Sparse representation of solutions of Kronecker product systems *

Sadegh Jokar † Volker Mehrmann†

31.07.08

Abstract

Three properties of matrices: the spark, the mutual incoherence and the restricted isometry property have recently been introduced in the context of compressed sensing. We study these properties for matrices that are Kronecker products and show how these properties relate to those of the factors. For the mutual incoherence we also discuss results for sums of Kronecker products.

Keywords Spark, mutual incoherence, restricted isometry property, compressed sensing, Kronecker product, sparse solution of linear systems.
AMS subject classification. 15A06, 65F50, 65F05, 15A15, 15A21.

Dedicated to Shmuel Friedland on the occasion of his 65th birthday.

1 Introduction

In this paper we discuss the computation of sparse solutions of underdetermined linear systems

$$Ax = b,$$

where $A \in \mathbb{R}^{m,n}$, with $m \leq n$ is given as a Kronecker product, i.e.

$$A = A_1 \otimes A_2 \otimes \ldots \otimes A_N, \quad A_i \in \mathbb{R}^{m_i,n_i}, i = 1, \ldots, N, \tag{1}$$

*Supported by *Deutsche Forschungsgemeinschaft* through the DFG Research Center Matheon *Mathematics for key technologies* in Berlin.

†Institut für Mathematik, TU Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany. `{jokar,mehrmann}@math.tu-berlin.de`.

or as a sum of Kronecker products

$$A = \sum_{j=1}^{M} A_1^j \otimes A_2^j \otimes \ldots \otimes A_N^j \quad A_i^j \in \mathbb{R}^{m_{i,j}, n_{i,j}}. \tag{2}$$

Since the solution is typically non-unique it is an important topic in many applications, in particular in optimal signal recovery and in compressed sensing, see e.g. [1, 3, 4, 5, 6, 9, 10, 13, 20] to find the sparsest solution,

$$\min \|x\|_0, \ \ s.t. \ Ax = b, \tag{3}$$

where $\|x\|_0$ denotes the number of nonzero entries of a vector x, see Section 2.

In general, the problem of finding the sparsest solution is known to be NP-hard [22]. However, in the context of compressed sensing, conditions have been derived on the size of the *support of* x, i.e. the number of nonzero elements of x, that allow one to compute the sparsest solution using ℓ_1-minimization via the so called *basis pursuit algorithm* [3, 5, 7, 8, 10, 11, 12], i.e, by computing

$$\min \|x\|_1, \ \ s.t. \ Ax = b, \tag{4}$$

where $\|x\|_1 = \sum_i |x_i|$.

Sufficient conditions for this approach to work are that some properties of the matrix A called *spark* [10, 25], *mutual incoherence* [7, 12] or the *restricted isometry property (RIP)* [2, 3, 4] are restricted. We will introduce these properties in Section 2.

For general matrices it is possible (though expensive) to determine the mutual incoherence, while analyzing the spark or the restricted isometry property is difficult. If, however, the matrix A has the form (1) then we show in Section 3 that these properties can be easily derived from the corresponding properties of the factors. For the mutual incoherence we can also extend these results to matrices of the form (2).

2 Notation and preliminaries

For $m, n \in \mathbb{N}$, where $\mathbb{N} = \{1, 2, \ldots\}$, we denote by $\mathbb{R}^{m,n}$ the set of real $m \times n$ matrices, by I_n the $n \times n$ identity matrix, and by $\langle \cdot, \cdot \rangle$ the Euclidean inner product in $\mathbb{R}^n$. For $1 \leq p \leq \infty$, the ℓ_p-norm of $x \in \mathbb{R}^n$ is defined by

$$\|x\|_p := \big(\sum_{j=1}^{n} |x_j|^p \big)^{\frac{1}{p}},$$

with the special case

$$\|x\|_\infty := \max_{j\in\{1,\dots,n\}} |x_j|,$$

if $p = \infty$. Finally, for $x \in \mathbb{R}^n$, we introduce the notation

$$\|x\|_0 := \#\operatorname{supp}(x),$$

where $\operatorname{supp}(x) := \{j \in \{1,\dots,n\} : x_j \neq 0\}$ is the *support* of x. Note that $\|\cdot\|_0$ is not a norm, since for $\alpha \neq 0$ we have $\|\alpha x\|_0 = \|x\|_0$. We use the term *k-sparse* for all vectors x such that $\|x\|_0 \leq k$.

