The beltway problem over orthogonal groups

Tamir Bendory, Dan Edidin, and Oscar Mickelin

Abstract

The classical beltway problem entails recovering a set of points from their unordered pairwise distances on the circle. This problem can be viewed as a special case of the crystallographic phase retrieval problem of recovering a sparse signal from its periodic autocorrelation. Based on this interpretation, and motivated by cryo-electron microscopy, we suggest a natural generalization to orthogonal groups: recovering a sparse signal, up to an orthogonal transformation, from its autocorrelation over the orthogonal group. If the support of the signal is collision-free, we bound the number of solutions to the beltway problem over orthogonal groups, and prove that this bound is exactly one when the support of the signal is radially collision-free (i.e., the support points have distinct magnitudes). We also prove that if the pairwise products of the signal's weights are distinct, then the autocorrelation determines the signal uniquely, up to an orthogonal transformation. We conclude the paper by considering binary signals and show that in this case, the collision-free condition need not be sufficient to determine signals up to orthogonal transformation.

1. Introduction

The beltway problem consists of recovering k points $t_1, \dots, t_k$ on the circle from knowing only the unordered set of their pairwise distances, measured along the circle [1, 2]. Clearly, the translation of any solution to the beltway problem by a global rotation or reflection is also a solution. Two solutions are called *equivalent* if they are related by a rotation and reflection. Two sets of points with identical difference sets are termed *homometric*. Generally, a given set of distances can give rise to non-equivalent solutions. Namely, there may exist homometric sets which are not equivalent. In [1], the maximum possible number $S(k)$ of non-equivalent and homometric sets, over all sets of k points on the unit circle, was bounded by $\exp\left(2^{\frac{\ln k}{\ln \ln k}}\right) \leq S(k) \leq \frac{1}{2}k^{k-2}$.

Crystallographic phase retrieval. The beltway problem originally arose in X-ray crystallography because of its connection to the phase retrieval problem. The (discrete) phase retrieval problem entails recovering a signal $x \in \mathbb{R}^n$, up to a sign, cyclic shift and reflection, from the magnitudes of its Fourier transform. It can be readily seen that this problem is equivalent to recovering x from its periodic autocorrelation [3],

$$A_\ell(x) = \sum_{i=0}^{n-1} x_i x_{(i+\ell) \bmod n}, \quad \ell = 0, \dots, n-1. \tag{1.1}$$

Note that the autocorrelation (1.1) is the second moment of $(g \cdot x)$, where g is a uniformly distributed element of the group of circular shifts $\mathbb{Z}_n$; this observation plays a pivotal role in this paper.

If $x \in \mathbb{R}^n$ is a binary signal, then the periodic autocorrelation $A(x)$ is determined by the cyclic difference set of the support of x, and two binary signals have the same autocorrelation if and only if their supports have the same cyclic difference sets. Thus, recovering a binary signal from its periodic autocorrelation (1.1) reduces to the beltway problem on the vertices of the regular

n-gon inscribed in the circle. This problem is important since it serves as an approximation of the physical model of X-ray crystallography technology, and thus can be thought of as a special case of the *crystallographic phase retrieval problem* [4, 5]. The general crystallographic phase retrieval problem is the problem of recovering a signal $x \in \mathbb{R}^n$, up to a cyclic shift and reflection, from its second moment, with the classical beltway problem corresponding to binary signals. If the entries of the signal lie in a small alphabet (which in X-ray crystallography corresponds to different types of atoms [5]), then recovering the signal from its autocorrelation is equivalent to checking if several partitions of its support have the same difference sets [6].

Existing uniqueness results for phase retrieval. The phase retrieval problem is typically ill-posed even for binary signals, and only very recently have uniqueness results been obtained. A subset of the authors of this paper established a conjecture that a sparse signal with generic entries is uniquely determined by its periodic autocorrelation as long as $k \leq n/2$ [7]. In particular, [7] provides a computational test to check the uniqueness for any pair (k, n). However, because of its heavy computational burden, the conjecture was affirmed only up to $n = 9$. In [8], it was shown that if the signal is sparse with respect to a generic basis in $\mathbb{R}^n$, then $k \leq n/2$ guarantees unique recovery for generic signals, and $k \leq n/4$ suffices to determine all signals. In [9], it was shown that *symmetric* signals can be recovered from their periodic autocorrelation for $k = O(n/\log^5 n)$. The computational aspects of the beltway problem were studied in [1, 2, 10, 11, 12].

Autocorrelations over orthogonal groups. The autocorrelation function (1.1) is the second moment of circular shifts of the signal, where the circular shifts are drawn from a uniform distribution. Analogously, one can define an autocorrelation over a compact group G as the expectation of all the products of two entries of $g \cdot x$, where g is drawn from a uniform distribution over G. The beltway problem over orthogonal groups is that of recovering a discretely supported signal from its autocorrelation over an orthogonal group. We formalize the problem rigorously in Section 2 and motivate it by single-particle cryo-electron microscopy: a prominent technology in structural biology to elucidate the spatial structure of biological molecules [13].

Collision-free support. In the classical beltway problem, the points $t_1, \ldots, t_k$ are termed *collision free* if each non-zero distance in the difference set of the points appears with multiplicity one. If we interpret the beltway problem as recovering a sparse signal from its periodic autocorrelation, then a signal is collision-free if there are no repeated differences in the signal's support. We generalize this definition to the beltway problem over orthogonal groups, see Definition 3.1. The collision-free hypothesis is essential to our analysis because, otherwise, it is not necessarily possible to determine the size of the support set from the autocorrelation, except in the binary case when all non-zero entries of the signal are one.

Main contributions. The contribution of this paper is three-fold. First, Section 2 formalizes the beltway problem over orthogonal groups. Second, we prove that the $\mathsf{O}(n)$-orbit of a sparse signal can be uniquely recovered from its autocorrelation if the signal satisfies the radially collision-free condition (i.e., the support points are on distinct spheres). If the signal satisfies an extension of the classical collision-free condition, we bound the number of orbits that correspond to a given autocorrelation. Section 3 presents the main results, which are proved in Section 4. The third contribution, in Section 5, is a detailed discussion on binary signals whose support is collision free but not radially collision-free. This case is more subtle. We derive conditions where the binary signal cannot be determined from its second moment, and conjecture when it can be determined. We also discuss the relationship with the turnpike problem.

