

Compressed Sensing

Gitta Kutyniok

Man braucht nur die Zeitung aufzuschlagen oder den Fernseher anzustellen und wird schon mit Meldungen zu "Big Data" oder der heutigen "Datenflut" konfrontiert. In der Tat produziert unsere Gesellschaft immer riesigere Datenmengen u.a. im Bereich der Medizin oder der Telekommunikation. Es wird prognostiziert, dass die Menge der digitalen Daten von 2005 bis 2020 um einen Faktor von ca. 300 wachsen wird, welches eine Größenordnung von ca. 40.000 Exabytes im Jahr 2020 bedeuten würde (1 Exabyte entspricht 1 Milliarde Gigabyte).

Es stellt sich die Frage, ob diese massiven Datenmengen wirklich vollständig akquiriert werden müssen. Geht man nach Galileo Galilei's Aussage "Measure what can be measured, ...", dann ist die Antwort ja. Allerdings sagte Albert Einstein gut 300 Jahre später: "Not everything that can be counted counts, ...". Und dieses Zitat enthält im Kern die heutige Annahme, dass der Informationsgehalt von Daten fast immer deutlich geringer ist, als deren Größe impliziert.

1 Ein neues Paradigma

In der mathematischen Signal- und Bildverarbeitung wird dieses Paradigma für digitale Daten $x \in \mathbb{R}^n$ folgendermaßen formuliert: Es existiert eine reelle $n \times N$ Matrix Φ ($N \geq n$) und ein k -sparse ($k \ll N$) Koeffizientenvektor $c \in \mathbb{R}^N$, d.h.

$$\|c\|_0 := \#\{i : c_i \neq 0\} \leq k,$$

so dass x darstellbar ist als

$$x = \Phi c.$$

x lässt sich somit als sparse Linearkombination der Spaltenvektoren $(\varphi_i)_{i=1}^N$ schreiben, und im einfachsten Fall $\Phi = I$ ist x selbst sparse, wobei die nicht-verschwindenden Einträge als Informationsgehalt interpretiert werden. Zahlreiche empirische Studien haben gezeigt, dass diese Annahme gerechtfertigt ist. Beispielsweise lassen sich Fotos normalerweise mittels eines Waveletsystems [7] sparse darstellen, siehe Abbildung 1. Auf dieser Tatsache beruht übrigens auch der Kompressionsstandard JPEG2000.

Akzeptiert man dieses Paradigma, so stellt sich als nächste Frage, ob nicht gleich nur der Informationsgehalt akquiriert werden kann. Zunächst erscheint dieses Vorhaben unmöglich, da die Positionen der nicht-verschwindenden Einträge unbekannt sind. Erstaunlicherweise konnte diese Frage aber kürzlich tatsächlich mittels der zu diesem Zweck von Candès, Romberg und Tao [4] und parallel von



Abbildung 1. *Originalbild und seine Waveletzerlegung*

Donoho [8] neu entwickelten mathematischen Methodik des Compressed Sensing positiv beantwortet werden.

Seit diesem Durchbruch hat Compressed Sensing einen einzigartigen Siegeszug nicht nur in die verschiedensten Gebiete der Mathematik, sondern auch in Anwendungsgebiete wie Elektrotechnik, Informatik, Medizin, Radartechnik oder auch Seismologie angetreten. In diesem Sinne hat sich Compressed Sensing zu einem hochgradig interdisziplinären Forschungsgebiet entwickelt. Auf der mathematischen Seite verwendet Compressed Sensing Methoden aus einer Vielzahl von Teilgebieten der Mathematik, insbesondere Angewandte Harmonische Analysis, Frame-Theorie, Geometrische Banachraumtheorie, Numerische Lineare Algebra, Optimierung, und der Theorie der Zufallsmatrizen; was diese Forschungsrichtung auch innerhalb der Mathematik aufgrund der methodischen Vielfalt außerordentlich spannend macht.

Im Folgenden werden wir nun tiefer in die Mathematik einsteigen und die grundsätzlichen Ideen und Resultate diskutieren. Für weitere Informationen verweisen wir auf die Bücher [11, 13]. Wir möchten auch bemerken, dass diese Theorie in den USA entstanden ist und wir deshalb gewisse Fachausdrücke im Englischen belassen werden.

