Maximum Entropy Sampling

Jon Lee
Mathematical Sciences Department
IBM T.J. Watson Research Center
Yorktown Heights, New York USA

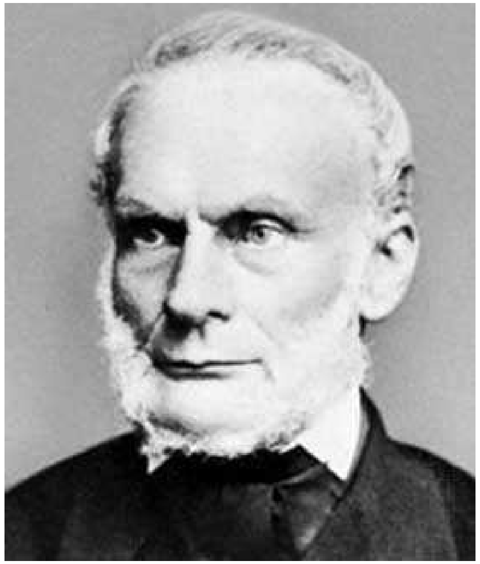
Information

"Chance and chance alone has a message for us. Everything that occurs out of necessity, everything expected, repeated day in and day out, is mute. Only chance can speak to us. We read its message much as gypsies read the images made by coffee grounds at the bottom of a cup."
- Milan Kundera (The Unbearable Lightness of Being)

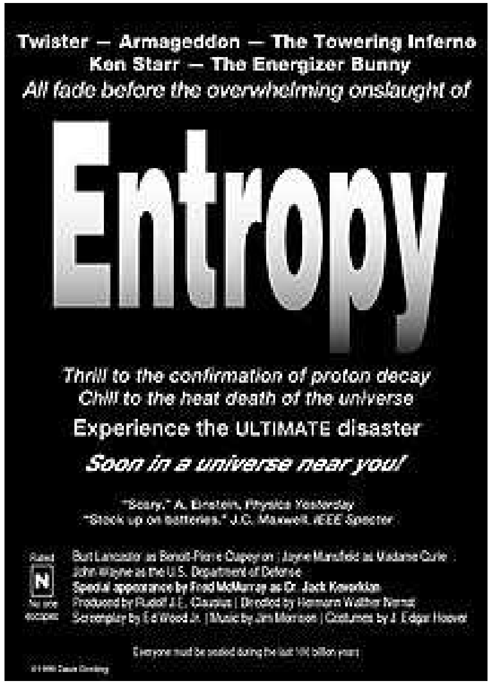
Entropy

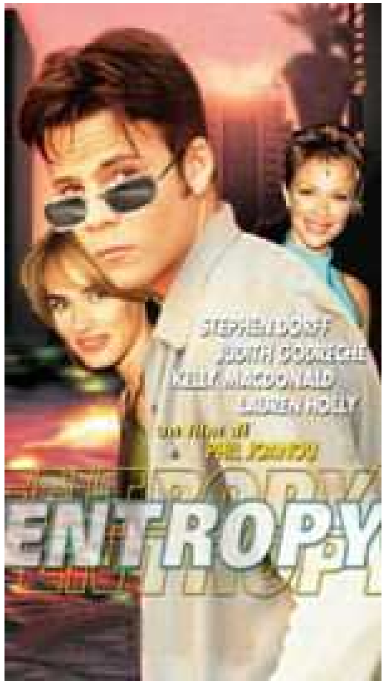
"I propose to name the magnitude $\mathcal{S}$ the entropy of the body from the Greek word $\eta\tau\rho o\pi\grave{\eta}$, a transformation. I have intentionally formed the word entropy so as to be as similar as possible to the word energy, since both these quantities, which are to be known by these names, as so nearly related to each other in their physical significance that a certain similarity in their names seemed to me advantageous ..."

— R. Clausius (1865)

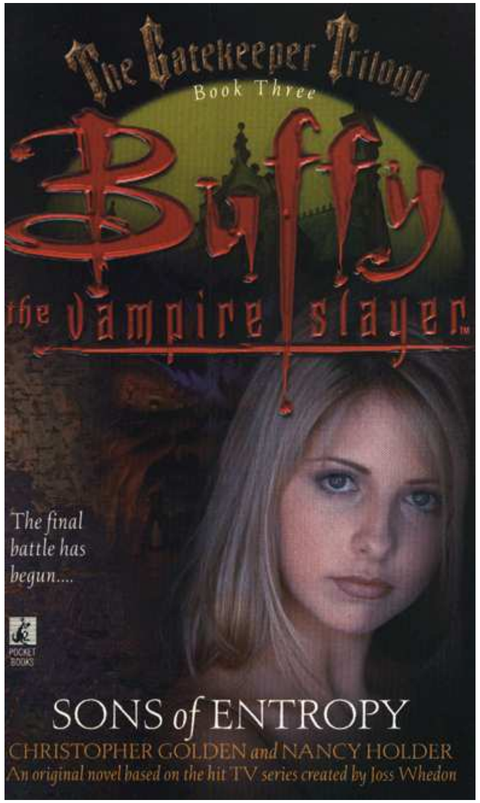


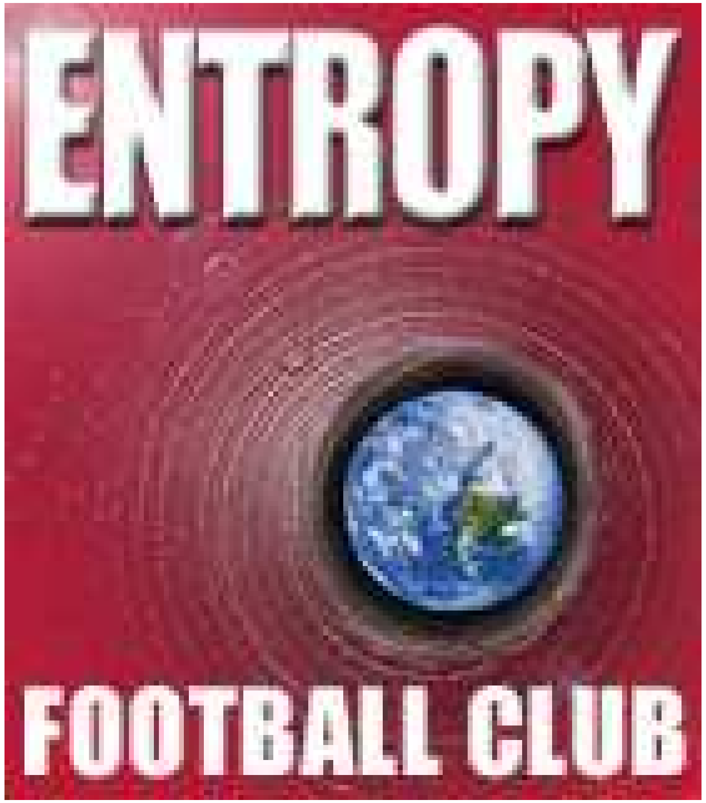
Entropy more recently...

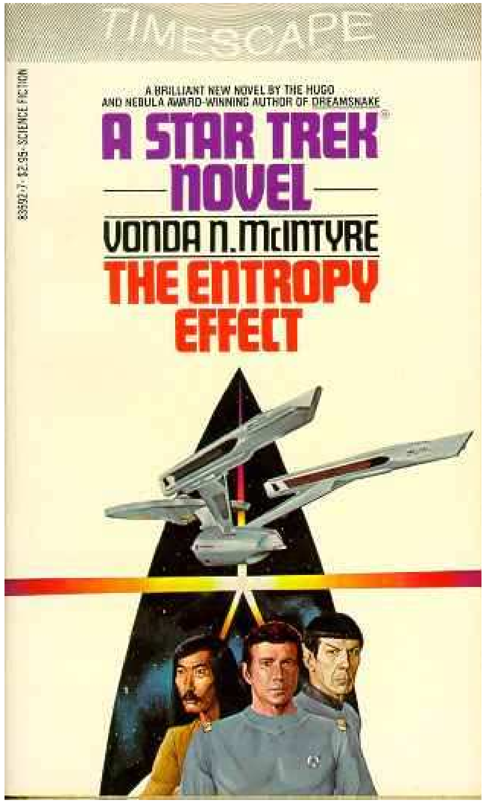




and more...


The Gatekeeper Trilogy
Book Three
Buffy
the vampire slayer
The final battle has begun....
SONS of ENTROPY
CHRISTOPHER GOLDEN and NANCY HOLDER
An original novel based on the hit TV series created by Joss Whedon


ENTROPY
FOOTBALL CLUB


TIMESCAPE
A BRILLIANT NEW NOVEL BY THE HUGO AND NEBULA AWARD-WINNING AUTHOR OF DREAMSNAKE
A STAR TREK NOVEL
VONDA N. McINTYRE
THE ENTROPY EFFECT

Maximum-Entropy Sampling

$N = \{1, 2, \ldots, n\}$

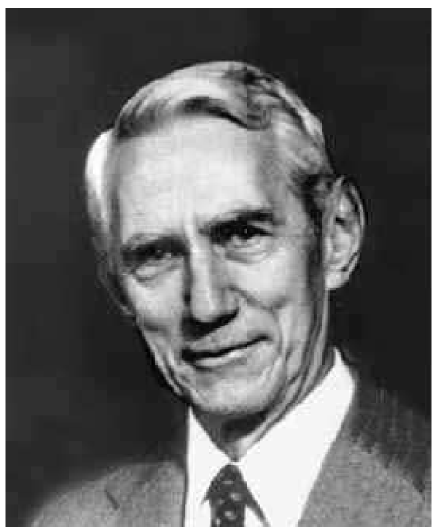
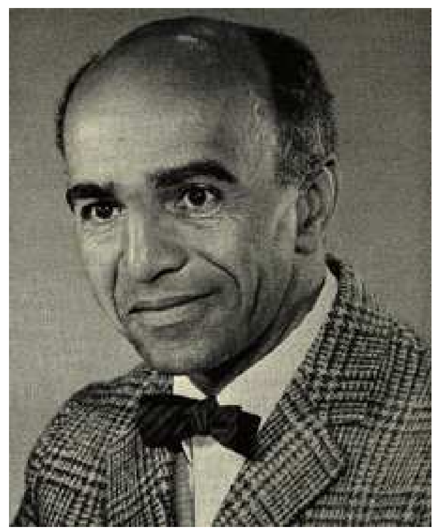
Random $Y_N = \{Y_j \ : \ j \in N\}$ with continuous denstity g_N

Goal: Choose $S \subset N$, with $|S| = s$, to maximize the "information" obtained about Y_N .

