb{C}^{N \times N}$ unitär. U, W heißen K -kohärent, falls

$$\sqrt{N} \max_{n,l=0,\dots,N-1} |\langle u_l, w_n \rangle| \leq K.$$

BEMERKUNG 10.3. Diskussion.

- (a) Unitarität $\Rightarrow (u_j)_0^{N-1}$ ist ONB, $(w_j)_0^{N-1}$ auch
 $\Rightarrow \sqrt{N} \max_{l,k} |\langle u_k, w_l \rangle| \stackrel{\text{CS}}{\leq} \sqrt{N} \|u\|_2 \|w\|_2 \leq \sqrt{N}$,
 $\Rightarrow K \leq \sqrt{N}$ universelle Schranke.

- (b) Sei $U = F_N$, $w = \text{Id} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{C}^{N \times N}$

$$\Rightarrow |\langle u_l, w_k \rangle| = \left| \frac{1}{\sqrt{N}} e^{2\pi i \frac{lk}{N}} \right| = \frac{1}{\sqrt{N}} \Rightarrow K = 1$$

$K = 1$ ist das kleinst mögliche K .

Demnach sind $U = F_N$ und $W = \text{Id}_N$ sind *optimal inkohärent*.

SATZ 10.4. Seien $V, W \in \mathbb{C}^{N \times N}$ zwei K -kohärente Matrizen. Sei $y \in \mathbb{C}^N \setminus \{0\}$ und $x, z \in \mathbb{C}^N$ mit $y = Vx = Wz$. Dann gelten

$$(57) \quad \|x\|_0 \|z\|_0 \geq \frac{N}{K^2}$$

$$(58) \quad \|x\|_0 + \|z\|_0 \geq \frac{2\sqrt{N}}{K}$$

Diese Unschärferelation besagt: Es ist nicht möglich, dass sowohl x als auch $\hat{x}$ auf wenigen Punkten konzentriert sind.

BEWEIS. $Vx = Wz \Rightarrow x = V^*Wz$. Dann

$$|x_k| = |(V^*Wz)_k| = \left| \sum_l (V^*W)_{k,l} z_l \right| \leq \sum_l \underbrace{|\langle w_l, v_k \rangle|}_{\max_{l,k} |\langle w_l, v_k \rangle| \leq \frac{k}{\sqrt{N}}} |z_l| \leq \frac{k}{\sqrt{N}} \|z\|_1$$

$$\Rightarrow \|x\|_1 \leq \|x\|_0 \frac{K}{\sqrt{N}} \|z\|_1.$$

$$\text{Analog: } z = W^*Vx \Rightarrow \|z\|_1 \leq \|z\|_0 \frac{K}{\sqrt{N}} \|x\|_1.$$

Multipliziere Ungl. miteinander und dividiere $\|x\|_1 \|z\|_1$

$$\begin{aligned} \Rightarrow 1 &\leq \|x\|_0 \|z\|_0 \frac{K^2}{N} \\ \Leftrightarrow \|x\|_0 \|z\|_0 &\geq \frac{N}{K^2} \\ \Leftrightarrow \sqrt{\|x\|_0 \|z\|_0} &\geq \frac{\sqrt{N}}{K} \\ \Rightarrow \frac{\|x\|_0 + \|z\|_0}{2} &\geq \sqrt{\dots} \geq \frac{\sqrt{N}}{K} \end{aligned}$$

Denn: arithmetisches Mittel $\geq$ geometrisches Mittel □

BEISPIEL 10.5. Für $V = F_N$, $W = \text{Id}$ erhalten wir:

$$\|x\|_0 \|\hat{x}\|_0 \geq N, \|x\|_0 + \|\hat{x}\|_0 \geq 2\sqrt{N}$$

Erfüllt beispielsweise für $\|x\|_0 \|\hat{x}\|_0 = \sqrt{N}$.

Die Schranke (Konstante) aus der Unschärfereferelation im letzten Satz kann man nicht verbessern, wie man anhand von “Kamm-Vektoren” (delta-train) sieht.

LINK: VORLESUNGSVIDEO 66 VOM 7.7.25

SATZ 10.6. Sei $N = n^2$ und $x \in \mathbb{C}^N$, der Vektor mit den Einträgen

$$x_i = \begin{cases} 1, & \text{falls } i = 0 \pmod n \\ 0, & \text{sonst} \end{cases}$$

Dann gilt:

$$\hat{x} = x \text{ und } \|x\|_0 = \|\hat{x}\|_0 = \sqrt{N} = n$$

BEWEIS. Für $j \in \{0, \dots, N-1 = n^2-1\}$ gilt

$$\begin{aligned} \hat{x}_j &= \frac{1}{\sqrt{N}} \sum_{l=0}^{n^2-1} x_l e^{2\pi i \frac{lj}{N}}, & \text{Betrag nur bei Vielfachem } l = k \cdot n \\ &= \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i \frac{kj}{n}} = \begin{cases} 1, & \text{falls } j = 0 \pmod n \\ 0, & \text{sonst} \end{cases} \end{aligned}$$

$$\Rightarrow \hat{x} = x, \text{ insbes.: } \|x\|_0 = \|\hat{x}\|_0 = n = \sqrt{N}. \quad \square$$

BEMERKUNG 10.7. Warum ist es wichtig, bei der Meßmatrix $A = \text{“zufällige partielle FT”}$ die Zeilen tatsächlich zufällig auszuwählen?

Beispiel der Kamm-Vektoren zeigt, warum manche deterministische sampling Mengen $T \subset \{0, \dots, N-1\}$ für die Rekonstruktion von sparsen Signalen aus den Fourierkoeffizienten $(\hat{x}_{j \in T})$ nicht geeignet sind.

Sei $N = n^2$ und x Kamm-Vek. wie oben. Wähle

$$T = \{0, \dots, n^2 - 1\} \setminus \underbrace{\{0, n, 2n, \dots, n^2 - n\}}_{\text{supp } x}$$

$\Rightarrow \hat{x}|_T \equiv 0$ für detlta-train, also Meßvektor $y = Ax = 0$.

Jeder “brauchbare”, insb. jeder lineare Rekonstruktionsalgorithmus würde $x^* \equiv 0$ ausgeben, falls Messung $y \equiv 0$.

Mit der Sampling Menge T kann ich x nicht vom Nullvektor unterscheiden.

Beachte Mächtigkeiten: $s = n = \|x\|_0$ und $m = \#T = n^2 - n = s^2 - s$

Also: In diesem Fall von T reichen nicht einmal $m = s^2 - s$ Messungen zur Rekonstruktion aus.

- Die Sampling Menge hat zuviel Struktur.
“Gute Sampling Mengen” besitzen möglichst wenig additive Struktur.
- Zufall ist einfachster Weg, um solche Strukturen bei Matrizen (meistens) zu vermeiden.