2 Das Compressed Sensing-Problem

Beginnen wir mit der Hauptproblemstellung im Forschungsgebiet des Compressed Sensing, und zwar der Lösung von hochgradig unterbestimmten linearen Gleichungssystemen, falls die gesuchte Lösung sparse ist: Sei $x \in \mathbb{R}^n$ ein sparser Vektor und A eine reelle $m \times n$ Matrix ($m \ll n$), so ist die Aufgabe, den Vektor x aus $y \in \mathbb{R}^m$ zu rekonstruieren, wobei

$$y = Ax. \quad (1)$$

Lässt sich x im allgemeineren Fall nur mittels einer Matrix Φ sparse darstellen, so lösen wir nach dem Koeffizientenvektor c auf, d.h.

$$y = A\Phi c. \quad (2)$$

Eine Sichtweise dieses Problems hatten wir schon diskutiert. In diesem Fall wird $A = (a_{i,j})_{i=1,j=1}^{m,n}$ als Messmatrix aufgefasst, die nur den Informationsgehalt von x mittels linearer Messungen $(\langle x, a_{i,\cdot} \rangle)_{i=1}^m$ durch die Zeilenvektoren abtastet. Aus einem anderen Blickwinkel – verbunden mit einer anderen Zielsetzung – kann dieses

Problem aber auch als Dimensionsreduktionsproblem aufgefasst werden. Eine weitere Interpretation ist die Suche nach einer effizienten Darstellung von y im Darstellungssystem der Spaltenvektoren $(a_{\cdot,j})_{j=1}^n$.

Im Folgenden werden wir der Einfachheit halber nur noch (1) betrachten, möchten aber betonen, dass gerade derzeit viel Forschungsenergie in theoretische und praktische Aspekte des erweiterten Problems (2) gesteckt wird.

3 Lösungsstrategien

Zunächst ist nun zu klären, ob die sparseste Lösung von (1) überhaupt eindeutig ist, d.h. ob sich das Minimierungsproblem

$$(P_0) \quad \min_x \|x\|_0 \text{ unter der Nebenbedingung } y = Ax$$

eindeutig lösen lässt. Eine einfache charakterisierende Bedingung kann mit Hilfe des Begriffes des *Spark* – zusammengesetzt aus “sparse” und “rank” – angegeben werden.

Definition 3.1. Sei A eine $m \times n$ Matrix. Dann ist der *Spark* von A , bezeichnet mit $\text{spark}(A)$, die minimale Anzahl linear abhängiger Spalten von A .

Das folgende Resultat liefert eine äquivalente Bedingung an den Grad der Sparsity von x .

Theorem 3.2 ([9]). Sei A eine $m \times n$ Matrix, und sei $k \in \mathbb{N}$. Dann sind die folgenden Bedingungen äquivalent.

- (i) Falls eine Lösung x von (P_0) k -sparse ist, so ist sie die eindeutige Lösung.
- (ii) $k < \text{spark}(A)/2$.

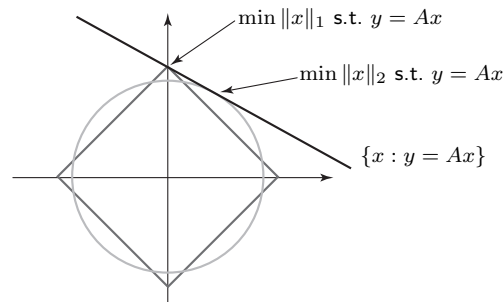
Allerdings ist das Problem (P_0) im Allgemeinen NP-hart, was sich durch die notwendige kombinatorische Suche erklärt, und lässt sich somit nicht effizient lösen. Als einen Ausweg haben Chen, Donoho und Saunders [5] $\|\cdot\|_0$ durch die ℓ_1 -Norm als nahestehende konvexe Norm ersetzt, was auf das sogenannte *Basis Pursuit Problem* führt:

$$(P_1) \quad \min_x \|x\|_1 \text{ unter der Nebenbedingung } y = Ax.$$

Dieses kann mittels Linearer Programmierung effizient gelöst werden. Ein Hauptproblem besteht nun darin zu zeigen, dass “ $\ell_0 = \ell_1$ ” gilt. Abbildung 2 illustriert, weshalb dies intuitiv plausibel ist.

