
A Nullspace Property for Subspace-Preserving Recovery

Mustafa D. Kaba¹ Chong You² Daniel P. Robinson³ Enrique Mallada⁴ René Vidal⁵

Abstract

Much of the theory for *classical sparse recovery* is based on conditions on the dictionary that are both necessary and sufficient (e.g., nullspace property) or only sufficient (e.g., incoherence and restricted isometry). In contrast, much of the theory for *subspace-preserving recovery*, the theoretical underpinnings for sparse subspace classification and clustering methods, is based on conditions on the subspaces and the data that are only sufficient (e.g., subspace incoherence and data inner-radius). This paper derives a *necessary and sufficient* condition for subspace-preserving recovery that is inspired by the classical nullspace property. Based on this novel condition, called here the *subspace nullspace property*, we derive equivalent characterizations that either admit a clear geometric interpretation that relates data distribution and subspace separation to the recovery success, or can be verified using a finite set of extreme points of a properly defined set. We further exploit these characterizations to derive new sufficient conditions, based on inner-radius and outer-radius measures and dual bounds, that generalize existing conditions and preserve the geometric interpretations. These results fill an important gap in the subspace-preserving recovery literature.

1. Introduction

Many machine learning problems involve the analysis of high-dimensional data whose intrinsic dimension is much lower than the ambient dimension. When such data comes from multiple classes, it can be well approximated by a union of low-dimensional subspaces, where each subspace

corresponds to one class. This has motivated the development of many supervised and unsupervised methods for learning a union of subspaces, a.k.a. the subspace classification and subspace clustering tasks (Vidal et al., 2016).

Among existing methods for subspace classification and clustering, sparse representation based methods (Wright et al., 2009; Elhamifar & Vidal, 2009; Soltanolkotabi & Candès, 2012; Soltanolkotabi et al., 2014; You & Vidal, 2015; Robinson et al., 2019) have drawn much attention due to their simplicity, broad theoretical guarantees and superior empirical performance. Such methods are based on expressing each data point as a linear combination of the other data points while enforcing ℓ_1 regularization on the representation coefficients. The idea is that the nonzero entries of the coefficient vector should correspond to data points that lie in the same subspace as the point being represented. Coefficient vectors that have that property are referred to as *subspace-preserving* (Vidal et al., 2016), a notion that plays a central role in establishing the correctness of sparse subspace classification and clustering methods. In particular, the recovery of subspace-preserving solutions is provably guaranteed when the subspaces are independent (Elhamifar & Vidal, 2009), disjoint (Elhamifar & Vidal, 2010) and non-trivially intersecting (Soltanolkotabi & Candès, 2012; Soltanolkotabi et al., 2014; You & Vidal, 2015). Recently, such guarantees have been further extended to a broader range of applications that deal with dimension-reduced data (Wang et al., 2015), corrupted data (Soltanolkotabi & Candès, 2012; Wang & Xu, 2016; You et al., 2017), data with missing entries (Tsakiris & Vidal, 2018), and affine subspaces (Li et al., 2018; You et al., 2019).

Most of the existing theoretical conditions for subspace-preserving recovery characterize the geometry of the union of subspaces (e.g., via subspace independence (Elhamifar & Vidal, 2009; Lu et al., 2012) and subspace incoherence (Soltanolkotabi & Candès, 2012; You & Vidal, 2015)) and the distribution of points in the subspaces (e.g., via inner-radius and circumradius (Soltanolkotabi & Candès, 2012; You & Vidal, 2015)). Such geometric conditions have clear geometric interpretations and provide an important understanding of the regimes in which sparse methods are applicable. However, such conditions are only sufficient and a tight characterization of subspace-preserving recovery has been missing in the existing literature.

¹eBay Inc., San Jose, CA, USA. ²Dept. of Elect. Eng. & Comp. Sci., University of California at Berkeley, Berkeley, CA, USA. ³Dept. of Ind. & Sys. Eng., Lehigh University, Bethlehem, PA, USA. ⁴Dept. of Elect. & Comp. Eng., Johns Hopkins University, Baltimore, MD, USA. ⁵MINDS & Dept. of Biom. Eng., Johns Hopkins University, Baltimore, MD, USA.. Correspondence to: Mustafa D. Kaba <mkaba@ebay.com>.

In this paper, we derive *necessary and sufficient* conditions for subspace-preserving recovery. Our conditions are inspired by the classical Nullspace Property (NSP) (Foucart & Rauhut, 2013, Defn. 4.1)¹, which has been widely used in classical sparse recovery, not only as a tool to obtain results of theoretical importance, but also as a powerful tool to obtain stable (Foucart & Rauhut, 2013, Thm. 4.14) and robust (Foucart & Rauhut, 2013, Thm. 4.19) recovery guarantees. However, the NSP appears not to have been explored in the subspace classification and clustering literature.² Arguably, one of the reasons is that necessary and sufficient conditions such as the NSP are often difficult to verify. This paper not only introduces an NSP that is successfully tailored for subspace-preserving recovery, but also derives equivalent characterizations of it based on the set of extreme points that reduce the verification of the condition to a finite set. Our primary goal is to present an in-depth theoretical analysis of subspace-preserving recovery through a nullspace-property-like condition, and not necessarily to construct computationally efficient tools. Hence, the conditions we derive can sometimes be computationally hard to verify, as is often the case with other conditions in the literature (Elhamifar & Vidal, 2009; Lu et al., 2012; Soltanolkotabi & Candès, 2012; You & Vidal, 2015).

More specifically, the contributions of this paper can be summarized as follows:

- We derive a necessary and sufficient condition for subspace-preserving recovery (Thm. 1) that we call the subspace nullspace property (SNSP). We show that the SNSP holds if and only if it holds on a certain subset of the columns of the data matrix \mathbf{X} , which has the potential of significantly facilitating the verification of the SNSP for the subspace classification task (Lemma 1).
- We provide a characterization of the SNSP based on a comparison of the symmetrized convex hull of the data points on a subspace and the symmetrized convex hull of the remaining data points, which provides a very clear geometric interpretation for the SNSP (Thm. 2). We derive a sufficient condition from this characterization, which is also geometric in nature and simplifies the decision to a comparison between the inner-radius and outer-radius of two compact convex sets (Cor. 2).
- By exploiting convexity of the recovery problem in the

¹Since the NSP is a necessary and sufficient condition, sufficient conditions like Restricted Isometry Property implies the NSP (Foucart & Rauhut, 2013, Sect. 6.2).

²As far as we know, (Robinson et al., 2019) is the only prior work that discusses a necessary and sufficient condition for subspace-preserving recovery and establishes a connection between their condition (based on dual certificate analysis) and the NSP. However, the exposition in (Robinson et al., 2019) is limited to a comment and an in-depth analysis is missing in the literature.

primal space, we provide a novel characterization of the SNSP that reduces its verification to a finite subset of the extreme points of the intersection of the nullspace of \mathbf{X} and the ℓ_1 -ball (Thm. 3). This characterization requires solving an ℓ_1 -minimization problem for each extreme point. By replacing the ℓ_1 -minimization problem by its dual and exploiting convexity in the dual space, we formulate a characterization of the SNSP that can be verified by a decision on a finite set only (Thm. 5).

