Generalized Scaling for the Constrained Maximum-Entropy Sampling Problem

Zhongzhu Chen[1], Marcia Fampa[2], and Jon Lee[1]

[1] University of Michigan, Ann Arbor, Michigan, USA `{zhongzhc,jonxlee}@umich.edu`
[2] Universidade Federal do Rio de Janeiro, Brasil
`fampa@cos.ufrj.br`

Abstract. The best techniques for the constrained maximum-entropy sampling problem, a discrete-optimization problem arising in the design of experiments, are via a variety of concave continuous relaxations of the objective function. A standard bound-enhancement technique in this context is *scaling*. We extend this technique to *generalized scaling*, we give mathematical results aimed at supporting algorithmic methods for computing optimal generalized scalings, and we give computational results demonstrating the usefulness of generalized scaling on benchmark problem instances.

Keywords: nonlinear 0/1-optimization · convex relaxation · maximum-entropy sampling

1 Introduction

Let C be a symmetric positive semidefinite matrix with rows/columns indexed from $N := \{1, 2, \ldots, n\}$, with $n > 1$. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. For $0 < s < n$, we define the *constrained maximum-entropy sampling problem*

$$z := \max \left\{ \operatorname{ldet} C[S(x), S(x)] \ : \ \mathbf{e}^\top x = s, \ x \in \{0, 1\}^n, \ Ax \leq b \right\}, \qquad \text{(CMESP)}$$

where $S(x)$ is the support of $x \in \{0, 1\}^n$, $C[S, S]$ is the principal submatrix of C indexed by S, and ldet is the natural logarithm of the determinant.

We refer to MESP when there are no constraints $Ax \leq b$, which was introduced in the "design of experiments" literature by [17]. MESP corresponds to the fundamental problem of choosing an s-subvector of a Gaussian random n-vector, so as to maximize the "differential entropy" (see [16]). MESP has been applied extensively in the field of environmental monitoring; see [11, Chapter 4], and the many references therein. Important for applications, the constraints $Ax \leq b$ of CMESP can model budget limitations, geographical considerations, and logical dependencies, for example. We assume $r := \operatorname{rank}(C) \geq s$, so that MESP always has a feasible solution with finite objective value.

CMESP serves as a nice example of a "non-factorable" mixed-integer nonlinear program. When C is a diagonal matrix, CMESP reduces to a general cardinality-constrained binary linear program. [1,2] established that when C is tridiagonal (or even when the support graph of C is a spider with a bounded number of legs), MESP is then polynomially solvable by dynamic programming.

[12] established that MESP is NP-hard and introduced a novel B&B (branch-and-bound) approach based on a spectral bound. [13] extended the spectral approach to CMESP. [5] and [6] developed a bound employing a novel convex relaxation. [3] developed the "BQP bound", using an extended formulation based on the Boolean quadric polytope. [4] introduced the "linx bound", based on a clever convex relaxation. [15] gave a novel "factorization bound" based on a somewhat mysterious convex relaxation. This was further developed by [14] and then [10]. [9] gave a methodology for combining multiple convex-optimization bounds to give improved bounds. All of these convex-optimization based bounds admit variable fixing methodology based on convex duality (see [11], for example). Another key idea for deriving bounds is "complementation". If C is invertible, we have

$$z = z(C^{-1}, n - s, -A, b - A\mathbf{e}) + \operatorname{ldet} C,$$

where $z(C^{-1}, n - s, -A, b - A\mathbf{e})$ denotes the optimal value of CMESP with C, s, A, b replaced by $C^{-1}, n - s, -A, b - A\mathbf{e}$, respectively. So we have a *complementary* CMESP problem and *complementary* bounds (i.e., bounds for the complementary problem plus $\operatorname{ldet} C$) immediately give us bounds on z. Some upper bounds on z also shift by $\operatorname{ldet} C$ under complementing, in which case there is no additional value in computing the complementary bound. Details on all of this can be found in [11].

Terminology. Throughout, we let $\Upsilon := (\gamma_1, \gamma_2, \ldots, \gamma_n)^{\top} \in \mathbb{R}_{++}^n$ be a "scaling vector". We refer to our bounds as *g-scaled* (i.e., *generalized scaled*), and when all elements of Υ are equal, we say *o-scaled* (i.e., *ordinary scaled*). If all elements of Υ are equal to 1, we say *un-scaled*.

Organization and contributions. In §2, we introduce the g-scaled BQP bound and establish its convexity in the log of the scaling vector, generalizing an important and practically-useful result (see [9, Thm. 11]). In §3, we introduce the g-scaled linx bound and establish its convexity in the log of the scaling vector, generalizing another very important and practically-useful result for o-scaling (see [9, Thm. 18]). These convexity results are key for the tractability of globally optimizing the scaling, something that we do not have for more general bound "masking" (see [7,8]). In §4, we introduce the g-scaled factorization bound, and we establish that g-scaling can significantly improve the factorization bound for CMESP, while the o-scaling cannot help it (see [10, Thm. 2.1]). We are also able to prove that for MESP, the all-ones vector is a stationary point for the bound as a function of the scaling vector. Therefore, g-scaling is unlikely to be helpful for MESP, similar to o-scaling. In §5, we present results of computational experiments, demonstrating the improvements on upper bounds and on the number of variables that can be fixed (using convex duality) due to g-scaling. In §6, we make some brief concluding remarks. In §7, we provide some proof sketches.