2. The beltway problem over orthogonal groups

2.1. Problem formulation

The formulation of the classical beltway problem as that of recovering a binary signal from its periodic autocorrelation leads to a natural generalization over orthogonal groups. Let x be a signal on $\mathbb{R}^n$; that is, a function $\mathbb{R}^n \to \mathbb{R}$. The second moment of the translated signal $g \cdot x$ with respect to the uniform (Haar) measure over $\mathsf{O}(n)$ is the function on $(\mathbb{R}^n)^2$

$$m_2(x)(\tau_1, \tau_2) = \int_{\mathsf{O}(n)} (g \cdot x)(\tau_1)(g \cdot x)(\tau_2) dg. \tag{2.1}$$

Note that m_2 is $\mathsf{O}(n)$-invariant, namely, $m_2(g \cdot x) = m_2(x)$ for any $g \in \mathsf{O}(n)$, so we can view $m_2(x)$ as a function $(\mathbb{R}^n)^2/\mathsf{O}(n) \to \mathbb{R}$. This, in turn, implies that we can only expect to determine the $\mathsf{O}(n)$-orbit of x from its second moment. In this case, the second moment is typically referred to as the *autocorrelation*.

More generally, it is possible to define the auto-correlation of a (compactly supported) distribution, such as a Dirac delta function, which will produce a distribution supported on G-invariant (compact) subsets of $(\mathbb{R}^n)^2$ (cf. [14, Section 2.1]). In detail, given a collection of points $t_1, \ldots, t_k \in \mathbb{R}^n$ and positive weights[1] $w_1, \ldots, w_k \in \mathbb{R}$, we define a k-sparse signal on $\mathbb{R}^n$ as

$$x = \sum_{i=1}^{k} w_i \delta_{t_i}, \tag{2.2}$$

where δ_t is a point mass located at $t \in \mathbb{R}^n$, represented by the Dirac delta function. With a slight abuse of notation, we occasionally refer to signals of the form (2.2) as *δ-functions*. There is an action of $\mathsf{O}(n)$ on δ-functions of the form (2.2), given by $g \cdot x = \sum_{i=1}^{k} w_i \delta_{g \cdot t_i}$.

In this case, we define the autocorrelation $m_2(x)$ as a distribution on $(\mathbb{R}^n)^2$, supported on the $\mathsf{O}(n)$-orbits $\overline{(t_i, t_j)} := \{(g \cdot t_i, g \cdot t_j) \,|\, g \in G\}$, where $i, j \in \{1, \ldots, k\}$ as follows. By [15, Example V.5.12] the (tensor) product $(g \cdot \delta_{t_i})(\tau_1)(g \cdot \delta_{t_j})(\tau_2)$ is the point mass $\delta_{(g \cdot t_i, g \cdot t_j)}(\tau_1, \tau_2)$ supported at $(g \cdot t_i, g \cdot t_j) \in (\mathbb{R}^n)^2$; We then define the integral of these point masses over $\mathsf{O}(n)$

$$\mu_{t_i, t_j}(\tau_1, \tau_2) = \int_{\mathsf{O}(n)} (g \cdot \delta_{t_i})(\tau_1)(g \cdot \delta_{t_j})(\tau_2) \, dg$$

to be the distribution with support the orbit $\overline{(t_i, t_j)}$ characterized by property that if f is a compactly supported smooth test function on $(\mathbb{R}^n)^2$ then

$$\int_{(\mathbb{R}^n)^2} f(\tau_1, \tau_2) \mu_{t_i, t_j}(\tau_1, \tau_2) \, d\tau_1 d\tau_2 = \int_{\mathsf{O}(n)} f(g \cdot t_i, g \cdot t_j) \, dg,$$

where the left hand side is the evaluation of the distribution μ_{t_i, t_j} on the test function $f(\tau_1, \tau_2)$, following the notation of [16, Section V.3]. The autocorrelation is then the distribution

$$m_2(x)(\tau_1, \tau_2) = \sum_{i,j=1}^{k} w_i w_j \mu_{t_i, t_j}(\tau_1, \tau_2). \tag{2.3}$$

We are now ready to define the beltway problem over orthogonal groups.

[1]The assumption that the weights are positive is physically motivated and allows us to streamline the exposition. From a mathematical perspective, if we assume that the weights are arbitrary, then we must also account for the fact that x and $-x$ have the same autocorrelation.

Definition 2.1. The beltway problem over an orthogonal group $\mathsf{O}(n)$ is the task of determining the $\mathsf{O}(n)$-orbit of a δ-function of the form (2.2) in $\mathbb{R}^n$ from its autocorrelation (2.3).

In the next section, we state our main results. We bound the number of orbits with the same second moment, and then derive conditions for there to be a one-to-one correspondence between the second moment and an orbit. The results are proved in Section 4. Before that, we motivate the beltway problem over orthogonal groups by relating it to the single-particle cryo-electron microscopy technology.

2.2. Motivation: Single-particle cryo-electron microscopy

The goal of single-particle cryo-electron microscopy is to estimate a 3-D function that represents the electrostatic potential of a biological molecule. The measurements are tomographic projections of the molecule, each rotated by a random element of the group of 3-D rotations $\mathsf{SO}(3)$ [13]. While tomographic projection is not a group action, it was shown that the second moment of cryo-EM is equivalent to (2.1) with $G = \mathsf{SO}(3)$ [17, 18]. Remarkably, it was proven that, in the high noise regime, the minimal number of samples required for accurate estimation (i.e., the sample complexity) is proportional to the lowest order moment that determines the molecular structure [19]. Moreover, it was shown that generic molecular structures are determined only by the third moment, implying that the sample complexity scales rapidly [20, 21, 22]. Thus, it is essential to identify classes of molecular structures that can be recovered from lower-order moments. In this context, [18, 14] have shown that if a molecular structure can be represented with only a few coefficients under some basis, then it can be recovered from the second moment, and hence fewer samples are necessary for recovery (i.e., an improved sample complexity). These findings were recently extended to any semi-algebraic prior, including sparsity and deep generative models [23, 24]. Using only the second moment also reduces the computational burden, and thus it also inspired designing a new class of algorithms, based on the method of moments [17, 25, 26, 14].

3. Main result

Before stating our main result, we introduce two definitions: an extension of the classical collision-free condition from the circle to $\mathbb{R}^n$, and the *radially collision-free* condition. The condition that a set $S = \{t_1, \ldots, t_k\}$ of points in $\mathbb{R}^n$ is radially collision-free is the assertion that all points of S have distinct magnitudes, a hypothesis which was used to prove a similar result for cryo-EM in [14]. To motivate the extension of the collision-free definition, note that a periodic signal in $\mathbb{R}^n$ can be viewed as a discrete function on S^1, supported at a subset S of the n-th roots of unity. On the circle, this means that if (t_i, t_j) and (t_ℓ, t_m) are two pairs of points in S, then there is no $g \in \mathsf{O}(2)$ such that $g \cdot \{t_i, t_j\} = \{t_\ell, t_m\}$, unless $\{i, j\} = \{\ell, m\}$. The definition below extends this idea from $\mathsf{O}(2)$ to $\mathsf{O}(n)$.