Definition 2.1 *[19, 21] The* Kronecker product *of $A = [a_{i,j}] \in \mathbb{R}^{p,q}$ and $B = [b_{i,j}] \in \mathbb{R}^{r,s}$ is denoted by $A \otimes B$ and is defined to be the block matrix*

$$A \otimes B := \begin{bmatrix} a_{1,1}B & \cdots & a_{1,q}B \\ \vdots & \ddots & \vdots \\ a_{p,1}B & \cdots & a_{p,q}B \end{bmatrix} \in \mathbb{R}^{pr,qs}.$$

Let $C = [c_1 \cdots c_r] \in \mathbb{R}^{q,r}$ with columns $c_i \in \mathbb{R}^q$, $1 \leq i \leq r$. Then,

$$\operatorname{vec}(C) := \begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix} \in \mathbb{R}^{qr}.$$

It is well known [21] that the matrix equation $AXB = C$, with matrices of appropriate dimensions, is equivalent to the linear system

$$(B^{\mathrm{T}} \otimes A)\operatorname{vec}(X) = \operatorname{vec}(C).$$

Furthermore, using the perfect shuffle permutation matrices Π_1, Π_2, we have that $\Pi_1(A \otimes B)\Pi_2 = B \otimes A$, see [21].

As our first special property we introduce the spark of a matrix.

Definition 2.2 *[10, 25] Let $A = [a_1, \dots, a_n] \in \mathbb{R}^{m,n}$, $2 \leq m \leq n$ have columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \dots, n$. The* spark *of A, denoted as* $\operatorname{spark}(A)$ *is defined as the cardinality of the smallest subset of linearly dependent columns of A.*

In other words, if all r-dimensional subsets of column vectors of A are linearly independent, but there exists a subset of $r + 1$ columns that are linearly dependent, then $\operatorname{spark}(A) = r + 1$. For convenience, if $m = n = 1$, we define $\operatorname{spark}(A) := 1$, and in the case where $m = n \geq 2$ and A is invertible, we set $\operatorname{spark}(A) := n + 1$. In general the spark and the rank of a matrix $A \in \mathbb{R}^{m,n}$ with $m \geq 2$, are related via

$$2 \leq \operatorname{spark}(A) \leq \operatorname{rank}(A) + 1.$$

Example 2.3 *If*

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix},$$

then $\operatorname{spark}(A) = \operatorname{rank}(A) + 1 = 3$. *On the other hand, if*

$$A = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

then $\operatorname{spark}(A) = 2$.

The quantity $\operatorname{spark}(A)$ can be used to derive necessary and sufficient conditions for the existence of sparse solutions.

Lemma 2.4 *[10, 16] Consider the linear system $Ax = b$ with $A \in \mathbb{R}^{m,n}$, $m \leq n$. A necessary and sufficient condition for the linear system $Ax = b$ to have a unique k-sparse solution x is that $k \leq \operatorname{spark}(A)/2$.*

The second property that we study is the mutual incoherence.

Definition 2.5 *[12] Let $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m,n}$, $m \leq n$ have columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \ldots, n$. Then the* mutual incoherence *$\mathcal{M}(A)$ is defined by*

$$\mathcal{M}(A) := \max_{i \neq j} |\langle a_i, a_j \rangle| = \max_{i \neq j} |(A^{\mathrm{T}} A)_{i,j}|.$$

Note that, since the columns of A are normalized, by the triangle inequality we always have $\mathcal{M}(A) \leq 1$. On the other hand, if A has orthonormal columns, then $\mathcal{M}(A) = 0$.

We have the following lower bound for $\mathcal{M}(A)$.

Lemma 2.6 *[24] Suppose that $A \in \mathbb{R}^{m,n}$, $m \leq n$ has columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \ldots, n$ and suppose further that A has full row rank. Then*

$$\mathcal{M}(A) \geq \sqrt{\frac{n-m}{m(n-1)}}.$$

The following lemma relates the sparsest solution as defined in (3) and the ℓ_1-solution as defined in (4) of the linear equation $Ax = b$ in terms of the mutual incoherence of a matrix A.

Lemma 2.7 *[10, 15] Suppose that $A \in \mathbb{R}^{m,n}$, $m \leq n$ has columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \ldots, n$. If b is a vector such that the equation $Ax = b$ has a solution satisfying*

$$\|x\|_0 < \frac{1 + \frac{1}{\mathcal{M}(A)}}{2},$$

then the ℓ_1-norm minimal solution in (4) coincides with the ℓ_0-minimal solution in (3).