4 Erfolgsgarantien

Befassen wir uns nun mit der Frage nach hinreichenden Bedingungen, das Compressed Sensing-Problem mittels (P_1) zu lösen. Hierbei spielen zwei Voraussetzungen eine herausragende Rolle: die *Sparsity* von x und die *Incoherence* von A .

Abbildung 2. Lösungen des ℓ_1 und ℓ_2 Minimierungsproblems

4.1 Mutual Coherence

Sehen wir uns ein erstes Resultat an, welches diese Beziehung sehr gut illustriert. Hierfür wird die sogenannte *Mutual Coherence* als Mass für die Incoherence benötigt, welche wir zunächst definieren.

Definition 4.1. Sei A eine $m \times n$ Matrix mit Spaltenvektoren $(a_i)_{i=1}^n$. Dann ist die *Mutual Coherence* $\mu(A)$ definiert durch

$$\mu(A) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}.$$

Dieser Wert ist offensichtlich nach oben beschränkt durch 1. Die untere Schranke hingegen ist die sogenannte *Welch Schranke* $((n-m)/(m(n-1)))^{1/2}$.

Diese Definition ermöglicht uns nun, den folgenden Satz zu formulieren.

Theorem 4.2 ([9, 10]). Sei A eine $m \times n$ Matrix, und sei $x \in \mathbb{R}^n \setminus \{0\}$ eine Lösung des zugehörigen Problems (P_0) mit

$$\|x\|_0 < \frac{1}{2}(1 + \mu(A)^{-1}). \quad (3)$$

Dann ist x die eindeutige Lösung von (P_0) und (P_1) .

Bedingung (3) zeigt sehr schön die Balance zwischen Sparsity und Mutual Coherence. Ist $\mu(A) = 0$, so ist eine exakte Rekonstruktion für alle Vektoren $x \in \mathbb{R}^n$ möglich. Je kohärenter die Spalten von A sind, desto weniger nicht-verschwindende Einträge in x sind erlaubt.

4.2 Restricted Isometry Property

Eine andere hinreichende Bedingung, die vielleicht noch intuitiver ist, ist die *Restricted Isometry Property*. Diese Eigenschaft misst, wie orthogonal Teilmengen von jeweils k Spalten einer $m \times n$ Matrix A sind. Man beachte, dass Orthogonalität simultan für *alle* dieser Teilmengen bei $m < n$ nicht möglich ist.

Definition 4.3. Sei A eine $m \times n$ Matrix. Dann hat A die *Restricted Isometry Property (RIP) der Ordnung k* , falls ein $\delta_k \in (0, 1)$ existiert, so dass für alle k -sparsen Vektoren $x \in \mathbb{R}^n$

$$(1 - \delta_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k) \|x\|_2^2$$

gilt.

Wie wir in Abschnitt 5.2 diskutieren werden, wird die RIP am besten von Zufallsmatrizen wie identisch unabhängig verteilten Gauß'schen Matrizen oder Bernoulli Matrizen erfüllt. In der Tat wurde der Begriff "Compressed Sensing" zunächst für Zufallsmatrizen als Messmatrizen verwandt, hat sich aber schon kurz darauf allgemein für die Rekonstruktion von sparsen Lösungen unterbestimmter linearer Gleichungssysteme eingebürgert.

Das folgende Resultat zeigt, dass exakte Rekonstruktion für k -sparse Vektoren x möglich ist, falls δ_{2k} nicht zu groß ist; man erkennt wieder die Balance zwischen der Sparsity von x und die Incoherence von A . Ferner zeigt dieses Resultat, dass Stabilität vorliegt in dem Sinne, dass eine sinnvolle Schranke für den Fehler der Rekonstruktion angegeben werden kann, falls x nur *fast* sparse ist. "Sinnvoll" bedeutet hier, dass der Fehler sich beschränken läßt durch den ℓ_1 -Fehler der besten k -Term Approximation von x , welche wir mit x_k bezeichnen. Man beachte, dass in diesem Fall x_k lediglich der Vektor x ist, bei dem alle bis auf die k betragsmäßig größten Einträge auf Null gesetzt wurden.

