Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices

Maryam Fazel[1] Haitham Hindi[2] Stephen P. Boyd[3]

Abstract

We present a heuristic for minimizing the rank of a positive semidefinite matrix over a convex set. We use the logarithm of the determinant as a smooth approximation for rank, and locally minimize this function to obtain a sequence of trace minimization problems. We then present a lemma that relates the rank of any general matrix to that of a corresponding positive semidefinite one. Using this, we readily extend the proposed heuristic to handle general matrices. We examine the vector case as a special case, where the heuristic reduces to an iterative ℓ_1-norm minimization technique.

As practical applications of the rank minimization problem and our heuristic, we consider two examples: minimum-order system realization with time-domain constraints, and finding lowest-dimension embedding of points in a Euclidean space from noisy distance data.

1 Introduction

We consider the general matrix Rank Minimization Problem (RMP) expressed as

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, X \\ \text{subject to} & X \in \mathcal{C}, \end{array} \tag{1}$$

where $X \in \mathbf{R}^{m\times n}$ is the optimization variable and $\mathcal{C}$ is a convex set, e.g., described by LMIs. It is well known that in general this problem is computationally hard to solve [VB96, §7.3]. The RMP arises in diverse areas such as control, system identification, statistics, signal processing, and computational geometry (many applications are cataloged in [Faz02]). Various heuristics have been developed to handle problems of this type, specially in the context of low-order controller design; see, e.g., [BG96, SIG98, Dav94].

In this paper we describe a new heuristic for rank minimization that unlike the existing methods, handles any general matrix and does not require a user-specified initial point. In practice, it is observed to yield low-rank solutions, and to require only a small number of convex (semidefinite) programs to be solved.

The outline of the paper is as follows. Section 2 states the semidefinite embedding lemma and its implications for the general RMP. Proofs are given in the appendices. Section 3 presents the log-det heuristic for the positive-semidefinite case, the general case, and the vector case. The last section discusses the applications and gives numerical examples.

2 The semidefinite embedding lemma

Consider the case where the matrix variable X in the RMP (1) is constrained to be positive semidefinite (PSD). The PSD cone has properties that aid us in finding a low-rank matrix; for example, such a matrix will always lie on the boundary of the cone. In fact, this is the basis of the analytical anti-centering and potential reduction methods discussed in [Dav94].

However, there are many applications where X is not necessarily PSD, or even square, making it important to find a way to deal with the general RMP in (1). We resolve this issue by showing that it is possible to associate with any nonsquare matrix X, a positive semidefinite matrix whose rank is exactly twice the rank of X. Thus, any general RMP can be embedded in a larger, positive semidefinite RMP. We refer to this as the *semidefinite embedding* lemma.

Lemma 1 *Let $X \in \mathbf{R}^{m\times n}$ be a given matrix. Then $\mathbf{Rank}\, X \le r$ if and only if there exist matrices $Y = Y^T \in \mathbf{R}^{m\times m}$ and $Z = Z^T \in \mathbf{R}^{n\times n}$ such that*

$$\mathbf{Rank}\, Y + \mathbf{Rank}\, Z \le 2r, \qquad \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \ge 0. \tag{2}$$

For the proof, see appendix A. This result means that minimizing the rank of a general nonsquare matrix X, problem (1), is equivalent to minimizing the rank of the semidefinite, block diagonal matrix $\mathbf{diag}(Y, Z)$:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\,\mathbf{Rank}\,\mathbf{diag}(Y,Z) \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \ge 0 \\ & X \in \mathcal{C}, \end{array} \tag{3}$$

with variables X, Y and Z. The equivalence is in the following sense: the tuple $(X^\star, Y^\star, Z^\star)$ is optimal

[1] Contact author. California Institute of Technology, e-mail: maryam@cds.caltech.edu

[2] Stanford University, e-mail: hhindi@stanford.edu

[3] Stanford University, e-mail: boyd@stanford.edu

Research supported in part by NSF (under ECS-9707111), and DARPA SEC (under F33615-99-C-3014).

for (3) if and only if $X^\star$ is optimal for problem (1), and the objective values in both problems are the same (which is why we keep the factor $\frac{1}{2}$ in the objective).

