Mathematical Programming
https://doi.org/10.1007/s10107-020-01590-2

FULL LENGTH PAPER

Series A

Role of sparsity and structure in the optimization landscape of non-convex matrix sensing

Igor Molybog[1] · Somayeh Sojoudi[1] · Javad Lavaei[1]

Received: 4 November 2019 / Accepted: 3 November 2020

Abstract
In this work, we study the optimization landscape of the non-convex matrix sensing problem that is known to have many local minima in the worst case. Since the existing results are related to the notion of restricted isometry property (RIP) that cannot directly capture the underlying structure of a given problem, they can hardly be applied to real-world problems where the amount of data is not exorbitantly high. To address this issue, we develop the notion of kernel structure property to obtain necessary and sufficient conditions for the inexistence of spurious local solutions for any class of matrix sensing problems over a given search space. This notion precisely captures the underlying sparsity and structure of the problem, based on tools in conic optimization. We simplify the conditions for a certain class of problems to show their satisfaction and apply them to data analytics for power systems.

Keywords Non-convex optimization · Spurious local minima · Matrix sensing

1 Introduction

Even under the ideal condition of no noise and zero approximation error, many highly-efficient machine learning techniques involve solving potentially hard or intractable computational problems while learning from data. In practice, they are tackled by heuristic optimization algorithms, based on relaxations or greedy principles. The lack of guarantees on their performance limits their use in applications with significant cost of an error, impacting our ability to implement progressive data analysis techniques

✉ Javad Lavaei
lavaei@berkeley.edu

Igor Molybog
igormolybog@berkeley.edu

Somayeh Sojoudi
sojoudi@berkeley.edu

[1] University of California, Berkeley, CA, USA

Published online: 20 November 2020

in crucial social and economic systems, such as healthcare, transportation, and energy production and distribution. Commonly, non-convexity is the main obstacle for a guaranteed learning of continuous parameters.

It is well known that many fundamental problems with a natural non-convex formulation are $\mathscr{NP}$-hard [27]. Sophisticated techniques for addressing this issue, like generic convex relaxations, may require working in an unrealistically high-dimensional space to guarantee the exactness of the solution. As a consequence of complicated geometrical structures, a non-convex function may contain an exponential number of saddle points and spurious local minima, and therefore local search algorithms may become trapped in any of those points. Nevertheless, empirical observations show positive results regarding the application of these approaches to several practically important instances. This provokes a major branch of research that aims to explain the success of experimental results in order to understand the boundaries of the applicability of the existing algorithms and develop new ones. A recent direction in non-convex optimization consists in studying how simple algorithms can solve potentially hard problems arising in data analysis applications. The most commonly applied class of such algorithms is based on *local search*, which will be the focus of this work. In some cases, prior information about the location of the solution is available, which significantly reduces the complexity of the search.

Consider searching over some given domain $\mathscr{X}$. For a twice continuously differentiable objective function $f : \mathscr{X} \to \mathbb{R}$ that reaches its global minimum f^*, if the point x attains $f(x) = f^*$, then we call it a *global minimizer*. The point x is said to be a *local minimizer* if $f(x) \leq f(x')$ holds for all x' within a local neighborhood of x. If x is a local minimizer, then it must satisfy the first- and second-order *necessary* optimality conditions. Conversely, a point x satisfying only the first-order condition is called a *first-order critical point,* while a point satisfying both of the conditions is called a *second-order critical point* or simply a *solution*. We call a solution *spurious* if it is not a global minimum. In this work, we study how the existence of a spurious solution depends on the size of the domain as well as the underlying structure of the problem.

The analysis of the landscape of the objective function around a global optimum may lead to an optimality guarantee for local search algorithms initialized sufficiently close to the solution [16,18,19,33,41,42]. Finding a good initialization scheme is highly problem-specific and difficult to generalize. Global analysis of the landscape is harder, but potentially more rewarding.

Both local and global convergence guarantees have been developed to justify the success of local search methods in various applications, such as dictionary learning [1], basic non-convex M-estimators [25], shallow [30] and deep [37] artificial neural networks with different activation [22] and loss [26] functions, phase retrieval [4,5,36] and more general matrix sensing problems [10,17]. Particularly, significant progress has been made towards understanding different variants of the *low-rank matrix recovery,* although explanations of the simplest version called *matrix sensing* are still under active development [6,7,10,24,39,43]. Given a linear sensing operator $\mathscr{A} : \mathbb{S}^n \to \mathbb{R}^m$ and a ground truth matrix $z \in \mathbb{R}^{n \times r}$ $(r < n)$, an instance of the rank-r matrix sensing

problem consists in minimizing over $\mathbb{R}^{n\times r}$ the nonconvex function

$$f_{z,\mathscr{A}}(x) = \left\|\mathscr{A}\left(xx^T - zz^T\right)\right\|_2^2 = \left\|\mathscr{A}\left(xx^T\right) - b\right\|_2^2, \tag{1}$$

where $b = \mathscr{A}(zz^T)$. We consider this function over a general set $\mathscr{X} \subseteq \mathbb{R}^{n\times r}$, although in this section we set $\mathscr{X} = \mathbb{R}^{n\times r}$. Recent work has generally found a certain assumption on the sensing operator to be sufficient for the matrix sensing problem to be "computationally easy to solve". Precisely, this assumption works with the notion of restricted isometry property (RIP).

Definition 1 *(Restricted isometry property)* The linear map $\mathscr{A} : \mathbb{S}^n \to \mathbb{R}^m$ is said to satisfy δ_r-RIP for some constant $\delta_r \in [0, 1)$ if there is $\gamma > 0$ such that

$$(1 - \delta_r)\|X\|_F^2 \leq \gamma\|\mathscr{A}(X)\|_2^2 \leq (1 + \delta_r)\|X\|_F^2$$

holds for all $X \in \mathbb{S}^n$ satisfying $\operatorname{rank}(X) \leq r$.

The existing results proving absence of spurious local minima using this notion (such as [2,10–12,28,31,32,43]) are based on a norm-preserving argument: the problem turns out to be a low-dimensional embedding of a canonical problem known to contain no spurious local minima. While the approach is widely applicable in its scope, it requires fairly strong assumptions on the data. In contrast, Zhang et al. [38,39] introduced a technique to find a certificate to guarantee that any given point cannot be a spurious local minimum of the problem of minimizing $f_{z,\mathscr{A}}$ over the set $\mathscr{X} \subseteq \mathbb{R}^{n\times r}$, where $z \in \mathscr{Z} \subseteq \mathbb{R}^{n\times r}$ and $\mathscr{A}$ satisfies δ_{2r}-RIP. Note that two different sets $\mathscr{X}$ and $\mathscr{Z}$ are involved here. Since $f_{z,\mathscr{A}}$ depends on z and $\mathscr{A}$, this introduces a class of optimization problems defined as

$$\left\{\underset{x\in\mathscr{X}}{\text{minimize}}\ f_{z,\mathscr{A}}(x) \;\middle|\; \mathscr{A} \text{ satisfies } \delta_{2r}\text{-RIP}, z \in \mathscr{Z}\right\}. \tag{Problem$^{\text{RIP}}$}$$

(Problem$^{\text{RIP}}$) consists of infinitely many instances of an optimization problem, each corresponding to some point z in $\mathscr{Z}$ and some operator $\mathscr{A}$ satisfying δ_{2r}-RIP. The state-of-the-art results for (Problem$^{\text{RIP}}$) are stated below.

Theorem 1 ([2,10,38]) *By taking $\mathscr{X} = \mathscr{Z} = \mathbb{R}^{n\times r}$, the following statements hold:*

- *If $\delta_{2r} < 1/5$, no instance of (Problem$^{\text{RIP}}$) has a spurious second-order critical point.*
- *If $r = 1$ and $\delta_2 < 1/2$, then no instance of (Problem$^{\text{RIP}}$) has a spurious second-order critical point.*
- *If $r = 1$ and $\delta_2 \geq 1/2$, then there exists an instance of (Problem$^{\text{RIP}}$) with a spurious second-order critical point.*

Non-existence of a spurious second-order critical point effectively means that any algorithm that converges to a second-order critical point is guaranteed to recover zz^T exactly. Examples of such algorithms include variants of the stochastic gradient descent

(SGD) that are known to avoid saddle or even spurious local minimum points under certain assumptions [8], and widely used in machine learning [3,20]. Besides SGD, many local search methods have been shown to be convergent to second-order critical points with high probability under mild conditions, including the classical gradient descent [21], alternating minimizations [23] and Newton's method [29]. In this paper, we present guarantees on the global optimality of the second-order critical points, which means that our results can be combined with any of the algorithms mentioned above to guarantee the global convergence.

Theorem 1 discloses the limits on the guarantees that the notion of RIP can provide. However, linear maps in applications related to physical systems, such as power system analysis, typically have no RIP constant smaller than 0.9, and yet the non-convex matrix sensing still manages to work on those instances. The gap between theory and practice motivates the following question.

What is the alternative property practical problems satisfy that makes them easy to solve via simple local search?

This question was studied earlier for special cases of matrix sensing, namely phase retrieval [31] and matrix completion [11]. In case of the phase retrieval problem, the alternative property consists in the particular distribution of the measurements operator. Regarding the matrix completion problem, the assumption includes conditions on the properties of the matrix being recovered along with conditions on the measurement operator itself. We address the above mentioned question by developing a new notion that deals with the measurement operator and precisely captures when a structured matrix recovery problem has no spurious solution over an arbitrary ball. We focus the analysis over a given ball since local search methods tend to search over a neighborhood rather than the entire space, based on prior knowledge. In Sect. 2, we motivate the need for a new notion replacing or improving RIP with real-world examples. Section 3 introduces some formal definitions and develops a mathematical framework to analyze spurious solutions and relate them to the underlying sparsity and structure of the problem, using techniques in conic optimization. Sections 4 and 5 give the theory behind this notion and examples of its application. In Sect. 6, we present numerical results on the application of the developed theory in a real-world problem appearing in power systems analysis. Concluding remarks are given in Sect. 7. Some of the proofs, technical details and lemmas are collected in the appendix.

1.1 Notation

$\mathbb{C}^n$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times r}$ denote the sets of complex and real n-dimensional vectors, and $n \times r$ matrices, respectively. $\mathbb{S}^n$ denotes the set of $n \times n$ symmetric matrices. $\mathrm{Tr}(A)$, $\|A\|_F$, $\lambda_{min}(A)$ and $\langle A, B\rangle$ are the trace of a square matrix A, its Frobenius norm, absolute smallest eigenvalue and the Frobenius inner product of matrices A and B of compatible sizes. The normal distribution with mean μ and covariance matrix Σ is denoted as $\mathcal{N}(\mu, \Sigma)$. In any linear space, $\mathbf{1}$ is a vector whose entries are all equal to 1 and I is the identity matrix. For $\omega \in \mathbb{R}^{n\times r}$ and $R \in \mathbb{R} \cup \{+\infty\}$, we define $\mathbb{B}_R(\omega) = \{a \in \mathbb{R}^{n\times r} : \|a - \omega\|_F \leq R\}$, $\bar{\mathbb{B}}_R(\omega) = \{a \in \mathbb{R}^{n\times r} : \|a - \omega\|_F < R\}$ and $\partial\mathbb{B}_R(\omega) = \{a \in \mathbb{R}^{n\times r} : \|a - \omega\|_F = R\}$. It follows from the definition

that $\partial\mathbb{B}_{+\infty}(\omega) = \emptyset$ and minimization over this set results in $+\infty$ for any objective function. For a square matrix A, we define the symetric part $\mathrm{Sym}(A) = (A + A^T)/2$. For a symmetric matrix A, its null space is denoted with $Ker(A)$. For square matrices $A_1, A_2, \ldots, A_n$, the matrix $diag(A_1, \ldots, A_n)$ is block-diagonal, with A_i's on the block diagonal. The notation $A \circ B$ refers to the Hadamard (entrywise) multiplication, and $A \otimes B$ refers to the Kronecker product of matrices. The vectorization operator $\mathrm{vec} : \mathbb{R}^{n\times r} \to \mathbb{R}^{nr}$ stacks the columns of a matrix into a vector. The *matricization* operator $\mathrm{mat}(\cdot)$ is the inverse of $\mathrm{vec}\,(\cdot)$. Let $\succeq$ denote the positive semidefinite sign.

For a linear operator $\mathscr{L} : \mathbb{R}^{n\times r} \to \mathbb{R}^m$, the adjoint operator is denoted by $\mathscr{L}^T : \mathbb{R}^m \to \mathbb{R}^{n\times r}$. The matrix $\mathbf{L} \in \mathbb{R}^{m\times nr}$ such that $\mathscr{L}(x) = \mathbf{L}\mathrm{vec}\,(x)$ is called the *matrix representation* of the linear operator $\mathscr{L}$. Bold letters are reserved for matrix representations of corresponding linear operators.

Sparsity pattern S of a set of matrices $\mathrm{M} \subset \mathbb{R}^{m\times n}$ is a subset of $\{1, \ldots, \max\{n, m\}\}^2$ such that $(i, j) \in S$ if and only if there is $X \in \mathrm{M}$ with the property that $X_{ij} \neq 0$. Given a sparsity pattern S, define its matrix representation $\mathbf{S} \in \mathbb{S}^{m\times n}$ as

$$\mathbf{S}_{ij} = \begin{cases} 0 & \text{if } (i, j) \in S, \\ 1 & \text{if } (i, j) \notin S, \end{cases}$$

The *orthogonal basis* of a given $m \times n$ matrix A (with $m \geq n$) is a matrix $P = \mathrm{orth}(A) \in \mathbb{R}^{m\times rank(A)}$ consisting of rank (A) orthonormal columns that span $\mathrm{range}(A)$:

$$P = \mathrm{orth}(A) \quad \Longleftrightarrow \quad PP^TA = A,\ P^TP = I_{\mathrm{rank}\,(A)}.$$

Positive part means $(\cdot)_+ = \max\{0, \cdot\}$, and eigenvalues in an arbitrary order are denoted by $\lambda_i(\cdot)$.