Remark 2.8 *Consider matrices of the form $A = [\Phi \ \Psi]$, where Φ and Ψ have orthonormal columns. If the sparsest solution x of $Ax = b$ satisfies*

$$\|x\|_0 < \frac{\sqrt{2} - \frac{1}{2}}{\mathcal{M}(A)},$$

then it has been shown in [14] that the solutions of the ℓ_1-norm minimization problem and ℓ_0-norm minimization problem coincide.

The third quantity that is important in the context of sparse recovery and compressed sensing is the *restricted isometry property.*

Definition 2.9 *[2, 3, 4, 5] Let $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m,n}$, $m \leq n$ have columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \ldots, n$.*

The k-restricted isometry constant of A *is the smallest number δ_k such that*

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

for all $x \in \mathbb{R}^n$ with $\|x\|_0 \leq k$.

The k-restricted isometry property requires that every set of columns of cardinality less than or equal to k approximately (with an error δ_k) behaves like an orthonormal basis.

The following lemma gives the relation between the sparsest solution (as defined in (3)) of a linear system $Ax = b$ and the ℓ_1-solution as defined in (4) in terms of the k-restricted isometry constant.

Lemma 2.10 *[2] Let $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m,n}$, $m \leq n$ have columns a_i that are normalized so that $\|a_i\|_2 = 1$, $i = 1, \ldots, n$.*

Suppose that

$$\delta_{2k} < \sqrt{2} - 1.$$

Then for all k-sparse solution vectors x of $Ax = b$ the solution of (4) is equal to the solution of (3).

After introducing the concepts of spark, mutual incoherence and k-restricted isometry property, in the next section we analyze these concepts for Kronecker product matrices.

3 Sparse representation and Kronecker Products of Matrices

In this section we study sparse solutions for linear system $Ax = b$, where the matrix A is given as a Kronecker product (1).

Our first result characterizes $\text{spark}(A \otimes B)$ in terms of $\text{spark}(A)$ and $\text{spark}(B)$. Note that if A, B have normalized columns then $A \otimes B$ has normalized columns as well.

Theorem 3.1 *Let $A = [a_1, \ldots, a_q] \in \mathbb{R}^{p,q}$ and $B = [b_1, \ldots, b_s] \in \mathbb{R}^{r,s}$ be rank-deficient matrices with normalized columns, i.e., $\|a_i\|_2 = 1$, $i = 1, \ldots, q$, $\|b_i\|_2 = 1$, $i = 1, \ldots, s$. Then*

$$\text{spark}(A \otimes B) = \text{spark}(B \otimes A) = \min\{\text{spark}(A), \text{spark}(B)\}. \tag{5}$$

If A is an invertible matrix and B is rank-deficient matrix, then

$$\text{spark}(A \otimes B) = \text{spark}(B). \tag{6}$$

If both A and B are square and invertible then

$$\text{spark}(A \otimes B) = (\text{spark}(A) - 1)(\text{spark}(B) - 1) + 1 = qs + 1.$$

Proof. Using the fact that $(B \otimes A)\text{vec}(X) = \Pi_1(A \otimes B)\Pi_2\text{vec}(X)$ and $\|\text{vec}(X)\|_0 = \|\Pi_2\text{vec}(X)\|_0$, we have $\text{spark}(A \otimes B) = \text{spark}(B \otimes A)$.

Consider first the case that A and B are rank-deficient. By the definition of $\text{spark}(B)$, there exists a vector $y \in \mathbb{R}^s$ with $\|y\|_0 = \text{spark}(B)$ such that $By = 0$. With

$$\hat{X} = [\ y \quad 0 \quad \cdots \quad 0\ ],$$

we have that $(A \otimes B)\text{vec}(\hat{X}) = 0$ and $\|\text{vec}(\hat{X})\|_0 = \|y\|_0 = \text{spark}(B)$. This means that $\text{spark}(A \otimes B) \leq \text{spark}(B)$. Using that $\text{spark}(A \otimes B) = \text{spark}(B \otimes A)$ and that also A is rank-deficient, we can apply the same argument as before and get $\text{spark}(A \otimes B) \leq \text{spark}(A)$. Therefore,

$$\text{spark}(A \otimes B) \leq \min\{\text{spark}(A), \text{spark}(B)\}. \tag{7}$$

