Hardness Results for Minimizing the Covariance of Randomly Signed Sum of Vectors

Peng Zhang
pz149@cs.rutgers.edu
Department of Computer Science
Rutgers University
November 29, 2022

Abstract

Given vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ with Euclidean norm at most 1 and $x_0 \in [-1, 1]^n$, our goal is to sample a random signing $x \in \{\pm 1\}^n$ with $E[x] = x_0$ such that the operator norm of the covariance of the signed sum of the vectors $\sum_{i=1}^n x(i)v_i$ is as small as possible. This problem arises from the algorithmic discrepancy theory and its application in the design of randomized experiments. It is known that one can sample a random signing with expectation $x_0$ and the covariance operator norm at most 1.

In this paper, we prove two hardness results for this problem. First, we show it is NP-hard to distinguish a list of vectors for which there exists a random signing with expectation $0$ such that the operator norm is 0 from those for which any signing with expectation $0$ must have the operator norm $\Omega(1)$. Second, we consider $x_0 \in [-1, 1]^n$ whose entries are all around an arbitrarily fixed $p \in [-1, 1]$. We show it is NP-hard to distinguish a list of vectors for which there exists a random signing with expectation $x_0$ such that the operator norm is 0 from those for which any signing with expectation $0$ must have the operator norm $\Omega((1 - |p|^2)^2)$.

1 Introduction

Given a list of $n$ vectors $V = v_1, \ldots, v_n \in \mathbb{R}^d$ and a vector $x_0 \in [-1, 1]^n$, our goal is to sample a random signing vector $x \in \{\pm 1\}^n$ with $E[x] = x_0$ such that the covariance of the signed sum of the vectors,

$$\text{Cov}(V, x) = \text{Cov} \left( \sum_{i=1}^n x(i)v_i \right) = \mathbb{E} \left[ \left( \sum_{i=1}^n (x(i) - x_0(i))v_i \right) \left( \sum_{i=1}^n (x(i) - x_0(i))v_i \right)^\top \right],$$

has the minimum operator norm. Here, $x(i)$ is the $i$th entry of $x$. Since the covariance scales quadratically with the maximum Euclidean norm of vectors $v_1, \ldots, v_n$, without loss of generality, we assume all $v_1, \ldots, v_n$ have Euclidean norms at most 1.

This problem arises from the algorithmic discrepancy theory and its application in the design of randomized experiments. A stronger version of the problem was first studied by Dadush, Garg, Lovett, and Nikolov [DGLN19], aiming to provide an algorithmic proof of Banaszczyk’s discrepancy problem [Ban98]. Here, the goal is to sample $x \in \{\pm 1\}^n$ with $E[x] = x_0$ such that $\sum_{i=1}^n (x(i) - x_0(i))v_i$ is $\sigma$-subgaussian. The $\sigma$-subgaussianity immediately implies the operator norm of $\text{Cov}(V, x)$, denoted by $||\text{Cov}(V, x)||$, is at most $\sigma^2$. Bansal, Dadush, Garg, and Lovett
designed a polynomial time algorithm, called the Gram-Schmidt Walk, that outputs a random $x \in \{\pm 1\}^n$ achieving $\sigma \leq \sqrt{40}$. This upper bound was then improved to $\sigma \leq 1$, by Harshaw, Sävje, Spielman, and Zhang [HSSZ19], which is tight and the equality holds when $n = d$ and $v_1, \ldots, v_n$ are the $n$ standard basis vectors in $\mathbb{R}^n$. Building on the Gram-Schmidt Walk algorithm, Harshaw, Sävje, Spielman, and Zhang [HSSZ19] proposed the Gram-Schmidt Walk Design to balance covariates in randomized experiments widely used in causal inference.

All the above upper bounds for $\|\text{Cov}(V, x)\|$, achieved by the Gram-Schmidt Walk, are independent of the optimal value, denoted by

$$C(V, x_0) \overset{\text{def}}{=} \min_{x \in \{\pm 1\}^n : E[x] = x_0} \|\text{Cov}(V, x)\|.$$  

It is natural to ask whether we can efficiently sample a random $x \in \{\pm 1\}^n$ with $E[x] = x_0$ such that $\|\text{Cov}(V, x)\|$ (approximately) equals the optimal value? In this paper, we prove strong hardness results for this question.

**Theorem 1.1.** There exists a constant $C_1 > 0$ such that given a list of vectors $V$ of Euclidean norm 1, it is NP-hard to distinguish whether $C(V, 0) = 0$ or $C(V, 0) > C_1$.