2 Motivating example

In this section, we motivate this work by offering a case study on data analytics for energy systems. The state of a power system can be modeled by a vector of complex voltages on the nodes (buses) of the network. Monitoring the state of a power system is obviously a necessary requirement for its efficient and safe operation. This crucial information should be inferred from some measurable parameters, such as the power that is generated and consumed at each bus or transmitted through a line. The power network can be modeled by a number of parameters grouped into the admittance matrix $Y \in \mathbb{C}^{n\times n}$. The state estimation problem consists in recovering the unknown voltage vector $v \in \mathbb{C}^n$ from the available measurements. In the noiseless scenario, these measurements are m real numbers of the form

$$v^* M_i v, \quad \forall\, i \in \{1, \ldots, m\}, \tag{2}$$

where $M_i \in \mathbb{C}^{n\times n}$ are sparse Hermitian matrices that are obtained from Y and model power flow, power injection, and voltage magnitude measurements. The sparsity pattern of the measurement matrices is determined by the topology of the network, while

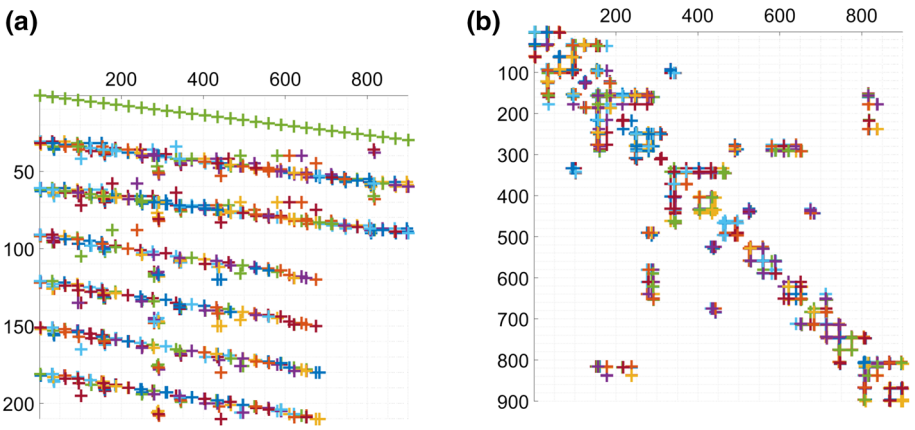

Fig. 1 Examples of the structure patterns of operators $\mathscr{A}$ (left plot) and $\mathscr{H}$ (right plot) in power system applications. The positions of the identical nonzero entries of a matrix are marked with the same markers

its nonzero entries are certain known functions of the entries of Y. Since the total number of nonzero elements in the matrices M_i exceeds the total number of parameters contained in Y, one can regard the mapping $Y \to \{M_i\}_{i=1}^m$ as an embedding from a low-dimensional space. For a detailed discussion on the problem formulation and approaches to its solution, please refer to e.g. Zhang et al. [40].

To formulate the problem as a low-rank matrix recovery, we introduce a sparse matrix $\mathbf{A} \in \mathbb{C}^{m \times n^2}$ with its i-th row equal to $\text{vec}(M_i)^T$. The measurement vector can be written as $\mathbf{A}\text{vec}(vv^T)$. To find v from the measurements, one may solve the non-convex optimization problem:

$$\underset{x \in \mathbb{C}^n, \|x-\omega\|_F \le R}{\text{minimize}} \left\| \mathbf{A}\text{vec}\,(xx^T - vv^T) \right\|^2 \tag{3}$$

where $\omega \in \mathbb{C}^n$ and $R \in \mathbb{R} \cup \{+\infty\}$ are some parameters determined by the prior knowledge about the solution v. In practice, this non-convex optimization problem is usually solved via local search methods, which converges to a second-order critical point at best. Since $f(x) = \|\mathbf{A}\text{vec}\,(xx^T - vv^T)\|_F^2 = \langle xx^T - vv^T, \mathbf{A}^T\mathbf{A}\text{vec}\,(xx^T - vv^T)\rangle$, the set of critical points of the problem is defined by the linear map represented with the matrix $\mathbf{H} = \mathbf{A}^T\mathbf{A}$, which thus is the key subject of the study. Problems arising in power systems analysis are based on operators that possess a specific structure. An example of a structure for the matrix $\mathbf{A}$ is given in Fig. 1a, and the structure of the corresponding $\mathbf{H}$ is described in Fig.1b. The respective power network will be considered in more details in Sect. 6. As discussed previously, given $\mathbf{H}$, it is practically important to know if there exist $v, x \in \mathbb{C}^n$ such that x is a critical point of (3) while $xx^T \neq vv^T$. Absence of these points proves that a local search method recovers v exactly, certifying safety of its use. For example, in case of unconstrained optimization ($R = +\infty$), the answer is affirmative if the optimal objective value of the following

problem is equal to zero:

$$\begin{aligned} \underset{v,\ x\in\mathbb{C}^n}{\text{minimize}} \ & \left\|\mathscr{A}(xx^T - vv^T)\right\|^2 \\ \text{subject to } & \nabla f(x) = 0 \\ & \nabla^2 f(x) \succeq 0 \end{aligned}$$

where ∇ and ∇^2 denote the gradient and Hessian operators. However, this is an $\mathscr{NP}$-hard problem in general and cannot be solved efficiently. Even if we were able to solve it, the sensing operator $\mathscr{A}$ could change over time without changing its structure, and therefore any conclusion made for a specific problem cannot be generalized to other instances of the problem for real-world applications where the data analysis is to be performed periodically. One way to circumvent this issue is to develop a sufficient condition for all mappings $\mathbf{H}$ with the same structure.

3 Introducing Kernel structure

Consider a linear operator $\mathscr{A} : \mathbb{S}^n \to \mathbb{R}^m$ with the matrix representation $\mathbf{A} \in \mathbb{R}^{m\times n^2}$ and a sparsity pattern $S_{\mathscr{A}}$. Assume that there is a set of hidden parameters $\xi \in \mathbb{R}^d$ with $d \ll m$ such that $\mathbf{A}$ is the image of ω in the space of a much higher dimension. In this way, $\mathscr{A}$ has a low-dimensional structure beyond sparsity, which is captured by $\mathbf{A} = \mathbf{A}(\xi)$ and $\mathbf{A}(0) = \mathbf{0}$. The motivating complex-valued example in Sect. 2 is a special case of this construction since it could be stated entirely with only real-valued vectors and matrices of a bigger size. We define the nonconvex objective

$$f : \mathbb{R}^{n\times r} \to \mathbb{R} \qquad \text{such that} \qquad f(x) = \left\|\mathscr{A}\left(xx^T - zz^T\right)\right\|^2$$

parametrized by $\mathscr{A}$ and $z \in \mathbb{R}^{n\times r}$. Its value is always nonnegative by construction, and the global minimum 0 is attainable. To emphasize the dependence on certain parameters, we will write them in the subscript. To align the minimization problem with the problem of reconstructing zz^T, we need to introduce a regularity assumption:

Assumption 1 For all $x, z \in \mathbb{R}^{n\times r}$:

$$\left\|\mathscr{A}\left(xx^T - zz^T\right)\right\| = 0 \text{ if and only if } xx^T = zz^T$$

We will rely on Assumption 1 throughout the paper. Note that this assumption is weaker than the assumption of existence of a $2r$-RIP constant, although it is stronger than the assumption of existence of a r-RIP constant. Another way to express the objective is

$$f(x) = \left\langle xx^T - zz^T, \mathscr{H}\left(xx^T - zz^T\right)\right\rangle.$$

Here, $\mathscr{H} = \mathscr{A}^T\mathscr{A}$ is the linear *kernel* operator that has the matrix representation $\mathbf{H} = \mathbf{A}^T\mathbf{A}$ and sparsity pattern $S_{\mathscr{H}}$. Namely, $(i, j) \in S_{\mathscr{H}}$ if and only if there exists k such that $(k, i) \in S_{\mathscr{A}}$ and $(k, j) \in S_{\mathscr{A}}$. Sparsity of $\mathscr{H}$ is controlled by the out-degree of the graph represented by $S_{\mathscr{A}}$, and tends to be low in applications like power systems. $S_{\mathscr{H}}$ is represented by a matrix $\mathbf{S}$, so that $S_{\mathscr{H}}$-sparse operators are exclusively those satisfying the linear equation $\mathscr{S}(\mathbf{H}) = \mathbf{S} \circ \mathbf{H} = \mathbf{0}$. Besides sparsity, $\mathscr{H}$ inherits the low-dimensional structure from $\mathscr{A}$, which can be captured by $\mathbf{H} = \mathbf{A}(\xi)^T\mathbf{A}(\xi) = \mathbf{H}(\xi)$ where $\xi \in \mathbb{R}^d$. This dependence can be locally approximated in the hidden parameter space with a linear one. More precisely, suppose that there is a linear operator $\mathscr{W}$ defined over $\mathbb{S}^{n^2}$ such that $\mathscr{W}(\mathbf{H}(\xi)) \approx 0$ for the values of ξ under consideration. Thus, from now on we focus exclusively on low-dimensional structures of the form $\mathscr{W}(\mathbf{H}) = 0$. Together, the sparsity operator $\mathscr{S}$ and the low-dimensional structure operator $\mathscr{W}$ form the combined structure operator $\mathscr{T} = (\mathscr{S}, \mathscr{W})$ that accumulates the structure of the kernel operator.

Definition 2 *(Kernel structure property or KSP)* Given a linear structure operator $\mathscr{T} : \mathbb{S}^{n^2} \to \mathbb{R}^t$, the linear map $\mathscr{A} : \mathbb{S}^n \to \mathbb{R}^m$ is said to satisfy $\mathscr{T}$-KSP if it satisfies Assumption 1 and

$$\mathscr{T}(\mathbf{A}^T\mathbf{A}) = 0$$

where $\mathbf{A}$ is the matrix representation of $\mathscr{A}$.

Notice that a particular sensing operator $\mathscr{A}$ can be kernel structured with respect to an entire family of structure operators, and we can possibly select any of them for our benefit in the following section.

4 Analysis based on KSP

Given a kernel structure $\mathscr{T}$, $\omega \in \mathbb{R}^{n\times r}$ and $R \in \mathbb{R} \cup \{+\infty\}$, we can state the problem under study as follows:

$$\left\{ \underset{x\in\mathbb{B}_R(\omega)}{\text{minimize}}\ f_{z,\mathscr{A}}(x) \;\middle|\; \mathscr{A} \text{ satisfies Assumption 1 and } \mathscr{T}\text{-KSP}, z \in \mathbb{B}_R(\omega) \right\}, \tag{Problem$^{\text{KSP}}$}$$

Note that (Problem$^{\text{KSP}}$) consists of infinitely many instances of an optimization problem, each corresponding to some point $z \in \mathbb{B}_R(\omega)$ and some operator $\mathscr{A}$ satisfying $\mathscr{T}$-KSP.

If X is regarded as an input and the operator $\mathscr{A}$ is regarded as a system with its output being $\mathscr{A}(X)$, the RIP constant aims at characterizing the input-output behavior of the system. This input-output relationship can also be controlled by imposing the following constraint on the matrix $\mathscr{H}$:

$$(1 - \delta)\mathscr{I} \preceq \mathscr{H} \preceq (1 + \delta)\mathscr{I},$$

where $\mathscr{I}$ is the identity operator. More precisely, the above inequality guarantees that the operator $\mathscr{A}$ will have an RIP constant less than or equal to δ. Inspired by this observation, we define the function $\mathbb{O}(x, z; \mathscr{T})$ to be the optimal objective value of the convex optimization problem:

$$\begin{aligned} \underset{\delta\in\mathbb{R},\mathscr{H}}{\text{minimum}} \quad & \delta \\ \text{subject to} \quad & \mathscr{L}_{x,z}(\mathscr{H}) = 0 && \text{(4a)} \\ & \mathscr{M}_{x,z}(\mathscr{H}) \succeq 0 && \text{(4b)} \\ & \mathscr{T}(\mathscr{H}) = 0 && \text{(4c)} \\ & (1-\delta)\mathscr{I} \preceq \mathscr{H} \preceq (1+\delta)\mathscr{I} && \text{(4d)} \end{aligned}$$

where $\mathscr{L}_{x,z}(\mathscr{H}) = \nabla f_{z,\mathscr{H}}(x)$ and $\mathscr{M}_{x,z}(\mathscr{H}) = \nabla^2 f_{z,\mathscr{H}}(x)$. This optimization is performed over all operators $\mathscr{H}$ satisfying the $\mathscr{T}$-KSP. We will later show that the function $\mathbb{O}$ sets an upper bound on the δ_{2r} such that none of the functions $f_{z;\mathscr{A}}$ with $\mathscr{A}$ satisfying $\mathscr{T}$-KSP and δ_{2r}-RIP has a spurious second-order critical point at x, provided that x is not on the boundary of the optimization domain $\mathbb{B}_R(\omega)$.

For completeness and for further reference within this paper, we calculate the analytic forms of the first- and second-order derivatives of $f_{z,\mathscr{H}}$ below. Introduce a vector $\mathbf{e}$ and a matrix $\mathbf{X}$ such that for all $u \in \mathbb{R}^{n\times r}$ it holds that

$$\mathbf{e} = \text{vec}\,(xx^T - zz^T), \qquad \mathbf{X}\text{vec}\,(u) = \text{vec}\,(xu^T + ux^T).$$

We write the operators $\mathscr{L}$, $\mathscr{M}$ and their transpose operators in closed form:

$$\begin{aligned} &\mathscr{L}_{x,z} : \mathbb{S}^{n^2} \to \mathbb{R}^{n\times r} && \mathscr{L}_{x,z}(\mathbf{H}) = 2\cdot \mathbf{X}^T\mathbf{H}\mathbf{e}, \\ &\mathscr{L}^T_{x,z} : \mathbb{R}^{n\times r} \to \mathbb{S}^{n^2} && \mathscr{L}^T_{x,z}(y) = \mathbf{e}y^T\mathbf{X}^T + \mathbf{X}y\mathbf{e}^T, \\ &\mathscr{M}_{x,z} : \mathbb{S}^{n^2} \to \mathbb{S}^{nr} && \mathscr{M}_{x,z}(\mathbf{H}) = [I_r \otimes \big(\text{mat}(\mathbf{H}\mathbf{e}) + \text{mat}(\mathbf{H}\mathbf{e})^T\big)] + \mathbf{X}^T\mathbf{H}\mathbf{X}, \\ &\mathscr{M}^T_{x,z} : \mathbb{S}^{nr} \to \mathbb{S}^{n^2} && \mathscr{M}^T_{x,z}(V) = \text{vec}\,(V)\mathbf{e}^T + \mathbf{e}\text{vec}\,(V)^T + \mathbf{X}V\mathbf{X}^T. \end{aligned}$$

Since $f_{z,\mathscr{H}}(x)$ is linear in $\mathscr{H}$, the operators $\mathscr{L}_{x,z}$ and $\mathscr{M}_{x,z}$ are both linear operators. Thus, the problem defining the function $\mathbb{O}$ is convex.

The function $\mathbb{O}$ is useful only for those points x that are located strictly inside the domain $\mathbb{B}_R(\omega)$. This is due to the fact the constraints (4a) and (4b) are meant to be optimality conditions for a point inside the domain $\mathbb{B}_R(\omega)$. For a point x that lies on the boundary, we define the corresponding function $\mathbb{O}^{\partial\mathbb{B}}(x, z; \mathscr{T}, \omega)$ as the optimal objective value of the convex optimization problem:

$$\begin{aligned} \underset{\delta,\mu\geq 0,\mathscr{H}}{\text{minimum}} \quad & \delta \\ \text{subject to} \quad & \mathscr{L}_{x,z}(\mathscr{H}) = -\mu(x-\omega) && \text{(5a)} \\ & P_{x-\omega}\mathscr{M}_{x,z}(\mathscr{H})P^T_{x-\omega} \succeq 0 && \text{(5b)} \\ & \mathscr{T}(\mathscr{H}) = 0 && \text{(5c)} \\ & (1-\delta)\mathscr{I} \preceq \mathscr{H} \preceq (1+\delta)\mathscr{I} && \text{(5d)} \end{aligned}$$

where $P_{x-\omega} \in \mathbb{R}^{(nr-1)\times nr}$ is the matrix of orthogonal projection onto the subspace orthogonal to $x-\omega$. The role of the function $\mathbb{O}^{\partial\mathbb{B}}$ is the same as of $\mathbb{O}$ but is applicable to only those values of x such that $\|x-\omega\| = R$. Note that (5a) and (5b) are the necessary optimal conditions for a solution on the boundary of $\mathbb{B}_R(\omega)$.

