Forty-Seventh Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 30 - October 2, 2009

Sparse and Low-Rank Matrix Decompositions

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky

Invited paper to Allerton 2009

***Abstract*— We consider the following fundamental problem: given a matrix that is the sum of an unknown sparse matrix and an unknown low-rank matrix, is it possible to exactly recover the two components? Such a capability enables a considerable number of applications, but the goal is both ill-posed and NP-hard in general. In this paper we develop (a) a new uncertainty principle for matrices, and (b) a simple method for exact decomposition based on convex optimization. Our uncertainty principle is a quantification of the notion that a matrix cannot be sparse while having diffuse row/column spaces. It characterizes when the decomposition problem is ill-posed, and forms the basis for our decomposition method and its analysis. We provide deterministic conditions – on the sparse and low-rank components – under which our method guarantees exact recovery.**

I. INTRODUCTION

Given a matrix formed by adding an unknown sparse matrix and an unknown low-rank matrix, we study the problem of decomposing the composite matrix into its sparse and low-rank components. Such a problem arises in a number of applications in statistical model selection, machine learning, system identification, computational complexity theory, and optics. In this paper we provide conditions under which the decomposition problem is well-posed, i.e., the sparse and low-rank components are fundamentally identifiable, and present tractable convex relaxations that recover the sparse and low-rank components exactly.

Main results: Formally let $C = A^\star + B^\star$ with $A^\star$ being a sparse matrix and $B^\star$ a low-rank matrix. Given C our goal is to recover $A^\star$ and $B^\star$ without any prior information about the sparsity pattern of $A^\star$, or the rank/singular vectors of $B^\star$. In the absence of additional conditions, this decomposition problem is clearly ill-posed. There are a number of situations in which a unique decomposition may not exist; for example the low-rank matrix $B^\star$ could itself be very sparse, making it hard to uniquely identify from another sparse matrix. In order to characterize when exact recovery is possible we develop a new notion of *rank-sparsity incoherence*, which relates the sparsity pattern of a matrix to its row/column spaces via an *uncertainty principle*. Our analysis is geometric in nature, with the tangent spaces to the algebraic varieties of sparse and low-rank matrices playing a prominent role.

Solving the decomposition problem is NP-hard in general. A reasonable first approach might be to minimize $\gamma|\mathrm{support}(A)| + \mathrm{rank}(B)$ subject to the constraint that $A+B = C$, where γ serves as a tradeoff between sparsity and rank. This problem is combinatorially complex and intractable to solve in general; we propose a tractable convex optimization problem where the objective is a convex relaxation of $\gamma|\mathrm{support}(A)|+\mathrm{rank}(B)$. We relax $|\mathrm{support}(A)|$ by replacing it with the ℓ_1 norm $\|A\|_1$, which is the sum of the absolute values of the entries of A. We relax $\mathrm{rank}(B)$ by replacing it with the *nuclear norm* $\|B\|_*$, which is the sum of the singular values of B. Notice that the nuclear norm can be viewed as an "ℓ_1 norm" applied to the singular values (recall that the rank of a matrix is the number of non-zero singular values). The ℓ_1 and nuclear norms have been shown to be effective surrogates for $|\mathrm{support}(\cdot)|$ and $\mathrm{rank}(\cdot)$, and a number of results give conditions under which these relaxations recover sparse [2], [7], [6], [5], [4] and low-rank [1], [8], [14] objects. Thus we aim to decompose C into its components $A^\star$ and $B^\star$ using the following convex relaxation:

$$(\hat{A}, \hat{B}) = \arg\min_{A,B} \quad \gamma\|A\|_1 + \|B\|_* \qquad \text{s.t.} \quad A + B = C. \tag{1}$$

One can transform (1) into a semidefinite program (SDP) [18], for which there exist polynomial-time general-purpose solvers. We show that under certain conditions on sparse and low-rank matrices $(A^\star, B^\star)$ the *unique optimum* of the SDP (1) is $(\hat{A}, \hat{B}) = (A^\star, B^\star)$. In fact

Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky are with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 (venkatc@mit.edu; parrilo@mit.edu; willsky@mit.edu). Sujay Sanghavi is with the Department of Electrical and Computer Engineering, University of Texas – Austin, Austin, TX 78712 (sanghavi@mail.utexas.edu). This work was supported by MURI AFOSR grant FA9550-06-1-0324, MURI AFOSR grant FA9550-06-1-0303, and NSF FRG 0757207.

the conditions for exact recovery are simply a mild tightening of the conditions for fundamental identifiability. Essentially these conditions require that the sparse matrix does not have support concentrated within a single row/column, while the low-rank matrix does not have row/column spaces closely aligned with the coordinate axes. An interesting feature of our conditions is that no assumptions are required on the magnitudes of the non-zero values of the sparse matrix $A^\star$ or the singular values of the low-rank matrix $B^\star$. We also describe a method to determine the trade-off γ numerically given C. We do not give detailed proofs of our results in this paper. Many of these results appear with proofs in a longer report [3].

Applications: We briefly outline various applications of our method; see [3] for more details. In a statistical model selection setting, the sparse matrix can correspond to a Gaussian graphical model [11] and the low-rank matrix can summarize the effect of latent, unobserved variables. Decomposing a given model into these simpler components is useful for developing efficient estimation and inference algorithms. In computational complexity, the notion of *matrix rigidity* [17] captures the smallest number of entries of a matrix that must be changed in order to reduce the rank of the matrix below a specified level (the changes can be of arbitrary magnitude). Bounds on the rigidity of a matrix have several implications in complexity theory [13]. Similarly, in a system identification setting the low-rank matrix represents a system with a small model order while the sparse matrix represents a system with a sparse impulse response. Decomposing a system into such simpler components can be used to provide a simpler, more efficient description. In optics, many real-world imaging systems are efficiently described as a sum of a diagonal matrix (representing a so-called "incoherent" system) and a low-rank matrix (representing a "coherent" component) [9]. Our results provide a tractable method to describe a composite optical system in terms of simpler component systems. More generally, our approach also extends the applicability of rank minimization, such as in problems in spectral data analysis.

II. Identifiability

As described in the introduction, the matrix decomposition problem is ill-posed in the absence of additional conditions. In this section we discuss and quantify the further assumptions required on sparse and low-rank matrices for this decomposition to be unique. Throughout this paper, we restrict ourselves to $n \times n$ matrices to avoid cluttered notation. All our analysis extends to rectangular $n_1 \times n_2$ matrices if we simply replace n by $\max(n_1, n_2)$.

A. Preliminaries

We begin with a brief description and properties of the algebraic varieties of sparse and low-rank matrices. An algebraic variety is the solution set of a system of polynomial equations [10]. Sparse matrices constrained by the size of their support can be viewed as algebraic varieties:

$$\mathcal{S}(m) \triangleq \{M \in \mathbb{R}^{n\times n} \mid |\text{support}(M)| \leq m\}. \quad (2)$$

The dimension of this variety is m. In fact $\mathcal{S}(m)$ can be thought of as a union of $\binom{n^2}{m}$ subspaces, with each subspace being aligned with m of the n^2 coordinate axes. To see that $\mathcal{S}(m)$ is a variety, we note that a union of varieties is also a variety and that each of the $\binom{n^2}{m}$ subspaces in $\mathcal{S}(m)$ can be described by a system of linear equations. For any matrix $M \in \mathbb{R}^{n\times n}$, the *tangent space* $\Omega(M)$ with respect to $\mathcal{S}(|\text{support}(M)|)$ at M is given by

$$\Omega(M) = \{N \mid \text{support}(N) \subseteq \text{support}(M),\ N \in \mathbb{R}^{n\times n}\}. \quad (3)$$

If $|\text{support}(M)| = m$ the dimension of $\Omega(M)$ is m. We view $\Omega(M)$ as a *subspace* in $\mathbb{R}^{n\times n}$.

