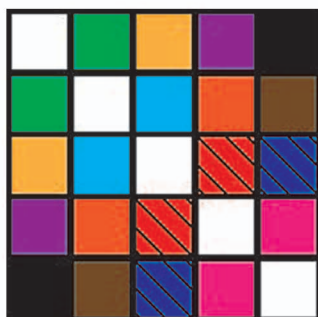


Euclidean Distance Matrices



[Essential theory, algorithms,
and applications]

Euclidean distance matrices (EDMs) are matrices of the squared distances between points. The definition is deceptively simple; thanks to their many useful properties, they have found applications in psychometrics, crystallography, machine learning, wireless sensor networks, acoustics, and more. Despite the usefulness of EDMs, they seem to be insufficiently known in the signal processing community. Our goal is to rectify this mishap in a concise tutorial. We review the fundamental properties of EDMs, such as rank or (non)definiteness, and show how the various EDM properties can be used to design algorithms for completing and denoising distance data. Along the way, we demonstrate applications to microphone position calibration, ultrasound tomography, room reconstruction from echoes, and phase retrieval. By spelling out the essential algorithms, we hope to fast-track the readers in applying EDMs to their own problems. The code for all of the described algorithms and to generate the figures in the article is available online at <http://lcav.epfl.ch/ivan.dokmanic>. Finally, we suggest directions for further research.

INTRODUCTION

Imagine that you land at Geneva International Airport with the Swiss train schedule but no map. Perhaps surprisingly, this may be sufficient to reconstruct a rough (or not so rough) map of the Alpine country, even if the train times poorly translate to distances or if some of the times are unknown. The way to do it

is by using EDMs; for an example, see “Swiss Trains (Swiss Map Reconstruction).”

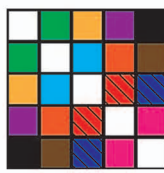
We often work with distances because they are convenient to measure or estimate. In wireless sensor networks, for example, the sensor nodes measure the received signal strengths of the packets sent by other nodes or the time of arrival (TOA) of pulses emitted by their neighbors [1]. Both of these proxies allow for distance estimation between pairs of nodes; thus, we can attempt to reconstruct the network topology. This is often termed *self-localization* [2]–[4]. The molecular conformation problem is another instance of a distance problem [5], and so is reconstructing a room’s geometry from echoes [6]. Less obviously, sparse phase retrieval [7] can be converted to a distance problem and addressed using EDMs.

Sometimes the data are not metric, but we seek a metric representation, as it happens commonly in psychometrics [8]. As a matter of fact, the psychometrics community is at the root of the development of a number of tools related to EDMs, including multidimensional scaling (MDS)—the problem of finding the best point set representation of a given set of distances. More abstractly, we can study EDMs for objects such as images, which live in high-dimensional vector spaces [9].

EDMs are a useful description of the point sets and a starting point for algorithm design. A typical task is to retrieve the original point configuration: it may initially come as a surprise that this requires no more than an eigenvalue decomposition (EVD) of a symmetric matrix. In fact, the majority of Euclidean distance problems require the reconstruction of the point set but always with one or more of the following twists:

- 1) The distances are noisy.
- 2) Some distances are missing.





3) The distances are unlabeled.

For examples of applications requiring solutions of EDM problems with different complications, see Figure 1.

There are two fundamental problems associated with distance geometry [10]: 1) given a matrix, determine whether it is an EDM and 2) given a possibly incomplete set of distances, determine whether there exists a configuration of points in a given embedding dimension—the dimension of the smallest affine space comprising the points—that generates the distances.

LITERATURE REVIEW

The study of point sets through pairwise distances, and that of EDMs, can be traced back to the works of Menger [11], Schoenberg [12], Blumenthal [13], and Young and Householder [14]. An important class of EDM tools was initially developed for the purpose of data visualization. In 1952, Torgerson introduced the notion of MDS [8]. He used distances to quantify the dissimilarities between pairs of objects that are not necessarily vectors in a metric space. Later, in 1964, Kruskal suggested the notion of stress as a measure of goodness of fit for nonmetric data [15], again representing experimental dissimilarities between objects.

A number of analytical results on EDMs were developed by Gower [16], [17]. In 1985 [17], he gave a complete characterization of the EDM rank. Optimization with EDMs requires adequate geometric intuitions about matrix spaces. In 1990, Glunt et al. [18] and Hayden et al. [19] provided insights into the structure of the convex cone of EDMs. An extensive treatise on EDMs with many original results and an elegant characterization of the EDM cone is provided by Dattorro [20].

In the early 1980s, Williamson, Havel, and Wüthrich developed the idea of extracting the distances between pairs of hydrogen atoms in a protein using nuclear magnetic resonance (NMR). The extracted distances were then used to reconstruct three-dimensional (3-D) shapes of molecules [5]. (Wüthrich received the Nobel Prize for chemistry in 2002.) The NMR spectrometer (together with some postprocessing) outputs the distances between the pairs of atoms in a large molecule. The distances are not specified for all atom pairs, and they are uncertain—i.e., given only up to an interval. This setup lends itself naturally to EDM treatment; for example, it can be directly addressed using MDS [21]. Indeed, the crystallography community also contributed a large number of important results on distance geometry. In a different biochemical application, comparing distance matrices yields efficient algorithms for comparing proteins from their 3-D structure [22].

In machine learning, one can learn manifolds by finding an EDM with a low embedding dimension that preserves the local geometry. Weinberger and Saul use it to learn image manifolds [9]. Other examples of using Euclidean distance geometry in machine learning are the results by Tenenbaum, De Silva, and Langford [23] on image understanding and handwriting recognition; Jain and

Saul [24] on speech and music; and Demaine et al. [25] on music and musical rhythms.

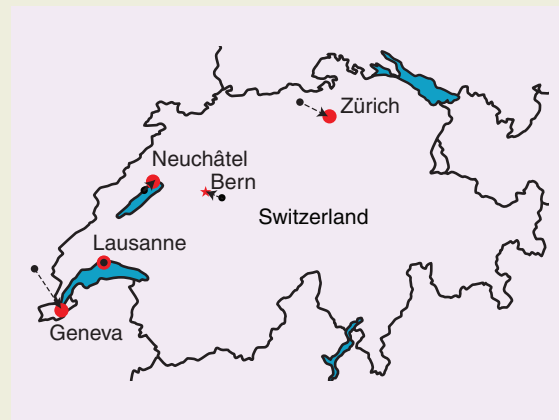
With the increased interest in sensor networks, several EDM-based approaches were proposed for sensor localization [2]–[4], [20]. The connections between EDMs, multilateration, and semidefinite programming are expounded in depth in [26], especially in the context of sensor network localization (SNL).

SWISS TRAINS (SWISS MAP RECONSTRUCTION)

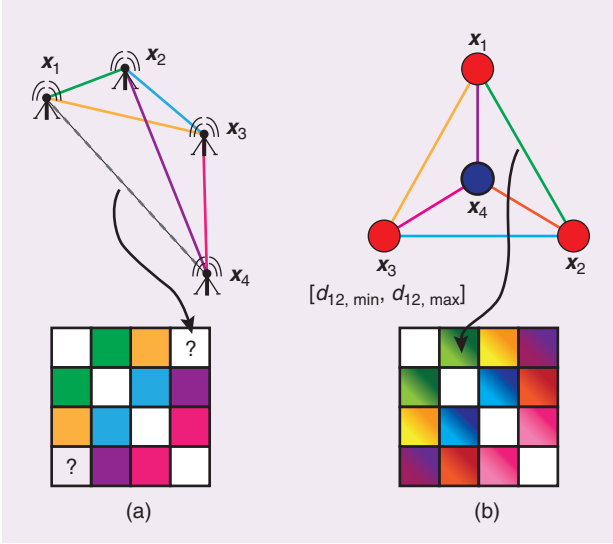
Consider the following matrix of the time in minutes it takes to travel by train between some Swiss cities (see Figure S1):

	L	G	Z	N	B
Lausanne	0	33	128	40	66
Geneva	33	0	158	64	101
Zürich	128	158	0	88	56
Neuchâtel	40	64	88	0	34
Bern	66	101	56	34	0

The numbers were taken from the Swiss railways timetable. The matrix was then processed using the classical MDS algorithm (Algorithm 1), which is basically an EVD. The obtained city configuration was rotated and scaled to align with the actual map. Given all of the uncertainties involved, the fit is remarkably good. Not all trains drive with the same speed, they have varying numbers of stops, and railroads are not straight lines (i.e., because of lakes and mountains). This result may be regarded as anecdotal, but, in a fun way, it illustrates the power of the EDM toolbox. Classical MDS could be considered the simplest of the available tools, yet it yields usable results with erroneous data. On the other hand, it might be that Swiss trains are just *that* good.



[FIGS1] A map of Switzerland with the true locations of five cities (red) and their locations estimated by using classical MDS on the train schedule (black).



[FIG1] Two real-world applications of EDMs. (a) SNL from estimated pairwise distances is illustrated with one distance missing because the corresponding sensor nodes are too far apart to communicate. (b) In the molecular conformation problem, we aim to estimate the locations of the atoms in a molecule from their pairwise distances. Here, because of the inherent measurement uncertainty, we know the distances only up to an interval.

Position calibration in ad hoc microphone arrays is often done with sources at unknown locations, such as hand claps, finger snaps, or randomly placed loudspeakers [27]–[29]. This gives us the distances (possibly up to an offset time) between the microphones and the sources and leads to the problem of multidimensional unfolding (MDU) [30].

All of the mentioned applications work with labeled distance data. In certain TOA-based applications, one loses the labels, i.e., the correct permutation of the distances. This issue arises when reconstructing the geometry of a room from echoes [6]. Another example of unlabeled distances is in sparse phase retrieval, where the distances between the unknown nonzero lags in a signal are revealed in its autocorrelation function (ACF) [7]. Recently, motivated by problems in crystallography, Gujarathi et al. published an algorithm for the reconstruction of Euclidean networks from unlabeled distance data [31].

OUR MISSION

We were motivated to write this tutorial after realizing that EDMs are not common knowledge in the signal processing community, perhaps for the lack of a compact introductory text. This is effectively illustrated by the anecdote that, not long before writing this article, one of the authors of this article had to add the (rather fundamental) rank property to the Wikipedia page on EDMs (search for “Euclidean distance matrix”). (We are working on improving that page substantially.) In a compact tutorial, we do not attempt to be exhaustive; much more thorough literature reviews are available in longer exposés on EDMs and distance geometry [10], [32], [33]. Unlike these works, which take the most general approach through graph realizations, we opt to show simple cases through examples and explain and spell out a

set of basic algorithms that anyone can use immediately. Two big topics that we discuss are not commonly treated in the EDM literature: localization from unlabeled distances and MDU (applied to microphone localization). On the other hand, we choose to not explicitly discuss the SNL problem as the relevant literature is abundant.

Implementations of all of the algorithms in this article are available online at <http://lcav.epfl.ch/ivan.dokmanic>. Our hope is that this will provide a solid starting point for those who wish to learn much more while inspiring new approaches to old problems.

FROM POINTS TO EDMs AND BACK

The principal EDM-related task is to reconstruct the original point set. This task is an inverse problem to the simpler forward problem of finding the EDM given the points. Thus, it is desirable to have an analytic expression for the EDM in terms of the point matrix. Beyond convenience, we can expect such an expression to provide interesting structural insights. We will define the notation as it becomes necessary—a summary is provided in Table 1.

Consider a collection of n points in a d -dimensional Euclidean space, ascribed to the columns of matrix $X \in \mathbb{R}^{d \times n}$, $X = [x_1, x_2, \dots, x_n]$, $x_i \in \mathbb{R}^d$. Then the squared distance between x_i and x_j is given as

$$d_{ij} = \|x_i - x_j\|^2, \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm. Expanding the norm yields

$$d_{ij} = (x_i - x_j)^\top (x_i - x_j) = x_i^\top x_i - 2x_i^\top x_j + x_j^\top x_j. \quad (2)$$

From here, we can read out the matrix equation for $D = [d_{ij}]$

$$\text{edm}(X) \stackrel{\text{def}}{=} \mathbf{1} \text{diag}(X^\top X)^\top - 2X^\top X + \text{diag}(X^\top X) \mathbf{1}^\top, \quad (3)$$

where $\mathbf{1}$ denotes the column vector of all ones and $\text{diag}(A)$ is the column vector of the diagonal entries of A . We see that $\text{edm}(X)$ is in fact a function of $X^\top X$. For later reference, it is convenient to define an operator $\mathcal{K}(G)$ similar to $\text{edm}(X)$, which operates directly on the Gram matrix $G = X^\top X$

$$\mathcal{K}(G) \stackrel{\text{def}}{=} \text{diag}(G) \mathbf{1}^\top - 2G + \mathbf{1} \text{diag}(G)^\top. \quad (4)$$

The EDM assembly formula (3) or (4) reveals an important property: because the rank of X is at most d (i.e., it has d rows), then the rank of $X^\top X$ is also at most d . The remaining two summands in (3) have rank one. By rank inequalities, the rank of a sum of matrices cannot exceed the sum of the ranks of the summands. With this observation, we proved one of the most notable facts about EDMs:

Theorem 1 (Rank of EDMs): *The rank of an EDM corresponding to points in \mathbb{R}^d is at most $d + 2$.*

This is a powerful theorem; it states that the rank of an EDM is independent of the number of points that generate it. In many

applications, d is three or less while n can be in the thousands. According to Theorem 1, the rank of such practical matrices is at most five. The proof of this theorem is simple, but, to appreciate that the property is not obvious, you may try to compute the rank of the matrix of nonsquared distances.

