RC 21679 (97681) (1 March 2000)
Computer Science/Mathematics

IBM Research Report

New Upper Bounds for Maximum-Entropy Sampling

Alan Hoffman

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Jon Lee

Department of Mathematics
University of Kentucky
Lexington, Kentucky

Joy Williams

Department of Mathematics
Earlham College
Richmond, Indiana

IBM Research Division
Almaden · Austin · Beijing · Delhi · Haifa · T.J. Watson · Tokyo · Zurich

NEW UPPER BOUNDS FOR MAXIMUM-ENTROPY SAMPLING

Alan HOFFMAN[1], Jon LEE[2] & Joy WILLIAMS[3]

February 2000

Abstract

We develop and experiment with new upper bounds for the constrained maximum-entropy sampling problem. Our partition bounds are based on Fischer's inequality. F urther new upper bounds combine the use of Fischer's inequality with previously developed bounds. We demonstrate this in detail by using the partitioning idea to strengthen the spectral bounds of Ko, Lee and Queyranne and of Lee. Computational evidence suggests that these bounds may be useful in solving problems to optimality in a branch-and-bound framework.

Keywords: experimental design, design of experiments, entropy, maximum-entropy sampling, spectral bound, Lagrangian, Fischer's inequality, branch-and-bound, matching, set partitioning.

[1]Dept. of Mathematical Sciences, T.J. Watson Research Center, IBM, Yorktown Heights, N.Y., U.S.A.

[2]Dept. of Mathematics, 715 POT, University of Kentucky, Lexington, KY 40506-0027, U.S.A.; CORE, 34 voie du Roman Pays, Université Catholique de Louvain, 1348 Louvain-la-Neuve, BELGIUM; Email: jlee@ms.uky.edu . WWW: http://www.ms.uky.edu/~jlee .

[3]Department of Mathematics, Earlham College, Richmond, Indiana, U.S.A.

Introduction

The constrained maximum-entropy sampling problem is an important problem in the *design of experiments*. For example, such problems arise in the design of monitoring networks (see Ko, Lee and Queyranne [KLQ95] and the references therein). Let C be an $n \times n$ real symmetric positive definite matrix with row (and column) indices $N = \{1, 2, \ldots, n\}$. For $S, T \subset N$, we let $C[S, T]$ denote the submatrix of C with row indices S and column indices T. We let $\det C[S, S]$ denote the determinant of the principal submatrix $C[S, S]$, with the convention that $\det C[\emptyset, \emptyset] = 1$. Let s be an integer satisfying $0 < s \leq n$. Let M be a finite index set. Let the a_{ij} and b_i be real numbers, for $i \in M$ and $j \in N$. The constrained maximum-entropy sampling problem is to solve

$$z := \max_{\substack{S \subset N: \\ |S| = s}} \ \ln \det \, C[S, S] \ ; \tag{1}$$

$$\text{subject to } \sum_{j \in S} a_{ij} \leq b_i \ , \ \forall \ i \in M \ . \tag{2}$$

In the context of the design of experiments, C is a covariance matrix of a set of Gaussian random variables, and, up to some constants, $\ln \det \, C[S, S]$ is the entropy of the set of random variables associated with S. So the constrained maximum-entropy sampling problem amounts to finding a most-informative s-subset of a set of n Gaussian covariates, subject to some side constraints.

The constrained maximum-entropy sampling problem is already NP-Hard in the important special case with $M = \emptyset$. Exact algorithms are based on the branch-and-bound framework (see [KLQ95]). One upper bound for z is the *spectral bound*

$$v := \sum_{l=1}^{s} \ \ln \, \lambda_l \, (C) \ \ , \tag{3}$$

where λ_l denotes the l^{th} greatest eigenvalue (see [KLQ95]). This bound does not take advantage of the side-constraints (2). Lee [Lee98] strengthened this bound, taking advantage of the side-constraints, using a Lagrangian methodology. Specifically, Lee introduced the *Lagrangian spectral bound*

$$v(A, b) := \min_{w \in \mathbb{R}_+^M} v(A, b, w) \ , \tag{4}$$

where

$$v(A, b, w) := \sum_{l=1}^{s} \ln \ \lambda_l (D_w C D_w) \ + \ \sum_{i \in M} w_i b_i \ , \tag{5}$$

and

$$D_w := \text{diag}_j \left\{ \exp \left\{ -\frac{1}{2} \sum_{i \in M} w_i a_{ij} \right\} \right\} . \tag{6}$$