Next the variety of rank-constrained matrices is defined as:

$$\mathcal{R}(k) \triangleq \{M \in \mathbb{R}^{n\times n} \mid \text{rank}(M) \leq k\}. \quad (4)$$

The dimension of this variety is $k(2n-k)$. To see that $\mathcal{R}(k)$ is a variety, note that the determinant of any $(k+1)\times(k+1)$ submatrix of a matrix in $\mathcal{R}(k)$ must be zero. As the determinant of any submatrix is a polynomial in the elements of the matrix, $\mathcal{R}(k)$ can be described as the solution set of a system of polynomial equations. For any matrix $M \in \mathbb{R}^{n\times n}$, the tangent space $T(M)$ with respect to $\mathcal{R}(\text{rank}(M))$ at M is the span of all matrices with either the same row-space as M or the same column-space as M. Specifically, let $M = U\Sigma V^T$ be the singular value decomposition (SVD) of M with $U, V \in \mathbb{R}^{n\times k}$, where $\text{rank}(M) = k$. Then we have that

$$T(M) = \{UX^T + YV^T \mid X, Y \in \mathbb{R}^{n\times k}\}. \quad (5)$$

The dimension of $T(M)$ is $k(2n-k)$. As before we view $T(M)$ as a *subspace* in $\mathbb{R}^{n\times n}$. Since both $T(M)$ and $\Omega(M)$ are subspaces of $\mathbb{R}^{n\times n}$, we can compare vectors in these subspaces. For more details and geometric intuition on these algebraic varieties and their tangent spaces, we refer the reader to our longer report [3].

B. Identifiability issues

We describe two situations in which identifiability issues arise. These examples suggest the kinds of additional conditions that are required in order to ensure that there exists a unique decomposition into sparse and low-rank matrices.

First let $A^\star$ be any sparse matrix and let $B^\star = e_i e_j^T$, where e_i represents the i'th standard basis vector. In this case the rank-1 matrix $B^\star$ is also very sparse, and a valid sparse-plus-low-rank decomposition might be $\hat{A} = A^\star + e_i e_j^T$ and $\hat{B} = 0$. Thus, we need conditions that ensure that the low-rank matrix is not too sparse. For any matrix M, consider the following quantity with respect to the tangent space $T(M)$:

$$\xi(M) \triangleq \max_{N \in T(M),\ \|N\| \leq 1} \|N\|_\infty. \tag{6}$$

Here $\|\cdot\|$ is the spectral norm (i.e., the largest singular value), and $\|\cdot\|_\infty$ denotes the largest entry in magnitude. Thus $\xi(M)$ being small implies that elements of the tangent space $T(M)$ cannot have their support concentrated in a few locations; as a result M cannot be very sparse. We formalize this idea by relating $\xi(M)$ to a notion of "incoherence" of the row/column spaces, where we view row/column spaces as being incoherent with respect to the standard basis if these spaces are not aligned closely with any of the coordinate axes. Letting $M = U\Sigma V^T$ be the singular value decomposition of M, we measure the incoherence of the row/column spaces of M as:

$$\mathrm{inc}(M) \triangleq \max \left[\max_i \|P_U e_i\|_2, \max_i \|P_V e_i\|_2 \right]. \tag{7}$$

Here $\|\cdot\|_2$ represent the vector ℓ_2 norm, and P_V, P_U denote projections onto the row/column spaces. Hence $\mathrm{inc}(M)$ measures the projection of the most "closely aligned" coordinate axis with the row/column spaces. For any rank-k matrix M we have that $\sqrt{\frac{k}{n}} \leq \mathrm{inc}(M) \leq 1$, where the lower bound is achieved (for example) if the row/column spaces span any k columns of an $n \times n$ orthonormal Hadamard matrix, while the upper bound is achieved if the row or column space contains a standard basis vector. Typically a matrix M with incoherent row/column spaces would have $\mathrm{inc}(M) \ll 1$. The following result shows that the more incoherent the row/column spaces of M, the smaller is $\xi(M)$.

Proposition 1: For any $M \in \mathbb{R}^{n \times n}$, we have that

$$\mathrm{inc}(M) \leq \xi(M) \leq 2\ \mathrm{inc}(M),$$

where $\xi(M)$ and $\mathrm{inc}(M)$ are defined in (6) and (7).