What really matters in Theorem 1 is the affine dimension of the point set, i.e., the dimension of the smallest affine subspace that contains the points, which is denoted by $\text{aff dim}(X)$. For example, if the points lie on a plane (but not on a line or a circle) in \mathbb{R}^3 , the rank of the corresponding EDM is four, not five. This will be made clear from a different perspective in the section “Essential Uniqueness,” as any affine subspace is just a translation of a linear subspace. An illustration for a one-dimensional (1-D) subspace of \mathbb{R}^2 is provided in Figure 2. Subtracting any point in the affine subspace from all of its points translates it to the parallel linear subspace that contains the zero vector.

ESSENTIAL UNIQUENESS

When solving an inverse problem, we need to understand what is recoverable and what is forever lost in the forward problem. Representing sets of points by distances usually increases the size of the representation. For most interesting n and d , the number of pairwise distances is larger than the size of the coordinate description, $(1/2)n(n-1) > nd$, so an EDM holds more scalars than the list of point coordinates. Nevertheless, some information is lost in this encoding such as the information about the absolute position and orientation of the point set. Intuitively, it is clear that rigid transformations (including reflections) do not change the distances between the fixed points in a point set. This intuitive fact is easily deduced from the EDM assembly formula (3). We have seen in (3) and (4) that $\text{edm}(X)$ is in fact a function of the Gram matrix $X^\top X$.

This makes it easy to show algebraically that rotations and reflections do not alter the distances. Any rotation/reflection can be represented by an orthogonal matrix $Q \in \mathbb{R}^{d \times d}$ acting on the points x_i . Thus, for the rotated point set $X_r = QX$, we can write

$$X_r^\top X_r = (QX)^\top (QX) = X^\top Q^\top QX = X^\top X, \quad (5)$$

where we invoked the orthogonality of the rotation/reflection matrix $Q^\top Q = I$.

Translation by a vector $b \in \mathbb{R}^d$ can be expressed as

$$X_t = X + b\mathbf{1}^\top. \quad (6)$$

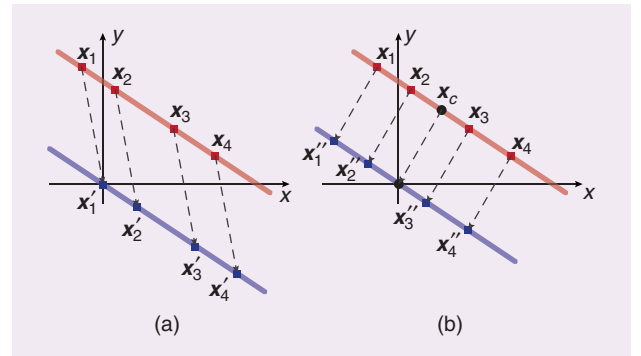
Using $\text{diag}(X^\top X_t) = \text{diag}(X^\top X) + 2X^\top b + \|b\|^2 \mathbf{1}$, one can directly verify that this transformation leaves (3) intact. In summary,

$$\text{edm}(QX) = \text{edm}(X + b\mathbf{1}^\top) = \text{edm}(X). \quad (7)$$

The consequence of this invariance is that we will never be able to reconstruct the absolute orientation of the point set using only the distances, and the corresponding degrees of freedom will be chosen freely. Different reconstruction procedures will lead to different realizations of the point set, all of them being rigid

[TABLE 1] A SUMMARY OF THE NOTATIONS.

SYMBOL	MEANING
n	NUMBER OF POINTS (COLUMNS) IN $X = [x_1, \dots, x_n]$
d	DIMENSIONALITY OF THE EUCLIDEAN SPACE
a_{ij}	ELEMENT OF A MATRIX A ON THE i TH ROW AND THE j TH COLUMN
D	AN EDM
$\text{edm}(X)$	AN EDM CREATED FROM THE COLUMNS IN X
$\text{edm}(X, Y)$	A MATRIX CONTAINING THE SQUARED DISTANCES BETWEEN THE COLUMNS OF X AND Y
$\mathcal{K}(G)$	AN EDM CREATED FROM THE GRAM MATRIX G
J	A GEOMETRIC CENTERING MATRIX
A_w	RESTRICTION OF A TO NONZERO ENTRIES IN w
W	MASK MATRIX, WITH ONES FOR OBSERVED ENTRIES
$S_{\mathbb{R}}^+$	A SET OF REAL SYMMETRIC POSITIVE-SEMIDEFINITE (PSD) MATRICES IN $\mathbb{R}^{n \times n}$
$\text{aff dim}(X)$	AFFINE DIMENSION OF THE POINTS LISTED IN X
$A \circ B$	HADAMARD (ENTRYWISE) PRODUCT OF A AND B
ε_{ij}	NOISE CORRUPTING THE (i, j) DISTANCE
e_i	i TH VECTOR OF THE CANONICAL BASIS
$\ A\ _F$	FROBENIUS NORM OF A , $(\sum_{ij} a_{ij}^2)^{1/2}$



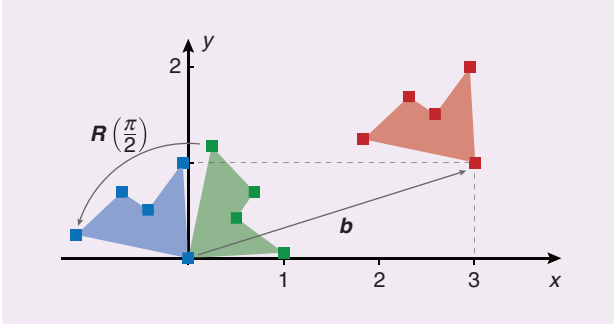
[FIG2] An illustration of the relationship between an affine subspace and its parallel linear subspace. The points $X = [x_1, \dots, x_4]$ live in an affine subspace—a line in \mathbb{R}^2 that does not contain the origin. In (a), the vector x_1 is subtracted from all the points, and the new point list is $X' = [0, x_2 - x_1, x_3 - x_1, x_4 - x_1]$. While the columns of X span \mathbb{R}^2 , the columns of X' only span a 1-D subspace of \mathbb{R}^2 —the line through the origin. In (b), we subtract a different vector from all points: the centroid $(1/4)X\mathbf{1}$. The translated vectors $X'' = [x'_1, \dots, x'_4]$ again span the same 1-D subspace.

transformations of each other. Figure 3 illustrates a point set under a rigid transformation; it is clear that the distances between the points are the same for all three shapes.

RECONSTRUCTING THE POINT SET FROM DISTANCES

The EDM equation (3) hints at a procedure to compute the point set starting from the distance matrix. Consider the following choice: let the first point x_1 be at the origin. Then, the first column of D contains the squared norms of the point vectors

$$d_{i1} = \|x_i - x_1\|^2 = \|x_i - 0\|^2 = \|x_i\|^2. \quad (8)$$



[FIG3] An illustration of a rigid transformation in 2-D. Here, the point set is transformed as $RX + b\mathbf{1}^\top$. The rotation matrix $R = [0 \ 1; -1 \ 0]$ (MATLAB notation) corresponds to a counterclockwise rotation of 90° . The translation vector is $b = [3, 1]^\top$. The shape is drawn for visual reference.

Consequently, we can construct the term $\mathbf{1} \text{diag}(X^\top X)$ and its transpose in (3), as the diagonal of $X^\top X$ contains exactly the norms squared $\|x_i\|^2$. Concretely,

$$\mathbf{1} \text{diag}(X^\top X) = \mathbf{1} d_1^\top, \quad (9)$$

where $d_1 = D\mathbf{e}_1$ is the first column of D . We thus obtain the Gram matrix from (3) as

$$G = X^\top X = -\frac{1}{2}(D - \mathbf{1} d_1^\top - d_1 \mathbf{1}^\top). \quad (10)$$

The point set can then be found by an EVD, $G = U\Lambda U^\top$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ with all eigenvalues λ_i nonnegative and U orthonormal, as G is a symmetric positive-semidefinite (PSD) matrix. Throughout this article, we assume that the eigenvalues are sorted in the order of decreasing magnitude, $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. We can now set $\widehat{X} \stackrel{\text{def}}{=} [\text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}), \mathbf{0}_{d \times (n-d)}] U^\top$. Note that we could have simply taken $\Lambda^{1/2} U^\top$ as the reconstructed point set, but if the Gram matrix really describes a d -dimensional point set, the trailing eigenvalues will be zeroes, so we choose to truncate the corresponding rows.

It is straightforward to verify that the reconstructed point set \widehat{X} generates the original EDM, $D = \text{edm}(X)$; as we have learned, \widehat{X} and X are related by a rigid transformation. The described procedure is called the *classical MDS*, with a particular choice of the coordinate system: x_1 is fixed at the origin.

In (10), we subtract a structured rank-2 matrix $(\mathbf{1} d_1^\top + d_1 \mathbf{1}^\top)$ from D . A more systematic approach to the classical MDS is to use a generalization of (10) by Gower [16]. Any such subtraction that makes the right-hand side of (10) PSD, i.e., that makes G a Gram matrix, can also be modeled by multiplying D from both sides by a particular matrix. This is substantiated in the following result.

Theorem 2 (Gower [16]): *D is an EDM if and only if*

$$-\frac{1}{2}(I - \mathbf{1} s^\top) D (I - s \mathbf{1}^\top) \quad (11)$$

is PSD for any s such that $s^\top \mathbf{1} = 1$ and $s^\top D \neq 0$.

Algorithm 1: The classical MDS.

```

1: function ClassicalMDS ( $D, d$ )
2:    $J \leftarrow I - (1/n)\mathbf{1}\mathbf{1}^\top$   $\triangleright$  Geometric centering matrix
3:    $G \leftarrow -(1/2)JDJ$   $\triangleright$  Compute the Gram matrix
4:    $U, [\lambda_i]_{i=1}^n \leftarrow \text{EVD}(G)$ 
5:   return  $[\text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}), \mathbf{0}_{d \times (n-d)}] U^\top$ 
6: end function

```

In fact, if (11) is PSD for one such s , then it is PSD for all of them. In particular, define the geometric centering matrix as

$$J \stackrel{\text{def}}{=} I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top. \quad (12)$$

Then, $-(1/2)JDJ$ being PSD is equivalent to D being an EDM. Different choices of s correspond to different translations of the point set.

The classical MDS algorithm with the geometric centering matrix is spelled out in Algorithm 1. Whereas so far we have assumed that the distance measurements are noiseless, Algorithm 1 can handle noisy distances too as it discards all but the d largest eigenvalues.

It is straightforward to verify that (10) corresponds to $s = \mathbf{e}_1$. Think about what this means in terms of the point set: $X\mathbf{e}_1$ selects the first point in the list, x_1 . Then, $X_0 = X(I - \mathbf{e}_1 \mathbf{1}^\top)$ translates the points so that x_1 is translated to the origin. Multiplying the definition (3) from the right by $(I - \mathbf{e}_1 \mathbf{1}^\top)$ and from the left by $(I - \mathbf{1} \mathbf{e}_1^\top)$ will annihilate the two rank-1 matrices, $\text{diag}(G)\mathbf{1}^\top$ and $\mathbf{1} \text{diag}(G)^\top$. We see that the remaining term has the form $-2X_0^\top X_0$, and the reconstructed point set will have the first point at the origin.

On the other hand, setting $s = (1/n)\mathbf{1}$ places the centroid of the point set at the origin of the coordinate system. For this reason, the matrix $J = I - (1/n)\mathbf{1}\mathbf{1}^\top$ is called the *geometric centering matrix*. To better understand why, consider how we normally center a set of points given in X : first, we compute the centroid as the mean of all the points,

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} X \mathbf{1}. \quad (13)$$

Second, we subtract this vector from all the points in the set

$$X_c = X - x_c \mathbf{1}^\top = X - \frac{1}{n} X \mathbf{1} \mathbf{1}^\top = X(I - \frac{1}{n} \mathbf{1} \mathbf{1}^\top). \quad (14)$$

In complete analogy with the reasoning for $s = \mathbf{e}_1$, we can see that the reconstructed point set will be centered at the origin.

ORTHOGONAL PROCRUSTES PROBLEM

Since the absolute position and orientation of the points are lost when going over to distances, we need a method to align the reconstructed point set with a set of anchors, i.e., points whose coordinates are fixed and known.

This can be achieved in two steps, sometimes called *Procrustes analysis*. Ascribe the anchors to the columns of Y , and suppose that we want to align the point set X with the columns of Y . Let X_a denote the submatrix (a selection of columns) of X that should be aligned with the anchors. We note that the number of anchors (the columns in X_a) is typically small compared with the total number of points (the columns in X).

In the first step, we remove the means y_c and $x_{a,c}$ from matrices Y and X_a , obtaining the matrices \bar{Y} , and \bar{X}_a . In the second step, termed *orthogonal Procrustes analysis*, we are searching for the rotation and reflection that best maps \bar{X}_a onto \bar{Y}

$$R = \arg \min_{Q: QQ^T = I} \|Q\bar{X}_a - \bar{Y}\|_F^2. \quad (15)$$

The Frobenius norm $\|\cdot\|_F$ is simply the ℓ^2 -norm of the matrix entries, $\|A\|_F^2 \stackrel{\text{def}}{=} \sum a_{ij}^2 = \text{trace}(A^T A)$.