It is possible to refine the result of Lemma 1 when X is known to have some structure:

Corollary 1 *If X has a block diagonal structure $X = \mathbf{diag}(X_1, \ldots, X_N)$, where $X_i \in \mathbf{R}^{m_i \times n_i}$, then without loss of generality, we may assume that the slack variables have the structure $Y = \mathbf{diag}(Y_1, \ldots, Y_N)$, where $Y_i = Y_i^T \geq 0 \in \mathbf{R}^{m_i \times m_i}$, and $Z = \mathbf{diag}(Z_1, \ldots, Z_N)$, where $Z_i = Z_i^T \geq 0 \in \mathbf{R}^{n_i \times n_i}$.*

To see this, note that $\mathbf{Rank}\, X = \sum_i \mathbf{Rank}\, X_i$ and apply Lemma 1 to each block to get

$$\begin{bmatrix} Y_i & X_i \\ X_i^T & Z_i \end{bmatrix} \geq 0 \quad i = 1, \ldots, N. \tag{4}$$

Corollary 2 *If X is symmetric, then without loss of generality, we can take $Y = Z$.*

See appendix B for the proof of the second corollary.

3 Log-det heuristic

We first state the log-det heuristic for the case of positive semidefinite matrices and propose an iterative linearization and minimization scheme for finding a local minimum. We then apply the log-det heuristic to the general (nonsquare) case using the semidefinite embedding lemma. We also consider the *cardinality minimization problem* as a special case of the RMP when the variable is a vector instead of a matrix, and discuss the log-det heuristic for this problem.

3.1 Positive semidefinite case

Consider the RMP with $X \in \mathbf{R}^{n \times n}$, $X \geq 0$. The log-det heuristic can be described as follows: rather than solving the RMP, use the function $\log\det(X + \delta I)$ as a *smooth surrogate* for $\mathbf{Rank}\, X$ and instead solve the problem

$$\begin{array}{ll} \text{minimize} & \log\det(X + \delta I) \\ \text{subject to} & X \in \mathcal{C}, \end{array} \tag{5}$$

where $\delta > 0$ can be interpreted as a small regularization constant (in practice we choose δ to be very small; see section 4 for numerical examples). The idea of using a log-det type function to obtain low-rank solutions to LMI problems is not entirely new—a similar idea also appears in the potential reduction method of [Dav94] for positive semidefinite matrices. However, we take a different approach to finding a local minimum of this function over the constraint set $\mathcal{C}$.

Note that the surrogate function $\log\det(X + \delta I)$ is not convex (in fact, it is concave). However, since it is smooth on the positive definite cone, it can be minimized (locally) using a local minimization method. We use iterative linearization to find a local minimum. Let X_k denote the kth iterate of the optimization variable X. The first-order Taylor series expansion of $\log\det(X + \delta I)$ about X_k is given by

$$\begin{array}{rcl} \log\det(X + \delta I) & \approx & \log\det(X_k + \delta I) + \\ & & \mathbf{Tr}(X_k + \delta I)^{-1}(X - X_k). \end{array} \tag{6}$$

Here we have used the fact that when $X > 0$, the gradient of $\log\det(X + \delta I)$ with respect to X is given by $\nabla_X \log\det(X + \delta I) = (X + \delta I)^{-1}$. Hence, one could attempt to minimize $\log\det(X + \delta I)$ over the constraint set $\mathcal{C}$ by iteratively minimizing the local linearization (6). This leads to

$$X_{k+1} = \operatorname*{argmin}_{X \in \mathcal{C}} \mathbf{Tr}(X_k + \delta I)^{-1} X. \tag{7}$$

The new optimal point is X_{k+1}, and we have ignored the constants in (6) because they do not affect the minimization. Note that at each iteration we solve a weighted trace minimization problem, with weights $W_k = (X_k + \delta I)^{-1}$. This is a semidefinite program in the variable X. If we choose $X_0 = I$, the first iteration of (7) is equivalent to minimizing the trace of X. It is shown in [FHB01] that the trace function is the *convex envelope* of the rank function over the set matrices with norm less than one. This result lends a theoretical support to the use of trace heuristic as an effective heuristic for the RMP. Therefore, we always pick $X_0 = I$, so that X_1 is the result of the trace heuristic, and the iterations that follow try to reduce the rank of X_1 further. In this sense, we can view this heuristic as a refinement of the trace heuristic.

Since the function $\log\det(X + \delta I)$ is concave in X, at each iteration its value decreases by an amount more than the decrease in the value of the linearized objective function. Based on this observation, it can be shown (e.g., using the global convergence theorem in [Lue84, p.187]) that the sequence $\{f(X_k)\}$ converges to a local minimum of $f(X) = \log\det(X + \delta I)$.