- We introduce bounds on the dual of the ℓ_1 -minimization problem when the columns of \mathbf{X} are assumed to have unit ℓ_p -norm (Prop. 1). These bounds allow us to simplify the aforementioned characterization of the SNSP based on extreme points to obtain a sufficient condition for subspace-preserving recovery that is simpler to check (Thm. 6).

2. Preliminaries and Problem Formulation

2.1. Notation and Preliminaries

The set of integers from 1 up to N is denoted as $[N] := \{1, \dots, N\}$. For any $\mathbf{c} \in \mathbb{R}^N$, the support of \mathbf{c} is denoted as $\text{Supp}(\mathbf{c}) := \{k \in [N] : c_k \neq 0\}$. The vector \mathbf{c} is called s -sparse if $|\text{Supp}(\mathbf{c})| \leq s$. For any index set $S \subseteq [N]$, the complement of S in $[N]$ is denoted by S^c . For a nonempty set $S \subseteq [N]$, the vector $\mathbf{c}_S \in \mathbb{R}^{|S|}$ denotes the part of \mathbf{c} that is supported on S . We use $\text{Pr}_S \in \mathbb{R}^{N \times N}$ to denote the matrix that projects onto the coordinates in S and sets all other coordinates to zero. For a matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ and an index set $S \subseteq [N]$, the matrix $\mathbf{X}_S \in \mathbb{R}^{D \times |S|}$ denotes the submatrix of \mathbf{X} consisting of the columns of \mathbf{X} indexed by S . Therefore, for all $\mathbf{c} \in \mathbb{R}^N$, we have $\mathbf{X} \text{Pr}_S \mathbf{c} = \mathbf{X}_S \mathbf{c}_S$. If $S = \{j\}$ for some j , we write \mathbf{x}_j instead of \mathbf{X}_S to refer to the j^{th} -column of \mathbf{X} . We prioritize the subscript over superscript so that $\mathbf{X}_S^\top \equiv (\mathbf{X}_S)^\top$, and *not* $(\mathbf{X}^\top)_S$. Finally, $\text{Null}(\mathbf{X})$ denotes the nullspace of the matrix \mathbf{X} , and $\mathbf{X}^{-1}(\cdot)$ denotes the inverse image under \mathbf{X} .

The ℓ_p -norm of a vector $\mathbf{x} \in \mathbb{R}^D$ is defined as $\|\mathbf{x}\|_p := (\sum_{k=1}^D |x_k|^p)^{\frac{1}{p}}$, where $|\cdot|$ denotes the absolute value. The unit ℓ_p -sphere is denoted by $\mathbb{S}_p^{D-1} := \{\mathbf{x} \in \mathbb{R}^D : \|\mathbf{x}\|_p = 1\}$ and the unit ℓ_p -ball by $\mathbb{B}_p^D := \{\mathbf{x} \in \mathbb{R}^D : \|\mathbf{x}\|_p \leq 1\}$.

The convex hull is denoted by $\text{conv}(\cdot)$. We denote the convex hull of the union of the columns of \mathbf{X} and $-\mathbf{X}$ by $\mathcal{K}(\mathbf{X})$. Sometimes we refer to it as the *symmetrized convex hull* of the columns of \mathbf{X} .

For a nonempty convex set $\mathcal{C} \subseteq \mathbb{R}^D$, the set of extreme points of \mathcal{C} is denoted as $\text{Ext}(\mathcal{C})$. These are precisely the points that cannot be written as a non-trivial convex combination of two distinct points in \mathcal{C} . The *affine hull* of \mathcal{C} , denoted by $\text{aff}(\mathcal{C})$, is the smallest affine set in \mathbb{R}^D that contains \mathcal{C} . The *relative interior*

(Rockafellar, 1970, p.44) of \mathcal{C} is defined as $\text{rinte}(\mathcal{C}) := \{\mathbf{x} \in \text{aff}(\mathcal{C}) : \exists \varepsilon > 0, (\mathbf{x} + \varepsilon \mathbb{B}_2^D) \cap \text{aff}(\mathcal{C}) \subseteq \mathcal{C}\}$. The *po-lar* (Rockafellar, 1970, p.125) of \mathcal{C} is defined as $\mathcal{C}^\circ := \{\mathbf{q} \in \mathbb{R}^D : \mathbf{q}^\top \mathbf{x} \leq 1 \text{ for all } \mathbf{x} \in \mathcal{C}\}$. Note that \mathcal{C}° is always a closed, convex set (Rockafellar, 1970, p.125).

We define the *inner- ℓ_p -radius* of a nonempty compact convex set $\mathcal{C} \subseteq \mathbb{R}^D$ containing the origin as the radius of the largest ℓ_p -ball (confined to the linear span of \mathcal{C}) one can inscribe inside \mathcal{C} , and denote it by $\tau_p(\mathcal{C})$. That is, $\tau_p(\mathcal{C}) := \max\{\alpha \in \mathbb{R}_{>0} : \alpha(\mathbb{B}_p^D \cap \text{span}(\mathcal{C})) \subseteq \mathcal{C}\}$, where $\text{span}(\mathcal{C})$ denotes the subspace spanned by \mathcal{C} and $\mathbb{R}_{>0}$ denotes the set of positive real numbers. Likewise, we define the *outer- ℓ_p -radius* of \mathcal{C} as the radius of the smallest ℓ_p -ball that contains \mathcal{C} , and denote it by $\mathfrak{R}_p(\mathcal{C})$. That is, $\mathfrak{R}_p(\mathcal{C}) := \min\{\beta \in \mathbb{R}_{>0} : \beta \mathbb{B}_p^D \supseteq \mathcal{C}\}$.

2.2. Sparse Subspace Classification and Clustering

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ be a matrix with nonzero columns drawn from a union of n subspaces $\bigcup_{i=1}^n \mathcal{S}_i \subset \mathbb{R}^D$ of dimension $d_i = \dim(\mathcal{S}_i)$. Let $\mathcal{P} := \{\mathcal{P}_i\}_{i=1}^n$ be a partitioning of $[N]$ defined according to the memberships of the columns of \mathbf{X} to the subspaces, i.e., $\mathcal{P}_i := \{k \in [N] : \mathbf{x}_k \in \mathcal{S}_i\}$.³ Assuming \mathcal{P} is *known*, the goal of *subspace classification* is to assign a new nonzero point $\mathbf{y} \in \mathcal{S}_i$ for some $i \in [n]$ to the subspace \mathcal{S}_i that it belongs.

