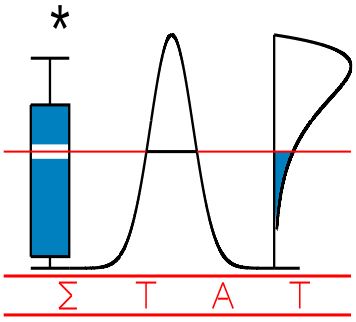
T E C H N I C A L

R E P O R T

0472

AN EVALUATION OF VARIOUS INITIAL CONFIGURATIONS AND SIMULATED ANNEALING-BASED ALGORITHMIC VARIANTS FOR HIERARCHICAL CLASSES ANALYSIS

CEULEMANS, E. and I. VAN MECHELEN



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

An evaluation of various initial configurations and simulated annealing-based algorithmic variants for hierarchical classes analysis

Eva Ceulemans & Iven Van Mechelen
Katholieke Universiteit Leuven

abstract

Hierarchical classes models are quasi-order retaining Boolean decomposition models for *N*-way *N*-mode binary data. These models are fitted to data by means of rationally started alternating least squares algorithms. The results of extensive simulation studies show that these algorithms, although succeeding quite well in recovering the underlying truth, end quite often in a local minimum. In this paper we investigate whether these local minima-problems are due to the starting configuration of the algorithms and/or the alternating character. In particular, the effect of using (1) other types of initial configurations and (2) simulated annealing-based algorithmic variants is evaluated. Initial simulation results for the original two-way HICLAS-model show that a (noisy) rationally started simulated annealing-based algorithm yields the most promising results. Directions for further research are discussed.

overview

1. HICLAS recapitulation
   a. model
   b. algorithm
2. the local minimum problem: JOC simulation study
3. two possible solutions
   a. changing initial configuration of ALS algorithm
   b. simultaneous optimization via simulated annealing
4. results
5. discussion

1.a. HICLAS recapitulation: model

- model for *I* by *J* binary **D**
- **D** is approximated by same-sized binary **M**
- **M** is decomposed into *I* by *R* binary **A** and *J* by *R* binary **B**

model matrix **M**						
	attributes					
objects	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

A		
objects	I	II
o1	1	1
o2	1	1
o3	1	0
o4	1	0
o5	1	1
o6	0	1
o7	0	0

model matrix **M**						
	attributes					
objects	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

B		
attributes	I	II
a1	1	0
a2	0	1
a3	0	0
a4	1	1
a5	0	1
a6	1	0

1.a. HICLAS recapitulation: model

- model for *I* by *J* binary **D**
- **D** is approximated by same-sized binary **M**
- **M** is decomposed into *I* by *R* binary **A** and *J* by *R* binary **B**
- **A** and **B** represent two types of relations in **M**
  - association
  - quasi-order

A

objects	I	II
o1	1	1
o2	1	1
o3	1	0
o4	1	0
o5	1	1
o6	0	1
o7	0	0

model matrix **M**

objects \ attributes	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

$$m_{ij} = \bigoplus_{r=1}^{R} a_{ir} b_{jr}$$

B

attributes	I	II
a1	1	0
a2	0	1
a3	0	0
a4	1	1
a5	0	1
a6	1	0

A		
objects	I	II
o1	1	1
o2	1	1
o3	1	0
o4	1	0
o5	1	1
o6	0	1
o7	0	0

model matrix **M**						
	attributes					
objects	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

$$m_{ij} = \bigoplus_{r=1}^{R} a_{ir} b_{jr}$$

B		
attributes	I	II
a1	1	0
a2	0	1
a3	0	0
a4	1	1
a5	0	1
a6	1	0

A		
objects	I	II
o1	1	1
o2	1	1
o3	1	0
o4	1	0
o5	1	1
o6	0	1
o7	0	0

model matrix **M**						
	attributes					
objects	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

$$\begin{cases} \mathbf{m}_{i.} \le \mathbf{m}_{i'.} \Leftrightarrow \mathbf{a}_{i.} \le \mathbf{a}_{i'.} \\ \mathbf{m}_{.j} \le \mathbf{m}_{.j'} \Leftrightarrow \mathbf{b}_{j.} \le \mathbf{b}_{j'.} \end{cases}$$