Theorem 4.4 ([3, 6]). *Sei A eine $m \times n$ Matrix, die die RIP der Ordnung $2k$ mit $\delta_{2k} < \sqrt{2}-1$ erfüllt. Seien ferner $x \in \mathbb{R}^n$ und \hat{x} eine Lösung des zugehörigen Problems (P_1) . Dann gilt*

$$\|x - \hat{x}\|_2 \leq C \cdot \left(\frac{\|x - x_k\|_1}{\sqrt{k}} \right)$$

für eine Konstante C abhängig von δ_{2k} .

Man fragt sich nun, weshalb hier δ_{2k} (statt δ_k) eine Rolle spielt. Ein intuitives Argument ist das Folgende: Ist RIP für eine Matrix A mit kleinem δ_{2k} erfüllt, so impliziert dies, dass die paarweisen Distanzen zwischen k -sparsen Vektoren $x_1, x_2 \in \mathbb{R}^m$ im Messraum \mathbb{R}^n erhalten bleiben im Sinne von

$$(1 - \delta_{2k}) \|x_1 - x_2\|_2^2 \leq \|Ax_1 - Ax_2\|_2^2 \leq (1 + \delta_{2k}) \|x_1 - x_2\|_2^2.$$

Es sei zum Abschluss dieses Abschnitts noch erwähnt, dass Eindeutigkeit der Lösung des ℓ_1 Minimierungsproblems mittels einer sogenannten *Null Space Property* exakt charakterisiert werden kann. Bei Interesse sei der Leser auf die vorhergenannten Bücher bzw. auf den Artikel [6] verwiesen, der diese Eigenschaft nicht nur für ℓ_1 Minimierung, sondern für allgemeine Decoder analysiert.

5 Eckpfeiler von Compressed Sensing

Aus den vorherigen Erörterungen lassen sich die zentralen Eckpfeiler des Compressed Sensing ersehen:

- Sparsity
- Messmatrix
- Rekonstruktionsalgorithmus

Sehen wir uns nun diese Aspekte und neuere Entwicklungen derer an, wobei wir auf technische Details verzichten werden.

5.1 Sparsity-Modelle

In Anwendungen ist das Vorliegen exakter Sparsity natürlich illusorisch – schon aufgrund eigentlich immer vorhandenen Rauschens. Deshalb wurde u.a. der Begriff *compressible* für Vektoren eingeführt.

Definition 5.1. Sei $1 \leq p < \infty$ und $r > 0$. Ein Vektor $x = (x_i)_{i=1}^n \in \mathbb{R}^n$ ist *p-compressible mit Konstante C und Rate r*, falls für alle $k \in \{1, \dots, n\}$

$$\min_{\tilde{x} \text{ k-sparse}} \|x - \tilde{x}\|_p \leq C \cdot k^{-r}$$

gilt.

Eine andere Erweiterung des Begriffs der Sparsity beruht auf der Beobachtung, dass die Menge der sparsen Vektoren eine Vereinigung von orthogonalen Unterräumen ist, siehe Abbildung 3.

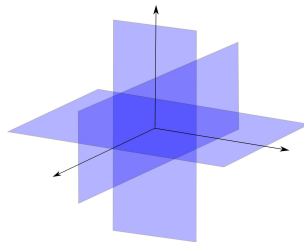


Abbildung 3. Menge der 2-sparsen Vektoren in \mathbb{R}^3

Dies führt auf das Signalmodell einer Vereinigung von Unterräumen, die zuerst in [18] vorgeschlagen wurde.

Zahlreiche erweiterte Sparsity-Modelle wurden im Laufe der Zeit eingeführt, oftmals von der beabsichtigten Anwendung inspiriert, u.a. *Block Sparsity*, *Fusion Frame Sparsity*, *Geometric Sparsity*, *Joint Sparsity*, etc. Diese Modelle für strukturierte Sparsity beinhalten zusätzliche Informationen über die Verteilung der nicht-verschwindenden Einträge, was Abschwächungen der Bedingungen an die Messmatrix erlaubt.