Lee demonstrated that $v(A, b, w)$ is convex in w, and he described descent methods for calculating the minimizing w in (4).

Other bounds are based on convex-programming formulations (see [AFLW96, AFLW99, Lee00]).

Branching consists of fixing an index j out of the solution (deleting row and column j of C) or fixing an index j in to the solution (pivoting in C on C_{jj}, decrementing s by one, and decreasing each b_i by a_{ij}. Lower bounds for fathoming are determined by heuristic search methods. See [KLQ95, Lee98, AFLW96, AFLW99] for more details.

In [AFLW96, AFLW99], we described how different bounds can be calculated by considering a "complementary problem" (when C is nonsingular). Specifically, since

$$\ln \det C[S, S] = \ln \det C + \ln \det C^{-1}[N \setminus S, N \setminus S]$$

(see for example [HJ85], Section 0.8.4 ("Minors of the inverse"), p. 21), we have the *complementary problem*

$$z = \ln \det C + \max_{\substack{N \setminus S \subset N: \\ |N \setminus S| = n-s}} \ln \det C^{-1}[N \setminus S, N \setminus S] \; ;$$
$$\text{subject to} \sum_{j \in N \setminus S} (-a_{ij}) \leq b_i - \sum_{j \in N} a_{ij} \;, \; \forall \; i \in M \; .$$

We can calculate a bound with respect to choosing the $n - s$ element set $N \setminus S$ for this problem, and then just add $\ln \det C$ to that bound. We note that the spectral bound for the complementary problem is always identical to the spectral bound for the original problem; but this is not the case for other bounds.

In Section 1, we develop new upper bounds based on Fischer's inequality. In Section 2, we demonstrate how to combine the use of Fischer's inequality with previously developed bounds. We demonstrate this in detail by strengthening the spectral bounds of Ko, Lee and Queyranne and of Lee. Computational experiments suggest that these bounds may be useful in solving problems to optimality in a branch-and-bound framework.

1 Partition Bounds

We base new methods for computing upper bounds on the following inequality of Fischer.

Lemma 1 *(***[Fis08] pp. 36–7; also see [HJ85] pp. 478–9: Theorem 7.8.3 and the exercise that follows it).** *Let B be a square symmetric positive-semidefinite matrix with rows and columns indexed from the cardinality s set S^*. Let $S_1, S_2, \ldots, S_s$ be a partition of S^* (note that we allow empty parts for convenience). Then*

$$\det B \leq \prod_{k=1}^{s} \det B[S_k, S_k] \ . \tag{7}$$

A sequence of refinements of the partition yields a nondecreasing sequence of upper bounds on $\det B$.

Let $\Pi = \{\pi_1, \pi_2, \ldots, \pi_s\}$ be a "partition" of s; that is, a multiset of non-negative integers such that $\sum_{k=1}^{s} \pi_k = s$. We introduce the *partition bound*

$$\psi(\Pi) := \max \sum_{k=1}^{s} \ln \det C[S_k, S_k] \ ; \tag{8}$$

$$\text{subject to } S_k \subset N \ , \ \forall\, k = 1, 2, \ldots, s \ ; \tag{9}$$

$$|S_k| = \pi_k \ , \ \forall\, k = 1, 2, \ldots, s \ ; \tag{10}$$

$$S_k \cap S_{k'} = \emptyset \ , \ \forall\, 1 \leq k < k' \leq s \ ; \tag{11}$$

$$\sum_{k=1}^{s} \sum_{j \in S_k} a_{ij} \leq b_i \ , \ \forall\, i \in M \ . \tag{12}$$

We note that when $\Pi = \{s, 0, 0, \ldots, 0\}$, we have $z = \psi(\Pi)$. Next, we establish that the partition bound is in fact an upper bound on z.