Entropy: $h(S) := -E[\ln g_S(Y_S)]$.

- R. Clausius (1865) — "entropy" (also Carnot and Kelvin in their versions of the 2nd law of thermodynamics).
- L. Boltzmann (1877) — statistical mechanics.
- C. Shannon (1948) — information theory.
- D. Blackwell (1951) — statistics.



Motivation: Environmental Monitoring

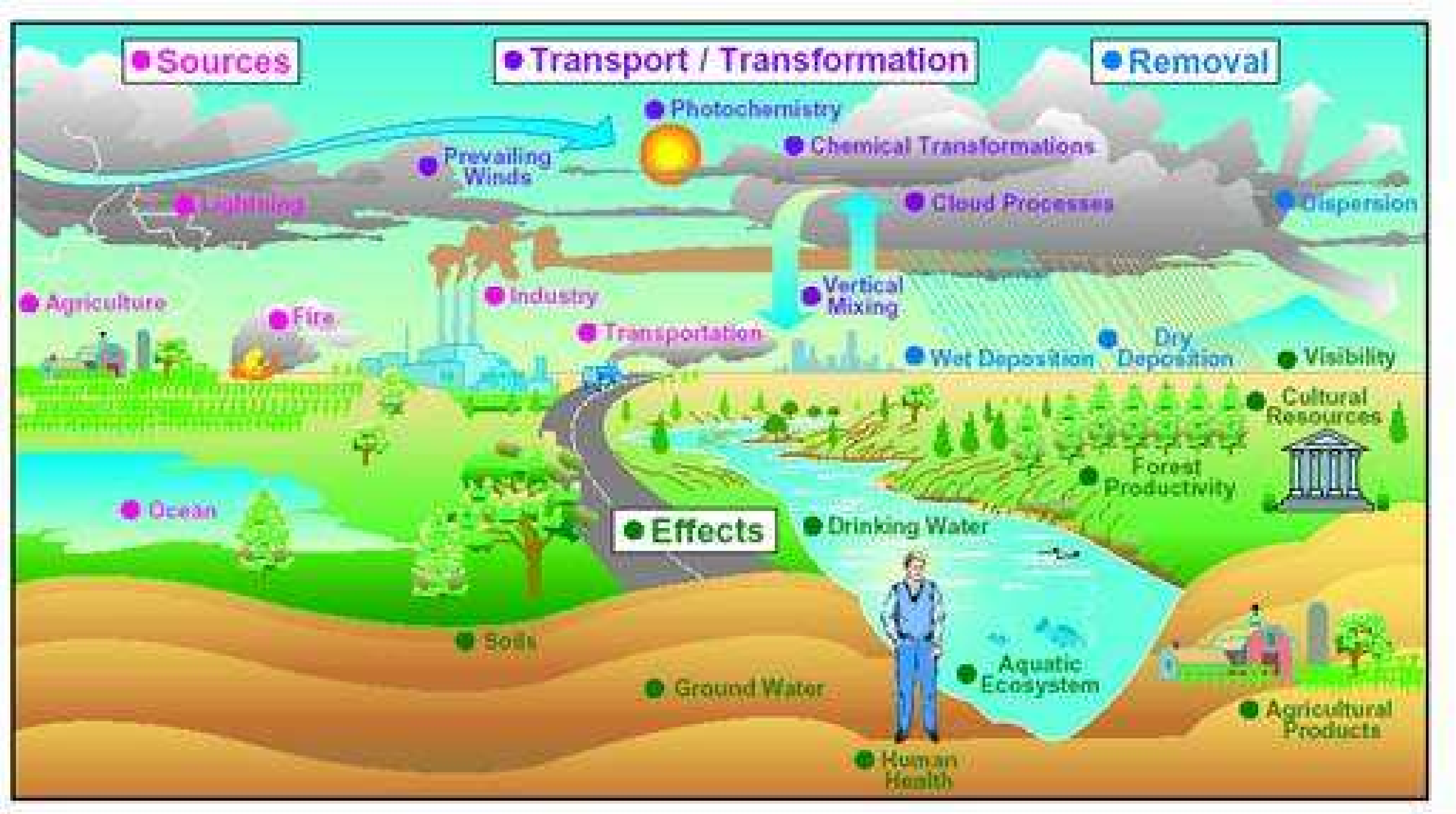
- Sites of emission $\Longrightarrow$ Causes
- Sites of deposition $\Longrightarrow$ *Effects**

$*$ *U.S. Clean Air Act* of 1990 and its revisions mandate *effects monitoring*

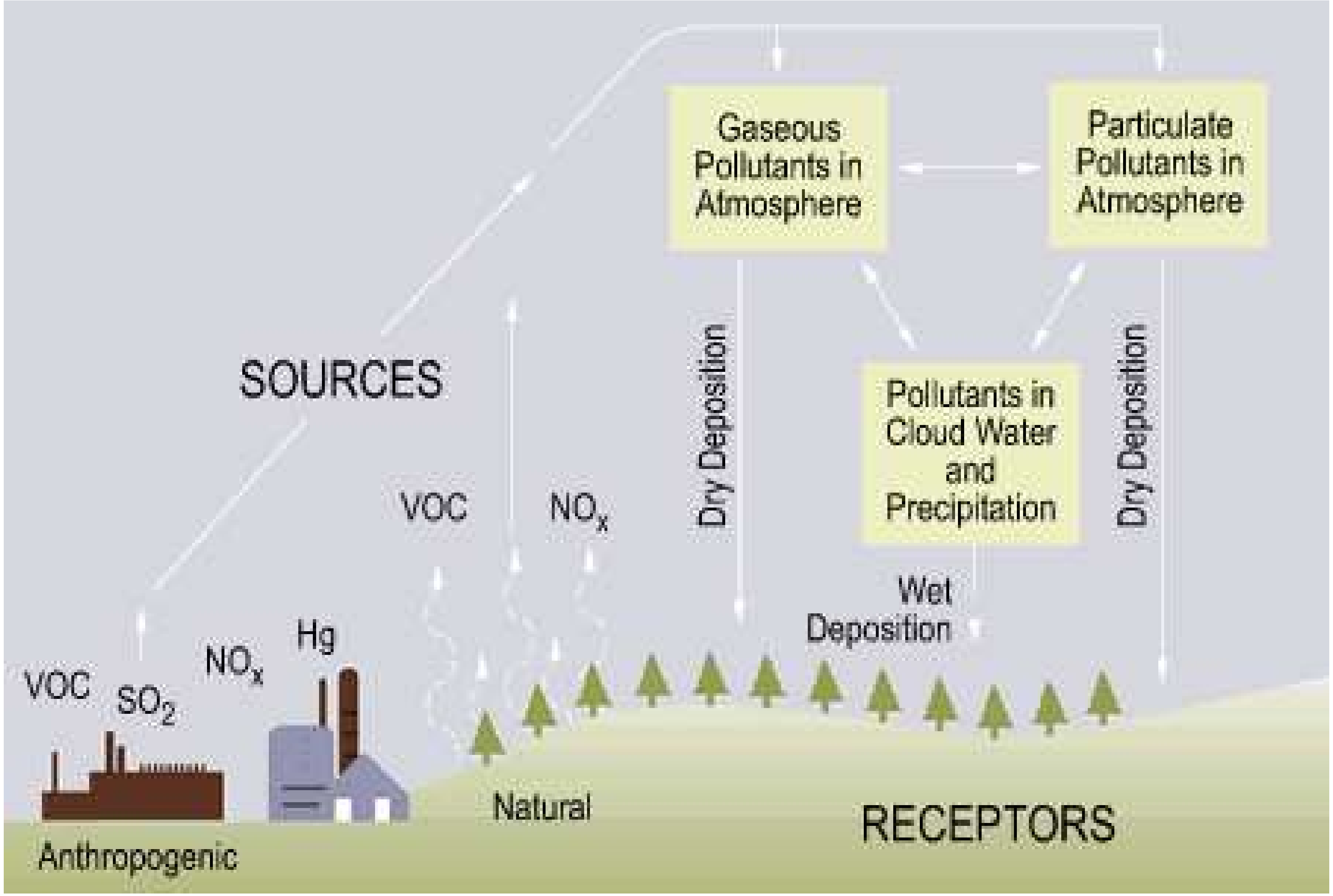
National **A**cidic **D**eposition **P**rogram/
National **T**rends **N**etwork
`nadp.sws.uiuc.edu`

1978 - 22 stations. 2004 - $>$ 220 stations.