Theorem 1.1 concerns $x_0 = 0$; that is, for every $i \in \{1, \ldots, n\}$, the marginal probability of $x(i)$ being 1 or $-1$ equals 1/2. One may expect that the hardness comes from the balanced marginal probability for each $x(i)$. If the marginal probability of each $x(i)$ changes towards 0 or 1, the problem may become more tractable. In particular, when $x_0 \in \{\pm 1\}^n$, it is clear that $C(V, x_0) = 0$ due to no randomness. Our Theorem 1.2 concerns $x_0 \neq 0$. It shows a gap, parameterized only by entries of $x_0$, between the covariance operator norms of two cases which are NP-hard to distinguish.

**Theorem 1.2.** There exists a constant $C_2 > 0$ such that the following holds: For any $p, q \in [-1, 1]$, there exists $x_0 \in \{p, p + (1 - |p|)q, p - (1 - |p|)q\}^n$ such that given a list of $n$ vectors $V$ of Euclidean norm at most 1, it is NP-hard to distinguish whether $C(V, x_0) = 0$ or $C(V, x_0) > C_2(1 - |p|)^2q^2$.

The parameters $p, q$ in Theorem 1.2 may depend on $d$ or $n$. When the parameter $q$ is a constant near 0, all the entries of $x_0$ are near $p$; Theorem 1.2 implies that it is NP-hard to distinguish whether $C(V, x_0) = 0$ or $C(V, x_0) = \Omega(1 - |p|)^2$. When $|p|$ increases, the gap between the two cases in Theorem 1.2 decreases. In particular, when $|p|$ goes to 1, the gap goes to 0.

**Proof ideas.** Our proofs of Theorem 1.1 and 1.2 build on reductions from the 2-2 Set-Splitting problem, for which Guruswami [Gur14] proved strong NP-hardness results. Roughly speaking, in the 2-2 Set-Splitting problem, we are given a universe $U$ and a family $S$ of 4-subsets of $U$, and our goal is to assign each element in the universe 1 or $-1$ to maximize the number of “split” sets in $S$ (a split set has half elements assigned 1 and half $-1$). The 2-2 Set-Splitting problem is closely related to the problem of signing vectors to minimize their discrepancy. Reducing from the 2-2 Set-Splitting problem, Charikar, Newman, and Nikolov [CNN11] proved NP-hardness results for minimizing Spencer’s discrepancy [Spe85, Ban10]; Spielman and Zhang [SZ22] proved NP-hardness results for minimizing Weaver’s discrepancy [Wea04, MSS15, BCMS19]. Our proofs are inspired by those from [CNN11] and [SZ22]. However, our constructions are different from these two papers due to the different notions of discrepancy. The proof of Theorem 1.1 is a direct reduction from the 2-2 Set-Splitting problem, together with an inequality between matrix operator norm and matrix trace. The proof of Theorem 1.2 is slightly more involved due to the requirement of the nonzero expectation of $x$. Comparing to the proof of Theorem 1.1, we introduce auxiliary input vectors so that we can construct a signing $x$ with the required expectation and covariance 0 whenever such a
signing exists, and we employ an orthogonal projection matrix to force the sum of signed auxiliary vectors is almost zero with a sufficiently large probability under which the signed sum of all the input vectors behave similarly to those in Theorem 1.1 (without auxiliary vectors).

Organization of the rest of the paper. In Section 2 we introduce some preliminaries and notations, and formally define the 2-2 Set-Splitting problem and its variants, and state the known hardness results. We prove Theorem 1.1 in Section 3 and Theorem 1.2 in Section 4.

2 Preliminaries and Notations

2.1 Matrices and Vectors

Given a vector $x \in \mathbb{R}^n$, we let $x(i)$ be the $i$th entry of $x$. Given a matrix $A \in \mathbb{R}^{m \times n}$, we let $A(i,j)$ be the $(i,j)$th entry of $A$. The Euclidean norm of $x$ is $\|x\| = \sqrt{\sum_{i=1}^{n} x(i)^2}$. The operator norm of $A$ is $\|A\| = \sup_{x \in \mathbb{R}^n} \frac{\|Ax\|}{\|x\|}$.

When $A$ is a square matrix, the trace of $A$ is the sum of the entries on its main diagonal, denoted by $\text{tr}(A)$. The trace of $A$ equals the sum of the eigenvalues of $A$. In addition, we will use $1_n$ for the all-1 vector in $n$ dimensions and $0_n$ for the all-0 vector, and use $J_{m \times n}$ for the all-1 matrix in $m \times n$ dimensions and $0_{m \times n}$ for the all-0 matrix. When the context is clear, we drop the subscription for dimensions. We will use $I$ for the identity matrix.