Proposition 1 $\quad z \leq \psi(\Pi)$.

Proof: Suppose that S^* is an optimal solution to (1–2). So $z = \ln \det C[S^*, S^*]$. Choose any partition $\mathcal{S} = \{S_1, S_2, \ldots, S_s\}$ of S^* satisfying (10); the conditions (9,11) are obviously satisfied. Moreover,

$$\sum_{k=1}^{s} \sum_{j \in S_k} a_{ij} = \sum_{j \in S^*} a_{ij} \ ,$$

so (12) is satisfied. Therefore, $\mathcal{S}$ is a feasible solution to the program (8-12). Now, applying Fischer's inequality (7) to $B = C[S^*, S^*]$, and considering the monotonicity of the logarithm, the result follows. $\square$

Next, we focus on situations where we can compute $\psi(\Pi)$ efficiently — either in a practical or theoretical sense. The simplest such situation is based on the finest partition, where we take $\Pi = \{1, 1, \ldots, 1\}$. In this case, we can recast the *diagonal bound* $\psi_1 := \psi(\Pi)$ as the optimal value of the following integer linear program:

$$\begin{aligned} \psi_1 = &\max \sum_{j \in N} (\ln\ C_{jj})\, x_j \ ; \\ &\text{subject to } \sum_{j \in N} x_j = s \ ; \\ &\sum_{j \in N} a_{ij} x_j \leq b_i \ , \ \forall\ i \in M \ ; \\ &x_j \in \{0, 1\}, \ \forall\ j \in N \ . \end{aligned}$$

General methods of integer linear programming can be applied to solve this bounding program (see [NW88]). When $M = \emptyset$, we obtain

$$\psi_1 = \sum_{l=1}^{s} \ln\ C_{[ll]} \ ,$$

where $C_{[ll]}$ is the l^{th} greatest diagonal element of C. So when $M = \emptyset$, we can calculate ψ_1 efficiently in the theoretical sense.

Example 1 *Let n be even, and let $s := n/2$. Let the nonzeros of C consist of the $n/2$ diagonal blocks*

$$\begin{pmatrix} C_{2l-1,2l-1} & C_{2l-1,2l} \\ C_{2l,2l-1} & C_{2l,2l} \end{pmatrix} := \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \ ,$$

for $l = 1, 2, \ldots, n/2$. It is easy to check that with $M = \emptyset$, we have $z = \psi_1 = 0$, while $v = \ln\ 2^{n/2}$. So here the diagonal bound is much better than the spectral bound.

Example 2 *Let $C := \epsilon I_n + 1_{n \times n}$, with $\epsilon > 0$. Clearly $\psi_1 = \ln\ (1+\epsilon)^s$, which tends to 0 as $\epsilon \to 0^+$. It is an exercise to check that $v = \ln\ (n+\epsilon)\epsilon^{s-1}$ and $z = \ln\ (s+\epsilon)\epsilon^{s-1}$ which both tend to $-\infty$ as $\epsilon \to 0^+$. So here the spectral bound is much better than the diagonal bound.*