3.2 General case

In order to extend the log-det heuristic to the general case, we use the semidefinite embedding lemma. Recall the equivalence between the RMP (1) and its PSD form (3). Since the matrix $\mathbf{diag}(Y, Z)$ is semidefinite, the log-det heuristic (5) can be directly applied to obtain

$$\begin{array}{ll} \text{minimize} & \log\det(\mathbf{diag}(Y, Z) + \delta I) \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\ & X \in \mathcal{C}. \end{array} \tag{8}$$

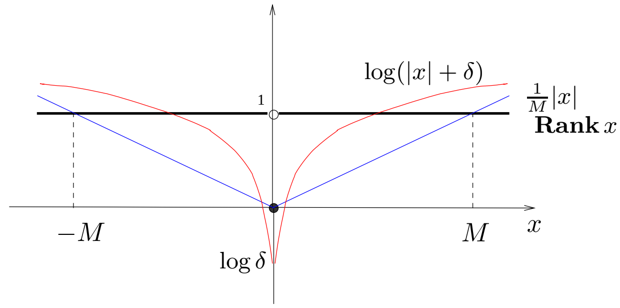

Figure 1: The rank, trace, and log-det objectives in the scalar case.

Linearizing as before, we obtain the following iterations for solving (8) locally:

$$\begin{array}{ll} & \mathbf{diag}(Y_{k+1}, Z_{k+1}) = \\ \text{argmin} & \mathbf{Tr}[(\mathbf{diag}(Y_k, Z_k) + \delta I)^{-1}\,\mathbf{diag}(Y, Z)] \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\ & X \in \mathcal{C}, \end{array} \tag{9}$$

where each iteration is an SDP in the variables X, Y and Z.

Figure 1 provides an intuitive interpretation for the heuristic. It shows the basic idea behind the $\mathbf{Tr}\, X$ and $\log\det(X + \delta I)$ approximations of $\mathbf{Rank}\, X$. Suppose $|x| \leq M$. The objective functions for the trace and log-det heuristics are shown for the scalar case, i.e., when $X = x \in \mathbf{R}$. Then, $\mathbf{Rank}\, x$ is simply equal to zero if $x = 0$ and is equal to 1 otherwise. After normalizing $|x|$ to one, the trace heuristic minimizes the sum of the singular values of the matrix [FHB01], which in this case will be $\sigma(x) = \frac{1}{M}|x|$.

3.3 Vector case: iterative ℓ_1-norm minimization

A useful special case of the RMP is the problem of minimizing the *cardinality*, i.e., the number of non-zero entries, of a vector over a convex set:

$$\begin{array}{ll} \text{minimize} & \mathbf{Card}\, x \\ \text{subject to} & x \in \mathcal{C}, \end{array} \tag{10}$$

where $x \in \mathbf{R}^n$ and $\mathcal{C}$ is a convex set. This is the same as finding the *sparsest* vector in $\mathcal{C}$. This problem comes up in many application areas. For example, in engineering design problems, x might represent some design variables and $\mathcal{C}$ the constraints and specifications. If $x_i = 0$ corresponds to an element or degree of freedom not used, then a sparse x corresponds to an efficient design, i.e., one that uses a small number of elements. The problem (10) is then to find the most efficient (or least complex) design that meets all the specifications.

This problem is a special case of the RMP where the matrix X is diagonal, i.e., $X = \mathbf{diag}\, x$. Therefore, we can specialize results about the rank problem to the cardinality problem. This is useful since it automatically provides us with heuristic solution methods for problem (10). To examine the log-det heuristic in the vector case, consider the special case of the diagonal rank minimization problem. We can show (see [Faz02]) that applying the log-det heuristic to this problem yields

$$\begin{array}{ll} \text{minimize} & \sum_i \log(y_i + \delta) \\ \text{subject to} & |x_i| \leq y_i \quad , \; i = 1, \ldots, n \\ & x \in \mathcal{C}, \end{array}$$

or, equivalently,

$$\begin{array}{ll} \text{minimize} & \sum_i \log(|x_i| + \delta) \\ \text{subject to} & x \in \mathcal{C}, \end{array}$$

where $x \in \mathbf{R}^n$ is the optimization variable. Similar to the matrix case, iterative linearization of the concave objective function gives the following heuristic for vector cardinality minimization:

$$x^{(k+1)} = \underset{x \in \mathcal{C}}{\text{argmin}} \sum_i \frac{|x_i|}{|x_i^{(k)}| + \delta}. \tag{11}$$

Note that if the initial point is chosen as $x^{(0)} = [1,\ 1, \ldots,\ 1]$, the first iteration will minimize $\sum_{i=1}^n |x_i|$, which is the ℓ_1-norm of x, denoted by $\|x\|_1$.