2.2 2-2 Set-Splitting Problem

Our proofs of Theorem 1.1 and 1.2 build on reductions from the 2-2 Set-Splitting problem. In the 2-2 Set-Splitting problem, we are given a universe $U = \{1,2,\ldots,n\}$ and a family of sets $S = \{S_1,\ldots,S_m\}$ in which each $S_j$ consists of 4 distinct elements from $U$. Our goal is to find an assignment of the $n$ elements in $U$, denoted by $z \in \{\pm 1\}^n$, to maximize the number of sets in $S$ in which the values of its elements sum up to 0. We say an assignment $z$ 2-2-splits (or simply, splits) a set $S_j \in S$ if $\sum_{i \in S_j} z(i) = 0$; we say $z$ unsplits $S_j$ if $\sum_{i \in S_j} z(i) \in \{\pm 2, \pm 4\}$. We say an instance of the 2-2 Set-Splitting problem is satisfiable if there exists an assignment that splits all the sets in $S$. We say an instance is $\gamma$-unsatisfiable if any assignment must unsplit at least $\gamma$ fraction of the sets in $S$. Given a number $b \geq 1$, a 2-2 Set-Splitting instance is called a $(b,2)$ Set-Splitting instance if each element in $U$ appears in at most $b$ sets in $S$. In a $(b,2,2)$ Set-Splitting instance, we have $4m \leq bn$.

Theorem 2.1 ([Gur04]). For any constant $\epsilon > 0$, there exists a constant $b$ such that it is NP-hard to distinguish satisfiable $(b,2,2)$ Set-Splitting instances from $(1/12 - \epsilon)$-unsatisfiable instances.

A similar hardness result holds for $b = 3$. We will need it for our constructions.

Theorem 2.2 ([SZ22]). There exists a constant $\gamma > 0$ such that it is NP-hard to distinguish satisfiable instances of the $(3,2,2)$ Set-Splitting problem from $\gamma$-unsatisfiable instances.

3 Proof of Theorem 1.1

In this section, we prove Theorem 1.1.
Given a \((3, 2, 2)\) Set-Splitting instance where \(|U| = n\) and \(|S| = m\), we will construct a list of \(N\) vectors \(V = v_1, \ldots, v_N \in \mathbb{R}^d\) each of Euclidean norm 1 such that (1) if the given \((3, 2, 2)\) Set-Splitting instance is satisfiable, then \(C(V, 0) = 0\), and (2) if the given \((3, 2, 2)\) Set-Splitting instance is \(\gamma\)-unsatisfiable, then \(C(V, 0) > C_1\).

For each element \(i \in U\), let \(A_i \subset \{1, \ldots, m\}\) consist of the indices of the sets that contain \(i\). For each element \(i\) that appears in exactly 1 set in \(S\) (that is, \(|A_i| = 1\)), we create 4 new sets and 2 new elements. For each element \(i\) that appears in 2 sets in \(S\), we create 5 new sets and 3 new elements. Let \(B_i\) be the set consisting of the indices of the newly created sets for element \(i\). Suppose there are \(n_1\) elements in \(U\) that appear in exactly 1 set in \(S\) and \(n_2\) elements that appear in 2 sets. We set
\[
d = m + 4n_1 + 5n_2 = m + 5n\quad\text{and}\quad N = n + 2n_1 + 3n_2 \leq 4n.
\]

Consider each element \(i \in U\). There are 3 cases depending on how many sets in \(S\) containing \(i\):

1. Element \(i\) appears in 3 sets in \(S\): We define \(v_i \in \mathbb{R}^d\) such that \(v_i(j) = \frac{1}{\sqrt{3}}\) for \(j \in A_i\) and \(v_i(j) = 0\) otherwise.

2. Element \(i\) appears in 1 set in \(S\): Suppose \(B_i = \{i_1, i_2, i_3, i_4\}\). We define \(v_i \in \mathbb{R}^d\) such that \(v_i(j) = \frac{1}{\sqrt{3}}\) for \(j \in A_i \cup \{i_1, i_2\}\) and \(v_i(j) = 0\) otherwise. We define two more vectors: (1) \(u_{i,1} \in \mathbb{R}^d\) such that \(u_{i,1}(i_1) = u_{i,1}(i_3) = u_{i,1}(i_4) = \frac{1}{\sqrt{3}}\) and \(u_{i,1}(j) = 0\) for all other \(j\)'s, and (2) \(u_{i,2} \in \mathbb{R}^d\) such that \(u_{i,2}(i_2) = -\frac{1}{\sqrt{3}}\) and \(u_{i,2}(i_3) = u_{i,2}(i_4) = \frac{1}{\sqrt{3}}\) and \(u_{i,1}(j) = 0\) for all other \(j\)'s.