Precipitation collected weekly; analyzed for: Hydrogen (acidity as pH — ‘acid rain’), Sulfate, Nitrate, Ammonia, Chloride, Calcium, Magnesium, Potassium, Sodium


Sources
Transport / Transformation
Removal
Photochemistry
Chemical Transformations
Prevailing Winds
Lightning
Cloud Processes
Dispersion
Agriculture
Fire
Industry
Transportation
Vertical Mixing
Wet Deposition
Dry Deposition
Visibility
Cultural Resources
Forest Productivity
Ocean
Effects
Drinking Water
Soils
Ground Water
Aquatic Ecosystem
Agricultural Products
Human Health

Wet vs. Dry


Gaseous Pollutants in Atmosphere
Particulate Pollutants in Atmosphere
SOURCES
Dry Deposition
Pollutants in Cloud Water and Precipitation
Dry Deposition
VOC
NO$_x$
Wet Deposition
Hg
NO$_x$
VOC
SO$_2$
Natural
RECEPTORS
Anthropogenic

NADP Networks

- NADP/NTN: National Trends Network
- NADP/AIRMoN: Atmospheric Integrated Research Monitoring Network
  - Designed to provide data with greater temporal resolution
  - Daily and event-based samples
  - 9 sites in the Eastern U.S. (including Ithaca N.Y.!)
- NADP/MDN: Mercury Deposition Network
  - Weekly samples
  - $\sim$ 70 sites

Typical Monitoring Site

ADS (N-CON Systems) $4.6K. . .

. . . and 4 workers

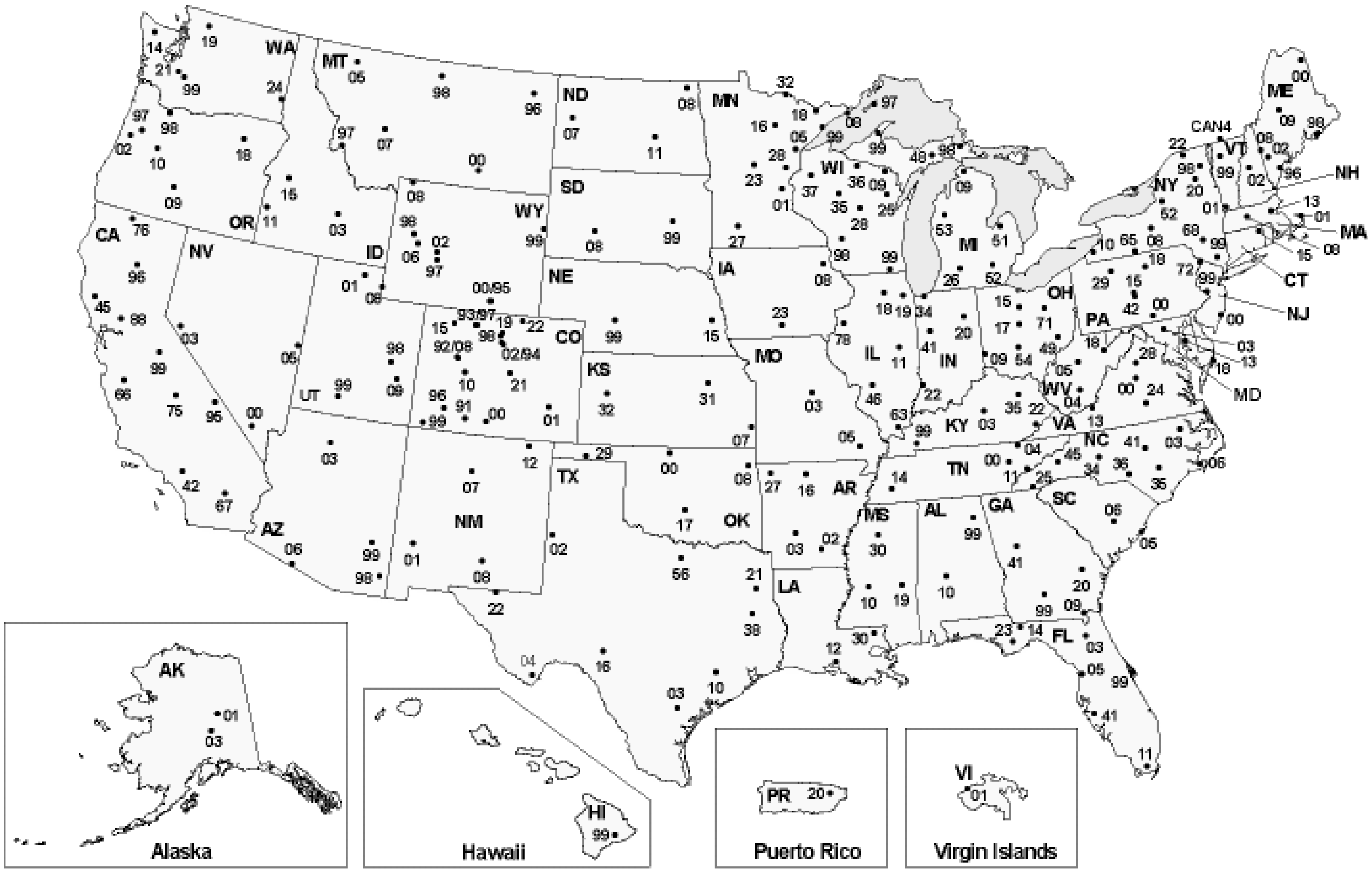
MDN (N-CON Systems)

TPC 3000 (Yankee Environ. Sys.)

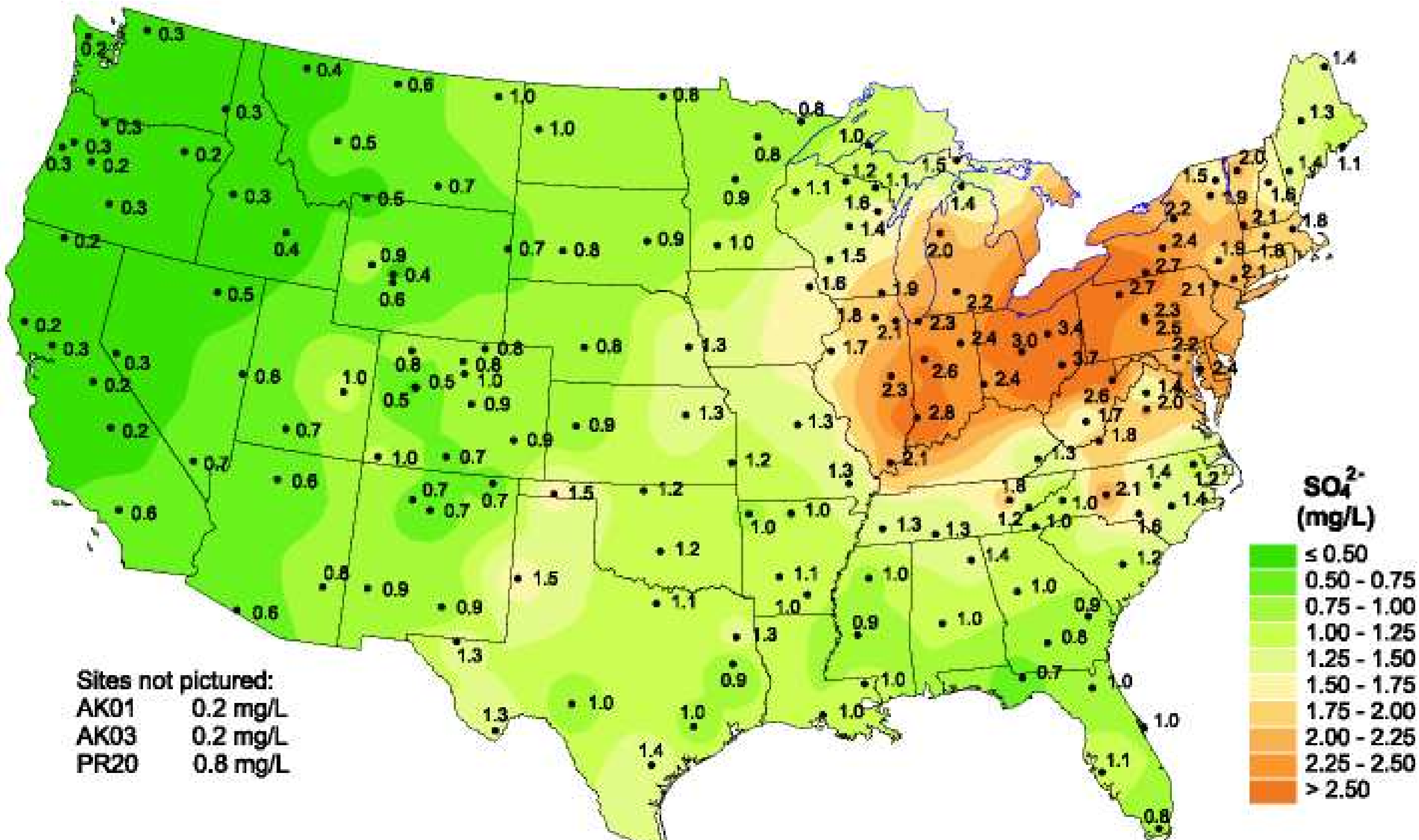
US Federal $

- YES has US federal funding of $300K to develop a new prototype over 2 years
- $3.5M federal funding for NTN ('99)
- $\sim$ $150M total US federal funding for environmental monitoring ('99)
  - much other monitoring focused on CO, NO_2, SO_2 and small particulate matter