LINK: VORLESUNGSVIDEO 67 VOM 8.7.25

SATZ 10.8 (RIP von inkohärenten Basen). Sei $A = \frac{1}{\sqrt{m}} R_T W^* V \in \mathbb{C}^{m \times N}$ aus K -kohärenten unitären U, W generierte Zufalls-Abtastmatrix mit $\#T = m$. Es ex. absolute Konstante $C > 0$ mit: Sind

$$(59) \quad \delta \in (0, 1) \text{ und } m \geq CK^2 \delta^{-2} s \ln^4(N)$$

erfüllt, dann gilt

$$\mathbb{P} \left(\delta_s(A) \leq \delta \right) \geq 1 - N^{-\ln^3(N)}$$

Beweis benutzt u.a. Dudleys Ungleichung für Erwartungswerte von Suprema von sub-Gauß-prozessen.

DEFINITION 10.9. Für einen stochastischen Prozess $X_t: (\Omega, \mathbb{P}) \rightarrow \mathbb{R}$, $t \in T$ definieren wir die Pseudometrik

$$d_T(s, t) := \left(\mathbb{E} |X_s - X_t|^2 \right)^{1/2}, \quad s, t \in T.$$

Die Überdeckungszahl $N(T, d, \varepsilon)$ ist definiert als die kleinste ganze Zahl N , sodass es eine Teilmenge $F \subseteq T$ mit $\text{card}(F) = N$ gibt und $\min_{s \in F} d(t, s) \leq \varepsilon$ für alle $t \in T$ gilt.

SATZ 10.10 (Dudleys Ungleichung). Sei X_t , $t \in T$, ein zentrierter subgaussischer Prozess. Dann gilt

$$\mathbb{E} \left[\sup_{t \in T} |X_t| \right] \leq 4\sqrt{2} \int_0^{\Delta(T)/2} \sqrt{\log(2N(T, d_T, u))} du,$$

wobei $\Delta(T) := \sup_{s, t \in T} d_T(s, t)$ den Durchmesser von T bezeichnet. Ist das Integral recht endlich, so existiert eine Version des Prozesses mit f.s. beschränkten und uniform stetigen Pfaden auf (T, d_T) .

Gelfand-Zahlen

LINK: VORLESUNGSVIDEO 68 VOM 8.7.25

- Kommen aus der geometrischen Funktionalanalysis und der Approximationstheorie.
- Beschreiben Approximations-Eigenschaft von Operatoren zwischen BR (Banachräumen) und Fehler von Approximations-Algorithmen.

Siehe z.B. [GG84] oder [CP88].

Hier:

- Analysiere Verbindung zwischen Gelfandzahlen von Einbettungsoperatoren und CS.
- Zeigt: Effizienz der Basisjagt kann i.w. durch keinen anderen Rekonstruktions-/Approximations-Algorithmus verbessert werden.

1. Definition Gelfand-Zahlen, Gelfand-Weiten, und Rekonstruktionsfehler

DEFINITION 11.1. Sei X ein Vektorraum und $M \subset X$ UVR. Dann ist die *Kodimension* von M def. als

$$\begin{aligned} \operatorname{codim} M &= \dim(\underbrace{X/M}_{\text{Quotientenraum}}) \\ &= \begin{cases} \dim X & \text{falls } \dim X < \infty \end{cases} - \dim M \end{aligned}$$

Additivität: $\operatorname{codim}(M_1) = k_1$, $\operatorname{codim}(M_2) = k_2$,

$$\max(k_1, k_2) \leq \operatorname{codim}(M_1 \cap M_2) \leq k_1 + k_2$$

Duale Darstellung: $M \subset X$ UVR

$$\operatorname{codim} M = m \Leftrightarrow \exists \text{ lin. unabh. } \lambda_1, \dots, \lambda_m \in X^*$$

so dass

$$\begin{aligned} M &= \{x \in X \mid \lambda_1(x) = 0 = \lambda_2(x) = \dots = \lambda_m(x)\} \\ &= \ker A, \text{ wobei } Ax = \begin{pmatrix} \lambda_1(x) \\ \vdots \\ \lambda_m(x) \end{pmatrix} \end{aligned}$$

X^* Dualraum und $\lambda_1, \dots, \lambda_m: X \rightarrow \mathbb{R}$ sind (stetige) lineare Funktionale

DEFINITION 11.2. Sei $T : X \rightarrow Y$ lin. Operator zwischen BR und $k \in \mathbb{N}$. Die k -te Gelfandzahl von T ist

$$c_k(T) = \inf_{\substack{M \subset X \text{ UVR} \\ \text{codim } M < k}} \sup_{\substack{x \in M \\ \|x\|=1}} \|Tx\| = \inf_{\substack{M \subset X \\ \text{codim } M < k}} \|T|_M\|$$

BEMERKUNG 11.3 (Eigenschaften der Gelfand-Zahlen).

- 1) $\|T\| = c_1(T) \geq c_2(T) \geq \dots \geq 0$
- 2) $c_{k+l-1}(T+S) \leq c_k(T) + c_l(S)$
- 3) $c_k(RST) \leq \|R\|c_k(S)\|T\|$
- 4) $\text{rank } T < k \Rightarrow c_k(T) = 0$
- 5) $c_k(\text{Id}) = 1$ für $\text{Id}: X \rightarrow Y = X$ mit $\dim X \geq k$.

Wir zeigen hier nur die zweite Eigenschaft.

Sei $\epsilon > 0$

$$\begin{aligned} x \in M \subset X, \text{codim } M < k &\Rightarrow \|Sx\| \leq (1 + \epsilon)c_k(S)\|x\| \\ x \in N \subset X, \text{codim } N < l &\Rightarrow \|Tx\| \leq (1 + \epsilon)c_l(S)\|x\| \end{aligned}$$

Für $x \in M \cap N$

$$\|(S+T)x\| \leq \|Sx\| + \|Tx\| \leq (1 + \epsilon)[c_k(S) + c_l(T)]\|x\|$$

Nun gilt:

$$\begin{aligned} \text{codim}(M \cap N) &\leq \text{codim } M + \text{codim } N \leq k - 1 + l - 1 \\ &< k + l - 1 \end{aligned}$$

Damit

$$c_{k+l-1}(S+T) \leq (1 + \epsilon)[c_k(S) + c_l(T)]$$

Nun Limes $\epsilon \rightarrow 0$.