3. Element \(i\) appears in 2 sets in \(S\): Suppose \(B_i = \{i_1, i_2, i_3, i_4, i_5\}\). We define \(v_i \in \mathbb{R}^d\) such that \(v_i(j) = \frac{1}{\sqrt{3}}\) for \(j \in A_i \cup \{i_1\}\) and \(v_i(j) = 0\) otherwise. We define three more vectors (1) \(u_{i,1} \in \mathbb{R}^d\) such that \(u_{i,1}(i_1) = u_{i,1}(i_2) = u_{i,1}(i_3) = \frac{1}{\sqrt{3}}\) and \(u_{i,1}(j) = 0\) for all other \(j\)'s, (2) \(u_{i,2} \in \mathbb{R}^d\) such that \(u_{i,2}(i_2) = u_{i,2}(i_4) = u_{i,2}(i_5) = \frac{1}{\sqrt{3}}\) and \(u_{i,2}(j) = 0\) for all other \(j\)'s, and (3) \(u_{i,3} \in \mathbb{R}^d\) such that \(u_{i,3}(i_3) = -\frac{1}{\sqrt{3}}\) and \(u_{i,3}(i_4) = u_{i,3}(i_5) = \frac{1}{\sqrt{3}}\) and \(u_{i,3}(j) = 0\) for all other \(j\)'s.

We let \(v_{n+1}, \ldots, v_N\) be the vectors \(u_{i,h}\)'s constructed above. We can check that all \(v_1, \ldots, v_N\) have Euclidean norm 1.

**Claim 3.1.** For any \(z \in \{\pm 1\}^n\), we can construct an \(y \in \{\pm 1\}^N\) such that (1) \(y(i) = z(i)\) for \(i \in \{1, \ldots, n\}\) and (2) \(\sum_{i=1}^N y(i)v_i(j) = 0\) for all \(j \in \{m+1, \ldots, d\}\).

**Proof.** We only need to determine the signs for the vectors \(u_{i,h}\)'s constructed for elements appearing in less than 3 sets in \(S\) and to check the coordinates of \(\sum_{i=1}^N y(i)v_i\) whose indices in \(B_i\)'s. Let \(i \in U\) be an element that appears in 1 set in \(S\). The subvectors of \(v_i, u_{i,1}, u_{i,2}\) restricted to the coordinates in \(B_i\) are:
\[
\begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix}, \quad \begin{pmatrix}
1 \\
0 \\
1
\end{pmatrix}, \quad \begin{pmatrix}
0 \\
-1 \\
1
\end{pmatrix}.
\]

We choose the signs in \(y\) for \(u_{i,1}, u_{i,2}\) to be \(-z(i)\) and \(z(i)\), respectively, which guarantees the signed sum of the \(v_1, u_{i,1}, u_{i,2}\) is 0 when restricted to \(B_i\). Since any other vector has 0 for the coordinates in \(B_i\), we have \(\sum_{i=1}^N y(i)v_i(j) = 0\) for \(j \in B_i\). Now, let \(i \in U\) be an element that
Thus, $\sum_{i=1}^{N} y(i)v_i = 0$. Let $z$ be such that $\sum_{i=1}^{n} z(i)v_i = 0$. We define a random vector $\mathbf{x} \in \{\pm 1\}^N$ such that $\mathbf{x} = \mathbf{y}$ with probability $1/2$ and $\mathbf{x} = -\mathbf{y}$ with probability $1/2$. Then, $\mathbb{E}[\mathbf{x}] = 0$ and $\text{Cov}(\mathbf{V}, \mathbf{x}) = 0$, that is, $C(\mathcal{V}, \mathbf{0}) = 0$.

Suppose the given $(3, 2)$ Set-Splitting instance is $\gamma$-unsatisfiable, meaning that for any assignment $\mathbf{z} \in \{\pm 1\}^n$, at least $\gamma$ fraction of the entries of $\sum_{i=1}^{n} z(i)v_i$ are in $\{\pm 2, \pm 4\}$. Then, for any $\mathbf{y} \in \{\pm 1\}^N$, at least

$$\frac{\gamma n}{N} \geq \frac{\gamma}{4}$$

fraction of the entries of $\sum_{i=1}^{N} y(i)v_i$ are in $\{\pm 2, \pm 4\}$. Then, for any random $\mathbf{x} \in \{\pm 1\}^N$ with $\mathbb{E}[\mathbf{x}] = 0$, let $\mathbf{w} = \sum_{i=1}^{N} x(i)v_i$,

$$\|\text{Cov}(\mathbf{w})\| = \left\|\mathbb{E}[\mathbf{ww}^T]\right\| \geq \frac{1}{d} \text{tr}\left(\mathbb{E}[\mathbf{ww}^T]\right) = \frac{1}{d} \mathbb{E}\left[\text{tr}(\mathbf{ww}^T)\right]$$

$$\geq \frac{1}{d} \cdot 4 \cdot \frac{\gamma N}{d} = \frac{\gamma N}{d} \geq \frac{4\gamma}{23}.$$  

The last inequality holds since $d \leq m + 5n \leq \frac{23m}{4} \leq \frac{23N}{4}$. That is, $C(\mathcal{V}, \mathbf{0}) > \frac{4\gamma}{23}$. If we can distinguish whether $C(\mathcal{V}, \mathbf{0}) = 0$ or $C(\mathcal{V}, \mathbf{0}) > \frac{4\gamma}{23}$, then we can distinguish whether a $(3, 2)$ Set-Splitting instance is satisfiable or $\gamma$-unsatisfiable, which is NP-hard by Theorem 2.2. This completes the proof of Theorem 1.1.