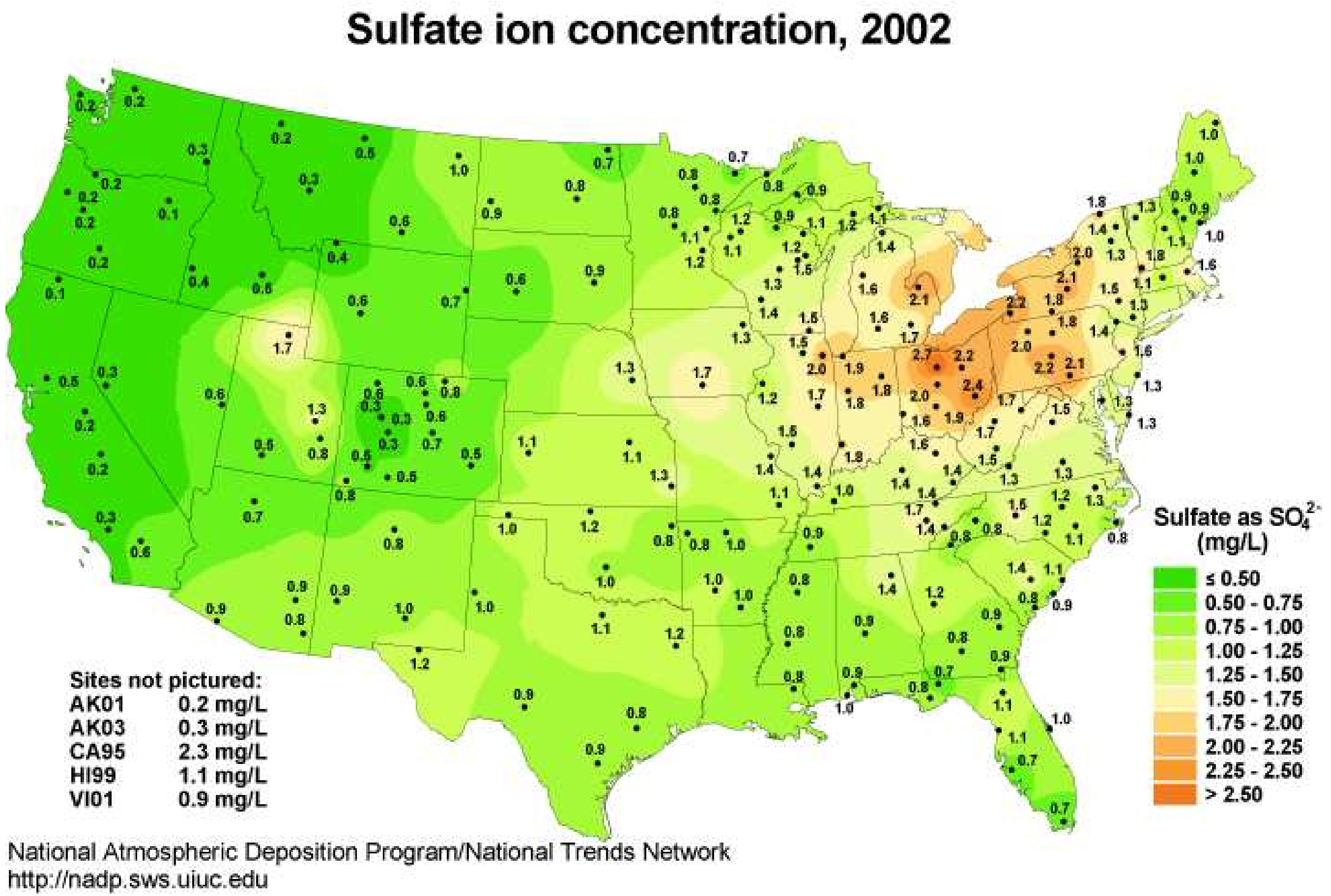
National Atmospheric Deposition Program
National Trends Network



Sulfate ion concentration, 1994



National Atmospheric Deposition Program/National Trends Network
http://nadp.sws.uiuc.edu


Sulfate ion concentration, 2002
Sulfate as SO_4^{2-} (mg/L)
≤ 0.50
0.50 - 0.75
0.75 - 1.00
1.00 - 1.25
1.25 - 1.50
1.50 - 1.75
1.75 - 2.00
2.00 - 2.25
2.25 - 2.50
> 2.50
Sites not pictured:
AK01 0.2 mg/L
AK03 0.3 mg/L
CA95 2.3 mg/L
HI99 1.1 mg/L
VI01 0.9 mg/L
National Atmospheric Deposition Program/National Trends Network
http://nadp.sws.uiuc.edu

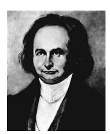
Data for Computational Experiments

- Environmental monitoring data: Courtesy of  Jim Zidek and co-workers at UBC — Monthly (logged) sulfate concentrations collected (weekly, over a 4-year period) at stations centered on the Ohio Valley

Nice Properties of Entropy

- The Gaussian  distribution maximizes the entropy for a given covariance matrix C
- Gaussian case: $h(S) = k_s + k \ln\det C[S,S]$
- Conditional Additivity:

$$h(N) = \overbrace{h(S)}^{\max} \overset{\Leftrightarrow}{+} \overbrace{h(N \setminus S|S)}^{\min}$$

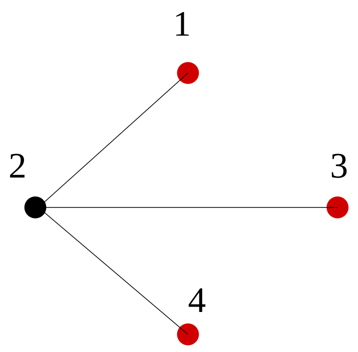
- Change coordinate systems: Entropy difference is log(Jacobian  of transformation)
- Submodularity: $h(S \cup T) + h(S \cap T) \le h(S) + h(T)$
- Complementation: $\ln\det C[S,S] = \ln\det C + \ln\det C^{-1}[N-S, N-S]$

Not-So-Nice Property

Proposition [KLQ '95]. The maximum-entropy sampling problem is NP-Hard (even for the Gaussian diagonally-dominant case)

Proof:

- **INDEPENDENT SET**: Does a simple undirected graph G on n vertices have an independent set of vertices of cardinality s ?
- Let $C := A(G) + 3nI$



$$\begin{pmatrix} 12 & 1 & 0 & 0 \\ 1 & 12 & 1 & 1 \\ 0 & 1 & 12 & 0 \\ 0 & 1 & 0 & 12 \end{pmatrix}$$

(KLQ '95) Branch . . .

- **Fixing j out of S:**
  $\Rightarrow$ Strike out row and column j : $C[N, N] \rightarrow$

$$C[N - j, N - j]$$

- **Fixing j in S:**

$\Rightarrow$  Schur complement of $C[j, j]$: $C[N, N] \rightarrow$

$$C[N - j, N - j] - C[N - j, j]C^{-1}[j, j]C[j, N - j]$$

(and solution/bounds are shifted by $\ln C[j, j]$).

. . . and Bound

- Lower bounds: Greedy, local-search rounding heuristics based on ....
- Upper bounds:
  - Spectral based bounds
    - KLQ ’95 (original B&B and spectral bound)
    - Lee ’98 (extension to side constraints)
    - Hoffman, Lee & Williams ’01 (spectral partition bounds)
    - LW ’03 (tightening HLW via ILP and matching)
    - Anstreicher, Lee ’04 (generalization of HLW)
  - NLP relaxation
    - Anstreicher, Fampa, Lee & Williams ’96 (continuous NLP relaxation and parallel B&B)

Complementary Bounds (AFLW '96)

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

- So a maximum entropy s-subset of N with respect to C is the complement of a maximum entropy $(n-s)$-subset of N with respect to C^{-1}
- So a bound on the complementary problem plus the entropy of the entire system is a bound on the original problem
- These complementary bounds can be quite effective

NLP Bound (AFLW '96)

$$\max f(x) := \ln\det\left(\mathrm{Diag}(x_j^{p_j})\; C\; \mathrm{Diag}(x_j^{p_j}) + \mathrm{Diag}(d_j^{x_j} - d_j x_j^{p_j})\right)$$

subject to $\sum_{j\in N} a_{ij}x_j \le b_i, \forall i;$ $\Longleftarrow$ CONSTRAINTS

$$\sum_{j\in N} x_j = s;$$

$$0 \le x_j \le 1, \forall j,$$

where the constants $d_j > 0$ and $p_j \ge 1$ satisfy $d_j \le \exp(p_j - \sqrt{p_j})$, and $\mathrm{Diag}(d_j) - C[N,N] \succeq 0$.