A closer look at this iterative procedure shows that in each step, a weighted ℓ_1-norm of the vector x is minimized, i.e., the objective function at each iteration is $\sum_i w_i^{(k)}|x_i|$, with $w_i^{(k)} = (|x_i^{(k)}| + \delta)^{-1}$. This yields an intuitive interpretation of the method: if $x_i^{(k)}$ is small, its weighting factor in the next minimization step, $w_i^{(k)}$, is large. So the small entries in x are generally pushed toward zero as far as the constraints on x allow, and thus yield a sparse solution.

The iterative ℓ_1 minimization procedure as a heuristic for obtaining sparse solutions, and its application to the problem of portfolio optimization with fixed transaction costs, is given in [LFB00]. Here we derived the same method as a special case of our log-det heuristic.

4 Applications

4.1 System realization with time-domain constraints

In this section, we discuss the problem of designing a low-order, discrete-time, linear time-invariant (LTI) dynamical system, directly from convex specifications on the first n time samples of its impulse response. Some typical specifications are bounds on the rise-time, settling-time, slew-rate, overshoot, etc. We show this problem can be posed as one of minimizing the rank of a Hankel matrix over a convex set. Denote by H_n the Hankel matrix with parameters $h_1, h_2, \ldots, h_{2n-1} \in \mathbf{R}$,

$$H_n = \begin{bmatrix} h_1 & h_2 & h_3 & \ldots & h_n \\ h_2 & h_3 & h_4 & \ldots & h_{n+1} \\ h_3 & h_4 & h_5 & \ldots & h_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n+1} & h_{n+2} & \ldots & h_{2n-1} \end{bmatrix}. \quad (12)$$

Results on system realization from partial impulse response exist (e.g., [Che84, Son90]), but are not directly applicable to our problem. We therefore begin by deriving the following result.

Fact 1 *Let $h_1, h_2, \ldots, h_n \in \mathbf{R}$ be given. Then there exists a minimum-order LTI system of order r with state space matrices $A \in \mathbf{R}^{r\times r}$, $b \in \mathbf{R}^{r\times 1}$ and $c \in \mathbf{R}^{1\times r}$, such that*

$$cA^{i-1}b = h_i \quad i = 1, \ldots, n,$$

if and only if

$$r = \min_{h_{n+1},\ldots,h_{2n-1}\in\mathbf{R}} \mathbf{Rank}\, H_n, \quad (13)$$

where H_n is a Hankel matrix whose first n parameters are the given $h_1, h_2, \ldots, h_n$, and whose last $n-1$ parameters, $h_{n+1}, \ldots, h_{2n-1} \in \mathbf{R}$, are free variables.

Proof: To prove the condition is necessary, let $H_n^\star$ be the minimum rank Hankel matrix in (13); we will construct A, b, c with the desired properties.

Suppose $r < n$. Since $\mathbf{Rank}\, H_n^\star = r$, there are at most r linearly independent rows in $H_n^\star$. So within the first $r+1$ rows, there must be at least one row that is linearly dependent on the rows above it. Suppose row $r'+1$ is the first such row, we show that the Hankel matrix in problem (13) will then have rank r'. To see this, let $H^\star_{r'+1,n}$ denote the top $r'+1$ rows, and solve $[\alpha_1 \ \ldots \ \alpha_{r'} \ \ -1]H^\star_{r'+1,n} = 0$ for the α_i. Then let

$$A = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{r'} \end{bmatrix},$$

$b = [h_1 \ \ldots \ h_{r'}]^T$, and $c = [1\ 0\ \ldots\ 0]$. By direct computation we see that $cA^{i-1}b = h_i$ for $i = 1, \ldots, n+r'$. Define $h_i = cA^{i-1}b$ for $i = n+r'+1, \ldots, 2n-1$, then from linear system theory (e.g., theorem 5.5.7 in [Son90]) we know that the resulting $n \times n$ Hankel matrix will have rank r'. Now since r is the minimum rank H_n can have, it follows that $r' = r$, and the above A, b, c satisfy the desired properties. (If $r = n$, all rows are linearly independent, and $\alpha_1, \ \ldots, \ \alpha_n$ can be chosen arbitrarily.)