4 Proof of Theorem 1.2

In this section, we prove Theorem 1.2.

Given a $(3, 2)$ Set-Splitting instance where $|U| = n$ and $|S| = m$, we will construct a list of vectors. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be the incidence matrix of the $(3, 2)$ Set-Splitting instance, where $A(j, i) = 1$ if element $i \in S_j$ and $A(j, i) = 0$ otherwise. Since each set in the $(3, 2)$ Set-Splitting instance has 4 distinct elements, each row of $\mathbf{A}$ has sum 4. Then, we define

$$M \overset{\text{def}}{=} \begin{pmatrix} A & -2I & -2I \\ 0 & I - \frac{1}{m}J & 0 \\ 0 & 0 & I - \frac{1}{m}J \end{pmatrix}.$$  

Let $\mathbf{P} \overset{\text{def}}{=} \mathbf{I} - \frac{1}{m}\mathbf{J}$. $\mathbf{P}$ is an orthogonal projection matrix onto the subspace orthogonal to $\mathbf{1}_m$. The dimensions of $M$ are $3m \times (n + 2m)$. Let $D = 3m$ and let $N = n + 2m$. We define a list of vectors $\mathcal{V} = \mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathbb{R}^D$ to be the columns of $M$. Note that each $\mathbf{v}_i$ has $O(1)$ Euclidean
norm. Dividing each \( v_i \) by the maximum norm among all \( v_1, \ldots, v_N \) yields a list of vectors with Euclidean norm at most 1. Without loss of generality, we assume \( p \geq 0 \). We define

\[
x_0 \overset{\text{def}}{=} \begin{pmatrix}
    p1_n \\
    (p + (1 - p)q)1_m \\
    (p - (1 - p)q)1_m
\end{pmatrix}.
\]

Denote \( \beta = (1 - p)q \). For any \( x \in \mathbb{R}^N \),

\[
\sum_{i=1}^N x(i)v_i = Mx.
\]

**Claim 4.1.** If \( x \in \mathbb{R}^N \) satisfies \( \mathbb{E}[x] = x_0 \), then \( \text{Cov}(Mx) = \mathbb{E} [Mxx^\top M^\top] \).

**Proof.** Note that

\[
\text{Cov}(Mx) = \mathbb{E} \left[ M(x - x_0)(x - x_0)^\top M^\top \right].
\]

It suffices to show that \( Mx_0 = 0 \).

\[
Mx_0 = \begin{pmatrix}
pA1_n - 2(p + \beta)1_m - 2(p - \beta)1_m \\
(p + \beta)\Pi1_m \\
(p - \beta)\Pi1_m
\end{pmatrix}.
\]

Since \( A1 = 41 \) and \( \Pi1 = 0 \), we have \( Mx_0 = 0 \).

Suppose the (3, 2-2) Set-Splitting instance is satisfiable, and let \( z \in \{\pm 1\}^n \) be an assignment that splits all the sets. We will show \( C(V, x_0) = 0 \), that is, there exists a random \( x \in \mathbb{R}^N \) such that \( \mathbb{E}[x] = x_0 \) and \( \text{Cov}(Mx) = 0 \). We let

\[
x = \begin{cases}
    1, & \text{with probability } p \\
    z \\
    1_m \\
    -1_m \\
    -z \\
    1_m \\
    -1_m \\
    -z \\
    1_m \\
    -1_m \\
    z \\
    1_m \\
    -1_m
\end{cases}, \quad \text{with probability } \frac{(1-p)(1+q)}{4} \\
\text{with probability } \frac{(1-p)(1+q)}{4} \\
\text{with probability } \frac{(1-p)(1-q)}{4} \\
\text{with probability } \frac{(1-p)(1-q)}{4} \\
\text{with probability } \frac{(1-p)(1-q)}{4} \\
\text{with probability } \frac{(1-p)(1-q)}{4}.
\]

**Claim 4.2.** \( \mathbb{E}[x] = x_0 \).

**Proof.** By our setting of \( x \):

\[
\mathbb{E}[x] = p1 + \frac{(1-p)(1+q)}{2} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = x_0.
\]

\( \square \)
Claim 4.3. $\text{Cov}(Mx) = 0$.