NLP Bound, cont'd

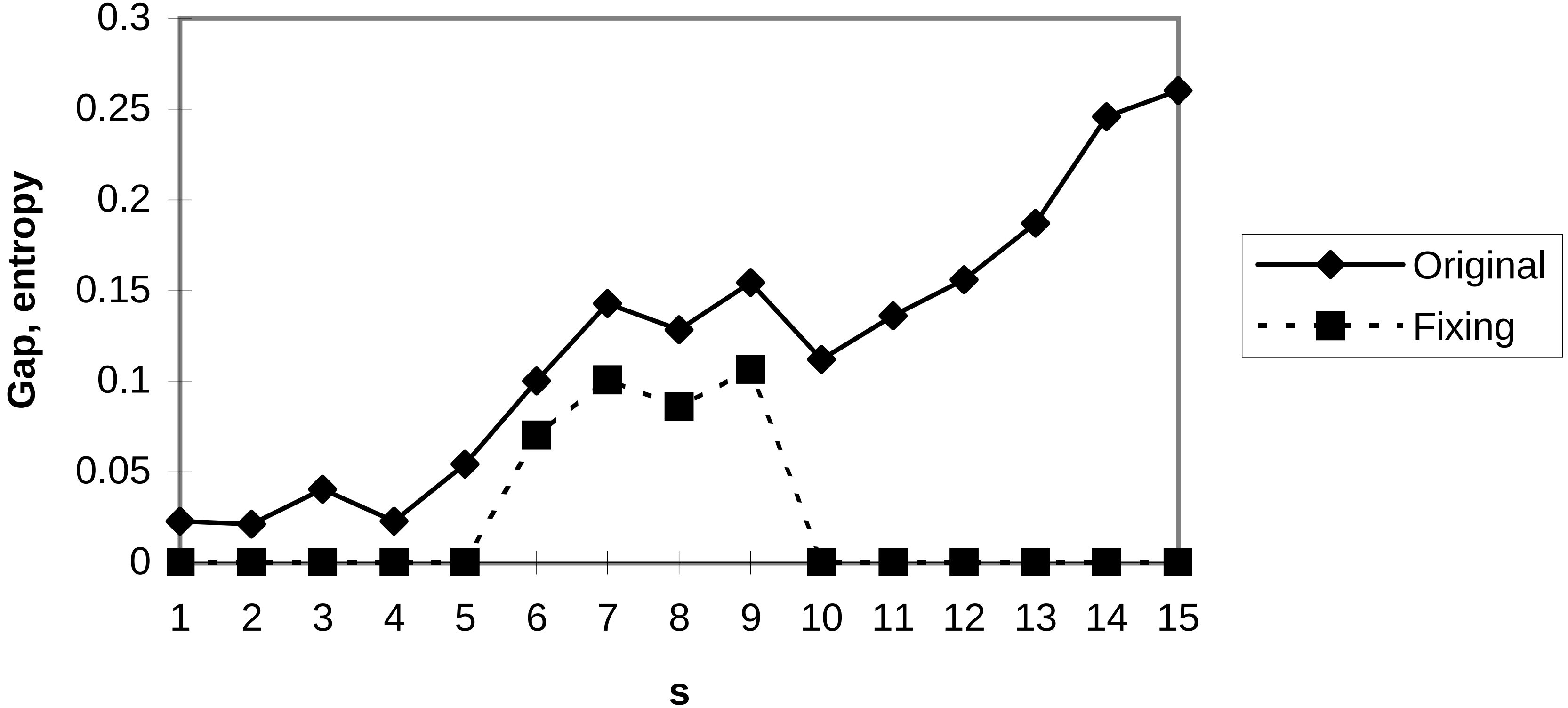
For $(\overbrace{1,1,\ldots,1}^{S},\overbrace{0,0,\ldots,0}^{N\setminus S})$

- $\mathrm{Diag}(d_j^{x_j} - d_j x_j^{p_j}) = \mathrm{Diag}(\overbrace{0,0,\ldots,0}^{S},\overbrace{1,1,\ldots,1}^{N\setminus S})$.
- $\mathrm{Diag}(x_j^{p_j})\, C \,\mathrm{Diag}(x_j^{p_j}) = \left(\begin{array}{c|c} C[S,S] & 0 \\ \hline 0 & 0 \end{array}\right)$

NLP Bound: Properties

- Assume $D \succeq C$, $p_j \geq 1$, $0 < d_j \leq \exp(p_j - \sqrt{p_j})$. Then f is concave for $0 < x \leq e$
- Assume that p and d satisfy the above, and $p' \geq p$. Let f' be defined as above, but using p' for p. Then $f'(x) \geq f(x) \; \forall \; 0 < x \leq e$
- Scaling C by γ adds $s \ln(\gamma)$ to the obj. Let
  $$f_\gamma(x) := \ln\det\left(\gamma X^{p/2}(C - D)X^{p/2} + (\gamma D)^x\right) - s\ln(\gamma)$$
  - Assume $I \succeq D \succeq C$, $p = e$. Then $f_\gamma(x) \geq f(x) \; \forall$ $0 \leq x \leq e$, $e^T x = s$ and $0 < \gamma \leq 1$
  - Assume $D \succeq C$, $D \succeq I$. Then $f_\gamma(x) \geq f(x) \; \forall$ $0 < x \leq e$, $e^T x = s$ and $\gamma \geq 1$, where p is chosen as above

NLP Bound: Fix and Re-Bound



NLP Bound: Parallel Experiments

Number of constraints	Number of processors 1 (seconds)	2	4	8 (speed-up factor)
0	62615	1.99	4.04	6.97
2	34619	2.03	4.10	8.04
5	5551	2.02	4.00	7.56
10	14815	1.95	4.00	7.15
15	12153	1.97	3.97	7.81

($n = 63,\ s = 31$; Circa ’97, Convex Exemplar, Lexington)

Diagonal Bound (HLW '01)

$$z \leq \sum_{l=1}^{s} \ln \operatorname{diag}_{[l]}(C)$$

- Entropy of any set is bounded by the sum of the entropies of n independent random variables having the same variances
- *Very* cheap to compute

Spectral Bound (KLQ '95)

$$z \leq \sum_{l=1}^{s} \ln \lambda_l(C)$$

- Determinant $=$ product of eigenvalues.
- Eigenvalue interlacing.

$$\begin{array}{ccc} \lambda_1 & \geq & \lambda_1' \\ \lambda_2 & \geq & \lambda_2' \\ \lambda_3 & \geq & \lambda_3' \\ & \vdots & \\ \lambda_s & \geq & \lambda_s' \end{array}$$

Problem #; $n/n{-}f/s{-}f$	Initial absolute entropy gap	UB calls	Max # active subproblems	Wall Seconds
$1; 52/16/8$	0.18149914	31	1	2
$2; 63/27/13$	0.56583546	323	15	7

(Circa ’92, MacFORTRAN, Mac IIci, Louvain-la-Neuve)

Lagrangian Spectral Bound (L ’98)

$$\min_{\pi \in \mathbb{R}^m_+} v(\pi)$$

where

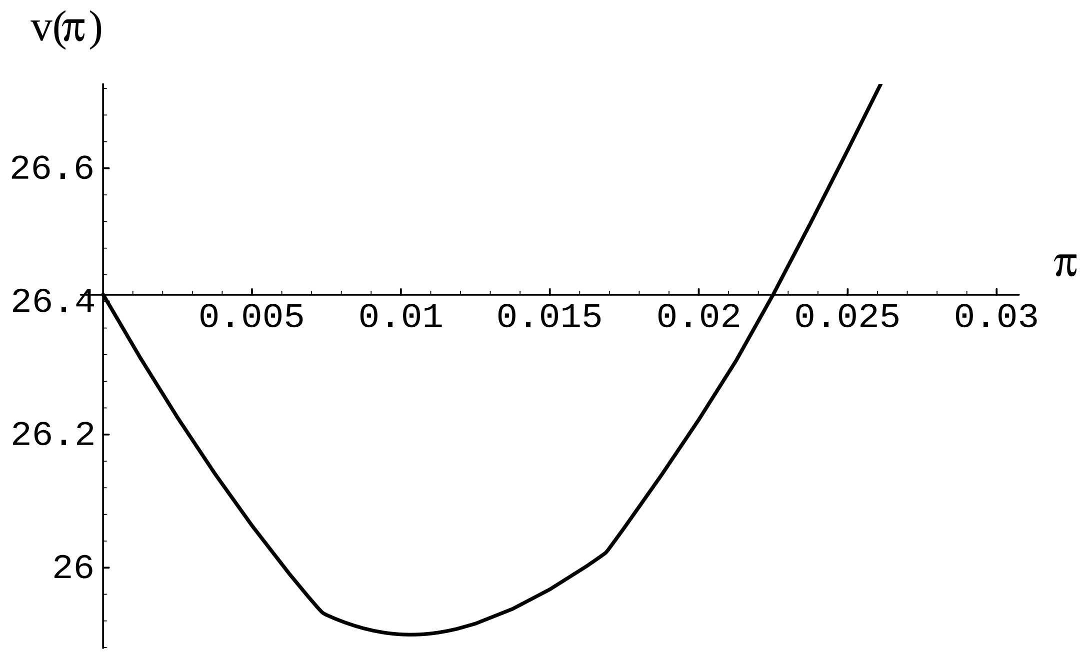
$$v(\pi) := \left\{ \sum_{l=1}^{s} \ln \lambda_l \left(D^\pi \; C \; D^\pi \right) + \sum_{i \in M} \pi_i b_i \right\},$$

and D^π is the diagonal matrix having

$$D^\pi_{jj} := \exp \left\{ -\frac{1}{2} \sum_{i \in M} \pi_i a_{ij} \right\}$$

Optimizing the Lagrangian Spectral Bound

- v_π is convex (in π)
- v_π is analytic when $\lambda_s\left(D^\pi\ C\ D^\pi\right) > \lambda_{s+1}\left(D^\pi\ C\ D^\pi\right)$



Optimizing the Bound, cont'd

- Let x^l be the eigenvector (of unit Euclidean norm) associated with λ_l.
- Define the *continuous solution* $\tilde{x} \in \mathbb{R}^N$ by $\tilde{x}_j := \sum_{l=1}^{s} \left(x_j^l\right)^2$, for $j \in N$.
- Define $\gamma \in \mathbb{R}^M$ by $\gamma_i := b_i - \sum_{j \in N} a_{ij}\tilde{x}_j$.
- If $\lambda_s > \lambda_{s+1}$, then γ is the gradient of f at π.
- Can incorporate this in a Quasi-Newton (or, with an expression for the Hessian, a Newton) method for finding the minimum.