To relax the δ_{2r}-RIP condition, we consider those operators that have a bounded effect on a linear subspace of limited-rank inputs. Indeed, for any $2r$ linearly independent vectors, the linear span of them is a linear subspace of the manifold of the $2r$-rank matrices. Thus, for any linear operator $\mathscr{P}$ from a $2r$-dimensional (or lower) vector space to $\mathbb{R}^{n^2}$, the following condition on $\mathscr{H}$ holds if $\mathscr{A}$ satisfies δ-RIP:

$$(1-\delta)\mathscr{P}^T\mathscr{P} \preceq \mathscr{P}^T\mathscr{H}\mathscr{P} \preceq (1+\delta)\mathscr{P}^T\mathscr{P}. \tag{6}$$

Based on this observation, we define the function $\mathbb{O}_P(x, z; \mathscr{T})$ as the optimal objective value of the following convex optimization problem:

$$\begin{aligned}
\underset{\delta\in\mathbb{R},\mathscr{H}}{\text{minimum}} \quad & \delta \\
\text{subject to} \quad & \mathscr{L}_{x,z}(\mathscr{H}) = 0 && \text{(7a)}\\
& \mathscr{M}_{x,z}(\mathscr{H}) \succeq 0 && \text{(7b)}\\
& \mathscr{T}(\mathscr{H}) = 0 && \text{(7c)}\\
& (1-\delta)\mathscr{P}^T\mathscr{P} \preceq \mathscr{P}^T\mathscr{H}\mathscr{P} \preceq (1+\delta)\mathscr{P}^T\mathscr{P} && \text{(7d)}
\end{aligned}$$

where $\mathscr{P}$ is the linear operator from $\mathbb{R}^{rank([x\ z])^2}$ to $\mathbb{R}^{n^2}$ that is represented by the matrix $\mathbf{P} = \text{orth}([x\ z]) \otimes \text{orth}([x\ z])$. Note that (7) is obtained from (4) by replacing its constraint (4d) with the milder condition (6). We will show that the function $\mathbb{O}_P$ sets a lower bound on the δ_{2r} such that none of the functions $f_{z;\mathscr{A}}$ with $\mathscr{A}$ satisfying $\mathscr{T}$-KSP and δ_{2r}-RIP has a spurious second-order critical point at x, provided that x is not on the boundary of the optimization domain $\mathbb{B}_R(\omega)$.

Similarly to $\mathbb{O}^{\partial\mathbb{B}}$, the function $\mathbb{O}_P^{\partial\mathbb{B}}(x, z; \mathscr{T}, \omega)$ is defined as the optimal objective value of the convex optimization problem:

$$\begin{aligned}
\underset{\delta,\mu\geq 0,\mathscr{H}}{\text{minimum}} \quad & \delta \\
\text{subject to} \quad & \mathscr{L}_{x,z}(\mathscr{H}) = -\mu(x-\omega) \\
& P_{x-\omega}\mathscr{M}_{x,z}(\mathscr{H})P_{x-\omega}^T \succeq 0 \\
& \mathscr{T}(\mathscr{H}) = 0 \\
& (1-\delta)\mathscr{P}^T\mathscr{P} \preceq \mathscr{P}^T\mathscr{H}\mathscr{P} \preceq (1+\delta)\mathscr{P}^T\mathscr{P}
\end{aligned}$$

which is designed to lower bound the constant δ_{2r} such that none of the functions $f_{z;\mathscr{A}}$ with $\mathscr{A}$ satisfying $\mathscr{T}$-KSP and δ_{2r}-RIP has a spurious second-order critical point at x, provided that x is on the boundary of $\mathbb{B}_R(\omega)$.

Now, we are ready to state one of the main results of this paper.

Theorem 2 (KSP necessary and sufficient conditions) *For all instances of* (Problem^{KSP}), *there are no spurious second-order critical points if*

$$\begin{cases} \mathbb{O}_P(x, z; \mathscr{T}) \equiv 1 \text{ over } \mathbb{B}_R(\omega) \times \mathbb{B}_R(\omega) \setminus \{xx^T = zz^T\} \\ \mathbb{O}_P^{\partial\mathbb{B}}(x, z; \mathscr{T}, \omega) \equiv 1 \text{ over } \partial\mathbb{B}_R(\omega) \times \mathbb{B}_R(\omega) \setminus \{xx^T = zz^T\} \end{cases} \tag{8}$$

and only if

$$\begin{cases} \mathbb{O}(x, z; \mathscr{T}) \equiv 1 \text{ over } \mathbb{B}_R(\omega) \times \mathbb{B}_R(\omega) \setminus \{xx^T = zz^T\} \\ \mathbb{O}^{\partial\mathbb{B}}(x, z; \mathscr{T}, \omega) \equiv 1 \text{ over } \partial\mathbb{B}_R(\omega) \times \mathbb{B}_R(\omega) \setminus \{xx^T = zz^T\} \end{cases} \tag{9}$$

This theorem is formally proven in the appendix. To elaborate on implications and practicality of the result, we present its application for a specific structure of the sensing operator below.

4.1 Ellipsoid norm: rank 1

In this subsection, we prove a special case of Theorem 2 for the ellipsoid norm objective function and $R = +\infty$. This proof first provides useful intuition behind the proof of the general case and then simplifies the conditions of Theorem 2 to show that they always hold for a specific class of operators.

Consider the ellipsoid norm of $xx^T - zz^T$ given by a full-rank matrix $Q \in \mathbb{R}^{n\times n}$, denoted with $h(x)$:

$$h(x) = \|Q(xx^T - zz^T)\|_F^2 = f_{z,\mathbf{A}}(x)$$

With no loss of generality, assume that $Q \in \mathbb{S}^n$ since $h(\cdot)$ really depends only on Q^TQ. The function can be implemented with a block-diagonal sensing operator matrix $\mathbf{A} = diag(Q, \ldots, Q) \in \mathbb{S}^{n^2}$, which generates a block-diagonal kernel matrix $\mathbf{H} = diag(QQ, \ldots, QQ)$. Thus, the kernel matrix is a block-diagonal matrix $\mathbf{H} = diag(H_{11}, \ldots, H_{nn}) \in \mathbb{S}^{n^2}$ with blocks of size $n \times n$ equal to each other; in other words, $H_{ii} = H_{jj}$ for all $i, j \in \{1, \ldots, n\}$. This generates a kernel structure. By applying the theory introduced above, we obtain the following result for the rank-one case.

Proposition 1 *Consider a kernel structure operator* $\mathscr{T} = (\mathscr{S}, \mathscr{W})$ *such that*

- $\mathscr{S}(\mathbf{H}) = \mathbf{0}$ *iff* $\mathbf{H} = diag(H_{11}, \ldots, H_{nn})$
- $\mathscr{W}(\mathbf{H}) = \mathbf{0}$ *iff* $H_{ii} = H_{jj}$ *for all* $i, j \in \{1, \ldots, n\}$,

Then, no instance of the (Problem^{KSP}) *with* $R = \infty$ *has a spurious second-order critical point over* $\mathbb{R}^n$.

The proposition implies that the function $h(x)$ can never have a spurious solution for rank-1 arguments. To prove this result, first notice that, Assumption 1 and the following lemma combined imply that H_{ii}, and, consequently, its decomposition $H_{ii} = QQ$, are full-rank matrices.

Lemma 1 *Given a constant* $\delta_r \in [0, 1)$, *the matrix* $Q \in \mathbb{S}^n$ *satisfies*

$$(1 - \delta_r)\|X\|_F^2 \leq \|QX\|_F^2 \leq (1 + \delta_r)\|X\|_F^2$$

for every X *such that* $\text{rank}\,(X) \leq r$ *only if* $\text{rank}\,(Q) = n$

Proof By contradiction, suppose that $u \in \text{Ker}(Q)$ and $u \neq 0$. Take $X = uu^T$ and observe that $QX = 0$, which contradicts that $(1 - \delta_r)\|X\|_F^2 \leq \|QX\|_F^2$. □

The following lemma provides a version the conditions (8) and (9) combined for this particular structure operator.

Lemma 2 *Given* $z \in \mathbb{R}^{n\times r}$, *a point* $x \in \mathbb{R}^{n\times r}$ *is not a first-order critical point of the function* $h(\cdot)$ *for an arbitrary full-rank matrix* Q *if and only if there is* $\lambda \in \mathbb{R}^{n\times r}$ *such that*

$$0 \neq \text{Sym}\left[(x\lambda^T + \lambda x^T)(xx^T - zz^T)\right] \succeq 0$$

Proof By expanding $h(x + u)$ as

$$\begin{aligned} h(x+u) = h(x) &+ \text{Tr}\left(2x^T\left(\left(xx^T - zz^T\right)M + M\left(xx^T - zz^T\right)\right)u\right) \\ &+ \text{Tr}\left(u^T\left(\left(xx^T - zz^T\right)M + M\left(xx^T - zz^T\right)\right)u\right. \\ &\left. + \left(xu^T + ux^T\right)M\left(xu^T + ux^T\right)\right) + o(|u|^2) \end{aligned}$$

one can arrive at a more specified expression for the second-order necessary conditions for local optimality:

$$\langle \nabla h(x), u\rangle = 2\left\langle Q\left(xx^T - zz^T\right), Q\left(xu^T + ux^T\right)\right\rangle = 0 \tag{10a}$$

$$\left\langle \nabla^2 h(x)u, u\right\rangle = 2\left\langle QQ\left(xx^T - zz^T\right), uu^T\right\rangle + \left\|Q(xu^T - ux^T)\right\|_F^2 \geq 0 \tag{10b}$$

for all $u \in \mathbb{R}^{n\times r}$. We re-arrange the first-order condition (10a):

$$\left(\left(xx^T - zz^T\right)M + M\left(xx^\top - zz^\top\right)\right)z = 0 \tag{11}$$

If the Eq. (11) does not hold for some $M \succ 0$, then x cannot be a critical point for that z and M. Consequently, the problem

$$\begin{aligned} \underset{M\in\mathbb{S}^n,\alpha\in\mathbb{R}}{\text{minimize}} \quad & -\alpha \\ \text{subject to} \quad & \left((zz^\top - xx^\top)M + M(zz^\top - xx^\top)\right)z = 0 \qquad (12a) \\ & M - \alpha\, I \succeq 0, \qquad (12b) \end{aligned}$$

is bounded from below by 0 if and only if the Eq. (11) does not hold for arbitrary $M \succ 0$. If x is critical for some $M \succ 0$, then it is unbounded.

The problem (12) is a semidefinite program with a zero duality gap, since $M = 0$ and $\alpha = -1$ constitute a strictly feasible primal point. We introduce the dual variable $\lambda \in \mathbb{R}^{n\times r}$ for the equality constraint (12a) and the dual variable $G \in \mathbb{S}^n$ for the positive semi-definite (PSD) constraint (12b). The dual problem can be written as

$$\max_{\lambda\in\mathbb{R}^{n\times r}, G\succeq 0} \min_{M\in\mathbb{S}^n,\alpha\in\mathbb{R}} \operatorname{Tr}\left[\left(2\operatorname{Sym}\left[\left(z\lambda^T + \lambda z^T\right)\left(zz^T - xx^T\right)\right] - G\right)M\right] + \alpha(\operatorname{Tr}(G) - 1)$$

The inner optimization problem has a finite solution if and only if

$$\begin{cases} G & = \left(z\lambda^\top + \lambda z^\top\right)\left(zz^\top - xx^\top\right) + \left(zz^\top - xx^\top\right)\left(z\lambda^\top + \lambda z^\top\right) \\ \operatorname{Tr}(G) & = 1 \end{cases}$$

The dual problem can be expressed as

$$\begin{aligned} \underset{\lambda\in\mathbb{R}^{n\times r}, G\in\mathbb{S}^n}{\text{maximize}} \quad & 0 \\ \text{subject to} \quad & G = \operatorname{Sym}\left[\left(z\lambda^\top + \lambda z^\top\right)\left(zz^\top - xx^\top\right)\right], \\ & \operatorname{Tr}(G) = 1, \\ & G \succeq 0 \end{aligned}$$

This is feasible if and only if the primal problem (12) is bounded. Consequently, it is feasible if and only if the point $x \in \mathbb{R}^{n\times r}$ is not a critical point of the function h for all $M \succ 0$.

To eliminate the condition on the trace, notice that a PSD matrix has a nonnegative trace that is equal to zero if and only if the matrix is the zero matrix. Since the constraints are homogeneous in G, the trace can always be normalized to 1. Thus, the dual feasibility is equivalent to the condition $0 \neq G \succeq 0$. This concludes the proof. □

This condition will be relaxed further for simplicity below.

Lemma 3 *Given $z \in \mathbb{R}^{n\times r}$, a point $x \in \mathbb{R}^{n\times r}$ is not a first-order critical point of the function $h(\cdot)$ for an arbitrary full-rank matrix Q if there are $T_1 \in \mathbb{R}^{r\times r}$ and $T_2 \in \mathbb{S}^r$ such that the matrix $T = \begin{bmatrix} 0 & T_1 \\ -T_1^T & T_2 \end{bmatrix}$ satisfies the relations*

$$0 \neq [-z \;\; x]\left(T^T P + PT\right)\begin{bmatrix} -z^T \\ x^T \end{bmatrix} \succeq 0 \tag{13}$$

where $P = \begin{bmatrix} z^T \\ x^T \end{bmatrix}[z \;\; x]$.

Proof Suppose that there exists T satisfying the condition of the lemma. Notice that

$$xT_1^T z^T + \frac{1}{2}xT_2x^T + zT_1x^T + \frac{1}{2}xT_2x^T = [-z\ \ x]\begin{bmatrix} 0 & -T_1 \\ T_1^T & T_2 \end{bmatrix}\begin{bmatrix} z^T \\ x^T \end{bmatrix}$$
$$= [z\ \ x]\begin{bmatrix} 0 & T_1 \\ -T_1^T & T_2 \end{bmatrix}\begin{bmatrix} -z^T \\ x^T \end{bmatrix}$$

and

$$xx^T - zz^T = [z\ \ x]\begin{bmatrix} -z^T \\ x^T \end{bmatrix} = [-z\ \ x]\begin{bmatrix} z^T \\ x^T \end{bmatrix}$$

We use the above formulas to expand the condition (13) and obtain

$$0 \neq \mathrm{Sym}\left[\left(x\left(zT_1 + \frac{xT_2}{2}\right)^T + \left(zT_1 + \frac{xT_2}{2}\right)x^T\right)(xx^T - zz^T)\right] \succeq 0,$$

The proof follows immediately from applying Lemma 2 with $\lambda = zT_1 + \frac{1}{2}xT_2$. □

To prove Proposition 1, we check the previous condition for all pairs of z and x.

Proof of Proposition 1 We start by proving that no point other than 0 and $\pm z$ can be a first-order critical point of the function h. Assume that $x \notin \{0, \pm z\}$. By Lemma 3, it is sufficient to prove that there are α and β in $\mathbb{R}$ such that the matrix $T = \begin{bmatrix} 0 & \alpha \\ -\alpha & \beta \end{bmatrix}$ satisfies

$$0 \neq G = [-z\ \ x]\left(T^T P + PT\right)\begin{bmatrix} -z^T \\ x^T \end{bmatrix} \succeq 0$$

where $P = \begin{bmatrix} z^T \\ x^T \end{bmatrix}[z\ \ x]$. Consider three scenarios for x and z :

Case 1 $x = \gamma z$: One can write

$$G = [-z\ \ x]\left(T^T P + PT\right)\begin{bmatrix} -z^T \\ x^T \end{bmatrix} = 2\gamma(\gamma^2 - 1)(2\alpha + \beta\gamma)zz^T zz^T$$

For $\alpha = \gamma(\gamma^2 - 1)$ and $\beta = 0$, it holds that $G = \left(2\gamma(\gamma^2 - 1)zz^T\right)^2 \succeq 0$. The matrix is nonzero for $x \notin \{0, \pm z\}$.