Definition 3.1. Let $S = \{t_1, \ldots, t_k\} \subset \mathbb{R}^n$ be a set of points.

1. We say that S is *collision-free* if for every pair of two-element subsets $\{t_i, t_j\}$ and $\{t_\ell, t_m\}$, $g \cdot \{t_i, t_j\} = \{t_\ell, t_m\}$ for some $g \in \mathsf{O}(n)$ if and only if $\{t_i, t_j\} = \{t_\ell, t_m\}$.
2. We say that S is *radially collision-free* if for every $t_i, t_j \in S$, $t_i = g \cdot t_j$ for some $g \in \mathsf{O}(n)$ if and only if $t_i = t_j$.

Clearly, any radially collision-free set is collision-free, but the converse need not be true. Note also that the radially collision-free condition has no analog in the classical beltway problem, where all points lie on a circle.

Our main result states that for any collision-free δ-function of the form (2.2), there are a finite number of $\mathsf{O}(n)$-orbits of δ-functions with the same second moment as x. Moreover, we provide an explicit upper bound for the number of possible $\mathsf{O}(n)$-orbits with the same second moment. Notably, this bound implies that if the support of x is radially collision-free, then the second moment uniquely determines the $\mathsf{O}(n)$-orbit of x.

To state our main result, we introduce some notation. Suppose that x is a δ-function of the form (2.2). Let $\{z_1, \ldots z_q\}$ be the set of distinct magnitudes of the vectors t_i, and suppose that r_p of the vectors have magnitude z_p with $\sum_p r_p = k$. In particular, the support of x is radially collision-free if and only if $q = k$, so all $r_p = 1$.

Theorem 3.2. *If $k \geq 3$, the number of $\mathsf{O}(n)$-orbits of δ-functions of the form* (2.2) *with the same second moment as $x = \sum_{i=1}^{k} w_i \delta_{t_i}$, whose support is collision-free, is at most*

$$\prod_{r_p \geq 2} \frac{\binom{r_p}{2}!}{r_p!} \prod_{a<b\leq q} (r_a r_b)! \tag{3.1}$$

In particular, if the support of x is radially collision-free (i.e., all $r_p = 1$), then the second moment determines x up to a translation by an element of $\mathsf{O}(n)$.

Importantly, Theorem 3.2 is independent of the non-zero weights w_i. Thus, it holds for the binary case when the weights are all ones—as in the classical beltway problem—as well as when the weights are either generic [7] or drawn from a finite alphabet [6]. When the weights are sufficiently generic, we can derive a stronger result: the orbit of any signal with collision-free support (not necessarily radially collision-free) is determined by the second moment.

Corollary 3.3. *Suppose that a δ-function of the form* (2.2) *has a collision-free support. If the pairwise products of the weights $\{w_i w_j\}_{i<j}$ are all distinct, then the second moment determines the $\mathsf{O}(n)$-orbit of x.*

Note that if the weights w_i take negative values, then the second moment is invariant to sign, namely, x and $-x$ result in the same second moment. Thus, in this case, x is at best determined up to an element of $\mathsf{O}(n) \times \mathbb{Z}_2$.

4. Proofs of Theorem 3.2 and Corollary 3.3

4.1. Characterization of the second moment in terms of Gram matrices

The set $\mathcal{V}_{n,k}$ of δ-functions supported at k-points can be identified with the quotient algebraic variety $U(n,k)/\Sigma_k$, where $U(n,k)$ is the set of k-tuples $((w_1, t_1), \ldots, (w_k, t_k))$, the t_k are distinct points in $\mathbb{R}^n$, none of the weights w_i are zero, and Σ_k is the symmetric group of permutations of $\{1, \ldots, k\}$.

The quotient map $\pi \colon U(n,k) \to \mathcal{V}_{n,k}$ is given explicitly by the formula

$$((w_1, t_1), \ldots, (w_k, t_k)) \mapsto \sum_{i=1}^{k} w_i \delta_{t_i}.$$

The set $U(n,k)$ is an $\mathsf{O}(n)$-invariant Zariski open set in the representation $V_{n,k} = (\mathbb{R} \times \mathbb{R}^n)^k$ of $\mathsf{O}(n)$, where $g \in O(n)$ acts by the rule $g \cdot ((w_1, t_1), \ldots, (w_k, t_k)) = ((w_1, g \cdot t_1), \ldots, (w_k, g \cdot t_k))$. With this action, the quotient map $\pi \colon U_{n,k} \to \mathcal{V}_{n,k}$ is $\mathsf{O}(n)$ equivariant, meaning that $\pi(g \cdot ((w_1, t_1), \ldots, (w_k, t_k)) = g \cdot \pi\left((w_1, t_1), \ldots, (w_k, t_k)\right)$.

Our strategy for proving Theorem 3.2 is to use the $\mathsf{O}(n)$-equivariance of the map π to relate the second moment of $\delta = \pi((w_1, t_1), \ldots, (w_k, t_k))$, as defined in (2.3), to the representation-theoretic second moment of the vector $((w_1, t_1), \ldots, (w_k, t_k)) \in V_{n,k}$, as defined in [18, Section 2.3]. To simplify notation, we denote elements of $V_{n,k}$ by pairs (W, X), where W is the $1 \times k$ matrix of weights $(w_1, \ldots, w_k)$ and $X = (t_1 \ldots t_k)$ is the $n \times k$ matrix whose columns are the vectors $t_1, \ldots, t_k$. The vector space $V_{n,k}$ decomposes as a representation of $\mathsf{O}(n)$, $V_{n,k} = V_0^k + V_1^k$, where V_0 is the trivial representation (i.e., $\mathsf{O}(n)$ acts trivially) and V_1 is the *defining representation* of $\mathsf{O}(n)$, corresponding to $\mathsf{O}(n)$ acting as rigid motions of $\mathbb{R}^n$. Under this decomposition, the matrix W corresponds to the projection to V_0^k, and the matrix X corresponds to the projection to V_1^k.