Spectral Partition Bound (HLW '00)

Let $\mathcal{N} = \{N_1, N_2, ..., N_n\}$ denote a partition of N. Let $C' = 0$ except for $C'[N_k, N_k] = C[N_k, N_k]$.

$$z \leq \sum_{l=1}^{s} \ln \lambda_l(C')$$

- Based on  "Fischer's Inequality"
- For $\mathcal{N} = \{\{1\}, \{2\}, \ldots, \{n\}\}$ we have the diagonal bound
- For $\mathcal{N} = \{N, \emptyset, \emptyset, \ldots, \emptyset\}$ we have the ordinary spectral bound
- As we partition N, the optimal value with respect to C' can not decrease *but the bound can decrease*

Sufficient Optimality Criterion

Let S be a subset of N, with $|S| = s$. If

$$\lambda_s(C[S,S]) \geq \max\{C_{jj} \ : \ j \in N \setminus S\},$$

then S is optimal

Proof. For $\mathcal{N} = \{S, \{s+1\}, \{s+2\}, ..., \{n\}\}$, the bound gives $\ln\det C[S,S]$ $\square$

Sufficient Optimality Criterion: Example

For $S := \{1, 2, \ldots, n/2\}$, let $C =$

$$\begin{array}{r} S: \\ N \setminus S: \end{array} \left(\begin{array}{c|c} nI + E & 0 \\ \hline 0 & \left(\frac{3n}{4}\right) I + E \end{array} \right),$$

and let $s := n/2$. Then

- $\Lambda(S) = \{3n/2, n, n, \ldots, n\}$;
- $\Lambda(N - S) = \{5n/4, 3n/4, 3n/4, ..., 3n/4\}$.
- spectral bound is $\ln\, (3n/2)(5n/4)n^{n/2-2}$;
- “diagonal” bound is $\ln\, (n+1)^{n/2}$;
- For $\mathcal{N} = \{S, \{n/2+1\}, \{n/2+2\}, ..., \{n\}\}$ the spectral partition bound gives $\ln\, (3n/2)n^{n/2-1} = \ln \det C[S, S]$.

Finding a Good Partition

1a. Let $\mathcal{N} = \{S, \{j_1\}, \{j_2\}, ..., \{j_{n-s}\}\}$, where S has high entropy.

1b. Or let $\mathcal{N} = \{N, \emptyset, \emptyset, \ldots, \emptyset\}$ (spectral bound).

1c. Or let $\mathcal{N} = \{\{1\}, \{2\}, \ldots, \{n\}\}$ ("diagonal" bound).

2. Use local search on the space of partitions.

Finding a Good Partition, cont'd

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)* $k \neq l$: $N_k \leftarrow N_k \cup N_l$, $N_l \leftarrow \emptyset$.

Experiments

	original			complementary		
	1a	1b	1c	1a	1b	1c
1	5.5121	5.7070	7.9250	3.3524	5.7070	3.2524
2a	4.5767	4.5793	5.0606	2.6555	2.6077	2.6294
2a–d	4.5767	4.5793	4.5774	2.6302	2.5211	2.6273

Entropy gaps (Ohio Valley sulfate data: $n = 63, s = 31$).

Observations

- Can get substantial improvement over starting partitions
- Complementation is valuable
- Swapping is valuable
- Sophisticated swaps sometimes help
- Robust across starting partition
- Expensive to compute

ILP Bound (LW '00)

$$
\begin{aligned}
g_s(\mathcal{N}) := \quad & \max \sum_{i=1}^{p} \sum_{k=1}^{|N_i|} f_k(N_i) x_k(i) \\
& \text{s.t. } \sum_{k=1}^{|N_i|} x_k(i) \le 1, \ \text{ for } i = 1, 2, \ldots, p; \\
& \sum_{i=1}^{p} \sum_{k=1}^{|N_i|} k x_k(i) = s \\
& x_k(i) \in \{0, 1\}, \ \text{ for } i = 1, 2, \ldots, p, \\
& \qquad\qquad\qquad\quad k = 1, 2, \ldots, |N_i|.
\end{aligned}
$$

ILP Bound, cont'd

- Refines the spectral partition bound.
- Calculate via dynamic programming (assuming $|N_i|$ is bounded):

  Boundary conditions:

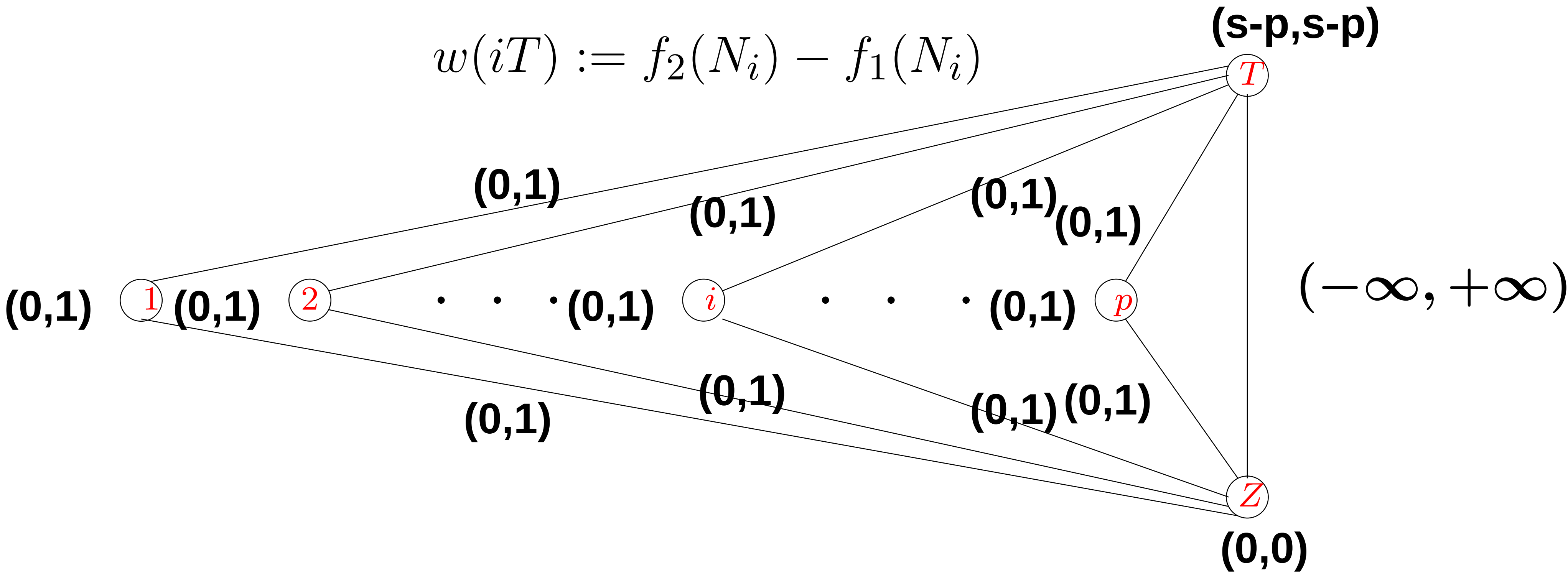
  $v_t(j) := -\infty$ when $\sum_{i=1}^{j} |N_i| < t \leq s$;
  $v_0(0) := 0$.

  $$v_t(j) = \max_{0 \leq k \leq \min\{|N_j|, t\}} \{f_k(N_j) + v_{t-k}(j-1)\}.$$

  Then $v_s(p) = g_s(\mathcal{N})$
- Calculate via Edmonds' min-weight matching algorithm when $|N_i| \leq 2$.

$|N_i| \leq 2$: Min-weight b-matching

$$w(iT) := f_2(N_i) - f_1(N_i)$$



$$w(iZ) := -f_1(N_i)$$

Bound := MinMatching $+ \sum_{i=1}^{p} f_1(N_i)$

Experiments, cont'd

INITIAL PARTITION (DIAG):

7.924975	3.252440

LOCAL SEARCH (2A):

Δ	g_s	$\bar{g}_s$
0	5.060646	2.629423
1	3.705539	1.600504
2	2.790565	1.235069
3	1.961030	1.235069

- Only calculated the $f_k(N_i)$ exactly for $k \leq \Delta$ and $k \geq n - \Delta$
- For $\Delta < k < n - \Delta$, we replaced $f_k(N_i)$ with the spectral upper bound $\sum_{l=1}^{k} \ln \lambda_l(N_i)$

Masked Spectral Bound (AL '04)

A *mask* is a (symmetric) $X \succeq 0$ having $\mathrm{diag}(X) = e$. The associated *masked spectral bound* is

$$\xi_{C,s}(X) := \sum_{l=1}^{s} \ln\left(\lambda_l\left(C \circ X\right)\right)$$

Special combinatorial cases:

- Spectral bound $X := E$
- Diagonal bound $X := I$
- Spectral partition bound $X := \mathrm{Diag}_i(E_i)$