Case 2 $z^T x = 0$: The matrix P takes the form $P = \begin{bmatrix} \|z\|^2 & 0 \\ 0 & \|x\|^2 \end{bmatrix}$. Therefore, for $\alpha = 0$ and $\beta = 1$, it holds that

$$G = 2\|x\|^2 xx^T \succeq 0$$

The matrix is nonzero for $x \neq 0$.

Case 3 $0 < (z^T x)^2 < \|z\|_2^2\|x\|_2^2$: By scaling, we can assume without loss of generality that $z^T x = 1$; thus $P = \begin{bmatrix} \|z\|_2^2 & 1 \\ 1 & \|x\|_2^2 \end{bmatrix}$. It is sufficient to show that $T^T P + PT \succ 0$ to

guarantee G to be nonzero and PSD. To show this, we use Sylvester's criterion. The upper-left corner of this matrix is equal to -2α. Moreover,

$$\det\left(T^T P + PT\right) = \left(-\left(\|x\|_2^2 - \|y\|_2^2\right)^2 - 4\right)\alpha^2 - 2\left(\|x\|_2^2 + \|y\|_2^2\right)\alpha\beta - \beta^2$$

For $\alpha = -1$, the discriminant of this quadratic polynomial with respect to β is equal to $D = 16(\|z\|_2^2\|x\|_2^2 - 1)$. By the strict Cauchy–Schwarz inequality in the assumption of the case, D is strictly greater than 0. Thus, there exists β such that the matrix is positive definite. This implies that none of $x \notin \{0, \pm z\}$ satisfies the first-order necessary condition of local optimality for an unconstrained problem. Assume that $x = 0$. The quadratic form on the Hessian at this point

$$\left\langle \nabla^2 h(0)u, u \right\rangle = -2\left\langle Qzz^T, uu^T \right\rangle$$

takes a negative value at $u = z$. Thus, it does not satisfy the second-order necessary condition of local optimality for an unconstrained problem. The points $x = \pm z$ are not spurious points, which concludes the proof. □

4.2 Ellipsoid norm: higher ranks

The function $h(\cdot)$ defined over $\mathbb{R}^{n\times r}$ is significantly harder to study analytically if $r > 1$. En empirical analysis of Theorem 2 allows us to make a conjecture.

Conjecture 1 *For the kernel structure operator introduced in Proposition 1, no instance of the* (Problem$^{\text{KSP}}$) *with* $\mathscr{Z} = \mathbb{R}^{n\times r}$ *has a spurious second-order critical points over* $\mathbb{R}^{n\times r}$ *for an arbitrary* r.

This conjecture is based on the evaluation of $\mathbb{O}(x, z; \mathscr{T})$ at 72,000 pairs of points $x, z \in \mathbb{R}^{8\times 3}$ randomly sampled from a standard Gaussian distribution. All of them have the optimal value 1. However, if we consider first-order critical points as well, then it is straightforward to find a counterexample. After dropping the constraint on $\mathscr{M}_{x,z}$ in the formulation of $\mathbb{O}(x, z; \mathscr{T})$ and $\mathbb{O}_P(x, z; \mathscr{T})$, one can formulate a statement similar to Theorem 2 tailored to first-order solutions. The following proposition presents a corollary of the result.

Proposition 2 *For the kernel structure operator introduced in Proposition 1, for every* $n \geq 8$ *and* $r > 1$, *there is* $z \in \mathbb{R}^{n\times r}$ *such that* (Problem$^{\text{KSP}}$) *has a spurious saddle point.*

Proof First, we prove it for $n = 8$ and $r = 2$ by a counterexample. Consider the two points

$$x = \begin{bmatrix} 0 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 0 \\ -1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad z = \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & 0 \\ 1 & 0 \\ 1 & -1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}$$

and find a matrix $\mathscr{H}$ that solves $\mathbb{O}(x, z; \mathscr{T})$ without the constraint on $\mathscr{M}_{x,z}$. This will result in $\mathscr{H}$ such that $\mathscr{M}_{x,z}(\mathscr{H})$ has both negative and positive eigenvalues. For larger values of n and r, one can fill up the extra entries with zeros and the proof carries over. □

The code for reproducing the result is available on-line[1]. It took 482 tosses to generate the counterexample of the matrices containing only ± 1 or 0 as their components. We used the uniform distribution over those matrices to generate the tosses.

4.3 DC power systems with acyclic topology

In Sect. 4.2, we studied one particular structure for the operator $\mathscr{A}$. Now, we analyze a real-world problem to highlight the role of the KSP. Recall that the power system discussed in Sect. 2 was an AC network for which the voltages were complex numbers. To simplify the computation, we analyze a DC system in this section, where all voltages are real-valued [13]. Assume that there are n nodes, associated with the unknown real-valued voltages $\tilde{x}_1, \ldots, \tilde{x}_n$. The power is measured at each node $i \in \{1, \ldots, n\}$ and is denoted as $\tilde{p}_i$, which can be calculated according to the formula:

$$p_i(\tilde{x}) = \sum_{j \in N(i)} \tilde{x}_i \left(\tilde{x}_i - \tilde{x}_j\right) \frac{1}{r_{ij}} = \tilde{p}_i,$$

where $N(i) \subset \{1, \ldots, n\}$ is the set of nodes adjacent to node i and $r_{ij} = r_{ji} > 0$ is the resistance of the line between nodes i and j. The least-squares formulation of the voltage recovery problem consists in minimization over the set $v \in \mathbb{B}_R(\mathbf{1})$ of

$$f(x) = \sum_{i=1}^{n} (p_i(x) - \tilde{p}_i)^2$$

which is a special case of the function (1). Let R be a number such that $2x_i > x_n$ for all $i \in \{1, \ldots, n-1\}$. In this subsection, we will demonstrate the application of

[1] github.com/igormolybog/matrix-sense-global.

our results on a specific topology of the network, although as discussed later on, our conclusion applies to any acyclic topology.

Suppose that the network possesses a star topology, meaning that each node $i \in \{1, \ldots, n-1\}$ is connected to only one node — namely, node n — and no others. This means that $N(i) = \{n\}$ if $i \neq n$ and $N(n) = \{1, \ldots, n-1\}$. As a result, the power measurements in this particular case can be written as

$$p_i(x) = x_i(x_i - x_n)\frac{1}{r_{in}}, \qquad i \in \{1, \ldots, n-1\}$$
$$p_n(x) = \sum_{j=1}^{n-1} x_n(x_n - x_j)\frac{1}{r_{jn}}$$

which generate a particular structure for the sensing operator. Solving $a_i \mathrm{vec}\,(xx^T - \tilde{x}\tilde{x}^T) = p_i$ for a_i, we conclude that the rows $a_1, \ldots, a_n$ of $\mathbf{A}$ can be written as

$$a_i = \xi_i \mathrm{vec}\,(E_{ii} - E_{ni}), \qquad i \in \{1, \ldots, n-1\}$$
$$a_n = -\mathrm{vec}\left(\xi_n E_{nn} + \sum_{j=1}^{n-1} \xi_j E_{jn}\right)$$

where E_{ij} is an $n \times n$ matrix with (i, j)-th entry equal to 1 and all other entries equal to zero, and where $\xi_i = \frac{1}{r_{in}} = \frac{1}{r_{ni}} > 0$ for $i \neq n$ and $\xi_n = -\sum_{j=1}^{n-1} \xi_j$. The corresponding kernel matrix is given by

$$\mathbf{H}_{star}(\xi) := \mathbf{H} = \mathbf{A}^T\mathbf{A} = \sum_{i=1}^{n} a_i^T a_i$$

As a result, $\mathbf{H}$ has a structured sparsity pattern: it is a block-diagonal matrix with n blocks $M_1, \ldots, M_n \in \mathbb{S}^n$ such that the first $n-1$ blocks each have only four nonzero entries:

$$M_i = \xi_i^2[E_{ii} - E_{in} - E_{ni} + E_{nn}] = \xi_i^2[e_n - e_i][e_n - e_i]^T, \qquad i \in \{1, \ldots, n-1\}$$

where e_i is the i-th column of the $n \times n$ identity matrix. The last block of $\mathbf{H}$ is a full matrix:

$$M_n = \xi\xi^T$$

The sparsity pattern of $\mathbf{H}$ is visualized for $n = 4$ in Fig. 2. The matrix $\mathbf{H} = \mathbf{H}_{star}(\xi)$ also has a low-dimensional structure: the $4(n-1) + n^2$ nonzero entries of $\mathbf{H}$ quadratically depend on only $n-1$ parameters $\xi_1, \ldots, \xi_{n-1}$. In Sect. 6, we will demonstrate how linearization of the structure can be applied, while here we provide an analytical proof that deals with the nonlinear low-dimensional structure directly. This proof sheds light on some of the ideas behind Theorem 2.

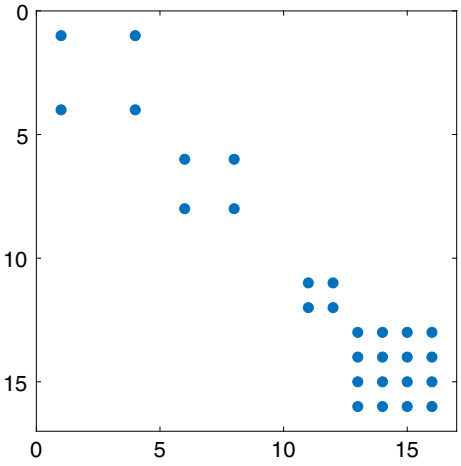

Fig. 2 Sparsity pattern of the matrix $\mathbf{H}$ corresponding to a DC power system with a star topology consisting of four buses

The 2-RIP constant of the sensing operators that correspond to star topology power networks does not exist due to their sparsity. Therefore, Theorem 1 cannot be applied. Nevertheless, we will show that the non-convex voltage recovery problem on a system with a star topology possesses no spurious local minima.

Proposition 3 *Consider the problem*

$$\left\{ \underset{x \in \bar{\mathbb{B}}_{1/3}(1)}{\text{minimize}} \; f_{z,\mathbf{A}}(x) \middle| \mathbf{H} = \mathbf{H}_{star}(\xi);\; z \in \bar{\mathbb{B}}_{1/3}(\mathbf{1}) \text{ and } \xi^T p(z) \neq 0 \right\},$$

or equivalently, (Problem$^{\text{KSP}}$) *under the additional constraint* $\xi^T p(z) \neq 0$. *No instance of this problem has a spurious second-order critical point.*

Proof Since $R < \frac{1}{3}$, it holds that $2x_i > x_n$ and $2z_i > z_n$ for all $i \in \{1, \ldots, n\}$. We are interested in the landscape of the function

$$\begin{aligned} h(x) &= \text{vec}\,(xx^T - zz^T)^T \mathbf{H} \text{vec}\,(xx^T - zz^T) \\ &= \sum_{i=1}^{n} (x_i x - z_i z)^T M_i (x_i x - z_i z) \end{aligned}$$

To find the first and second derivatives, consider

$$\begin{aligned} h(x+u) = &\sum_{i=1}^{n} (x_i x - z_i z + x_i u + u_i x + u_i u)^T \\ &M_i\,(x_i x - z_i z + x_i u + u_i x + u_i u) \\ = h(x) &+ 2 \sum_{i=1}^{n} (x_i u + u_i x)^T\, M_i\,(x_i x - z_i z) \end{aligned}$$

$$+\sum_{i=1}^{n}\left(x_i u+u_i x\right)^T M_i\left(x_i u+u_i x\right)$$
$$+2\sum_{i=1}^{n}(u_i u)^T M_i\left(x_i x-z_i z\right)+o\left(|u|^2\right)$$

Selecting the term that is linear in u, the gradient takes the form

$$\begin{aligned}\nabla_x h(x) &= 2\sum_{i=1}^{n}\left[x_i M_i\left(x_i x-z_i z\right)+\mathrm{Tr}\left[x^T M_i\left(x_i x-z_i z\right)\right]e_i\right]\\ &= 2\sum_{i=1}^{n-1}\Big[\xi_i^2 x_i(e_n-e_i)(e_n-e_i)^T(x_i x-z_i z)\\ &\qquad +\mathrm{Tr}\left[x^T(e_n-e_i)(e_n-e_i)^T(x_i x-z_i z)\right]e_i\\ &\qquad +2x_i\xi\xi^T(x_i x-z_i z)+\mathrm{Tr}\left[x^T\xi\xi^T(x_i x-z_i z)\right]e_i\Big]\end{aligned}$$

which can be written in the compact form

$$\nabla_x h(x)=B(x)[p(x)-p(z)]$$

where the (i,j)-th component of the $n\times n$ matrix $B(x)$ is

$$B_{ij}=\begin{cases}\xi_i(2x_i-x_n) & \text{if } i=j,\ i\neq n\\ \sum_{s=1}^{n-1}\xi_i(2x_n-x_s) & \text{if } i=j=n\\ -\xi_i x_i & \text{if } i\neq j,\ i=n \text{ or } i\neq j,\ j=n\\ 0 & \text{otherwise}\end{cases}$$

and $p(x)$ is a vector with its i-th component eqal to $p_i(x)$. If $B(x)$ is non-singular at a point x, then x is a first-order critical point of $h(x)$ in the open ball $\bar{\mathbb{B}}_{1/3}(\mathbf{1})$ if and only if $p(x)-p(z)=\mathscr{A}(xx^T-zz^T)=0$, which implies that it is a global minimum. Therefore, it is essential to identify all points x such that $\det(B)=0$.