To show the condition is also sufficient, note that given A, b, and c, letting $h_i = cA^{i-1}b$ for $i = n+1, \ldots, 2n-1$ yields $\mathbf{Rank}\, H_n = \mathbf{Rank}\, H_r = r$. No other assignment to $h_{n+1}, \ldots, h_{2n-1}$ can give a lower rank for H_n, since this would contradict that A, b, c gives the minimum-order system realizing $h_1, \ldots, h_n$. $\square$

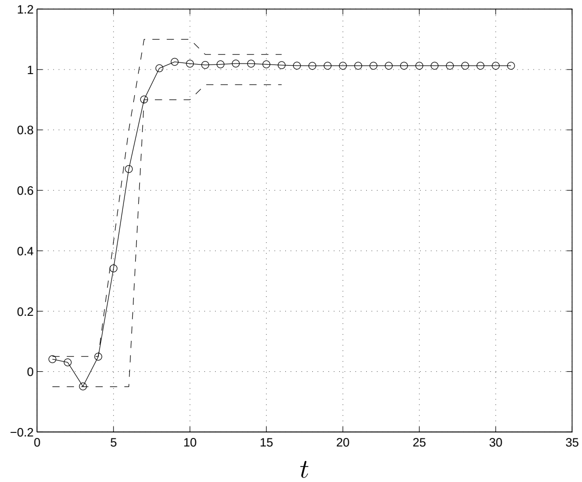

Figure 2: Step response specifications (dashed) and actual step response obtained after 5 iterations of the log-det heuristic.

In other words, Fact 1 states that there exists a linear time invariant system of order r whose first n impulse response samples are $h_1, \ldots, h_n$, if and only if the minimal-rank Hankel matrix has rank r. Note that the constraints are only on the first n samples, even though $h_{n+1}, \ldots, h_{2n-1}$ also appear in the Hankel matrix. These extra variables are left free in the optimization. Thus, they are chosen in a way so as to minimize the overall rank of the Hankel matrix.

To see how the result above can be used to design low-order systems directly from time-domain specifications, consider the specifications on the step response shown in Figure 2. The goal is to find the minimum-order system whose step response fits in the region defined by the dashed lines, up to the 16th sample. The dashed lines are meant to capture a typical set of time-domain step response specifications: certain rise-time, slew-rate, overshoot, and settling characteristics and an approximate delay of four samples. The problem can be expressed as

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, H_n \\ \text{subject to} & l_i \le s_i \le u_i, \quad i = 1, \ldots, n \\ & h_{n+1}, \ldots, h_{2n-1} \in \mathbf{R}, \end{array} \quad (14)$$

where $s_k = \sum_{i=1}^{k} h_i$ denote the terms in the step response, and l_i and u_i are, respectively, samples of the lower and upper time domain specifications (shown by the dashed lines). In this example $n = 16$.

This problem is an RMP with no analytical solution. Note also that the optimization variable H_n is not positive semidefinite. We apply the generalized log-det

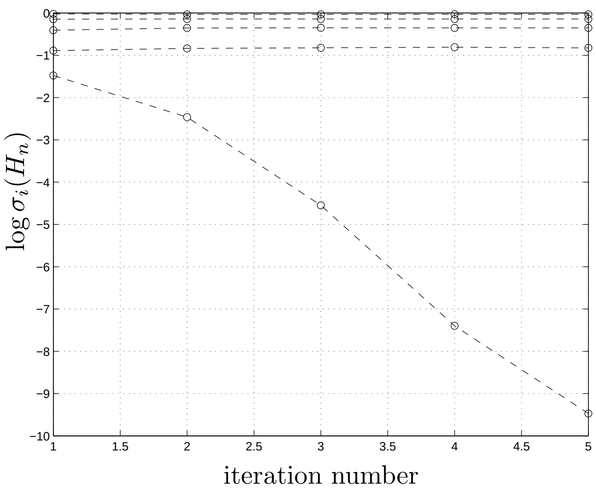

Figure 3: Log of the singular values $\sigma_1, \ldots, \sigma_5$ of H_n at each iteration.

heuristic we described to this problem. Because of the approximate four-sample delay specification, we do not expect that the specifications can be met by a system of order less than four. After five iterations of the log-det heuristic, a fourth-order system is obtained with the step response shown in Figure 2. Thus, all the specifications can be met by a linear time-invariant system of order exactly four. In this example, we set $\delta = 10^{-6}$. Figure 3 shows the logarithm of the nonzero Hankel singular values. We see that the rank of the 16×16 matrix H_n drops to 5 after the first iteration, and the next four iterations bring the rank to 4, which in this case appears to be the global minimum.