Validity

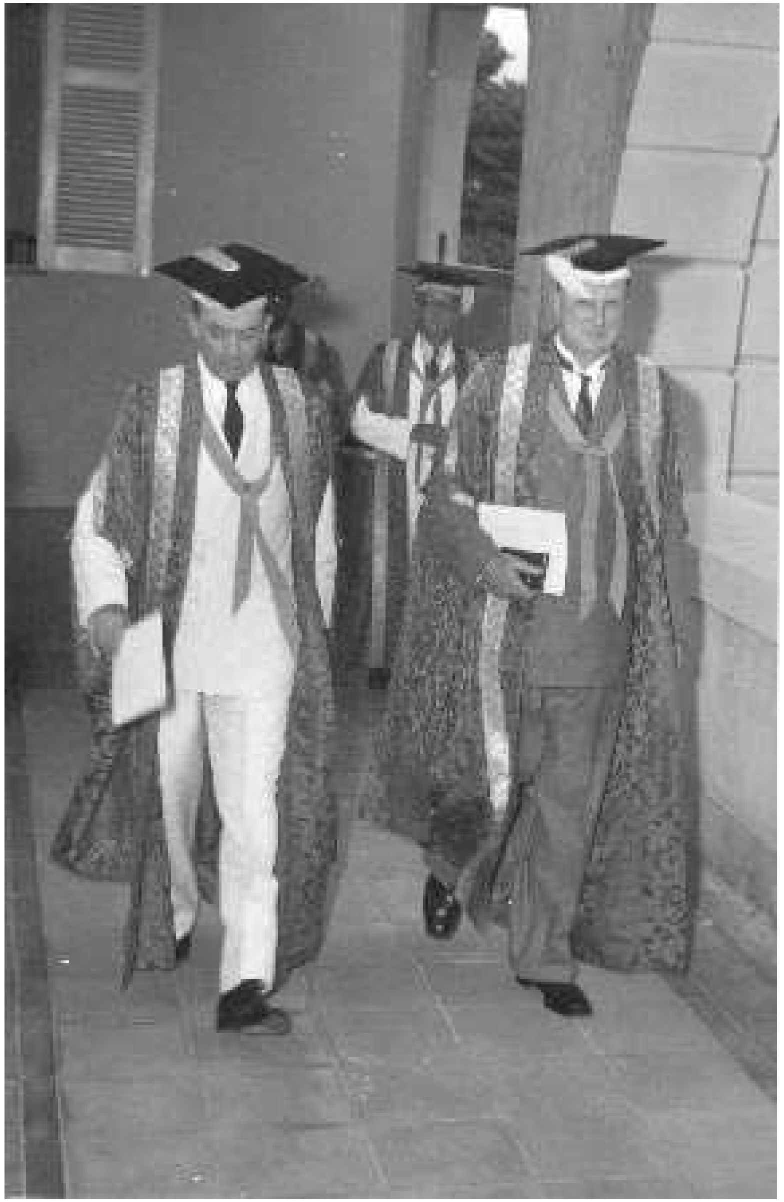
Based on

- $\det A = \prod_l \lambda_l(A)$
- *"Oppenheim's Inequality"*

$$\det\ A \leq \det\ A \circ B / \prod_{j=1}^{n} B_{jj}\ ,$$

where $A \succeq 0$ and $B \succeq 0$

- the eigenvalue inequalities $\lambda_l(A) \geq \lambda_l(B)$, where $A \succeq 0$, and B is a principal submatrix of A



Optimizing the Mask

- Set of masks is a convex set
- $\xi_{C,s}(X)$ is not generally a convex function, so we seek a good local minimizer
- For $\tilde{X} \succeq 0$, let $u_l(C \circ \tilde{X})$ be an eigenvector, of Euclidean norm 1, associated with $\lambda_l(C \circ \tilde{X})$. Then, as long as $\lambda_s(C \circ \tilde{X}) > \lambda_{s+1}(C \circ \tilde{X})$, the gradient of $\xi_{C,s}(\cdot)$ at $\tilde{X}$ is the matrix

  $$\nabla_X \xi_{C,s}(\tilde{X}) = C \circ \sum_{l=1}^{s} \lambda_l(C \circ \tilde{X}) u_l(C \circ \tilde{X}) u_l(C \circ \tilde{X})'$$

- When $\lambda_s(C \circ \tilde{X}) = \lambda_{s+1}(C \circ \tilde{X})$, $\xi_{C,s}(\cdot)$ is not differentiable at $\tilde{X}$. Optimal mask problem corresponds to minimizing a nondifferentiable, nonconvex function with a $\succeq$-constraint

Affine Scaling Heuristic

- For a given $\tilde{X} \succ 0$ with $\mathrm{diag}(\tilde{X}) = e$, let $G = \nabla_X \xi_{C,s}(\tilde{X})$, and consider the linear SDP

  $$\min \{ G \bullet X \ : \ \mathrm{diag}(X) = e, \ X \succeq 0 \}$$

  (where "•" is inner product)

- The affine scaling direction D at $\tilde{X}$ is given by

  $$D := \tilde{X}\big(G - \mathrm{Diag}(u)\big)\tilde{X},$$

  where $u = (\tilde{X} \circ \tilde{X})^{-1} \, \mathrm{diag}(\tilde{X} G \tilde{X})$

- Given the direction D and $0 < \beta < 1$, we consider a step of the form

  $$X := \tilde{X} - \alpha \beta^k D$$

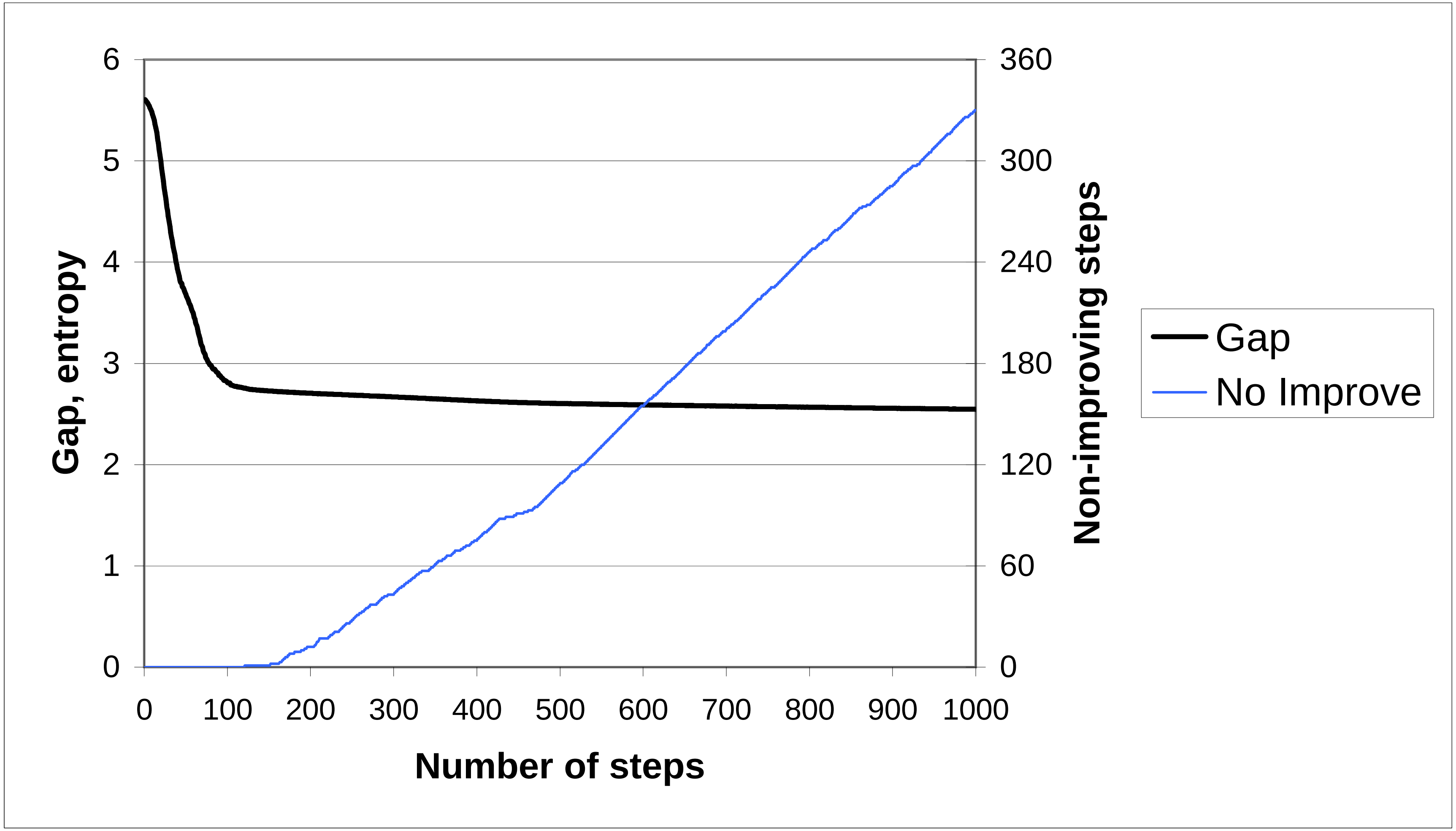
Affine Scaling Heuristic, cont'd.