With a slight abuse of notation, we denote $B(x)$ with the shorthand notation B. Let β_n denote the (n,n)-th entry of B, which is equal to $\sum_{s=1}^{n-1}\xi_s(2x_n-x_s)$. Represent the matrix B as a block matrix: $B=\begin{bmatrix}B' & b^T\\ b & B''\end{bmatrix}$ with the scalar $B'=\xi_1(2x_1-x_n)$ and the $(n-1)$-dimensional vector $b^T=[0,\ldots,0,-\xi_1 x_1]$. Since $x\in\bar{\mathbb{B}}_{1/3}(\mathbf{1})$, we have $B'\neq 0$. One can write:

$$\det(B)=0 \iff \det(B''-B'^{-1}bb^T)=0$$

The new $(n-1)\times(n-1)$ matrix $B''-B'^{-1}bb^T$ is equal to B'' in all components but its $(n-1,n-1)$-th entry, which changes to

$$\begin{aligned}\beta_{n-1} &= -\xi_1\frac{x_1x_n}{2x_1-x_n}+\beta_n\\ &= -\xi_1\frac{x_1x_n}{2x_1-x_n}+\xi_1(2x_n-x_1)+\sum_{s=2}^{n-1}\xi_s(2x_n-x_s)\\ &= -2\xi_1\frac{(x_1-x_n)^2}{2x_1-x_n}+\sum_{j=2}^{n-1}\xi_s(2x_n-x_s)\end{aligned}$$

Repeating the above matrix reduction argument $n-1$ times yields that $\det(B)=0$ if and only if $\beta_1=0$, where

$$\beta_1=-2\sum_{i=1}^{n-1}\xi_i\frac{(x_i-x_n)^2}{2x_i-x_n}$$

Since $2x_i-x_n>0$, we conclude that $\beta_1=0$ if and only if $x_1=\cdots=x_n$. Therefore, there are no spurious solutions outside the set $\{x: x_1=\cdots=x_n\}$. To study this set, we derive the Hessian of $h(x)$ by extracting from $h(x+u)$ the term that is quadratic in u :

$$\begin{aligned}\nabla^2_{xx}h(x)=\sum_{i=1}^{n}\Big[&x_i^2M_i+x_i\left(e_ix^TM_i+M_ixe_i^T\right)+e_ix^TM_ixe_i^T\\ &+M_i\left(x_ix-z_iz\right)e_i^T+e_i\left(x_ix-z_iz\right)^TM_i\Big]\end{aligned}$$

and substitute $x_1=\cdots=x_n=x'$ or $x=x'\mathbf{1}$. At the same time, we substitute $M_i=\xi_i^2[e_n-e_i][e_n-e_i]^T$ and $M_n=\xi\xi^T$, and note that $(e_n-e_i)^T\mathbf{1}=\mathbf{1}^T(e_n-e_i)=0$ and $\xi^T\mathbf{1}=\mathbf{1}^T\xi=0$ by construction. After simplification, we obtain that

$$\begin{aligned}\nabla^2_{xx}h(x)\big|_{x=x'\mathbf{1}}=x'^2\Bigg[&\sum_{i=1}^{n-1}\xi_i^2(e_n-e_i)(e_n-e_i)^T+\xi\xi^T-\left(\xi e_n^T+e_n\xi^T\right)z_nz^T\xi\\ &-\sum_{i=1}^{n-1}\xi_i^2\left((e_n-e_i)e_i^T+e_i(e_n-e_i)^T\right)z_i(z_n-z_i)\Bigg]\end{aligned}$$

Consider the quadratic form $q(s,t)=[s\ \ldots\ s\ t]\nabla^2_{xx}h(x)\big|_{x=x'\mathbf{1}}[s\ \ldots\ s\ t]^T$, where $[s\ \ldots\ s\ t]\in\mathbb{R}^n$. One can write:

$$
\begin{aligned}
q(s,t) &= \sum_{i=1}^{n-1} \xi_i^2 (t-s)^2 + \left(\sum_{i=1}^{n-1} s\xi_i + t\xi_n\right)^2 \\
&\quad -2tz_n \left(\sum_{i=1}^{n-1} s\xi_i + t\xi_n\right)\left(\sum_{i=1}^{n-1} z_i\xi_i + z_n\xi_n\right) \\
&\quad -2s(t-s)\sum_{i=1}^{n-1} \xi_i^2 z_i (z_n - z_i) \\
&= (t-s)^2 \left[\sum_{i=1}^{n-1} \xi_i^2 + \left(\sum_{i=1}^{n-1} \xi_i\right)^2\right] \\
&\quad +2(t-s)\left[t\left(\sum_{i=1}^{n-1} \xi_i\right)\left(\sum_{i=1}^{n-1} \xi_i z_n (z_i - z_n)\right) - s\left(\sum_{i=1}^{n-1} \xi_i^2 z_i (z_n - z_i)\right)\right] \\
&= c_1(t-s)^2 + 2(t-s)(c_2 t - c_3 s) \\
&= (c_1 + 2c_2)t^2 - 2(c_1 + c_2 + c_3)st + (c_1 + 2c_3)s^2
\end{aligned}
$$

where c_1, c_2 and c_3 are some constants introduced to shorten the expression. If $q(s,t)$ takes negative values, then $\nabla^2_{xx} h(x)\big|_{x=x'\mathbf{1}}$ has a negative eigenvalue, and therefore $x = x'\mathbf{1}$ cannot be a spurious second-order critical point.

Now, consider the polynomials $q(1,t)$ and $q(s,1)$. Suppose that $\sum_{i=1}^{n-1} \xi_i^2 \geq \left(\sum_{i=1}^{n-1} \xi_i\right)^2$ and consider any point $z \in \bar{\mathbb{B}}_{1/3}(\mathbf{1})$. Since $|z_n(z_i - z_n)| \leq \frac{4}{3} \cdot \frac{2}{3} = \frac{8}{9}$, it must hold that

$$
c_1 + 2c_2 > \sum_{i=1}^{n-1} \xi_i^2 + \left(\sum_{i=1}^{n-1} \xi_i\right)^2 - 2\frac{8}{9}\left(\sum_{i=1}^{n-1} \xi_i\right)^2 = \sum_{i=1}^{n-1} \xi_i^2 - \frac{6}{9}\left(\sum_{i=1}^{n-1} \xi_i\right)^2 \geq 0
$$

Thus, $q(1,t)$ is a polynomial of order 2 with respect to t with a positive leading term. Suppose that $\sum_{i=1}^{n-1} \xi_i^2 < \left(\sum_{i=1}^{n-1} \xi_i\right)^2$. Similarly,

$$
c_1 + 2c_3 > \sum_{i=1}^{n-1} \xi_i^2 + \left(\sum_{i=1}^{n-1} \xi_i\right)^2 - 2\frac{8}{9}\sum_{i=1}^{n-1} \xi_i^2 = \left(\sum_{i=1}^{n-1} \xi_i\right)^2 - \frac{6}{9}\sum_{i=1}^{n-1} \xi_i^2 \geq 0
$$

and thus $q(s,1)$ is a polynomial of order 2 with respect to s with a positive leading term. At least one of the polynomials $q(1,t)$ or $q(s,1)$ has a positive leading term. Both $q(1,t)$ and $q(s,1)$ have the same determinant and thus the argument to be made below can be made for any of them. Without loss of generality, assume that $q(1,t)$ has a positive leading term and takes negative values if and only if its determinant is strictly positive:

$$
4(c_1 + c_2 + c_3)^2 - 4(c_1 + 2c_2)(c_1 + 2c_3) > 0.
$$

It is positive if and only if $(c_2 - c_3)^2 > 0$, which is equivalent to $c_2 \neq c_3$. After substituting $c_2 = (\sum_{i=1}^{n-1} \xi_i)(\sum_{i=1}^{n-1} \xi_i z_n(z_i - z_n)) = -(\sum_{i=1}^{n-1} \xi_i) p_n(z)$ and $c_3 = (\sum_{i=1}^{n-1} \xi_i^2 z_i(z_n - z_i)) = -\sum_{i=1}^{n-1} \xi_i p_i(z)$, one can guarantee that $\nabla^2_{xx} h(x)\big|_{x=x'\mathbf{1}}$ has a negative eigenvalue unless

$$\xi^T p(z) = 0.$$

Otherwise, $q(1, t)$ only reaches zero at $t = 1$ and never crosses it. □

The technique of using the properties of the Schur complement to reduce the dimension of a matrix one by one can be applied to any arbitrary network with an acyclic topology. For any such network, the gradient $\nabla_x h(x)$ also takes the form $\nabla_x h(x) = B[p(x) - p(z)]$, but with a different matrix B. Applying elimination of the rows and columns of B that correspond to the leaves first, then to the first layer of parent nodes, then to the second layer of parent nodes and so forth leads to a similar result on the location of first-order critical points. Thus, in a similar way, the conclusion of Proposition 3 can be proven for any arbitrary acyclic network, but for a different value of the radius R that may not be analytically calculable. However, the proof is not generalizable to networks with cycles. The proof of Proposition 3 was based on analyzing the first- and second-order optimality conditions and exploiting the properties of the operator $\mathscr{A}$ that benefits from both sparsity and a low-dimensional structure. The ideas used in the proof help the reader understand Theorem 2. Since Proposition 3 does not apply to networks with cycles, one may instead use Theorem 2 to numerically evaluate the inexistence of spurious solutions for any particular cyclic network. This will be carried out in Sect. 6.

5 Combining KSP with RIP

After fixing the hyperparameters $\omega \in \mathbb{R}^{n\times r}$ and $R \in \mathbb{R} \cup \{+\infty\}$ together with the kernel structure of the sensing operators and the RIP constant, we can state the problem under study in this section as follows:

$$\left\{ \underset{x\in\mathbb{B}_R(\omega)}{\text{minimize}} f_{z,\mathscr{A}}(x) \;\middle|\; \mathscr{A} \text{ satisfies } \delta_{2r}\text{-RIP and } \mathscr{T}\text{-KSP},\ z \in \mathbb{B}_R(\omega) \right\}, \tag{Problem$^{\text{KSP+RIP}}$}$$

Note that (Problem$^{\text{KSP+RIP}}$) consists in the minimization of a class of functions $f_{z,\mathscr{A}}$ that correspond to some point $z \in \mathbb{B}_R(\omega)$ and some operator $\mathscr{A}$ that satisfies $\mathscr{T}$-KSP and δ_{2r}-RIP simultaneously. This is a generalization of both (Problem$^{\text{RIP}}$) and (Problem$^{\text{KSP}}$). For (Problem$^{\text{KSP+RIP}}$), we provide necessary and sufficient conditions for having no spurious second-order critical point, and consequently no spurious local minimum.

Theorem 3 (KSP+RIP necessary and sufficient conditions) *For all instances of* (Problem$^{\text{KSP+RIP}}$)*, there are no spurious second-order critical points if*

$$\begin{cases} \delta_{2r} < \min\limits_{\substack{x\in\mathbb{B}_R(\omega), z\in\mathbb{B}_R(\omega)\\ xx^T\neq zz^T}} \mathbb{O}_P(x,z;\mathscr{T}) & \text{(14a)} \\ \delta_{2r} < \min\limits_{\substack{x\in\partial\mathbb{B}_R(\omega), z\in\mathbb{B}_R(\omega)\\ xx^T\neq zz^T}} \mathbb{O}_P^{\partial\mathbb{B}}(x,z;\mathscr{T},\omega) & \text{(14b)} \end{cases}$$

and only if

$$\begin{cases} \delta_{2r} < \min\limits_{\substack{x\in\mathbb{B}_R(\omega), z\in\mathbb{B}_R(\omega)\\ xx^T\neq zz^T}} \mathbb{O}(x,z;\mathscr{T}) & \text{(15a)} \\ \delta_{2r} < \min\limits_{\substack{x\in\partial\mathbb{B}_R(\omega), z\in\mathbb{B}_R(\omega)\\ xx^T\neq zz^T}} \mathbb{O}^{\partial\mathbb{B}}(x,z;\mathscr{T},\omega) & \text{(15b)} \end{cases}$$

Following the results of Zhang et al. [38], the necessary and sufficient conditions coincide for the trivial structure operator $\mathscr{T}\equiv 0$ and $R=+\infty$.

5.1 Robustness

Consider the scenario where the measurements are corrupted with independent and identically distributed Gaussian noise. More precisely, we assume that the measurement vector b is corrupted by an additive noise that can be written as $\mathscr{A}(V)$ for some random matrix V that is probably full rank (since $\mathscr{A}(\cdot)$ is from the high-dimensional space $\mathbb{S}^n$ to the presumably low-dimensional space $\mathbb{R}^m$, we just need the mild surjectivity assumption). In this section, we show that the resulting recovery error can be bounded with high probability. For simplicity, we consider the case $R=+\infty$, but a similar argument can be used to analyze the case with a finite radius.

Theorem 4 *Consider* (Problem$^{\text{KSP+RIP}}$) *with $R=+\infty$ for which the condition* (14a) *holds. Let $V\in\mathbb{S}^n$ be a random matrix of arbitrary rank. Define the noisy recovery loss*

$$g(x)=\left\|\mathscr{A}(xx^T-zz^T+V)\right\|.$$

For every $p\in(0,1)$ and $\varepsilon>0$, there exists $\sigma=\sigma(p,\varepsilon;\mathscr{A},z)>0$ such that for $V\sim\mathscr{N}(0,\sigma^2 I)$, with probability at least p, every second-order critical point $x^\star$ of $g(x)$ satisfies $\|x^\star x^{\star T}-zz^T\|<\varepsilon$

Proof Expand the recovery loss:

$$\begin{aligned} g(x) &= \left\langle xx^T-zz^T+V,(xx^T-zz^T+V)\right\rangle \\ &= f(x)+\left\langle V,\mathscr{H}(xx^T-zz^T)+xx^T-zz^T\right\rangle+\langle V,\mathscr{H}(V)\rangle \end{aligned}$$

We outline the proof below:

– Split $\mathbb{R}^n$ into four regions according to the behavior of $f(x)$ associated with the noiseless scenario:
  1. ε-neighborhood of the second-order critical points of $f(x)$
  2. some neighborhood of the remaining first-order critical points of $f(x)$
  3. inner compact region where the value of $\|\nabla f\|$ is bounded by some positive constants from below and from above
  4. Outer region, where $\|\nabla f\|$ is large;
– Show the existence of σ such that there are no second-order critical points of $g(x)$ in regions 2, 3 and 4 with high probability;
– Conclude that the only region that contains the second-order critical points of $g(x)$ with high probability is region 1, which coincides with the set $\{x : \|xx^T - zz^T\| < \varepsilon\}$.

The illustration of the regions used in the proof can be found in Fig. 3. To prove formally, first calculate the gradient

$$\nabla_x g(x) = \nabla_x f(x) + \nabla_x \left\langle V, \mathscr{H}(xx^T) + xx^T \right\rangle$$

For $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, r\}$, one can write:

$$\frac{\partial}{\partial x_{ij}} \langle V, xx^T \rangle = \left\langle V, e_i x_j^T + x_j e_i^T \right\rangle = 2 e_i^T V x_j$$

where e_i is the i-th column of the $n \times n$ identity matrix. Moreover,

$$\frac{\partial}{\partial x_{ij}} \langle V, \mathscr{H}(xx^T) \rangle = \langle V, \mathscr{H}(e_i x_j^T + x_j e_i^T) \rangle$$

The Hessian can also be written as

$$\nabla^2_{xx} g(x) = \nabla^2_{xx} f(x) + \nabla^2_{xx} \left\langle V, \mathscr{H}\left(xx^T\right) + xx^T \right\rangle$$

Similarly, for $i' \in \{1, \ldots, n\}$ and $j' \in \{1, \ldots, r\}$, we have

$$\frac{\partial^2}{\partial x_{ij} \partial x_{i'j'}} \left\langle V, xx^T \right\rangle = 2\delta_{jj'} \langle V, E_{ii'} \rangle$$

where $\delta_{jj'} \in \mathbb{R}$ is defined as $\delta_{jj'} = \begin{cases} 1 \text{ if } j = j' \\ 0 \text{ otherwise,} \end{cases}$ and $E_{ii'} \in \mathbb{R}^{n \times n}$ is a matrix whose (i, i')-th entry is 1 and other entries are 0. Similarly,

$$\frac{\partial^2}{\partial x_{ij} \partial x_{i'j'}} \left\langle V, \mathscr{H}(xx^T) \right\rangle = \delta_{jj'} \langle V, \mathscr{H}(E_{ii'} + E_{i'i}) \rangle$$

By assumption, there exists $\gamma > 0$ such that $\frac{1-\delta_{2r}}{\gamma}\|xx^T - zz^T\|_F^2 \leq \|\mathscr{A}(xx^T - zz^T)\|_F^2$ for all x. This implies that $f(x)$ is a coercive functions of x for any given $\mathscr{A}$ and z. Moreover, $\|\nabla_x f(x)\|$ is also a coercive function. To show this, using the notation from Sect. 4, consider

$$\begin{aligned}\left\langle \frac{x}{\|x\|_F}, \nabla_x f(x) \right\rangle &= \frac{2}{\|x\|_F}\langle \text{vec}\,(x), \mathbf{X}^T\mathbf{He}\rangle \\ &= \frac{2}{\|x\|_F}\langle \mathbf{X}\text{vec}\,(x), \mathbf{He}\rangle \\ &= \frac{2}{\|x\|_F}\langle xx^T + xx^T, \mathscr{H}(xx^T - zz^T)\rangle \\ &= \frac{4}{\|x\|_F}[f(x) + \langle zz^T, \mathscr{H}(xx^T - zz^T)\rangle]\end{aligned}$$

Knowing that $f(x)$ grows as fast as $\|xx^T\|_F^2 = \|x\|_F^4$ and $-\langle zz^T, \mathscr{H}(xx^T - zz^T)\rangle$ grows at most as fast as $\|x\|_F^2$, we conclude that $\langle \frac{x}{\|x\|_F}, \nabla_x f(x)\rangle \to \infty$ as $\|x\| \to \infty$, which implies that $\|\nabla_x f(x)\| \to \infty$ as $\|x\| \to \infty$.