By [18, Theorem 2.3], the second moment of a pair $(W, X) \in V_{n,k}$ is equivalent to the pair of symmetric $k \times k$ matrices $(W^T W, X^T X)$. Thus, we can identify the second moment on $V_{n,k}$ as the function $m_2 \colon V_{n,k} \to \mathrm{Sym}^k \mathbb{R} \times \mathrm{Sym}^k \mathbb{R}, (W, X) \mapsto (W^T W, X^T X)$. By [18, Theorem 2.3], the ambiguity group of the second moment on the representation $V_{n,k} = V_0^k \oplus V_1^k$ is the group $O(1) \times O(n)$, since V_0 and V_1 have dimensions 1 and n, respectively. (Note that in our formulation the ambiguity groups are orthogonal groups rather than unitary groups because we work with real representations.) In other words, $m_2(W', X') = m_2(W, X)$ if and only if $W' = \pm W$ and $X' = gX$ for some $g \in \mathsf{O}(n)$. However, because we assume that the weight vector W is positive we can ignore the sign ambiguity.

Definition 4.1. For a symmetric matrix A in $\mathbb{R}^{k \times k}$, and a weight vector $W \in \mathbb{R}^k$ define

$$\mathcal{T}(A, W) := \{(A_{ii}, A_{jj}, A_{ij}, w_i w_j) : 1 \leq i < j \leq k\}.$$

Proposition 4.2. *If $x = \sum_{i=1}^k w_i \delta_{t_i} \in \mathcal{V}_{n,k}$ is collision free, then the second moment $m_2(x)(\tau_1, \tau_2)$ is equivalent to the data $\mathcal{T}(X^T X, W)$ where $W = (w_1, \ldots, w_k)$.*

Proof. As explained in Section 2.1, the second moment is the distribution on $\mathbb{R}^n \times \mathbb{R}^n$

$$m_2(x)(\tau_1, \tau_2) = \sum_{i,j=1}^k w_i w_j \mu_{t_i, t_j}(\tau_1, \tau_2),$$

where $\mu_{t_i,t_j}(\tau_1, \tau_2)$ is supported on the orbit $\overline{(t_i, t_j)}$. If we assume that the points $t_1, \ldots, t_k$ are collision free, then the set of $\mathsf{O}(n)$-orbits $\{\overline{(t_i, t_j)}\}_{i \leq j}$ are all distinct, so we can determine this set of $\mathsf{O}(n)$-orbits from the support of the distribution $m_2(x)(\tau_1, \tau_2)$ on $(\mathbb{R}^n)^2$, and we can likewise determine the products of the weights $w_i w_j$ by evaluating $m_2(x)(\tau_1, \tau_2)$ on a smooth approximation to thecharacteristic function $\mathbf{1}_{\overline{(t_i,t_j)}}$ that equals 1 on $\overline{(t_i, t_j)}$. On the other hand, the $\mathsf{O}(n)$-orbit of the pair (t_i, t_j) is uniquely determined by the triple of real numbers $(t_i \cdot t_i, t_j \cdot t_j, t_i \cdot t_j)$. Thus, if $A = X^T X$ is the Gram matrix of X, then the orbit $\overline{(t_i, t_j)}$ determines the triple of entries (A_{ii}, A_{jj}, A_{ij}). Conversely, given the set $\mathcal{T}(XX^T, W)$ there is a unique $O(n)$-orbit $\overline{(t_i, t_j)}$ such that $(s_i \cdot s_i, s_j \cdot s_j, s_i \cdot s_j) = (A_{ii}, A_{jj}, A_{ij})$ for all $(s_1, s_2) \in \overline{(t_i, t_j)}$. Thus, we can reconstruct the distribution $\sum_{i,j=1}^k w_i w_j \mu_{t_i,t_j}(\tau_1, \tau_2)$ from the data $\mathcal{T}(XX^T, W)$. □

4.2. Proof of Theorem 3.2

Proof. Proposition 4.2 implies that if the support of x is collision-free, then there are a finite number of $\mathsf{O}(n)$-orbits of δ-functions y with the same second moment. This follows since if $y = \sum_{i=1}^k w_i' \delta_{s_i} \in \mathcal{V}_{n,k}$ is another δ-function with the same second moment as $x = \sum w_i \delta_{t_i}$, then $\mathcal{T}(X^T X, W) = \mathcal{T}(Y^T Y, W')$ where $(W, X) = ((w_1, \ldots, w_k), (t_1 \ \ldots \ t_k))$ and $(W', Y) = ((w_1', \ldots, w_k'), (s_1 \ \ldots \ s_k))$. We complete the proof of Theorem 3.2 by bounding this number.

Since $A = X^TX$ and $B = Y^TY$ have the same set of diagonal entries with multiplicity, we can, after reordering the supports, assume that $\|t_i\| = \|s_i\|$ for all i and that $\|t_i\| \le \|t_{i+1}\|$ for $i = 1, \dots, k$. By assumption, $\|t_i\|$ takes on q different values as i ranges from 1 to k. Let $\mathcal{P}_1, \dots, \mathcal{P}_q$ be the partition of $\{1, \dots, k\}$ such that $\|t_j\|$ is constant for all $j \in \mathcal{P}_\ell$ and $|\mathcal{P}_\ell| = r_\ell$.

The symmetric matrices $A = X^TX$ and $B = Y^TY$ have the same set of entries and identical diagonals. For any $i \le j$, the triplet (A_{ii}, A_{jj}, A_{ij}) represents an orbit in the support of $m_2(x)$. Likewise, for any $k \le l$, the triple (B_{kk}, B_{ll}, B_{kl}) represents an orbit in the support of $m_2(y)$. Since x and y have the same second moments, the sets of triples $\{(A_{ii}, A_{jj}, A_{ij})\}_{i,j}$ and $\{(B_{kk}, B_{ll}, B_{kl})\}_{k,l}$ must be the same. Let B_{kl} be an entry of B with $k \le l$, and with $k \in \mathcal{P}_a, l \in \mathcal{P}_b$. Because we have ordered the diagonal, we know that $a \le b$ and that if $B_{kl} = A_{ij}$ for some $i \le j$, then $(i, j) \in \mathcal{P}_a \times \mathcal{P}_b$. Hence, if $a < b$, the entry B_{kl} can take $r_a r_b$ possible values corresponding to pairs in the product $\mathcal{P}_a \times \mathcal{P}_b$. On the other hand, if $a = b$, then B_{kl} can take $\binom{r_a}{2}$ possible values corresponding to pairs $i \le j \in \mathcal{P}_a \times \mathcal{P}_a$. Thus, there are at most $\prod_{r_p \ge 2} \binom{r_p}{2}! \prod_{a<b\le q} (r_n r_m)!$ possible matrices $B = Y^TY$ corresponding to Gram matrices of δ-functions y with the same second moment as x. However, if we reorder the points t_i with the same magnitude we do not change the diagonal and obtain a matrix Y which is a permutation of X, and the Gram matrix $B = Y^TY$ corresponds to a δ-function y with the same $\mathsf{O}(n)$-orbit as x. Since there are $\prod_{p=1}^q r_p!$ reorderings of the support set $t_1, \dots, t_k$ which preserve the magnitudes, we see that there are at most $\prod_{r_p \ge 2} \frac{\binom{r_p}{2}!}{r_p!} \prod_{a<b\le q} (r_a r_b)!$ possible $\mathsf{O}(n)$-orbits of δ-functions with the same second moment as x. □

