Sparse Signal Processing: Subspace Clustering and System Identification

Author(s):
Heckel, Reinhard Wolfram

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010252760 →

DISS. ETH NO. 22148

SPARSE SIGNAL PROCESSING: SUBSPACE CLUSTERING AND SYSTEM IDENTIFICATION

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

REINHARD WOLFRAM HECKEL

Dipl.-Ing., University of Ulm
born on 08.04.1986
citizen of
Germany

accepted on the recommendation of

Prof. Helmut Bölcskei,	examiner
Prof. Holger Boche,	co-examiner
Prof. Peter Bühlmann,	co-examiner

2014

Abstract

An important problem in statistics, machine learning, and modern signal processing is to recover information of limited complexity, or, more specifically, of low-dimensional structure, from seemingly few data. Often, this amounts to recover a sparse signal, i.e., a signal which is non-zero at few locations only, by solving an under-determined system of linear equations. A by now well known example is compressive sampling [CW08], a signal processing technique for efficiently acquiring and reconstructing certain signals from far fewer measurements than necessary for reconstruction with traditional methods. Compressive sampling relies on the insight that many real-world signal have a sparse representation in some basis. In this thesis, we will use ideas from sparse signal processing to cluster high dimensional data points, to identify sparse linear operators, and to recover sparse signals with certain block-structure.

In the first part of this thesis, we consider the problem of clustering noisy and incompletely observed high-dimensional data points into a union of low-dimensional subspaces and a set of outliers. The number of subspaces, their dimensions, and their orientations are assumed unknown. This problem is known as subspace clustering and has applications in, e.g., unsupervised learning, image processing, disease detection and computer vision [Vid11]. We propose a simple low-complexity subspace clustering algorithm, termed thresholding-based subspace clustering (TSC), which applies spectral clustering to an adjacency matrix obtained by thresholding the correlations between data points. In other words, the adjacency matrix is constructed from

the nearest neighbors of each data point in spherical distance. A statistical performance analysis shows that TSC succeeds even when the subspaces intersect and that it exhibits robustness to additive noise. Specifically, our results reveal an explicit tradeoff between the affinity of the subspaces and the tolerable noise level. We furthermore prove that TSC succeeds even when the data points are incompletely observed with the number of missing entries allowed to be (up to a log-factor) linear in the ambient dimension. We also propose a simple scheme that provably detects outliers, and present numerical results on real and synthetic data.

In practice, it is often desirable to first project the high-dimensional data points to be clustered into a lower-dimensional space and to perform the clustering task there; this reduces storage requirements and computational cost. We quantify analytically the impact of dimensionality-reduction through random projection on the performance of the sparse subspace clustering (SSC) algorithm [EV13] and the TSC algorithm. Both SSC and TSC are based on the principle that a data point in a low-dimensional subspace can be sparsely represented through points in the same subspace. We find that for both algorithms dimensionality reduction down to the order of the subspace dimensions is possible without incurring significant performance degradation. Our results only require the random projection matrix to satisfy the Johnson-Lindenstrauss property and are therefore general enough to apply to structured random projection matrices that allow for efficient implementations.

In the second part of this thesis, we consider the problem of identifying a linear deterministic operator from its response to a given probing signal. For the large class of linear operators that can be represented as a continuous weighted superposition of time-frequency shift operators we show that stable identifiability is possible if and only if the support area of the operator's spreading function satisfies $\Delta \leq 1/2$. This result holds for an arbitrary (possibly fragmented) support region of the spreading function, does not impose limitations on the total extent of the support region, and, most importantly, does not require the support region to be known prior to identification. Previous results [Kai62; Bel69; KP05; PW06a] assumed the support region of the spreading function to be known prior to identification, which is often impossible to realize in practice. E.g, for wireless channels, knowing the support region would amount to knowing the delays and Doppler shifts induced by the scatterers in the propagation medium. Furthermore, we prove that stable identifiability of *almost all operators* is possible if $\Delta < 1$. This result is surprising as it says that there is no penalty for not knowing the support region of the spreading function prior to identification. Identification algorithms that provably recover all operators with $\Delta \leq 1/2$, and almost all operators with $\Delta < 1$ are presented.