For an arbitrary $K' > 0$, define the set $C_{K'} = \{x | f(x) \leq K', \|\nabla_x f(x)\| \leq K'\}$. It is compact due to the coerciveness. The difference $\nabla_x g(x) - \nabla_x f(x)$ is linear in both V and x, while $\nabla_x f(x)$ is cubic in x. Noting that $\|\nabla_x g(x)\| \geq \|\nabla_x f(x)\| - \|\nabla_x f(x) - \nabla_x g(x)\|$, one can conclude that $\|\nabla_x g(x)\|$ is also a coercive function. Therefore, for any $p_K \in (0, 1)$ there exist K and $\sigma = \sigma_K$ such that $\|\nabla_x g(x)\| > 0$ over $\mathbb{R}^n \backslash C_K$ with probability p_K. Select $p_K = \sqrt[3]{p}$ and fix the corresponding K and σ_K.

The set O_{fo} of first-order critical points of $f(x)$ is closed due to the closed graph theorem. Moreover, it is bounded due to coercievness of $f(x)$, and thus compact even when $R = +\infty$. Denote the set of second-order critical points of $f(x)$ with $O_{min} \subseteq O_{fo}$. It coincides with the set of global minimizers of $f(x)$ since the condition (14a) holds and Theorem 3 can be utilized. Define $U_{min} = \cup_{x \in O_{min}} \bar{\mathbb{B}}_\varepsilon(x)$, and the set $O_{rest} = O_{fo} \backslash U_{min}$ that is compact. Note that the minimum eigenvalue of $\nabla_{xx}^2 f(x)$ is strictly negative for every $x \in O_{rest}$. Since minimum eigenvalue is a continious function, there exists $\bar{\lambda} < 0$ such that $\min_{x \in O_{rest}} \lambda_{min}(\nabla_{xx}^2 f(x)) = \bar{\lambda}$. By continuity of $\nabla_{xx}^2 f(x)$ with respect to x, there exists $\xi > 0$ such that $\lambda_{min}(\nabla_{xx}^2 f(x)) < \frac{\bar{\lambda}}{2}$ for all $x \in U_{rest} = \cup_{x' \in O_{rest}} \bar{\mathbb{B}}_\xi(x')$. The difference of Hessians $\nabla_{xx}^2 g(x) - \nabla_{xx}^2 f(x)$ is linear in V and constant in x. Therefore, for any $\psi > 0$ and $p_\psi \in (0, 1)$, there exists σ_ψ such that with probability p_ψ, it holds that $\|\nabla_{xx}^2 g(x) - \nabla_{xx}^2 f(x)\|_F < \psi$ for all $x \in C_K$. Select $\psi = \frac{\bar{\lambda}}{3}$ and $p_\psi = \sqrt[3]{p}$, and fix the corresponding σ_ψ. Notice that under $\sigma \leq \sigma_\psi$, with probability p_ψ there are no second-order critical points of $g(x)$ in U_{rest}.

Denote $U_{fo} = U_{rest} \cup U_{min}$ and notice that $C_K \backslash U_{fo}$ is a compact set that contains no first-order critical points of $f(x)$. Therefore, there exists $\rho > 0$ such that $\|\nabla_x f(x)\| > \rho$ for all $x \in C_K \backslash U_{fo}$. Due to the continuity of $\nabla_x g(x) - \nabla_x f(x)$, for any $\phi > 0$ and $p_\phi \in (0, 1)$, there exists $\sigma = \sigma_\phi$ such that with probability p_ϕ it holds that $\|\nabla_x g(x) - \nabla_x f(x)\|_F < \phi$ for all $x \in C_K$. Select $\phi = \rho$ and $p_\phi = \sqrt[3]{p}$, and

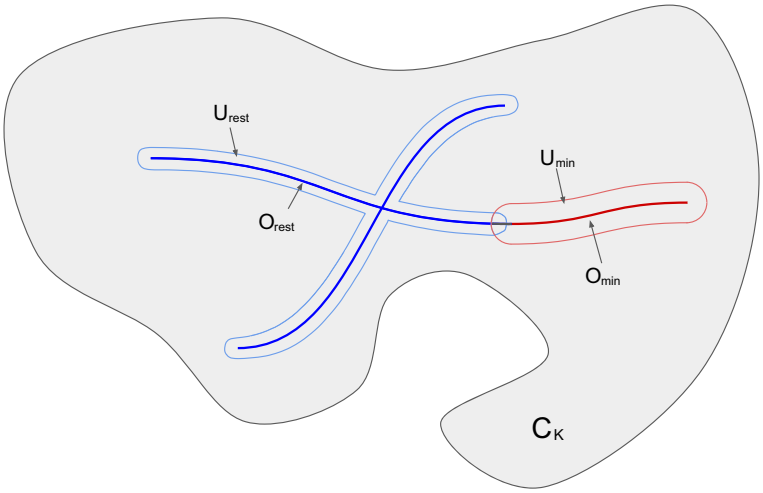

Fig. 3 Schematic of the domain of the function $f(x)$ with highlighted regions. The grey area denotes the compact region C_K. The bold lines denote the set of first-order critical points named O_{fo} whose subset shown in red corresponds to the set of global minimizers named O_{min}, while the blue part corresponds to O_{rest}. The area countered by the red shaded line is the ε-neighborhood of O_{min}, namely U_{min}, while the area countered by the blue shaded line is the ξ-neighborhood of O_{rest}, namely U_{rest}. The proof finds that with high probability there are no second-order critical points of $g(x)$ outside of C_K (outer region 4), or inside U_{rest} (region 2), or inside $C_K \backslash [U_{rest} \cup U_{min}]$. Therefore, all such points must be located inside U_{min}. (colour figure online)

fix the corresponding σ_ϕ. Notice that under $\sigma \leq \sigma_\phi$, with probability p_ϕ there are no second-order critical points of $g(x)$ in $C_K \backslash U_{fo}$.

To conclude the proof, select $\sigma < \min\{\sigma_K, \sigma_\psi, \sigma_\phi\}$ and observe that with probability at least $p_K \times p_\psi \times p_\phi = p$ there are no second-order critical points of $g(x)$ in the set $\mathbb{R}^n \backslash U_{min} = [\mathbb{R}^n \backslash C_K] \cup U_{rest} \cup [C_K \backslash U_{fo}]$. □

5.2 Sparse structure and normalization

Due to Theorem 1 for the rank-1 case, the instances of (Problem$^{\text{KSP+RIP}}$) have no spurious solutions with $\mathscr{T} \equiv 0$ as long as δ_2 is upper bounded by $\frac{1}{2}$. In this subsection, we are concerned with the question of how much sparsity can impact the best bound on RIP that certifies global convergence. Formally, we set $\mathscr{W} \equiv 0$ and $\mathscr{T} \equiv \mathscr{S}$ and find a tighter upper bound on δ_2. After enforcing sparsity, it is natural to expect that the bound grows and becomes less restrictive. However, this turns out not to be the case.

Let $n = 2$ and $r = 1$, and consider the smallest sparsity pattern possible for $\mathscr{H} = \mathscr{A}^T \mathscr{A} \succ 0$. It consists exclusively of elements (i, i), and thus enforces $\mathbf{H}$ to be diagonal. Consider the point x with respect to the instance of the problem given by z and $\mathbf{A}$ as in the example below:

Example 1 Assume that

$$x = (1, 1); \quad z = (\sqrt{2}, -\sqrt{2}); \quad \mathbf{A} = \text{diag}(\sqrt{3}, 1, 1, \sqrt{3})$$

Then, x is spurious for $f_{z,\mathbf{A}}$ since it satisfies the second-order necessary conditions:

$$\nabla f_{z,\mathbf{A}}(x) = 0, \quad \nabla^2 f_{z,\mathbf{A}}(x) = 16 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \succeq 0$$

which makes it a spurious second-order critical point (note that $xx^T \neq zz^T$). Notice that $\mathscr{H} = \mathscr{A}^T \mathscr{A}$ is indeed diagonal. Moreover, for all $X \in \mathbb{S}^2$, the operator $\mathscr{A}$ satisfies the tight bound $\|X\|_F^2 \leq \|\mathscr{A}(X)\|^2 = \| \begin{bmatrix} \sqrt{3} & 1 \\ 1 & \sqrt{3} \end{bmatrix} \circ X \| \leq 3\|X\|_F^2$. Therefore, the largest number δ_2 for this instance is equal to $1/2$, which coincides with the upper bound for unstructured problems. Somewhat counter-intuitively, the tight bound established in Zhang et al. [38,39] holds even when a very restrictive sparsity pattern of the kernel operator is enforced. Nevertheless, for an arbitrary low-dimensional structure $\mathscr{W}$, a tighter sparsity constraint entails a less restrictive bound on the RIP constant as discussed below.

Proposition 4 *If the sparsity pattern S has a sub-pattern S' meaning that $S' \subset S$, then $\mathbb{O}(x, z; \mathscr{W}, \mathscr{S}') \leq \mathbb{O}(x, z; \mathscr{W}, \mathscr{S})$ for all $x, y \in \mathbb{R}^{n \times r}$. Thus, the necessary bound on the RIP constant for $\mathbf{H}$ with S' is not more restrictive than the bound for $\mathbf{H}$ with S.*

In other words, a more restrictive assumption on the sparsity of the kernel operator can only push the upper bound on the RIP constant higher up. Consequently, Example 1 shows that there is no sparsity pattern of cardinality greater than 3 that can itself compensate the lack of isometry. Note that the example is given for the case $n = 2$, but there is a straightforward extension to an arbitrary n by adding zero components to x and z. It is common in practice to normalize the rows of the sensing matrix before proceeding to recovery. In the context of power systems, it is expressed as $x^T M_i x \rightarrow \frac{x^T M_i x}{\|M_i\|_F}$. For Example 1, after normalization, $\mathbf{A}$ turns into the identity. The corresponding instance of the problem is known to have no spurious critical points. This illustrates how normalization helps to improve the isometry property of the sensing operator and removes the spurious second-order critical points out of the corresponding instance of the problem. Normalization in this case can be regarded as inducing structure on top of sparsity.

6 Numerical results

It is desirable to numerically study the non-convex matrix recovery in problems with a structured sensing operator. The objective is to show how the general theory developed in Sect. 3 can be applied to a real-world problem, namely the power system state estimation discussed in Sect. 2. In general, optimization problems in (14) and (15) are non-convex. Thus, we propose to use Bayessian optimization [9] in order to obtain a numerical estimation of their solutions. We have empirically observed that Bayesian

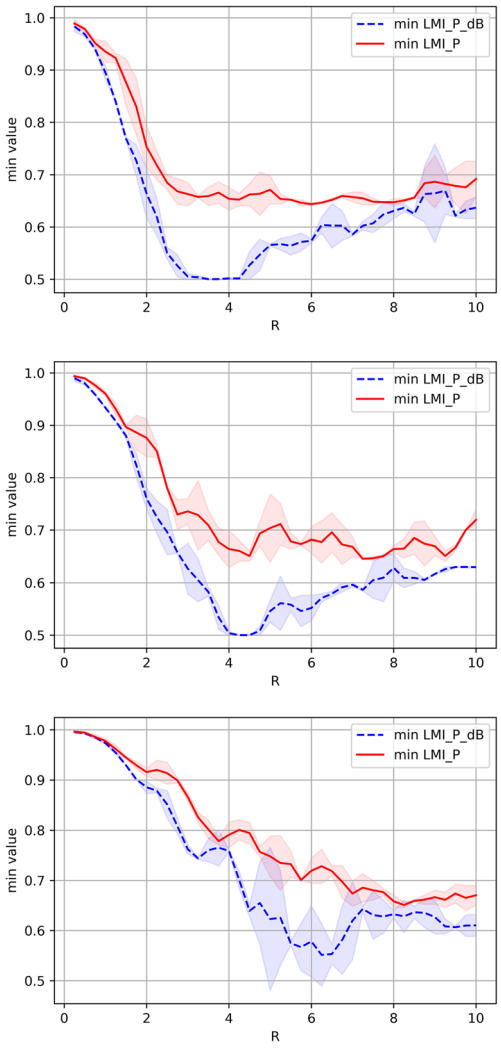

Fig. 4 The outcome of the minimization of $\mathbb{O}_P(x, z)$ and $\mathbb{O}_P^{\partial\mathbb{B}}(x, z)$ with the Bayessian optimization toolbox. The resulting value is the approximation of the right-hand side of the inequalities in (14) and can be used in Theorem 3 to estimate the lower bound on the sufficient RIP constant for global optimality. The values of the radius of the domain ball $\mathbb{B}_R(\omega)$ are on the x-axis, and the corresponding approximations of $\min \mathbb{O}_P(x, z)$ and $\min \mathbb{O}_P^{\partial\mathbb{B}}(x, z)$ are on the y-axis. The red line depicts the lowest observed value of the function $\mathbb{O}_P(x, z)$ and the blue dashed line depicts the minimum value of the function $\mathbb{O}_P^{\partial\mathbb{B}}(x, z)$

optimization tends to obtain the same optimal solution to this problem much faster than random shooting or cross-entropy.

6.1 Power systems

In this section, we focus our attention on three networks named `case9`, `case14` and `case30` that are provided in the MATPOWER package [44]. For `case9`, the number of buses is $n = 9$ and there are $m = 63$ possible power measurements that can be collected, while we have $n = 14$ and $m = 98$ for `case14` and have $n = 30$ and $m = 210$ for `case30`. We denote the corresponding sensing operators with $\mathscr{A}^9$, $\mathscr{A}^{14}$ and $\mathscr{A}^{30}$. Both matrices $\mathbf{A}^{30}$ and $\mathbf{H}^{30}$ are visualized in Fig. 1.

We linearize the low-dimensional structure that was discussed in Sect. 4.3. Repetition of the nonzero entries of $\mathbf{H}^{\cdot}$ (after some scaling) is considered as a form of low-dimensional structure, instead of the nonlinear dependence on the admittance. For example, if the entries (i, j) and (i', j') of $\mathscr{H}^{30}$ are equal, then $\mathscr{W}^{30}$ is constructed to be such that its kernel consists of matrices, for which the entries (i, j) and (i', j') are equal.