The solution to (15), found by Schönemann in his Ph.D. thesis [34], is given by the singular value decomposition (SVD). Let $\bar{X}_a \bar{Y}^T = U\Sigma V^T$; then, we can continue computing (15) as follows:

$$\begin{aligned} R &= \arg \min_{Q: QQ^T = I} \|Q\bar{X}_a\|_F^2 + \|\bar{Y}\|_F^2 - \text{trace}(Y^T Q\bar{X}_a) \\ &= \arg \min_{\bar{Q}: \bar{Q}\bar{Q}^T = I} \text{trace}(\bar{Q}\Sigma), \end{aligned} \quad (16)$$

where $\bar{Q} \stackrel{\text{def}}{=} V^T Q U$ and we used the orthogonal invariance of the Frobenius norm and the cyclic invariance of the trace. The last trace expression in (16) is equal to $\sum_{i=1}^n \sigma_i \bar{q}_{ii}$. Noting that \bar{Q} is also an orthogonal matrix, its diagonal entries cannot exceed 1. Therefore, the maximum is achieved when $\bar{q}_{ii} = 1$ for all i , meaning that the optimal \bar{Q} is an identity matrix. It follows that $R = VU^T$.

Once the optimal rigid transformation has been found, the alignment can be applied to the entire point set as

$$R(X - x_{a,c} \mathbf{1}^T) + y_c \mathbf{1}^T. \quad (17)$$

COUNTING THE DEGREES OF FREEDOM

It is interesting to count how many degrees of freedom there are in different EDM-related objects. Clearly, for n points in \mathbb{R}^d , we have

$$\#_X = n \times d \quad (18)$$

degrees of freedom: if we describe the point set by the list of coordinates, the size of the description matches the number of degrees of freedom. Going from the points to the EDM (usually) increases the description size to $(1/2)n(n-1)$, as the EDM lists the distances between all the pairs of points. By Theorem 1, we know that the EDM has rank at most $d+2$.

Let us imagine for a moment that we do not know any other EDM-specific properties of our matrix except that it is symmetric, positive, zero-diagonal (or hollow), and that it has rank $d+2$. The purpose of this exercise is to count the degrees of freedom associated with such a matrix and to see if their number matches the intrinsic

number of the degrees of freedom of the point set, $\#_X$. If it did, then these properties would completely characterize an EDM. We can already anticipate from Theorem 2 that we need more properties: a certain matrix related to the EDM—as given in (11)—must be PSD. Still, we want to see how many degrees of freedom we miss.

We can do the counting by looking at the EVD of a symmetric matrix, $D = U\Lambda U^T$. The diagonal matrix Λ is specified by $d+2$ degrees of freedom because D has rank $d+2$. The first eigenvector of length n takes up $n-1$ degrees of freedom due to the normalization; the second one takes up $n-2$, as it is in addition orthogonal to the first one; for the last eigenvector, number $(d+2)$, we need $n-(d+2)$ degrees of freedom. We do not need to count the other eigenvectors because they correspond to zero eigenvalues. The total number is then

$$\begin{aligned} \#_{\text{DOF}} &= \underbrace{(d+2)}_{\text{Eigenvalues}} + \underbrace{(n-1) + \dots + [n-(d+2)]}_{\text{Eigenvectors}} - \underbrace{n}_{\text{Hollowness}} \\ &= n \times (d+1) - \frac{(d+1) \times (d+2)}{2}. \end{aligned}$$

For large n and fixed d , it follows that

$$\frac{\#_{\text{DOF}}}{\#_X} \sim \frac{d+1}{d}. \quad (19)$$

Therefore, even though the rank property is useful and we will show efficient algorithms that exploit it, it is still not a *tight* property (with symmetry and hollowness included). For $d=3$, the ratio (19) is $(4/3)$, so loosely speaking, the rank property has 30% too many determining scalars, which we need to set consistently. In other words, we need 30% more data to exploit the rank property than

we need to exploit the full EDM structure. We can assert that, for the same amount of data, the algorithms perform at least $\approx 30\%$ worse if we only exploit the rank property without EDMness.

The one-third gap accounts for various geometrical constraints that must be satisfied. The redundancy in the EDM representation is what makes denoising and completion algorithms possible, and thinking in terms of degrees of freedom gives us a fundamental understanding of what is achievable. Interestingly, the previous discussion suggests that for large n and large $d = o(n)$, little is lost by only considering rank.

Finally, in the previous discussion, for the sake of simplicity we ignored the degrees of freedom related to absolute orientation. These degrees of freedom, which are not present in the EDM, do not affect the large n behavior.

SUMMARY

Let us summarize what we have achieved in this section:

- We explained how to algebraically construct an EDM given the list of point coordinates.
- We discussed the essential uniqueness of the point set; information about the absolute orientation of the points is irretrievably lost when transitioning from points to an EDM.

MISSING ENTRIES ARISE BECAUSE
OF THE LIMITED RADIO RANGE
OR BECAUSE OF THE NATURE
OF THE SPECTROMETER.

- We explained classical MDS—a simple EVD-based algorithm (Algorithm 1) for reconstructing the original points—along with discussing parameter choices that lead to different centroids in reconstruction.
- Degrees of freedom provide insight into scaling behavior. We showed that the rank property is satisfactory, but there is more to it than just rank.

EDMs AS A PRACTICAL TOOL

We rarely have a perfect EDM. Not only are the entries of the measured matrix plagued by errors, but often we can measure just a subset. There are various sources of error in distance measurements: we already know that in NMR spectroscopy, we get intervals instead of exact distances. Measuring the distance using received powers or TOAs is subject to noise, sampling errors, and model mismatch.

Missing entries arise because of the limited radio range or because of the nature of the spectrometer. Sometimes the nodes in the problem at hand are asymmetric by definition; in microphone calibration, we have two types: microphones and calibration sources. This results in a particular block structure of the missing entries (see Figure 4 for an illustration).

It is convenient to have a single statement for both EDM approximation and EDM completion as the algorithms described in this section handle them at once.

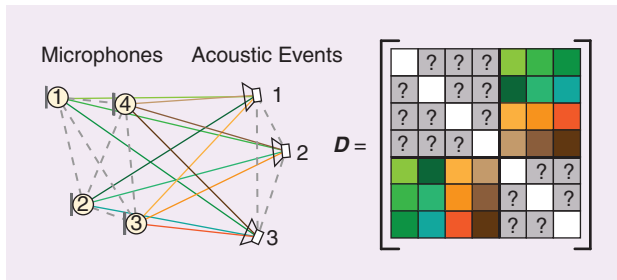
Problem 1: Let $D = \text{edm}(X)$. We are given a noisy observation of the distances between $p \leq (1/2)n(n-1)$ pairs of points from X . That is, we have a noisy measurement of $2p$ entries in D

$$\tilde{d}_{ij} = d_{ij} + \varepsilon_{ij}, \quad (20)$$

for $(i, j) \in E$, where E is some index set and ε_{ij} absorbs all errors. The goal is to reconstruct the point set \hat{X} in the given embedding dimension, so that the entries of $\text{edm}(\hat{X})$ are close in some metric to the observed entries \tilde{d}_{ij} .

To concisely write down completion problems, we define the mask matrix W as follows:

$$w_{ij} \stackrel{\text{def}}{=} \begin{cases} 1, & (i, j) \in E \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$



[FIG4] The microphone calibration as an example of MDU. We can measure only the propagation times from acoustic sources at unknown locations to microphones at unknown locations. The corresponding revealed part of the EDM has a particular off-diagonal structure, leading to a special case of EDM completion.

This matrix then selects elements of an EDM through a Hadamard (entrywise) product. For example, to compute the norm of the difference between the observed entries in A and B , we write $\|W \circ (A - B)\|$. Furthermore, we define the indexing A_W to mean the restriction of A to those entries where W is nonzero. The meaning of $B_W \leftarrow A_W$ is that we assign the observed part of A to the observed part of B .

EXPLOITING THE RANK PROPERTY

Perhaps the most notable fact about EDMs is the rank property established in Theorem 1: the rank of an EDM for points living in \mathbb{R}^d is at most $d+2$. This leads to conceptually simple algorithms for EDM completion and denoising. Interestingly, these algorithms exploit only the rank of the EDM. There is no explicit Euclidean geometry involved, at least not before reconstructing the point set.

We have two pieces of information: a subset of potentially noisy distances and the desired embedding dimension of the point configuration. The latter implies the rank property of the EDM that we aim to exploit. We may try to alternate between enforcing these two properties and hope that the algorithm produces a sequence of matrices that converges to an EDM. If it does, we have a solution. Alternatively, it may happen that we converge to a matrix with the correct rank that is not an EDM or that the algorithm never converges. The pseudocode is listed in Algorithm 2.

Algorithm 2: The alternating rank-based EDM completion.

```

1: function RankCompleteEDM( $W, \tilde{D}, d$ )
2:    $D_W \leftarrow \tilde{D}_W$  ▷ Initialize observed entries
3:    $D_{11^c-W} \leftarrow \mu$  ▷ Initialize unobserved entries
4:   repeat
5:      $D \leftarrow \text{EVThreshold}(D, d+2)$ 
6:      $D_W \leftarrow \tilde{D}_W$  ▷ Enforce known entries
7:      $D_I \leftarrow 0$  ▷ Set the diagonal to zero
8:      $D \leftarrow (D)_+$  ▷ Zero the negative entries
9:   until Convergence or MaxIter
10:  return  $D$ 
11: end function

12: function EVThreshold( $D, r$ )
13:    $U, [\lambda_i]_{i=1}^n \leftarrow \text{EVD}(D)$ 
14:    $\Sigma \leftarrow \text{diag}(\lambda_1, \dots, \lambda_r, \underbrace{0, \dots, 0}_{n-r \text{ times}})$ 
15:    $D \leftarrow U \Sigma U^T$ 
16:  return  $D$ 
17: end function

```

A different, more powerful approach is to leverage algorithms for low-rank matrix completion developed by the compressed sensing community. For example, OptSpace [35] is an algorithm for recovering a low-rank matrix from noisy, incomplete data. Let us take a look at how OptSpace works. Denote by $M \in \mathbb{R}^{m \times n}$ the rank- r matrix that we seek to recover, by $Z \in \mathbb{R}^{m \times n}$ the measurement noise, and by $W \in \mathbb{R}^{m \times n}$ the mask corresponding to the

measured entries; for simplicity, we chose $m \leq n$. The measured noisy and incomplete matrix is then given as

$$\widetilde{M} = W \circ (M + Z). \quad (22)$$

Effectively, this sets the missing (nonobserved) entries of the matrix to zero. OptSpace aims to minimize the following cost function:

$$F(A, S, B) \stackrel{\text{def}}{=} \frac{1}{2} \|W \circ (\widetilde{M} - ASB^T)\|_F^2, \quad (23)$$

where $S \in \mathbb{R}^{r \times r}$, $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{n \times r}$, such that $A^T A = B^T B = I$. Note that S need not be diagonal.

The cost function (23) is not convex, and minimizing it is a priori difficult [36] because of many local minima. Nevertheless, Keshavan, Montanari, and Oh [35] show that using the gradient descent method to solve (23) yields the global optimum with high probability, provided that the descent is correctly initialized.

Let $\widetilde{M} = \sum_{i=1}^m \sigma_i a_i b_i^T$ be the SVD of \widetilde{M} . Then, we define the scaled rank- r projection of \widetilde{M} as $\widetilde{M}_r \stackrel{\text{def}}{=} \alpha^{-1} \sum_{i=1}^r \sigma_i a_i b_i^T$. The fraction of observed entries is denoted by α so that the scaling factor compensates the smaller average magnitude of the entries in \widetilde{M} in comparison with M . The SVD of \widetilde{M}_r is then used to initialize the gradient descent, as detailed in Algorithm 3.

Algorithm 3: OptSpace [35].

```

1: function OptSpace( $\widetilde{M}$ ,  $r$ )
2:    $\widetilde{M} \leftarrow \text{Trim}(\widetilde{M})$ 
3:    $\widetilde{A}, \widetilde{\Sigma}, \widetilde{B} \leftarrow \text{SVD}(\alpha^{-1} \widetilde{M})$ 
4:    $A_0 \leftarrow$  First  $r$  columns of  $\widetilde{A}$ 
5:    $B_0 \leftarrow$  First  $r$  columns of  $\widetilde{B}$ 
6:    $S_0 \leftarrow \underset{S \in \mathbb{R}^{r \times r}}{\text{argmin}} F(A_0, S, B_0) \quad \triangleright \text{Eq. (23)}$ 
7:    $A, B \leftarrow \underset{A^T A = B^T B = I}{\text{argmin}} F(A, S_0, B) \quad \triangleright \text{See the note below}$ 
8:   return  $AS_0 B^T$ 
9: end function
 $\triangleright$  Line 7: gradient descent starting at  $A_0, B_0$ 
```

Two additional remarks are due in the description of OptSpace. First, it can be shown that the performance is improved by zeroing the overrepresented rows and columns. A row (respectively, column) is overrepresented if it contains more than twice the average number of observed entries per row (respectively, column). These heavy rows and columns bias the corresponding singular vectors and values, so (perhaps surprisingly) it is better to throw them away. We call this step “Trim” in Algorithm 3.