4.2 Euclidean distance matrix problems

Euclidean distance matrix (EDM) problems deal with constructing configurations of points from information about interpoint (Euclidean) distances. A simple example is reconstruction of the geographical map of a set of cities given pairwise inter-city distances [BG97, p. 16].

A matrix $D \in \mathbf{R}^{n\times n}$ is called a *Euclidean distance matrix* if there exist points $x_1, \ldots, x_n$ in $\mathbf{R}^r$ such that $D_{ij} = \|x_i - x_j\|^2$. We refer to the dimension of the space in which the points lie, r, as the *embedding dimension.* Let $X \in \mathbf{R}^{r\times n}$ denote the matrix containing the x_i as columns, i.e., $X = [x_1 \ \ldots \ x_n]$. The relation between the matrix of inner products $B = X^T X$ and the distance matrix D is then

$$D = \mathbf{diag}\, B\ \mathbf{1}^T + \mathbf{1}\ (\mathbf{diag}\, B)^T - 2B,$$

where

$$D_{ij} = \|x_i\|^2 + \|x_j\|^2 - 2x_i^T x_j = B_{ii} + B_{jj} - 2B_{ij}.$$

Let $V = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ be the projection matrix onto the hyperplane $\mathbf{1}^T x = 0$. Multiplying a vector by V "centers" the vector by subtracting the mean of all coordinates from each coordinate, i.e., by shifting the origin to the centroid of the points. Multiplying D by V on both sides yields

$$\begin{aligned} VDV &= V(\mathbf{diag}\, B\ \mathbf{1}^T + \mathbf{1}\ (\mathbf{diag}\, B)^T - 2B)V \\ &= -2VBV \\ &= -2\tilde{X}^T\tilde{X}, \end{aligned}$$

where $\tilde{X} = XV$, and columns of $\tilde{X}$ are the centered x_is. The matrix $-\frac{1}{2}VDV$ is sometimes called the double-centered distance matrix.

Schoenberg in 1935 [Sch35] gave the necessary and sufficient conditions for a matrix to be an EDM with given embedding dimension. In our notation, this result shows that $D = D^T \in \mathbf{R}^{n\times n}$ is an EDM with embedding dimension r if and only if the following hold:

$$D_{ii} = 0, \tag{15}$$

$$VDV \preceq 0, \tag{16}$$

$$\mathbf{Rank}(VDV) \le r. \tag{17}$$

Problems involving EDMs arise in a variety of fields, In Multi-Dimensional Scaling (MDS) in statistics, such problems occur in extracting the underlying geometric structure of distance data. In computational chemistry, they come up in inferring the 3-dimensional structure of a molecule (molecular conformation) from information about its interatomic distances (see, e.g., [Tro00]).

If the EDM D is known exactly, the corresponding configuration of points (up to a unitary transform) can be obtained by finding a square-root of $-\frac{1}{2}VDV$. However, in practice, typically only partial data, noisy measurements or incomplete information on D are available. It is often desired to find an EDM that not only is consistent with the measurements, but also requires the smallest number of coordinates to represent the data, i.e., has the smallest embedding dimension. This problem can be expressed as the RMP

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}(-VDV) \\ \text{subject to} & D_{ii} = 0 \\ & -VDV \succeq 0 \\ & D \in \mathcal{C}, \end{array} \tag{18}$$

where $\mathcal{C}$ is a convex set denoting the prior information on D. For example, we may have interval constraints on the distances, i.e.,

$$L_{ij} \le D_{ij} \le U_{ij},$$

where matrices L and U denote the lower and upper bounds. Another common constraint is for D to be close to the measured distance matrix $\hat{D}$ (e.g., in matrix 2-norm or Frobenius norm),

$$\|D - \hat{D}\|_{2,F} \le \epsilon,$$

where ϵ is a given tolerance. The measure of closeness can also be the Lipschitz distance, which is defined as

$$\delta(D, \hat{D}) = \log\left(\max_{i,j} \frac{D_{ij}}{\hat{D}_{ij}} \max_{i,j} \frac{\hat{D}_{ij}}{D_{ij}}\right).$$