Example: If $M \in \mathbb{R}^{n \times n}$ is a full-rank matrix or a matrix such as $e_i e_j^T$, then $\xi(M) = 1$. Thus a bound on the incoherence of the row/column spaces of M is important in order to bound ξ.

Next consider the scenario in which $B^\star$ is any low-rank matrix and $A^\star = -v e_1^T$ with v being the first column of $B^\star$. Thus, $C = A^\star + B^\star$ has zeros in the first column, $\mathrm{rank}(C) = \mathrm{rank}(B^\star)$, and C has the same column space as $B^\star$. A reasonable sparse-plus-low-rank decomposition in this case might be $\hat{B} = B^\star + A^\star$ and $\hat{A} = 0$. Here $\mathrm{rank}(\hat{B}) = \mathrm{rank}(B^\star)$. Requiring that a sparse matrix $A^\star$ have "bounded degree" (i.e., few non-zero entries per row/column) avoids such identifiability issues. For any matrix M, we define the following quantity with respect to the tangent space $\Omega(M)$:

$$\mu(M) \triangleq \max_{N \in \Omega(M),\ \|N\|_\infty \leq 1} \|N\|. \tag{8}$$

The quantity $\mu(M)$ being small for a matrix implies that the *spectrum* of any element of the tangent space $\Omega(M)$ is not too "concentrated", i.e., the singular values of these elements are not too large. We show in the following proposition that a sparse matrix M with "bounded degree" (a small number of non-zeros per row/column) has small $\mu(M)$.

Proposition 2: Let $M \in \mathbb{R}^{n \times n}$ be any matrix with at most $\mathrm{deg}_{\max}(M)$ non-zero entries per row/column, and with at least $\mathrm{deg}_{\min}(M)$ non-zero entries per row/column. With $\mu(M)$ as defined in (8), we have that

$$\mathrm{deg}_{\min}(M) \leq \mu(M) \leq \mathrm{deg}_{\max}(M).$$

Example: Note that if $M \in \mathbb{R}^{n \times n}$ has full support, i.e., $\Omega(M) = \mathbb{R}^{n \times n}$, then $\mu(M) = n$. Therefore, a constraint on the number of zeros per row/column provides a useful bound on μ. We emphasize here that simply bounding the number of non-zero entries in M does not suffice; the *sparsity pattern* also plays a role in determining the value of μ.

III. Rank-Sparsity Uncertainty Principle and Exact Recovery

In this section we show that sparse matrices $A^\star$ with small $\mu(A^\star)$ and low-rank matrices $B^\star$ with small $\xi(B^\star)$ are identifiable given $C = A^\star + B^\star$, and can in fact be exactly recovered using the SDP (1).

A. Tangent-space identifiability

Before analyzing whether $(A^\star, B^\star)$ can be recovered in general (for example, using the SDP (1)), we ask a simpler question. Suppose that we had prior information about the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$, in addition to being given $C = A^\star + B^\star$. Can we then *uniquely* recover $(A^\star, B^\star)$ from C? Assuming such prior knowledge of the tangent spaces is unrealistic in practice as it is equivalent to assuming prior knowledge of the support of $A^\star$ and the row/column spaces of $B^\star$; however, we obtain useful insight into the kinds of conditions required on sparse and low-rank matrices for exact decomposition. Given this knowledge of the tangent spaces, a necessary and sufficient condition for unique recovery is that the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$ intersect *transversally*:

$$\Omega(A^\star) \cap T(B^\star) = \{0\}. \tag{9}$$

That is, the subspaces $\Omega(A^\star)$ and $T(B^\star)$ have a trivial intersection. The sufficiency of this condition for unique decomposition is easily seen. For the necessity part, suppose for the sake of a contradiction that a non-zero matrix M belongs to $\Omega(A^\star) \cap T(B^\star)$; one can add and subtract M from $A^\star$ and $B^\star$ respectively while still having a valid decomposition, which violates the uniqueness requirement. In fact the transverse intersection of the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$ described in (9) is also one of the conditions required for $(A^\star, B^\star)$ to be the *unique* optimum of the SDP (1) [3]. The following proposition provides a simple condition in terms of $\mu(A^\star)$ and $\xi(B^\star)$ for the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$ to intersect transversally.