We assume that the columns of $\mathbf{X}_{\mathcal{P}_i}$ span \mathcal{S}_i , hence any point $\mathbf{y} \in \mathcal{S}_i$ can be expressed as a linear combination of the columns of $\mathbf{X}_{\mathcal{P}_i}$, i.e., there exists a $\mathbf{c} \in \mathbb{R}^N$ with support in \mathcal{P}_i such that $\mathbf{y} = \mathbf{X}\mathbf{c}$. Given $\mathbf{y} \in \mathcal{S}_i$, a solution \mathbf{c} to the system $\mathbf{y} = \mathbf{X}\mathbf{c}$ is called *subspace-preserving* if and only if $\text{Supp}(\mathbf{c}) \subseteq \mathcal{P}_i$. Note that such a vector \mathbf{c} need not be unique, since in general the number of columns of $\mathbf{X}_{\mathcal{P}_i}$ will be larger than the dimension of \mathcal{S}_i . Nonetheless, assuming that the largest subspace dimension $\bar{d} = \max_i d_i$ is small relative to N , it follows that all subspace-preserving vectors \mathbf{c} must be \bar{d} -sparse. This motivated (Wright et al., 2009) to tackle the subspace classification problem by solving the basis pursuit problem

$$\min_{\mathbf{c} \in \mathbb{R}^N} \|\mathbf{c}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{X}\mathbf{c}, \quad (1)$$

where the ℓ_1 norm is used as a proxy for the sparsity of \mathbf{c} . To understand the correctness of this approach, we consider the following theoretical question concerning subspace-preserving recovery:

(Q) What are the necessary and sufficient conditions on the union of subspaces $\bigcup_{i=1}^n \mathcal{S}_i$ and the data \mathbf{X} such that for all $i \in [n]$ and all $\mathbf{y} \in \mathcal{S}_i$, all solutions to (1) are subspace-preserving?

³We implicitly assume that no column of \mathbf{X} lies at the intersection of two distinct subspaces from the union, which is not a major assumption for data coming from real applications.

The answer to **(Q)** provides necessary and sufficient conditions under which any test data point $\mathbf{y} \in \bigcup_{i=1}^n \mathcal{S}_i$ is correctly classified by assigning it to the subspace determined by $\text{Supp}(\mathbf{c})$.⁴

Closely related to subspace classification is the problem of *subspace clustering*, where we assume that the subspaces and the membership of the columns of \mathbf{X} to their respective subspaces $\{\mathcal{S}_i\}$ are *unknown* and the goal is to segment the columns of \mathbf{X} into their respective subspaces, i.e., to find \mathcal{P} . The work of (Elhamifar & Vidal, 2013) addressed the subspace clustering problem by solving a modified basis pursuit problem where each data point is expressed as a combination of “other” data points. More precisely, for each $j \in [N]$, the authors propose to solve

$$\min_{\mathbf{c} \in \mathbb{R}^{N-1}} \|\mathbf{c}\|_1 \text{ s.t. } \mathbf{x}_j = \mathbf{X}_{-j}\mathbf{c},$$

where \mathbf{X}_{-j} denotes the submatrix of \mathbf{X} consisting of all but the j^{th} column. For each j , the solution of this optimization problem provides a sparse representation of the data point \mathbf{x}_j in terms of the other data points, which is then used to cluster the data. The conditions under which such a representation is subspace-preserving can help guarantee the correct segmentation of the data points. Hence, finding an answer to **(Q)** has important implications in subspace clustering setting. In this paper, we focus on answering **(Q)** for the subspace classification problem, and refer the reader to (You & Vidal, 2015; Robinson et al., 2019) for a detailed discussion of the implications of such answers in the subspace clustering setting.

We end this section by a comparison of the subspace classification problem and the recovery of the block sparse signals. Over the years researchers studied the different aspects of this problem in detail. For instance, there is work on exact recovery of block sparse signals (Jenatton et al., 2011; Bajwa et al., 2015) and the recovery of the support of the unknown block sparse signal (Obozinski et al., 2011; Kolar et al., 2011). A nullspace characterization of the recovery of block sparse signals is also available (Stojnic et al., 2009). However, the existing work on block sparse signals differs from the problem we study here in one major way: We are neither interested in the correct recovery of the support nor the correct recovery of the signal itself. Even if the minimizer of (1) has a support different than the original signal, the recovery attempt can be deemed successful in the setting we study. To the best of our knowledge, the problem **(Q)** was never studied in the literature addressing the recovery of block sparse signals.

⁴Note that if $\mathbf{y} \equiv \mathbf{0}$, then the solution to (1) is $\bar{\mathbf{c}} \equiv \mathbf{0}$. Therefore, the support of $\bar{\mathbf{c}}$ is the empty set, and we have $\emptyset = \text{Supp}(\bar{\mathbf{c}}) \subseteq \mathcal{P}_i$ for all $i \in [n]$. That is, when $\mathbf{y} \equiv \mathbf{0}$, the solution to (1) is always subspace-preserving. In particular, $\mathbf{y} \equiv \mathbf{0}$ is assigned to all subspaces, which is in accordance with the fact that all subspaces contain the origin.

3. A Nullspace Property for Subspace-preserving Recovery

This section provides an initial answer to **(Q)** based on a necessary and sufficient condition that is inspired by the classical NSP for sparse recovery. We call this condition the *Subspace Nullspace Property* (SNSP) in homage to its ties with subspace classification and clustering, and the NSP.

Definition 1 (SNSP). *We denote the set of vectors in the nullspace of \mathbf{X} whose support is not contained in any single element of the partition \mathcal{P} by $\text{Null}(\mathbf{X}, \mathcal{P})$. That is,*

$$\text{Null}(\mathbf{X}, \mathcal{P}) := \{\eta \in \text{Null}(\mathbf{X}) : \text{Supp}(\eta) \not\subseteq P, \forall P \in \mathcal{P}\}. \quad (2)$$

We say that \mathbf{X} satisfies the SNSP if and only if for all $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$ and $P \in \mathcal{P}$, we have

$$\min_{\mathbf{z}: \mathbf{X}_P \mathbf{z} = \mathbf{X}_P \eta_P} \|\mathbf{z}\|_1 < \|\eta_{P^c}\|_1. \quad (3)$$

We find it informative to briefly discuss how we interpret the SNSP. It tells us that the vectors that are relevant in characterizing the subspace-preserving recovery are those vectors in $\text{Null}(\mathbf{X})$ whose supports are not contained within a single element of the partition \mathcal{P} . This is the first level of attention SNSP introduces. Then, (3) further rules out certain vectors in $\text{Null}(\mathbf{X}, \mathcal{P})$: For any $\bar{\eta} \in \text{Null}(\mathbf{X}, \mathcal{P})$, if there is $\hat{\eta} \in \text{Null}(\mathbf{X}, \mathcal{P})$ with $\bar{\eta}_{P^c} = \hat{\eta}_{P^c}$ and $\|\mathbf{X}_P \hat{\eta}_P\|_1 < \|\mathbf{X}_P \bar{\eta}_P\|_1$, then it is enough to check (3) for $\hat{\eta}$ only. Hence, the variation in $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$ introduced by the nullspace of \mathbf{X}_P is irrelevant, unless it is reducing the ℓ_1 -norm of η_P .

When \mathbf{X}_P has full column rank, the only solution to $\mathbf{X}_P \mathbf{z} = \mathbf{X}_P \eta_P$ is $\mathbf{z} = \eta_P$, and the condition in (3) becomes $\|\eta_P\|_1 < \|\eta_{P^c}\|_1$. Such a condition is closely related to the classical NSP of order s , which requires $\|\eta_P\|_1 < \|\eta_{P^c}\|_1$ to hold for all s -sparse patterns, i.e. for all P with $|P| \leq s$. Therefore, the SNSP differs from the NSP in two directions. First, it does not require the columns of \mathbf{X}_P to be linearly independent, thus allowing for non-unique sparse solutions. This makes sense for subspace classification and clustering, where the goal is to determine the correct subspace rather than a unique sparse representation, hence the specific support of the solution is irrelevant as long as it gives the correct subspace. Second, the SNSP is verified over the partitioning \mathcal{P} of the data according to the membership to the subspaces, while the NSP is verified with respect to all s -sparse patterns.