DEFINITION 11.4 (Gelfandweite/Gelfand width). Sei Y ein normierter Raum, $C \subset Y$

$$d^m(C, Y) := \inf_{\substack{M \subset Y \\ \text{codim } M \leq k}} \sup_{x \in C \cap M} \|x\|$$

heißt m -te Gelfandweite .

$$d^0(C, Y) = \sup_{x \in C} \|x\| \leq \text{diam } C$$

Ist $T : X \rightarrow Y$ linear, stetig, $C = T(B_1 \subset X) \subset Y$, dann ist

$$d^m(C, Y) = c_{m-1}(T).$$

DEFINITION 11.5 (Fehler in der Approx.-Th.). Sei $C \subset Y$. Definiere kleinstmöglicher Rekonstruktionsfehler bei m linearen Messungen:

$$E^m(C, Y) := \inf_{\mathcal{A} = \Phi \circ L_m} \sup_{x \in C} \|x - \mathcal{A}x\|_Y = \inf_{\Phi: \mathbb{R}^m \rightarrow Y} \inf_{L_m = (\lambda_1, \dots, \lambda_m) \in (Y^*)^m} \sup_{x \in C} \|x - \Phi \circ L_m x\|_Y,$$

wobei

- $x \in Y$ Signal,
- $L_m : Y \rightarrow \mathbb{R}^m$ stetige, lineare Mess/Informationsabbildung,

- $\Phi: \mathbb{R}^m \rightarrow Y$ beliebige Rekonstruktionsabbildung.

Die Abb. L_m hat die Form

$$L_m(x) = (\lambda_1(x), \dots, \lambda_m(x))$$

λ_i lineare Funktionale $\in Y^*$. Normalerweise sind λ_i fest.

BEMERKUNG. *Aber:* Bei der Messung kann man auch adaptiv vorgehen und die j -te Messung in Abhängigkeit von den bisherigen Messdaten $\Lambda_1, \dots, \Lambda_{j-1}$ designen

$$\lambda_j(x) = \Lambda_{j, \Lambda_1(x), \dots, \Lambda_{j-1}(x)}(x) = \Lambda_j[\Lambda_1(x), \dots, \Lambda_{j-1}(x)](x)$$

In diesem Fall nennt man die Messungen $\lambda_1, \dots, \lambda_m$ adaptiv.

Achtung, sie sind linear im letzten Argument (x), aber nicht notwendigerweise bzgl. der als Parameter übergebenen Daten $\lambda_1(x), \dots, \lambda_{j-1}(x)$.

D.h. für jeden Vektor $p \in \mathbb{R}^{j-1}$ ist die Abbildung $\Lambda_j[p]: Y \rightarrow \mathbb{R}$ linear und stetig, aber die Abbildung $\mathbb{R}^{j-1} \ni p \mapsto \Lambda_j[p](x)$ ist nicht notwendigerweise linear.

DEFINITION 11.6 (Adaptiver Fehler in der Approx.-Th.). Für adaptive $N_m: Y \rightarrow \mathbb{R}^m$ wie in der Bemerkung definieren wir den Fehler von $\mathcal{A}$ als

$$E_{\text{ada}}^m(C, Y) := \inf_{\Phi: \mathbb{R}^m \rightarrow Y} \inf_{N_m: Y \rightarrow \mathbb{R}^m \text{ adaptiv}} \sup_{x \in C} \|x - \mathcal{A}x\|_Y$$

Da jedes m -Tupel von linearen Messung auch als m adaptive Messungen aufgefasst werden kann (Ignoriere bisherige Messwerte!), gilt

$$(60) \quad \left\{ \begin{array}{c} \Lambda_1, \dots, \Lambda_m \in Y^* \\ \text{adaptive} \end{array} \right\} \supset \{ \Lambda_1, \dots, \Lambda_m \in Y^* \text{ beliebig} \}$$

bzw.

$$(61) \quad \{N_m: Y \rightarrow \mathbb{R}^m \text{ adaptiv}\} \supset \{L_m: Y \rightarrow \mathbb{R}^m \text{ stetig \& linear}\}$$

Die erhöhte Flexibilität bei adaptiven Messungen wecken die Erwartung, dass der zugehörige Approximationsfehler $E_{\text{ada}}^m(C, Y)$ kleiner ist als der $E^m(C, Y)$ bei herkömmlichen linearen Messungen.

SATZ 11.7. Sei Y ein normierter Raum, sei $C \subset Y$. Dann gilt

- (a) $E_{\text{ada}}^m(C, Y) \leq E^m(C, Y)$
- (b) Falls C (Punkt-)symmetrisch ist (d.h. $C = -C$), dann gilt

$$d^m(C, Y) \leq E_{\text{ada}}^m(C, Y)$$

- (c) Falls zusätzlich $C + C \subset a \cdot C$, $a > 0$, dann gilt

$$E^m(C, Y) \leq a \cdot d^m(C, Y)$$

wobei $a > 0$ positive Konstante

BEMERKUNG 11.8. Falls C so ist, dass $C + C \subset a \cdot C$ und $C = -C$, gilt insgesamt

$$d^m(C, Y) \leq E_{\text{ada}}^m(C, Y) \leq E^m(C, Y) \leq a \cdot d^m(C, Y)$$

Kann ein solches C existieren?

Ja, z.B. $C = T(B_X)$, wobei $B_X = B_1(0) \subset X$ und $T: X \rightarrow Y$ linearer Operator. Wegen

$B_X = -B_X \Rightarrow C = -C$. Wegen $B_X + B_X = 2B_X$ auch $T(B_X) + T(B_X) = T(2B_X) = 2T(B_X)$, also $\Rightarrow a = 2$.

BEWEIS VON SATZ 11.7. Teil (a) ist wegen (61) und der Definitionen klar.

Bei Teil (b) ist unter Voraussetzung $C = -C$ z.z.:

$$\inf_{\substack{M \subset Y \\ \text{codim } M \leq m}} \sup_{x \in C \cap M} \|x\| \leq \inf_{\substack{\mathcal{A} = \Phi \circ N_m \\ N_m \text{ adaptiv}}} \sup_{x \in C} \|x - \mathcal{A}x\|$$

Wir betrachten N_m adaptiv, d.h.

$$\begin{aligned} \lambda_1(x) &= \Lambda_1(x) \\ \lambda_2(x) &= \Lambda_{2, \Lambda_1(x)}(x) \\ &\vdots \\ \lambda_m(x) &= \Lambda_{m, \Lambda_1(x), \Lambda_2(x), \dots, \Lambda_{m-1}(x)}(x) \end{aligned}$$

Wir definieren

$$L_m x = \left(\Lambda_1(x), \Lambda_{2,0}(x), \dots, \Lambda_{m,0,\dots,0}(x) \right)$$

(d.h. $\lambda_1, \dots, \lambda_m$ so designt als ob vorherige Messungen alle $0, \dots, 0$)

$$M = \ker L_m = \ker \Lambda_1 \cap \ker \Lambda_{2,0} \cap \dots \cap \ker \Lambda_{m,0,\dots,0}.$$

Höchstens um m Dimensionen reduziert.

$$\Rightarrow \text{codim}(M) \leq m$$

$$\Rightarrow d^m(C, Y) \leq \sup_{x \in C \cap \ker L_m} \|x\|$$

Nur noch zz.

$$\sup_{x \in C \cap \ker L_m} \|x\| \leq \sup_{x \in C} \|x - \mathcal{A}x\|$$

für jeden adaptiven Algorithmus $\mathcal{A} = \Phi \circ N_m$.