Notation. $\operatorname{Diag}(x) \in \mathbb{R}^{n \times n}$ makes a diagonal matrix from $x \in \mathbb{R}^n$. $\operatorname{diag}(X) \in \mathbb{R}^n$ extracts the diagonal of $X \in \mathbb{R}^{n \times n}$. We let $\mathbb{S}_+^n$ (resp., $\mathbb{S}_{++}^n$) be the set of positive semidefinite (resp., definite) symmetric matrices of order n. We let $\lambda_\ell(M)$ be the ℓ-th greatest eigenvalue of $M \in \mathbb{S}_+^n$. We denote by $\mathbf{e}$ an all-ones vector. For matrices A and B with the same shape, $A \circ B$ is the Hadamard (i.e., element-wise) product. We denote natural logarithm by log, and apply it component-wise to vectors.

2   BQP bound

We define the convex set

$$P(n, s) := \left\{ (x, X) \in \mathbb{R}^n \times \mathbb{S}^n \; : \; X - xx^{\top} \succeq 0, \; \operatorname{diag}(X) = x, \; \mathbf{e}^{\top} x = s, \; X\mathbf{e} = sx \right\}.$$

For $\Upsilon \in \mathbb{R}_{++}^n$, $x \in [0, 1]^n$ and $X \in \mathbb{S}_+^n$, we define

$$f_{\text{BQP}}(x, X; \Upsilon) := \operatorname{ldet}\left((\operatorname{Diag}(\Upsilon) C \operatorname{Diag}(\Upsilon)) \circ X + \operatorname{Diag}(\mathbf{e} - x) \right) - 2 \sum_{i=1}^n x_i \log \gamma_i$$

and the *g-scaled BQP bound*

$$z_{\text{BQP}}(\Upsilon) := \max \left\{ f_{\text{BQP}}(x, X; \Upsilon) \; : \; (x, X) \in P(n, S), \; Ax \leq b \right\}. \tag{BQP}$$

Note that we can interpret this bound as applying the un-scaled BQP bound to the symmetrically-scaled matrix $\operatorname{Diag}(\Upsilon) C \operatorname{Diag}(\Upsilon)$, and then correcting by $-2 \sum_{i=1}^n x_i \log \gamma_i$.

Theorem 1
1.i. $z \leq z_{BQP}$;

1.ii. For all $\varUpsilon \in \mathbb{R}^n_{++}$, $f_{BQP}(x, X; \varUpsilon)$ is concave on the feasible region of BQP;
1.iii. $z_{BQP}(\varUpsilon)$ is convex in $\log \varUpsilon$.

The BQP bound was first analyzed and developed in [3], establishing Thm. 1.*i* for $\varUpsilon = \mathbf{e}$. Thm. 1.*ii* is a result of [3], with details filled in by [11]. Thm. 1.*iii* significantly generalizes a result of [9], where it is established only for o-scaling: i.e., on $\{\varUpsilon = \gamma \mathbf{e} \ : \ \gamma \in \mathbb{R}_{++}\}$. The proof of Thm. 1.*iii* requires new ideas (see the proof sketch in the Appendix). Additionally, the result is quite important as it enables the use of readily available quasi-newton methods (like BFGS) for finding the globally optimal g-scaling for the BQP bound.

3 linx bound

For $\varUpsilon \in \mathbb{R}^n_{++}$ and $x \in [0,1]^n$, we define

$$f_{\text{linx}}(x; \varUpsilon) := \tfrac{1}{2} \left(\text{ldet} \left(\text{Diag}(\varUpsilon) C \, \text{Diag}(x) C \, \text{Diag}(\varUpsilon) + \text{Diag}(\mathbf{e} - x) \right) \right) - \sum_{i=1}^n x_i \log \gamma_i$$

and the *g-scaled linx bound*

$$z_{\text{linx}}(\varUpsilon) := \max \left\{ f_{\text{linx}}(x; \varUpsilon) \ : \ \mathbf{e}^\top x = s, \ 0 \le x \le \mathbf{e}, \ Ax \le b \right\}. \tag{linx}$$

Note that we *cannot* interpret this bound as applying the un-scaled linx bound to the row-scaled matrix $\text{Diag}(\varUpsilon)C$, because we would lose symmetry.

Theorem 2
2.i. $z \le z_{linx}$;
2.ii. For all $\varUpsilon \in \mathbb{R}^n_{++}$, $f_{linx}(x; \varUpsilon)$ is concave on the feasible region of linx;
2.iii. $z_{linx}(\varUpsilon)$ is convex in $\log \varUpsilon$.

The linx bound was first analyzed and developed in [4], establishing Thm. 2.*i* for $\varUpsilon = \mathbf{e}$. Thm. 2.*ii* is a result of [4], with details filled in by [11]. Thm. 2.*iii* generalizes a result of [9], where it is established only for o-scaling: i.e., on $\{\varUpsilon = \gamma \mathbf{e} \ : \ \gamma \in \mathbb{R}_{++}\}$. The proof of Thm. 2.*iii* requires new ideas (see the proof sketch in the Appendix). Additionally, the result is quite important as it enables the use of readily available quasi-newton methods (like BFGS) for finding the globally optimal g-scaling for the linx bound.

