Maximum-Entropy Sampling and the Boolean Quadric Polytope

Kurt M. Anstreicher*

December 14, 2017

Abstract

We consider a bound for the maximum-entropy sampling problem (MESP) that is based on solving a max-det problem over a relaxation of the Boolean Quadric Polytope (BQP). This approach to MESP was first suggested by Christoph Helmberg over 15 years ago, but has apparently never been further elaborated or computationally investigated. We find that the use of a relaxation of BQP that imposes semidefiniteness and a small number of equality constraints gives excellent bounds on many benchmark instances. These bounds can be further tightened by imposing additional inequality constraints that are valid for the BQP. Duality information associated with the BQP-based bounds can be used to fix variables to 0/1 values, and also as the basis for the implementation of a "strong branching" strategy. A branch-and-bound algorithm using the BQP-based bounds solves some benchmark instances of MESP to optimality for the first time.

Keywords: Maximum-entropy sampling, semidefinite programming, semidefinite optimization, Boolean quadric polytope.

Mathematics Subject Classification:

1 Introduction

Let C be an $n \times n$ symmetric positive definite matrix and let s be an integer with $0 < s < n$. For subsets S and T of $N := \{1, 2, \ldots, n\}$, we let $C[S, T]$ denote the submatrix of C having rows indexed by S and columns indexed by T. The *maximum-entropy sampling problem* is

$$\text{MESP}: \qquad z(C, s) := \max \{\text{ldet } C[S, S] \ : \ S \subset N, \ |S| = s\},$$

where ldet denotes the natural logarithm of the determinant. The MESP was introduced in [15] and applied to the design of environmental monitoring networks in [6, 18]. In a typical application, C is a sample covariance matrix obtained from time-series observations of an environmental variable at n locations, and it is desired to choose s locations from which to conduct subsequent data collection so as to maximize the information obtained. The

*Department of Management Sciences, University of Iowa, Iowa City, IA, 52242, USA. Email: `kurt-anstreicher@uiowa.edu`.

use of entropy as a selection criterion, together with the assumption that values at the n locations are drawn from a multivariate normal distribution, then leads naturally to MESP because $\text{ldet}\, C[S,S]$ is, up to constants, the entropy of the Gaussian random variables having covariance matrix $C[S,S]$.

The MESP is an interesting example of a nonlinear discrete optimization problem, and algorithms for the problem have been the subject of considerable investigation; see for example the survey articles [10, 11]. Exact algorithms to compute a maximum-entropy design use the "branch-and-bound" (B&B) framework, for which a key ingredient is the methodology for producing an upper bound on $z(C,s)$. (Subsequent nodes in the B&B tree corresponding to indeces being fixed into or out of S result in problems of the same form as MESP, but with different data (C', s').) A fast method that can provide a reasonably sharp upper bound on $z(C,s)$ is critical to the success of such an approach. A variety of different bounding methods have been developed and investigated [1, 2, 3, 4, 8, 9, 12], and several of these methods have been incorprated into complete B&B algorithms. Recent results using the optimized"masked spectral bound" [1, 4] are perhaps the most promising so far, although computing this bound requires the approximate solution of a nondifferentiable, nonconvex eigenvalue problem over a semidefiniteness constraint.

In this paper we consider a bound for the MESP that is based on solving a max-det (or max-ldet) problem over a relaxation of the Boolean Quadric Polytope (BQP). This bound was suggested by Christoph Helmberg over 15 years ago [10, 11] but has apparently never been elaborated or computationally investigated. In the next section we describe the basic BQP-based bound, which can be written as the maximization of a concave function over semidefiniteness and linear equality constraints. The BQP bound, like the NLP bound [2, 3] is sensitive to a scaling of the covariance matrix C. We describe an initial scaling that has performed well computationally, as well as a simple updating procedure that can be used to improve the scaling of C. Computational results using a benchmark data set with $n = 63$ show that the BQP bound performs very well compared to a number of other bounds for MESP, including the masked spectral bound [1, 4]. In Section 3 we consider strengthening the basic BQP bound by adding additional linear inequalities that are valid for the BQP. The addition of a large number of linear inequalities to the nonlinear SDP results in a problem that quickly becomes computationally challenging. We describe an approach that avoids these computational difficulties by using a linearization of the objective, resulting in a linear SDP which is computationally more tractable and still produces a rigorous bound. We illustrate the effect of adding additional inequality constraints using the same problems with $n = 63$ that were considered in Section 2. In Section 4 we describe variable-fixing logic that can potentially be used to fix variables in the BQP relaxation to 0/1 values. For the original BQP bound, which is based on a nonlinear SDP, this logic is specialized for the particular structure of the problem. Variable fixing can also be based on the linearized SDP, in which case a procedure based on existing theory for fixing variables in SDP relaxations [7] can be adapted. In Section 5 we describe the implementation of a complete B&B algorithm using the BQP-based bounds. This algorithm has state-of-the-art performance on the instances with $n = 63$, and also obtains the first optimal solutions of some instances based on a larger benchmark data set with $n = 124$. Statistics from the B&B solution process illustrate the considerable effect that the variable-fixing procedures described in Section 4 have when implemented within a B&B framework.

NOTATION: I is an identity matrix, E is a matrix of all ones and e is a vector of all ones, with the dimension implicit for all three. For a square matrix X, $\mathrm{diag}(X)$ is the vector of diagonal entries of X while for a vector x, $\mathrm{Diag}(x)$ is the diagonal matrix such that $x = \mathrm{diag}(\mathrm{Diag}(x))$. $X \succeq 0$ denotes that a matrix X is symmetric and positive semidefinite. $A \circ B$ denotes the Hadamard (i.e., element-wise) product, and $A \bullet B$ denotes the matrix inner product $A \bullet B = \mathrm{tr}(AB^T)$.