Beachte, dass für $x \in M$ gilt $L_m(X) = (0, 0, \dots, 0)^T$, damit

$$\Lambda_1(x) = 0, \Lambda_{2, \Lambda_1(x)=0}(x) = 0, \Lambda_{3, \Lambda_1(x)=0, \Lambda_2(x)=0} = 0, \dots$$

insgesamt $N_m(x) = 0$ und $\mathcal{A}(x) = \Phi(N_m(x)) = \Phi(0)$.

Wir betrachten $x \in C \cap M = C \cap \ker L_m$

$$\sup_{y \in C} \|y - \mathcal{A}y\| \geq \|x - \mathcal{A}x\| = \|x - \Phi(N_m(x))\| = \|x - \Phi(0)\|$$

Wegen $-x \in C \cap \ker L_m$

$$\sup_{y \in C} \|y - \mathcal{A}y\| \geq \|-x - \Phi(N_m(-x))\| = \|-x - \Phi(0)\|$$

Dann:

$$\begin{aligned} \|x\| &= \left\| \frac{1}{2}(x - \Phi(0)) + \frac{1}{2}(x + \Phi(0)) \right\| \\ &\leq \frac{\|x - \Phi(0)\| + \|x + \Phi(0)\|}{2} \leq \sup_{y \in C} \|y - \mathcal{A}y\| \end{aligned}$$

Beide Ungl. zusammen:

$$\begin{aligned} \Rightarrow d^m(C, Y) &\leq \sup_{x \in C \cap \ker L_m} \|x\| \leq \sup_{y \in C} \|y - \mathcal{A}y\| \text{ für jeden adaptiven Algorithmus} \\ \Rightarrow d^m(C, Y) &\leq E_{\text{ada}}^m(C, Y) \end{aligned}$$

Bei Teil (c) wird zusätzlich $\mathcal{C} + \mathcal{C} \subset a\mathcal{C}$ angenommen.

Sei $M \subset Y$ mit $\text{codim } M \leq m$. Wähle $L = L_m : Y \rightarrow \mathbb{R}^m$ linear mit $M = \ker L$.

Für $y \in L(\mathcal{C})$ wähle $\Phi(y) \in \mathcal{C} \cap L^{-1}(y)$ beliebig und setze $\mathcal{A} = \Phi \circ L$.

$$\begin{aligned} E^m(\mathcal{C}, Y) &\leq \sup_{x \in \mathcal{C}} \|x - Ax\| = \sup_{x \in \mathcal{C}} \|x - \Phi(L(x))\| \\ &\leq \sup_{x \in \mathcal{C}} \sup_{z \in \mathcal{C} \cap L^{-1}(L(x))} \|x - z\| \end{aligned}$$

$$x \in \mathcal{C}, z \in \mathcal{C} \cap L^{-1}(L(x)) \Rightarrow x - z \in \mathcal{C} + (-\mathcal{C}) \subset a \cdot \mathcal{C}$$

$$L(x - z) = L(x) - L(z) = L(x) - L(L^{-1}(L(x))) = 0 \quad \Rightarrow x - z \in \ker L = M$$

$$\Rightarrow E^m(\mathcal{C}, Y) \leq \sup_{u \in a \cdot \mathcal{C} \cap M} \|u\| = \sup_{v \in \mathcal{C} \cap M} \|av\| = a \sup_{v \in \mathcal{C} \cap M} \|v\|$$

Dies gilt für jedes $M \subset Y$ mit $\text{codim } M \leq m$:

$$\inf_{\text{codim } M \leq m} E^m(\mathcal{C}, Y) \leq \inf_{\text{codim } M \leq m} a \sup_{v \in \mathcal{C} \cap M} \|v\| \leq a \cdot d^m(\mathcal{C}, Y)$$

□

2. Zusammenhang zwischen Gelfand-Weiten und CS

LINK: VORLESUNGSVIDEO 69 VOM 14.7.25

Um abstrakte Abschätzungen zu Rekonstruktionsalgorithmen mit CS in Zusammenhang zu bringen, betrachten wir nun speziell

$$\begin{aligned} B_1^N \subset \ell_1^N = \ell_1(\bar{N}) &=: X && \text{Einheitskugel in einem BR} \\ \ell_p^N = \ell_p(\bar{N}) &=: Y && \text{ein BR} \\ T: \ell_1^N = X \rightarrow Y = \ell_p^N, && \mathcal{C} := T(B_1^N) \end{aligned}$$

Wie bereits in Bemerkung 11.8 diskutiert

$$\begin{aligned} -B_1^N &= B_1^N && \Rightarrow -\mathcal{C} = \mathcal{C} \\ B_1^N + B_1^N &= 2B_1^N && \Rightarrow \mathcal{C} + \mathcal{C} \subset 2\mathcal{C} \end{aligned}$$

Daher liefert Satz 11.7

$$d^m(\mathcal{C}, Y) \leq E_{\text{ada}}^m(\mathcal{C}, Y) \leq E^m(\mathcal{C}, Y) \leq a \cdot d^m(\mathcal{C}, Y)$$

Da alle Größen gleich sind (bis auf einen Faktor zwei), reicht es eine obere und untere Schranke (als Funktion der Parameter m, N, p) an jeweils eine davon herzuleiten.

Um „nicht zu vergessen“ wofür $\mathcal{C}$ steht schreiben wir stattdessen B_1^N , weil es ja tatsächlich diese Menge, nur in einem anderen Raum ist.

Zunächst

2.1. Obere Schranke an $E^m(B_1^N, \ell_p^N)$. Vorüberlegungen zu ‚einfachen‘ Fällen:

Ist $p = 1$ und $m < N$, so folgt $d^m(B_1^N, \ell_1^N) = 1$ mit Bemerkung 11.3 oder nach Definition des Einheitsballs.

Ist $p > 2$, so folgt $\|x\|_p \leq \|x\|_2$ und damit

$$d^m(B_1^N, \ell_p^N) \leq d^m(B_1^N, \ell_2^N)$$

Demnach konzentrieren wir uns auf den Parameterbereich $p \in (1, 2]$.

SATZ 11.9. *Es existiert eine Konstante $C_0 > 0$, so dass für alle $p \in (1, 2]$ und $m, N \in \mathbb{N}$ mit $m < N$ gilt*

$$d^m(B_1^N, \ell_p^N) \leq C_0 \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p}$$

Wir können

$$C_0 = 4\sqrt{3}c' = 4\sqrt{3}g^{-1}(16 \cdot 144 \cdot (1 + \ln 9))$$

wählen, wobei die Funktion g in Bemerkung 11.11 definiert ist.

Bemerke: Hier kommt kein s vor.