The next result shows that SNSP gives an answer to **(Q)**.⁵

Theorem 1. *The solution to (1) is subspace-preserving for all $i \in [n]$ and $\mathbf{y} \in \mathcal{S}_i$ if and only if \mathbf{X} satisfies the SNSP. That is, for all $i \in [n]$ and for all $\bar{\mathbf{c}} \in \mathbb{R}^N$ supported in \mathcal{P}_i ,*

the support of any minimizer of

$$\min_{\mathbf{c}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{c}} \|\mathbf{c}\|_1 \quad (4)$$

is contained in \mathcal{P}_i if and only if \mathbf{X} satisfies the SNSP.

Proof. \Rightarrow : Let $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$ and $P \in \mathcal{P}$. Since $\mathbf{0} = \mathbf{X} \eta = \mathbf{X} (\text{Pr}_P(\eta) + \text{Pr}_{P^c}(\eta))$, we have

$$\mathbf{X} \text{Pr}_P(\eta) = \mathbf{X} \text{Pr}_{P^c}(-\eta). \quad (5)$$

Note that $\text{Supp}(\text{Pr}_P(\eta)) \subseteq P \in \mathcal{P}$. Then, by the hypothesis, the support of any minimizer of

$$\min_{\mathbf{z}: \mathbf{X} \text{Pr}_P(\eta) = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 \quad (6)$$

is a subset of P . Therefore,

$$\min_{\mathbf{c}: \mathbf{X}_P(\eta_P) = \mathbf{X}_P(\mathbf{c})} \|\mathbf{c}\|_1 = \min_{\mathbf{z}: \mathbf{X} \text{Pr}_P(\eta) = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1. \quad (7)$$

By (5), $\text{Pr}_{P^c}(-\eta)$ is feasible for (6). Moreover, $\text{Supp}(\text{Pr}_{P^c}(-\eta)) \neq \emptyset$, because $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$. Hence, $\text{Supp}(\text{Pr}_{P^c}(-\eta)) \not\subseteq P$. Therefore, $\text{Pr}_{P^c}(-\eta)$ cannot be a minimizer of (6). So,

$$\min_{\mathbf{z}: \mathbf{X} \text{Pr}_P(\eta) = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 < \|\text{Pr}_{P^c}(-\eta)\|_1 = \|\eta_{P^c}\|_1.$$

Combining with (7), we get $\min_{\mathbf{c}: \mathbf{X}_P(\eta_P) = \mathbf{X}_P(\mathbf{c})} \|\mathbf{c}\|_1 < \|\eta_{P^c}\|_1$. Hence, \mathbf{X} satisfies the SNSP.

\Leftarrow : Suppose that \mathbf{X} satisfies the SNSP. Let $\bar{\mathbf{c}} \in \mathbb{R}^N$ with $\text{Supp}(\bar{\mathbf{c}}) \subseteq P$ for some $P \in \mathcal{P}$. Suppose on the contrary that (4) has a minimizer $\hat{\mathbf{c}}$ with $\text{Supp}(\hat{\mathbf{c}}) \setminus P \neq \emptyset$. Let $\eta := \bar{\mathbf{c}} - \hat{\mathbf{c}}$ and note that $\eta \in \text{Null}(\mathbf{X})$ and $\text{Supp}(\eta) \setminus P \neq \emptyset$. If there exists $Q \in \mathcal{P}$ with $\text{Supp}(\eta) \subseteq Q$, then necessarily $\bar{\mathbf{c}}_P = \hat{\mathbf{c}}_P$, and so $\|\bar{\mathbf{c}}\|_1 = \|\hat{\mathbf{c}}_P\|_1 = \|\hat{\mathbf{c}}\|_1 < \|\hat{\mathbf{c}}\|_1$. This contradicts $\hat{\mathbf{c}}$ being a minimizer of (4). Hence, we can assume that $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$, so that

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 = \min_{\mathbf{z}: \mathbf{X} \text{Pr}_P(\bar{\mathbf{c}}) = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1,$$

since $P \supseteq \text{Supp}(\bar{\mathbf{c}})$. Setting $\mathbf{y} := \mathbf{z} - \text{Pr}_P(\hat{\mathbf{c}})$, we obtain

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 = \min_{\mathbf{y}: \mathbf{X} \text{Pr}_P(\bar{\mathbf{c}} - \hat{\mathbf{c}}) = \mathbf{X} \mathbf{y}} \|\mathbf{y} + \text{Pr}_P(\hat{\mathbf{c}})\|_1.$$

Then, by the triangle inequality,

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 \leq \min_{\mathbf{y}: \mathbf{X}_P(\eta_P) = \mathbf{X} \mathbf{y}} \|\mathbf{y}\|_1 + \|\hat{\mathbf{c}}_P\|_1.$$

Shrinking the constraint set gives us

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 \leq \min_{\mathbf{c}: \mathbf{X}_P(\eta_P) = \mathbf{X}_P(\mathbf{c})} \|\mathbf{c}\|_1 + \|\hat{\mathbf{c}}_P\|_1.$$

The SNSP implies

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 < \|\eta_{P^c}\|_1 + \|\hat{\mathbf{c}}_P\|_1.$$

Since $P \supseteq \text{Supp}(\bar{\mathbf{c}})$, we finally get

$$\min_{\mathbf{z}: \mathbf{X} \bar{\mathbf{c}} = \mathbf{X} \mathbf{z}} \|\mathbf{z}\|_1 < \|-\hat{\mathbf{c}}_{P^c}\|_1 + \|\hat{\mathbf{c}}_P\|_1 = \|\hat{\mathbf{c}}\|_1$$

which is again a contradiction, since $\hat{\mathbf{c}}$ is a minimizer of (4). So, we must have $\text{Supp}(\hat{\mathbf{c}}) \subseteq P$. \square

⁵All missing proofs can be found in the supplementary material.

3.1. Verification of the SNSP on a Submatrix

One can work with a submatrix of \mathbf{X} and still verify the SNSP for \mathbf{X} . Specifically, one can discard the data points from subspace \mathcal{S}_i that do not correspond to extreme points of the symmetrized convex hull of the points in \mathcal{S}_i , and verify the SNSP for the matrix defined by the remaining points. Formally, recall that for each $\mathcal{P}_i \in \mathcal{P}$, the convex set $\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})$ is the convex hull of the union of the columns of $\mathbf{X}_{\mathcal{P}_i}$ and $-\mathbf{X}_{\mathcal{P}_i}$. Let $\tilde{\mathcal{P}}_i := \{l \in \mathcal{P}_i : \mathbf{x}_l \in \text{Ext}(\mathcal{K}(\mathbf{X}_{\mathcal{P}_i}))\}$ be the collection of indices associated with columns of \mathbf{X} that are extreme points of $\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})$, which must be nonempty. Let $\tilde{\mathcal{P}} := \{\tilde{\mathcal{P}}_i\}$ and $\tilde{\mathbf{X}} := \mathbf{X}_{\tilde{\mathcal{P}}}$ so that $\tilde{\mathbf{X}}$ consists of all columns of \mathbf{X} whose indices are in $\tilde{\mathcal{P}}$. The SNSP can now be verified on $\tilde{\mathbf{X}}$.