Proof. By Claim 4.1,
$$\text{Cov}(Mx) = \mathbb{E} \left[ Mxx^\top M^\top \right].$$
We will show that $Mx = 0$ always holds. We check all the vectors in the support of $x$. If $x = 1$, then
$$Mx = \begin{pmatrix} A1 - 21 \\ \Pi1 \\ \Pi1 \end{pmatrix} = 0.$$
If $x = \begin{pmatrix} \pm z \\ 1 \\ -1 \end{pmatrix}$, then
$$Mx = \begin{pmatrix} \pm Az - 21 + 21 \\ \Pi1 \\ -\Pi1 \end{pmatrix} = 0,$$
where we use the fact $Az = 0$. Similarly, if $x = \begin{pmatrix} \pm z \\ -1 \\ 1 \end{pmatrix}$, then $Mx = 0$.

Suppose the $(3,2)$ Set-Splitting instance is $\gamma$-unsatisfiable, that is, for any $z \in \{-1,1\}^n$, at least $\gamma$ fraction of the entries of $Az$ are in $\{-2,2,-4,4\}$. We will show $C(V, x_0) = \Omega(\beta^2)$, that is, for any random $x \in \{-1,1\}^n$ satisfying $\mathbb{E}[x] = x_0$, the operator norm of $\text{Cov}(Mx)$ is $\Omega(\beta^2)$. We write $x$ as
$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$
where $x_1 \in \{-1,1\}^n$ and $x_2, x_3 \in \{-1,1\}^m$. Then,
$$Mx = \begin{pmatrix} Ax_1 - 2(x_2 + x_3) \\ \Pi x_2 \\ \Pi x_3 \end{pmatrix}.$$
The following claim splits $\|\text{Cov}(Mx)\|$ into three terms.

Claim 4.4. $\|\text{Cov}(Mx)\| \geq \frac{1}{D}\max \left\{ \mathbb{E} \left[ \|Ax_1 - 2(x_2 + x_3)\|^2 \right], \mathbb{E} \left[ \|\Pi x_2\|^2 \right], \mathbb{E} \left[ \|\Pi x_3\|^2 \right] \right\}$.

Proof. By Claim 4.1
$$\|\text{Cov}(Mx)\| = \left\| \mathbb{E} \left[ Mxx^\top M^\top \right] \right\|$$
$$\geq \frac{1}{D}\text{tr} \left( \mathbb{E} \left[ Mxx^\top M^\top \right] \right)$$
$$= \frac{1}{D}\mathbb{E} \left[ \text{tr} \left( Mxx^\top M^\top \right) \right]$$
$$= \frac{1}{D}\mathbb{E} \left[ \|Mx\|^2 \right]$$
$$= \frac{1}{D} \left( \mathbb{E} \left[ \|Ax_1 - 2(x_2 + x_3)\|^2 \right] + \mathbb{E} \left[ \|\Pi x_2\|^2 \right] + \mathbb{E} \left[ \|\Pi x_3\|^2 \right] \right)$$
$$\geq \frac{1}{D}\max \left\{ \mathbb{E} \left[ \|Ax_1 - 2(x_2 + x_3)\|^2 \right], \mathbb{E} \left[ \|\Pi x_2\|^2 \right], \mathbb{E} \left[ \|\Pi x_3\|^2 \right] \right\}.$$
We will show that at least one of the three terms in the rightmost-hand side is sufficiently large. We first look at the last two terms \( \| \Pi x_2 \|^2 \) and \( \| \Pi x_3 \|^2 \). Let \( y \in \{ \pm 1 \}^m \) be any vector. Then,

\[
\| \Pi y \|^2 = \left\| y - \frac{y^\top 1}{m} \cdot 1 \right\|^2.
\]

Let \( \alpha(y) = \frac{y^\top 1}{m} \). Then, \( \| \Pi y \|^2 = (1 - \alpha(y)^2)m \).

**Claim 4.5.** If \( \mathbb{E}[\alpha(x_2)^2] \) or \( \mathbb{E}[\alpha(x_3)^2] \) smaller than \( 1 - \frac{\gamma^2}{24} \), then \( \text{Cov}(Mx) = \Omega(\gamma \beta^2) \).

**Proof.** Without loss of generality, we assume \( \mathbb{E}[\alpha(x_2)^2] < 1 - \frac{\gamma^2}{24} \). The argument for \( \mathbb{E}[\alpha(x_3)^2] \) is the same. By Claim 4.4,

\[
\| \text{Cov}(Mx) \| \geq \frac{1}{D} \mathbb{E} \left[ \| \Pi x_2 \|^2 \right] = \frac{(1 - \mathbb{E}[\alpha(y)^2])m}{D} = \Omega(\gamma \beta^2).
\]

In the rest of the proof, we assume that both \( \mathbb{E}[\alpha(x_2)^2] \) and \( \mathbb{E}[\alpha(x_3)^2] \) are at least \( 1 - \frac{\gamma^2}{24} \). We will show that under this assumption, with probability \( \Omega(\beta) \), a large fraction of the entries of \( x_2 + x_3 \) are 0, and thus \( \| Ax_1 - 2(x_2 + x_3) \|^2 \approx \| Ax_1 \|^2 \). This implies \( \mathbb{E} \left[ \| Ax_1 - 2(x_2 + x_3) \|^2 \right] = \Omega(\beta m) \).