Based on this property, we form the linear operators $\mathscr{T}^9$, $\mathscr{T}^{14}$ and $\mathscr{T}^{30}$. All of the matrices in their kernel subspace are rank deficient. In this case, Theorem 3 can only provide us with the trivial upper bound on the RIP: $\delta_2 < 1$. However, this operator will allow us to use Theorem 3 to find a less conservative bound on RIP to certify the inexistence of spurious solutions for the structured mapping. Although the power system state estimation aims to find a complex vector, it is straightforward to verify that $\langle a + \sqrt{-1}b, \mathbf{H}(a + \sqrt{-1}b)\rangle = \langle a, \mathbf{H}a\rangle + \langle b, -\mathbf{H}b\rangle$ for any real vectors a and b as well as a real symmetric matrix $\mathbf{H}$. Therefore, it is enough to consider (14) over $\mathscr{X} = \mathscr{Z} = \mathbb{R}^n$.

The purpose of the experiment is to study the dependence of δ_{2r} that is sufficient for the absence of spurious solutions in (Problem$^{\text{KSP+RIP}}$) on the radius R of the ball domain. Intuitively, one would expect the dependence to be monotonically decreasing, since the larger the domain is, the more solutions can appear there with some being spurious. However, this is not exactly what can be observed. Figure 4 shows the right-hand side of the inequalities in (14) from Theorem 3 for a range of values of R for three structure operators: $\mathscr{T}^9$, $\mathscr{T}^{14}$ and $\mathscr{T}^{30}$. In these experiments, the vector ω has the unit entries. The red line provides a guarantee on no spurious solutions in the interior of the domain, while the blue dashed line takes care of the spurious solutions on the boundary. Indeed, the red curve decreases monotonically and converges to a value around 0.64 for all of the experiments, while the blue dashed line decreases to 0.5 and recovers back to the same value afterwards. It turns out that 0.64 is the bound on δ_{2r} for $R = +\infty$ in each of the cases as well. This interesting behavior can be explained qualitatively. Consider a toy example with three cases in Fig. 5, where the domain grows from Case I to Case III. There are no spurious solutions in case I, whereas one appears in case II and disappears in case III. Notice that the spurious solution can only appear on the boundary, which motivates the steady behavior of the red curve in Fig. 4. Recall that the threshold 0.5 is valid for the trivial structure operator $\mathscr{T} \equiv 0$ and $R = +\infty$ and the blue curve never goes below it. Therefore, the constructed conditions of the absence of spurious local optimality are strictly superior to the previously known bound.

The above simulations were based on the networks provided in the package MATPOWER 7.0b1 [44]. Keeping the structure of a network, we set the parameters of the lines equal to each other to be able to better visualize the operator $\mathscr{H}$. All of the presented simulations were performed using the MATLAB bayesopt toolbox, and the MATLAB modeling toolbox CVX [14,15] with SDPT3 [34,35] as the underlying solver.

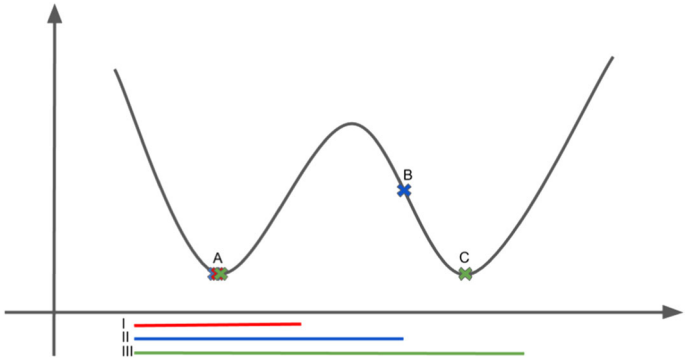

Fig. 5 Illustration for the local solution on the boundary. Three cases are considered, each marked with a different color. The colored intervals along the x-axis depict the domain in each of the cases, while the colored crosses denote the local solutions (colour figure online)

6.2 Synthetic data

In this subsection, we present numerical studies of the matrix recovery problem for structured sensing operators obtained from random ensembles. For simplicity, we set $R = +\infty$. In this section, the smallest value of δ_r such that $\mathscr{A}$ satisfies the δ_r-RIP property is referred to as *the best RIP constant* of the map $\mathscr{A}$.

Recall that the structure operator is defined by two operators stack together: $\mathscr{T} = (\mathscr{S}, \mathscr{W})$. Here, $\mathscr{W}$ captures the underlying structure that is not captured by the sparsity operator $\mathscr{S}$. We consider the same form of this operator as in the experiment on power systems data. Given the matrix representation $\mathbf{H}$ of the kernel operator, denote the unique nonzero values in this matrix with the scalars $h_1, \ldots, h_{d_\mathscr{W}}$. It means that $\mathbf{H}$ is representable in the form $\mathbf{H} = h_1\mathbf{E}_1 + \cdots + h_{d_\mathscr{W}}\mathbf{E}_{d_\mathscr{W}}$, where $\mathbf{E}_i$ is a matrix of the same size as $\mathbf{H}$ with 0 and 1 entries. The operator $\mathscr{W}$ that we use in this section is any operator that has the subspace $\{\beta_1\mathbf{E}_1 + \ldots + \beta_{d_\mathscr{W}}\mathbf{E}_{d_\mathscr{W}} \,|\beta_1, \cdots, \beta_{d_\mathscr{W}} \in \mathbb{R}\}$ as its kernel.

We introduce a distribution $RS(p_0, U)$ over the space of structure operators by describing the sampling scheme below. First, we generate the measurement structure matrix $\mathbf{A}_{\text{st}}$ such that each of its components takes value 0 with probability p_0 and any of the values $1, \ldots, U$ with the equal probability of $\frac{1-p_0}{U}$. We then form the kernel structure matrix as $\mathbf{H}_{\text{st}} = \mathbf{A}_{\text{st}}^T\mathbf{A}_{\text{st}}$ and construct the sparsity operator $\mathscr{S}$ and the extra structure operator $\mathscr{W}$ as discussed before. The obtained structure operator $\mathscr{T}$ is such that the operator represented with $\mathbf{A}_{\text{st}}$ satisfies the $\mathscr{T}$-KSP. Note that the average sparsity of $\mathbf{A}_{\text{st}}$ is p_0 and the number of unique nonzero values is U with high probability, which implies that p_0 is a parameter qualifying the amount of sparsity structure in the problem, and U is a parameter qualifying the amount of additional structure.

Figure 6 depicts the estimated sufficient RIP to guarantee the existence of no spurious second-order critical points in random problems with different values for the sparsity (p_0) and the unique counter (U). The sufficient RIP is obtained from Theorem 3 by imposing the KSP. Observe that the sparsity and the additional structure (the

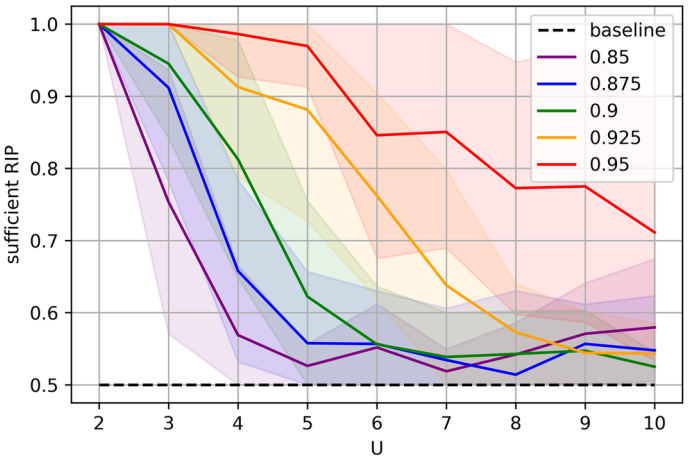

Fig. 6 The average of sufficient best RIP constant obtained from the developed analytic framework (Theorem 3) for random structures generated from the distribution $RS(p_0, U)$ (each colored line stands for one specific value of p_0), compared with the baseline method from Theorem 1 (shown as black and dashed). Shaded area represents the standard deviation window (colour figure online)

number of unique nonzero values in the measurement matrix in this particular case) both have a significant impact on the sufficient RIP. Note that a higher p_0 means more sparsity and a lower U means more extra structure. Although it was observed theoretically that sparsity alone could not guarantee an increase in the sufficient best RIP constant, it appears to be an important characteristic when combined with the additional structure. Even for structures with a considerably low sparsity (0.85), the tight extra structure ($U = 2$) has the sufficient best RIP of 1, which is a counter-intuitive result. The sufficient RIP seems to decay exponentially as we relax extra structure by increasing U, but with different bases for different p_0. This behavior coincides with the one predicted in Proposition 4. If the goal is to make the RIP higher than a certain threshold, the amount of extra structure needed to achieve this reduces dramatically with the increase of the sparsity structure.

The experiment demonstrates that our method can be successfully applied to matrix sensing with randomly generated structure. The key takeaway from this experiment is that our method captures the trade-off between the sparsity and the low-dimensional structural properties of a given mapping. It shows that imposing restrictions on structure significantly affects the sufficient RIP, which leads to certifying the absence of spurious solutions under far less restrictive requirements (by improving the previous RIP bound 0.5 for arbitrary mappings).

7 Conclusion

The paper is concerned with the theoretical explanation of the recent empirical success of solving the low-rank matrix sensing problem via nonconvex optimization. It is known that under an RIP assumption on the sensing operator, the optimization

problem has no spurious local minima. This assumption is too strong for real-world applications where the amount of data cannot be sufficiently high. Aside from that, it does not account for the prior knowledge about the solution that is available in different applications. We develop the notion of Kernel structure property (KSP) based on linear matrix inequalities, which can be used instead or combined with RIP in this context. KSP explains how the inherent structure of an operator contributes to the inexistence of spurious local minima over the entire space or a given ball. As a special case, we study sparse sensing operators that have a low-dimensional representation. Using KSP, we obtain novel necessary and sufficient conditions for having no spurious solutions over a compact set for the matrix sensing problem, and demonstrate them in analytical and numerical studies.

Acknowledgements This work was supported by grants from ARO, ONR, AFOSR, and NSF.

Appendix

In this part, we will prove Theorems 2 and 3 via showing the nonexistence of a counterexample. Specifically, given $\mathscr{T}$, for a point x and a parameter value z, we aim to find a value $\delta_{2r}^{x,z}$ for which the following claim holds:

"There exists $\mathscr{A}$ that satisfies $\mathscr{T}$-KSP and δ_{2r}-RIP such that x is a second-order critical point of $f_{z,\mathscr{A}}$ if and only if $\delta_{2r} > \delta_{2r}^{x,z}$"

$\mathscr{X}$ is equal to $\mathbb{B}_R(\omega)$ in the notation from Sect. 1. The conditions for a point x to be a second-order critical point of a function f over $\mathbb{B}_R(\omega)$ can be expressed in the compact form:

$$\begin{cases} \nabla f(x) = 0, \\ \nabla^2 f(x) \succeq 0 \end{cases} \text{ if } x \notin \partial\mathbb{B}_R(\omega) \text{ or } \begin{cases} \exists\, \mu \leq 0 : \ \nabla f(x) = \mu(x-\omega), \\ P_{x-\omega} \nabla^2 f(x) P_{x-\omega}^T \succeq 0 \end{cases} \text{ if } x \in \partial\mathbb{B}_R(\omega)$$

where $P_{x-\omega} \in \mathbb{R}^{(nr-1)\times nr}$ is the matrix of orthogonal projection onto the subspace orthogonal to $x - \omega$. With that in mind, we construct the two functions: $\delta(x, z)$ and $\partial\delta(x, z)$ via the following optimization procedures:

$$\begin{aligned} \delta(x,z) \quad \equiv \quad & \underset{\delta_{2r}\in\mathbb{R},\mathscr{A}}{\text{minimum}} \quad \delta_{2r} \\ & \text{subject to} \quad \mathscr{L}_{x,z}(\mathscr{A}^T\mathscr{A}) = 0 \\ & \qquad\qquad\ \ \mathscr{M}_{x,z}(\mathscr{A}^T\mathscr{A}) \succeq 0 \\ & \qquad\qquad\ \ \mathscr{T}(\mathscr{A}^T\mathscr{A}) = 0 \\ & \qquad\qquad\ \ \mathscr{A} \text{ satisfies } \delta_{2r}\text{-RIP}. \end{aligned}$$

$$\begin{aligned}\partial\delta(x,z) \quad \equiv \quad & \underset{\delta_{2r},\mu\in\mathbb{R},\mu\geq 0,\mathscr{A}}{\text{minimum}} \quad \delta_{2r} \\ & \text{subject to} \quad \mathscr{L}_{x,z}(\mathscr{A}^T\mathscr{A}) = -\mu(x-\omega) \\ & \qquad\qquad\quad P_{x-\omega}\mathscr{M}_{x,z}(\mathscr{A}^T\mathscr{A})P_{x-\omega}^T \succeq 0 \\ & \qquad\qquad\quad \mathscr{T}(\mathscr{A}^T\mathscr{A}) = 0 \\ & \qquad\qquad\quad \mathscr{A} \text{ satisfies } \delta_{2r}\text{-RIP.}\end{aligned}$$

In each of the problems, the first two constraints represent the requirement that x is a second-order critical point of $f_{z,\mathscr{A}}$, the third constraint takes care of the KSP, and the last one is the RIP. It is straightforward to verify that $\min\{\delta, \partial\delta\}$ takes the value of the desired $\delta_{2r}^{x,z}$. Minimization of $\delta_{2r}^{x,z}$ over $\{x \in \mathscr{X}, z \in \mathscr{Z} : xx^T \neq zz^T\}$ gives $\delta_{2r}^{\star}$ such that (Problem$^{\text{KSP+RIP}}$) with δ_{2r} has an instance with a spurious second-order critical point if and only if $\delta_{2r} > \delta_{2r}^{\star}$.

Suppose that we are able to find $\overline{\delta_{2r}^{x,z}}$ and $\underline{\delta_{2r}^{x,z}}$ such that $\underline{\delta_{2r}^{x,z}} \leq \delta_{2r}^{x,z} \leq \overline{\delta_{2r}^{x,z}}$ for all $x \in \mathscr{X}, z \in \mathscr{Z}$. Then,

$$\underline{\delta^{\star}} = \min_{\substack{x\in\mathscr{X},z\in\mathscr{Z}\\ xx^T\neq zz^T}} \underline{\delta_{2r}^{x,z}} \leq \min_{\substack{x\in\mathscr{X},z\in\mathscr{Z}\\ xx^T\neq zz^T}} \delta_{2r}^{x,z} \leq \min_{\substack{x\in\mathscr{X},z\in\mathscr{Z}\\ xx^T\neq zz^T}} \overline{\delta_{2r}^{x,z}} = \overline{\delta^{\star}}.$$

This inequality shows that $\delta_{2r} \geq \underline{\delta_{2r}^{\star}}$ is a sufficient, and $\delta_{2r} \leq \overline{\delta_{2r}^{\star}}$ is a necessary condition for the absence of spurious second-order critical points in the instances of the problem (Problem$^{\text{KSP+RIP}}$). Now, it is desirable to show that $\min\{\mathbb{O}_P^{\partial\mathbb{B}}(x,z;\mathscr{T},\omega), \mathbb{O}_P(x,z;\mathscr{T})\}$ can serve as $\underline{\delta_{2r}^{x,z}}$, and $\min\{\mathbb{O}^{\partial\mathbb{B}}(x,z;\mathscr{T},\omega), \mathbb{O}(x,z;\mathscr{T})\}$ can serve as $\overline{\delta_{2r}^{x,z}}$.