Let $C = A \otimes B$, then every column of C has the form $c_j = a_{u_j} \otimes b_{v_j}$. To prove equality in (5), we assume w.l.o.g. that

$$\text{spark}(B) \leq \text{spark}(A). \tag{8}$$

Then by (7) we have $\text{spark}(A \otimes B) \leq \text{spark}(B)$. Suppose now that

$$\text{spark}(A \otimes B) = \ell < \text{spark}(B). \tag{9}$$

This implies, in particular, that any set of ℓ columns of B is linearly independent, while there exist scalars $\lambda_1, \ldots, \lambda_\ell$ not all 0 and indices $u_1, \ldots, u_\ell$ where $u_i \neq u_j$ for all $i \neq j$, and $v_1, \ldots, v_\ell$ such that

$$\sum_{j=1}^{\ell} (a_{u_j} \otimes b_{v_j})\lambda_j = \sum_{j=1}^{\ell} (\lambda_j a_{u_j}) \otimes b_{v_j} = 0.$$

In this sum there may occur repeated copies of vectors b_j, so without loss of generality we may assume the indices v_i are numbered so that

$$\underbrace{v_1 = \cdots = v_{k_1}}_{g_1} < \underbrace{v_{k_1+1} = \cdots = v_{k_2}}_{g_2} < \cdots < \underbrace{v_{k_{t-1}+1} = \cdots = v_{k_t}}_{g_t}.$$

Therefore, we have

$$(\sum_{j=1}^{k_1} \lambda_j a_{u_j}) \otimes b_{g_1} + (\sum_{j=k_1+1}^{k_2} \lambda_j a_{u_j}) \otimes b_{g_2} + \cdots + (\sum_{j=k_{t-1}+1}^{k_t} \lambda_j a_{u_j}) b_{g_t} = 0, \quad (10)$$

where $k_t = \ell$. Since $b_{g_1}, \ldots, b_{g_t}$ are linearly independent, it follows that for all $1 \leq i \leq t$ we have

$$\sum_{j=k_{i-1}+1}^{k_i} \lambda_j a_{u_j} = 0,$$

where $k_0 = 0$. This contradicts the assumption in (8) that

$$\ell < \operatorname{spark}(B) \leq \operatorname{spark}(A),$$

because the u_j are pairwise distinct and at least one of the coefficients λ_j is nonzero.

Now suppose that A is invertible and B is rank-deficient. Then with the same argument as above, we have $\operatorname{spark}(A \otimes B) \leq \operatorname{spark}(B)$. Let $X = [x_1, \ldots, x_q] \neq 0$, such that $\|\operatorname{vec}(X)\|_0 = \operatorname{spark}(A \otimes B)$ and $(A \otimes B)\operatorname{vec}(X) = 0$. This implies that $BXA^{\mathrm{T}} = 0$, and, since A is invertible we have $BX = 0$, while on the other hand $X \neq 0$. Thus there exists at least one index i such that $x_i \neq 0$ and $Bx_i = 0$. Hence,

$$\operatorname{spark}(B) \leq \|x_i\|_0 \leq \|\operatorname{vec}(X)\|_0 = \operatorname{spark}(A \otimes B),$$

and therefore $\operatorname{spark}(A \otimes B) = \operatorname{spark}(B)$.

For the case where both A and B are invertible, $A \otimes B$ is invertible as well, see [21]. Therefore,

$$\operatorname{spark}(A \otimes B) = \operatorname{rank}(A \otimes B) + 1 = qs + 1.$$

□

We immediately have the following corollary of Theorem 3.1.

Corollary 3.2 *Consider rank-deficient matrices* $\{A_i\}_{i=1}^N$ *with normalized columns. Then*

$$\text{spark}(A_1 \otimes \ldots \otimes A_N) = \min_{1 \le i \le N} \{\text{spark}(A_i)\}.$$

By combining Lemma 2.4 and Corollary 3.2 we get the following Corollary:

Corollary 3.3 *Consider a linear system* $(A_1 \otimes \ldots \otimes A_N)x = b$ *with rank-deficient matrices* $A_i \in \mathbb{R}^{p_i, q_i}$ *that have normalized columns. A necessary and sufficient condition for this linear system to have a unique* k*-sparse solution* x *is that*

$$k \le \frac{\min\limits_{1 \le i \le N} \{\text{spark}(A_i)\}}{2}.$$

Remark 3.4 *Corollary 3.3 is saying that if one of the matrices* A_j *has small* spark *then we can only uniquely recover vectors of the sparsity up to* $\text{spark}(A_j)/2$ *in the linear system* $(A_1 \otimes \ldots \otimes A_N)x = b$.

Similar to the analysis of $\text{spark}(A \otimes B)$, we can also obtain an estimate of $\mathcal{M}(A \otimes B)$ in terms of $\mathcal{M}(A)$ and $\mathcal{M}(B)$.