Lemma 1. *The data matrix \mathbf{X} satisfies the SNSP if and only if $\tilde{\mathbf{X}}$ (see the previous paragraph) satisfies the SNSP.*

Although Lemma 1 shows that one can safely replace the data matrix \mathbf{X} with its submatrix $\tilde{\mathbf{X}}$ for the purpose of verifying the SNSP, in this paper we do not assume that this substitution takes place. The reason for this is two fold: 1) There is no major difference for our theoretical development. 2) If the data is normalized with the ℓ_p -norm where $p \notin \{1, \infty\}$, then no reduction takes place since every column of the data matrix \mathbf{X} is an extreme point in this case. However, sparse subspace classification might benefit immensely from this reduction when the data set is large and either not normalized or normalized with the ℓ_1 or ℓ_∞ -norm. In this case, rather than verifying the SNSP for the data matrix \mathbf{X} , we can verify it on the much smaller submatrix of extreme points. Such a reduction in matrix size can speed up ℓ_1 -recovery significantly.

Since the SNSP characterizes when subspace-preserving recovery is possible, we devote the remainder of the paper to the derivation of alternative characterizations of the SNSP. These characterizations are either geometrically interpretable (thus providing insight to the nature of subspace-preserving recovery) or verifiable by considering certain *finite* sets of points (thus opening the door to developing practical approaches for certifying when the SNSP holds).

4. A Geometrically Interpretable Characterization of SNSP

In this section we introduce a characterization of the SNSP that allows for a clear geometric interpretation. Our main result in this section reads as follows.

Theorem 2. *The matrix \mathbf{X} satisfies the SNSP if and only if for all $i \in [n]$, we have*

$$\mathcal{S}_i \cap \mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c}) \subseteq \text{rinte}(\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})). \quad (8)$$

Proof. For any $\eta \in \text{Null}(\mathbf{X})$ it holds that $\mathbf{X}_P(\eta_P) =$

$-\mathbf{X}_{P^c}(\eta_{P^c})$. Combining this with Defn. 1, we have that \mathbf{X} satisfies the SNSP if and only if for all $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$ and for all $P \in \mathcal{P}$, we have

$$z : \mathbf{X}_{P^c} \left(-\frac{\eta_{P^c}}{\|\eta_{P^c}\|_1} \right) = \mathbf{X}_P(z) \quad \|z\|_1 < 1. \quad (9)$$

Since the optimization problem in (9) is feasible, we can replace it with its dual to obtain

$$\max_{w \in (\mathbf{X}_P^\top)^{-1} \mathbb{B}_\infty^{|\mathcal{P}|}} -\frac{1}{\|\eta_{P^c}\|_1} \eta_{P^c}^\top \mathbf{X}_{P^c}^\top w < 1. \quad (10)$$

That is, \mathbf{X} satisfies the SNSP if and only if for all $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$ and for all $P \in \mathcal{P}$, we have

$$\begin{aligned} \mathbf{X}_{P^c} \left(\frac{\eta_{P^c}}{\|\eta_{P^c}\|_1} \right) &\in \text{rinte} \left[\left((\mathbf{X}_P^\top)^{-1} \mathbb{B}_\infty^{|\mathcal{P}|} \right)^\circ \right] \\ &= \text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right). \end{aligned}$$

where the inclusion follows from the definition of the polar operator, and the equality follows from (Rockafellar, 1970, Cor. 16.3.2). After defining the set

$$\Lambda_P := \{ \mathbf{X}_{P^c}(\eta_{P^c}) : \eta \in \text{Null}(\mathbf{X}, \mathcal{P}) \text{ and } \|\eta_{P^c}\|_1 = 1 \},$$

we conclude that \mathbf{X} satisfies the SNSP if and only if for all $P \in \mathcal{P}$, we have $\Lambda_P \subseteq \text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right)$.

Next, we claim that for all $P \in \mathcal{P}$, the following holds

$$\Lambda_P \subseteq \text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{S}_1^{|\mathcal{P}^c|-1} \subseteq \Lambda_P \cup \{0\}. \quad (11)$$

The first inclusion is straightforward. In order to show why the second inclusion also holds, we let $z \in \text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{S}_1^{|\mathcal{P}^c|-1}$. Hence, there exists $\eta \in \text{Null}(\mathbf{X})$ such that $z = \mathbf{X}_P(-\eta_P) = \mathbf{X}_{P^c}(\eta_{P^c})$, with $\|\eta_{P^c}\|_1 = 1$. If $\eta \in \text{Null}(\mathbf{X}, \mathcal{P})$, then $z \in \Lambda_P$, and we are done. If not, there exists $Q \in \mathcal{P}$, with $Q \neq P$, such that $\text{Supp}(\eta) \subseteq Q$. But if this is the case, we necessarily have $z = \mathbf{X}_P(-\eta_P) = 0$. Hence, the second inclusion holds.

Since we always have $0 \in \text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right)$, by (11) we conclude that \mathbf{X} satisfies the SNSP if and only if for all $P \in \mathcal{P}$, we have $\text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{S}_1^{|\mathcal{P}^c|-1} \subseteq \text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right)$. The set $\text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right)$ is convex. The convex hull of $\{0\} \cup \left(\text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{S}_1^{|\mathcal{P}^c|-1} \right)$ is $\text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{B}_1^{|\mathcal{P}^c|}$. Hence, we conclude that \mathbf{X} satisfies the SNSP, if and only if, for all $P \in \mathcal{P}$, we have $\text{im}(\mathbf{X}_P) \cap \mathbf{X}_{P^c} \mathbb{B}_1^{|\mathcal{P}^c|} \subseteq \text{rinte} \left(\mathbf{X}_P \mathbb{B}_1^{|\mathcal{P}|} \right)$. Since, for all $i \in [n]$, we have $\text{im}(\mathbf{X}_{\mathcal{P}_i}) = \mathcal{S}_i$, $\mathbf{X}_{\mathcal{P}_i^c} \mathbb{B}_1^{|\mathcal{P}_i^c|} = \mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c})$ and $\mathbf{X}_{\mathcal{P}_i} \mathbb{B}_1^{|\mathcal{P}_i|} = \mathcal{K}(\mathbf{X}_{\mathcal{P}_i})$, we obtain the result. \square

Thm. 2 has a straightforward geometric interpretation. It tells us that subspace-preserving recovery always succeeds, if and only if, for any $i \in [n]$, the intersection of the symmetrized convex hull of the columns of $\mathbf{X}_{\mathcal{P}_i^c}$ and the subspace \mathcal{S}_i , is contained in the relative interior of the symmetrized convex hull of the columns of $\mathbf{X}_{\mathcal{P}_i}$. Although, to the best of our knowledge, this theorem did not appear in the literature before, it is possible to find work that makes similar arguments (Elhamifar & Vidal, 2013, p.8). We refer the reader to Fig. 1 for an illustration of Thm. 2 and comparison with similar arguments existing in the literature.