Second, the minimization of (23) does not have to be performed for all variables at once. In [35], the authors first solve the easier, convex minimization for S , and then with the optimizer S fixed, they find the matrices A and B using the gradient descent. These steps correspond to lines 6 and 7 of Algorithm 3. For an application of OptSpace in the calibration of ultrasound measurement rigs, see “Calibration in Ultrasound Tomography.”

CALIBRATION IN ULTRASOUND TOMOGRAPHY

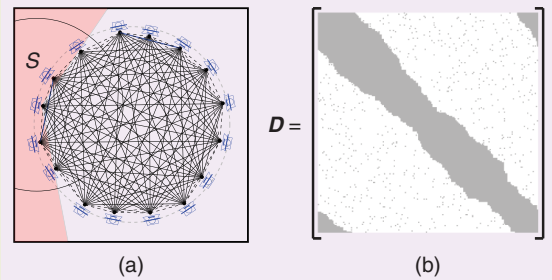
The rank property of EDMs, introduced in Theorem 1, can be leveraged in the calibration of ultrasound tomography devices. An example device for diagnosing breast cancer is a circular ring with thousands of ultrasound transducers placed around the breast [37]. The setup is shown in Figure S2(a).

Because of manufacturing errors, the sensors are not located on a perfect circle. This uncertainty in the positions of the sensors negatively affects the algorithms for imaging the breast. Fortunately, we can use the measured distances between the sensors to calibrate their relative positions. We can estimate the distances by measuring the times of flight (TOF) between pairs of transducers in a homogeneous environment, e.g., in water.

We cannot estimate the distances between all pairs of sensors because the sensors have limited beamwidths. (It is hard to manufacture omnidirectional ultrasonic sensors.) Therefore, the distances between the neighboring sensors are unknown, contrary to typical SNL scenarios where only the distances between nearby nodes can be measured. Moreover, the distances are noisy and some of them are unreliably estimated. This yields a noisy and incomplete EDM whose structure is illustrated in Figure S2(b).

Assuming that the sensors lie in the same plane, the original EDM produced by them would have a rank less than five. We can use the rank property and a low-rank matrix completion method, such as OptSpace (Algorithm 3), to complete and denoise the measured matrix [38]. Then, we can use the classical MDS in Algorithm 1 to estimate the relative locations of the ultrasound sensors.

For the reasons mentioned previously, SNL-specific algorithms are suboptimal when applied to ultrasound calibration. An algorithm based on the rank property effectively solves the problem and enables one to derive upper bounds on the performance error calibration mechanism, with respect to the number of sensors and the measurement noise. The authors in [38] show that the error vanishes as the number of sensors increases.



[FIGS2] (a) Ultrasound transducers lie on an approximately circular ring. The ring surrounds the breast and after each transducer fires an ultrasonic signal, the sound speed distribution of the breast is estimated. A precise knowledge of the sensor locations is needed to have an accurate reconstruction of the enclosed medium. (b) Because of the limited beamwidth of the transducers, noise, and imperfect TOF estimation methods, the measured EDM is incomplete and noisy. The gray areas show the missing entries of the matrix.

MULTIDIMENSIONAL SCALING

MDS refers to a group of techniques that, given a set of noisy distances, find the best fitting point conformation. It was originally proposed in psychometrics [8], [15] to visualize the (dis) similarities between objects. Initially, MDS was defined as the problem of representing distance data, but now the term is commonly used to refer to methods for solving the problem [39].

Various cost functions were proposed for solving MDS. In the section “Reconstructing the Point Set from Distances,” we already encountered one method: the classical MDS. This method minimizes the Frobenius norm of the difference between the input Gram matrix and the Gram matrix of the points in the target embedding dimension.

The Gram matrix contains inner products, but it is better to work directly with the distances. A typical cost function represents the dissimilarity of the observed distances and the distances between the estimated point locations. An essential observation is that the feasible set for these optimizations is not convex (i.e., EDMs with embedding dimensions smaller than $n - 1$ lie on the boundary of a cone [20], which is a nonconvex set).

A popular dissimilarity measure is *raw stress* [40], defined as the value of

$$\text{minimize}_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j) \in E} (\sqrt{\text{edm}(X)_{ij}} - \sqrt{d_{ij}})^2, \quad (24)$$

where E defines the set of revealed elements of the distance matrix D . The objective function can be concisely written as $\|W \circ (\sqrt{\text{edm}(X)} - \sqrt{D})\|_F^2$; a drawback of this cost function is that it is not globally differentiable. The approaches described in the literature comprise iterative majorization [41], various methods using convex analysis [42], and steepest descent methods [43].

Another well-known cost function, first studied by Takane, Young, and De Leeuw [44], is termed *s-stress*,

$$\text{minimize}_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j) \in E} (\text{edm}(X)_{ij} - d_{ij})^2. \quad (25)$$

Again, we write the objective concisely as $\|W \circ (\text{edm}(X) - D)\|_F^2$. Conveniently, the s-stress objective is globally differentiable, but a disadvantage is that it puts more weight on errors in larger distances than on errors in smaller ones. Gaffke and Mathar [45] propose an algorithm to find the global minimum of the s-stress function for embedding dimension $d = n - 1$. EDMs with this embedding dimension exceptionally constitute a convex set [20], but we are typically interested in embedding dimensions much smaller than n . The s-stress minimization in (25) is not convex for $d < n - 1$. It was analytically shown to have saddle points [46], but interestingly, no analytical nonglobal minimizer has been found [46].

Browne proposed a method for computing s-stress based on Newton–Raphson root finding [47]. Glunt reports that the method

by Browne converges to the global minimum of (25) in 90% of the test cases in his data set [48]. (While the experimental setup of Glunt [48] is not detailed, it was mentioned that the EDMs were produced randomly.)

The cost function in (25) is separable across points i and across coordinates k , which is convenient for distributed implementations. Parhizkar [46] proposed an alternating coordinate descent method that leverages this separability by updating a single coordinate of a particular point at a time. The s-stress function restricted to the

k th coordinate of the i th point is a fourth-order polynomial

$$f(x; \alpha^{(i,k)}) = \sum_{\ell=0}^4 \alpha_{\ell}^{(i,k)} x^{\ell}, \quad (26)$$

where $\alpha^{(i,k)}$ lists the polynomial coefficients for the i th point and the k th coordinate. For example, $\alpha_0^{(i,k)} = 4 \sum_j w_{ij}$, that is, four times the number of points connected to point i . Expressions for the remaining coefficients are given in [46]; in the pseudocode (Algorithm 4), we assume that these coefficients are returned by the function “GetQuadricCoeffs,” given the noisy incomplete matrix \bar{D} , the observation mask W , and the dimensionality d . The global minimizer of (26) can be found analytically by calculating the roots of its derivative (a cubic). The process is then repeated over all coordinates k and points i until convergence. The resulting algorithm is remarkably simple yet empirically converges fast. It naturally lends itself to a distributed implementation. We spell it out in Algorithm 4.

Algorithm 4: Alternating descent [46].

```

1: function AlternatingDescent( $\bar{D}, W, d$ )
2:    $X \in \mathbb{R}^{d \times n} \leftarrow X_0 = 0$  ▷ Initialize the point set
3:   repeat
4:     for  $i \in \{1, \dots, n\}$  do ▷ Points
5:       for  $k \in \{1, \dots, d\}$  do ▷ Coordinates
6:          $\alpha^{(i,k)} \leftarrow \text{GetQuadricCoeffs}(W, \bar{D}, d)$ 
7:          $x_{i,k} \leftarrow \text{argmin}_x f(x; \alpha^{(i,k)})$  ▷ Eq. (26)
8:       end for
9:     end for
10:  until Convergence or MaxIter
11:  return  $X$ 
12: end function

```

When applied to a large data set of random, noiseless, and complete distance matrices, Algorithm 4 converges to the global minimum of (25) in more than 99% of the cases [46].

SEMIDEFINITE PROGRAMMING

Recall the characterization of EDMs (11) in Theorem 2. It states that D is an EDM if and only if the corresponding geometrically centered Gram matrix $-(1/2)JDJ$ is PSD. Thus, it establishes a

THE S-STRESS OBJECTIVE
IS EVERYWHERE DIFFERENTIABLE,
BUT A DISADVANTAGE IS
THAT IT FAVORS LARGE
OVER SMALL DISTANCES.

one-to-one correspondence between the cone of EDMs, denoted by EDM^n and the intersection of the symmetric positive-semidefinite cone \mathbb{S}_+^n with the geometrically centered cone \mathbb{S}_c^n . The latter is defined as the set of all symmetric matrices whose column sum vanishes,

$$\mathbb{S}_c^n = \{G \in \mathbb{R}^{n \times n} \mid G = G^\top, G\mathbf{1} = \mathbf{0}\}. \quad (27)$$

We can use this correspondence to cast EDM completion and approximation as semidefinite programs. While (11) describes an EDM of an n -point configuration in any dimension, we are often interested in situations where $d \ll n$. It is easy to adjust for this case by requiring that the rank of the centered Gram matrix be bounded. One can verify that

$$\left. \begin{array}{l} D = \text{edm}(X) \\ \text{affdim}(X) \leq d \end{array} \right\} \iff \left\{ \begin{array}{l} -\frac{1}{2}JDJ \succeq 0 \\ \text{rank}(JDJ) \leq d, \end{array} \right. \quad (28)$$

when $n \geq d$. That is, EDMs with a particular embedding dimension d are completely characterized by the rank and definiteness of JDJ .

Now we can write the following rank-constrained semidefinite program for solving Problem 1:

$$\begin{array}{ll} \underset{G}{\text{minimize}} & \|W \circ (\bar{D} - \mathcal{K}(G))\|_F^2 \\ \text{subject to} & \text{rank}(G) \leq d \\ & G \in \mathbb{S}_+^n \cap \mathbb{S}_c^n. \end{array} \quad (29)$$

The second constraint is just shorthand for writing $G \succeq 0$, $G\mathbf{1} = \mathbf{0}$. We note that this is equivalent to MDS with the s-stress cost function thanks to the rank characterization (28).

Unfortunately, the rank property makes the feasible set in (29) nonconvex, and solving it exactly becomes difficult. This makes sense, as we know that s-stress is not convex. Nevertheless, we may relax the hard problem by simply omitting the rank constraint and hope to obtain a solution with the correct dimensionality:

$$\begin{array}{ll} \underset{G}{\text{minimize}} & \|W \circ (\bar{D} - \mathcal{K}(G))\|_F^2 \\ \text{subject to} & G \in \mathbb{S}_+^n \cap \mathbb{S}_c^n. \end{array} \quad (30)$$

We call (30) a semidefinite relaxation (SDR) of the rank-constrained program (29).

The constraint $G \in \mathbb{S}_c^n$, or equivalently, $G\mathbf{1} = \mathbf{0}$, means that there are no strictly positive definite solutions. (G has a nullspace, so at least one eigenvalue must be zero.) In other words, there exist no strictly feasible points [32]. This may pose a numerical problem, especially for various interior point methods. The idea is then to reduce the size of the Gram matrix through an invertible transformation, somehow removing the part of it responsible for the nullspace. In what follows, we describe how to construct this smaller Gram matrix.

A different, equivalent way to phrase the multiplicative characterization (11) is the following statement: a symmetric hollow matrix D is an EDM if and only if it is negative semidefinite on $\{\mathbf{1}\}^\perp$ (on all vectors \mathbf{t} such that $\mathbf{t}^\top \mathbf{1} = 0$). Let us construct an

orthonormal basis for this orthogonal complement—a subspace of dimension $(n-1)$ —and arrange it in the columns of matrix $V \in \mathbb{R}^{n \times (n-1)}$. We demand

$$\begin{aligned} V^\top \mathbf{1} &= \mathbf{0} \\ V^\top V &= I. \end{aligned} \quad (31)$$

There are many possible choices for V , but all of them obey that $VV^\top = I - (1/n)\mathbf{1}\mathbf{1}^\top = J$. The following choice is given in [2]:

$$V = \begin{bmatrix} p & p & \cdots & p \\ 1+q & q & \cdots & q \\ q & 1+q & \cdots & q \\ \vdots & \cdots & \ddots & \vdots \\ q & q & \cdots & 1+q \end{bmatrix}, \quad (32)$$

where $p = -1/(n + \sqrt{n})$ and $q = -1/\sqrt{n}$.

With the help of the matrix V , we can now construct the sought Gramian with reduced dimensions. For an EDM $D \in \mathbb{R}^{n \times n}$,

$$\mathcal{G}(D) \stackrel{\text{def}}{=} -\frac{1}{2}V^\top DV \quad (33)$$

is an $(n-1) \times (n-1)$ PSD matrix. This can be verified by substituting (33) in (4). Additionally, we have that

$$\mathcal{K}(V\mathcal{G}(D)V^\top) = D. \quad (34)$$

Indeed, $H \mapsto \mathcal{K}(VHV^\top)$ is an invertible mapping from \mathbb{S}_+^{n-1} to EDM^n whose inverse is exactly \mathcal{G} . Using these notations, we can write down an equivalent optimization program that is numerically more stable than (30) [2],

$$\begin{array}{ll} \underset{H}{\text{minimize}} & \|W \circ (\bar{D} - \mathcal{K}(VHV^\top))\|_F^2 \\ \text{subject to} & H \in \mathbb{S}_+^{n-1}. \end{array} \quad (35)$$

On the one hand, with the previous transformation, the constraint $G\mathbf{1} = \mathbf{0}$ became implicit in the objective, as $VHV^\top \mathbf{1} \equiv \mathbf{0}$ by (31); on the other hand, the feasible set is now the full semidefinite cone \mathbb{S}_+^{n-1} .