B		
attributes	I	II
a1	1	0
a2	0	1
a3	0	0
a4	1	1
a5	0	1
a6	1	0

model matrix **M**						
	attributes					
objects	a1	a2	a3	a4	a5	a6
o1	1	1	0	1	1	1
o2	1	1	0	1	1	1
o3	1	0	0	1	0	1
o4	1	0	0	1	0	1
o5	1	1	0	1	1	1
o6	0	1	0	1	1	0
o7	0	0	0	0	0	0

A		
objects	I	II
o1	1	1
o2	1	1
o3	1	0
o4	1	0
o5	1	1
o6	0	1
o7	0	0

$$\begin{cases} \mathbf{m}_{i.} \le \mathbf{m}_{i'.} \Leftrightarrow \mathbf{a}_{i.} \le \mathbf{a}_{i'.} \\ \mathbf{m}_{.j} \le \mathbf{m}_{.j'} \Leftrightarrow \mathbf{b}_{j.} \le \mathbf{b}_{j'.} \end{cases}$$

B		
attributes	I	II
a1	1	0
a2	0	1
a3	0	0
a4	1	1
a5	0	1
a6	1	0

1.b. HICLAS recapitulation: algorithm

- loss function: $f(\mathbf{M}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(d_{ij} - m_{ij} \right)^2$

- algorithm consists of two steps
  - ALS procedure: look for **A**, **B** that minimize loss function
  - closure operation: modify **A**, **B** so as to represent the quasi order relations in **M**

- ALS procedure
  - starts from initial configuration for **A** (resp. **B**): $\mathbf{A}^0$
  - initial configuration can be rational, random or user provided
  - conditional upon $\mathbf{A}^0$, $\mathbf{B}^1$ is estimated conditionally optimal by means of branch-and-bound Boolean regression
  - conditional upon $\mathbf{B}^1$, $\mathbf{A}^1$ is estimated by means of Boolean regression
  - conditional updating of **A** and **B** continues until no further improvement in the loss function is observed

2. JOC simulation study

- 3 types of *I* by *J* matrices involved
  - true matrices **T**
  - data matrices **D**: **T** perturbed with error
  - model matrices **M**: yielded by rationally started ALS algorithm
- 3 independent variables were orthogonally crossed
  - size: 15 by 25, 20 by 20, 80 by 20, 40 by 40 (+ 2 big sizes)
  - rank underlying **T**: 3, 5, 8
  - error level: 0, 5, 10, 15, 20, 25 %
- 25 replications per cell

- of the 1800 analyses, 1295 or 72% ended in a solution with $f(\mathbf{M}) \leq$ number of discrepancies **T**-**D**
- as number of discrepancies **T**-**D** constitutes an upper bound for $f(\mathbf{M})$ of global minimum: in at least 28% of the cases rationally started ALS algorithm ends in local minimum

3.a. changing initial configuration of ALS algorithm

- often, one deals with local minima problems by means of a multistart procedure
- implies rerunning the algorithm using a user-specified number of random starts (i.e., 100) and retaining the best solution only
- two types of random starts:
  - pseudo-random start: the columns of $\mathbf{A}^0$ consist of randomly chosen columns of **D**
  - truly-random start: a_{ir}^0 are independent realizations of a Bernoulli variable with the probability parameter depending on the number of ones in **D**

- however, (1) vastness of HICLAS solution space (i.e., 2^{IR} possible initial configurations) and (2) generally, low number of iterations before convergence is reached => will random multistart procedure solve the problem?

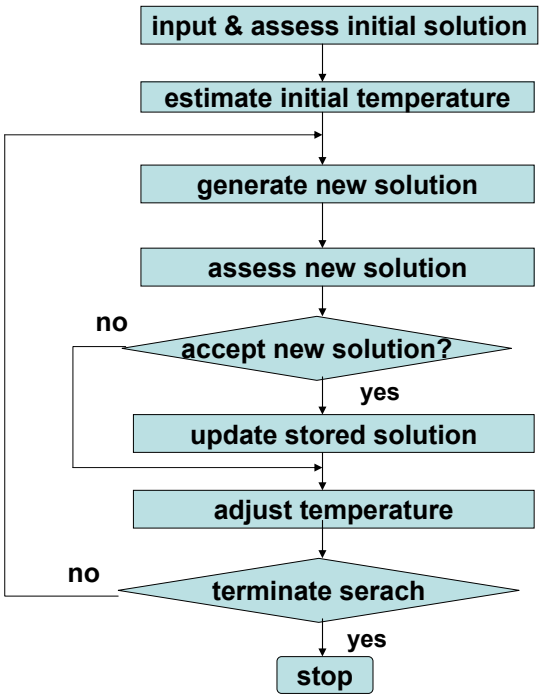
- therefore, smart random multistart procedure: rational start perturbed with a small amount of error (i.e., 10 %)