5.2 Messmatrizen

Obwohl jede Anwendung ihre eigenen Anforderungen an die Messmatrix stellt und folglich deren Wahl oftmals sehr einschränkt, kann man sich fragen, welche Matrizen allgemein optimal wären. Wir analysieren im Folgenden die RIP bzgl. optimaler Matrizen.

Die ersten Analysen betrachteten Zufallsmatrizen. In [1] konnte gezeigt werden, dass $m \times n$ identisch unabhängig verteilte Gauß'sche Matrizen sowie Bernoulli Matrizen die RIP der Ordnung k mit δ_k erfüllen, falls für die Anzahl der Messwerte

$$m \gtrsim \delta_k^{-2} k \log\left(\frac{n}{k}\right)$$

gilt. Diese Abhängigkeit von k und n ist in der Tat optimal, wie Argumente mit Gelfand-Weiten von ℓ_1 -Kugeln zeigen [12].

Interessanterweise ist die Konstruktion von Zufallsmatrizen, die die RIP mit hinreichend kleinem δ_k erfüllen, eng mit dem Johnson-Lindenstrauss Lemma [14] aus dem Jahr 1984 verbunden. Damals hat dieses Resultat nicht viel Aufsehen erregt, wohingegen es heutzutage ein zentrales Hilfsmittel in der Theorie des Compressed Sensing ist. Das Resultat lautet wie folgt.

Theorem 5.2 (Johnson-Lindenstrauss Lemma [14]). *Seien $\varepsilon \in (0, 1)$, $x_1, \dots, x_p \in \mathbb{R}^n$, und sei $m = O(\varepsilon^{-2} \log p)$ eine natürliche Zahl. Dann existiert eine Lipschitz Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit*

$$(1-\varepsilon)\|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2^2 \leq (1+\varepsilon)\|x_i - x_j\|_2^2$$

für alle $i, j \in \{1, \dots, p\}$.

Solche Abbildungen f lassen sich mit Hilfe von *concentration inequalities* in hochdimensionalen Räumen konstruieren. Die "Äquivalenz" dieser Konstruktion zu der Konstruktion von RIP Matrizen mit kleinem δ_k konnte in [1, 16] gezeigt werden.

Verwendet man hingegen deterministische Matrizen, so zeigt sich, dass es leider ungleich schwerer ist, $m \gtrsim k^2$ zu unterbieten, was gerade in hohen Dimensionen die Verwendung dieser Matrizen eher unattraktiv macht. Die beste derzeit bekannte deterministische Konstruktion wurde kürzlich von einem weiteren Fields-Medaillenträger, der auf diesem Gebiet arbeitet, gezeigt, und zwar von Bourgain und seinen Koauthoren, die $m \gtrsim k^{2-\alpha}$ mit sehr kleiner Konstante $\alpha > 0$ erreichten [2].

5.3 Rekonstruktionsalgorithmen

Konvexe Optimierungsprobleme, insbesondere das Minimierungsproblem (P_1) , sind die am häufigsten analysierten Rekonstruktionsalgorithmen und liefern in der Regel die theoretisch stärksten Rekonstruktionsgarantien. Beim Vorliegen verrauschter Messwerte beachte man, dass statt (P_1) in diesem Fall das Problem

$$\min_x \|x\|_1 \quad \text{unter der Nebenbedingung} \quad \|Ax - y\|_2^2 \leq \varepsilon$$

gelöst wird, wobei $\varepsilon > 0$ abhängig vom Rauschanteil gewählt wird.