We will need the following properties about \( \alpha(y) \) for any vector \( y \in \{ \pm 1 \}^m \). When the context is clear, we use \( \alpha \) for \( \alpha(y) \).

**Claim 4.6.** For any \( \delta > 0 \), \( \mathbb{E}[\alpha^2] \leq \mathbb{E}[|\alpha| \leq \delta] \cdot \delta^2 + 1 - \mathbb{E}[|\alpha| \leq \delta] \).

**Proof.** Note that

\[
\mathbb{E}[\alpha^2] \leq \mathbb{E}[|\alpha| \leq \delta] \cdot \delta^2 + 1 - \mathbb{E}[|\alpha| \leq \delta] \cdot \delta^2 + 1 - \mathbb{E}[|\alpha| \leq \delta].
\]

Thus,

\[
\mathbb{E}[\alpha^2] \leq \frac{1 - \mathbb{E}[\alpha^2]}{1 - \delta^2}.
\]

**Claim 4.7.** For any \( \delta > 0 \),

\[
\mathbb{P}(\alpha > \delta) \geq \frac{\mathbb{E}[\alpha] + \delta(1 - 2 \mathbb{P}(|\alpha| \leq \delta))}{1 + \delta}, \quad \mathbb{P}(\alpha < -\delta) \geq \frac{-\mathbb{E}[\alpha] + \delta(1 - 2 \mathbb{P}(|\alpha| \leq \delta))}{1 + \delta}.
\]

**Remark.** Since \( -1 \leq \mathbb{E}[\alpha] \leq 1 \), the above two lower bounds are both smaller than 1.

**Proof.** We introduce some notations:

\[
\pi = \mathbb{P}(|\alpha| \leq \delta), \quad \pi_+ = \mathbb{P}(\alpha > \delta), \quad \pi_- = \mathbb{P}(\alpha < -\delta).
\]

Let \( 1_{|\alpha| > \delta} \) be the indicator such that \( 1_{|\alpha| > \delta} = 1 \) if \( |\alpha| > \delta \) and \( 1_{|\alpha| > \delta} = 0 \) otherwise. Then,

\[
\mathbb{E} \left[ \alpha 1_{|\alpha| > \delta} \right] = \mathbb{E}[\alpha] - \mathbb{E} \left[ \alpha 1_{|\alpha| \leq \delta} \right] \geq \mathbb{E}[\alpha] - \delta \pi.
\]
On the other hand,
\[ E[\alpha 1_{|\alpha| > \delta}] \leq \pi_+ - \delta\pi_- = \pi_+ - \delta(1 - \pi_+ - \pi) = (1 + \delta)\pi_+ + \delta\pi - \delta. \]

Combining the above two inequalities:
\[ \pi_+ \geq \frac{E[\alpha] + \delta(1 - 2\pi)}{1 + \delta}. \]

To lower bound \( \pi_- \), we note that
\[ E[\alpha 1_{|\alpha| > \delta}] \leq E[\alpha] + \delta\pi. \]

On the other hand,
\[ E[\alpha 1_{|\alpha| > \delta}] \geq \delta\pi_+ - \pi_- = \delta(1 - \pi - \pi_-) - \pi_- = -(1 + \delta)\pi_- - \delta\pi + \delta. \]

Combining the above two inequalities:
\[ \pi_- \geq \frac{-E[\alpha] + \delta(1 - 2\pi)}{1 + \delta}. \]

We choose \( \delta = 1 - \frac{\gamma\beta}{10} \). Here, we choose this number to simplify our calculation; we can choose \( \delta \) to be any number such that \( \gamma - (1 - \delta) \) is greater than a positive constant. By our choice for \( x_0 \),
\[ E[\alpha(x_2)] = p + \beta \text{ and } E[\alpha(x_3)] = p - \beta. \]