LINK: VORLESUNGSVIDEO 70 VOM 14.7.25

Für den Beweis benötigen wir ein Lemma aus der Analysis einer Veränderlichen:

LEMMA 11.10. *Seien $m, N, s \in \mathbb{N}$ mit $s, m \leq N$.*

(a) *Gilt $m \geq s$ und für ein $C > 0$*

$$(62) \quad m \geq C \cdot s \cdot \ln \left(\frac{eN}{s} \right)$$

so folgt

$$(63) \quad m \geq C \cdot s \cdot \ln \left(\frac{eN}{m} \right)$$

(b) *Gilt für eine $C \geq e$ Ungleichung (63) so folgt*

$$(64) \quad m \geq c'' \cdot s \cdot \ln \left(\frac{eN}{s} \right) \quad \text{mit } c'' = \frac{C}{1 + \ln C}$$

Also sind die Bedingungen (62) und (63) „fast“ äquivalent.

BEWEIS. Zu (a):

$$m \geq s \Rightarrow \frac{1}{s} \geq \frac{1}{m} \text{ und } \ln \text{ ist isoton}$$

Zu (a): Die Annahme ist

$$m \geq C \cdot s \cdot \ln \left(\frac{eN}{m} \frac{s}{s} \right) \quad \text{wobei } C \geq e$$

woraus

$$(65) \quad m \geq C \cdot s \cdot \ln \left(\frac{eN}{s} \right) + C \cdot \frac{m}{m} s \cdot \ln \left(\frac{s}{m} \right)$$

folgt. Annahme impliziert auch:

$$0 < \frac{s}{m} \leq \frac{1}{C \ln \underbrace{\frac{eN}{m}}_{\geq e}} \leq \frac{1}{C} \leq \frac{1}{e}$$

Nun ist die Funktion $f: (0, \infty) \rightarrow \mathbb{R}, x \mapsto x \ln x$ stetig, antiton auf $(0, 1/e]$ und isoton auf $(1/e, \infty)$. Also

$$\frac{s}{m} \ln \frac{s}{m} = f\left(\frac{s}{m}\right) \geq f\left(\frac{1}{C}\right) = \frac{1}{C} \ln\left(\frac{1}{C}\right) = -\frac{\ln C}{C}$$

Einsetzen in (65) ergibt:

$$m \geq C \cdot s \cdot \ln\left(\frac{eN}{s}\right) - \emptyset \cdot m \cdot \left(\frac{\ln C}{\emptyset}\right)$$

$\Leftrightarrow$

$$m(1 + \ln C) \geq C \cdot s \cdot \ln\left(\frac{eN}{s}\right)$$

$\Leftrightarrow$

$$m \geq g(C) \cdot s \cdot \ln\left(\frac{eN}{s}\right) \quad \text{falls wir setzen } g(C) := \frac{C}{1 + \ln C}$$

□

BEMERKUNG 11.11 (Analysiere Funktion g).

$$g: [1, \infty) \rightarrow \mathbb{R}, g(x) = \frac{x}{1 + \ln x}$$

ist glatt auf $(1, \infty)$, stetig auf $[1, \infty)$ und streng isoton wegen $g'(x) = \frac{\ln x}{(1 + \ln x)^2} > 0$. Wegen $\lim_{x \rightarrow \infty} g(x) = \infty$ ist

$$g: [1, \infty) \rightarrow [g(1), \infty) = [1, \infty)$$

bijektiv und damit die Inverse $G := g^{-1}$ wohldefiniert.

Wir brauchen später folgenden Vergleichswert:

$$\begin{aligned} e &< g^{-1}(16 \cdot 144 \cdot (1 + \ln 9)) \\ \Leftrightarrow \frac{e}{2} &= \frac{e}{1 + \ln e} = g(e) < 16 \cdot 144 \cdot (1 + \ln 9) \end{aligned}$$

LINK: VORLESUNGSVIDEO 71 VOM 14.7.25

BEWEIS VON SATZ 11.9. $\|x\|_p \leq \|x\|_1$ für $p \geq 1$ impliziert

$$d^m(B_1^N, \ell_p^N) \leq d^0(B_1^N, \ell_p^N) = 1 = 1^{1-1/p}$$

Falls $m \leq c' \ln \frac{eN}{m}$ (tatsächlich wählen wir $c' = g^{-1}(16 \cdot 144 \cdot (1 + \ln 9))$, insbesondere $c' > 1$) so folgt

$$1^{1-1/p} \leq \left(c' \frac{\ln \frac{eN}{m}}{m} \right)^{1-1/p}$$

Somit insgesamt für ‚kleine‘ m und $p \geq 1$:

$$d^m(B_1^N, \ell_p^N) \leq \min \left\{ 1, c' \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p} \leq c'^{1-1/p} \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p} \leq c' \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p}$$

wie gewünscht. Wir haben an diese Stelle die Abschätzung ‚verkompliziert‘ damit sie zu derjenigen im kompletären Bereich $m > c' \ln \frac{eN}{m}$ passt, der auch viel spannender ist. Setze

$$(66) \quad s := \left\lfloor \frac{m}{c' \ln \frac{eN}{m}} \right\rfloor \quad \text{womit} \quad \frac{m}{2c' \ln \frac{eN}{m}} < s \leq \frac{m}{c' \ln \frac{eN}{m}}$$

Nun gibt Lemma 11.10 (b) aufgrund von $c' \geq e$

$$(67) \quad m \geq c'' s \ln \frac{eN}{s} \quad \text{mit } c'' := \frac{c'}{1 + \ln c'} = g(c') = 16 \cdot 144 \cdot (1 + \ln 9)$$

Nun wenden wir Satz 7.9 mit den Parametern

$$(68) \quad s \in \mathbb{N}, s \leq N, \epsilon = \frac{2}{e} \in (0, 1), \delta_{2s} \leq \delta = \frac{1}{2}$$

Diese Wahl impliziert mit $C = 144 \cdot (1 + \ln 9)$

$$\begin{aligned} & C\delta^{-2} \left(2s \ln \frac{eN}{s} + \ln \frac{2}{\epsilon} \right) \\ &= C4 \left(2s \ln \frac{eN}{s} + \ln e \right) \\ &\leq 4C \left(2s \ln \frac{eN}{s} + s \ln \frac{eN}{s} \right) \\ &= 12C s \ln \frac{eN}{s} < m \quad \text{wegen (67)} \end{aligned}$$

Damit ist die Implikation von Satz 7.9

$$\mathbb{P} \left(\delta_{2s}(A) \leq \frac{1}{2} \right) \geq 1 - \epsilon = 1 - \frac{2}{e} > 0$$