Eine andere Klasse von Rekonstruktionsalgorithmen sind Greedy-Algorithmen. Diese approximieren die Einträge des zu rekonstruierenden Vektors x iterativ und sind deutlich schneller und einfacher zu implementieren als konvexe Optimierungsprobleme. Ein prominenter Vertreter dieser Klasse ist *Orthogonal Matching Pursuit*, eingeführt in [20]. Die Philosophie dieses Algorithmus ist es, in jeder Iteration einen neuen Eintrag von x zu berechnen, indem das derzeitige Residuum mit den Spaltenvektoren von A verglichen wird, die derzeitige Indexmenge des rekonstruierten Vektors um den besten Index erweitert wird und das neue Residuum durch orthogonale Projektion von y auf den durch die ausgewählten Indizes gegebenen Unterraum von \mathbb{R}^n bestimmt wird.

6 Einige Anwendungsbeispiele

Lassen Sie uns mit einigen Anwendungsbeispielen schließen. Das erste Beispiel zeigt einen der beeindruckenden Erfolge von Compressed Sensing im Bereich der Medizin, genauer der Magnetresonanztomographie (MRT). In der MRT wird das Bild durch das Abtasten von dessen Fouriertransformation an bestimmten Werten gemessen und anschließend rekonstruiert. Hierfür wird oftmals ein quadratisches Gitter verwendet. Allerdings ist die Datenakquisition extrem langsam, und jeder, der schon einmal in einem MRT-Scanner gelegen hat, kann dies bestätigen. Es ist offensichtlich, dass dies insbesondere bei der Behandlung von Kindern ein großes Problem darstellt. Da die Bilder sparse z.B. in einem Waveletsystem sind und die Datenakquisition aus linearen Messwerten besteht, bietet sich die Möglichkeit an, Compressed Sensing Methoden anzuwenden. In der Tat konnte in [19] mit diesem Zugang die benötigte Zeit für die Datenakquisition um einen Faktor 6 reduziert werden, siehe Abbildung 4. Dies wird im

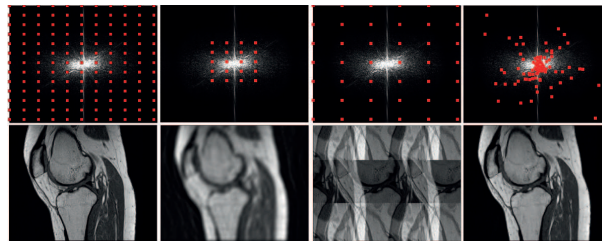


Abbildung 4. *Verschiedene Abtaststrategien bei MRT im Fourierbereich; die rechte Abbildung entspricht der Compressed Sensing Methodik*

Wesentlichen durch die Wahl zufälliger Abtastwerte im Fourierbereich und Rekonstruktion mittels

$$\min_x \|\Phi^T x\|_1 \text{ unter der Nebenbedingung } \|\mathcal{F}_S x - y\|_2 \leq \epsilon$$

erreicht, wobei \mathcal{F}_S die Akquisition von Messwerten der Fouriertransformation in den Punkten in S , ϵ die Genauigkeit der Rekonstruktion zu den gemessenen Daten und Φ ein Waveletsystem ist.

Das zweite Beispiel zeigt, dass Compressed Sensing auch auf Probleme, die auf den ersten Blick nichts mit Messungen zu tun zu haben scheinen, angewandt werden kann. In der Datenverarbeitung ist ein zentrales Problem das Fehlen von Daten, da manchmal Daten durch die Übertragung oder andere Einflüsse verlorengehen oder erst gar nicht akquiriert werden können. Sind die Originaldaten x aber sparse in einem Darstellungssystem Φ und kennt man den Bereich, in dem Daten fehlen (und sei P_K die orthogonale Projektion auf diesen Bereich), so können die Daten durch Lösung des Minimierungsproblems

$$\min_c \|c\|_1 \text{ unter der Nebenbedingung } P_K x = P_K \Phi c$$

durch $\Phi \hat{c}$ rekonstruiert werden, wobei \hat{c} die Lösung des Minimierungsproblems ist. Eine Illustration aus dem Ge-

biet der Bildverarbeitung unter Verwendung von sogenannten Shearlets [17] findet sich in Abbildung 5. Für Algorithmus und Theorie verweisen wir auf [15].