2 The BQP bound

The Boolean Quadric Polytope (BQP) [13] is usually defined as $\mathrm{Co}\{x_i, x_i x_j, 1 \le i < j \le n \,|\, x \in \{0,1\}^n\}$, where $\mathrm{Co}\{\cdot\}$ denotes the convex hull. For our purposes it is convenient to consider elements of the BQP to be matrices $X \succeq 0$, and so we define BQP to be $\mathrm{Co}\{xx^T \,|\, x \in \{0,1\}^n\}$. The two definitions are clearly equivalent; the second repeats the variables above and below the diagonal of X and uses the fact that $x = \mathrm{diag}(xx^T)$ if $x \in \{0,1\}^n$. A relaxation of the MESP using the BQP can then be written

$$\begin{aligned} \max \quad & \mathrm{ldet}(C \circ X + I - \mathrm{Diag}(x)) \\ \text{s.t.} \quad & e^T x = s \\ & X \in \mathrm{BQP},\ x = \mathrm{diag}(X). \end{aligned} \tag{1}$$

Note that the above problem is indeed a relaxation and not an exact representation of MESP because MESP corresponds to maximizing the objective over the extreme points of the BQP. However, since the objective in (1) is concave, the solution will generally *not* be at an extreme point. The problem (1) is not tractable as written because the constraint that $X \in \mathrm{BQP}$ cannot be exactly enforced. However, by replacing this constraint with constraints on X that are valid for BQP we can obtain a further relaxation that is implementable. The basic relaxation that we will consider is

$$\begin{aligned} \max \quad & \mathrm{ldet}(C \circ X + I - \mathrm{Diag}(x)) \\ \text{s.t.} \quad & e^T x = s,\ Xe = sx \\ & X \succeq xx^T,\ x = \mathrm{diag}(X). \end{aligned} \tag{2}$$

Note that the constraints $X \succeq xx^T,\ x = \mathrm{diag}(X)$ of (2) imply that $0 \le x \le e$. The equality constraints $Xe = sx$ in (2), which were not included in Helmberg's original formulation, can be viewed as reformulation-linearization technique (RLT) [14] constraints obtained by multiplying both sides of the constraint $e^T x = s$ in turn by each x_i, and then replacing terms of the form $x_i x_j$ with X_{ij}. We find that these additional constraints have a small effect on the solution value in (2), but can have a significant effect on the variable-fixing logic considered in Section 4.

Duality for determinant maximization problems is described in [17], but the form of the problems considered there is somewhat different than (2) making the results awkward to apply. To derive the dual to (2) it is convenient to re-write the problem using matrices

$$Y = \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix}, \qquad \widehat{C} = \begin{pmatrix} 0 & 0 \\ 0 & C - I \end{pmatrix}.$$

When working with Y and $\widehat{C}$ we index rows and columns from 0 to n, so for example $Y_{00} = 1$.

Then (2) can be written in the form

$$\begin{array}{ll} \max & \operatorname{ldet}(\widehat{C} \circ Y + I) \\ \text{s.t.} & A_i \bullet Y = b_i, \quad i = 1, \ldots, m, \\ & Y \succeq 0, \end{array} \tag{3}$$

where each A_i is a symmetric matrix, and the constraints $A_i \bullet Y = b_i$, $i = 1, \ldots, 2(n+1)$ express the equality constraints of (2) and the additional constraint $Y_{00} = 1$. To obtain the dual of (3) we form the Lagrangian

$$L(u, T, Y) = \operatorname{ldet}(\widehat{C} \circ Y + I) + T \bullet Y + \sum_{i=1}^{m} u_i (b_i - A_i \bullet Y) \tag{4}$$

and consider the Lagrangian dual problem

$$\min_{u, T \succeq 0} \max_{Y} L(u, T, Y).$$

Since the Lagrangian is concave in Y for any fixed (u, T), the gradient in Y must be zero at the maximizer, and therefore

$$(\widehat{C} \circ Y + I)^{-1} \circ \widehat{C} + T - \sum_{i=1}^{m} u_i A_i = 0.$$

Let

$$S = (\widehat{C} \circ Y + I)^{-1} \succ 0. \tag{5}$$

We then have

$$S \circ \widehat{C} - \sum_{i=1}^{m} u_i A_i = -T \preceq 0, \tag{6}$$

and

$$\begin{aligned} (T - \sum_{i=1}^{m} u_i A_i) \bullet Y &= -((\widehat{C} \circ Y + I)^{-1} \circ \widehat{C}) \bullet Y \\ &= -(\widehat{C} \circ Y + I)^{-1} \bullet (\widehat{C} \circ Y) \\ &= -(\widehat{C} \circ Y + I)^{-1} \bullet (\widehat{C} \circ Y + I) + S \bullet I \\ &= \operatorname{tr}(S) - (n+1). \end{aligned} \tag{7}$$

Substituting (7) into (4), at the maximizer Y we then have

$$L(u, T, Y) = b^T u - \operatorname{ldet}(S) + \operatorname{tr}(S) - (n+1),$$

and the Lagrangian dual can be written as the problem

$$\begin{array}{ll} \min & b^T u - \operatorname{ldet}(S) + \operatorname{tr}(S) - (n+1) \\ \text{s.t.} & S \circ \widehat{C} - \sum_{i=1}^{m} u_i A_i \preceq 0. \end{array} \tag{8}$$

Note that the constraint in (8) is homogenous, so the variables (u, S) can be scaled by any positive factor. Examining the effect of such a scaling on the objective of (8) it is easy to calculate that to minimize the objective we should have $b^T u + \text{tr}(S) = n + 1$. As a result, (8) can be written in the alternative form

$$\begin{array}{ll} \min & -\,\text{ldet}(S) \\ \text{s.t.} & \text{tr}(S) + b^T u = n+1 \\ & S \circ \widehat{C} - \displaystyle\sum_{i=1}^{m} u_i A_i \preceq 0. \end{array}$$

In applying a bound for MESP, for example the BQP bound of (2), there are two transformations of the data C that can be applied to potentially improve the bound obtained: applying a scale factor to C, and working with the complementary problem.

Scale factor. Note that for the MESP with data matrix C, and a positive scalar γ, $z(\gamma C, s) = z(C, s) + s \ln \gamma$. As a result, in applying any bound for MESP we are free to scale C by a value γ and then adjust the resulting bound by subtracting $s \ln \gamma$. Some bounds for the MESP, including the eigenvalue bound [9] and the masked spectral bound [1, 4] are invariant to such a scaling operation. However the BQP bound (2), like the NLP bound [2, 3] is sensitive to the choice of scaling factor. To evaluate the possibility of improving the bound by changing the scale factor it is useful to consider the function

$$v(\gamma, X) = \text{ldet}((\gamma C - I) \circ X + I) - s \ln \gamma.$$

Using the well-known fact [5, p.75] that $\partial\, \text{ldet}(X)/\partial X = X^{-1}$ we have

$$\frac{\partial}{\partial \gamma} v(\gamma, X) = F(\gamma, X)^{-1} \bullet (C \circ X) - s/\gamma,$$

where $F(\gamma, X) = (\gamma C - I) \circ X + I$. For γ to be a minimizer of $v(\gamma, X)$ we then require that $\gamma F(\gamma, X)^{-1} \bullet (C \circ X) = s$, which is equivalent to

$$F(\gamma, X)^{-1} \bullet (I - \text{Diag}(x)) = n - s. \tag{9}$$

There are several ways in which (9) can be used to attempt to improve the scale factor γ. For example, if X is the solution of (2) for some γ, we can use the fact [5, p.64] that

$$\frac{\partial}{\partial \gamma} F(\gamma, X)^{-1} \bullet (I - \text{Diag}(x)) = -\,\text{diag}\left(F^{-1}(\gamma, X)(C \circ X) F^{-1}(\gamma, X)\right)^T (e - x) \tag{10}$$

to make a first-order correction to γ in an attempt to satisfy (9). Note that (10) implies that the left-hand side of (9) is monotonically decreasing in γ, so (9) has a unique solution. However, if γ is changed then (2) must be re-solved with the new scaling factor applied, so X may also change.