Another situation that we can exploit is when $\Pi = \{2, 2, \ldots, 2, 0, 0, \ldots, 0\}$ for even s and $\Pi = \{2, 2, \ldots, 2, 1, 0, 0, \ldots, 0\}$ for odd s. Again, we can recast this *matching bound* as the optimal value of an integer linear program. We let N^2 be the set of all subsets T of N satisfying $|T| = 2$. For all $T \subset N^2$, we define binary variables y_T. The form of the program depends on the parity of s. If s is even, then our bound $\psi_2 := \psi(\Pi)$ is

$$\begin{aligned}
\psi_2 = &\max \sum_{T \in N^2} (\ln \det C[T,T])\, y_T \ ; \\
&\text{subject to } \sum_{T \in N^2} y_T = s/2 \ ; \\
&\sum_{\substack{T \in N^2 \ : \\ j \in T}} y_T \le 1 \ , \forall \ j \in N \ ; \\
&\sum_{T \in N^2} \left(\sum_{j \in T} a_{ij} \right) y_T \le b_i \ , \ \forall \ i \in M \ ; \\
&y_T \in \{0,1\}, \ \forall \ T \in N^2 \ .
\end{aligned}$$

General methods of integer linear programming may be applied to solve this bounding program. Constrained maximum-entropy sampling problems are already quite difficult at $n = 100$, so this approach might be quite reasonable. We note that for the important case in which $M = \emptyset$, there is a theoretically-efficient algorithm for solving this program. We simply define a complete graph with vertex set N and edge set N^2. For edge $T \in N^2$, we assign weight $\ln \det C[T,T]$. Then we simply find a maximum-weight matching having cardinality $s/2$, hence the moniker "matching bound".

If s is odd, we define additional binary variables x_j, for $j \in N$. In this case, we calculate our bound by solving the integer linear program

$$\begin{aligned}
\psi_2 = &\max \sum_{j \in N} (\ln C_{jj})\, x_j + \sum_{T \in N^2} (\ln \det C[T,T])\, y_T \ ; \\
&\text{subject to } \sum_{j \in N} x_j = 1 \ ; \\
&\sum_{T \in N^2} y_T = (s-1)/2 \ ; \\
&x_j + \sum_{\substack{T \in N^2 \ : \\ j \in T}} y_T \le 1 \ , \forall \ j \in N \ ;
\end{aligned}$$

$$\sum_{j\in N} a_{ij}x_j + \sum_{T\in N^2}\left(\sum_{j\in T} a_{ij}\right) y_T \le b_i \ , \ \forall\ i \in M \ ;$$
$$x_j \in \{0,1\},\ \forall\ j \in N \ ;$$
$$y_T \in \{0,1\},\ \forall\ T \in N^2 \ .$$

Again, when $M = \emptyset$, we can recast this program using matchings. We start with the same graph as before, but now we incorporate an additional vertex 0. We join vertex 0 to each other vertex $j \in N$. Then, for each $j \in N$, we give weight $K + \ln C_{jj}$ to the edge $\{0, j\}$, for sufficiently large K. Then, we again find a maximum-weight matching having cardinality $s/2$. For large enough K, the optimal weight matching will include exactly one of the edges meeting vertex 0. The optimal weight for this matching will exceed ψ_2 by exactly K.

Example 3 *Continuing with the matrix from Example 2, we calculate* $\psi_2 = \ln\ (2\epsilon + \epsilon^2)^{s/2}$ *which tends to* $-\infty$ *as* $\epsilon \to 0^+$*. So here the matching bound is much better than the diagonal bound. Furthermore, holding* ϵ *constant (at some small positive value), and letting* n *increase, we can make the matching bound do much better than the spectral bound as well.*

Although ψ_2 is harder to calculate than ψ_1, we do have $\psi_2 \le \psi_1$, so it may be worth the extra effort, in the context of branch-and-bound, to calculate ψ_2.

We performed some computational experiments using environmental monitoring data (see [KLQ95]). In the tables, the "bars" indicate bounds applied to the complementary problem. The first problem has $n = 48$ and no side constraints. In Table 1 we display the gaps (i.e., upper bound minus optimal entropy) $\psi_k - z$ and $\bar{\psi}_k - z$. This gives an indication of the behavior of the partition bound ψ_k and the complementary partition bound $\bar{\psi}_k$ as k increases. For small values of k, we can observe how ψ_k (resp. $\bar{\psi}_k$) does better than $\bar{\psi}_k$ (resp. ψ_k) when s is small (resp. large) relative to n. Unfortunately, the improvement in the bound is often rather slight as k increases, while the difficulty in calculating the bound grows very quickly. We note that experience has shown that a gap of up to perhaps 3 indicates that a problem might be solved to optimality by branch-and-bound within a reasonable amount of time.