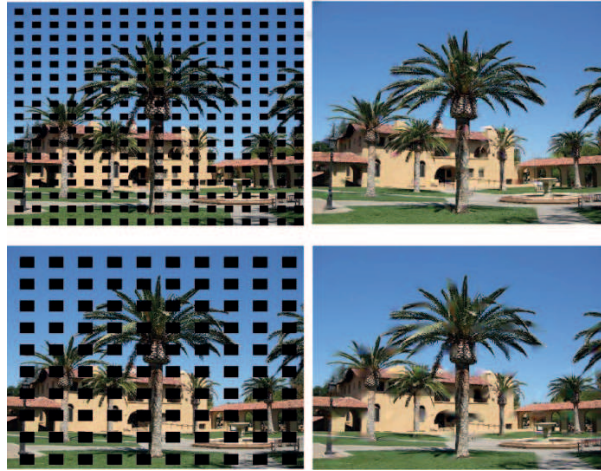


Abbildung 5. *Linke Spalte: Bilder mit fehlenden Bereichen; Rechte Spalte: Zugehörige rekonstruierte Bilder*

7 Ein Ausblick in die Zukunft...

Welches sind nun die derzeit wichtigsten Forschungsrichtungen auf dem Gebiet des Compressed Sensing? Ein Schwerpunkt liegt sicher auf konkreten *Anwendungen* z.B. in der Biologie, Elektrotechnik, Kommunikationstechnologie, Medizin, Optik oder Radartechnologie. Ferner werden oftmals Daten als Matrizen modelliert, und die Sparsity Annahme wird durch *low rank* Annahmen ersetzt. Dies führt auf das Teilgebiet des *Matrix Completion* oder allgemeiner *Tensor Completion*. Ferner ist – wie wir gesehen haben – die Frage noch völlig offen, ob man deterministische Matrizen konstruieren kann, für die die gleiche geringe Anzahl von Messwerten ausreichend ist wie für die diskutierten Zufallsmatrizen, und diverse *strukturierte Zufallsmatrizen* werden derzeit untersucht. Interessant ist derzeit auch die Entwicklung eines sinnvollen Analogons zu Compressed Sensing im Kontinuierlichen.

Literatur

- [1] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin, A simple proof of the Restricted Isometry Property for random matrices. *Constr. Approx.*, 28:253-263, 2008.
- [2] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova, Explicit constructions of rip matrices and related problems. *Duke Math. J.*, 159:145-185, 2011.
- [3] E. J. Candès, The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. I*, 346:589–592, 2008.
- [4] E. J. Candès, J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207-1223, 2006.

- [5] S. S. Chen, D. L. Donoho, and M. A. Saunders, Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1998.
- [6] A. Cohen, W. Dahmen, and R. DeVore, Compressed sensing and best k-term approximation. *J. Am. Math. Soc.*, 22:211–231, 2009.
- [7] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
- [8] D. L. Donoho, Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.
- [9] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization, *Proc. Natl. Acad. Sci. USA*, 100:2197–2202, 2003.
- [10] M. Elad and A. M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 48:2558–2567, 2002.
- [11] Y. C. Eldar and G. Kutyniok, eds., *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.
- [12] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, The Gelfand widths of ℓ_p -balls for $0 < p \leq 1$. *J. Complexity*, 26:629–640, 2010.
- [13] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.
- [14] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math*, 26:189–206, 1984.
- [15] E. J. King, G. Kutyniok, and X. Zhuang, Analysis of Inpainting via Clustered Sparsity and Microlocal Analysis. *J. Math. Imaging Vis.* 48:205–234, 2014.
- [16] F. Krahermer and R. Ward, New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *SIAM J. Math. Anal.*, 43:1269–1281, 2011.
- [17] G. Kutyniok and D. Labate, eds., *Shearlets: Multiscale Analysis for Multivariate Data*. Birkhäuser, Boston, 2012.
- [18] Y. Lu and M. Do, Sampling signals from a union of subspaces. *IEEE Signal Proc. Mag.*, 25:41–47, 2008.
- [19] K. Lustig, D. L. Donoho, and J. M. Pauly, Sparse mri: The application of compressed sensing for rapid mr imaging. *Magn. Reson. Med.*, 6:118295, 2007.
- [20] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers*, 1:40–44, 1993.