Complementary problem. For any set $S \subset N = \{1, 2, \ldots, n\}$ let $\overline{S}$ denote the complementary set $\overline{S} = N \setminus S$. Using the identity

$$\text{ldet}\, C[S, S] = \text{ldet}\, C + \text{ldet}\, C^{-1}[\overline{S}, \overline{S}]$$

we have the identity $z(C, s) = z(C^{-1}, n - s) + \text{ldet}\, C$. As a result, given an instance of MESP we are free to instead work with the "complementary problem" that replaces C with C^{-1} and s with $n - s$. Any bound applied to the complementary problem, adjusted by

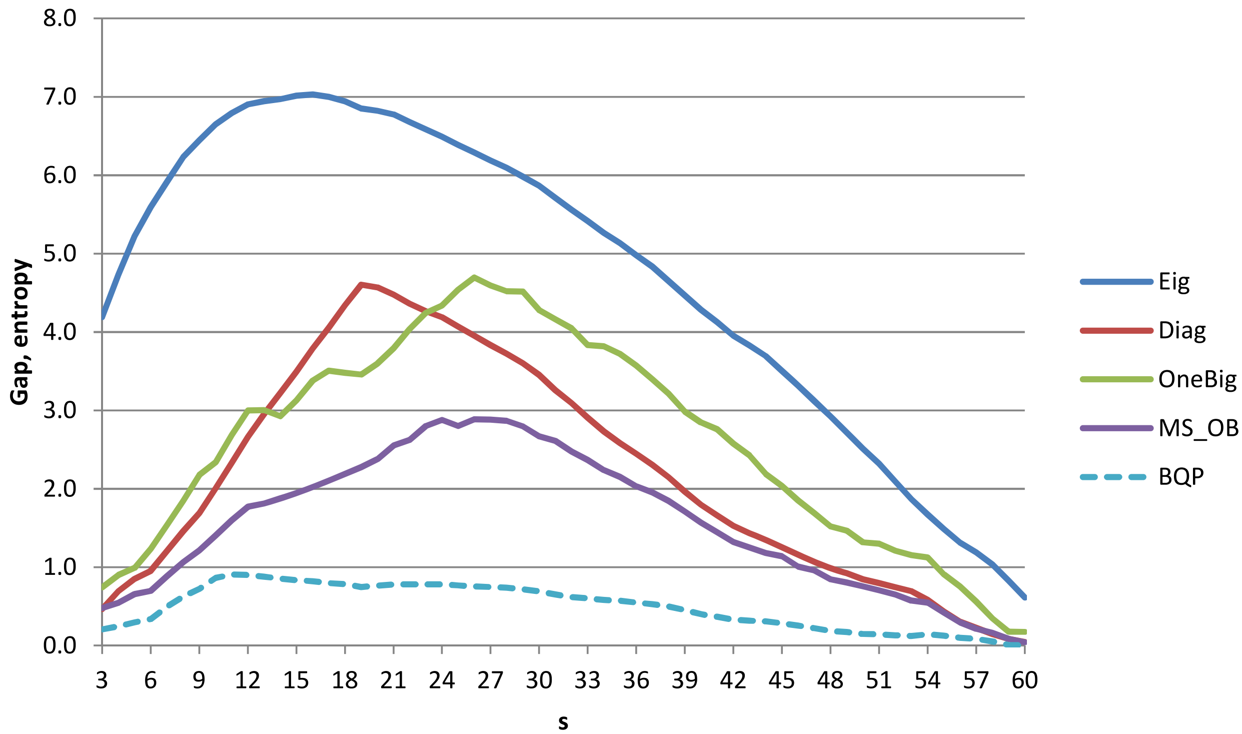

Figure 1: Entropy gaps for bounds applied to $n = 63$ instances

subtracting $\operatorname{ldet} C$, then gives a bound for the original problem. The eigenvalue bound [9] is the same when applied to the original and complementary problems, but most other bounds for MESP can differ when applied to the two problems. It is typically the case that the bound based on the original problem is better for small values of s while the bound based on the complementary problem is better for large values of s, but where exactly one bound becomes better than the other can only be determined by computing both.

To evaluate the basic BQP bound (2) we consider MESP problems using a matrix C with $n = 63$, from [6]. This data has been previously used to evaluate several other bounds for MESP, facilitating compársions between the various bounds. In Figure 1 we give the gap in objective between several bounds and the best feasible solution generated by a heuristic [9] for $s = 3, 4, \ldots, 60$. The eigenvalue, diagonal and "One Big" bounds are all special cases of the Masked Spectral bound. The eigenvalue bounds is the same for the original and complementary problems; for all other bounds the value reported is the better of the two. The values for the optimized Masked Spectral bound (MS) are taken from the compuational results of [4]; these are somewhat better than values previously computed in [1] when the bound for the original problem is better than the bound for the complementary problem ($n \leq 26$). Values for the BQP bound are reported using a scale factor based on the initial choice $\gamma = 1/\operatorname{diag}(C)_{[s]}$, where $\operatorname{diag}(C)_{[s]}$ is the s'th largest component of $\operatorname{diag}(C)$, followed by one first-order correction based on (9) and (10), requiring the solution of two problems of the form (2). The BQP bound is computed using SDPT3 version 4.0 [16], which supports max-ldet terms in the objective. For problems of this size the CPU time required to solve a single instance of (2) on a PC with an Intel i7-6700 CPU running at 3.40 GHz, with 16G of RAM and a 64-bit OS is approximately 5 seconds. It is clear from the figure that the performance of the BQP bound is outstanding on these problems.