Proposition 3: For any two matrices $A^\star$ and $B^\star$, we have that

$$\mu(A^\star)\xi(B^\star) < 1 \Rightarrow \Omega(A^\star) \cap T(B^\star) = \{0\},$$

where $\xi(B^\star)$ and $\mu(A^\star)$ are defined in (6) and (8), and the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$ are defined in (3) and (5).

Thus, both $\mu(A^\star)$ and $\xi(B^\star)$ being small implies that the spaces $\Omega(A^\star)$ and $T(B^\star)$ intersect transversally; consequently, we can exactly recover $(A^\star, B^\star)$ given $\Omega(A^\star)$ and $T(B^\star)$. In the following section we show that a slight tightening of the condition in Proposition 3 for identifiability is also sufficient to guarantee exact recovery of $(A^\star, B^\star)$ using the SDP (1).

Another important consequence of Proposition 3 is that we have an elementary proof of the following rank-sparsity uncertainty principle.

Theorem 1: For any matrix $M \neq 0$, we have that

$$\xi(M)\mu(M) \geq 1,$$

where $\xi(M)$ and $\mu(M)$ are as defined in (6) and (8) respectively.

Proof: Given any $M \neq 0$ it is clear that $M \in \Omega(M) \cap T(M)$, i.e., M is an element of both tangent spaces. However $\mu(M)\xi(M) < 1$ would imply from Proposition 3 that $\Omega(M) \cap T(M) = \{0\}$, which is a contradiction. Consequently, we must have that $\mu(M)\xi(M) \geq 1$. □

Hence, for *any* matrix $M \neq 0$ both $\mu(M)$ and $\xi(M)$ cannot be small. Note that Proposition 3 is an assertion involving μ and ξ for (in general) *different* matrices, while Theorem 1 is a statement about μ and ξ for the *same* matrix. Essentially the uncertainty principle asserts that no matrix can be too sparse while having "incoherent" row and column spaces. An extreme example is the matrix $e_i e_j^T$, which has the property that $\mu(e_i e_j^T)\xi(e_i e_j^T) = 1$.

B. Exact recovery using semidefinite program

Our main result is the following simple, deterministic sufficient condition for exact recovery using the SDP (1).

Theorem 2: Given $C = A^\star + B^\star$, if

$$\mu(A^\star)\xi(B^\star) < \frac{1}{8}$$

then the *unique optimum* $(\hat{A}, \hat{B})$ of (1) is $(A^\star, B^\star)$ for the following range of γ:

$$\gamma \in \left(\frac{\xi(B^\star)}{1 - 6\mu(A^\star)\xi(B^\star)}, \frac{1 - 4\mu(A^\star)\xi(B^\star)}{\mu(A^\star)} \right).$$

The proof essentially involves verifying the subgradient optimality conditions of the SDP (1) [15], [3]. Comparing with Proposition 3, we see that the condition for exact recovery is only slightly stronger than that required for identifiability. Therefore sparse matrices $A^\star$ with small $\mu(A^\star)$ and low-rank matrices $B^\star$ with small $\xi(B^\star)$ can be recovered exactly from $C = A^\star + B^\star$ using a tractable convex program.

Using Propositions 1 and 2 along with Theorem 2 we have the following result, which gives more concrete classes of sparse and low-rank matrices that can be exactly recovered.

Corollary 3: Suppose $A^\star$ and $B^\star$ are such that $\deg_{\max}(A^\star)\ \mathrm{inc}(B^\star) < \frac{1}{16}$, where these quantities are defined in Propositions 1 and 2. Then given $C = A^\star + B^\star$ the unique optimum of the SDP (1) is $(\hat{A}, \hat{B}) = (A^\star, B^\star)$ for a range of γ (which can be computed from Propositions 1 and 2, and Theorem 2).