Table 2 compares these partition bounds to previous bounds for a data set having $n = 124$; "Id,Di,Tr" refer to particular parameter choices for some convex-programming bounds (see [AFLW99]) . We note that for s not too small nor too large, these problems are beyond our current capability to solve, so we have tabulated bounds rather than gaps. For this data set, ψ_2 (resp. $\bar{\psi}_2$) offers no significant improvement over ψ_1 (resp. $\bar{\psi}_1$). Furthermore, for small (resp. large) values of s, ψ_1 (resp. $\bar{\psi}_1$), which is very cheap to compute, is competitive with the other bounds.

$\psi_k - z$	$s = 12$	$s = 24$	$s = 36$
$k = 1$	2.960525	6.346888	11.544246
$k = 2$	2.683129	6.208118	11.335711
$k = 3$	2.298198	5.915404	10.871107

$\bar{\psi}_k - z$	$s = 12$	$s = 24$	$s = 36$
$k = 1$	4.180991	2.372326	0.716902
$k = 2$	4.180985	2.372316	0.716723
$k = 3$	4.180968	2.372173	0.712107

Table 1: Gaps ($n = 48$)

$s =$	10	20	30	40	50	60	70	80	90	100	110	120
$\psi_1 =$	44.21	81.45	116.15	149.08	180.55	210.37	237.55	262.83	285.32	303.53	316.93	326.84
$\psi_2 =$	44.21	81.45	116.15	149.08	180.55	210.37	237.55	262.83	285.32	303.53	316.93	326.84
$\bar{\psi}_1 =$	290.12	293.06	291.39	286.14	277.95	267.12	253.88	237.72	217.31	191.56	160.27	122.15
$\bar{\psi}_2 =$	290.12	293.06	291.39	286.13	277.95	267.12	253.87	237.70	217.30	191.54	160.23	122.12
$Id =$	47.23	92.25	136.30	179.18	220.62	260.21	297.36	331.15	359.97	380.65	385.18	337.13
$\overline{Id} =$	470.81	490.05	478.64	454.27	422.68	386.47	347.06	305.28	261.69	216.66	170.47	123.30
$Di =$	63.89	118.57	167.33	210.72	250.84	287.90	321.88	352.40	378.13	395.44	395.35	337.95
$\overline{Di} =$	339.16	370.54	375.75	370.41	359.03	343.66	324.58	299.97	269.12	231.78	186.36	130.63
$Tr =$	52.08	98.13	140.33	180.55	219.16	255.87	290.14	321.12	347.29	365.47	367.81	319.11
$\overline{Tr} =$	358.40	387.92	390.61	382.37	368.17	349.62	327.09	298.42	263.67	223.61	178.50	127.16
$v =$	50.35	90.57	124.08	151.67	173.51	189.90	198.75	199.76	193.85	180.41	159.64	125.40

Table 2: Bounds ($n = 124$)

2 Spectral Partition Bounds

Let $\mathcal{N} = \{N_1, N_2, ..., N_n\}$ denote any partition of N (since the partition has n parts, we are allowing, for convenience, empty parts). For $k = 1, 2, \ldots, n$, let $\Lambda(N_k)$ be the multiset of $|N_k|$ eigenvalues of $C[N_k, N_k]$. Let $\Lambda(\mathcal{N})$ denote the multiset union of $|N| = n$ elements from the sets $\Lambda(N_k)$. For a multiset $\Lambda(\cdot)$, $\Lambda_l(\cdot)$ denotes the l^{th} greatest element. We define the *spectral partition bound*

$$\phi(\mathcal{N}) := \sum_{l=1}^{s} \ln \ \Lambda_l(\mathcal{N}) \ . \tag{13}$$

Proposition 2 $\quad z \leq \phi(\mathcal{N})$.