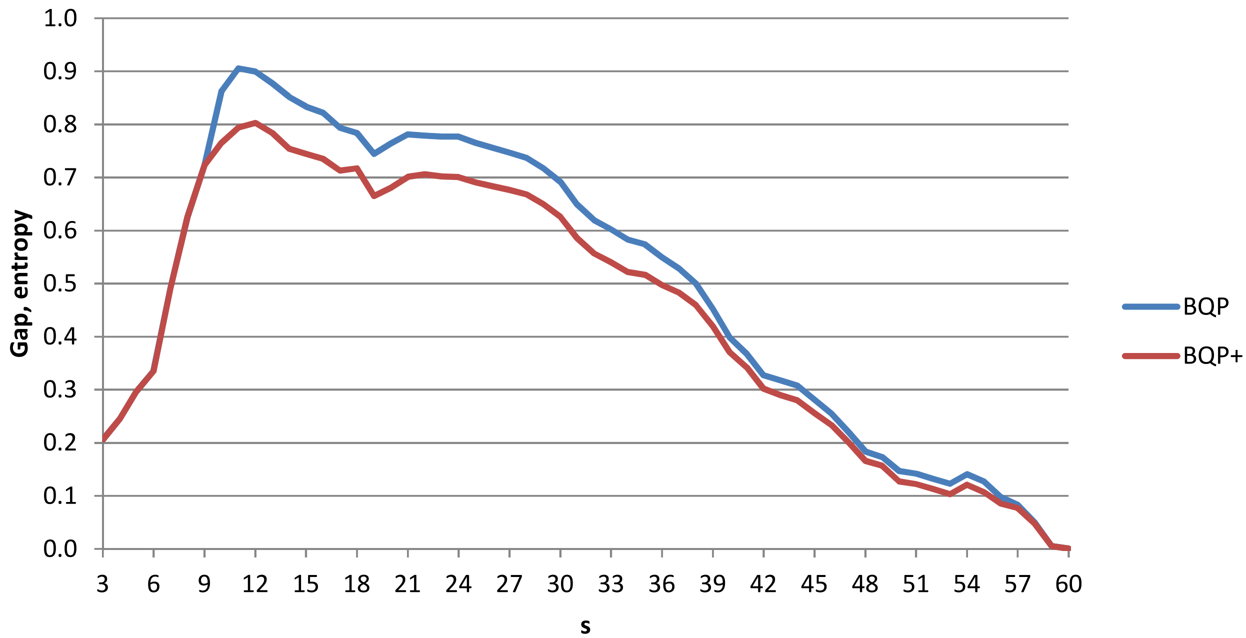

Figure 2: Effect of adding inequality constraints on $n = 63$ instances

3 Adding inequality constraints

One approach to improving the basic BQP bound for MESP from the previous section is to add inequality constraints that are valid for the Boolean Quadric Polytope [13]. Such constraints include the RLT constraints

$$X_{ij} \geq 0, \quad X_{ij} \geq 1 - x_i - x_j, \quad X_{ij} \leq x_i, \quad X_{ij} \leq x_j, \tag{11}$$

for all $i < j$, as well as the triangle inequalities

$$\begin{aligned} x_i + x_j + x_k &\leq X_{ij} + X_{ik} + X_{jk} + 1, \\ X_{ij} + X_{ik} - X_{jk} &\leq x_i, \\ X_{ij} + X_{jk} - X_{ik} &\leq x_j, \\ X_{ik} + X_{jk} - X_{ij} &\leq x_k, \end{aligned}$$

for all $i < j < k$. Due to the large number of such constraints (especially the triangle inequalities) it is inadvisable to add all of the RLT constraints and/or triangle inequalities *a priori*. A more efficient methodology is to first solve the basic BQP relaxation, and then generate a managable number of violated inequalities to add to the relaxation. This process can be repeated for several "rounds" so as to avoid generating a large number of redundant inequalities that would seriously degrade the performance of an interior-point solver such as SDPT3.

We have applied this approach to the instances with $n = 63$ described in the previous section. Although the SDPT3 solver performs well on the basic BQP relaxation with a log-det objective and added equality constraints, we find that the solver often encounters numerical problems with a modest number (a few hundred) of added linear inequalities. To circumvent this problem we consider a linearization of the objective that still produces a rigorous bound, but allows us to solve a purely linear SDP as opposed to a problem with the nonlinear log-det objective. Note that since $\text{ldet}(\cdot)$ is concave, we have

$$\text{ldet}(\widehat{C} \circ Y + I) \leq \text{ldet}(\widehat{C} \circ \overline{Y} + I) + G(\overline{Y}) \bullet (Y - \overline{Y}) \tag{12}$$

for any $\overline{Y}$ such that $(\widehat{C} \circ \overline{Y} + I) \succ 0$, where $G(\overline{Y}) = (\widehat{C} \circ \overline{Y} + I)^{-1} \circ \widehat{C}$ is the gradient of $\text{ldet}(\widehat{C} \circ Y + I)$ at $Y = \overline{Y}$. Using (12) we can obtain a rigorous bound by solving the linear SDP

$$\begin{array}{ll} \max & G(\overline{Y}) \bullet Y \\ \text{s.t.} & A_i \bullet Y = b_i, \quad i = 1, \ldots, m, \\ & B_i \bullet Y \leq d_i, \quad i = 1, \ldots, p, \\ & Y \succeq 0, \end{array} \tag{13}$$

where $\overline{Y}$ is the solution of the original BQP relaxation, and each constraint $B_i \bullet Y \leq d_i$ corresponds to an RLT or triangle inequality. In addition to avoiding numerical difficulties when inequalities are added to the log-det problem, note that (12) is valid for *any* $\overline{Y}$ with $(\widehat{C} \circ \overline{Y} + I) \succ 0$, so that the linearized problem (13) can be used to obtain a rigorous bound even if $\overline{Y}$ does not satisfy the desired optimality conditions for (3). In Figure 2 we illustrate the effect of adding inequality constraints, using the linearized problem (13) where $\overline{Y}$ is the solution of (3), adding RLT and triangle inequalities in two "rounds" where on each round the tolerance for violated inequalities is set so as to generate a maximum of 2,000 constraints. (For the cases $s = 3, \ldots, 60$ the median number of inequality constraints p was about 3,000, with a maximum of 3,852 and a minimum of 142.) In Figure 2 the gaps based on (12) and (13) are labeled BQP+. As can be seen from the figure, the addition of inequality constraints has very little effect on the instances with $s \leq 9$, for which the original bound was better than the bound using the complementary problem. For the remaining instances, where the bound using the complementary problem is better, the addition of inequality constraints has a relatively small absolute effect but does reduce the gap by a nontrivial fraction for the cases where the gap from the original BQP relaxation was the highest.

4 Fixing variables

Because the BQP bound is based on a convex programming relaxation, convex duality can be used to devise variable-fixing logic that may permit some variables to be set at 0/1 values if the gap between the relaxation objective value and optimum (or best known) objective value is sufficiently small. Such logic has the potential to substantially reduce the amount of branching in a B&B algorithm, and is a significant feature for the BQP bound compared to other bounds (including the eigenvalue and masked spectral bounds) for which no such logic is possible. Variable-fixing logic *is* possible for the NLP bound [2, 3] and was a significant contributor to the computational success reported in [3].