Insbesondere existiert zumindest eine ‚zufällig ausgewürfelte‘ Gaußmatrix $A = A^\omega \in \mathbb{R}^{m \times N}$ mit

$$\delta_{2s}(A) \leq \frac{1}{2} < \frac{4}{\sqrt{41}} \approx 0,62 \dots$$

INTERMEZZO. Damit sogar

$$\delta_{2s}(A) \leq \frac{1}{2} < \frac{4}{\sqrt{41}} \approx 0,62 \dots$$

und es gilt für Rekonstruktionsalgorithmus $\mathcal{A}_m$ mittels Basisjagd (d.h. $\ell - 1$ -Optimierungsproblem (P_1)) aufgrund der Messdaten von Messmatrix A

$$\|x - \mathcal{A}_m(x)\|_p = \|x - x^\# \|_p \leq \frac{c}{s^{1-1/p}} \sigma_s(x)_1 \leq \frac{c}{s^{1-1/p}} \|x\|_1$$

Da dies ein zulässiger Rekonstruktionsalgorithmus ist und $\|x\|_1 \leq 1$ für $x \in B_1^n$

$$d^m(B_1^N, \ell_p^N) \leq E^m(B_1^N, \ell_p^N) \leq \frac{c}{s^{1-1/p}}$$

und mit (66) folgt weiter

$$\leq \frac{c}{\left(\frac{m}{2c' \ln \frac{eN}{m}}\right)^{1-1/p}} \leq \left(\frac{2c' c^{p/p-1} \ln \frac{eN}{m}}{m}\right)^{1-1/p}$$

was schon eine Abschätzung von richtigem Typ ist, aber nicht so präzise, wie gewünscht.

Stattdessen gehen wir folgendermassen vor und

Ahmen dabei Teil des Beweises von Satz 6.1 nach.
Zerlege

$$\bar{N} = \bigcup_{k \geq 0}^{\circ} S_k$$

in Teilmengen der Mächtigkeit $\#S_k = s$, bis auf die letzte, so dass

- S_0 enthält die Indizes der größten Werte $|x_i|$
- S_1 enthält die Indizes der nächstgrößten Werte $|x_i|$
- etc.

Damit

$$|x_j| \geq |x_i| \quad \text{für } j \in S_{k-1}, i \in S_k, k \geq 1$$

Lemma 6.2 besagt dann

$$\|x_{S_k}\|_2 \leq \frac{1}{\sqrt{s}} \|x_{S_{k-1}}\|_1, \quad k \geq 1$$

Für $x \in M := \ker A$ folgt:

$$\begin{aligned}
\|x\|_p &\leq \sum_{k \geq 0} \|x_{S_k}\|_p \\
\text{Hölder} \quad &\leq \sum_{k \geq 0} s^{1/p-1/2} \|x_{S_k}\|_2 \\
\text{RIP} \quad &\leq \sum_{k \geq 0} s^{1/p-1/2} \frac{1}{\sqrt{1-\delta}} \|Ax_{S_k}\|_2 \\
&= \frac{s^{1/p-1/2}}{\sqrt{1-\delta}} \left[\|Ax_{S_0}\|_2 + \sum_{k \geq 1} \|Ax_{S_k}\|_2 \right] \\
&= \frac{s^{1/p-1/2}}{\sqrt{1-\delta}} \left[- \sum_{k \geq 1} \|Ax_{S_k}\|_2 + \sum_{k \geq 1} \|Ax_{S_k}\|_2 \right] \\
\Delta\text{-Ungl} \quad &\leq \frac{s^{1/p-1/2}}{\sqrt{1-\delta}} 2 \sum_{k \geq 1} \|Ax_{S_k}\|_2 \\
\text{RIP} \quad &\leq 2s^{1/p-1/2} \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{k \geq 1} \|x_{S_k}\|_2 \\
&\leq 2s^{1/p-1/2} \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{k \geq 1} \frac{1}{\sqrt{s}} \|x_{S_{k-1}}\|_1 \\
&\leq \frac{2}{s^{1-1/p}} \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{k \geq 1} \|x_{S_{k-1}}\|_1
\end{aligned}$$

Wegen der Zerlegung folgt $\sum_{k \geq 1} \|x_{S_{k-1}}\|_1 = \|x\|_1$ und mit (66)

$$\begin{aligned}
\|x\|_p &\leq \frac{2}{s^{1-1/p}} \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \|x\|_1 \\
(66) \quad &\leq 2 \left(\frac{2c' \ln \frac{eN}{m}}{m} \right)^{1-1/p} \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \|x\|_1 \\
(\delta = \frac{1}{2}) \quad &\leq 4\sqrt{3} \left(\frac{c' \ln \frac{eN}{m}}{m} \right)^{1-1/p} \|x\|_1
\end{aligned}$$

Für $x \in B_1^N \cap M$ folgt

$$\|x\|_p \leq 4\sqrt{3} \left(\frac{c' \ln \frac{eN}{m}}{m} \right)^{1-1/p}$$

Also gilt für $m \geq c' \ln \frac{eN}{m}$

$$d^m(B_1^N, \ell_p^N) \leq 4\sqrt{3} \left(\frac{c' \ln \frac{eN}{m}}{m} \right)^{1-1/p} \leq \underbrace{4\sqrt{3}c'}_{=C_0} \left(\frac{\ln \frac{eN}{m}}{m} \right)^{1-1/p}$$

mit

$$C_0 = 4\sqrt{3}c' = 4\sqrt{3}g^{-1}(16 \cdot 144 \cdot (1 + \ln 9))$$

□

LINK: VORLESUNGSVIDEO 72 VOM 15.7.25

2.2. Untere Schranke an Gelfand-Weiten $d^m(B_1^N, \ell_p^N)$. Für $p \in (1, \infty]$ wollen wir untere Schranken an $d^m(B_1^N, \ell_p^N)$ herleiten.

Der Beweis nutzt wesentlich folgendes (rigorose und präzise) Resultat an die mindestens notwendige Anzahl von Messungen, um erfolgreiche Basisjagd zu garantieren.

(Verschiedentlich haben wir schon solche Resultat bewiesen, angekündigt und diskutiert.)

SATZ 11.12. Seien $m, N \in \mathbb{N}$, $m \leq N$, und $A \in \mathbb{R}^{m \times N}$ derart, dass jeder $2s$ -dünne Vektor $x \in \mathbb{R}^N$ ein Minimierer von

$$\min_{z \in \mathbb{R}^N} \|z\|_1 \text{ mit } Az = Ax$$

ist. Dann gilt

$$m \geq \frac{1}{\ln 9} \cdot s \cdot \ln \left(\frac{N}{4s} \right)$$

Für den Beweis brauchen wir ein (kombinatorisches Lemma:)

LEMMA 11.13 (Kombinatorisches Lemma:). Seien $s, N \in \mathbb{N}$ mit $s < N$. Dann existieren $n \geq \left(\frac{N}{4s}\right)^{s/2}$ Teilmengen $S_1, \dots, S_n \subset \overline{N}$ sd.:

- (1) $|S_j| = s \quad \forall j = 1, \dots, n$
- (2) $|S_j \cap S_i| < \frac{s}{2} \quad \forall i \neq j$

BEWEIS. Als Übung. □

BEWEIS VON SATZ 11.12. Betrachte Quotienten-Vektorraum

$$X := \ell_1^N / \ker A = \{[x] = x + \ker A \mid x \in \mathbb{R}^N\}$$

mit Quotientennorm

$$\|[x]\| = \inf_{v \in \ker A} \|x - v\|_1$$

Für $z = x - v$ mit $x \in \mathbb{R}_{2s}^N$ und $v \in \ker A$ haben wir $Ax = Az$ und damit auch $\|z\|_1 \geq \|x\|_1$, weil x nach Annahme Minimierer ist.