Still, as Krislock and Wolkowicz mention [32], by omitting the rank constraint, we allow the points to move about in a larger space, so we may end up with a higher-dimensional solution even if there is a completion in dimension d .

There exist various heuristics for promoting lower rank. One such heuristic involves the trace norm—the convex envelope of rank. The trace or nuclear norm is studied extensively by the compressed sensing community. In contrast to the common wisdom in compressed sensing, the trick here is to maximize the trace norm, not to minimize it. The mechanics are as follows: maximizing the sum of squared distances between the points will stretch the configuration as much as possible, subject to available constraints. But stretching favors smaller affine dimensions (e.g., imagine pulling out a roll of paper or stretching a bent string). Maximizing the sum of squared distances can be rewritten as maximizing the sum of norms in a centered point configuration—but that is exactly the trace of the Gram matrix $G = -(1/2)JDJ$ [9]. This idea has been

successfully put to work by Weinberger and Saul [9] in manifold learning and by Biswas et al. in SNL [49].

Noting that $\text{trace}(\mathbf{H}) = \text{trace}(\mathbf{G})$ because $\text{trace}(\mathbf{J}\mathbf{D}\mathbf{J}) = \text{trace}(\mathbf{V}^\top \mathbf{D}\mathbf{V})$, we write the following SDR:

$$\begin{aligned} & \underset{\mathbf{H}}{\text{maximize}} \quad \text{trace}(\mathbf{H}) - \lambda \|\mathbf{W} \circ (\bar{\mathbf{D}} - \mathcal{K}(\mathbf{V}\mathbf{H}\mathbf{V}^\top))\|_F \\ & \text{subject to} \quad \mathbf{H} \in \mathbb{S}_+^{n-1}. \end{aligned} \quad (36)$$

Here, we opted to include the data fidelity term in the Lagrangian form, as proposed by Biswas et al. [49], but it could also be moved to constraints. Finally, in all of the above relaxations, it is straightforward to include upper and lower bounds on the distances. Because the bounds are linear constraints, the resulting programs remain convex; this is particularly useful in the molecular conformation problem. A MATLAB/CVX [50], [51] implementation of the SDR (36) is given in Algorithm 5.

Algorithm 5: SDR (MATLAB/CVX).

```

1: function EDM = sdr_complete_edm(D, W, lambda)
2:
3: n = size(D, 1);
4: x = -1/(n + sqrt(n));
5: y = -1/sqrt(n);
6: V = [y*ones(1, n - 1); x*ones(n - 1) + eye(n - 1)];
7: e = ones(n, 1);
8:
9: cvx_begin sdP
10:  variable G(n - 1, n - 1) symmetric;
11:  B = V*G*V';
12:  E = diag(B)*e' + e*diag(B)' - 2*B;
13:  maximize trace(G) - lambda * norm(W .* (E - D), 'fro');
14:  subject to
15:      G >= 0;
16: cvx_end
17:
18: EDM = diag(B)*e' + e*diag(B)' - 2*B;
```

**MULTIDIMENSIONAL UNFOLDING:
A SPECIAL CASE OF COMPLETION**

Imagine that we partition the point set into two subsets and that we can measure the distances between the points belonging to different subsets but not between the points in the same subset. MDU [30] refers to this special case of EDM completion.

MDU is relevant for the position calibration of ad hoc sensor networks, particularly of microphones. Consider an ad hoc array of m microphones at unknown locations. We can measure the distances to k point sources, also at unknown locations, for example, by emitting a pulse. (We assume that the sources and the microphones are synchronized.) We can always permute the points so that the matrix assumes the

structure shown in Figure 4, with the unknown entries in two diagonal blocks. This is a standard scenario described, for example, in [27].

One of the early approaches to metric MDU is that of Schönemann [30]. We go through the steps of the algorithm and then explain how to solve the problem using the EDM toolbox. The goal is to make a comparison and emphasize the universality and simplicity of the introduced tools.

Denote by $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_m]$ the unknown microphone locations and by $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_k]$ the unknown source locations. The distance between the i th microphone and the j th source is

$$\delta_{ij} = \|\mathbf{r}_i - \mathbf{s}_j\|^2, \quad (37)$$

so that, in analogy with (3), we have

$$\Delta = \text{edm}(\mathbf{R}, \mathbf{S}) = \text{diag}(\mathbf{R}^\top \mathbf{R}) \mathbf{1}^\top - 2\mathbf{R}^\top \mathbf{S} + \mathbf{1} \text{diag}(\mathbf{S}^\top \mathbf{S}), \quad (38)$$

where we overloaded the edm operator in a natural way. We use Δ to avoid confusion with the standard Euclidean \mathbf{D} . Consider now two geometric centering matrices of sizes m and k , denoted as \mathbf{J}_m and \mathbf{J}_k . Similar to (14), we have

$$\mathbf{R}\mathbf{J}_m = \mathbf{R} - \mathbf{r}_c \mathbf{1}^\top, \quad \mathbf{S}\mathbf{J}_k = \mathbf{S} - \mathbf{s}_c \mathbf{1}^\top. \quad (39)$$

This means that

$$\mathbf{J}_m \Delta \mathbf{J}_k = \bar{\mathbf{R}}^\top \bar{\mathbf{S}} \stackrel{\text{def}}{=} \bar{\mathbf{G}} \quad (40)$$

is a matrix of the inner products between vectors $\tilde{\mathbf{r}}_i$ and $\tilde{\mathbf{s}}_j$. We used tildes to differentiate this from the real inner products between \mathbf{r}_i and \mathbf{s}_j because in (40), the points in $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}$ are referenced to different coordinate systems. The centroids \mathbf{r}_c and \mathbf{s}_c generally do not coincide. There are different ways to decompose $\bar{\mathbf{G}}$ into a product of two full rank matrices, call them \mathbf{A} and \mathbf{B}

$$\bar{\mathbf{G}} = \mathbf{A}^\top \mathbf{B}. \quad (41)$$

We could, for example, use the SVD, $\bar{\mathbf{G}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ and set $\mathbf{A}^\top = \mathbf{U}$ and $\mathbf{B} = \mathbf{\Sigma}\mathbf{V}^\top$. Any two such decompositions are linked by some invertible transformation $\mathbf{T} \in \mathbb{R}^{d \times d}$

$$\bar{\mathbf{G}} = \mathbf{A}^\top \mathbf{B} = \bar{\mathbf{R}}^\top \mathbf{T}^{-1} \mathbf{T} \bar{\mathbf{S}}. \quad (42)$$

We can now write down the conversion rule from what we can measure to what we can compute

$$\begin{aligned} \mathbf{R} &= \mathbf{T}^\top \mathbf{A} + \mathbf{r}_c \mathbf{1}^\top \\ \mathbf{S} &= (\mathbf{T}^{-1})^\top \mathbf{B} + \mathbf{s}_c \mathbf{1}^\top, \end{aligned} \quad (43)$$

where \mathbf{A} and \mathbf{B} can be computed according to (41). Because we cannot reconstruct the absolute position of the point set, we can arbitrarily set $\mathbf{r}_c = \mathbf{0}$, and $\mathbf{s}_c = \alpha \mathbf{e}_1$. Recapitulating, we have that

$$\Delta = \text{edm}(\mathbf{T}^\top \mathbf{A}, (\mathbf{T}^{-1})^\top \mathbf{B} + \alpha \mathbf{e}_1 \mathbf{1}^\top), \quad (44)$$

and the problem is reduced to computing T and α so that (44) holds, or in other words, so that the right-hand side is consistent with the data Δ . We reduced MDU to a relatively small problem: in 3-D, we need to compute only ten scalars. Schönmann [30] provides an algebraic method to find these parameters and mentions the possibility of least squares, while Crocco, Del Bue, and Murino [27] propose a different approach using nonlinear least squares.

This procedure seems quite convoluted. Rather, we see MDU as a special case of matrix completion, with the structure illustrated in Figure 4.

More concretely, represent the microphones and the sources by a set of $n = k + m$ points ascribed to the columns of matrix $X = [R \ S]$. Then, $\text{edm}(X)$ has a special structure as seen in Figure 4,

$$\text{edm}(X) = \begin{bmatrix} \text{edm}(R) & \text{edm}(R, S) \\ \text{edm}(S, R) & \text{edm}(S) \end{bmatrix}. \quad (45)$$

We define the mask matrix for MDU as

$$W_{\text{MDU}} \stackrel{\text{def}}{=} \begin{bmatrix} 0_{m \times m} & 1_{m \times k} \\ 1_{k \times m} & 0_{k \times k} \end{bmatrix}. \quad (46)$$

With this matrix, we can simply invoke the SDR in Algorithm 5. We could also use Algorithm 2 or Algorithm 4. The performance of different algorithms is compared in the next section.

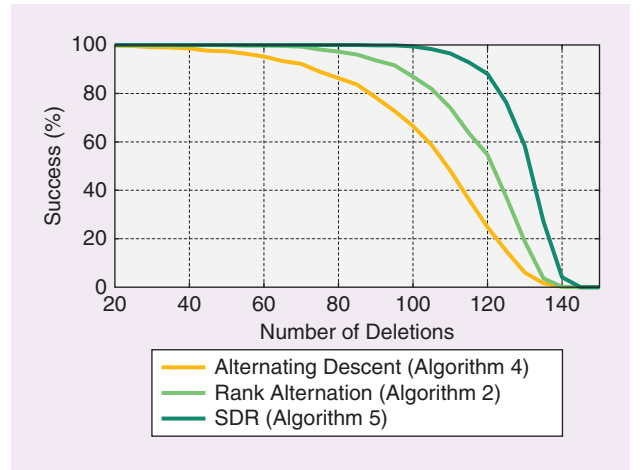
It is worth mentioning that SNL-specific algorithms that exploit the particular graph induced by limited range communication do not perform well on MDU. This is because the structure of the missing entries in MDU is in a certain sense opposite to the one of SNL.

PERFORMANCE COMPARISON OF ALGORITHMS

We compare the described algorithms in two different EDM completion settings. In the first experiment (Figures 5 and 6), the entries to delete are chosen uniformly at random. The second experiment (Figures 7 and 8) tests the performance in MDU, where the nonobserved entries are highly structured. In Figures 5 and 6, we assume that the observed entries are known exactly, and we plot the success rate (percentage of accurate EDM reconstructions) against the number of deletions in the first case and the number of calibration events in the second case. Accurate reconstruction is defined in terms of the relative error. Let D be the true and \hat{D} the estimated EDM. The relative error is then $\|\hat{D} - D\|_F / \|D\|_F$, and we declare success if this error is below 1%.

To generate Figures 6 and 8, we varied the amount of random, uniformly distributed jitter added to the distances, and for each jitter level, we plotted the relative error. The exact values of intermediate curves are less important than the curves for the smallest and largest jitter and the overall shape of the ensemble.

A number of observations can be made about the performance of algorithms. Notably, OptSpace (Algorithm 3) does not perform well for randomly deleted entries when $n = 20$; it was designed for larger matrices. For this matrix size, the mean relative reconstruction error achieved by OptSpace is the worst of all algorithms (Figure 6). In fact, the relative error in the noiseless case was rarely below the success



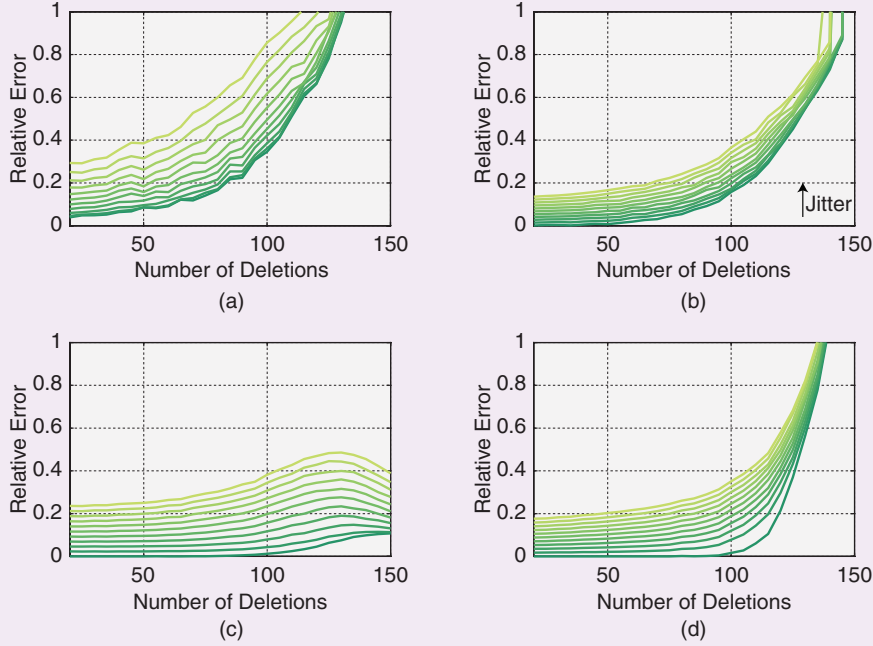
[FIG5] A comparison of different algorithms applied to completing an EDM with random deletions. For every number of deletions, we generated 2,000 realizations of 20 points uniformly at random in a unit square. The distances to delete were chosen uniformly at random among the resulting $(1/2) * 20 * (20 - 1) = 190$ pairs; 20 deletions correspond to $\approx 10\%$ of the number of distance pairs and to 5% of the number of matrix entries; 150 deletions correspond to $\approx 80\%$ of the distance pairs and to $\approx 38\%$ of the number of matrix entries. Success was declared if the Frobenius norm of the error between the estimated matrix and the true EDM was less than 1% of the Frobenius norm of the true EDM.

threshold (set to 1%), so we omitted the corresponding near-zero curve from Figure 5. Furthermore, OptSpace assumes that the pattern of missing entries is random; in the case of a blocked deterministic structure associated with MDU, it never yields a satisfactory completion.