For the BQP bound there are at least two different methodologies that can be applied to devise variable-fixing logic. We will describe both in some detail since our experience is that neither methodology dominates the other in computational experiments.

Methodology I. One approach to fixing variables is based on the dual problem (8). Consider for example adding the constraint $e_i e_i^T \bullet Y = 1$ to (3), corresponding to the constraint $x_i = 1$

in (2)[1]. The dual problem (8) with this added constraint can then be written as

$$\begin{array}{ll} \min & b^T u + \omega - \operatorname{ldet}(S) + \operatorname{tr}(S) - (n+1) \\ \text{s.t.} & S \circ \widehat{C} \preceq \omega e_i e_i^T + \sum_{i=1}^m u_i A_i. \end{array} \tag{14}$$

Let $(\bar{u}, \overline{S})$ denote the solution to (8). Assuming that $\widehat{C}_{ii} \neq 0$, for any $\omega \neq 0$ a feasible solution to (14) is then $u = \bar{u}$,

$$S = \overline{S} + \frac{\omega}{\widehat{C}_{ii}} e_i e_i^T = \overline{S}^{1/2} \left(I + \frac{\omega}{\widehat{C}_{ii}} \overline{S}^{-1/2} e_i e_i^T \overline{S}^{-1/2} \right) \overline{S}^{1/2}.$$

For such an S the objective value in (14) is equal to

$$z(\omega) = z(0) + \omega \left(1 + \frac{1}{\widehat{C}_{ii}} \right) - \ln \left(1 + \frac{\omega \sigma_i}{\widehat{C}_{ii}} \right), \tag{15}$$

where $z(0)$ is the solution objective value in (8) and $\sigma = \operatorname{diag}(\overline{S}^{-1})$. A straightforward differentiation shows that the minimum in (15) occurs at

$$\omega^* = \widehat{C}_{ii} \left(\frac{1}{\widehat{C}_{ii} + 1} - \frac{1}{\sigma_i} \right).$$

Substituting this value into (15), recalling that $\widehat{C}_{ii} = C_{ii} - 1$, then obtains $z(\omega^*) = z(0) + \delta_i^1$, where

$$\delta_i^1 = \ln \left(\frac{C_{ii}}{\sigma_i} \right) + 1 - \frac{C_{ii}}{\sigma_i}.$$

Note that if the primal problem (3) has been solved, then the solution dual matrix $\overline{S}$ is given by (5), and therefore $\sigma_i = C_{ii}\bar{x}_i + 1 - \bar{x}_i$ for each i. Finally, the above derivation assumed that $\widehat{C}_{ii} \neq 0$, but if $\widehat{C}_{ii} = 0$ then the above formulas correctly give $\omega^* = 0$ and $z(\omega^*) = z(0)$, the latter because (5) and $\widehat{C}_{ii} = 0$ together imply that $\sigma_i = 1$.

Let v be the objective value corresponding to a feasible solution of MESP. If $z(0) + \delta_i^1 < v$, then we have a proof that $Y_{ii} = x_i = 0$ in any optimal solution of (2), and in this case the instance of MESP can be reduced by deleting the ith row and column of C.

Similar logic can be applied to evaluate the effect of setting $x_i = 0$ in (2). In this case the constraint added to (3) is $e_i e_i^T \bullet Y = 0$, the objective in (14) does not contain the term ω, and the expression for $z(\omega)$ differs from (15) by not containing the term ω. Differentiating the modified expression for $z(\omega)$ to obtain the optimal ω results in

$$\omega^* = \widehat{C}_{ii} \left(\frac{\sigma_i - 1}{\sigma_i} \right),$$

and substituting in to obtain the objective value in the modified version of (14) results in $z(\omega^*) = z(0) + \delta_i^0$, where

$$\delta_i^0 = \ln \left(\frac{1}{\sigma_i} \right) + 1 - \frac{1}{\sigma_i}.$$

[1] Recall that in (3) we index the rows and colums of Y, $\widehat{C}$ and S beginning with zero.

In this case $z(0)+\delta_i^0 < v$ implies that $x_i = 1$ in any optimal solution of (2). In MESP, fixing a set of indeces $F \subset N$ into S can be accomplished [9] by forming the reduced problem where C is replaced by the Shur complement

$$C[N', N'] - C[N', F]\, C[F, F]^{-1}\, C[F, N'], \tag{16}$$

where $N' = N \setminus F$, s is replaced with $s' = s - |F|$, and the objective is adjusted by the constant $\text{ldet}(C_{FF})$.

Methodology II. An alternative approach for variable fixing is based on using the linearized problem (13), with or without the additional inequality constraints $B_i \bullet Y \le d_i$, $i = 1 \ldots, p$. General theory for variable fixing in linear SDP problems is considered in [7]. For a problem with constraints corresponding to the BQP, the approach in [7] would be to first transform the problem to an equivalent one with constraints in max-cut form. In our application, the presence of the trace constraint $e^T \operatorname{diag}(X) = s$ facilitates applying the variable-fixing logic without applying this transformation. We will work with the dual of (13), which is the linear SDP problem

$$\begin{aligned} \min \quad & b^T u + d^T w \\ \text{s.t.} \quad & G(\overline{Y}) - \sum_{i=1}^{m} u_i A_i - \sum_{i=1}^{p} w_i B_i \preceq 0 \\ & w \ge 0. \end{aligned} \tag{17}$$

First consider the logic to impose the constraint $x_i = 1$. Similar to the approach taken in Methodology I, we add the constraint $e_i e_i^T \bullet Y = 1$ to (13). Let $(\bar{u}, \bar{w})$ be an optimal solution to (17), and let μ be the dual variable corresponding to this new constraint. Assume that the trace constraint is written in the form $e^T \operatorname{diag}(Y) = s+1$, that this is the equality constraint with $i = 1$, and let $u_1 = \bar{u}_1 + \omega$. We consider a restricted dual problem of the form:

$$\begin{aligned} \min \quad & b^T \bar{u} + d^T \bar{w} + (s+1)\omega - \mu \\ \text{s.t.} \quad & G(\overline{Y}) - \sum_{i=1}^{m} \bar{u}_i A_i - \sum_{i=1}^{p} \bar{w}_i B_i - \omega I + \mu e_i e_i^T \preceq 0 \\ & \omega \ge 0. \end{aligned} \tag{18}$$