3.b. simulated annealing

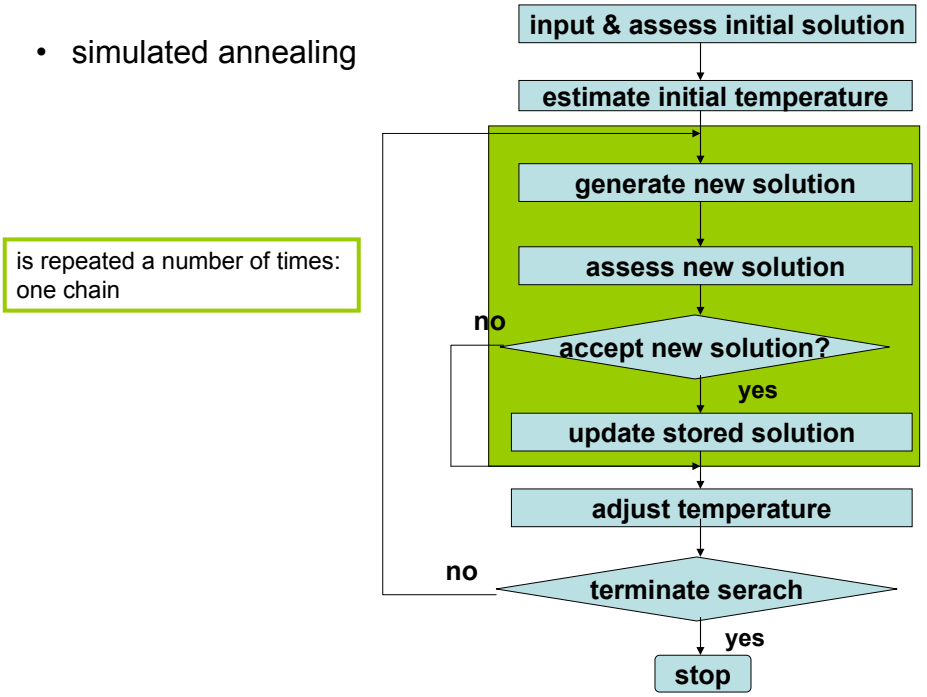
- a random-search technique, which exploits the analogy between the way that liquids freeze into a minimum energy structure (the annealing process) and the search for a minimum in a more general system

- annealing:
  - at high temperatures, the molecules of a liquid move freely with respect to one another
  - if the liquid is cooled slowly, thermal mobility is lost and the atoms are able to form a pure crystal; this crystal is the state of minimum energy
  - if the liquid is cooled quickly, the obtained crystal will contain imperfections; such a crystal has a somewhat higher energy