Note that Thm. 2 does not assume that the columns of \mathbf{X} are normalized. Therefore, Thm. 2 suggests that an ℓ_p -normalization of the columns of \mathbf{X} may affect subspace-preserving recovery. Intuitively, one would like the right-hand-side of (8) to be as large as possible, which suggests that an ℓ_∞ -normalization of the columns of \mathbf{X} may be a good option. However, this normalization also enlarges the left-hand-side of (8), and predicting the total effect of this normalization can be non-trivial. In fact, an ℓ_∞ -normalization can sometimes be counter-productive (see Example 1 in the next section). On the other hand, an ℓ_2 -normalization has an advantage. Suppose that we are given two data matrices \mathbf{X} and $\hat{\mathbf{X}}$, and moreover that $\hat{\mathbf{X}} = \Phi \mathbf{X}$ for some orthogonal matrix $\Phi \in \mathbb{R}^{D \times D}$. (In special cases, Φ can be realized as a rotation matrix.) If we normalize \mathbf{X} and $\hat{\mathbf{X}}$ using the ℓ_p -norm for $p \neq 2$, then Thm. 2 tells us that it is possible that the normalized \mathbf{X} satisfies the SNSP, while the normalized $\hat{\mathbf{X}}$ does not. This is an undesirable effect since the two (unscaled) data sets differ only by a rotation. However, if we normalize with the ℓ_2 -norm, this problem does not occur, meaning that the decision on the normalized \mathbf{X} and normalized $\hat{\mathbf{X}}$ are the same.

Corollary 1. *Let $\Phi \in \mathbb{R}^{D \times D}$ be an orthogonal matrix. Let \mathbf{X}_{ℓ_2} and $(\Phi \mathbf{X})_{\ell_2}$ denote the matrices obtained after normalizing the columns of \mathbf{X} and $\Phi \mathbf{X}$ with the ℓ_2 -norm, respectively. Then, \mathbf{X}_{ℓ_2} satisfies the SNSP, if and only if, $(\Phi \mathbf{X})_{\ell_2}$ satisfies the SNSP.*

Proof. The result follows from Thm. 2 and the fact that the matrix Φ preserves the ℓ_2 -norm. \square

To the best of our knowledge Thm. 2 provides one of the most geometrically interpretable necessary and sufficient conditions for subspace-preserving recovery in the literature. However, in general, it can be difficult to determine whether a convex set is contained in another, so that the verification of (8) can be difficult. Therefore, sufficient conditions that reduce the computational burden can be valuable. With this motivation in mind, we introduce the next corollary, which reduces the verification of the SNSP to a comparison of two numbers: An inner-radius and an outer-radius.

Corollary 2. *Suppose that the columns of \mathbf{X} have unit ℓ_p -norm. The data matrix \mathbf{X} satisfies the SNSP if, for all $i \in [n]$, we have*

$$\mathfrak{R}_p(\mathcal{S}_i \cap \mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c})) < \mathfrak{r}_p(\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})), \quad (12)$$

i.e. if, for all $i \in [n]$, the outer- ℓ_p -radius of $\mathcal{S}_i \cap \mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c})$ is strictly smaller than the inner- ℓ_p -radius of $\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})$.

Proof. Since inequality (12) implies (8), the result immediately follows from Thm. 2. \square

We end this section by connecting Cor. 2 to existing geometric results for subspace-preserving recovery in the literature.

Relation with results for independent subspaces. It is shown in (Elhamifar & Vidal, 2009) that subspace-preserving recovery is guaranteed if the collection of subspaces is linearly independent. Here, we demonstrate that this result is implied by Cor. 2. Recall that a collection of subspaces $\{\mathcal{S}_k\}_{k=1}^n$ is called *independent* if the dimension of the span of $\{\mathcal{S}_k\}_{k=1}^n$ is equal to the sum of the dimensions of the individual subspaces \mathcal{S}_k , i.e. if we have $\dim(\bigoplus_{k=1}^n \mathcal{S}_k) = \sum_{k=1}^n \dim(\mathcal{S}_k)$. If $\{\mathcal{S}_k\}_{k=1}^n$ is a union of independent subspaces, then the left-hand-side of (12) is 0 for every $i \in [n]$, and so (12) is satisfied trivially.

Relation with results characterized by incoherence measures. Many of the existing geometric conditions for subspace-preserving recovery are characterized by inradius and certain *incoherence* metrics (Soltanolkotabi & Candès, 2012; You & Vidal, 2015; You et al., 2016). In particular, You & Vidal (2015) shows that under the assumption that the columns of \mathbf{X} have unit ℓ_2 -norm, subspace-preserving recovery is guaranteed if the following principal recovery condition is satisfied:

$$\mu(\mathcal{S}_i, \mathbf{X}_{\mathcal{P}_i^c}) < \mathfrak{r}_2(\mathcal{K}(\mathbf{X}_{\mathcal{P}_i})), \quad \forall i \in [n], \quad (13)$$

where $\mu(\mathcal{S}_i, \mathbf{X}_{\mathcal{P}_i^c})$ is the incoherence between the subspace \mathcal{S}_i and the points $\mathbf{X}_{\mathcal{P}_i^c}$ defined as $\mu(\mathcal{S}_i, \mathbf{X}_{\mathcal{P}_i^c}) = \max_{\mathbf{x} \in \mathcal{S}_i - \{0\}} \|\mathbf{X}_{\mathcal{P}_i^c}^\top \mathbf{x}\|_\infty / \|\mathbf{x}\|_2$. Note that the right hand side of (13) is the same as that of (12). On the other hand, the incoherence on the left hand side of (13) is related to the outer-radius on the left hand side of (12) as follows:

$$\begin{aligned} \mu(\mathcal{S}_i, \mathbf{X}_{\mathcal{P}_i^c}) &= \mathfrak{R}_2(\text{Pr}_{\mathcal{S}_i}(\mathbf{X}_{\mathcal{P}_i^c})) \\ &= \mathfrak{R}_2(\text{Pr}_{\mathcal{S}_i}(\mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c}))) \geq \mathfrak{R}_2(\mathcal{S}_i \cap \mathcal{K}(\mathbf{X}_{\mathcal{P}_i^c})), \end{aligned} \quad (14)$$

where $\text{Pr}_{\mathcal{S}_i}(\cdot)$ denotes projection onto subspace \mathcal{S}_i . Therefore, the condition in (13) implies (i.e., is more restrictive than) the condition in (12), showing that Cor. 2 provides a stronger result than the result in You & Vidal (2015).