Let $\overline{T} = \sum_{i=1}^m \bar{u}_i A_i + \sum_{i=1}^p \bar{w}_i B_i - G(\overline{Y}) \succeq 0$. Ignoring the constant term in the objective, (18) can be written in the form

$$\begin{aligned} \min \quad & (s+1)\omega - \mu \\ \text{s.t.} \quad & \omega I - \mu e_i e_i^T \succeq -\overline{T} \\ & w \ge 0. \end{aligned}$$

Assume that $\overline{T} = Q \Lambda Q^T$, where Q is an orthonormal matrix and Λ is a diagonal matrix with $\lambda = \operatorname{diag}(\Lambda) \ge 0$. The constraint $\omega I - \mu e_i e_i^T \succeq -\overline{T}$ is then equivalent to $\omega I - \mu q q^T \succeq -\Lambda$, where $q = Q^T e_i$. Assuming that $\mu > 0, \omega > 0$, this constraint is in turn equivalent to the constraints

$$\begin{aligned}
\omega I + \Lambda &\succeq \mu q q^T \\
I &\succeq \mu(\omega I + \Lambda)^{-1/2} q q^T (\omega I + \Lambda)^{-1/2} \\
q^T(\omega I + \Lambda)^{-1} q &\leq \frac{1}{\mu} \\
\mu &\leq \left(\sum_{i=0}^{n} \frac{q_i^2}{\omega + \lambda_i} \right)^{-1}.
\end{aligned}$$

Let $z(0) = \text{ldet}(\widehat{C} \circ \overline{Y} + I) - G(\overline{Y}) \bullet \overline{Y} + b^T \bar{u} + d^T \bar{w}$, and for $\omega > 0$ define

$$z(\omega) = z(0) + (s+1)\omega - \left(\sum_{i=0}^{n} \frac{q_i^2}{\omega - \lambda_i} \right)^{-1}.$$

We can perform a one-dimensional linesearch on $z(\omega)$ to find an approximate minimizer ω^*; note that $z(\cdot)$ is convex since $z(\omega)$ (ignoring a constant) is the minimum in (18) as a function of ω. Let $\delta_i^1 = z(\omega^*) - z(0)$. If $z(0) + \delta_i^1 < v$, where v is the objective value for a feasible solution to MESP, then we have a proof that $x_i = 0$ in any optimal solution of MESP.

The logic to impose the constraint $x_i = 0$ is identical to the above with $e_i - e_0$ replacing the vector e_i throughout. Note that $(e_i - e_0)^T Y (e_i - e_0) = 1 - x_i$, so the constraint $Y \bullet (e_i - e_0)(e_i - e_0)^T = 1$ is equivalent to $x_i = 0$, and in this case $z(0) + \delta_i^0 < v$ where $\delta_i^0 = z(\omega^*) - z(0)$ provides a proof that $x_i = 1$ in any optimal solution of MESP.

It is important to note that in the case where no inequality constraints are added ($p = 0$), all of the information required to implement Methodology II is already available from the solution of (3). The reason for this is simply that the solution $\overline{Y}$ of (3) must also solve (13). It is obvious that with $p = 0$, (13) is a relaxation of (3), but since (13) is based on the gradient of the objective at $\overline{Y}$, if the objective value in (13) was less than $G(\overline{Y}) \bullet \overline{Y}$ then it would be possible to construct a solution in (3) whose objective value was less than that of $\overline{Y}$. In particular, when $p = 0$, the matrix $\overline{T}$ is available as a dual variable from the solution of (3) via (6). When $p = 0$ we can then compute δ_i^1 for each i using both methodologies and take the lower (more negative) value, and similarly for each δ_i^0. In our experience neither methodology dominates the other and it is worthwhile to execute both. In addition to their use in fixing variables, the δ_i^1 and δ_i^0 values can be used as the basis for a strong branching strategy within a B&B framework, as described in the next section.

For the instances with $n = 63$, we find that the variable-fixing logic is capable of fixing some variables to 0/1 values when the gap between the BQP bound and the known value v is less than .50. For example, for the instances with $s = 40$, 45 , 50 and 55, applying the BQP bound to the complementary problem, the number of variables that can be fixed to value 1 is 4, 21, 32 and 41, respectively (no variables can be fixed to value 0).

5 Branch and Bound implementation

We have implemented the BQP-based bounds in a complete B&B algorithm for solving instances of MESP to optimality. In this section we describe some important features of our B&B implentation and give computational results on benchmark problems.

The B&B tree is initialized with a root node corresponding to the given MESP, and a best-known value (BKV) obtained from a set of heuristics for MESP [9]. We maintain a queue of unfathomed nodes from the tree, which is processed using depth-first search. Each node corresponds to a subproblem with some variables fixed to 0/1 values. When a node is removed from the queue we form the reduced MESP problem obtained by eliminating variables fixed to 0, and using (16) to fix variables to 1. The resulting problem is an instance of MESP where $s' \leq s$ indeces must be chosen from $n' \leq n$ candidate indeces. If $s' = 1$ or $s' = n' - 1$ then we enumerate the possible solutions. If a solution is found with objective value better than the BKV then the BKV is updated and the set of corresponding indeces saved. In either case the node is discarded.

If $1 < s' < n' - 1$ we first check if the bound inherited from the node's parent is less than the BKV. If so (which is possible since the BKV may have been updated since the node was placed on the queue) then the node is discarded. Otherwise we compute a bound by solving a problem of the form (2). We use the same bound (original or complementary) as was used for the node's parent, using a scaling factor γ that is inherited from the node's parent. Only one problem of the form (2) is solved to compute the bound; we compute an update to γ as described in Section 2 and pass this updated value to child nodes. If the resulting bound is less than the BKV then we discard the node. Otherwise we apply the variable-fixing logic described in Section 4 to fix additional variables to 0/1 values, if possible. Following additional variable-fixing we are left with the problem of choosing s'' indeces from n'' candidate indeces. If $s'' \leq 1$ or $s'' \geq n'' - 1$ then we enumerate the possible solutions, update the BKV if applicable, and discard the node.

$1 < s'' < n'' - 1$ then we will replace the node with two children by branching on one index. To choose the branching index we use a strong branching strategy based on the δ_i^0 and δ_i^1 values described in Section 4. Let $\delta_{\min}$ be the minimum of all of the δ_i^0 and δ_i^1 values for the remaining variables, and let $i_{\min}$ be the index for which this is minimum is attained. Initially we branch on $i_{\min}$, putting the "easy" node on the queue first and the "hard" node second. If $\delta_{\min} = \delta_{i_{\min}}^0$ then the easy node has $x_{i_{\min}} = 0$ and the hard node has $x_{i_{\min}} = 1$, with the reverse if $\delta_{\min} = \delta_{i_{\min}}^1$. Putting the hard node on second and using depth-first search induces an initial "dive" in the B&B tree. This dive eventually produces a feasible solution and typically leads to rapid update of the BKV when the initial BKV is not the optimal value. After the initial dive is complete, resulting in a node being fathomed, we continue to use the strong branching criterion if the value $\delta_{\min}$ is sufficiently negative, but switch the order and place the hard node on the queue first and the easy node second. We use a simple criterion based on the absolute value of $\delta_{\min}$ as well as the value of $\delta_{\min}$ relative to the current gap to decide if $\delta_{\min}$ is sufficiently negative, and if not we simply branch on the variable i with the minimum value of $|\bar{x}_i - .5|$ in the solution of (2).