- simulated annealing



- simulated annealing

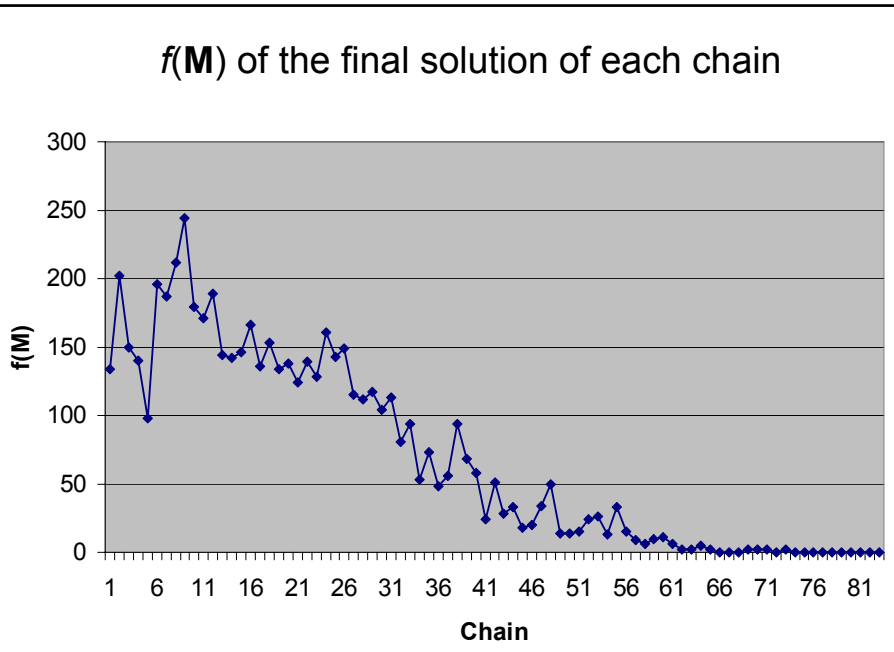


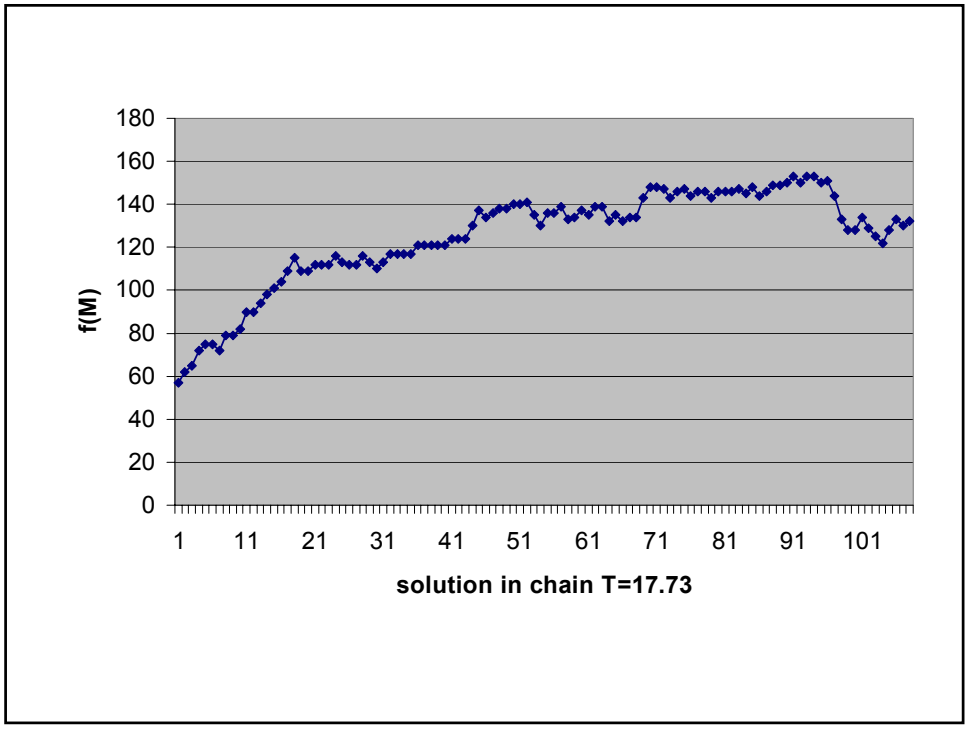
- a solution: **A** and **B**
- initial solution: rational, noisy rational, pseudo-random, truly random
- generation of new solutions: alter the entries of **A** and **B** with probability 1/((*I*+*J*)**R*)
- assessing a solution: calculate the loss function value *f*(**M**)
- acceptation decision:
  - better solution is always accepted
  - worse solution is accepted with probability