4 Factorization bound

Lemma 3 *(see [15, Lem. 14]) Let $\lambda \in \mathbb{R}^k_+$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_k$, and let $0 < s \le k$. There exists a unique integer ι, with $0 \le \iota < s$, such that $\lambda_\iota > \frac{1}{s-\iota} \sum_{\ell=\iota+1}^k \lambda_\ell \ge \lambda_{\iota+1}$, with the convention $\lambda_0 = +\infty$.*

Now, suppose that $\lambda \in \mathbb{R}^k_+$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_k$. Given an integer s with $0 < s \le k$, let ι be the unique integer defined by Lem. 3. We define $\phi_s(\lambda) := \sum_{\ell=1}^\iota \log \lambda_\ell + (s - \iota) \log \left(\frac{1}{s-\iota} \sum_{\ell=\iota+1}^k \lambda_\ell \right)$. Next, for $X \in \mathbb{S}^k_+$, we define $\varGamma_s(X) := \phi_s(\lambda_1(X), \ldots, \lambda_k(X))$.

   Suppose that the rank of C is $r \ge s$. Then we factorize $C = FF^\top$, with $F \in \mathbb{R}^{n \times k}$, for some k satisfying $r \le k \le n$. Now, for $\varUpsilon \in \mathbb{R}^n_{++}$ and $x \in [0,1]^n$, we define $F_{\text{DDFact}}(x; \varUpsilon) := \sum_{i=1}^n \gamma_i x_i F_{i \cdot}^\top F_{i \cdot}$. Finally, we define $f_{\text{DDFact}}(x; \varUpsilon) := \varGamma_s(F_{\text{DDFact}}(x; \varUpsilon)) - \sum_{i=1}^n x_i \log \gamma_i$ and the *g-scaled factorization bound*

$$z_{\text{DDFact}}(\varUpsilon) := \max \left\{ f_{\text{DDFact}}(x; \varUpsilon) \ : \ \mathbf{e}^\top x = s, \ 0 \le x \le \mathbf{e}, \ Ax \le b \right\}. \tag{DDFact}$$

Noticing that $F_{\mathrm{DDFact}}(x;\Upsilon) := F^\top \mathrm{Diag}\left(\sqrt{\Upsilon}\right)\mathrm{Diag}(x)\,\mathrm{Diag}\left(\sqrt{\Upsilon}\right)F$, we can interpret this bound as applying the un-scaled DDFact bound to the symmetrically-scaled matrix $\mathrm{Diag}\left(\sqrt{\Upsilon}\right)C\,\mathrm{Diag}\left(\sqrt{\Upsilon}\right)$, and then correcting by $-\sum_{i=1}^n x_i \log \gamma_i$.

Definition 4 *For any x feasible to DDFact, suppose the eigenvalues of $F_{DDFact}(x;\Upsilon)$ are $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_k = 0$, where $r \in [s,k]$ and $F_{DDFact}(x;\Upsilon) = Q\,\mathrm{Diag}(\lambda)Q$ with an orthonormal matrix Q. Define $\beta(\lambda) := (\beta_1, \beta_2, \ldots, \beta_k)^\top$ such that*

$$\beta_i := \frac{1}{\lambda_i},\ \forall\ i \in [1,\iota],\ \ \beta_i := \frac{s - \iota}{\sum_{i \in [\iota+1,k]} \lambda_i},\ \forall\ i \in [\iota+1,k],$$

where ι is the unique integer in Lemma 3.

Theorem 5

5.i. $z \leq z_{DDFact}$;

5.ii. For all $\Upsilon \in \mathbb{R}_{++}^n$, $f_{DDFact}(x;\Upsilon)$ is concave on the feasible region of DDFact;

5.iii. For all $\Upsilon \in \mathbb{R}_{++}^n$ and $x \geq 0$ in the domain of $f_{DDFact}(x;\Upsilon)$, let $T(x;\Upsilon) :=$ $\mathrm{diag}\left(F_{DDFact}(x;\Upsilon)Q\,\mathrm{Diag}\left(\beta(\lambda)\right)Q^\top F_{DDFact}(x;\Upsilon)^\top\right) - \log\Upsilon$ where $Q, \beta(\lambda)$ are defined in 4, then

$$\lim_{\substack{\|\hat{x}-x\| \to 0\ :\\ \hat{x} \geq 0\ is\ in\ the\ domain\\ of\ f_{DDFact}(x;\Upsilon)}} \frac{\left|f_{DDFact}(\hat{x};\Upsilon) - f_{DDFact}(x;\Upsilon) - T(x;\Upsilon)^\top(\hat{x}-x)\right|}{\|\hat{x}-x\|} = 0.$$

5.iv. For all x feasible in the domain of $f_{DDFact}(x;\Upsilon)$, $f_{DDFact}(x;\Upsilon)$ is differentiable in Υ at all $\Upsilon \in \mathbb{R}_{++}^n$. In particular, for MESP, let x^* to be one optimal solution to DDFact, then we have

$$\left.\frac{\partial f_{DDFact}(x^*;\Upsilon)}{\partial \Upsilon}\right|_{\Upsilon=\mathbf{e}} = 0.$$