Let \( E \) be the event that both \( \alpha(x_2) > \delta \) and \( \alpha(x_3) < -\delta \) happen. By union bound,
\[
\Pr(E) \geq \Pr(\alpha(x_2) > \delta) + \Pr(\alpha(x_3) < -\delta) - 1
\geq \frac{E[\alpha(x_2)] - E[\alpha(x_3)] + 2\delta - 2\delta(\Pr(|\alpha(x_2)| \leq \delta) + \Pr(|\alpha(x_3)| \leq \delta))}{1 + \delta} - 1 \tag{by Claim 4.7}
\geq \frac{2\beta + 2\delta - \frac{2\delta}{1 - 2\pi}(2 - E[\alpha(x_2)^2] - E[\alpha(x_3)^2])}{1 + \delta} - 1 \tag{by Claim 4.6}
\geq \frac{1}{1 + \delta} \left( 2\beta + 1 - \frac{\gamma\beta}{10} - 4 \left( 1 - \frac{\gamma\beta}{10} \right) \left( \frac{\gamma\beta^2/24}{1 - (1 - \gamma\beta/10)^2} - 1 \right) \right)
\geq \frac{1}{1 + \delta} \left( 2\beta - \frac{\gamma\beta}{10} - \left( 1 - \frac{\gamma\beta}{10} \right) \beta \right)
\geq \frac{\beta}{2} \left( 1 - \frac{\gamma}{10} \right). \tag{by our setting of \( \delta \) and assumption on \( E[\alpha(x_2)^2], E[\alpha(x_3)^2] \)}
\]

Assume event \( E \) happens. At least
\[ \frac{1 + \alpha(x_2)}{2} > 1 - \frac{\gamma\beta}{24} \]

fraction of the entries of \( x_2 \) are 1; at least
\[ \frac{1 - \alpha(x_3)}{2} > 1 - \frac{\gamma\beta}{24} \]
fraction of the entries of $x_3$ are $-1$. Thus, at least $1 - \frac{\gamma}{12}$ fraction of the entries of $x_2 + x_3$ are 0. Note that among these 0-valued entries of $x_2 + x_3$, at least $\gamma(1 - \frac{\beta}{12})$ fraction of the entries of $Ax_1$ are in $\{ \pm 2, \pm 4 \}$. In this case,

$$\|Mx\|^2 \geq \|Ax_1 - 2(x_2 + x_3)\|^2 \geq 4\gamma \left(1 - \frac{\beta}{12}\right)m.$$  

By Claim 4.4,

$$\|\text{Cov}(Mx)\| \geq \frac{1}{D}E\left[\|Ax_1 - 2(x_2 + x_3)\|^2\right]$$

$$\geq \frac{1}{D} \Pr(E) \cdot 4\gamma \left(1 - \frac{\beta}{12}\right)m$$

$$\geq \frac{1}{5m} \frac{\beta}{2} \left(1 - \frac{\gamma}{10}\right) \cdot 4\gamma \left(1 - \frac{\beta}{12}\right)m$$

$$= \Omega(\beta).$$

Together with Claim 4.5, a $\gamma$-unsatisfiable $(3, 2)$-Set-Splitting instance leads to $C(V, x_0) = \Omega(\beta^2)$.

If we can distinguish whether $C(V, x_0) = 0$ or $C(V, x_0) = \Omega(\beta^2)$, then we can distinguish whether a $(3, 2)$-Set-Splitting instance is satisfiable or $\gamma$-unsatisfiable, which is NP-hard by Theorem 2.2. This completes the proof of Theorem 1.2.

References

[Ban98] Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 3–10. IEEE, 2010.

[BCMS19] Marcin Bownik, Pete Casazza, Adam W Marcus, and Darrin Speegle. Improved bounds in weaver and feichtinger conjectures. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2019(749):267–293, 2019.

[BDGL18] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The gram-schmidt walk: a cure for the banaszczyk blues. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 587–597, 2018.

[CNN11] Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1607–1614. SIAM, 2011.

[DGLN19] Daniel Dadush, Shashwat Garg, Shachar Lovett, and Aleksandar Nikolov. Towards a constructive version of banaszczyk’s vector balancing theorem. *Theory of Computing*, 15(1):1–58, 2019.

[Gur04] Venkatesan Guruswami. Inapproximability results for set splitting and satisfiability problems with no mixed clauses. *Algorithmica*, 38(3):451–469, 2004.

[HSSZ19] Christopher Harshaw, Fredrik Sävje, Daniel Spielman, and Peng Zhang. Balancing co-variates in randomized experiments with the gram–schmidt walk design. *arXiv preprint arXiv:1911.03071*, 2019.
[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison—singer problem. *Annals of Mathematics*, pages 327–350, 2015.

[Spe85] Joel Spencer. Six standard deviations suffice. *Transactions of the American mathematical society*, 289(2):679–706, 1985.

[SZ22] Daniel A Spielman and Peng Zhang. Hardness results for weaver’s discrepancy problem. *arXiv preprint arXiv:2205.01482*, 2022.

[Wea04] Nik Weaver. The kadison–singer problem in discrepancy theory. *Discrete mathematics*, 278(1-3):227–239, 2004.