Theorem 3.5 *Consider matrices* $A = [a_1, \ldots, a_{n_1}] \in \mathbb{R}^{m_1, n_1}$ *and* $B = [b_1, \ldots, b_{n_2}] \in \mathbb{R}^{m_2, n_2}$ *with normalized columns. Then*

$$\mathcal{M}(A \otimes B) = \max\{\mathcal{M}(A), \mathcal{M}(B)\}.$$

Proof. Suppose that $C = A \otimes B$ and $C = [c_1 \cdots c_n]$, where $c_i \in \mathbb{R}^m$, $m = m_1 m_2$ and $n = n_1 n_2$. Then we have $\mathcal{M}(C) = \max_{i \ne j} |\langle c_i, c_j \rangle|$. Since $c_i = a_p \otimes b_q$ and $c_j = a_r \otimes b_s$ for some p, q, r, s, using properties of the Kronecker product [21], we have

$$\langle c_i, c_j \rangle = \langle a_p \otimes b_q, a_r \otimes b_s \rangle = \langle a_p, a_r \rangle \cdot \langle b_q, b_s \rangle. \tag{11}$$

By Definition 2.5 and (11) we then have

$$\begin{aligned} \mathcal{M}(C) = \mathcal{M}(A \otimes B) &= \max_{\substack{p,q,r,s \\ (p,q) \ne (r,s)}} |\langle a_p, a_r \rangle \cdot \langle b_q, b_s \rangle| \\ &= \max_{\substack{p,q,r,s \\ p \ne r, q \ne s}} \{|\langle a_p, a_r \rangle \cdot \langle b_q, b_s \rangle|, |\langle a_p, a_r \rangle|, |\langle b_q, b_s \rangle|\}. \end{aligned} \tag{12}$$

On the other hand, since the matrices A and B have normalized columns, we have

$$|\langle a_p, a_r\rangle \cdot \langle b_q, b_s\rangle| \leq |\langle a_p, a_r\rangle|,$$

and similarly

$$|\langle a_p, a_r\rangle \cdot \langle b_q, b_s\rangle| \leq |\langle b_q, b_s\rangle|.$$

Therefore, from (12) we have

$$\begin{aligned}\mathcal{M}(A \otimes B) &= \max_{\substack{p,q,r,s \\ p\neq r, q\neq s}} \{|\langle a_p, a_r\rangle|, |\langle b_q, b_s\rangle|\} \\ &= \max\{\max_{p\neq r}|\langle a_p, a_r\rangle|, \max_{q\neq s}|\langle b_q, b_s\rangle|\} \\ &= \max\{\mathcal{M}(A), \mathcal{M}(B)\}.\end{aligned}$$

□

A direct consequence of this theorem is the following Corollary.

Corollary 3.6 *Consider matrices $\{A_i\}_{i=1}^N$ with normalized columns and let $A = A_1 \otimes \ldots \otimes A_N$. Then,*

$$\mathcal{M}(A) = \max_{1\leq i\leq n} \mathcal{M}(A_i).$$

Corollary 3.6 shows that if one of the matrices A_i has a large mutual incoherence, then it will dominate the mutual incoherence of A, regardless of all the other factors in the Kronecker product.

We also have a result that relates the k-restricted isometry constant of $\delta_k^{A\otimes B}$ to those of δ_k^A and δ_k^B.

Theorem 3.7 *Let $A \in \mathbb{R}^{p,q}$ and $B \in \mathbb{R}^{r,s}$ have normalized columns. Then*

$$\delta_k^{A\otimes B} = \delta_k^{B\otimes A} \geq \max\{\delta_k^A, \delta_k^B\}. \tag{13}$$

Proof. Using the fact that $B \otimes A = \Pi_1(A \otimes B)\Pi_2$, where Π_1 and Π_2 are permutation matrices we have

$$\|\mathrm{vec}(X)\|_2^2 = \|\Pi_2 \mathrm{vec}(X)\|_2^2,$$

and

$$\|(B \otimes A)\mathrm{vec}(X)\|_2^2 = \|\Pi_1(A \otimes B)\Pi_2 \mathrm{vec}(X)\|_2^2 = \|(A \otimes B)(\Pi_2 \mathrm{vec}(X))\|_2^2.$$