The DDFact bound was first analyzed and developed in [15], establishing Thm. 5.*i* for $\Upsilon = \mathbf{e}$, and developed further in [14]. We note that the o-scaled factorization bound for CMESP is invariant under the scale factor (see [10]), so the use of any type of scaling in the context of the DDFact bound is completely new. Thms. 5.*iii-iv* are the first differentiablity results of any type for the DDFact bound. The proof methods (sketched in the Appendix) are quite technical and novel. Furthermore, they explain the success of our quasi-newton based methods for calculating optimal g-scalings for the DDFact bound, not anticipated by previous works which exposed only subgradients connected to DDFact. As we will see in §5, g-scaling can improve the DDFact bound for CMESP. These observations and Thm. 5.*iv* leave open the interesting question of whether g-scaling can help the DDFact bound for MESP; we can interpret Thm. 5.*iv* as a partial result toward a negative answer.

5   Numerical results

We experimented on benchmark instances of MESP, using three covariance matrices that have been extensively used in the literature, with $n = 63, 90, 124$ (see, e.g., [12,13,6,3,4]). For testing CMESP, we included five side constraints $a_i^\top x \leq b_i$, for $i = 1, \ldots, 5$, in MESP. As there is no benchmark data for the side constraints, we have generated them randomly. For each n, the left-hand side of constraint i is given by a uniformly-distributed random vector a_i with integer components between -2 and 2. The right-hand side of the constraints was selected so that, for every s considered in the experiment, the best known solution of the instance of MESP is violated by at least one constraint.

For each n, we considered instances of MESP and CMESP with a wide range of s. We ran our experiments under Windows, on an Intel Xeon E5-2667 v4 @ 3.20 GHz processor equipped with 8 physical cores (16 virtual cores) and 128 GB of RAM. We implemented our code in `Matlab` using the solvers `SDPT3` v. 4.0 for BQP, and `Knitro` v. 12.4 for linx and DDFact, and optimizing scaling vectors Υ using a BFGS algorithm, and the o-scaling parameters γ using the Newton's method. Besides solving the relaxations to get upper bounds for our test instances of MESP and CMESP, we compute lower bounds with a heuristic of [13, Sec. 4] and then a local search (see [12, Sec. 4]).

In Fig. 1, we show the impact of g-scaling on the linx bound for MESP on the three benchmark covariance matrices. For the $n = 63$ matrix, we also show the impact of g-scaling on the BQP bound. The DDFact and complementary DDFact bounds are only considered in the experiments for CMESP, as the g-scaling methodology was only able to improve these bounds when side constraints were added to MESP. The plots on the left in Fig. 1 present the "integrality gap decrease ratios", given by the difference between the integrality gaps using o-scaling and the integrality gaps using g-scaling, divided by the integrality gaps using o-scaling. The integrality gaps are given by the difference between the upper bounds computed with the relaxations and lower bounds given by heuristic solutions. We see that larger n leads to larger maximum ratios. We also see that the g-scaling methodology is effective in reducing all bounds evaluated, especially the linx bound. Even for the most difficult instances, with intermediate values of s, we have some improvement on the bounds, which can be effective in the branch-and-bound context where the bounds would ultimately be applied. The plots on the right in Fig. 1 present the integrality gaps, and we see that even when the integrality gaps given by the o-scaling are less than 1, g-scaling can reduce them.

In Fig. 2, we show for CMESP, similar results to the ones shown in Fig. 1, except that now we also present the effect of g-scaling on the DDFact and the complementary DDFact bounds. We see from the integrality gap decrease ratios that when side constraints are added to MESP, the g-scaling is, in general, more effective in reducing the gaps given by o-scaling. We also see that, it is particularly effective in reducing the DDFact and complementary DDFact bounds. Especially for the $n = 124$ matrix, we see a significant reduction on the gaps given by complementary DDFact and DDFact, for s smaller and greater than 50, respectively.

We also investigated how the improvement of g-scaling over o-scaling for the linx bound can increase the possibility of fixing variables in MESP and CMESP. The methodology for fixing variables is based on convex duality and has been applied since the first convex relaxation was proposed for these problems in [5]. When a lower bound for each problem is available, the dual solution of the relaxation can potentially be used to fix variables at $0/1$ values (see [11], for example). This is an important feature in the B&B context. The methodology may be able to fix a number of variables when the relaxation generates a strong bound, and in doing so, it reduces the size of the successive subproblems and improves the bounds computed for them.