Bounding the Lipschitz distance by some ϵ results in the following set of linear constraints on D:

$$\frac{D_{kl}}{\hat{D}_{kl}} \leq (\exp\epsilon)\frac{D_{ij}}{\hat{D}_{ij}}, \quad \text{for all } i, j, k, l.$$

Another type of constraint comes up when only the "order" of the measured distances is considered, rather than the absolute distances themselves. This happens, for example, in non-metric MDS in psychometrics, where the data are human judgments on a pair of stimuli. The human mind may distort distances in a monotonic fashion; therefore only the information on the order of distances is reliable. The order information translates simply to linear inequality constraints on the entries of D, which is convex. Thus, many useful constraints on D are convex and can be easily handled.

The log-det heuristic can be applied to the RMP (18). Our numerical experiments show that they work well, yielding EDMs with very low embedding dimensions.

As a numerical example, we consider 30 randomly generated points in $\mathbf{R}^5$, with all coordinates distributed uniformly over the interval $[0, 1]$. Let $\hat{D}$ be the matrix of squared distances corrupted by additive Gaussian noise, with zero mean and covariance 0.01. This matrix has full rank (with probability one) because of the additive noise, which obscures the underlying geometric structure. We would like to find the D close to $\hat{D}$ in Frobenuis norm, with the smallest embedding dimension. This can be expressed as the RMP

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}(-VDV) \\ \text{subject to} & D_{ii} = 0, \\ & -VDV \geq 0 \\ & \|D - \hat{D}\|_F \leq \epsilon, \end{array}$$

where we assume the tolerance ϵ to be $0.05\|\hat{D}\|_F$. Applying the log-det heuristic to this problem results in a D with an embedding dimension of 5 after 2 iterations, which is exactly the dimension of the underlying space in this case. In this example we set $\delta = 10^{-6}$.

5 Conclusions

We presented the log-det heuristic for matrix rank minimization, and showed it can be applied to any general matrix using the embedding lemma in section 2. In the vector case, where the RMP reduces to finding the sparsest vector in a convex set, we showed the heuristic reduces to iterative ℓ_1-norm minimization.

We showed how to express two problems involving Hankel and Euclidean distance matrices as RMPs. One of our goals here is to point out how several central problems in different fields can be cast as an RMP. We then applied the heuristic to some numerical examples. In these (and other applications given in [Faz02]), the heuristic has been observed to converge in a few iterations (typically 5 or 6), and to yield very low rank solutions.

A Proof of lemma 1

We show each direction separately:

($\Rightarrow$) Suppose that $\mathbf{Rank}\, X = r_0 \leq r$. Then X can be factored as $X = L\,R$, where $L \in \mathbf{R}^{m\times r_0}$ and $R \in \mathbf{R}^{r_0\times n}$, and $\mathbf{Rank}\, L = \mathbf{Rank}\, R = r_0$. Set Y and Z to be the rank r_0 matrices $L\,L^T$ and $R^T\,R$, respectively. Then we have

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} = \begin{bmatrix} L \\ R^T \end{bmatrix} \begin{bmatrix} L^T & R \end{bmatrix} \geq 0.$$

($\Leftarrow$) To prove the other direction, we begin with the following lemma [BEFB94, p.28], which is a generalization of the well known Schur complement condition for positive semidefiniteness [GL89]:

Let X, Y, and Z be real matrices of appropriate dimensions. Then we have the following equivalence:

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \quad \Leftrightarrow \quad \left\{ \begin{array}{ll} \text{(i)} & Y \geq 0 \\ \text{(ii)} & X^T(I - YY^\dagger) = 0 \\ \text{(iii)} & Z - X^TY^\dagger X \geq 0 \end{array} \right., \tag{19}$$

where $Y^\dagger$ denotes the Moore-Penrose psuedoinverse of Y. Also recall that for any $X \in \mathbf{R}^{m\times n}$,

$$\mathbf{Rank}\, X = n - \dim\mathcal{N}(X) = m - \dim\mathcal{N}(X^T). \tag{20}$$

Now given that conditions (i), (ii) and (iii) hold, our goal is to show that they imply $\mathbf{Rank}\, Y \geq \mathbf{Rank}\, X$ and $\mathbf{Rank}\, Z \geq \mathbf{Rank}\, X$.