In addition to the computation of a bound based on the type of bound (original or complementary) used by a node's parent, we periodically compute the other bound (complementary or original, respectively) to check if the other bound is possibly better. This can occur due to the fixing of variables, and as described in [4] is important to consider when using a bound like the BQP bound, or the masked spectral bound, which differs for the original and complementary problems. On the other hand excessive computation of both bounds would substantially increase the total computation time. We use a simple criterion based on the number of fixed variables and depth in the tree to decide when to check the other bound.

We first consider the instances of MESP with $n = 63$ described in Section 2. These

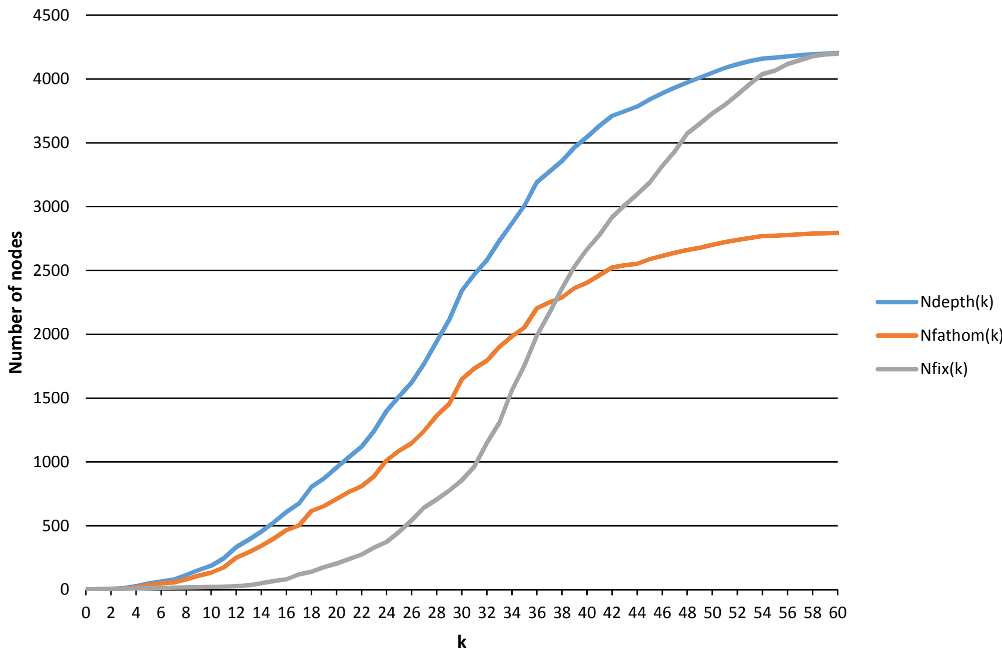

Figure 3: Cumulative B&B tree statistics for $n = 63$, $s = 31$ instance

problems were previously solved by a B&B algorithm using the NLP-based bound in [2, 3]. Our B&B algorithm using the BQP-based bound solves any instance of these problems in less than one hour on a 3.4 GHz PC. For example, the instance with $s = 31$ is solved in 2,438 seconds using 4,203 nodes in the B&B tree, with over 85% of the total time dedicated to the SDPT3 solver. This total time is far superior to the result using the masked spectral bound in [4], which required 65 hours on a 2.4 GHz PC, and is almost identical to the time required in [3] after adjusting for the difference in CPU clock speed. As shown in Figure 1 the initial gap for the BQP-based bound for these instances is largest at $s = 11$, but the instance with $s = 11$ requires only 764 seconds and 1,481 nodes. In general instances of MESP are most difficult to solve to optimality when s is approximately equal to $n/2$, which is consistent with the inherent combinatorial complexity of the problem.

In Figure 3 we give statistics for the B&B tree from the solution of the instance with $n = 63$, $s = 31$. For each k, Ndepth(k) is the cumulative number of nodes at depth less than or equal to k, where a node is at depth k if it has been obtained as the result of branching on k variables. Nfathom(k) is the cumulative number of nodes fathomed at depth less than or equal to k; the overall fraction of nodes fathomed is almost exactly 2/3. Nfix(k) is the number of nodes having no more than k fixed variables. Note that a node at depth k has k variables fixed due to branching, so we would have Nfix(k)=Ndepth(k) if there were no additional fixed variables. The figure shows the substantial effect of the variable-fixing logic described in Section 4. For example, the median depth of the nodes in the B&B tree is 29, with 2,113 of 4,203 nodes at depth less than or equal to 29. However there were only 775 nodes with 29 or fewer fixed variables, and the median number of fixed variables was 37. In addition to possibly improving the bounds obtained, the larger number of fixed variables has a considerable effect in speeding up our bound computations based on (2) due to the reduced number of problem variables.

In addition to the problems with $n = 63$ we also considered a set of MESP instances