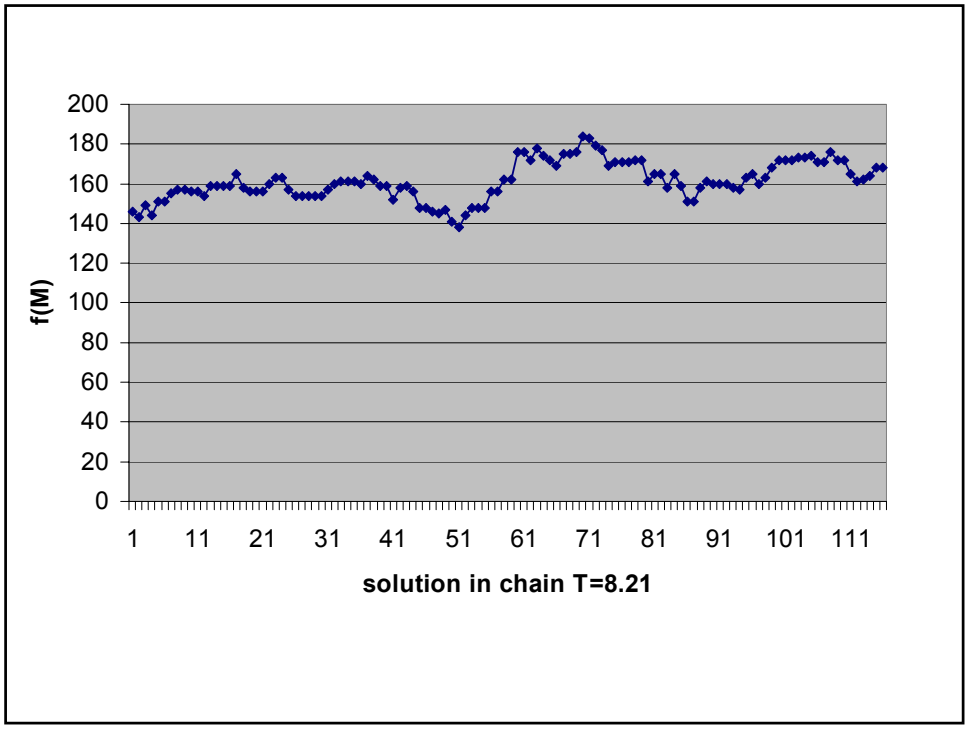
$$p = \exp((f(\mathbf{M})_{curr} - f(\mathbf{M})_{new}) / T)$$

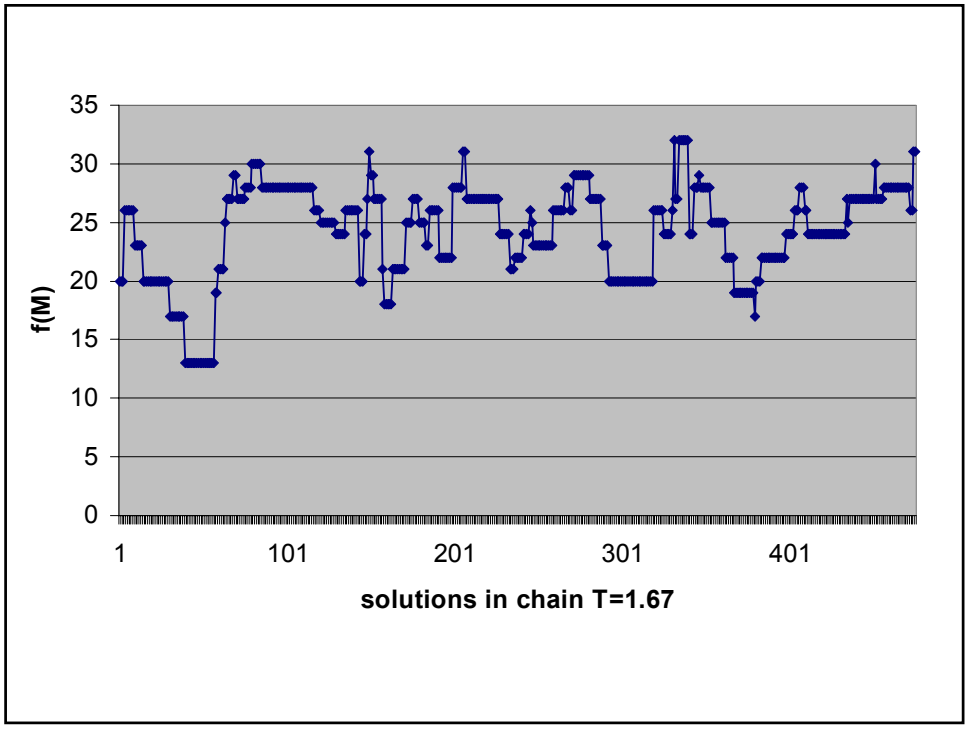
- estimating the initial temperature T_0:
  - a suitable initial temperature results in an average probability of accepting a 'worse' solution of .8
  - is estimated by conducting an initial search in which all increases are accepted
  - $$T_0 = \frac{\text{average increase in loss function}}{\ln(.8)}$$