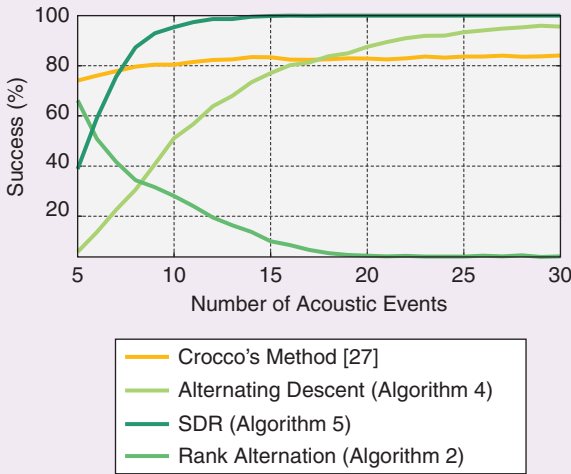
On the other hand, when the unobserved entries are randomly scattered in the matrix, and the matrix is large—in the ultrasonic calibration example, the number of sensors n was 200 or more—OptSpace is a very fast and attractive algorithm. To fully exploit OptSpace, n should be even larger, in the thousands or tens of thousands.

SDR (Algorithm 5) performs well in all scenarios. For both the random deletions and the MDU, it has the highest success rate and it behaves well with respect to noise. Alternating coordinate descent (Algorithm 4) performs slightly better in noise for a small number of deletions and a large number of calibration events, but Figures 5 and 7 indicate that, for certain realizations of the point set, it gives large errors. If the worst-case performance is critical, SDR is a better choice. We note that, in the experiments involving the SDR, we have set the multiplier λ in (36) to the square root of the number of missing entries. This simple choice was empirically found to perform well.

The main drawback of SDR is the speed; it is the slowest among the tested algorithms. To solve the semidefinite program, we used CVX [50], [51], a MATLAB interface to various interior point methods. For larger matrices (e.g., $n = 1,000$), CVX runs out of memory on a desktop computer, and essentially never finishes. MATLAB implementations of alternating coordinate descent, rank alternation (Algorithm 2), and OptSpace are all much faster.



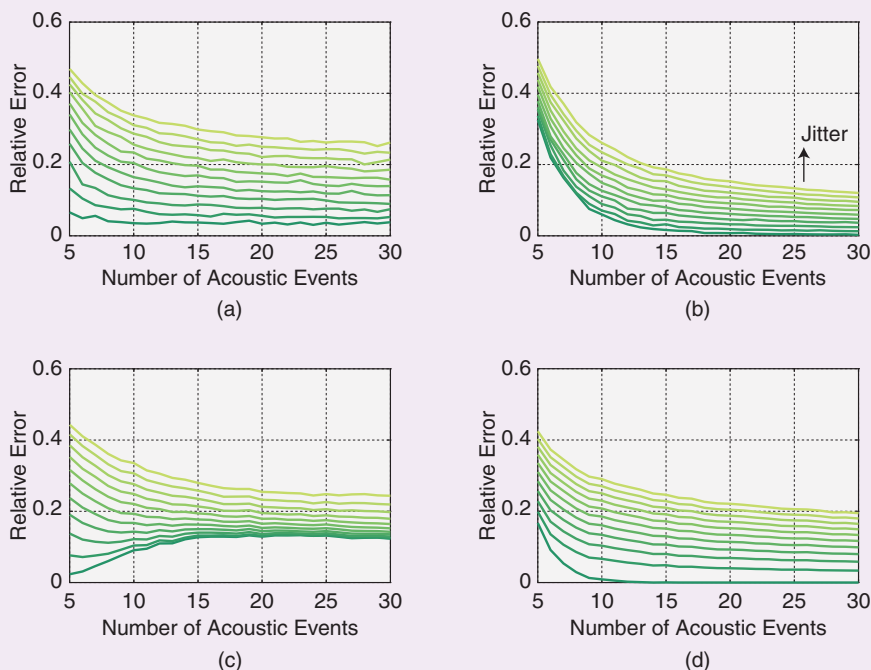
[FIG6] A comparison of different algorithms applied to completing an EDM with random deletions and noisy distances. For every number of deletions, we generated 1,000 realizations of 20 points uniformly at random in a unit square. In addition to the number of deletions, we varied the amount of jitter added to the distances. Jitter was drawn from a centered uniform distribution, with the level increasing in the direction of the arrow, from $\mathcal{U}[0, 0]$ (no jitter) for the darkest curve at the bottom, to $\mathcal{U}[-0.15, 0.15]$ for the lightest curve at the top, in 11 increments. For every jitter level, we plotted the mean relative error $\|\hat{D} - D\|_F / \|D\|_F$ for all algorithms. (a) OptSpace (Algorithm 3). (b) Alternating descent (Algorithm 4). (c) The rank alternation (Algorithm 2). (d) SDR (Algorithm 5).



[FIG7] A comparison of different algorithms applied to MDU with a varying number of acoustic events k . For every number of acoustic events, we generated 3,000 realizations of $m = 20$ microphone locations uniformly at random in a unit cube. The percentage of the missing matrix entries is given as $(k^2 + m^2) / (k + m)^2$ so that the ticks on the abscissa correspond to [68, 56, 51, 50, 51, 52] % (nonmonotonic in k with the minimum for $k = m = 20$). Success was declared if the Frobenius norm of the error between the estimated matrix and the true EDM was less than 1% of the Frobenius norm of the true EDM.

The microphone calibration algorithm by Crocco [27] performs equally well for any number of acoustic events. This may be explained by the fact that it always reduces the problem to ten unknowns. It is an attractive choice for practical calibration problems with a smaller number of calibration events. The algorithm's success rate can be further improved if one is prepared to run it for many random initializations of the nonlinear optimization step.

Interesting behavior can be observed for the rank alternation in MDU. Figures 7 and 8 show that, at low noise levels, the performance of the rank alternation becomes worse with the number of acoustic events. At first glance, this may seem counterintuitive, as more acoustic events means more information; one could simply ignore some of them and perform at least equally well as with fewer events. But this reasoning presumes that the method is aware of the geometrical meaning of the matrix entries; on the contrary, rank alternation is using only rank. Therefore, even if the percentage of the observed matrix entries grows until a certain point, the size of the structured blocks of unknown entries grows as well (and the percentage of known entries in columns/rows corresponding to acoustic events decreases). This makes it harder for a method that does not use geometric relationships to complete the matrix. A loose comparison can be made to image inpainting: If the pixels are missing randomly, many methods will do a good job, but if a large patch is missing, we cannot do much without additional structure (in our case geometry) no matter how large the rest of the image is.



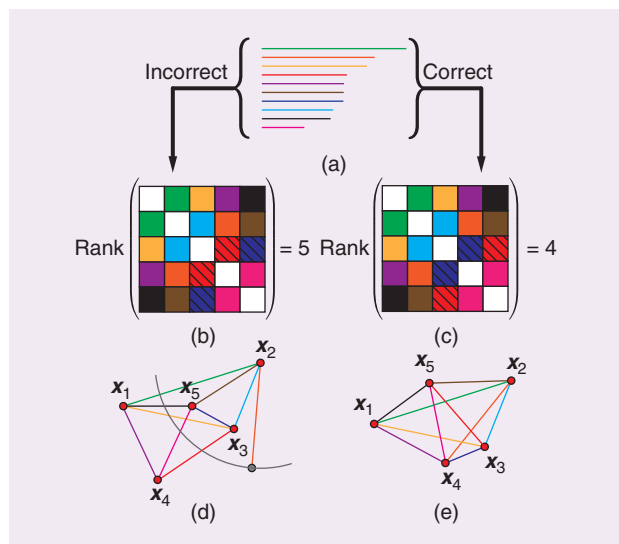
[FIG8] A comparison of different algorithms applied to MDU with a varying number of acoustic events k and noisy distances. For every number of acoustic events, we generated 1,000 realizations of $m = 20$ microphone locations uniformly at random in a unit cube. In addition to the number of acoustic events, we varied the amount of random jitter added to the distances, with the same parameters as in Figure 6. For every jitter level, we plotted the mean relative error $\|\hat{D} - D\|_F / \|D\|_F$ for all algorithms. (a) Crocco's method [27]. (b) Alternating descent (Algorithm 4). (c) Rank alternation (Algorithm 2) and SDR (Algorithm 5).

To summarize, for smaller and moderately sized matrices, the SDR seems to be the best overall choice. For large matrices, the SDR becomes too slow and one should turn to alternating coordinate descent, rank alternation, or OptSpace. Rank alternation is the simplest algorithm, but alternating coordinate descent performs better. For very large matrices (n on the order of thousands or tens of thousands), OptSpace becomes the most attractive solution. We note that we deliberately refrained from making detailed running time comparisons due to the diverse implementations of the algorithms.

SUMMARY

In this section, we discussed:

- the problem statement for EDM completion and denoising and how to easily exploit the rank property (Algorithm 2)
- standard objective functions in MDS, raw stress and s-stress, and a simple algorithm to minimize s-stress (Algorithm 4)
- different SDRs that exploit the connection between EDMs and PSD matrices
- MDU and how to solve it efficiently using EDM completion
- performance of the introduced algorithms in two very different scenarios: EDM completion with randomly unobserved entries and EDM completion with a deterministic block structure of unobserved entries (MDU).



[FIG9] An illustration of the uniqueness of EDMs for unlabeled distances. A set of unlabeled distance (a) is distributed in two different ways in a tentative EDM with embedding dimension two (b) and (c). The correct assignment yields the matrix with the expected rank (c), and the point set is easily realized in the plane (e). On the contrary, swapping just two distances [the hatched squares in (b) and (c)] makes it impossible to realize the point set in the plane (d). Triangles that do not coincide with the swapped edges can still be placed, but in the end, we are left with a hanging orange stick that cannot attach itself to any of the five nodes.

EDM PERSPECTIVE ON SPARSE PHASE RETRIEVAL (THE UNEXPECTED DISTANCE STRUCTURE)

In many cases, it is easier to measure a signal in the Fourier domain. Unfortunately, it is common in these scenarios that we can only reliably measure the magnitude of the Fourier transform (FT). We would like to recover the signal of interest from just the magnitude of its FT, hence the name *phase retrieval*. X-ray crystallography [54] and speckle imaging in astronomy [55] are classic examples of phase retrieval problems. In both of these applications, the signal is spatially sparse. We can model it as

$$f(x) = \sum_{i=1}^n c_i \delta(x - x_i), \quad (S1)$$

where c_i are the amplitudes and x_i are the locations of the n Dirac deltas in the signal. In what follows, we discuss the problem on 1-D domains, that is, for $x \in \mathbb{R}$, knowing that a multidimensional phase retrieval problem can be solved by solving multiple 1-D problems [7].

Note that measuring the magnitude of the FT of $f(x)$ is equivalent to measuring its ACF. For a sparse $f(x)$, the ACF is also sparse and is given as

$$a(x) = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \delta(x - (x_i - x_j)), \quad (S2)$$

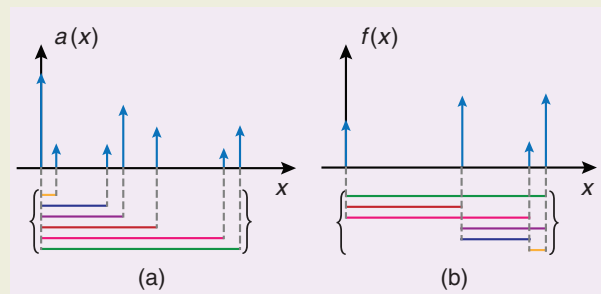
where we note the presence of differences between the locations x_i in the support of the ACF. As $a(x)$ is symmetric, we do not know the order of x_i and so we can only know these differences up to a sign, which is equivalent to knowing the distances $\|x_i - x_j\|$ (Figure S3).

For the following reasons, we focus on the recovery of the support of the signal $f(x)$ from the support of the ACF $a(x)$: 1) in certain applications, the amplitudes c_i may be all equal, thus limiting their role in the reconstruction and 2) knowing the support of $f(x)$ and its ACF is sufficient to exactly recover the signal $f(x)$ [7].

The recovery of the support of $f(x)$ from the one of $a(x)$

corresponds to the localization of a set of n points from their unlabeled distances: we have access to all the pairwise distances but we do not know which pair of points corresponds to any given distance. This can be recognized as an instance of the turnpike problem, whose computational complexity is believed not to be NP-hard but for which no polynomial time algorithm is known [56].

From an EDM perspective, we can design a reconstruction algorithm recovering the support of the signal $f(x)$ by labeling the distances obtained from the ACF such that the resulting EDM has a rank that is less than or equal to three. This can be regarded as unidimensional scaling with unlabeled distances, and the algorithm to solve it is similar to echo sorting (Algorithm 6).