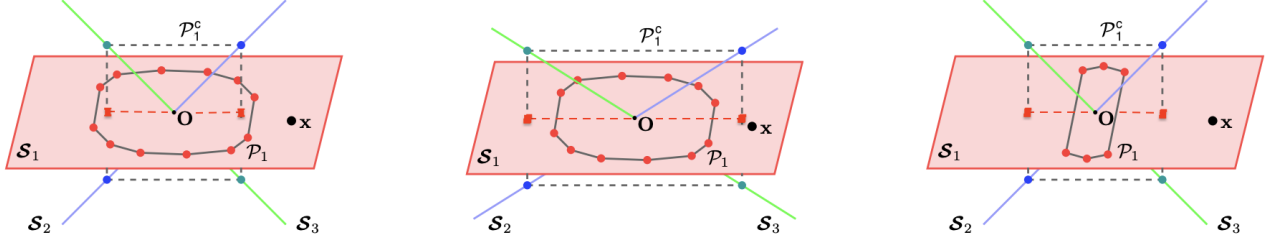


Figure 1. Elhamifar & Vidal (2013) provide an interpretation of their results on subspace-preserving recovery in terms of these drawings, when the collection of the subspaces is disjoint. For a point $x \in \mathcal{S}_1 \cap (\mathcal{S}_2 \oplus \mathcal{S}_3)$ they argue that subspace-preserving recovery in \mathcal{S}_1 succeeds in the case shown on the left because $\min\{\alpha : x \in \alpha\mathcal{K}(\mathbf{X}_{\mathcal{P}_1})\} < \min\{\alpha : x \in \alpha\mathcal{K}(\mathbf{X}_{\mathcal{P}_1^c})\}$. They argue that recovery fails in the two other cases (shown in the middle and the right) because this inequality is reversed. According to their interpretation, the failure is due to smaller angles between the subspaces (middle) and degenerate data distribution (right). Our interpretation of these cases based on Thm. 2 is as follows: We make no specific reference to points in $\mathcal{S}_1 \cap (\mathcal{S}_2 \oplus \mathcal{S}_3)$. We argue that subspace-preserving recovery in \mathcal{S}_1 succeeds in the case shown on the left because $\mathcal{S}_1 \cap \mathcal{K}(\mathbf{X}_{\mathcal{P}_1^c}) \subset \text{rinte}(\mathcal{K}(\mathbf{X}_{\mathcal{P}_1}))$, whereas it fails in the next two cases because this strict inclusion fails to hold. (Drawings are borrowed from (Elhamifar & Vidal, 2013) for comparison.)

5. Reduction of the Verification of SNSP to a Decision on Finite Sets

In the previous section we developed a characterization of the SNSP that admits a clear geometric interpretation. We also noted that condition (8) in Thm. 2 can be difficult to verify. In this section we introduce equivalent characterizations of SNSP that reduces the verification of SNSP to decisions on finite sets. These finite sets correspond to extreme points of certain compact convex sets. For instance, the next theorem shows that the SNSP admits a characterization in terms of the extreme points of $\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N$. Before stating the theorem, we introduce the following new notation:

$$\begin{aligned} & \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P}) \\ & := \{\eta \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N) : \text{Supp}(\eta) \not\subseteq P, \forall P \in \mathcal{P}\}. \end{aligned}$$

That is, $\text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$ is the collection of all extreme points of $\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N$ whose support is not contained in any $P \in \mathcal{P}$. The next result reduces the verification of the SNSP from $\text{Null}(\mathbf{X}, \mathcal{P})$ to $\text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$.

Theorem 3. *The matrix \mathbf{X} satisfies the SNSP if and only if, for all $\eta \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$ and all $P \in \mathcal{P}$, we have*

$$\|\eta_P\|_1 + \min_{\mathbf{X}_P(\eta_P) = \mathbf{X}_P(z)} \|z\|_1 < 1. \quad (15)$$

While Thm. 3 shows that the verification of SNSP can be reduced to a decision on a finite set, there is still an ℓ_1 -minimization problem that persists. We address this issue in the next couple of results.

Theorem 4. *The matrix \mathbf{X} satisfies the SNSP if and only if, for all $P \in \mathcal{P}$ and all $\eta \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$, we have*

$$\|\eta_P\|_1 + \max_{w \in \text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|}} \eta_P^\top w < 1. \quad (16)$$

Proof. Since there is zero duality gap for the ℓ_1 -minimization problem in (15), we can replace it by its dual, namely $\max_{w \in \text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|}} (\eta_P)^\top w$, which gives the desired result. \square

Now that we have converted the minimization problem to a maximization problem by using its dual, we can reduce the problem further to a decision on a finite set, as shown by the next result.

Theorem 5. *The matrix \mathbf{X} satisfies the SNSP if and only if, for all $\eta \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$ and all $P \in \mathcal{P}$, we have*

$$\|\eta_P\|_1 + \max_{w \in \text{Ext}(\text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|})} \eta_P^\top w < 1. \quad (17)$$

Proof. Since the optimization problem in (16) is the maximization of a linear objective function over a compact convex feasible set $\text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|}$, the optimal value must occur at an extreme point of $\text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|}$. This gives the desired result. \square

Thm. 5 is our final characterization of the SNSP. We stress that it does not require the solution of a *continuous* optimization problem or verification on an infinite set. With Thm. 5, the SNSP can be verified in a finite number of steps by checking (17) at all $(\eta, w) \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P}) \times \text{Ext}(\text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^{|\mathcal{P}|})$ and each $P \in \mathcal{P}$. We illustrate this approach with a simple example.

Example 1. *In this example, we consider the union of three subspaces $\{\mathcal{S}_i\}_{i=1}^3$ in \mathbb{R}^3 . We assume that \mathcal{S}_1 and \mathcal{S}_2 are lines through the origin spanned by $\mathbf{x}_1 := [0 \ 1 \ 1]^\top$ and $\mathbf{x}_2 := [0 \ 1 \ 1/2]^\top$, respectively, and \mathcal{S}_3 is the xy -plane.*

Furthermore, we assume that we have four points from this union $\{\mathbf{x}_i\}_{i=1}^4$ that are given by \mathbf{x}_1 and \mathbf{x}_2 , as defined above, $\mathbf{x}_3 := [1 \ 1 \ 0]^\top$ and $\mathbf{x}_4 := [-1 \ 1 \ 0]^\top$. Note that the columns of the data matrix

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4] = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1/2 & 0 & 0 \end{bmatrix}$$

have unit ℓ_∞ -norm, and we have the partitioning $\{\mathcal{P}_k\}_{k=1}^3$ given by $\mathcal{P}_1 = \{1\}$, $\mathcal{P}_2 = \{2\}$ and $\mathcal{P}_3 = \{3, 4\}$. The nullspace of \mathbf{X} is 1-dimensional, and spanned by $\eta = [1/4 \ -1/2 \ 1/8 \ 1/8]^\top$. Then, we trivially have $\text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^3) = \{\pm\eta\}$. Moreover, since the support of η is not contained in any \mathcal{P}_k , we also have $\text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^3, \mathcal{P}) = \{\pm\eta\}$. A simple observation reveals that $\text{Ext}(\text{im}(\mathbf{X}_{\mathcal{P}_1}^\top) \cap \mathbb{B}_\infty^2) = \{\pm 1\}$, $\text{Ext}(\text{im}(\mathbf{X}_{\mathcal{P}_2}^\top) \cap \mathbb{B}_\infty^1) = \{\pm 1\}$, and $\text{Ext}(\text{im}(\mathbf{X}_{\mathcal{P}_3}^\top) \cap \mathbb{B}_\infty^1) = \{\pm [1 \ 1]^\top, \pm [1 \ -1]^\top\}$. We see that (17) is violated for \mathcal{P}_2 . Hence, \mathbf{X} does not satisfy the SNSP.

Interestingly, if we normalize the columns of \mathbf{X} by the ℓ_2 -norm to obtain a new matrix \mathbf{X}_{ℓ_2} , then the nullspace of \mathbf{X}_{ℓ_2} is spanned by $\tilde{\eta} := [2 \ -\sqrt{10} \ 1 \ 1]^\top$ and \mathbf{X}_{ℓ_2} satisfies the SNSP. Therefore, the effect of normalization can be counter-intuitive as we already pointed out in the previous section.