- The initial step parameter α corresponds to a fixed fraction of either a “short step" or a “long step"
- The short step is based on the limit of the Dikin ellipsoid that is used to define D
- The long step is based on the feasible region $X \succeq 0$
- We attempt a step with $k = 0$, and we accept the resulting X if $\xi_{C,s}(X) < \xi_{C,s}(\tilde{X})$
- If not, we retract by incrementing k a limited number of times in an attempt to decrease $\xi_{C,s}(\cdot)$
- For the highest allowed k, we accept X even if $\xi_{C,s}(X) > \xi_{C,s}(\tilde{X})$

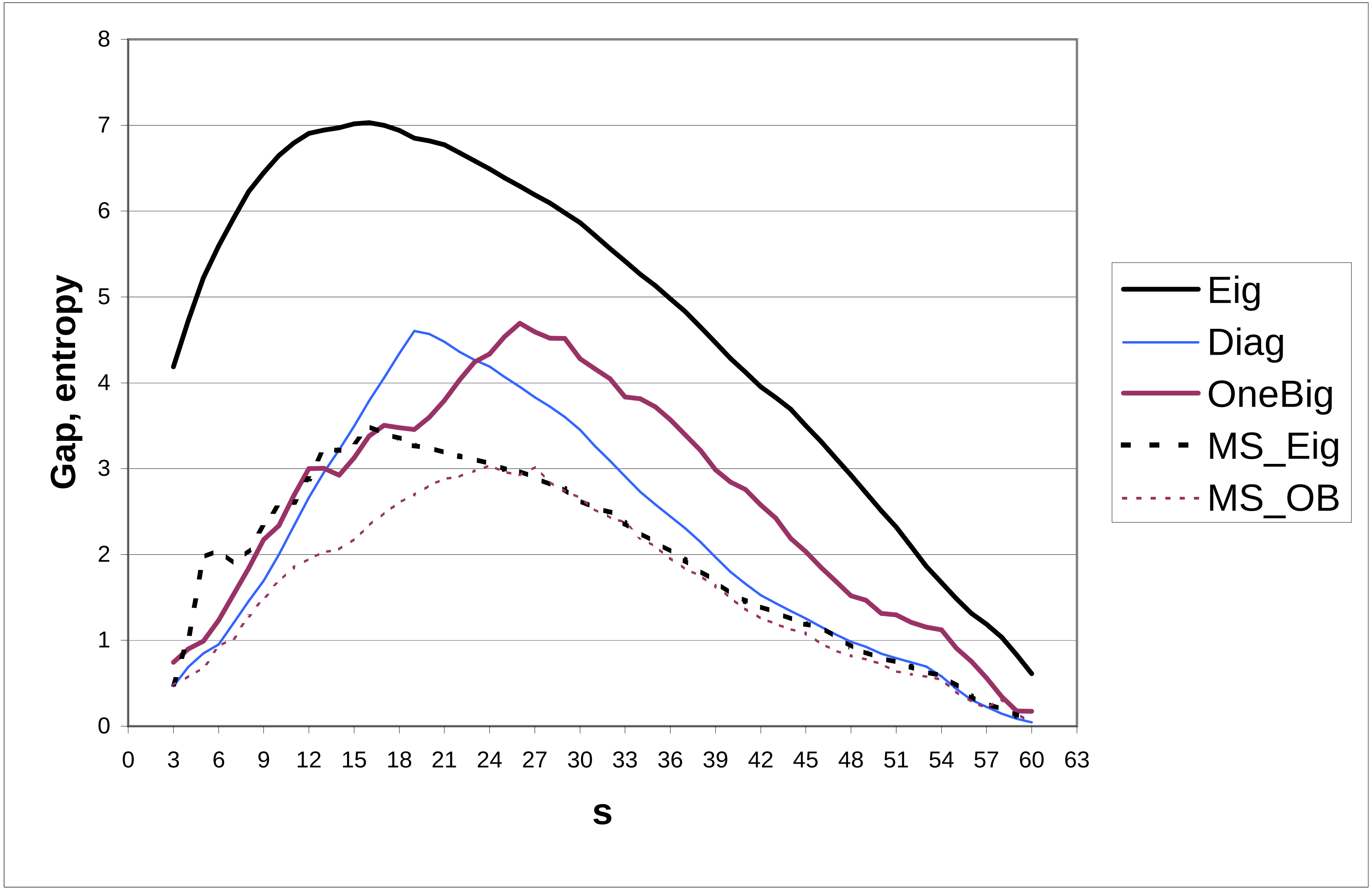
Computational Experiments

- Implemented in MATLAB
- Used both original and complementary bounds

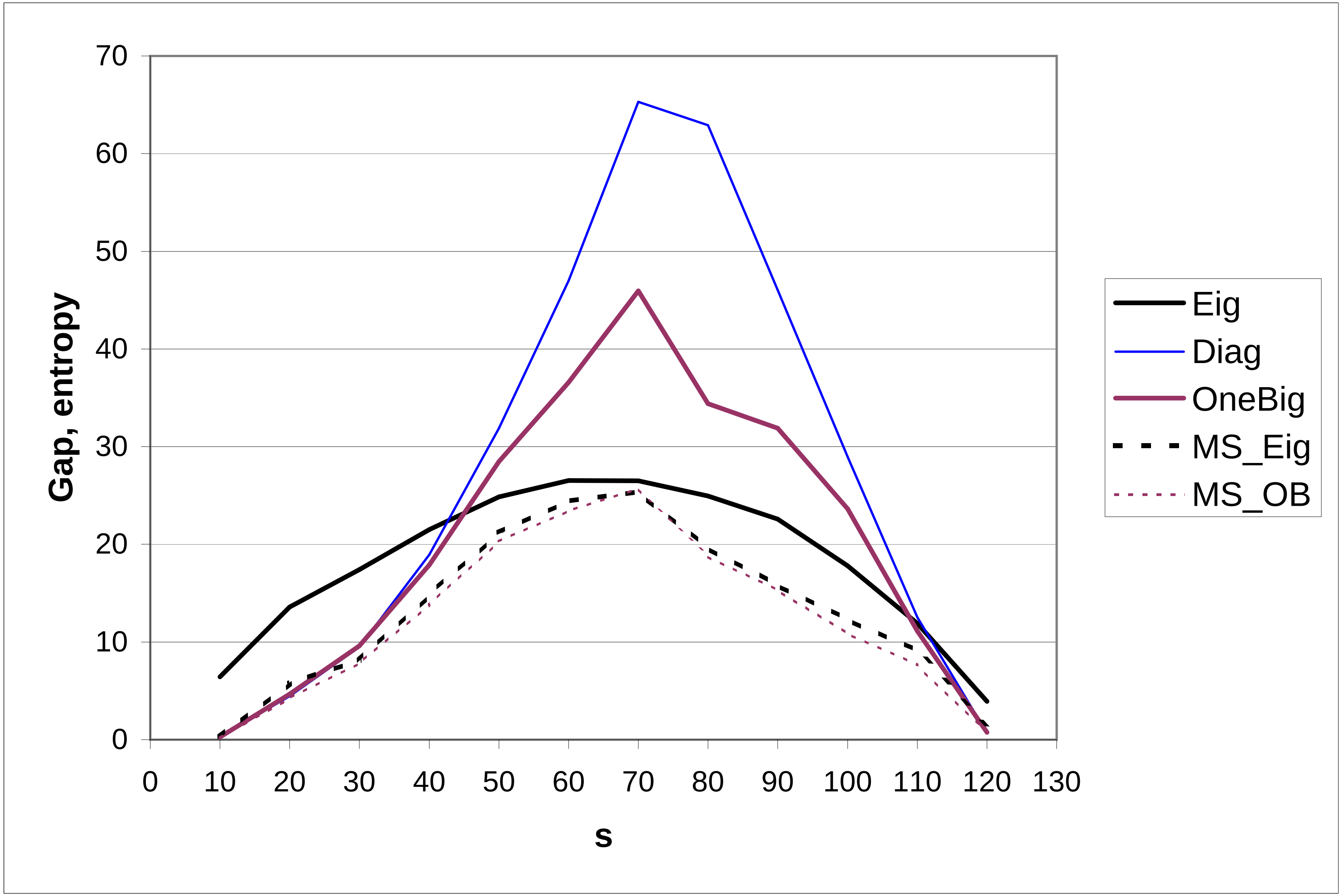
Decrease in bound: $n = 63$, $s = 31$



Comparison of bounds: $n = 63$



Comparison of bounds: $n = 124$



Variations on the Theme

Applying Oppenheim's inequality slightly differently, we obtain the different bounds:

- $u := \min\left\{\prod_{l=1}^{s}\lambda_l(C\circ X)/\prod_{l=1}^{s}\mathrm{diag}_{[l]}(X)\ :\ X\succeq 0\right\},$
  where $\mathrm{diag}_{[l]}(X) = l$-th least component of $\mathrm{diag}(X)$
- $v := \min\left\{\prod_{l=1}^{s}\lambda_l(C\circ X)\ :\ X\succeq 0,\ \mathrm{diag}(X)=e\right\}$
- $w := \min\left\{\prod_{l=1}^{s}\lambda_l(C\circ \hat{X})\ :\ X\succeq 0,\ \hat{X}_{ij} := \frac{X_{ij}}{\sqrt{X_{ii}X_{jj}}}\right\}$

Proposition $u \le v = w$

Some References

- Anstreicher and Lee. A masked spectral bound for maximum-entropy sampling. In A. Di Bucchianico, H. Läuter and H.P. Wynn, eds., “MODA 7 - Advances in Model-Oriented Design and Analysis”, Contrib. to Stat., Springer, Berlin, 2004
- Lee. Maximum entropy sampling. In A.H. El-Shaarawi and W.W. Piegorsch, eds., “Encyclopedia of Environmetrics”. Wiley, 2001
- Lee. Semidefinite programming in experimental design. In H. Wolkowicz, R. Saigal and L. Vandenberghe, eds., “Handbook of Semidefinite Programming”, International Ser. in Oper. Res. and Manag. Sci., Vol. 27, Kluwer, 2000