Remark 4.3. Note that the bound given in (3.1) merely determines the number of permutations of $A = X^TX$, which could possibly be the Gram matrix of a matrix Y, which is not a permutation of X. However, in any example, the number of possible matrices Y with $\mathcal{T}(Y^TY) = \mathcal{T}(X^TX)$ will be further limited by the fact that the permutation of A must necessarily be a positive semidefinite matrix of rank ℓ, where ℓ is the dimension of the subspace spanned by the vectors in the support of the function x. However, we do not know a way to use this constraint to reduce the bound uniformly over all collision-free $x \in \mathcal{V}_{n,k}$; see Example 5.5.

4.3. Proof of Corollary 3.3

Proof. If $x = \sum_{i=1}^k w_i \delta_{t_i}$, then by Proposition 4.2 the autocorrelation $m_2(x)(\tau_1, \tau_2)$ determines the $\mathsf{O}(n)$-orbits $\overline{(t_i, t_j)}$ and the values of $w_i w_j$. In particular, if all pairwise products $w_i w_j$ are distinct, then any δ-function $y = \sum w_i' s_i$ with the same second moment as x must have (after possibly reordering the support point s_i), $w_i' = w_i$ and $\overline{t_i, t_j} = \overline{s_i, s_j}$, since $\overline{t_i, t_j}$ is the unique orbit where the integral with respect to $m_2(x)(\tau_1, \tau_2)$ of a smooth approximation to its characteristic function is $w_i w_j$ and we assume that the weights are positive. In particular, if $X = (t_1; \dots\ t_k)$ and $Y = (s_1\ \dots\ s_k)$, then for any pair $i \le j$, $(X^TX)_{ij} = (Y^TY)_{ij}$; i.e., X and Y have the same Gram matrices so x and y are orthogonally equivalent. □

5. The beltway problem over orthogonal groups for binary signals

Theorem 3.2 implies that the $\mathsf{O}(n)$-orbit of any δ-function whose support is radially collision-free is uniquely determined by its second moment. Likewise, by Corollary 3.3, if the support of x is only collision-free but the pairwise products of the weights are distinct, then the $\mathsf{O}(n)$-orbit of x is determined from the second moment. In this section, we discuss the problem of determining a *binary* δ-function in $\mathbb{R}^n$ from its autocorrelation over the orthogonal group $\mathsf{O}(n)$, when the support is collision-free but not radially collision-free. In this case, the situation is more nuanced, and we divide the discussion into four subsections. In Section 5.1 we prove that if the support of x consists of linearly independent points, then we cannot expect to recover the $\mathsf{O}(n)$-orbit of x from its second moment. In Sections 5.2 and 5.3 we discuss the case where the support set lies on a sphere. In Section 5.2 we prove that if $k \leq n$, then we cannot determine the $\mathsf{O}(n)$-orbit of x from its second moment. Conversely, in Section 5.3, we establish a conjecture that if $k > n$ then an $\mathsf{O}(n)$-orbit of a generic binary δ-function is determined by the second moment. Finally, in Section 5.4 we discuss the connection of this problem with the *turnpike* problem.

5.1. Linearly independent supports

In this section, we show that if the support of a binary δ-function x consists of linearly independent points in $\mathbb{R}^n$, then we cannot expect to recover the $\mathsf{O}(n)$-orbit of x from its second moment if at least two of the points have the same magnitude.

Proposition 5.1. *Consider the set of $\mathsf{O}(n)$-orbits of binary δ-functions $x = \sum_i^k \delta_{t_i}$, with $k \leq n$, such that $S = \{t_1, \ldots, t_k\}$ is collision-free and $\|t_i\| = \|t_j\|$ for some $i \neq j$. Then, there exists a Zariski dense subset $\mathcal{W}$ such that for all $x \in \mathcal{W}$, the $\mathsf{O}(n)$-orbit of x is not determined by its second moment.*

Proof. Let $\mathcal{U}$ be the set of δ-functions satisfying the hypotheses of the proposition. The subset of $\mathcal{U}' \subset \mathcal{U}$ corresponding to δ-functions x whose supports are linearly independent is Zariski open and hence dense. Thus, it suffices to prove that there is a Zariski dense subset of $\mathcal{U}'$ such that if $x \in \mathcal{U}'$, the $\mathsf{O}(n)$-orbit of x is not determined from its second moment.

Suppose $x = \sum_{i=1}^k \delta_{t_i} \in \mathcal{U}'$. After reordering the points, we may assume that $\|t_1\| = \|t_2\|$. Also, after applying a suitable orthogonal transformation, we may assume that the matrix $X = (t_1 \ldots t_k)$ is upper triangular; i.e., $t_i = (t_{i1}, \ldots, t_{ii}, 0, \ldots, 0)^T$ for $i = 1, \ldots, k$. In particular, we may assume that the last $n-k$ rows of X are zero. It follows that the $k \times k$ Gram matrix $X^T X$ is the same as the Gram matrix obtained by deleting the last $n-k$ rows of the triangular matrix X. In other words, we can, without loss of generality, assume that $k = n$.