Assume, without loss of generality, that $\mathbf{Rank}\, Y \leq \mathbf{Rank}\, Z$ (if this were not the case, we could write the conditions in (19) with Y and Z interchanged). From condition (ii) of (19), since $(I - YY^\dagger)$ is a projection operator for $\mathcal{N}(Y)$, it follows that

$$\mathcal{N}(X^T) \supseteq \mathcal{N}(Y) \quad \Rightarrow \quad \dim\mathcal{N}(X^T) \geq \dim\mathcal{N}(Y).$$

Using (20), we conclude that $\mathbf{Rank}\, Y \geq \mathbf{Rank}\, X^T = \mathbf{Rank}\, X$.□

B Proof of corollary 2

To prove that when X is symmetric we can take $Y = Z$ in (3), we will show that given any feasible Y and Z, we can construct a matrix W that is feasible when inserted in place of Y and Z in (3) and yields an equal or smaller objective value.

Again assume, without loss of generality, that $\mathbf{Rank}\, Y \leq \mathbf{Rank}\, Z$. Now let α be a positive real number and consider the matrix αY. Then for any $\alpha > 0$, $\mathbf{Rank}\, \alpha Y = \mathbf{Rank}\, Y$ and αY satisfies conditions (i) and (ii) of (19). If we can show that for some $\alpha_0 > 0$ condition (iii) is also satisfied, then we can take $W = \alpha_0 Y$ and we are done.

Toward that end, consider the expression for condition (iii), with αY in place of Y and Z. Noting that $(\alpha Y)^\dagger = \frac{1}{\alpha} Y^\dagger$, we can be write this as

$$\alpha^2 Y - XY^\dagger X \geq 0. \tag{21}$$

Recall that $Y^\dagger$ can be decomposed as

$$Y^\dagger = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & \\ & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1 \Sigma^{-1} U_1^T, \tag{22}$$

where Σ contains the nonzero eigenvalues of Y, U_1 and U_2 are orthonormal matrices that span the range space of Y, $\mathcal{R}(Y)$, and the nullspace of Y, $\mathcal{N}(Y)$, respectively, and satisfy the identity:

$$U_2 U_2^T + U_1 U_1^T = I. \tag{23}$$

Note that when X is symmetric, condition (ii) in (19) is equivalent to $XU_2 = 0$. Using this relation, and pre- and post-multiplying (21) by $[U_1\, U_2]^T$ and $[U_1\, U_2]$, respectively, we see that (21) holds if and only if the following equivalent condition holds:

$$\alpha^2 \Sigma - U_1^T X Y^\dagger X U_1 \geq 0.$$

This condition can be satisfied by any $\alpha^2 \geq \lambda_{\max}(\Sigma^{-1/2} U_1^T X Y^\dagger X U_1 \Sigma^{-1/2})$. $\square$

References

[BEFB94] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[BG96] E. Beran and K. Grigoriadis. A combined alternating projection and semidefinite programming algorithm for low-order control design. In *Proceedings of IFAC 96*, volume C, pages 85–90, July 1996.

[BG97] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 1997.

[Che84] C. T. Chen. *Linear System Theory and Design*. Holt, Rinehart and Winston, 1984.

[Dav94] J. David. *Algorithms for analysis and design of robust controllers*. PhD thesis, Kat. Univ. Leuven, ESAT, 3001 Leuven, Belgium, 1994.

[Faz02] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Dept. of Elec. Eng., Stanford University, 2002.

[FHB01] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, volume 6, pages 4734–4739, June 2001.

[GL89] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, Baltimore, second edition, 1989.

[LFB00] M. S. Lobo, M. Fazel, and S. Boyd. Portfolio optimization with linear and fixed transaction costs and bounds on risk. Submitted to *Operations Research*, June 2000.

[Lue84] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Mass., 2nd edition, 1984.

[Sch35] I. J. Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Annals of Mathematics*, 36(3):724–732, July 1935.

[SIG98] R. E. Skelton, T. Iwasaki, and K. Grigoriadis. *A Unified Algebraic Approach to Linear Control Design*. Taylor and Francis, 1998.

[Son90] E. D. Sontag. *Mathematical Control Theory*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, 1990.

[Tro00] M. W. Trosset. Distance matrix completion by numerical optimization. *Computational Optimization and Applications*, 17(1):11–22, October 2000.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.