Therefore $\delta_k^{A\otimes B} = \delta_k^{B\otimes A}$. To prove the assertion, it is sufficient to prove that $\delta_k^{A\otimes B} \geq \delta_k^B$, the proof that $\delta_k^{A\otimes B} \geq \delta_k^A$ follows analogously. We know that δ_k^B is the smallest constant such that, for all x with $\|x\|_0 \leq k$, we have

$$(1 - \delta_k^B)\|x\|_2^2 \leq \|Bx\|_2^2 \leq (1 + \delta_k^B)\|x\|_2^2.$$

For any x with $\|x\|_0 \leq k$, we can construct the matrix $X = [\ x \quad 0 \quad \cdots \quad 0\ ]$, with $\|\mathrm{vec}(X)\|_0 \leq k$. Since A has normalized columns, we have

$$\|(A \otimes B)(\mathrm{vec}(X))\|_2^2 = \sum_{i=1}^{p} a_{i,1}^2 \|Bx\|_2^2 = \|Bx\|_2^2, \tag{14}$$

and

$$\|\mathrm{vec}(X)\|_2^2 = \|x\|_2^2. \tag{15}$$

On the other hand $\delta_k^{A\otimes B}$ is the smallest constant such that

$$(1 - \delta_k^{A\otimes B})\|\mathrm{vec}(X)\|_2^2 \leq \|(A \otimes B)(\mathrm{vec}(X))\|_2^2 \leq (1 + \delta_k^{A\otimes B})\|\mathrm{vec}(X)\|_2^2,$$

and for the special class of k-sparse vectors $\mathrm{vec}(X)$ from (14) and (15) we have

$$(1 - \delta_k^{B})\|\mathrm{vec}(X)\|_2^2 \leq \|(A \otimes B)(\mathrm{vec}(X))\|_2^2 \leq (1 + \delta_k^{B})\|\mathrm{vec}(X)\|_2^2,$$

where δ_k^B is the smallest constant for this special class of k-sparse vectors. Therefore, for general k-sparse vectors, we have

$$\delta_k^{A\otimes B} \geq \delta_k^B.$$

$\square$

Remark 3.8 *Note that for $k = 2$, equality holds in (13), since for a given normalized matrix A, we have $\delta_2^A = \mathcal{M}(A)$. Therefore, by Theorem 3.5 it follows that*

$$\delta_2^{A\otimes B} = \max\{\delta_2^A, \delta_2^B\}.$$

For $k \geq 3$, however the inequality may be strict. For example if $A = [I_2\ H_2]$ and $B = [I_4\ H_4]$, where I_n is the identity matrix of order n and H_n is the Hadamard *matrix of order n, see e.g. [17, 18], then $\delta_3^{A\otimes B} = 1.0545 >$* $\max\{1, \frac{1}{\sqrt{2}}\}$*. Here the k-restricted isometry constants of these matrices were calculated using the singular value decomposition for all submatrices consisting of* 3 *columns.*

We have the obvious corollary.

Corollary 3.9 *Suppose that matrices A_i for $i = 1, \ldots, N$ have normalized columns. Then*

$$\delta_k^{A_1\otimes\ldots\otimes A_N} \geq \max_{1\leq i\leq N}\{\delta_k^{A_i}\}.$$

According to Lemma 2.10, if the restricted isometry constant δ_{2k} is small enough ($\delta_{2k} < \sqrt{2}-1$), then one can recover all k-sparse solutions using ℓ_1-minimization. On the other hand, Corollary 3.9 implies that if the k-restricted isometry constant δ_k of A is small (for example less than 1/2), then A can not be written as a Kronecker product of matrices A_i with smaller sizes.

4 Sums of Kronecker products

In many applications, in particular in finite difference or finite element discretizations of partial differential equations in more than one space dimension [23], linear systems with matrices that are sums of Kronecker products arise.

It is then an obvious question whether the spark, the mutual incoherence and the k-restricted isometry property for sums of Kronecker products can be related to that of the summands.

Unfortunately, in general we do not have a nice relation between spark$(A+B)$ and spark(A), spark(B).

Example 4.1 *Let E_n denote the $n \times n$ matrix of all ones. If*

$$A = \begin{bmatrix} I_5 & E_5 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and

$$B = \begin{bmatrix} [1\ 2\ 3\ 4\ 5]^{\mathrm{T}}[1\ 2\ 3\ 4\ 5] & I_5 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix},$$

then $5 = \text{spark}(A+B) > \text{spark}(A) + \text{spark}(B) = 2+2$.

On the other hand if $A = I_2 \otimes I_2$ and $A+B = \frac{1}{2}(E_2 \otimes E_2)$ then $2 = \text{spark}(A+B) < \text{spark}(A) + \text{spark}(B) = 5+5$.

For the mutual incoherence the situation is better. We introduce the following concept of diagonal and off-diagonal mutual incoherence.