Damit

$$(69) \quad \|[x]\| = \|x\|_1 \text{ für } x \in \mathbb{R}_{2s}^N$$

Seien $S_1, \dots, S_n \subset \bar{N}$ wie in kombinatorischen Lemma 11.13. Definiere: $x_1, \dots, x_n \in \mathbb{R}^N$ durch

$$x_j := \frac{1}{s} \mathbb{1}_{S_j}$$

$$\text{d. h. } x_j = (x_j(1), \dots, x_j(N)) \in \mathbb{R}_s^N \text{ und } x_j(i) = \begin{cases} 1/s, & i \in S_j \\ 0, & \text{sonst} \end{cases}.$$

Dann $\|x_j\|_1 = 1$, weil $|S_j| = s$, sowie

$$|x_k(i) - x_j(i)| = \begin{cases} 1/s, & i \in S_k \Delta S_j := (S_k \cup S_j) \setminus (S_k \cap S_j) \\ 0, & \text{sonst} \end{cases}$$

insbesondere

$$\|x_k - x_j\|_1 = \sum_{i=1}^N |x_k^i - x_j^i| = \frac{1}{s} |S_k \Delta S_j| > \frac{1}{s} \cdot s = 1, \quad \forall 1 \leq j \neq k \leq n$$

nach Konstruktion der $S_1, \dots, S_n$

Wegen $x_k - x_j \in \mathbb{R}_{2s}^N$ folgt aus (69)

$$\|[x_k - x_j]\| = \|[x_k] - [x_j]\| = \|x_k - x_j\|_1 > 1$$

Nun verwenden wir wieder ein Volumen-Vergleichsargument:

- $[x_1], \dots, [x_n]$ liegen auf der Einheitskugel $\mathbb{S} \subset X$
- $\|[x_i] - [x_j]\| > 1$
- $\dim X = r := \text{rank}(A) \leq m$

$\Rightarrow$ Die Bälle $B_{1/2}([x_j]) \subset X$ sind alle disjunkt und in $B_{3/2}([0])$ enthalten.

$$n \cdot \frac{1}{2^r} \text{Vol}(B) = \text{Vol} \left(\bigcup_j^n (B_{1/2}([x_j])) \right) \leq \text{Vol}(B_{3/2}(0)) = \left(\frac{3}{2}\right)^r \text{Vol}(B)$$

Hier steht B für den Einheitsball.

Daraus folgt $n \leq 3^r \leq 3^m$ und zusammen mit dem kombinatorischen Lemma 11.13

$$\left(\frac{N}{4s}\right)^s \leq n^2 \leq 9^m$$

somit wie gewünscht

$$\Rightarrow m \geq \frac{1}{\ln 9} s \ln \frac{N}{4s}$$

□

LINK: VORLESUNGSVIDEO 73 VOM 15.7.25

Nun können wir die untere Schranke an $d^m(B_1^N, \ell_p^N)$ formulieren und beweisen:

SATZ 11.14. Seien $1 < p \leq \infty$ und $m, N \in \mathbb{N}$ mit $m < N$.

(a) Für

$$c' := \frac{2}{1 + 4 \ln 9} \approx 0,2043 \dots < 1 \quad \text{und} \quad \mu := \min \left\{ 1, \frac{c' \ln \left(\frac{eN}{m} \right)}{m} \right\}$$

gilt:

$$\frac{\mu^{1-1/p}}{2^{2-1/p}} \leq d^m(B_1^N, \ell_p^N)$$

(b) Es existiert eine Konstante $c > 0$, so dass

$$c \cdot \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p} \leq d^m(B_1^N, \ell_p^N)$$

Mögliche Wahl

$$c := \frac{\min\{1, c'\}^{1-1/p}}{2^{2-1/p}} \geq \frac{\min\{1, c'\}}{4} = \frac{c'}{4} \approx 0,0510 \dots > \frac{1}{20}$$

BEWEIS. Widerspruchsbeweis der Beh. (a): Nehmen an:

$$d^m(B_1^N, \ell_p^N) < \frac{\mu^{1-1/p}}{2^{2-1/p}} \quad \left[\mu := \min\left\{1, \frac{c' \ln \left(\frac{eN}{m}\right)}{m}\right\} \right]$$

Nach Def. von d^m ex. UVR $M \subset \mathbb{R}^N$ mit $\text{codim } M \leq m$ und

$$(70) \quad \|v\|_p < \frac{\mu^{1-1/p}}{2^{1-1/p}} \|v\|_1 \quad \text{für alle } v \in M, v \neq 0$$

Sei $A \in \mathbb{R}^{m \times N}$ so, dass $\ker A = M$. Setze $s = \lfloor \frac{1}{\mu} \rfloor \geq 1$. Damit

$$\frac{1}{2\mu} < s \leq \frac{1}{\mu} \quad \text{insbesondere } \frac{\mu}{2} \leq \frac{1}{2s}$$

Mit (70) folgt

$$(71) \quad \|v\|_p < \frac{1}{2} \frac{\mu^{1-1/p}}{2^{1-1/p}} \|v\|_1 \leq \frac{1}{2} \left(\frac{1}{2s} \right)^{1-1/p} \|v\|_1, \quad v \in \ker A \setminus \{0\}$$

Dies liefert zusammen mit der Hölder-Ungleichung

$$\begin{aligned} \|v\|_1 &\leq N^{1-1/p} \|v\|_p \\ 1 &< \frac{1}{2} \left(\frac{1}{2s} \right)^{1-1/p} N^{1-1/p} \end{aligned}$$

und damit $2s < N$.

Für beliebiges $S \subset \overline{N}$ mit $\#S \leq 2s$ folgt allgemein mit Hölder

$$\|v_S\|_1 \leq (2s)^{1-1/p} \|v_S\|_p \leq (2s)^{1-1/p} \|v\|_p$$

und für $v \in \ker A \setminus \{0\}$ wegen (71) sogar

$$< \frac{1}{2} \|v\|_1$$

$\Rightarrow A$ hat die NRE der Ordnung $2s$

$\Rightarrow$ Jeder $2s$ -dünne Vektor kann mit Basisjagd rekonstruiert werden.