Proof: Let S be any subset of N having cardinality s.

$$\begin{aligned}
&\ln \det C[S,S] \\
&\qquad \le \sum_{k=1}^{n} \ln \det C[S\cap N_k, S\cap N_k] && (14)\\
&\qquad = \sum_{k=1}^{n} \sum_{\lambda\in\Lambda(S\cap N_k)} \ln\ \lambda && (15)\\
&\qquad \le \sum_{k=1}^{n} \sum_{l=1}^{|S\cap N_k|} \ln\ \Lambda_l(N_k) && (16)\\
&\qquad \le \sum_{l=1}^{s} \ln\ \Lambda_l(\mathcal{N})\ . && (17)
\end{aligned}$$

We note that the inequality (14) holds by Fischer's inequality (7), the equation (15) holds since the product of all eigenvalues of a matrix is its determinant, the inequality (16) holds by the eigenvalue interlacing inequalities (see, for example, [HJ85] pp. 185–6: Theorem 4.3.8; also see [KLQ95]), and the inequality(17) holds by allowing S to range over subsets of N having cardinality s. □

We note that

- $\phi(\{N,\emptyset,\emptyset,\ldots,\emptyset\}) = v$ (thus subsuming the bound of [KLQ95]);
- $\phi(\{\{1\},\{2\},\ldots,\{n\}\}) = \psi_1$.

Next, we demonstrate a revealing situation in which neither of these two partitions is best possible.

Example 4 *Let $s := n/2$, and let $S := \{1,2,\ldots,n/2\}$; so $N\setminus S = \{n/2+1, n/2+2,\ldots,n\}$. Let $C[S,S] := nI_s + 1_{s\times s}$, let $C[N-S,N-S] := (3n/4)I_s + 1_{s\times s}$, and let $C[S,N\setminus S] := C[N\setminus S,S] := 0_{s\times s}$. It is an exercise to check that $\Lambda(S) = \{3n/2,n,n,\ldots,n\}$, $\Lambda(N-S) = \{5n/4,3n/4,3n/4,...,3n/4\}$, and obviously $\Lambda(N) = \Lambda(S)\cup\Lambda(N-S)$. Therefore $v = \phi(\{N,\emptyset,\emptyset,\ldots,\ \emptyset\}) = \phi(\{S,N\setminus S,\emptyset,\emptyset,\ldots,\emptyset\}) = \ln\ (3n/2)(5n/4)n^{n/2-2}$; note that we picked up the $5n/4$ from $\Lambda(N-S)$. We also have $\phi(\{\{1\},\{2\},\ldots,\{n\}\}) = \psi_1 = \ln\ (n+1)^{n/2}$; note that we did not pick up any of $\Lambda(N-S)$, but the bound is deteriorated by chopping up S and using Fischer's inequality. Finally, we observe that $\phi(\{S,\{n/2+1\},\{n/2+2\},...,\{n\}\}) = \ln \det C[S,S]$, since all of the diagonal entries of $C[N-S,N-S]$ are $3n/4+1$, which is less than all eigenvalues of*

$C[S,S]$. So this last partition establishes that S has maximum entropy, while the other partitions mentioned above do not.

Example 4 suggests the following sufficient optimality criterion.

Proposition 3 *Let S be a feasible subset of N. If*

$$\Lambda_s(S) \geq \max\{C_{jj} \ : \ j \in N \setminus S\},$$

then $z = \ln \det C[S,S]$.