Definition 4.2 *Suppose that $A = [a_1, \dots, a_n]$, $B = [b_1, \dots, b_n] \in \mathbb{R}^{m,n}$, $m \le n$, have normalized columns. Then the* off-diagonal mutual incoherence $\mathcal{M}_{OD}(A,B)$ of A and B *is defined via*

$$\mathcal{M}_{OD}(A,B) := \max_{i \neq j} |\langle a_i, b_j \rangle|$$

and the diagonal mutual incoherence $\mathcal{M}_D(A,B)$ of A and B *is defined via*

$$\mathcal{M}_D(A,B) := \max_{i} |\langle a_i, b_i \rangle|.$$

Remark 4.3 *Note that in Definition 4.2 the order of the columns is important. Note further that in the special case that* $A = B$ *we have* $\mathcal{M}_{OD}(A, A) = \mathcal{M}(A)$ *and* $\mathcal{M}_D(A, A) = 1$.

Then we have the following theorem.

Theorem 4.4 *Let* $A = [a_1, \ldots, a_n], B = [b_1, \ldots, b_n] \in \mathbb{R}^{m,n}$ *be matrices with normalized columns and suppose that* $\mathcal{M}_D(A, B) \neq 1$. *Then,*

$$\mathcal{M}(A + B) \leq \frac{\mathcal{M}(A) + 2\mathcal{M}_{OD}(A, B) + \mathcal{M}(B)}{2(1 - \mathcal{M}_D(A, B))}. \tag{16}$$

Proof. For $i \neq j$, by the triangle inequality we have that

$$\begin{aligned} |\langle a_i + b_i, a_j + b_j \rangle| &\leq |\langle a_i, a_j \rangle| + |\langle a_i, b_j \rangle| + |\langle b_i, a_j \rangle| + |\langle b_i, b_j \rangle| \\ &\leq \mathcal{M}(A) + 2\mathcal{M}_{OD}(A, B) + \mathcal{M}(B) \end{aligned} \tag{17}$$

and

$$\|a_i + b_i\|_2^2 = 2 + 2\langle a_i, b_i \rangle \geq 2(1 - |\langle a_i, b_i \rangle|) \geq 2(1 - \mathcal{M}_D(A, B)). \tag{18}$$

Combining (17) and (18), we get

$$\mathcal{M}(A + B) = \max_{i \neq j} \frac{|\langle a_i + b_i, a_j + b_j \rangle|}{\|a_i + b_i\|_2 \|a_j + b_j\|_2} \leq \frac{\mathcal{M}(A) + 2\mathcal{M}_{OD}(A, B) + \mathcal{M}(B)}{2(1 - \mathcal{M}_D(A, B))}.$$

□

Note that the inequality (16) also holds if $\mathcal{M}_D(A, B) = 1$, if we define the right side to be infinite in this case.

Remark 4.5 *The bound in Theorem 4.4 is sharp. For example if*

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = I_2$$

then

$$\mathcal{M}(A + B) = 1, \quad \mathcal{M}(A) = \mathcal{M}(B) = \mathcal{M}_D(A, B) = 0$$

and

$$\mathcal{M}_{OD}(A, B) = 1.$$

Theorem 4.4 immediately extends to more than one summand.

Corollary 4.6 *Consider matrices $A^i \in \mathbb{R}^{m,n}$, $1 \leq i \leq M$ with normalized columns. If*

$$M - 2 \sum_{1 \leq i < j \leq M} \mathcal{M}_D(A^i, A^j) > 0,$$

then

$$\mathcal{M}(\sum_{i=1}^{M} A^i) \leq \frac{\sum_{i=1}^{M} \mathcal{M}(A^i) + 2 \sum_{1 \leq i < j \leq M} \mathcal{M}_{OD}(A^i, A^j)}{M - 2 \sum_{1 \leq i < j \leq M} \mathcal{M}_D(A^i, A^j)}.$$

In order to apply these results to sums of Kronecker products of the form $A = \sum_{j=1}^{M} A_1^j \otimes \ldots \otimes A_N^j$, we introduce the abbreviation

$$\mathcal{U}(\sum_{j=1}^{M} A^j) := \frac{\sum_{j=1}^{M} \mathcal{M}(A^j) + 2 \sum_{1 \leq i < j \leq M} \mathcal{M}_{OD}(A^i, A^j)}{M - 2 \sum_{1 \leq i < j \leq M} \mathcal{M}_D(A^i, A^j)}.$$

We have the following Corollary.