[FIGS3] A graphical representation of the phase retrieval problem for 1-D sparse signals. (a) We measure the ACF of the signal and we recover a set of distances (sticks in Figure 9) from its support. (b) These are the unlabeled distances between all the pairs of Dirac deltas in the signal $f(x)$. We exactly recover the support of the signal if we correctly label the distances.

UNLABELED DISTANCES

In certain applications, we can measure the distances between the points, but we do not know the correct labeling. That is, we know all the entries of an EDM, but we do not know how to arrange them in the matrix. As illustrated in Figure 9(a), we can imagine having a set of sticks of various lengths. The task is to work out the correct way to connect the ends of different sticks so that no stick is left hanging open-ended.

In this section, we exploit the fact that, in many cases, distance labeling is not essential. For most point configurations, there is no other set of points that can generate the corresponding set of distances up to a rigid transformation.

Localization from unlabeled distances is relevant in various calibration scenarios where we cannot tell apart distance measurements belonging to different points in space. This can occur when we measure the TOAs of echoes, which correspond to the distances between the microphones and the image sources (ISs) (see Figure 10) [6], [29]. Somewhat surprisingly, the same

problem of unlabeled distances appears in sparse phase retrieval; see “EDM Perspective on Sparse Phase Retrieval (the Unexpected Distance Structure).”

No efficient algorithm currently exists for localization from unlabeled distances in the general case of noisy distances. We should mention, however, a recent polynomial-time algorithm (albeit of a high degree) by Gujarathi et al. [31] that can reconstruct relatively large point sets from unordered, noiseless distance data.

At any rate, the number of assignments to test is sometimes small enough that an exhaustive search does not present a problem. We can then use EDMs to find the best labeling. The key to the unknown permutation problem is the following fact.

Theorem 3: Draw $x_1, x_2, \dots, x_n \in \mathbb{R}^d$ independently from some absolutely continuous probability distribution (e.g., uniformly at random) on $\Omega \subseteq \mathbb{R}^d$. Then, with probability 1, the obtained point configuration is the unique (up to a rigid

transformation) point configuration in Ω that generates the set of distances $\{\|x_i - x_j\|, 1 \leq i < j \leq n\}$.

This fact is a simple consequence of a result by Boutin and Kemper [52] who provide a characterization of point sets reconstructable from unlabeled distances. Figure 9(b) and (c) shows two possible arrangements of the set of distances in a tentative EDM; the only difference is that the two hatched entries are swapped. But this simple swap is not harmless: there is no way to attach the last stick in Figure 9(d) while keeping the remaining triangles consistent. We could do it in a higher embedding dimension, but we insist on realizing it in the plane.

What Theorem 3 does not tell us is how to identify the correct labeling. But we know that for most sets of distances, only one (correct) permutation can be realized in the given embedding dimension. Of course, if all the labelings are unknown and we have no good heuristics to trim the solution space, finding the correct labeling is difficult, as noted in [31]. Yet there are interesting situations where this search is feasible because we can augment the EDM point by point. We describe one such situation next.

HEARING THE SHAPE OF A ROOM

An important application of EDMs with unlabeled distances is the reconstruction of the room shape from echoes [6]. An acoustic setup is shown in Figure 10(a), but one could also use radio signals. Microphones pick up the convolution of the sound emitted by the loudspeaker with the room impulse response (RIR), which can be estimated by knowing the emitted sound. An example RIR recorded by one of the microphones is illustrated in Figure 10(b), with peaks highlighted in green. Some of these peaks are first-order echoes coming from different walls, and some are higher-order echoes or just noise.

Echoes are linked to the room geometry by the image source (IS) model [53]. According to this model, we can replace echoes by ISs—mirror images of the true sources across the corresponding walls. The position of the IS of s corresponding to wall i is computed as

$$\tilde{s}_i = s + 2 \langle p_i - s, n_i \rangle n_i, \quad (47)$$

where p_i is any point on the i th wall and n_i is the unit normal vector associated with the i th wall [see Figure 10(a)].

A convex room with planar walls is completely determined by the locations of first-order ISs [6], so by reconstructing their locations, we actually reconstruct the room's geometry.

We assume that the loudspeaker and the microphones are synchronized so that the times at which the echoes arrive directly correspond to distances. The challenge is that the distances—the green peaks in Figure 10(b)—are unlabeled: it might happen that the k th peak in the RIR from microphone 1 and the k th peak in the RIR from microphone 2 come from different walls, especially

for larger microphone arrays. Thus, we have to address the problem of echo sorting to group peaks corresponding to the same IS in RIRs from different microphones.

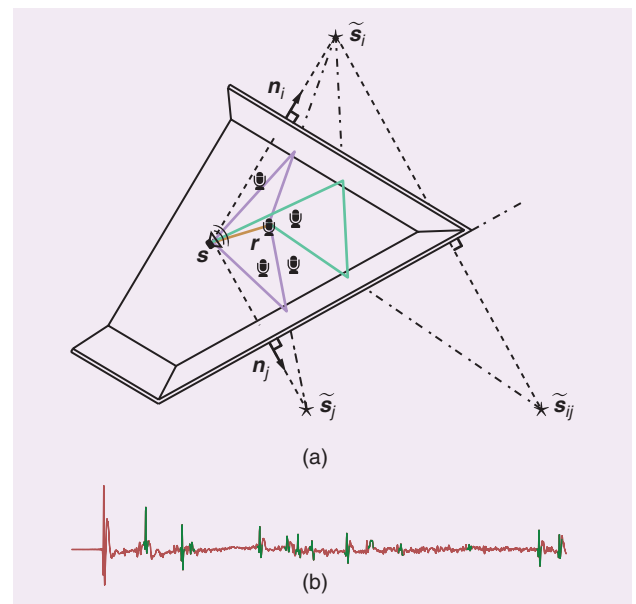
Assuming that we know the pairwise distances between the microphones $R = [r_1, \dots, r_m]$, we can create an EDM corresponding to the microphone array. Because echoes correspond to ISs, and ISs are just points in space, we attempt to grow that EDM by adding one point—an IS—at a time. To do that, we pick one echo from every microphone's impulse response, augment the EDM based on echo arrival times, and check how far the augmented matrix is from an EDM with embedding dimension three, as we work in 3-D space. The distance from an EDM is measured

with the s-stress cost function. It was shown in [6] that a variant of Theorem 3 applies to ISs when microphones are positioned at random. Therefore, if the augmented matrix satisfies the EDM properties, almost surely we have found a good IS. With probability 1, no other combination of points could have generated the used distances.

The main reason for using EDMs and s-stress instead of, for instance, the rank property is that we get robust algorithms. The echo arrival times are corrupted with various errors, and relying on the rank is too brittle. It was verified experimentally [6] that EDMs and s-stress yield a very robust filter for the correct combinations of echoes.

Thus, we may try all feasible combinations of echoes and expect to get exactly one “good” combination for every IS that is

AN IMPORTANT APPLICATION OF EDMs WITH UNLABELED DISTANCES IS THE RECONSTRUCTION OF THE ROOM SHAPE FROM ECHOES.



[FIG10] (a) An illustration of the IS model for first- and second-order echoes. Vector n_i is the outward-pointing unit normal associated with the i th wall. The stars denote the IS, and \tilde{s}_{ij} is the IS corresponding to the second-order echo. The sound rays corresponding to first reflections are shown in purple, and the ray corresponding to the second-order reflection is shown in green. (b) The early part of a typical recorded RIR.

“visible” in the impulse responses. In this case, as we are only adding a single point, the search space is small enough to be rapidly traversed exhaustively. Geometric considerations allow for a further trimming of the search space: because we know the diameter of the microphone array, we know that an echo from a particular wall must arrive at all the microphones within a temporal window corresponding to the array’s diameter.

The procedure is as follows: collect all echo arrival times received by the i th microphone in the set T_i and fix $t_1 \in T_1$ corresponding to a particular IS. Then, Algorithm 6 finds echoes in other microphones’ RIRs that correspond to this same IS. Once we group all the peaks corresponding to one IS, we can determine its location by multilateration (e.g., by running the classical MDS) and then repeat the process for other echoes in T_1 .

To get a ballpark idea of the number of combinations to test, suppose that we detect 20 echoes per microphone and that the diameter of the five-microphone array is 1 m. (We do not need to look beyond early echoes corresponding to at most three bounces; this is convenient as echoes of higher orders are challenging or impossible to isolate.) Thus, for every peak time $t_1 \in T_1$, we have to look for peaks in the remaining four microphones that arrived within a window around t_1 of length $2 \times (1 \text{ m}/343 \text{ m/s})$, where 343 m/s is the speed of sound. This is approximately 6 ms, and in a typical room, we can expect about five early echoes within a window of that duration. Thus, we have to compute the s-stress for $20 \times 5^4 = 12,500$ matrices of size 6×6 , which can be done in a matter of seconds (or less) on a desktop computer. In fact, once we assign an echo to an IS, we can exclude it from further testing, so the number of combinations can be further reduced.

Algorithm 6: Echo sorting [6].

```

1: function EchoSort ( $R, t_1, \dots, T_m$ )
2:    $D \leftarrow \text{edm}(R)$ 
3:    $s_{\text{best}} \leftarrow +\text{Inf}$ 
4:   for all  $t = [t_2, \dots, t_m]$ , such that  $t_i \in T_i$  do
5:      $d \leftarrow c \cdot [t_1, t^\top]^\top \quad \triangleright c \text{ is the sound speed}$ 
6:      $D_{\text{aug}} \leftarrow \begin{bmatrix} D & d \\ d^\top & 0 \end{bmatrix}$ 
7:     if  $s\text{-stress}(D_{\text{aug}}) < s_{\text{best}}$  then
8:        $s_{\text{best}} \leftarrow s\text{-stress}(D_{\text{aug}})$ 
9:        $d_{\text{best}} \leftarrow d$ 
10:    end if
11:  end for
12:  return  $d_{\text{best}}$ 
13: end function

```

Algorithm 6 was used to reconstruct rooms with centimeter precision [6] with one loudspeaker and an array of five microphones. The same algorithm also enables a dual application: indoor localization of an acoustic source using only one microphone—a feat not possible if we are not in a room [57].

[TABLE 2] APPLICATIONS OF EDMs WITH DIFFERENT TWISTS.

APPLICATION	MISSING DISTANCES	NOISY DISTANCES	UNLABELED DISTANCES
WIRELESS SENSOR NETWORKS	✓	✓	×
MOLECULAR CONFORMATION	✓	✓	×
HEARING THE SHAPE OF A ROOM	×	✓	✓
INDOOR LOCALIZATION	×	✓	✓
CALIBRATION	✓	✓	×
SPARSE PHASE RETRIEVAL	×	✓	✓

SUMMARY

To summarize this section:

- We explained that for most point sets, the distances they generate are unique; there are no other point sets generating the same distances.
- In room reconstruction from echoes, we need to identify the correct assignment of the distances to ISs. EDMs act as a robust filter for echoes coming from the same IS.
- Sparse phase retrieval can be cast as a distance problem, too. The support of the ACF gives us the distances between the deltas in the original signal. Echo sorting can be adapted to solve the problem from the EDM perspective.

IDEAS FOR FUTURE RESEARCH

Even problems that at first glance seem to have little to do with EDMs sometimes reveal a distance structure when you look closely. A good example is sparse phase retrieval.

The purpose of this article is to convince the reader that EDMs are powerful objects with a multitude of applications (Table 2 lists various flavors) and that they should belong to any practitioner’s toolbox. We have an impression that the power of EDMs and the associated algorithms has not been sufficiently recognized in the signal processing community, and our goal is to provide a good starting reference. To this end, and perhaps to inspire new research directions, we list several EDM-related problems that we are curious about and believe are important.

DISTANCE MATRICES ON MANIFOLDS

If the points lie on a particular manifold, what can be said about their distance matrix? We know that if the points are on a circle, the EDM has a rank of three instead of four, and this generalizes to hyperspheres [17]. But what about more general manifolds? Are there invertible transforms of the data or of the Gram matrix that yield EDMs with a lower rank than the embedding dimension suggests? What about different distances, e.g., the geodesic distance on the manifold? The answers to these questions have immediate applications in machine learning, where the data can be approximately assumed to be on a smooth surface [23].

PROJECTIONS OF EDMs ON LOWER-DIMENSIONAL SUBSPACES

What happens to an EDM when we project its generating points to a lower-dimensional space? What is the minimum number of

projections that we need to be able to reconstruct the original point set? The answers to these questions have a significant impact on imaging applications such as X-ray crystallography and seismic imaging. What happens when we only have partial distance observations in various subspaces? What are the other useful low-dimensional structures on which we can observe the high-dimensional distance data?

EFFICIENT ALGORITHMS FOR DISTANCE LABELING

Without application-specific heuristics to trim down the search space, identifying correct labeling of the distances quickly becomes an arduous task. Can we identify scenarios for which there are efficient labeling algorithms? What happens when we do not have the labeling, but we also do not have the complete collection of sticks? What can we say about the uniqueness of incomplete unlabeled distance sets? Some of the questions have been answered by Gujarathi et al. [31], but many remain. The quest is on for faster algorithms as well as algorithms that can handle noisy distances.

In particular, if the noise distribution on the unlabeled distances is known, what can we say about the distribution of the reconstructed point set (taking in some sense the best reconstruction over all labelings)? Is it compact, or can we jump to totally wrong assignments with positive probability?