In Table 1, we show the impact of using g-scaled linx, compared to o-scaled linx, on an iterative procedure where we solve linx, DDFact, and complementary DDFact, fixing variables at $0/1$ whenever possible. In both cases, we update the scaling parameter every time we solve linx. For o-scaling, we optimize the scalar γ by applying Newton steps until the absolute value of the derivative is less than 10^{-10}. For g-scaling, we optimize the vector Υ by applying up to 10 BFGS steps, taking $\gamma \mathbf{e}$ as a starting point. We limit the number of BFGS steps in this experiment to get closer to what might be practical within B&B. We present in the columns of Table 1, the following information from left to right: The problem considered, n, the range of s considered, the scaling, the number of instances solved (one for each s considered), the number of instances on which we could fix at least one variable ("inst fix"), the total number of variables fixed on all instances solved ("var fix"), the %-improvement of g-scaling over o-scaling for the two last statistics. Additionally, to better understand how well our methods works for MESP as n grows, we also experimented with a covariance

Fig. 1: Comparison between g-scaling and o-scaling for MESP

Fig. 2: Comparison between g-scaling and o-scaling for CMESP

matrix of order $n = 300$, which is a principal submatrix of the covariance matrix of order $n = 2000$ used as a benchmark in the literature (see [14,10]). First, we see that, except for the number of instances of MESP with $n = 124$ and $n = 300$ on which we could fix variables, there is always an improvement. The improvement becomes very significant when side constraints are considered. We note that the number of variables fixed, reported on Table 1, refers only to the root nodes of the B&B algorithm and indicates a promising approach to reduce the B&B enumeration.

	n	s	scaling	s	Number of inst fix	var fix	Improvement inst fix	var fix
MESP	63	[2,62]	o	61	41	1123		
			g	61	42	1140	2.44%	1.51%
	90	[2,89]	o	88	41	1741		
			g	88	42	1790	2.44%	2.81%
	124	[2,123]	o	122	35	3322		
			g	122	35	3353	0.00%	0.93%
	300	[80,120]	o	41	41	8382		
			g	41	41	10753	0.00%	28.3%
CMESP	63	[3, 52]	o	50	22	371		
			g	50	28	537	27.27%	44.74%
	90	[4, 87]	o	84	26	606		
			g	84	37	1048	42.31%	72.94%
	124	[11, 110]	o	100	9	197		
			g	100	33	1120	266.67%	468.53%

Table 1: Impact of g-scaling on variable fixing

The experiments with the fixing methodology show that g-scaling can effectively lead to a positive impact on the solution of MESP and CMESP, especially of the latter.

6 Conclusion

We have seen that g-scaling can lead to improvements in upper bounds and variable fixing for MESP and very good improvements for CMESP. In future work, we will implement this in an efficient manner, within a B&B algorithm. In that context, it is important to efficiently use parent scaling vectors to warm-start the optimization of scaling vectors for children (see [4]). An open question is whether g-scaling can help the DDFact bound for MESP. Thm. 5.*iv* is a partial result toward a negative answer. Finally, there is another convex-optimization bound, the so-called "NLP bound" (see [6]), and it appears to be more difficult to get mathematical results on optimizing a g-scaling version of that bound; but this is a good direction to explore.

7 Appendix: Proof sketches

Proof sketch [Thm. 1]

1.i: Suppose that the optimal solution to CMESP is x^*, let $X^* := x^* (x^*)^\top$. Then $(x^*, X^*) \in P(n, S)$, and $f_{\mathrm{BQP}} (x^*, X^*; \Upsilon) = \operatorname{ldet} C [S(x^*), S(x^*)]$. Thus $z_{\mathrm{BQP}}(\Upsilon) \geq f_{\mathrm{BQP}} (x^*, X^*; \Upsilon) = z$.

1.ii: This is essentially a result of [3], with details filled in by [11].

1.iii: Let $F_{\mathrm{BQP}}(x, X; \Upsilon) := (\operatorname{Diag}(\Upsilon) C \operatorname{Diag}(\Upsilon)) \circ X + \operatorname{Diag}(\mathbf{e} - x)$ and $A_{\mathrm{BQP}}(X; \Upsilon) := (\operatorname{Diag}(\Upsilon) C \operatorname{Diag}(\Upsilon)) \circ X$. Then given (x, X) in the domain of $f_{\mathrm{BQP}}(x, X; \Upsilon)$ and feasible to

BQP, we have

$$\frac{\partial f_{\mathrm{BQP}}^2\,(x,X;\Upsilon)}{\partial\,(\log\Upsilon)^2}$$

$$= \quad 4\,\mathrm{Diag}(x-\mathbf{e})\,\mathrm{Diag}\left(\mathrm{diag}\left(F_{\mathrm{BQP}}(x,X;\Upsilon)^{-1}\right)\right)$$
$$- 4\,\mathrm{Diag}(x-\mathbf{e})\left(F_{\mathrm{BQP}}(x,X;\Upsilon)^{-1}\circ F_{\mathrm{BQP}}(x,X;\Upsilon)^{-1}\right)\mathrm{Diag}(x-\mathbf{e}).$$

When $x < \mathbf{e}$ and $X \succ 0$, let $D_{\mathrm{BQP}}(x) := (\mathrm{Diag}(\mathbf{e}-x))^{1/2} \succ 0$ and further, $E_{\mathrm{BQP}}(x,X;\Upsilon) := (D_{\mathrm{BQP}}(x))^{-1}\,A_{\mathrm{BQP}}(X;\Upsilon)\,(D_{\mathrm{BQP}}(x))^{-1} \succ 0$. It can be shown that

$$\frac{\partial f_{\mathrm{BQP}}^2\,(x,X;\Upsilon)}{\partial\,(\log\Upsilon)^2} = 4\,(E_{\mathrm{BQP}}(x,X;\Upsilon)+I)^{-1}\circ\left((E_{\mathrm{BQP}}(x,X;\Upsilon))^{-1}+I\right)^{-1} \succ 0.$$