Therefore sparse matrices with bounded degree (i.e., support not too concentrated in any row/column) and low-rank matrices with row/column spaces not closely aligned with the coordinate axes can be uniquely decomposed. We emphasize here that our results provide *deterministic* sufficient conditions for exact recovery. We also note that these conditions only involve the sparsity pattern of $A^\star$ and the row/column spaces of $B^\star$. There is *no dependence* on the non-zero entries of $A^\star$ or the singular values of $B^\star$. The reason for this is that the subgradient optimality conditions for (1) only involve the tangent spaces $\Omega(A^\star)$ and $T(B^\star)$, and not the specific non-zero entries of $A^\star$ or the singular values of $B^\star$ [3].

IV. Simulation Results

We confirm the theoretical predictions in this paper with some simple experimental results. In these experiments we generate a random rank-k matrix $B^\star$ in $\mathbb{R}^{n \times n}$

as follows: we generate random $X, Y \in \mathbb{R}^{n\times k}$ with i.i.d. Gaussian entries and set $B^\star = XY^T$. We generate a random m-sparse matrix $A^\star$ by choosing the support set of size m uniformly at random, and setting the values within this support to be i.i.d. Gaussian. All our simulations were performed using YALMIP [12] and the SDPT3 software [16] for solving SDPs.

We begin by presenting a heuristic to choose the trade-off parameter γ. Based on Theorem 2 we know that exact recovery is possible for a *range* of γ. Therefore, one can simply check the stability of the solution $(\hat{A}, \hat{B})$ as γ is varied without prior knowledge of the appropriate value for γ. To formalize this scheme we consider the following equivalent SDP for $t \in [0, 1]$:

$$\begin{aligned}(\hat{A}_t, \hat{B}_t) = \arg\min_{A,B} \quad & t\|A\|_1 + (1-t)\|B\|_* \\ \text{s.t.} \quad & A + B = C.\end{aligned} \tag{10}$$

There is a one-to-one correspondence between (1) and (10) given by $t = \frac{\gamma}{1+\gamma}$. The benefit of (10) is that the range of valid parameters is compact, i.e., $t \in [0, 1]$, as opposed to (1) where $\gamma \in [0, \infty)$. Let tol_t be defined as:

$$\mathrm{tol}_t = \frac{\|\hat{A}_t - A^\star\|_F}{\|A^\star\|_F} + \frac{\|\hat{B}_t - B^\star\|_F}{\|B^\star\|_F}, \tag{11}$$

where $(\hat{A}_t, \hat{B}_t)$ is the solution of (10), and $\|\cdot\|_F$ is the Frobenius norm. We compute the difference between solutions for some t and $t-\epsilon$ as follows:

$$\mathrm{diff}_t = (\|\hat{A}_{t-\epsilon} - \hat{A}_t\|_F) + (\|\hat{B}_{t-\epsilon} - \hat{B}_t\|_F), \tag{12}$$

where $\epsilon > 0$ is some small fixed stepsize, say $\epsilon = 0.01$. We generate a random $A^\star \in \mathbb{R}^{25\times 25}$ that is 25-sparse and a random $B^\star \in \mathbb{R}^{25\times 25}$ with rank $= 2$ as described above. Given $C = A^\star + B^\star$, we solve (10) for various values of t. Figure 1 on the left shows two curves – one is tol_t and the other is diff_t. Clearly we do not have access to tol_t in practice. However, we see that diff_t is near-zero in exactly three regions. For sufficiently small t the optimal solution to (10) is $(\hat{A}_t, \hat{B}_t) = (A^\star + B^\star, 0)$, while for sufficiently large t the optimal solution is $(\hat{A}_t, \hat{B}_t) = (0, A^\star + B^\star)$. As seen in the figure, diff_t stabilizes for small and large t. The third "middle" range of stability is where we typically have $(\hat{A}_t, \hat{B}_t) = (A^\star, B^\star)$. Notice that outside of these three regions diff_t is not close to 0 and in fact changes rapidly. Therefore if a reasonable guess for t (or γ) is not available, one one could solve (10) for a range of t and choose a solution corresponding to the "middle" range in which diff_t is stable and near zero.