In the third and finial part of this thesis, we consider the recovery of block-sparse signals, i.e., sparse signals that have nonzero entries occurring in blocks, from an underdetermined system of linear equations. The block-sparse model includes a number of structured sparsity models, in particular the multiple measurement vector (MMV), and the generalized MMV (GMMV) model, where the measurement matrices are allowed to differ across measurements. We derive probabilistic recovery guarantees showing that—under certain (mild) conditions on the measurement matrix—a ℓ_2/ℓ_1-norm minimization program fails with a probability that decays exponentially in the length of the blocks. This result continues to hold when the measurement is subject to (bounded) noise. Our results evaluated for the GMMV case show that recovery performance does not suffer from the individual measurements being taken through different measurement matrices. What is more, recovery performance typically benefits from diversity in the measurement matrices.

Kurzfassung

Informationen von geringer Komplexität, oder genauer gesagt, von niedrigdimensionaler Struktur aus wenigen Daten wiederherzustellen, ist ein wichtiges Problem in Statistik, in maschinellem Lernen und in moderner Signalverarbeitung. Oft bedeutet dies ein dünnbesetztes Signal, d. h. ein Signal, das nur an wenigen Stellen ungleich Null ist, durch das Lösen eines unterbestimmten Gleichungssystems wiederherzustellen. Ein Beispiel ist compressive sampling, [CW08], eine Signalverarbeitungstechnik für die effiziente Akquisition und Rekonstruktion bestimmter Signale von weitaus weniger Messungen, als für die Rekonstruktion mit traditionellen Methoden benötigt wird. Compressive sampling beruht auf der Einsicht, dass viele Signale in der realen Welt eine dünnbesetzte Darstellung in einer bestimmten Basis besitzen. In dieser Dissertation benutzen wir Ideen aus dem Gebiet der Wiederherstellung dünnbesetzter Signale, um hochdimensionale Datenpunkte zu clustern, dünnbesetzte lineare Operatoren zu identifizieren und ein dünnbesetztes Signal mit einer bestimmten Blockstruktur wiederherzustellen.

In dem ersten Teil dieser Dissertation betrachten wir das Clustern von verrauschten und nur zum Teil beobachteten hochdimensionalen Datenpunkten in eine Vereinigung von niedrigdimensionalen Unterräumen und ein Set von Ausreissern. Die Anzahl der Unterräume, ihre Dimensionen und ihre Richtungen sind unbekannt. Dieses Problem ist als "subspace clustering" bekannt und hat z. B. Anwendungen in unüberwachtem Lernen, der Bildverarbeitung, der Detektion von Krankheiten und "computer vision" [Vid11]. Wir stellen einen ein-

fachen subspace clustering Algorithmus von niedriger Komplexität vor, welchen wir "thresholding-based subspace clustering (TSC)" nennen. TSC wendet spektrales Clustern auf eine Adjazenzmatrix an, welche durch thresholding der Korrelationen zwischen den Datenpunkten erzeugt wird. In anderen Worten, diese Matrix wird aus den nächstgelegenen Nachbarn (in sphärischer Distanz) eines jedes Datenpunktes konstruiert. Eine statistische Analyse zeigt, dass TSC sogar das richtige Ergebnis liefern kann, wenn die Unterräume überlappen. Zudem ist TSC unempfindlich gegen additives Rauschen. Insbesondere zeigen unsere Resultate eine explizite Wechselbeziehung zwischen der Ähnlichkeit der Unterräume und dem tolerablen Rauschpegel. Wir zeigen darüber hinaus, dass TSC sogar erfolgreich clustern kann, wenn die Datenpunkte nur teilweise beobachtet werden und die Anzahl von fehlenden Einträgen (bis auf einen logarithmischen Term) linear in der einbettenden Dimension ist. Ausserdem stellen wir eine simple Methode vor, um Ausreisser zu detektieren, und wir präsentieren numerische Resultate mit echten und synthetischen Daten.