Now, consider matrices of the form $Y = (t_1 \; \ldots t_{k-1} \; s_k)$. The Gram matrices $X^T X$ and $Y^T Y$ differ only in the k-th row and column. Consider the set of s_k, which satisfy the equations $t_1 \cdot s_k = t_2 \cdot t_k$, $t_2 \cdot s_k = t_1 \cdot t_k$, $t_i \cdot s_k = t_i \cdot t_k$ for $i = 3, \ldots, k-1$, and $s_k \cdot s_k = t_k \cdot t_k$. Because the matrix $X = (t_1 \ldots t_{k-1})$ is triangular, if we write $s_k = (s_{k1} \ldots s_{kk})^T$, the first $k-1$ equations in the system above give a unique solution for $(s_{k1}, \ldots, s_{k,k-1})$. The last coordinate s_{kk} is determined, up to a sign, by the equation $s_k \cdot s_k = t_k \cdot t_k$. However, this equation will not have a solution if $\sum_{i=1}^k t_{ki}^2 < \sum_{l=1}^{k-1} s_{kl}^2$. However, since $s_{k1}, \ldots s_{k,k-1}$ do not depend on t_{kk}, we know that for fixed $t_{k1}, \ldots, t_{k,k-1}$ and t_{kk} sufficiently large the system will have a solution. Moreover, this solution is unique up to the action of $\mathsf{O}(k)$. By construction, if $Y = (t_1 \; \ldots t_{k-1} \; s_k)$, $\mathcal{T}(Y^T Y) = \mathcal{T}(X^T X)$ but the Gram matrix $Y^T Y$ cannot be obtained from a permutation of X. Hence, if $y = \delta_{t_1} + \ldots + \delta_{t_{k-1}} + \delta_{s_k}$, then $m_2(y) = m_2(x)$ but y is not orthogonally equivalent to x. □

5.2. δ-functions supported on spheres with $k \leq n$

Let us consider binary δ-functions of the form $\sum_{i=1}^k \delta_{t_i}$, where $S = \{t_1, \ldots, t_k\}$ is a collision-free subset of points in S^{n-1}. By Theorem 3.2, we know that there can be up to $\frac{\binom{k}{2}!}{k!}$ possible non-orthogonally equivalent δ-functions with the same second moment as x (including x). When $k \leq n$, then a general collection of points on S^{n-1} are linearly independent and the method used in the proof of Proposition 5.1 yields the following result that bounds the number of solutions from below.

Proposition 5.2. *Consider the set of binary δ-functions on the sphere whose support is collision-free and assume $3 < k \leq n$. Then, there is a Zariski dense subset $\mathcal{W}$ so that if $x \in \mathcal{W}$, then there are at least $(k-1)!$ non-equivalent binary δ-functions with the same second moment as x (including x).*

Example 5.3. We performed the following numerical experiment with $k = n = 4$. We constructed $10,000$ random 4×4 triangular matrices $X = (t_1\ t_2\ t_3\ t_4)$ with each $\|t_i\| = 1$ as follows. We took $t_1 = (1, 0, 0, 0)^T$, $t_2 = (t_{21}, t_{22}, 0, 0)^T$ with (t_{21}, t_{22}) uniformly (Haar) sampled on S^1, $t_3 = (t_{31}, t_{32}, t_{33}, 0)^T$ with $(t_{31}, t_{32}, t_{33})^T$ uniformly sampled on S^2, and $t_4 = (t_{41}, t_{42}, t_{43}, t_{44})^T$ uniformly sampled on S^3. We found that in approximately 14% of the sampled matrices, any non-trivial permutation of the first three entries of the fourth column of the Gram matrix X^TX yields the Gram matrix of another triangular matrix $(t_1\ t_2\ t_3\ s_4)$ with $\|s_4\| = 1$; this yields δ-functions $y = \delta_{t_1} + \delta_{t_2} + \delta_{t_3} + \delta_{s_4}$ supported on S^3 which are not orthogonally equivalent to $x = \sum_{i=1}^{4} \delta_{t_i}$. Note that Theorem 3.2 implies that there are at most $6!/4! = 30$ possible δ-functions with the same second moment as $x = \delta_{t_1} + \delta_{t_2} + \delta_{t_3} + \delta_{t_4}$ (including x).

5.3. δ-functions supported on spheres with $k > n$

When $k > n$, the support of a δ-function consists of linearly dependent points. In this case, we pose the following conjecture.

Conjecture 5.4. *Suppose that $k > n$. Then, the $\mathsf{O}(n)$-orbit of a generic binary δ-function $x = \sum_{i=1}^{k} \delta_{t_i}$ with support on the sphere S^{n-1} is determined by its second moment.*

If $x = \sum_{i=1}^{k} \delta_{t_i}$ is supported on S^n and has a collision-free support, then the second moment determines all pairwise inner products $t_i \cdot t_j$. Since $\|t_i\| = 1$ for all i, this is equivalent to knowing the pairwise distance $\|t_i - t_j\|$. Thus, the problem of recovering the $\mathsf{O}(n)$-orbit of this δ-function is equivalent to recovering the $\mathsf{O}(n)$-orbit of $S = \{t_1, \ldots t_k\}$ from their pairwise distances. The problem of determining a set of points in $\mathbb{R}^n$ from their pairwise distances (up to rigid motions) was studied in [27], where the authors prove that if $k \geq n + 2$, then there is a hypersurface in $\mathrm{Sym}^k \mathbb{R}^n$ such that any set of k points in the complement of this hypersurface can be recovered, up to a rigid motion, from its set of pairwise distances. This result can be viewed as giving evidence that Conjecture 5.4 has a positive answer, at least when $k \geq n + 2$.

5.4. δ-functions supported on S^1 and the turnpike problem

The *turnpike problem* is the problem of recovering a set of points $S = \{a_1, \ldots, a_k\} \subset \mathbb{R}$, up to a translation and reflection, from their pairwise distances $|a_i - a_j|$. This problem arises in multiple applications, including angle-of-arrival estimation [28], identifiability of quantum systems [29], protein sequencing [30, 31, 32], error-correcting codes for polymer-based data storage [33] and DNA mapping [34, 35]. In the latter, the turnpike problem is known in the literature as the partial digest problem. The turnpike problem is equivalent to the problem of recovering a sparse signal from its aperiodic autocorrelation, which has been studied in depth in the phase retrieval literature (but is different from the crystallographic phase retrieval problem) [36, 37, 38]. In [1], the maximum possible number $H(k)$ of non-equivalent and homometric sets of k points in $\mathbb{R}$ was bounded by $\frac{1}{2}k^{0.8107144} \leq H(k) \leq \frac{1}{2}k^{1.2324827}$. If the support is collision-free, then the difference set always determines the points (up to a shift and reflection), with the exception of when the points t_i belong to an explicit 2-dimensional subspace of points when $k = 6$ [39, 40, 41, 42, 36].

If M is at least the maximum distance between the points in S, then the set S can be embedded in the half-circle by the map $a_i \mapsto t_i = (\cos(\pi a_i/M), \sin(\pi a_i/M))$. This follows since the t_i's lie in a half circle $0 \leq \pi(|a_i - a_j|/M) \leq \pi$ and the inner product $t_i \cdot t_j = \cos(\pi(a_i - a_j)/M)$ determines

$|a_i - a_j|$. Thus, the problem of recovering $S = \{t_1, \ldots, t_k\}$ from its pairwise distances is equivalent to the problem of recovering the $\mathsf{O}(2)$-orbit of the binary δ-function $x = \sum_{i=1}^{k} \delta_{t_i}$ from its second moment.