Lemma 4 *The following statements hold all $x \in \mathscr{X}$ and $z \in \mathscr{Z}$:*

$$\mathbb{O}_P(x,z) \leq \delta(x,z) \leq \mathbb{O}(x,z) \tag{16a}$$

$$\mathbb{O}_P^{\partial\mathbb{B}}(x,z) \leq \partial\delta(x,z) \leq \mathbb{O}^{\partial\mathbb{B}}(x,z) \tag{16b}$$

Proof Here, we show only inequality (16a) since (16b) can be shown similarly. Notice that for $P = \text{orth}([x,z])$, the following sequence of inclusions holds:

$$\{PYP^T : Y \in \mathbb{S}^{rank([x\ z])}\} \subseteq \{X \in \mathbb{S}^n : \text{rank}\,(X) \leq 2r\} \subseteq \mathbb{S}^n. \tag{17}$$

Let $(\mathscr{H}^*, \delta^*)$ denote the minimizer of the problem corresponding to $LMI(x,z)$. By the defenition of the $\mathbb{O}$ function, for every $X \in \mathbb{S}^n$ it holds that

$$(1-\delta^*)\|X\|_F^2 \leq \langle X, \mathscr{H}^*(X)\rangle = \|\mathscr{A}^*(X)\|^2 \leq (1+\delta^*)\|X\|_F^2$$

The decomposition $\mathscr{H}^* = \mathscr{A}^{*T}\mathscr{A}^*$ exists because $\mathscr{H}^* \succeq 0$. If the inequality holds for all $X \in \mathbb{S}^n$, it must hold when $rank(X) \leq 2r$, as noticed in (17). Thus, we conclude that the pair $(\mathscr{A}^*, \delta^*)$ is feasible for the problem defining $\delta(x,z)$. This proves the upper

bound. Similarly, if $(\mathscr{A}_*, \delta_*)$ is the minimizer of the problem defining $\delta(x, z)$, then by (17), the pair $(\mathscr{A}_*^T \mathscr{A}_*, \delta_*)$ is feasible for the problem defining $\mathbb{O}_P(x, z)$. This can be verified after rewriting the last constraint of the problem defining $\mathbb{O}_P$ in the form

$$\begin{aligned}(1-\delta)\|PYP^T\|_F^2 &\le \langle PYP^T, \mathscr{A}_*^T\mathscr{A}_*(PYP^T)\rangle \\ &= \|\mathscr{A}_*(PYP^T)\|^2 \le (1+\delta)\|PYP^T\|_F^2\end{aligned}$$

for all $Y \in \mathbb{S}^{rank(x,z)}$. It is important to notice that the same argument works for an arbitrary choice of $P \in \mathbb{R}^{n\times d}$ with $d \le 2r$. □

The above lemma completes the proof of Theorem 3. Theorem 2 follows by substituting 1 in the right-hand sides of 14 and 15. Notice that the linearity of the gradient and the Hessian with respect to the kernel operation matrix is the only property of the objective function that has been extensively used here. It can be exploited for generalization of the developed theory.

Proof of Proposition 4 We write the dual of the problem defining the function $\mathbb{O}$ as:

$$\underset{y,\lambda,U_1\succeq 0,U_2\succeq 0,V\succeq 0}{\text{maximize}} \quad \mathrm{Tr}[U_1 - U_2] \tag{18a}$$

$$\text{subject to} \quad \mathrm{Tr}[U_1 + U_2] = 1, \tag{18b}$$

$$\mathscr{L}_{x,z}^T(y) - \mathscr{M}_{x,z}^T(V) + \mathscr{T}^T(\lambda) = U_1 - U_2 \tag{18c}$$

This problem is the exact reformulation of

$$\underset{y\in\mathbb{R}^{n\times r},V\succeq 0,\mu\in\mathbb{R}^t}{\text{maximize}} \quad \frac{\sum_{i=1}^d(-\lambda_i(\mathscr{L}_{x,z}^T(y) - \mathscr{M}_{x,z}^T(V) + \mathscr{T}^T(\mu)))_+}{\sum_{i=1}^d(+\lambda_i(\mathscr{L}_{x,z}^T(y) - \mathscr{M}_{x,z}^T(V) + \mathscr{T}^T(\mu)))_+} \tag{19}$$

For details, please refer to Lemma 14 in Zhang, Sojoudi, and Lavaei [38]. Both primal and dual problems are bounded and the dual is strictly feasible. Recall vector $\mathbf{e}$ and matrix $\mathbf{X}$ from Sect. 4, where for all $u \in \mathbb{R}^{n\times r}$ it holds that

$$\mathbf{e} = \mathrm{vec}\,(xx^T - zz^T), \qquad \mathbf{X}\mathrm{vec}\,(u) = \mathrm{vec}\,(xu^T + ux^T)$$

A strictly feasible point of (18) can be chosen as $y = 0, \lambda = 0, V = \varepsilon I, U_1 = \eta I - \varepsilon W$ and $U_2 = \eta I + \varepsilon W$, where $2\eta = n^{-2}$, $2W = r[\mathrm{vec}\,(I)\mathbf{e}^T + \mathbf{e}\mathrm{vec}\,(I)^T] - \mathbf{X}\mathbf{X}^T$, and ε is sufficiently small to ensure that both U_1 and U_2 are PSD. Consequently, Slater's condition and strong duality hold, and thus the solution of (19) coincides with $\mathbb{O}(x, z)$. If $\mathscr{T} = (\mathscr{S}, \mathscr{W})$, then $\mathscr{T}^T(u, \mathbf{T}) = \mathscr{W}^T(u) + \mathscr{S}^T(\mathbf{T})$. At the same time, if $\mathscr{S}$ is represented by the matrix $\mathbf{S}$, then $\mathscr{S}(\mathbf{T}) = \mathscr{S}^T(\mathbf{T}) = \mathbf{S} \circ \mathbf{T}$. Let $\mathbf{S}$ and $\mathbf{S}'$ be the matrix representations of $\mathscr{S}$ and $\mathscr{S}'$, respectively. $S' \subset S$ means that there exists S^Δ such that $S = S' \cup S^\Delta$ and $\mathbf{S}' = \mathbf{S} + \mathbf{S}^\Delta$. It is straightforward to verify that for every $\mathbf{R} \in \mathbb{S}^{n^2}$ there exists $\mathbf{T} \in \mathbb{S}^{n^2}$ such that $\mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{R} = \mathbf{S}' \circ \mathbf{T}$. The opposite is also true: for every $\mathbf{T} \in \mathbb{S}^{n^2}$ there exists $\mathbf{R} \in \mathbb{S}^{n^2}$ such that $\mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{R} = \mathbf{S}' \circ \mathbf{T}$.

We introduce the short-hand notation $\mathscr{O}(y, V, u) = \mathscr{L}^T_{x,z}(y) - \mathscr{M}^T_{x,z}(V) + \mathscr{W}^T(u)$. One can write:

$$
\begin{aligned}
&\mathbb{O}(x, z; \mathscr{W}, \mathscr{S}') \\
&\quad = \underset{y\in\mathbb{R}^{n\times r}, V\succeq 0, u\in\mathbb{R}^l, \mathbf{T}\in\mathbb{S}^{n^2}}{\text{minimize}} \frac{\sum_{i=1}^d (-\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S}' \circ \mathbf{T}))_+}{\sum_{i=1}^d (+\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S}' \circ \mathbf{T}))_+} \\
&\quad = \underset{y\in\mathbb{R}^{n\times r}, V\succeq 0, u\in\mathbb{R}^l, \mathbf{T}\in\mathbb{S}^{n^2}, \mathbf{R}\in\mathbb{S}^{n^2}}{\text{minimize}} \frac{\sum_{i=1}^d (-\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{R}))_+}{\sum_{i=1}^d (+\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{R}))_+} \\
&\quad \leq \underset{y\in\mathbb{R}^{n\times r}, V\succeq 0, u\in\mathbb{R}^l, \mathbf{T}\in\mathbb{S}^{n^2}}{\text{minimize}} \frac{\sum_{i=1}^d (-\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{0}))_+}{\sum_{i=1}^d (+\lambda_i(\mathscr{O}(y, V, u) + \mathbf{S} \circ \mathbf{T} + \mathbf{S}^\Delta \circ \mathbf{0}))_+} \\
&\quad = \mathbb{O}(x, z; \mathscr{W}, \mathscr{S})
\end{aligned}
$$

This completes the proof. □

References

1. Agarwal, A., et al.: Learning sparsely used overcomplete dictionaries via alternating minimization. SIAM J. Optim. **26**(4), 2775–2799 (2016)
2. Bhojanapalli, S., Behnam, N., Srebro, N.: Global optimality of local search for low rank matrix recovery. In: Advances in Neural Information Processing Systems, pp. 3873–3881 (2016)
3. Bottou, L., Bousquet, O.: The tradeoffs of large scale learning. In: Advances in Neural Information Processing Systems,pp. 161–168 (2008)
4. Candes, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via Wirtinger flow: theory and algorithms. IEEE Trans. Inf. Theory **61**(4), 1985–2007 (2015)
5. Chen, Y., et al.: Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval. Math. Program. **176**, 5–37 (2019)
6. Chen, Y., et al.: Noisy matrix completion: understanding statistical guarantees for convex relaxation vianonconvex optimization. arXiv preprint arXiv:1902.07698 (2019)
7. Chi, Y., Lu, Y.M., Chen, Y.: Nonconvex optimization meets low-rank matrix factorization: an overview. arXiv preprint arXiv:1809.09573 (2018)
8. Daneshmand, H., et al.: Escaping saddles with stochastic gradients. arXiv preprint arXiv:1803.05999 (2018)
9. Frazier, P.I.: A tutorial on bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)
10. Ge, R., Jin, C., Zheng, Y.: No spurious local minima in nonconvex low rank problems: a unified geometric analysis. In: Proceedings of the 34th International Conference on Machine Learning, vol. 70, pp. 1233–1242. JMLR. org (2017)
11. Ge, R., Lee, J.D., Ma, T.: Matrix completion has no spurious local minimum. In: Advances in Neural Information Processing Systems, p. 29732981 (2016)
12. Ge, R., et al.: Escaping from saddle points: online stochastic gradient for tensor decomposition. arXiv:1503.02101 [cs.LG] (2015)
13. Ghosh, A., Boyd, S., Saberi, A.: Minimizing effective resistance of a graph. SIAM Rev. **50**(1), 37–66 (2008)
14. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1. (2014)
15. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, p. 95110. Springer-Verlag Limited (2008)
16. Jain, P., Netrapalli, P., Sanghavi, S.: Low-rank matrix completion using alternating minimization. In: Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, pp. 665–674. ACM (2013)

17. Josz, C., et al.: A theory on the absence of spurious solutions for noncon-vex and nonsmooth optimization. In: Advances in Neural Information Processing Systems, pp. 2441–2449 (2018)
18. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from a few entries. IEEE Trans. Inf. Theory **56**(6), 2980–2998 (2010a)
19. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from noisy entries. J. Mach. Learn. Res. **11**(jul), 2057–2078 (2010b)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
21. Lee, J.D., et al.: Gradient descent only converges to minimizers. In: Conference on Learning Theory, pp. 1246–1257 (2016)
22. Li, D., Ding, T., Sun, R.: Over-parameterized deep neural networks have no strict local minima for any continuous activations. arXiv preprint arXiv:1812.11039 (2018)
23. Li, Q., Zhu, Z., Tang, G.: Alternating minimizations converge to second-order optimal solutions. In: International Conference on Machine Learning, pp. 3935–3943 (2019)
24. Li, Xingguo, et al.: Symmetry, saddle points, and global optimization landscape of nonconvex matrix factorization. IEEE Trans. Inf. Theory **65**(6), 3489–3514 (2019)
25. Mei, S., Bai, Y., Montanari, A.: The landscape of empirical risk for non-convex losses. arXiv preprint arXiv.1607.06534 (2016)
26. Nouiehed, M. Razaviyayn, M.: Learning deep models: critical points and local openness. arXiv preprint arXiv:1803.02968 (2018)
27. Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is NP-hard. J. Glob. Optim. **1**(1), 15–22 (1991). https://doi.org/10.1007/BF00120662
28. Park, D., et al.: Non-square matrix sensing without spurious local minima via the Burer–Monteiro approach. arXiv preprint arXiv:1609.03240 (2016)
29. Paternain, S., Mokhtari, A., Ribeiro, A.: A newton-based method for nonconvex optimization with fast evasion of saddle points. SIAM J. Optim. **29**(1), 343–368 (2019)
30. Soltanolkotabi, M.: Learning relus via gradient descent. In: Advances in Neural Information Processing Systems, 2007–2017 (2017)
31. Sun, J., Qu, Q., Wright, J.: A geometric analysis of phase retrieval. Found. Comput. Math. **18**(5), 1131–1198 (2018)
32. Sun, J., Qu, Q., Wright, J.: Complete dictionary recovery using nonconvex optimization. In: International Conference on Machine Learning, pp. 2351–2360 (2015)
33. Sun, R., Luo, Z.-Q.: Guaranteed matrix completion via non-convex factorization. IEEE Trans. Inf. Theory **62**(11), 6535–6579 (2016)
34. Toh, K.-C., Todd, M.J., Tütüncü, R.H.: SDPT3: a MAT-LAB software package for semidefinite programming, version 1.3. Optim. Methods Softw. **11**(1–4), 545–581 (1999)
35. Tütüncü, R.H., Toh, K.-C., Todd, M.J.: Solving semidefinite–quadratic–linear programs using SDPT3. Math. Program. **95**(2), 189–217 (2003)
36. Vaswani, N., Nayer, S., Eldar, Yonina C.: Low-rank phase retrieval. IEEE Trans. Signal Process. **65**(15), 4059–4074 (2017)
37. Yun, C., Sra, S., Jadbabaie, A.: Global optimality conditions for deep neural networks. In: International Conference on Learning Representations (2018)
38. Zhang, R.Y., Sojoudi, S., Lavaei, J.: Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery. J. Mach. Learn. Res. **20**(114), 1–34 (2019)
39. Zhang, R., et al.: How much restricted isometry is needed in nonconvex matrix recovery? In: Advances in Neural Information Processing Systems, p. 55915602 (2018)
40. Zhang, Y., Madani, R., Lavaei, J.: Conic relaxations for power system state estimation with line measurements. IEEE Trans. Control Netw. Syst. **5**(3), 1193–1205 (2018)
41. Zhao, T., Wang, Z., Liu, H.: A nonconvex optimization framework for low rank matrix estimation. In: Advances in Neural Information Processing Systems, pp. 559–567 (2015)
42. Zheng, Q., Lafferty, J.: A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. In: Advances in Neural Information Processing Systems, pp. 109–117 (2015)
43. Zhu, Zhihui, et al.: Global optimality in low-rank matrix optimization. IEEE Transactions on Signal Processing **66**(13), 3614–3628 (2018)

44. Zimmerman, R.D., Murillo-Sanchez, C.E., Thomas, R.J.: MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans. Power Syst. **26**(1), 12–19 (2011)