In der Praxis ist es oft wünschenswert, die hochdimensionalen Daten zunächst in einen niederdimensionaleren Raum zu projizieren und sie dann zu clustern. Das reduziert die Kosten, die Daten zu speichern und sie zu verarbeiten. Wir quantifizieren analytisch den Einfluss der Dimensionsreduktion durch zufällige Projektionen auf die Qualität des Clustering-Ergebnisses von dem "sparse subspace clustering (SSC)"-Algorithmus [EV13] und dem TSC-Algorithmus. Sowohl SSC als auch TSC basieren auf dem Prinzip, dass ein Datenpunkt in einem niedrigdimensionalen Unterraum durch wenige andere Punkte aus demselben Unterraum dargestellt werden kann. Unsere Ergebnisse zeigen, dass für beide Algorithmen eine Dimensionsreduktion bis auf die Grössenordnung der Dimensionen der Unterräume möglich ist - ohne merkbare Verschlechterung der Clustering-Lösung. Unsere Resultate setzen nur voraus, dass die zufällige Projektionsmatrix eine Johnson-Lindenstrauss-Eigenschaft erfüllt, und gelten somit für strukturierte Zufallsmatrizen, welche effizient implementiert werden können.

Im zweiten Teil dieser Dissertation betrachten wir die Identifikation von einem linearen deterministischen Operator von seiner Antwort auf ein Testsignal. Wir betrachten die grosse Klasse von linearen Operatoren, welche als eine (kontinuierlich) gewichtete Superposition von Zeit-Frequenz-Verschiebungsoperatoren dargestellt werden kann. Für diese Klasse von Operatoren zeigen wir, dass stabile Identifikation möglich ist, wenn und nur wenn die Fläche des Trägers der Streufunktion $\Delta \leq 1/2$ erfüllt. Dieses Resultat gilt für beliebige (möglicherweise fragmentierte) Träger der Streufunktion und schränkt das totale Ausmass des Trägers nicht ein. Vor allem aber verlangt dieses Resultat nicht, dass der Träger der Streufunktion bekannt ist. Frühere Ergebnisse [Kai62; Bel69; KP05; PW06a] nahmen an, dass der Träger der Streufunktion vor der Identifikation des Operators bekannt ist, was oftmals in der Praxis nicht zu realisieren ist. Den Träger zu kennen, bedeutet für drahtlose Kanäle z. B., dass die Verzögerungen und Doppler-Verschiebungen der Streuer in dem Ausbreitungsmedium bekannt sind. Zudem beweisen wir, dass *fast alle Operatoren* stabil identifiziert werden können, wenn $\Delta < 1$. Das ist überraschend, da somit keine Kosten dafür anfallen, dass der Träger der Streufunktion unbekannt ist. Schliesslich präsentieren wir Algorithmen, die alle Operatoren beweisbar identifizieren können, für die $\Delta \leq 1/2$ gilt, und fast alle Operatoren, für die $\Delta < 1$ gilt.

Im dritten und letzten Teil dieser Dissertation betrachten wir die Wiederherstellung von Block-dünnbesetzten Signalen von einem unterbestimmten System von linearen Gleichungen. Block-dünnbesetzte Signale sind dünnbesetzte Signale, in welchen die Nicht-Null-Einträge in Blocks vorkommen. Block-dünnbesetzte Signale beinhalten eine Reihe von Modellen von dünnbesetzten Signalen mit Struktur, z. B. das multiple measurement vector (MMV) und das generalized MMV (GMMV) Problem, bei welchen die (Mess) Matrizen über verschiedene Messungen variieren dürfen. Wir zeigen probabilistische Resultate, welche die Wiederherstellung der Signale garantieren und zeigen, dass - unter bestimmten (schwachen) Bedingungen, die Messmatrix betreffend - ein ℓ_2/ℓ_1-Norm-Minimierungsprogramm nur mit einer Wahrscheinlichkeit versagt, die exponentiell in der Anzahl der Länge der Blocks abfällt. Dieses Resultat gilt auch, wenn die Messungen von beschränktem Rauschen betroffen sind. Unsere Resultate ausgewertet für den GMMV-Fall zeigen, dass die Wiederherstellungsqualität der Signale nicht darunter leidet, dass die einzelnen Messungen durch verschiedene Messmatritzen erhalten werden. Vielmehr profitiert die Wiederherstellungsqualität in der Regel von der Vielfalt in den Messmatritzen.