Note that the set $S = \{a_1, \ldots, a_k\} \subset \mathbb{R}$ is collision-free if and only if the corresponding set $T = \{t_1, \ldots, t_k\} \subset S^1$ is collision-free. The result of [39] implies that the only collision-free subsets of the real line that cannot be determined, up to a rigid motion, from their pairwise differences are six element sets of the form $P = \{0, a, b-2a, 2b-2a, 2b, 3b-a\}$ or $Q = \{0, a, 2a+b, a+2b, 2b-a, 3b-a\}$, where a, b are real numbers. In this case, the sets P, Q have the same difference sets but are not equivalent under rigid transformations. Translated to S^1, this implies that the only $\mathsf{O}(2)$-orbits of binary δ-functions with collision-free supports lying in a half-circle which cannot be determined from their second moments are the pairs of δ functions $x = \sum_{t_i \in T} \delta_{t_i}$ and $y = \sum_{s_i \in S} \delta_{s_i}$, where $S = \{(\cos(t/2M), \sin(t/2M))\}_{t \in P}$ and $T = \{(\cos(t/2M), \sin(t/2M))\}_{t \in Q}$.

Example 5.5. The sets of integer points $P = \{0, 1, 8, 11, 13, 17\}$ and $Q = \{0, 1, 4, 10, 12, 17\}$ are collision-free and have the same difference sets, but are not equivalent. If we embed these points in the half-circle by the map $\mathbb{Z} \to S^1$, $n \mapsto (\cos(\pi n/17), \sin(\pi n/17))$, then we obtain two sets of points $S = \{s_1, \ldots, s_6\}$ and $T = \{t_1, \ldots, t_6\}$ such that $x = \sum_{i=1}^{6} \delta_{t_i}$ and $y = \sum_{i=1}^{6} \delta_{s_i}$ are not orthogonally equivalent but have the same second moments. With this ordering of the points, the corresponding Gram matrices are

$$A = \begin{pmatrix} 1.0 & 0.98 & 0.74 & -0.27 & -0.60 & -1.0 \\ 0.98 & 1. & 0.85 & -0.092 & -0.45 & -0.98 \\ 0.74 & 0.85 & 1. & 0.45 & 0.092 & -0.74 \\ -0.27 & -0.092 & 0.45 & 1. & 0.93 & 0.27 \\ -0.60 & -0.45 & 0.092 & 0.93 & 1. & 0.60 \\ -1.0 & -0.98 & -0.74 & 0.27 & 0.60 & 1.0 \end{pmatrix}, \quad B = \begin{pmatrix} 1.0 & 0.98 & 0.092 & -0.45 & -0.74 & -1.0 \\ 0.98 & 1. & 0.27 & -0.27 & -0.60 & -0.98 \\ 0.092 & 0.27 & 1. & 0.85 & 0.60 & -0.092 \\ -0.45 & -0.27 & 0.85 & 1. & 0.93 & 0.45 \\ -0.74 & -0.60 & 0.60 & 0.93 & 1. & 0.74 \\ -1.0 & -0.98 & -0.092 & 0.45 & 0.74 & 1.0 \end{pmatrix}$$

which have the same set of entries, but one cannot be obtained from the other by the action of the permutation group S_6 which simultaneously permutes rows and columns.

By Theorem 3.2, there are at most $14!/6! \approx 1.21 \times 10^8$ possible δ-functions with the same second moment. However, in the example with six points on S^1 we actually only obtain two such functions. The reason is that in this case, there is only one non-trivial permutation of the Gram matrix $X^T X$ that can be factored as $Y^T Y$, where Y is a 2×6 matrix. Indeed, a general permutation of $X^T X$ need not be semi-definite nor have rank 2. For example, the matrix

$$C = \begin{pmatrix} 1.0 & 0.74 & 0.98 & -0.27 & -0.60 & -1.0 \\ 0.74 & 1. & 0.85 & -0.092 & -0.45 & -0.98 \\ 0.98 & 0.85 & 1. & 0.45 & 0.092 & -0.74 \\ -0.27 & -0.092 & 0.45 & 1. & 0.93 & 0.27 \\ -0.60 & -0.45 & 0.092 & 0.93 & 1. & 0.60 \\ -1.0 & -0.98 & -0.74 & 0.27 & 0.60 & 1.0 \end{pmatrix}$$

has eigenvalues $\{3.9, 2.1, 0.30, -0.28, 0, 0\}$, so it is not positive semi-definite and has rank 4.

Remark 5.6. The result of [39] is only relevant for binary δ-functions with collision free support lying in a half circle. We expect that there are other examples of binary δ-functions with collision-free support on S^1 which cannot be recovered from their second moments.

Acknowledgments

The authors are grateful to Tanya Christiansen for helpful discussions, as well as to the referee for finding an error in an earlier version of the manuscript. T.B. and D.E. are supported by the BSF grant no. 2020159. T.B. is also supported in part by the NSF-BSF grant no. 2019752, and the ISF grant no.1924/21. D.E. was also supported by NSF-DMS 2205626. O.M. was supported by the ARO W911NF-17-1-0512 and the Simons Foundation Math + X Investigator award 506976.

References

[1] S. S. Skiena, W. D. Smith, P. Lemke, Reconstructing sets from interpoint distances, in: Proceedings of the sixth annual symposium on Computational geometry, 1990, pp. 332–339.

[2] P. Lemke, S. S. Skiena, W. D. Smith, Reconstructing sets from interpoint distances, Discrete and Computational Geometry: The Goodman-Pollack Festschrift (2003) 597–631.

[3] T. Bendory, D. Edidin, Algebraic theory of phase retrieval, Not. AMS 69 (9) (2022) 1487–1495.

[4] V. Elser, The complexity of bit retrieval, IEEE Transactions on Information Theory 64 (1) (2017) 412–428.

[5] V. Elser, T.-Y. Lan, T. Bendory, Benchmark problems for phase retrieval, SIAM Journal on Imaging Sciences 11 (4) (2018) 2429–2455.

[6] T. Bendory, D. Edidin, I. Gonzalez, Finite alphabet phase retrieval, Applied and Computational Harmonic Analysis 66 (2023) 151–160.

[7] T. Bendory, D. Edidin, Toward a mathematical theory of the crystallographic phase retrieval problem, SIAM Journal on Mathematics of Data Science 2 (3) (2020) 809–839.