ANALYTICAL LOCAL MINIMUM OF S-STRESS

Everyone agrees that there are many, but, to the best of our knowledge, no analytical minimum of s-stress has yet been found.

CONCLUSIONS

We hope that we have succeeded in showing how universally useful EDMs are and that readers will be inspired to dig deeper after coming across this material. Distance measurements are so common that a simple, yet sophisticated tool like EDMs deserves attention. A good example is the SDR: even though it is generic, it is the best-performing algorithm for the specific problem of ad hoc microphone array localization. Continuing research on this topic will bring new revolutions like it did in the 1980s in crystallography. Perhaps the next one will be fueled by solving the labeling problem.

ACKNOWLEDGMENTS

We would like to thank Dr. Farid M. Naini for his help in expediting the numerical simulations. We would also like to thank the anonymous reviewers for their numerous insightful suggestions that have improved the article. Ivan Dokmanić and Juri Ranieri were supported by the ERC Advanced Grant—Support for Frontier Research—SPARSAM, number 247006. Ivan Dokmanić was also supported by the Google Ph.D. Fellowship.

AUTHORS

Ivan Dokmanić (ivan.dokmanic@epfl.ch) is a Ph.D. candidate in the Audiovisual Communications Laboratory (LCAV) at the École Polytechnique Fédérale de Lausanne (expected graduation in May 2015). His interests include inverse problems, audio and acoustics, signal processing for sensor arrays/networks, and fundamental signal processing. He was previously a teaching assistant at the

University of Zagreb, a codec developer for MainConcept AG, and a digital audio effects designer for Little Endian Ltd. During the summer of 2013, he was a research intern with Microsoft Research in Redmond, Washington, where he worked on ultrasonic sensing. For his work on room shape reconstruction using sound, he received the Best Student Paper Award at the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing. In 2014, he received a Google Ph.D. fellowship.

Reza Parhizkar (reza.parhizkar@gmail.com) received his B.Sc. degree in electrical engineering from Sharif University, Tehran, Iran, in 2003 and his M.Sc. and Ph.D. degrees in communication systems from the École Polytechnique Fédérale de Lausanne in 2009 and 2013. He was an intern at the Nokia research center, Lausanne, and Qualcomm, Inc., San Diego, California. He is currently the head of research in maxx red AG-Zug, Switzerland, where he works on optimization frameworks in finance. His work on sensor calibration for ultrasound tomography devices won the Best Student Paper Award at the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing. His Ph.D. thesis “Euclidean Distance Matrices: Properties, Algorithms, and Applications” was nominated by the thesis committee for the ABB Best Ph.D. Thesis Award in 2014. His research interests include mathematical signal processing, Euclidean distance matrices, and finance.

Juri Ranieri (juri.ranieri@epfl.ch) received his M.S. and B.S. degrees in electronic engineering in 2009 and 2007, respectively, from the Università di Bologna, Italy. From July to December 2009, he was a visiting student at the Audiovisual Communications Laboratory (LCAV) at the École Polytechnique Fédérale de Lausanne (EPFL). From January to August 2010, he was with IBM Zurich to investigate the lithographic process as a signal processing problem. From September 2010 to September 2014, he was with the doctoral school at EPFL, where he obtained his Ph.D. degree at LCAV under the supervision of Prof. Martin Vetterli and Prof. Amina Chebira. From April to July 2013, he was an intern at Lyric Labs of Analog Devices, Cambridge, Massachusetts. His main research interests are inverse problems, sensor placement, and sparse phase retrieval.

Martin Vetterli (martin.vetterli@epfl.ch) received his engineering degree from Eidgenössische Technische Hochschule, Zürich, Switzerland, his M.S. degree from Stanford University, California, and his doctorate degree from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He was on the faculty of Columbia University and the University of California, Berkeley, before joining EPFL. He is currently the president of the National Research Council of the Swiss National Science Foundation. His research interests are in signal processing and communications, e.g., wavelet theory and applications, sparse sampling, joint source-channel coding, and sensor networks. He received the Best Paper Awards of the IEEE Signal Processing Society (1991, 1996, and 2006) and the IEEE Signal Processing Technical and Society Awards (2002 and 2010). He is a Fellow of the IEEE, the Association for Computing Machinery, and the European Association for Signal Processing. He is the coauthor of *Wavelets and Subband Coding* (1995), *Signal Processing for Communications* (2008), and *Foundations of Signal Processing* (2014). He is a highly cited researcher in engineering (Thomson ISI Web of Science, 2007 and 2014).

REFERENCES

- [1] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 22, no. 4, pp. 54–69, July 2005.
- [2] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving Euclidean distance matrix completion problems via semidefinite programming," *Comput. Optim. Appl.*, vol. 12, nos. 1–3, pp. 13–30, Jan. 1999.
- [3] L. Doherty, K. Pister, and L. El Ghaoui, "Convex position estimation in wireless sensor networks," in *Proc. IEEE Conf. Computer Communications (INFOCOM)*, 2001, vol. 3, pp. 1655–1663.
- [4] P. Biswas and Y. Ye, "Semidefinite programming for ad hoc wireless sensor network localization," in *Proc. ACM/IEEE Int. Conf. Information Processing in Sensor Networks*, 2004, pp. 46–54.
- [5] T. F. Havel and K. Wüthrich, "An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution," *J. Mol. Biol.*, vol. 182, no. 2, pp. 281–294, 1985.
- [6] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. Natl. Acad. Sci.*, vol. 110, no. 30, pp. 12186–12191, June 2013.
- [7] J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli, "Phase retrieval for sparse signals: Uniqueness conditions," *IEEE Trans. Inform. Theory*, arXiv:1308.3058v2.
- [8] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.
- [9] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. II-988–II-995, 2004.
- [10] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM Rev.*, vol. 56, no. 1, pp. 3–69, 2014.
- [11] K. Menger, "Untersuchungen über allgemeine metrik," *Math. Ann.*, vol. 100, no. 1, pp. 75–163, Dec. 1928.
- [12] I. J. Schoenberg, "Remarks to Maurice Frechet's article 'Sur la définition axiomatique d'une classe d'espaces distancés vectoriellement applicable sur l'espace de Hilbert,'" *Ann. Math.*, vol. 36, no. 3, p. 724, July 1935.
- [13] L. M. Blumenthal, *Theory and Applications of Distance Geometry*. Oxford, U.K.: Clarendon Press, 1953.
- [14] G. Young and A. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, no. 1, pp. 19–22, 1938.
- [15] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [16] J. C. Gower, "Euclidean distance geometry," *Math. Sci.*, vol. 7, pp. 1–14, 1982.
- [17] J. C. Gower, "Properties of Euclidean and non-Euclidean distance matrices," *Linear Algebra Appl.*, vol. 67, pp. 81–97, June 1985.
- [18] W. Glunt, T. L. Hayden, S. Hong, and J. Wells, "An alternating projection algorithm for computing the nearest Euclidean distance matrix," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 4, pp. 589–600, 1990.
- [19] T. L. Hayden, J. Wells, W.-M. Liu, and P. Tarazaga, "The cone of distance matrices," *Linear Algebra Appl.*, vol. 144, pp. 153–169, Jan. 1990.
- [20] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*. Palo Alto, California: Meboo, 2011.
- [21] M. W. Trosset, "Applications of multidimensional scaling to molecular conformation," *Comp. Sci. Stat.*, vol. 29, pp. 148–152, 1998.
- [22] L. Holm and C. Sander, "Protein structure comparison by alignment of distance matrices," *J. Mol. Biol.*, vol. 233, no. 1, pp. 123–138, Sept. 1993.
- [23] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [24] V. Jain and L. Saul, "Exploratory analysis and visualization of speech and music by locally linear embedding," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, PA, 2004, vol. 2, pp. II-988–II-995.
- [25] E. D. Demaine, F. Gomez-Martin, H. Meijer, D. Rappaport, P. Taslakian, G. T. Toussaint, T. Winograd, and D. R. Wood, "The distance geometry of music," *Comput. Geom.*, vol. 42, no. 5, pp. 429–454, July 2009.
- [26] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Math. Program.*, vol. 109, nos. 2–3, pp. 367–384, Mar. 2007.
- [27] M. Crocco, A. D. Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 660–673, 2012.
- [28] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. Int. Workshop HSC*, Las Vegas, NV, 2008, pp. 2445–2448.
- [29] I. Dokmanić, L. Daudet, and M. Vetterli, "How to localize ten microphones in one fingersnap," in *Proc. European Signal Processing Conf. (EUSIPCO)*, pp. 2275–2279, 2014.
- [30] P. H. Schönemann, "On metric multidimensional unfolding," *Psychometrika*, vol. 35, no. 3, pp. 349–366, 1970.
- [31] S. R. Gujarathi, C. L. Farrow, C. Glosser, L. Granlund, and P. M. Duxbury, "Ab-initio reconstruction of complex Euclidean networks in two dimensions," *Phys. Rev. E*, vol. 89, no. 5, p. 053311, 2014.
- [32] N. Krislock and H. Wolkowicz, "Euclidean distance matrices and applications," in *Handbook on Semidefinite, Conic and Polynomial Optimization*. Boston, MA: Springer, Jan. 2012, pp. 879–914.
- [33] A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, *Distance Geometry: Theory, Methods, and Applications*. New York, NY: Springer Science & Business Media, Dec. 2012.
- [34] P. H. Schönemann, "A solution of the orthogonal procrustes problem with applications to orthogonal and oblique rotation," Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign, 1964.
- [35] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2980–2998, June 2010.
- [36] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 952–960.
- [37] N. Duric, P. Littrup, L. Poulo, A. Babkin, R. Pevzner, E. Holsapple, O. Rama, and C. Glide, "Detection of breast cancer with ultrasound tomography: First results with the computed ultrasound risk evaluation (CURE) prototype," *J. Med. Phys.*, vol. 34, no. 2, pp. 773–785, 2007.
- [38] R. Parhizkar, A. Karbasi, S. Oh, and M. Vetterli, "Calibration using matrix completion with application to ultrasound tomography," *IEEE Trans. Signal Processing*, vol. 61, no. 20, pp. 4923–4933, July 2013.
- [39] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer, 2005.
- [40] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.
- [41] J. De Leeuw, "Applications of convex analysis to multidimensional scaling," in *Recent Developments in Statistics*, J. Barra, F. Brodeau, G. Romier, and B. V. Cutsem, Eds. Amsterdam: North Holland Publishing Company, 1977, pp. 133–146.
- [42] R. Mathar and P. J. F. Groenen, "Algorithms in convex analysis applied to multidimensional scaling," in *Symbolic-Numeric Data Analysis and Learning*, E. Diday and Y. Lechevallier, Eds. Hauppauge, New York: Nova Science Publishers, 1991, pp. 45–56.
- [43] L. Guttman, "A general nonmetric technique for finding the smallest coordinate space for a configuration of points," *Psychometrika*, vol. 33, no. 4, pp. 469–506, 1968.
- [44] Y. Takane, F. Young, and J. De Leeuw, "Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features," *Psychometrika*, vol. 42, no. 1, pp. 7–67, 1977.
- [45] N. Gaffke and R. Mathar, "A cyclic projection algorithm via duality," *Metrika*, vol. 36, no. 1, pp. 29–54, 1989.
- [46] R. Parhizkar, "Euclidean distance matrices: Properties, algorithms and applications," Ph.D. dissertation, School of Computer and Communication Sciences, Ecole Polytechnique Federale de Lausanne (EPFL), 2013.
- [47] M. Browne, "The young-householder algorithm and the least squares multidimensional scaling of squared distances," *J. Classif.*, vol. 4, no. 2, pp. 175–190, 1987.
- [48] W. Glunt, T. L. Hayden, and W.-M. Liu, "The embedding problem for predistance matrices," *Bull. Math. Biol.*, vol. 53, no. 5, pp. 769–796, 1991.
- [49] P. Biswas, T. C. Liang, K. C. Toh, Y. Ye, and T. C. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 4, pp. 360–371, 2006.
- [50] M. Grant and S. Boyd. (2014, Mar.). CVX: MATLAB software for disciplined convex programming, version 2.1. [Online]. Available: <http://cvxr.com/cvx>
- [51] M. Grant and S. Boyd. (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control* (ser. Lecture Notes in Control and Information Sciences). V. Blondel, S. Boyd, and H. Kimura, Eds. New York: Springer-Verlag, pp. 95–110. Available: http://stanford.edu/~boyd/graph_dcp.html
- [52] M. Boutin and G. Kemper, "On reconstructing N-point configurations from the distribution of distances or areas," *Adv. Appl. Math.*, vol. 32, no. 4, pp. 709–735, May 2004.
- [53] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.
- [54] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A*, vol. 7, no. 3, pp. 394–411, Mar. 1990.
- [55] W. Beavers, D. E. Dudgeon, J. W. Beletic, and M. T. Lane, "Speckle imaging through the atmosphere," *Lincoln Lab. J.*, vol. 2, no. 2, pp. 207–228, 1989.
- [56] S. S. Skiena, W. D. Smith, and P. Lemke, "Reconstructing sets from inter-point distances," in *Proc. ACM Symposium on Computational Geometry (SCG)*, 1990, pp. 332–339.
- [57] R. Parhizkar, I. Dokmanić, and M. Vetterli, "Single-channel indoor microphone localization," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.