On the one hand, given $\Upsilon > 0$, $\frac{\partial f_{\mathrm{BQP}}^2(x,X;\Upsilon)}{\partial(\log\Upsilon)^2}$ is analytical on (x,X) in the domain of $f_{\mathrm{BQP}}\,(x,X;\Upsilon)$. On the other hand, the feasible set of BQP is compact. Therefore, given (x,X) in the domain of $f_{\mathrm{BQP}}\,(x,X;\Upsilon)$ and feasible to BQP, there exists $\epsilon > 0$ such that $\mathcal{N}_{(x,X)} := \{(x',X') : \|x-x'\| \le \epsilon\}$ $\cap\{\text{domain of } f_{\mathrm{BQP}}\,(x,X;\Upsilon)\}\cap\{\text{feasible set to BQP}\}$ is compact. This implies that if $\frac{\partial f_{\mathrm{BQP}}^2(x,X;\Upsilon)}{\partial(\log\Upsilon)^2}$ $\prec 0$, then $\exists\,(x',X') \in \mathcal{N}_{(x,X)}$ such that $x' < \mathbf{e}$ and $\frac{\partial f_{\mathrm{BQP}}^2(x',X';\Upsilon)}{\partial(\log\Upsilon)^2} \prec 0$, a contradiction. So, for each fixed (x,X) such above, $f_{\mathrm{BQP}}\,(x,X;\Upsilon)$ is convex in $\log\Upsilon$. Because $z_{\mathrm{BQP}}(\Upsilon)$ is the pointwise maximum over all $(x,X) \in P(n,x)$, it is convex in $\log\Upsilon$.   □

Proof sketch [Thm. 2]

2.i: Suppose that the optimal solution to CMESP is x^*; then we can show $f_{\mathrm{linx}}(x^*;\Upsilon) = \mathrm{ldet}\,C\,[S(x^*),S(x^*)]$. Thus $z_{\mathrm{linx}}(\Upsilon) \ge f_{\mathrm{linx}}(x^*;\Upsilon) = z$.

2.ii: This is essentially a result of [4], with details filled in by [11].

2.iii: Let $F_{\mathrm{linx}}(x;\Upsilon) := \mathrm{Diag}(\Upsilon)C\,\mathrm{Diag}(x)C\,\mathrm{Diag}(\Upsilon) + \mathrm{Diag}(\mathbf{e}-x)$ and $A_{\mathrm{linx}}(x;\Upsilon) := \mathrm{Diag}(\Upsilon)C\,\mathrm{Diag}(x)C\,\mathrm{Diag}(\Upsilon)$. Let $D_{\mathrm{linx}}(x) := (\mathrm{Diag}(\mathbf{e}-x))^{1/2}$ and $E_{\mathrm{linx}}(x;\Upsilon) := (D_{\mathrm{linx}}(x))^{-1}\,A_{\mathrm{linx}}(x;\Upsilon)\,(D_{\mathrm{linx}}(x))^{-1}$ when $x < \mathbf{e}$. Then similar to 1.iii.   □

Proof sketch [Thm. 5]

5.i: This is essentially a result of [10].

5.ii: This is essentially a result of [15], with details filled in by [11].

5.iii: Based on [14, Proposition 2] and [18, Theorem 2.4.18], we can show that for $x, \hat{x}$ in the domain of $f_{\mathrm{DDFact}}(x;\Upsilon)$, the directional derivative of $f_{\mathrm{DDFact}}(x;\Upsilon)$ at x in direction $\frac{\hat{x}-x}{\|\hat{x}-x\|}$ is $T(x;\Upsilon)^\top\left(\frac{\hat{x}-x}{\|\hat{x}-x\|}\right)$ where

$$T(x;\Upsilon) := \mathrm{diag}\left(F_{\mathrm{DDFact}}(x;\Upsilon)Q\,\mathrm{Diag}\,(\beta(\lambda))\,Q^\top F_{\mathrm{DDFact}}(x;\Upsilon)^\top\right) - \log\Upsilon.$$

We first show two preliminary results:

(a) It can be shown that $f_{\mathrm{DDFact}}(x;\Upsilon)$ is continuous on its domain. Then, because the feasible region of DDFact is compact, given x, $\exists\,\tilde{r} > 0$ such that $\forall r \le \tilde{r}$, $\mathcal{B}_r(x) := \{y : \|y-x\| \le r\}$ is included in the domain of $f_{\mathrm{DDFact}}(x;\Upsilon)$. Furthermore, the intersection of the feasible region of DDFact and $\mathcal{B}_r(x)$ is compact and included in the domain of $f_{\mathrm{DDFact}}(x;\Upsilon)$, denoted as $\mathcal{N}_x^r$, which implies uniform continuity of $f_{\mathrm{DDFact}}(x;\Upsilon)$ on $\mathcal{N}_x^r$.

(b) Let $\mathcal{C}(x) := \{y : \|y-x\| = 1\}$. $\forall\epsilon > 0$, by the Heine-Borel Theorem, $\exists$ a finite set $F \subset \mathcal{C}(x)$ such that $\forall y \in \mathcal{C}(x)$, $\exists u \in F$ such that $\|y-u\| < \epsilon$.