Damit sagt uns Satz 11.12, dass

$$m \geq \frac{1}{\ln 9} \cdot s \cdot \ln \left(\frac{N}{4s} \right)$$

Aus 1.17 & Folgerung 1.18 haben wir:

$$[A \text{ hat die NRE der Ordnung } 2s \Rightarrow \ker A \cap \mathbb{R}_{2(2s)}^N = \{0\}] \Rightarrow$$

$m \geq \text{rank } A \geq 2(2s) = 4s$. Jetzt:

$$\begin{aligned} m &\geq \frac{1}{\ln 9} \cdot s \cdot \ln \left(\frac{N}{m} e \right) \geq \frac{1}{\ln 9} s \ln \left(\frac{eN}{m} \right) - \frac{s}{\ln 9} \ln e \\ &\stackrel{s > \frac{1}{2}\mu}{>} \frac{1}{2 \ln 9} \cdot \frac{1}{\mu} \ln \left(\frac{eN}{m} \right) - \frac{m}{4 \ln 9} \end{aligned}$$

Umstellen

$$\Rightarrow m > \left(1 + \frac{1}{4 \ln 9} \right)^{-1} \left(\frac{1}{2 \ln 9} \right) \frac{1}{\mu} \ln \left(\frac{eN}{m} \right)^{\mu \leq c' \frac{\ln(eN/m)}{m}} \geq \text{const.} \frac{m}{c'} \dots = m$$

Widerspruch, weil

$$\frac{\text{const.}}{c'} = \frac{\frac{1}{2 \ln 9}}{1 + \frac{1}{4 \ln 9}} \frac{1 + 4 \ln 9}{2} = \frac{1}{1 + \frac{1}{4 \ln 9}} \frac{1}{4 \ln 9} (1 + 4 \ln 9) = \frac{1 + 4 \ln 9}{1 + 4 \ln 9} = 1$$

□

Insgesamt haben wir bewiesen:

SATZ 11.15. Seien $m, N \in \mathbb{N}$ mit $m < N$.

(a) Für $1 \leq p \leq 2$ gilt

$$\frac{1}{20} \cdot \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p} \leq d^m(B_1^N, \ell_p^N) \leq C_0 \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p}$$

(b) Für $2 < p \leq \infty$ gilt

$$\frac{1}{20} \cdot \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1-1/p} \leq d^m(B_1^N, \ell_p^N) \leq C_0 \min \left\{ 1, \frac{\ln \frac{eN}{m}}{m} \right\}^{1/2}$$

jeweils mit

$$C_0 = 4\sqrt{3}g^{-1}(16 \cdot 144 \cdot (1 + \ln 9)) \quad \text{wobei} \quad g(x) = \frac{x}{1 + \ln x}$$

Für das CS sind auch die Mengen B_q^N mit $q \in (0, 1)$ als Modelle für komprimierbare Vektoren interessant. Auch in diesen Fällen gibt es Abschätzungen der Gelfand-Weiten.

SATZ 11.16 ([FPRU10] Foucart, Pajor, Rauhut & Ullrich, 2010). Für $0 < p \leq 1$ und $p < q \leq 2$ existieren Konstanten $c_{p,q}, C_{p,q} > 0$, die nur von p und q abhängen, so dass für $m < N$ gilt:

(72)

$$c_{p,q} \min \left\{ 1, \frac{\ln(N/m) + 1}{m} \right\}^{1/p-1/q} \leq d^m(B_p^N, \ell_q^N) \leq C_{p,q} \min \left\{ 1, \frac{\ln(N/m) + 1}{m} \right\}^{1/p-1/q},$$

und falls $p < 1$, dann gilt zusätzlich:

(73)

$$c_{p,q} \min \left\{ 1, \frac{\ln(N/m) + 1}{m} \right\}^{1/p-1/q} \leq d^m(B_{p,\infty}^N, \ell_q^N) \leq C_{p,q} \min \left\{ 1, \frac{\ln(N/m) + 1}{m} \right\}^{1/p-1/q}.$$

Hier steht $B_{p,\infty}^N$ für den ‚Einheitsball‘ bezüglich der schwachen Quasinorm

$$\|x\|_{p,+\infty} := \sup_{t>0} |\{j \in \overline{N} \mid |x_j| \geq t\}|^{1/p} \cdot t.$$

Index

- (P_0) , 19
- (P_q) , 31
- $(P_{0,\eta})$, 30
- (s, t) -restricted orthogonality constant, 74
- $A^\dagger$, 49
- A_S , 20
- B_p^N , 12
- $E^m(\cdot, \cdot)$, 108
- $E_{\text{ada}}^m(\cdot, \cdot)$, 109
- K -kohärent, 103
- $\mathbb{K}_S^N$, 12
- $\mathbb{K}_s^N$, 12
- $\|\cdot\|_{p,+\infty}$, 15
- $\delta_s(\cdot)$, 73
- ℓ_1 -Kohärenzfunktion, 63
- ℓ_q -robuste NRE der Ordnung s mit Konstanten ρ, τ , 43
- $\|A\|_{p \rightarrow q} := \sup_{\|x\|_p=1} \|Ax\|_q$, 53
- $\|\cdot\|_*$, 53
- μ , 63
- μ_1 , 63
- $\overline{N}$, 11
- $\preceq_P$, 28
- $\text{sgn } z$, 46
- σ_j , 55
- $\sigma_s(\cdot)_p$, 14
- $\theta_{s,t}(\cdot)$, 74
- $\text{tr}(\cdot)$, 54
- $\|A\|$, 53
- $c_k(\cdot)$, 108
- $d^m(\cdot, \cdot)$, 108
- x_S , 20
- (NR_S) , 36
- $(\text{RNR}_S^{\rho,\tau})$, 41
- $(\text{RNR})_{s,q}^{\rho,\tau}$, 43
- 53
- NP-schwer**, 28
- NP-vollständig**, 28
- NP**, 28
- P**, 28
- Basisjagd, 34
- beste s -Term Approximation, 13
- codim , 107
- equiangular tight frame (ETF), 65
- Gaußsche Zufallsmatrix, 91
- Gelfandweite, 108
- Gelfandzahl, 108
- gleichwinkliges System, 65
- Kleinstmöglicher Rekonstruktionsfehler, 108
- Kodimension, 107
- Kohärenz, 63
- Moore-Penrose-Pseudoinverse, 49
- nicht-wachsende Umordnung, 14
- Nullraumeigenschaft, 36
- Nullraumeigenschaft der Ordnung s , 37
- Rademacher-ZV, 96
- Reduzierte Singulärwertzerlegung, SVD, 56
- RIP-Konstante der Ordnung s , 73
- robuste NRE bzgl. S , 41
- s -sparse, 11
- schwache ℓ_p -Norm, 15
- singulärer Vektor, 55
- Singulärwert, 55
- stabile NRE, 39
- stabile NRE der Ordnung s mit Konstante ρ , 39