Proof: We simply observe that under the hypothesis we have $\phi(\{S,\{s+1\},\{s+2\},...,\{n\}\}) = \ln \det C[S,S]$. $\Box$

Next, we turn to the issue of finding the best partition $\mathcal{N}$. That is,

$$\text{MIN}\Phi: \quad \min\{\phi(\mathcal{N}) \ : \ \mathcal{N} \text{ is a partition of } N\}.$$

As we do not know a good algorithm for MINΦ, we suggest a heuristic which is outlined in Figure 1. We experimented with the heuristic of Figure 1 on an example from [AFLW99] having $n = 63, s = 31$. For the local-search of Step 2, we repeatedly evaluated the spectral bound for $\mathcal{O}(n^2)$ "nearby partitions" and selected the move that achieved the best improvement. Specifically, we considered the moves described in Figure 2. At a first pass, we just used the single-element moves 2a until no further improvement was possible. Then we proceeded further using 2a–d. We worked with the complementary problem as well as the original. The results are displayed in Table 3. Since we have the optimal value from [AFLW99], we have subtracted that value from the bounds to obtain the gaps. The first row consists of the gaps after Step 1. The second row consists of the gaps after all 2a moves were completed. The third row consists of the gaps after all 2a–d moves were completed. Also, for Step 1a, since we wish to see how the bounding idea performs when we have an extremely good heuristic for finding an S with high entropy, we actually used an optimal S. The results are exceptionally good. The best bound obtained (2.521062) is much better than the ordinary spectral bound ($v = 5.707025$) and also significantly better than the best bound obtained ($\bar{\psi}_1 = 3.252440$) without applying the local search procedure indicated in Step 2. In addition, our results suggests that our particular local search moves (see Figure 2) are rather robust, since the final bound obtained from local search procedure does not depend very much on the initial partition selected in Step 1. Finally, we mention that for the best bound obtained (i.e., $\bar{\phi}(\mathcal{N})$ starting with 1b), the final partition had block sizes of: 3, 4, 4, 5, 5, 7, 9, 11, 15.

1a. Use heuristic methods to find an $S \subset N$, $|S| = s$ with a high value of $\ln \det C[S,S]$ (e.g., greedy and interchange heuristics for the problem are discussed in [KLQ95]); then take the initial partition $\mathcal{N} = \{S, \{j_1\}, \{j_2\}, ..., \{j_{n-s}\}\}$, where $N \setminus S = \{j_1, j_2, \ldots, j_{n-s}\}$ (and use Proposition 3 to attempt to establish the optimality of S).

1b. Alternatively, we could try the initial partition $\mathcal{N} = \{N, \emptyset, \emptyset, \ldots, \emptyset\}$ which would guarantee that we do no worse than the spectral bound v.

1c. Alternatively, we could try the initial partition $\mathcal{N} = \{\{1\}, \{2\}, \ldots, \{n\}\}$ which would guarantee that we do no worse than the diagonal bound ψ_1.

2. Use a local search method on the space of partitions, to decrease $\phi(\mathcal{N})$.

Figure 1: MINΦ Heuristic

2a. *(single-element move)* $j \in N_k$, $l \neq k$: $N_k \leftarrow N_k - j$, $N_l \leftarrow N_l + j$.

2b. *(two-element switch)* $j \in N_k$, $i \in N_l$, $l \neq k$: $N_k \leftarrow N_k - j + i$, $N_l \leftarrow N_l - i + j$.

2c. *(one new two-block or two new one-blocks)* $j \in N_k$, $i \in N_l$, $i \neq j$, $N_h = \emptyset$, $N_g = \emptyset$: $N_k \leftarrow N_k - j$, $N_l \leftarrow N_l - i$, $N_h \leftarrow N_h + i$, $N_g \leftarrow N_g + j$.

2d. *(merge two blocks)* $k \neq l$: $N_k \leftarrow N_k \cup N_l$, $N_l \leftarrow \emptyset$.