Corollary 4.7 *Consider a linear system of the form*

$$(\sum_{j=1}^{M} A_1^j \otimes \ldots \otimes A_N^j) x = b,$$

where the matrices A_i^j are of appropriate dimensions and have normalized columns. Suppose that there exists a solution x with the sparsity

$$\|x\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mathcal{U}(\sum_{j=1}^{M} A_1^j \otimes \ldots \otimes A_N^j} \right).$$

Then this is the unique solution with this sparsity which can be recovered using ℓ_1-minimization as defined in (4).

Proof. By applying Lemma 2.7 and Corollary 4.6, we have that

$$\|x\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mathcal{U}(\sum_{j=1}^{M} A_1^j \otimes \ldots \otimes A_N^j} \right)$$

implies that

$$\|x\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mathcal{M}(\sum_{j=1}^{M} A_1^j \otimes \ldots \otimes A_N^j} \right)$$

and therefore by Lemma 2.7 the sparse solution is unique. $\square$

Example 4.8 *Consider a sum of Kronecker products* $C = I \otimes A + A \otimes I$ *as they for example arise in the finite difference approximation of boundary value problems for* $2D$ *elliptic PDEs. Then it is easy to see that*

$$\mathcal{M}(C) \leq \mathcal{U}(C) = \frac{2\mathcal{M}(A) + 2\mathcal{M}_{OD}(I \otimes A, A \otimes I)}{2 - 2\mathcal{M}_{OD}(I \otimes A, A \otimes I)}.$$

Especially, if A *is a* $1D$ *finite difference matrix, e.g.*

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix},$$

then we have

$$0.4237 = \mathcal{M}(C) \leq \mathcal{U}(C) = 0.5615.$$

For the k-restricted isometry property it is an open problem to establish relationships between that of a sum of Kronecker products and the summands.

5 Conclusion

We have analyzed the recently introduced concepts of the spark, the mutual incoherence and the k-restricted isometry property of matrix in Kronecker product form to that of the Kronecker factors.

Acknowledgment

We thank J. Gagelman, O. Holtz, M. Pfetsch, C. Van Loan, Z. Xu and H. Yserentant for fruitful discussions on this topic.

References

[1] E. J. Candés, *Compressive sampling*, in Proc. International Congress of Mathematics, Madrid, Spain, 2006, pp. 1433–1452.

[2] E. J. Candès, *The restricted isometry property and its implications for compressed sensing.* Technical Report, California Institute of Technology, 2008.

[3] E. J. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math. **6**, no. 2 (2006), pp. 227–254.

[4] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52**, no. 2 (2006), pp. 489–509.

[5] E. J. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59**, no. 8 (2006), pp. 1207–1223.

[6] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51**, no. 12 (2005), pp. 4203–4215.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decmposition by basis pursuit*, SIAM J. Sci. Comput. **20**, no. 1 (1999), pp. 33–61.

[8] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation.* Preprint, 2006.

[9] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52**, no. 4 (2006), pp. 1289–1306.

[10] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization*, Proc. Natl. Acad. Sci. USA **100**, no. 5 (2003), pp. 2197–2202.

[11] D. L. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52**, no. 1 (2006), pp. 6–18.

[12] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory **47**, no. 7 (2001), pp. 2845–2862.

[13] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit*, Tech. Report 2006-02, Stanford, Department of Statistics, 2006.

[14] M. Elad and A. M. Bruckstein, *A generalized uncertainty principle and sparse representation in pairs of bases*, IEEE Trans. Inform. Theory **48**, no. 9 (2002), pp. 2558–2567.

[15] J. J. Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Trans. Inform. Theory **51**, no. 10 (2005), pp. 1601–1608.

[16] R. Gribonval and M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory **49**, no. 12 (2003), pp. 3320–3325.

[17] J. Hadamard, *Résolution d'une question relative aux déterminants*, Bull. Des Sciences Math. **17**, no. 2 (1893), pp. 240–246.

[18] K. J. Horadam, *Hadamard Matrices and Their Applications*, Princton University Press, Princeton, NJ, 2007.

[19] R. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[20] S. Jokar and M. Pfetsch, *Exact and approximate sparse solutions of underdetermined linear equations.* SIAM J. Sci. Comput., To appear, 2007.

[21] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press, New York, 2nd ed., 1985.

[22] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput. **24**, no. 2 (1995), pp. 227–234.

[23] G. Strang, *Computational Science and Engineering*, Wellesley-Cambridge Press, 2007.

[24] T. Strohmer and R. Heath, *Grassmannian frames with applications to coding and communication*, Appl. Comput. Harmon. Anal. **14**, no. 3 (2003), pp. 257–275.

[25] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50**, no. 10 (2004), pp. 2231–2242.