So far, in this section, we have derived necessary and sufficient conditions for the SNSP to hold that reduce its verification to a decision on a finite set. Still, these characterizations can be computationally involved, which makes it desirable to obtain sufficient conditions that are simpler to verify. To derive such a condition, we need an upper bound on the dual of the ℓ_1 -minimization problem, which we provide next.

Proposition 1. *Let $\Psi \in \mathbb{R}^{r \times s}$ be a matrix with columns of unit ℓ_p -norm and $\bar{\mathbf{y}} \in \mathbb{R}^s$. Then,*

$$\max_{w \in \text{im}(\Psi^\top) \cap \mathbb{B}_\infty^s} \bar{\mathbf{y}}^\top w \leq \frac{\|\Psi \bar{\mathbf{y}}\|_p}{\tau_p(\Psi \mathbb{B}_1^s)}. \quad (18)$$

If $p \in \{1, \infty\}$ and $\tau_p(\Psi \mathbb{B}_1^s) = 1$, then equality holds.

Our characterizations in terms of extreme points require that we either solve an ℓ_1 -minimization problem or check the extreme points of $\text{im}(\mathbf{X}_P^\top) \cap \mathbb{B}_\infty^D$ to verify the SNSP. The bound in Prop. 1 allows us to simplify these characterizations and get rid of this requirement, as shown next.

Theorem 6. *Suppose that the columns of \mathbf{X} have unit ℓ_p -norm. The matrix \mathbf{X} satisfies the SNSP if, for all $\eta \in \text{Ext}(\text{Null}(\mathbf{X}) \cap \mathbb{B}_1^N, \mathcal{P})$ and all $P \in \mathcal{P}$, we have*

$$\|\eta_P\|_1 + \frac{\|\mathbf{X}_P(\eta_P)\|_p}{\tau_p(\mathcal{K}(\mathbf{X}_P))} < 1. \quad (19)$$

Proof. The result is an immediate consequence of Thm. 4 and Prop. 1 since $\mathcal{K}(\mathbf{X}_P) = \mathbf{X}_P \mathbb{B}_1^D$. \square

Before we end this section we would like to note that if the dimension of the nullspace of \mathbf{X} or the dimension of the individual subspaces $\{\mathcal{S}_i\}$ are large, then the number of extreme points can be large. However, numerical evidence suggests that making decisions based on a random sampling of extreme points can lead to satisfactory results (Kaba et al., 2021). The investigation of such an approach for subspace-preserving recovery is beyond the scope of this paper, so we leave it as future work.

6. Conclusion

In this paper we have provided several necessary and sufficient conditions for subspace-preserving recovery, which is at the heart of the analysis of subspace classification and clustering algorithms in machine learning. Our characterizations are based on a condition that is similar to the classical Nullspace Property (NSP), which we called Subspace Nullspace Property (SNSP). To the best of our knowledge, the SNSP is the first NSP-like condition for subspace-preserving recovery that has been extensively studied. We showed that some formulations of the SNSP give rise to a clear geometric interpretation or allow us to reduce the verification of the SNSP to decisions on finite sets. These finite sets correspond to certain extreme points, and hence, they also admit a clear geometric description. We also showed that all of our equivalent conditions can further be simplified to obtain sufficient conditions that are geometrically intuitive and computationally simpler.

The classical NSP plays an important role in sparse recovery literature. It is arguably the most popular necessary and sufficient condition that guarantees the success of ℓ_1 -minimization in the recovery of sparse signals. However, NSP-like conditions were mostly overlooked in the subspace-preserving recovery literature; the results in this paper fill that gap.

References

- Bajwa, W. U., Duarte, M. F., and Calderbank, R. Conditioning of random block subdictionaries with applications to block-sparse recovery and regression. *IEEE Transactions on Information Theory*, 61(7):4060–4079, 2015.
- Elhamifar, E. and Vidal, R. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2790–2797, 2009.
- Elhamifar, E. and Vidal, R. Clustering disjoint subspaces via sparse representation. In *IEEE International Conference*

- on *Acoustics, Speech, and Signal Processing*, pp. 1926–1929, 2010.
- Elhamifar, E. and Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11): 2765–2781, 2013.
- Foucart, S. and Rauhut, H. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser Basel, 2013.
- Jenatton, R., Audibert, J.-Y., and Bach, F. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*, 12:2777–2824, 2011.
- Kaba, M. D., Zhao, M., Vidal, R., Robinson, D. P., and Mallada, E. What is the largest sparsity pattern that can be recovered by 1-norm minimization? *IEEE Transactions on Information Theory*, 67(5):3060–3074, 2021.
- Kolar, M., Lafferty, J., and Wasserman, L. Union support recovery in multi-task learning. *Journal of Machine Learning Research*, 12(7), 2011.
- Li, C.-G., You, C., and Vidal, R. On geometric analysis of affine sparse subspace clustering. *IEEE Journal on Selected Topics in Signal Processing*, 12(6):1520–1533, 2018.
- Lu, C.-Y., Min, H., Zhao, Z.-Q., Zhu, L., Huang, D.-S., and Yan, S. Robust and efficient subspace segmentation via least squares regression. In *European Conference on Computer Vision*, pp. 347–360, 2012.
- Obozinski, G., Wainwright, M. J., Jordan, M. I., et al. Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*, 39(1):1–47, 2011.
- Robinson, D. P., Vidal, R., and You, C. Basis pursuit and orthogonal matching pursuit for subspace-preserving recovery: Theoretical analysis. *arXiv preprint arXiv:1912.13091*, 2019.
- Rockafellar, R. T. *Convex Analysis*. Princeton University Press, 1970.
- Soltanolkotabi, M. and Candès, E. J. A geometric analysis of subspace clustering with outliers. *Annals of Statistics*, 40(4):2195–2238, 2012.
- Soltanolkotabi, M., Elhamifar, E., and Candès, E. J. Robust subspace clustering. *Annals of Statistics*, 42(2):669–699, 2014.
- Stojnic, M., Parvaresh, F., and Hassibi, B. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Transactions on Signal Processing*, 57(8):3075–3085, 2009.
- Tsakiris, M. C. and Vidal, R. Theoretical analysis of sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pp. 4975–4984, 2018.
- Vidal, R., Ma, Y., and Sastry, S. *Generalized Principal Component Analysis*. Springer Verlag, 2016.
- Wang, Y., Wang, Y., and Singh, A. A deterministic analysis of noisy sparse subspace clustering for dimensionality-reduced data. In *International Conference on Machine Learning*, pp. 1422–1431, 2015.
- Wang, Y.-X. and Xu, H. Noisy sparse subspace clustering. *Journal of Machine Learning Research*, 17(12): 1–41, 2016.
- Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, Feb. 2009.
- You, C. and Vidal, R. Geometric conditions for subspace-sparse recovery. In *International Conference on Machine Learning*, pp. 1585–1593, 2015.
- You, C., Robinson, D. P., and Vidal, R. Scalable sparse subspace clustering by orthogonal matching pursuit. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3918–3927, 2016.
- You, C., Robinson, D. P., and Vidal, R. Provable self-representation based outlier detection in a union of subspaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4323–4332, 2017.
- You, C., Li, C.-G., Robinson, D. P., and Vidal, R. Is an affine constraint needed for affine subspace clustering? In *IEEE International Conference on Computer Vision*, 2019.