Figure 2: Local Search Moves

The spectral partition bound does not take advantage of the side-constraints. We can improve the spectral partition bound by adapting the Lagrangian

	original			complementary		
	1a	1b	1c	1a	1b	1c
1	5.512078	$v = 5.707025$	$\psi_1 = 7.924975$	3.352356	$v = 5.707025$	$\bar{\psi}_1 = 3.252440$
2a	4.576737	4.579252	5.060646	2.655492	2.607664	2.629423
2a–d	4.576737	4.579252	4.577380	2.630182	2.521062	2.627294

Table 3: Gaps using local search $(n = 63, s = 31)$

methodology employed in [Lee98]. As before, we consider a partition $\mathcal{N} = \{N_1, N_2, ..., N_n\}$ of N and define the diagonal matrix D_w by (6). Now we let $\Lambda(N_k, A, w)$ denote the multiset of $|N_k|$ eigenvalues of $(D_w C D_w)[N_k, N_k]$ and $\Lambda(\mathcal{N}, A, w)$ denote the multiset union of $|N| = n$ elements from the sets $\Lambda(N_k, A, w)$. We introduce the *Lagrangian spectral partition bound*

$$\phi(\mathcal{N}, A, b) := \min_{w \in \mathbb{R}_+^M} \phi(\mathcal{N}, A, b, w) \ ,$$

where

$$\phi(\mathcal{N}, A, b, w) := \sum_{l=1}^{s} \ln \ \Lambda_l(\mathcal{N}, A, w) \ + \sum_{i \in M} w_i b_i \ .$$

Following the ideas in [Lee98], we can demonstrate the following results.

Proposition 4 $z \leq \phi(\mathcal{N}, A, b)$.

Proposition 5 *The function $\phi(\mathcal{N}, A, b, w)$ is convex in w.*

Also, again adapting an idea in [Lee98], we can derive an expression for subgradients of $\phi(\mathcal{N}, A, b, w)$.

Finally, we mention that although we chose to concentrate on spectral bounds, other bounds (for example, those in [AFLW96, AFLW99, Lee00]) can also be strengthened using the partitioning idea and Fischer's inequality. We will discuss experiments with some of these possibilities in a forthcoming paper in which we report on results of incorporating our new bounds in a branch-and-bound code for the exact solution of constrained maximum-entropy sampling problems.

Acknowledgments

Jon Lee's research was partially supported by the Department of Mathematical Sciences, T.J. Watson Research Center, IBM (summer 1999), a CORE

Fellowship (1999–2000), and Sabbatical and Scholarly Leaves from the University of Kentucky (1999–2000). Part of this research was carried out using the facilities of the High-Performance Computing Laboratory of the College of Arts and Sciences at the University of Kentucky, supported in part by NSF Grant DMS-9508543.

References

[AFLW96] Kurt M. Anstreicher, Marcia Fampa, Jon Lee, and Joy Williams. Continuous relaxations for constrained maximum-entropy sampling. In *Integer Programming and Combinatorial Optimization (Vancouver, BC, 1996)*, volume 1084 of *Lecture Notes in Computer Science*, pages 234–248. Springer, Berlin, 1996.

[AFLW99] Kurt M. Anstreicher, Marcia Fampa, Jon Lee, and Joy Williams. Using continuous nonlinear relaxations to solve constrained maximum-entropy sampling problems. *Mathematical Programming, Series A*, 85(2):221–240, 1999.

[Fis08] Ernst S. Fischer. Über den Hadamardschen determinantensatz. *Archiv für Mathematik und Physik*, 13(3):32–40, 1908.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[KLQ95] Chun-Wa Ko, Jon Lee, and Maurice Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.

[Lee98] Jon Lee. Constrained maximum-entropy sampling. *Operations Research*, 46(5):655–664, 1998.

[Lee00] Jon Lee. Semidefinite-programming in experimental design. In Henry Wolkowicz, Romesh Saigal and Lieven Vandenberghe, editors, *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research and Management Science*. Kluwer, 2000.

[NW88] George L. Nemhauser and Laurence A. Wolsey. *Integer and combinatorial optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1988. A Wiley-Interscience Publication.