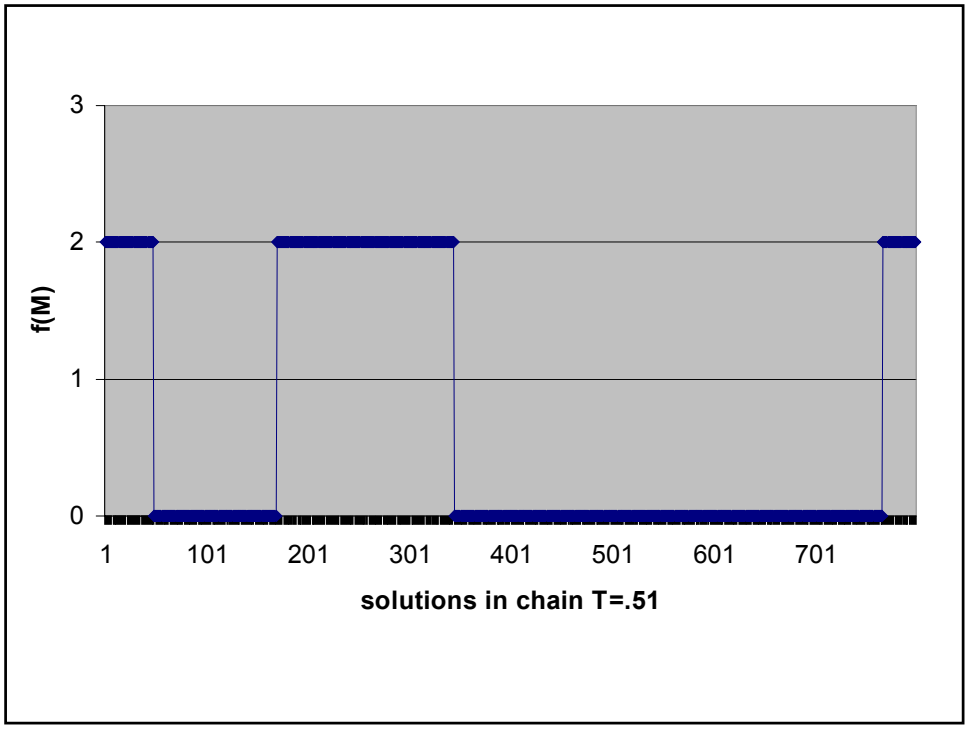
- length of chain: $(I+J)*2^R$ solutions or $.1*((I+J)*2^R)$ solutions accepted
- adjusting the temperature: $T_{k+1}=.95T_k$
- termination criterion: T_k has a very low value or the final solution of ten subsequent chains is the same




180
160
140
120
100
80
60
40
20
0
f(M)
1
11
21
31
41
51
61
71
81
91
101
solution in chain T=17.73


200
180
160
140
120
100
80
60
40
20
0
f(M)
1
11
21
31
41
51
61
71
81
91
101
111
solution in chain T=8.21


35
30
25
20
15
10
5
0
f(M)
1
101
201
301
401
solutions in chain T=1.67


3
2
1
0
f(M)
1
101
201
301
401
501
601
701
solutions in chain T=.51

4. results

- 8 analyses of JOC simulated data sets
- percentage of data sets with $f(\mathbf{M}) \leq$ number of discrepancies **T**-**D**

	algorithm	
start	ALS	simulated annealing
rational	71.9	95.1
pseudo-random	93.2	76.4
truly-random	94.6	69.8
noisy rational	94.9	94.4

- percentage of data sets with lowest $f(\mathbf{M})$

	algorithm	
start	ALS	simulated annealing
rational	19.1	63.7
pseudo-random	51.4	34.1
truly-random	51.0	31.3
noisy rational	51.2	61.3

- badness of recovery of the association relation

	algorithm	
start	ALS	simulated annealing
rational	.099	.083
pseudo-random	.087	.091
truly-random	.086	.095
noisy rational	.087	.084

- goodness of recovery of the quasi order relations

	algorithm	
start	ALS	simulated annealing
rational	.880	.906
pseudo-random	.902	.896
truly-random	.904	.892
noisy rational	.903	.905

5. discussion

- for ALS, almost no difference between multistart procedures, which all work better than rational start
- for SA, best results with rational and noisy rational start
- hence, best combination: perform rationally started ALS procedure and use output as input for SA
- question: difference between random ALS and random SA due to 100 starts for ALS vs 1 start for SA?
- what about other global optimization procedures, like genetic algorithms and tabu search?

- ALS results do not generalize to three-way case, noisy rational > rational > random
- three-way SA?