Now we are ready to prove Thm. 5.iii. We will assume that $T(x;\varUpsilon) \neq 0$ for simplicity. First, by the uniform continuity in (a), given $\epsilon > 0$ and $r \leq \tilde{r}$, $\exists \delta \in (0,\epsilon)$ such that $\forall x_1, x_2 \in \mathcal{N}_x^r$ with $\|x_1 - x_2\| \leq \frac{\delta}{\|T(x;\varUpsilon)\|}$, we have $|f_{\mathrm{DDFact}}(x_1;\varUpsilon) - f_{\mathrm{DDFact}}(x_2;\varUpsilon)| < \epsilon$. Second, by (b), $\exists F_\epsilon$ such that $\forall y \in \mathcal{C}(x)$, $\exists u \in \mathcal{C}(x)$ such that $\|y - u\| < \frac{\delta}{\|T(x;\varUpsilon)\|\cdot\tilde{r}}$. Third, by the existence of directional derivatives of $f_{\mathrm{DDFact}}(x;\varUpsilon)$ at x, $\forall r \leq \tilde{r}$ small enough, we have $\forall u \in F_\epsilon, t \leq r$, $\left|f_{\mathrm{DDFact}}(x + tu;\varUpsilon) - f_{\mathrm{DDFact}}(x;\varUpsilon) - tT(x;\varUpsilon)^\top u\right| < \epsilon$. Fourth, $\forall \hat{x} \in \mathcal{N}_x^r$, $\frac{\hat{x}}{\|\hat{x}\|} \in \mathcal{C}(x)$ and $\|\hat{x}\| \leq r \leq \tilde{r}$, and by the second argument, $\exists u \in F_\epsilon$ such that $\|\hat{x} - \|\hat{x}\| \cdot u\| = \|\hat{x}\| \cdot \|\hat{x}/\|\hat{x}\| - u\| < \|\hat{x}\| \cdot \frac{\delta_1}{\|T(x;\varUpsilon)\|\cdot\tilde{r}} \leq \frac{\delta}{\|T(x;\varUpsilon)\|}$.

In all, given $\epsilon > 0$, $\exists r \leq \tilde{r}$ and F_ϵ such that $\forall \hat{x} \in \mathcal{N}_x^r$, $\exists u \in F_\epsilon$ such that

$$
\begin{aligned}
&\left|f_{\mathrm{DDFact}}(\hat{x};\varUpsilon) - f_{\mathrm{DDFact}}(x;\varUpsilon) - T(x;\varUpsilon)^\top(\hat{x} - x)\right| \\
={} &\ |f_{\mathrm{DDFact}}(\hat{x};\varUpsilon) - f_{\mathrm{DDFact}}(\|\hat{x}\| \cdot u;\varUpsilon)| \\
&+ \left|f_{\mathrm{DDFact}}(\|\hat{x}\| \cdot u;\varUpsilon) - f_{\mathrm{DDFact}}(x;\varUpsilon) - T(x;\varUpsilon)^\top(\|\hat{x}\| \cdot u - x)\right| \\
&+ \left|T(x;\varUpsilon)^\top(\|\hat{x}\| \cdot u - \hat{x})\right| \\
<{} &\ \epsilon + \epsilon + \frac{\delta}{\|T(x;\varUpsilon)\|} \cdot \|T(x;\varUpsilon)\| \ < \ 3\epsilon,
\end{aligned}
$$

which implies the result.

5.iv: By switching the role of x and $\varUpsilon$, we can show that for any x in the domain of $f_{\mathrm{DDFact}}(x;\varUpsilon)$, there is a vector $\tilde{T}(x;\varUpsilon) \in \mathbb{R}^n$ such that

$$
\lim_{\substack{\|h\| \to 0 \,:\\ \varUpsilon + h > 0}} \frac{|f_{\mathrm{DDFact}}(x;\varUpsilon + h) - f_{\mathrm{DDFact}}(x;\varUpsilon) - \tilde{T}(x;\varUpsilon)^\top h|}{\|h\|} = 0.
$$

When $\varUpsilon > 0$ falls into the interior of the positive cone, the above result is equivalent to $f_{\mathrm{DDFact}}(x;\mathbf{e})$ being differentiable in $\varUpsilon$.

Letting $T(x^*;\varUpsilon)$ be as defined in the proof of Thm. 5.iii, the remaining result is equivalent to $x^* \circ (T(x^*;\mathbf{e}) - \mathbf{e}) = 0$, which is further equivalent to

$$
(T(x^*;\mathbf{e}))_i = 1, \ \forall x_i^* > 0.
$$

Suppose that σ is a permutation of $1, \cdots, n$ such that $(T(x^*;\mathbf{e}))_{\sigma(1)} \geq \cdots \geq (T(x^*;\mathbf{e}))_{\sigma(n)}$. By [14] and KKT conditions for DDFact, we have

$$
\sum_{i \in \{1,2,\ldots,n\}} x^*_{\sigma(i)} (T(x^*;\mathbf{e}))_{\sigma(i)} = \sum_{i \in \{1,2,\ldots,s\}} (T(x^*;\mathbf{e}))_{\sigma(i)} = s.
$$