Next we generate random 25×25 rank-k matrices $B^\star$ and m-sparse matrices $A^\star$ as described above, for various values of k and m. The goal is to recover $(A^\star, B^\star)$ from $C = A^\star + B^\star$ using the SDP (1). We declare success in recovering $(A^\star, B^\star)$ if $\mathrm{tol}_\gamma < 10^{-3}$ (where tol_γ is defined analogous to tol_t in (11)). Figure 1 on the right shows the success rate in recovering $(A^\star, B^\star)$ for various values of m and k (averaged over 10 experiments for each m, k). Thus we see that one can recover sufficiently sparse $A^\star$ and sufficiently low-rank $B^\star$ from $C = A^\star + B^\star$ using (1).

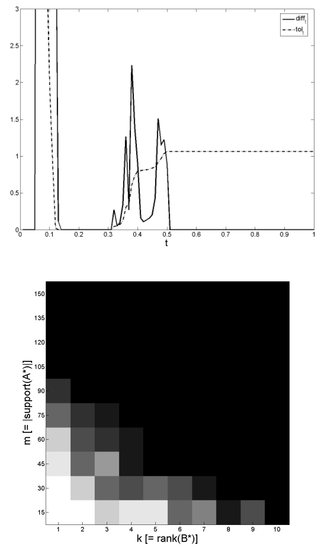

Fig. 1. (Left) Comparison between tol_t and diff_t for a randomly generated 25×25 example with $\mathrm{support}(A^\star) = 25$ and $\mathrm{rank}(B^\star) = 2$. (Right) We generate random m-sparse $A^\star$ and random rank-k $B^\star$ of size 25×25, and attempt to recover $(A^\star, B^\star)$ from $C = A^\star + B^\star$ using (1). For each value of m, k we repeated this procedure 10 times. The figure shows the probability of success in recovering $(A^\star, B^\star)$ using (1) for various values of m and k. White represents a probability of success of 1, while black represents a probability of success of 0.

V. Discussion

This paper studied the problem of exactly decomposing a given matrix $C = A^\star + B^\star$ into its sparse and low-rank components $A^\star$ and $B^\star$. Based on a notion of rank-sparsity incoherence, we characterized fundamental identifiability as well as exact recovery using a tractable convex program; the incoherence property relates the sparsity pattern of a matrix and its row/column spaces via a new uncertainty principle. Our results have applications in fields as diverse as machine learning, complexity theory, optics, and system identification. Our work opens many interesting research avenues: *(a)* modifying our method for specific applications and characterizing the resulting improved performance, *(b)* understanding rank-sparsity uncertainty principles more generally, and *(c)* developing lower-complexity decomposition algorithms that take advantage of special structure in (1), which general-purpose SDP solvers do not.

References

[1] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Submitted for publication*, 2008.

[2] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2), 2006.

[3] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *Technical report*, 2009.

[4] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 2006.

[5] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6), 2006.

[6] D. L. Donoho and M. Elad. Optimal sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, 100, 2003.

[7] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7), 2001.

[8] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[9] M. Fazel and J. Goodman. Approximations for partially coherent optical imaging systems. *Technical Report, Department of Electrical Engineering, Stanford University*, 1998.

[10] J. Harris. *Algebraic Geometry: A First Course*. Springer-Verlag, 1995.

[11] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[12] J. Lfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, 2004.

[13] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *36th IEEE Symposium on Foundations of Computer Science (FOCS)*, 1995.

[14] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *Submitted to SIAM Review*, 2007.

[15] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996.

[16] K. C. Toh, M. J. Todd, and R. H. Tutuncu. Sdpt3 - a matlab software package for semidefinite-quadratic-linear programming. In *Available from http://www.math.nus.edu.sg/ mattohkc/sdpt3.html*.

[17] L. G. Valiant. Graph-theoretic arguments in low-level complexity. In *6th Symposium on Mathematical Foundations of Computer Science*, 1977.

[18] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1), 1996.