[8] D. Edidin, A. Suresh, The generic crystallographic phase retrieval problem, arXiv preprint arXiv:2307.06835 (2023).

[9] S. Ghosh, P. Rigollet, Sparse multi-reference alignment: Phase retrieval, uniform uncertainty principles and the beltway problem, Foundations of Computational Mathematics (2022) 1–48.

[10] P. Duxbury, C. Lavor, L. Liberti, L. L. de Salles-Neto, Unassigned distance geometry and molecular conformation problems, Journal of Global Optimization (2022) 1–10.

[11] Y. Disser, S. S. Skiena, Geometric reconstruction problems, in: Handbook of Discrete and Computational Geometry, Chapman and Hall/CRC, 2017, pp. 897–913.

[12] S. Huang, I. Dokmanić, Reconstructing point sets from distance distributions, IEEE Transactions on Signal Processing 69 (2021) 1811–1827.

[13] T. Bendory, A. Bartesaghi, A. Singer, Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities, IEEE signal processing magazine 37 (2) (2020) 58–76.

[14] T. Bendory, Y. Khoo, J. Kileel, O. Mickelin, A. Singer, Autocorrelation analysis for cryo-EM with sparsity constraints: Improved sample complexity and projection-based algorithms, Proceedings of the National Academy of Sciences 120 (18) (2023) e2216507120.

[15] L. Hörmander, The analysis of linear partial differential operators. I, Vol. 256 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1983, Distribution theory and Fourier analysis.

[16] M. Reed, B. Simon, I: Functional analysis, Vol. 1, Academic press, 1981.

[17] Z. Kam, The reconstruction of structure from electron micrographs of randomly oriented particles, Journal of Theoretical Biology 82 (1) (1980) 15–39.

[18] T. Bendory, D. Edidin, The sample complexity of sparse multireference alignment and single-particle cryo-electron microscopy, SIAM Journal on Mathematics of Data Science 6 (2) (2024) 254–282.

[19] E. Abbe, J. M. Pereira, A. Singer, Estimation in the group action channel, in: 2018 IEEE International Symposium on Information Theory (ISIT), IEEE, 2018, pp. 561–565.

[20] A. Perry, J. Weed, A. S. Bandeira, P. Rigollet, A. Singer, The sample complexity of multireference alignment, SIAM Journal on Mathematics of Data Science 1 (3) (2019) 497–517.

[21] T. Bendory, N. Boumal, C. Ma, Z. Zhao, A. Singer, Bispectrum inversion with application to multireference alignment, IEEE Transactions on signal processing 66 (4) (2017) 1037–1050.

[22] A. S. Bandeira, B. Blum-Smith, J. Kileel, J. Niles-Weed, A. Perry, A. S. Wein, Estimation under group actions: recovering orbits from invariants, Applied and Computational Harmonic Analysis 66 (2023) 236–319.

[23] T. Bendory, N. Dym, D. Edidin, A. Suresh, Phase retrieval with semi-algebraic and ReLU neural network priors, arXiv preprint arXiv:2311.08833 (2023).

[24] T. Bendory, N. Dym, D. Edidin, A. Suresh, A transversality theorem for semi-algebraic sets with application to signal recovery from the second moment and cryo-EM, arXiv preprint arXiv:2405.04354 (2024).

[25] N. Sharon, J. Kileel, Y. Khoo, B. Landa, A. Singer, Method of moments for 3D single particle ab initio modeling with non-uniform distribution of viewing angles, Inverse Problems 36 (4) (2020) 044003.

[26] E. Levin, T. Bendory, N. Boumal, J. Kileel, A. Singer, 3D ab initio modeling in cryo-EM by autocorrelation analysis, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1569–1573.

[27] M. Boutin, G. Kemper, Which point configurations are determined by the distribution of their pairwise distances?, International Journal of Computational Geometry & Applications 17 (01) (2007) 31–43.

[28] C. R. Karanam, B. Korany, Y. Mostofi, Magnitude-based angle-of-arrival estimation, localization, and target tracking, in: 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), IEEE, 2018, pp. 254–265.

[29] D. Burgarth, K. Yuasa, Identifiability of open quantum systems, Physical Review A 89 (3) (2014) 030302.

[30] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, S. Pan, On reconstructing a string from its substring compositions, in: 2010 IEEE International Symposium on Information Theory, IEEE, 2010, pp. 1238–1242.

[31] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, S. Pan, Quadratic-backtracking algorithm for string reconstruction from substring compositions, in: 2014 IEEE International Symposium on Information Theory, IEEE, 2014, pp. 1296–1300.

[32] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, S. Pan, String reconstruction from substring compositions, SIAM Journal on Discrete Mathematics 29 (3) (2015) 1340–1371.

[33] R. Gabrys, S. Pattabiraman, O. Milenkovic, Mass error-correction codes for polymer-based data storage, in: 2020 IEEE International Symposium on Information Theory (ISIT), IEEE, 2020, pp. 25–30.

[34] R. M. Karp, Mapping the genome: some combinatorial problems arising in molecular biology, in: Proceedings of the twenty-fifth annual ACM symposium on Theory of computing, 1993, pp. 278–285.

[35] S. S. Skiena, G. Sundaram, A partial digest approach to restriction site mapping, Bulletin of Mathematical Biology 56 (1994) 275–294.

[36] J. Ranieri, A. Chebira, Y. M. Lu, M. Vetterli, Phase retrieval for sparse signals: Uniqueness conditions, arXiv preprint arXiv:1308.3058 (2013).

[37] K. Jaganathan, Y. C. Eldar, B. Hassibi, Phase retrieval: An overview of recent developments, Optical Compressive Imaging (2016) 279–312.

[38] T. Bendory, R. Beinert, Y. C. Eldar, Fourier phase retrieval: Uniqueness and algorithms, in: Compressed Sensing and its Applications: Second International MATHEON Conference 2015, Springer, 2017, pp. 55–91.

[39] A. Bekir, S. W. Golomb, There are no further counterexamples to S. Piccard's theorem, IEEE transactions on information theory 53 (8) (2007) 2864–2867.

[40] G. S. Bloom, A counterexample to a theorem of S. Piccard, Journal of Combinatorial Theory, Series A 22 (3) (1977) 378–379.

[41] G. S. Bloom, S. W. Golomb, Applications of numbered undirected graphs, Proceedings of the IEEE 65 (4) (1977) 562–570.

[42] S. Piccard, Sur les ensembles de distances des emsembles de points d'un espace euclidien, Neuchâtel: Secrétariat de l'Université, 1939.