On the other hand, if $x^*_{\sigma(i)} = 1$, we have

$$
\begin{aligned}
(T(x^*;\mathbf{e}))_{\sigma(i)} &= F_{\sigma(i)\cdot} Q \operatorname{Diag}(\beta(\lambda)) Q^\top F_{\sigma(i)\cdot}^\top \ \leq \ F_{\sigma(i)\cdot} (F_{\mathrm{DDFact}}(x;\mathbf{e}))^\dagger F_{\sigma(i)\cdot}^\top \\
&= F_{\sigma(i)\cdot} \left(F_{\sigma(i)\cdot}^\top F_{\sigma(i)\cdot} + \sum_{j \neq \sigma(i)} x_j^* F_{j\cdot}^\top F_{j\cdot}\right)^\dagger F_{\sigma(i)\cdot}^\top \\
&\leq F_{\sigma(i)\cdot} \left(F_{\sigma(i)\cdot}^\top F_{\sigma(i)\cdot}\right)^\dagger F_{\sigma(i)\cdot}^\top \ = \ 1
\end{aligned}
$$

where the first inequality is due to $Q \operatorname{Diag}(\beta(\lambda)) Q^\top F_{\mathrm{DDFact}}(x;\varUpsilon) \preceq I$ and that the two matrices can be simultaneously diagonalized by Q, and the second inequality is by the Sherman–Morrison formula for the pseudo-inverse.

The above two formulae, together with the KKT conditions and $\sum_{i \in [n]} x^*_{\sigma(i)} = s$, imply that $(T(x^*;\mathbf{e}))_{\sigma(1)} = \cdots = (T(x^*;\mathbf{e}))_{\sigma(s)} = 1$, and $\forall i > s$ such that $x^*_{\sigma(i)} > 0$, $(T(x^*;\mathbf{e}))_{\sigma(i)} = (T(x^*;\mathbf{e}))_{\sigma(s)} = 1$, which finishes the proof. $\qquad\square$

References

1. Al-Thani, H., Lee, J.: Tridiagonal maximum-entropy sampling and tridiagonal masks. LAGOS 2021 proceedings, Procedia Computer Science **195**, 127–134 (2021)
2. Al-Thani, H., Lee, J.: Tridiagonal maximum-entropy sampling and tridiagonal masks (2021), preprint at: `http://arxiv.org/abs/2112.12814`
3. Anstreicher, K.M.: Maximum-entropy sampling and the Boolean quadric polytope. Journal of Global Optimization **72**(4), 603–618 (2018)
4. Anstreicher, K.M.: Efficient solution of maximum-entropy sampling problems. Operations Research **68**(6), 1826–1835 (2020)
5. Anstreicher, K.M., Fampa, M., Lee, J., Williams, J.: Continuous relaxations for constrained maximum-entropy sampling. In: Integer Programming and Combinatorial Optimization (Vancouver, BC, 1996), Lecture Notes in Computer Science, vol. 1084, pp. 234–248. Springer, Berlin (1996)
6. Anstreicher, K.M., Fampa, M., Lee, J., Williams, J.: Using continuous nonlinear relaxations to solve constrained maximum-entropy sampling problems. Mathematical Programming, Series A **85**(2), 221–240 (1999)
7. Anstreicher, K.M., Lee, J.: A masked spectral bound for maximum-entropy sampling. In: mODa 7—Advances in model-oriented design and analysis, pp. 1–12. Contrib. Statist., Physica, Heidelberg (2004)
8. Burer, S., Lee, J.: Solving maximum-entropy sampling problems using factored masks. Mathematical Programming **109**(2-3, Ser. B), 263–281 (2007)
9. Chen, Z., Fampa, M., Lambert, A., Lee, J.: Mixing convex-optimization bounds for maximum-entropy sampling. Mathematical Programming, Series B **188**, 539–568 (2021)
10. Chen, Z., Fampa, M., Lee, J.: On computing with some convex relaxations for the maximum-entropy sampling problem. INFORMS Journal on Computing (2023), `https://doi.org/10.1287/ijoc.2022.1264`
11. Fampa, M., Lee, J.: Maximum-Entropy Sampling: Algorithms and Application. Springer International Publishing (2022), `https://doi.org/10.1007/978-3-031-13078-6`
12. Ko, C.W., Lee, J., Queyranne, M.: An exact algorithm for maximum entropy sampling. Operations Research **43**(4), 684–691 (1995)
13. Lee, J.: Constrained maximum-entropy sampling. Operations Research **46**(5), 655–664 (1998)
14. Li, Y., Xie, W.: Best principal submatrix selection for the maximum entropy sampling problem: Scalable algorithms and performance guarantees (2020), preprint at: `https://arxiv.org/abs/2001.08537`
15. Nikolov, A.: Randomized rounding for the largest simplex problem. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing. pp. 861–870 (2015)
16. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal **27**(3), 379–423 (1948)
17. Shewry, M.C., Wynn, H.P.: Maximum entropy sampling. Journal of Applied Statistics **46**, 165–170 (1987)
18. Zalinescu, C.: Convex Analysis in General Vector Spaces. World Scientific (2002)