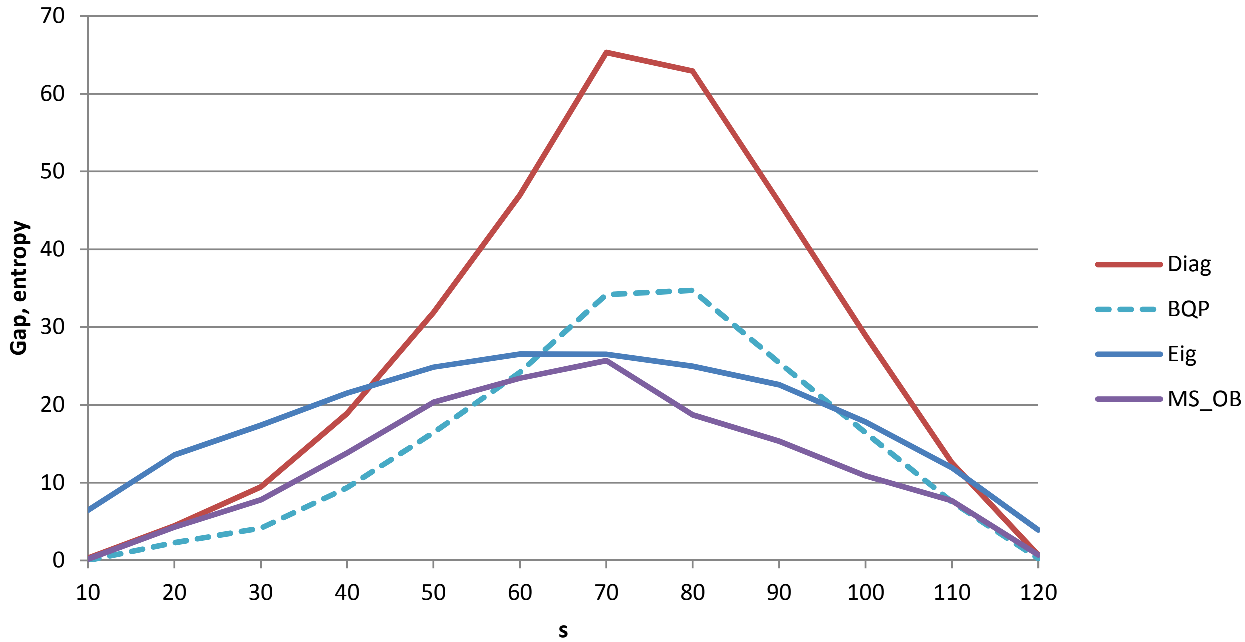

Figure 4: Entropy gaps for bounds applied to $n = 124$ instances

with $n = 124$, introduced in [8]. In Figure 4 we give the root gap to the best value obtained by a heuristic for these instances with $s = 10, 20, \ldots, 120$ for several different bounds. It is shown in [8] that the NLP bounds [2, 3] are very poor on these problems. The BQP bound is better than any previously known bound for these instances for $s \leq 50$ and $s \geq 110$, but is worse than some other bounds (in particular the masked spectral bound) for $60 \leq s \leq 100$. Comparing Figures 1 and 4 it is clear that the root gaps are much higher for the instances with $n = 124$, which means that we can expect that these problems will be much more difficult to solve to optimality. Since the root gaps are much larger than those for the instances with $n = 63$, one might also hope that the addition of inequalities as described in Section 3 might have a larger effect. Unfortunately this is not the case; we find that adding additional inequalities from the BQP has a relatively small effect on the bound, as shown for the problems with $n = 63$ in Figure 2.

To our knowledge the only instance with $n = 124$ that is reported as solved to optimality in the literature is the problem with $s = 115$, which was solved in under one hour of computation using the masked spectral bound [4]. For our B&B algorithm based on the BQP bound we have found that it is practical to solve these instances to optimality when the root gap is below 5. In Table 1 we give solution statistics for the instances with $s = 15$, 20, 25, 30 and 115. To our knowledge the problem with $n = 124$, $s = 30$ is currently the most difficult instance of MESP that has been solved to optimality.

Acknowledgement

I am grateful to Jon Lee for several very valuable conversations on the topic of this paper, and to Sam Burer for providing details of the computational results from [4].

Table 1: B&B statistics for $n = 124$ instances

s	Heuristic Value	Solution Value	Root Gap	Nodes	Time (hours)
15	61.558	61.889	0.788	173	0.1
20	76.989	77.827	2.284	700	1.0
25	92.176	92.828	3.203	4,126	3.2
30	106.674	106.700	4.142	93,652	117.0
115	137.188	137.299	2.921	1,819	0.6

References

[1] K. Anstreicher and J. Lee. A masked spectral bound for maximum-entropy sampling. In *mODa 7—Advances in model-oriented design and analysis*, Contrib. Statist., pages 1–10. Physica, Heidelberg, 2004.

[2] K. M. Anstreicher, M. Fampa, J. Lee, and J. Williams. Continuous relaxations for constrained maximum-entropy sampling. In *Integer Programming and Combinatorial Optimization (Vancouver, BC, 1996)*, volume 1084 of *Lecture Notes in Comput. Sci.*, pages 234–248. Springer, Berlin, 1996.

[3] K. M. Anstreicher, M. Fampa, J. Lee, and J. Williams. Using continuous nonlinear relaxations to solve constrained maximum-entropy sampling problems. *Mathematical Programming*, 85(2, Ser. A):221–240, 1999.

[4] S. Burer and J. Lee. Solving maximum-entropy sampling problems using factored masks. *Mathematical Programming Series B*, 109:263–281, 2007.

[5] A. Graham. *Kronecker Products and Matrix Calculus: with Applications*. Ellis Horwood, 1981.

[6] P. Guttorp, N. Le, P. Sampson, and J. Zidek. Using entropy in the redesign of an environmental monitoring network. Technical Report 116, The Department of Statistics, The University of British Columbia, 1992.

[7] C. Helmberg. Fixing variables in semidefinite relaxations. *SIAM Journal on Matrix Analysis and Applications*, 21(3):952–969, 2000.

[8] A. Hoffman, J. Lee, and J. Williams. New upper bounds for maximum-entropy sampling. In *mODa 6—Advances in model-oriented design and analysis (Puchberg/Schneeberg, 2001)*, Contrib. Statist., pages 143–153. Physica, Heidelberg, 2001.

[9] C.-W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.

[10] J. Lee. Semidefinite programming in experimental design. In R. Saigal, H. Wolkowicz, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research and Management Science*, pages 528–532. Kluwer Acad. Publ., Boston, MA, 2000.

[11] J. Lee. Maximum-entropy sampling. In A. H. El-Shaarawi and W. W. Piegorsch, editors, *Encyclopedia of Environmetrics*, volume 3, pages 1229–1234. John Wiley & Sons Inc., 2001.

[12] J. Lee and J. Williams. A linear integer programming bound for maximum-entropy sampling. *Mathematical Programming*, 94(2-3, Ser. B):247–256, 2003.

[13] M. Padberg. The Boolean quadric polytope: some characteristics, facets and relatives. *Mathematical Programming Series B*, 45:139–172, 1989.

[14] H. Sherali and W. Adams. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems.* Kluwer, 1998.

[15] M. Shewry and H. P. Wynn. Maximum entropy sampling. *Journal of Applied Statistics*, 46:165–170, 1987.

[16] K.-C. Toh, M. J. Todd, and R. H. T́ut́uńcú. On the implementation and usage of SDPT3 - a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0. In M. Anjos and J. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research and Management Science*, pages 715–754. Springer, 2012.

[17] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.

[18] S. Wu and J. V. Zidek. An entropy based review of selected NADP/NTN network sites for 1983-86. *Atmospheric Environment*, 26A:2089–2103, 1992.