

Research Article

A Simulated Annealing Algorithm for D-Optimal Design for 2-Way and 3-Way Polynomial Regression with Correlated Observations

Chang Li¹ and Daniel C. Coster²

¹ Business School, Shandong University of Political Science and Law, 63 East Jiefang Road, Jinan, Shandong 250014, China

² Department of Mathematics and Statistics, Utah State University, Logan, UT 84341, USA

Correspondence should be addressed to Chang Li; wisdom101126@hotmail.com

Received 10 November 2013; Revised 1 March 2014; Accepted 1 March 2014; Published 26 March 2014

Academic Editor: Li Weili

Copyright © 2014 C. Li and D. C. Coster. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Much of the previous work in D-optimal design for regression models with correlated errors focused on polynomial models with a single predictor variable, in large part because of the intractability of an analytic solution. In this paper, we present a modified, improved simulated annealing algorithm, providing practical approaches to specifications of the annealing cooling parameters, thresholds, and search neighborhoods for the perturbation scheme, which finds approximate D-optimal designs for 2-way and 3-way polynomial regression for a variety of specific correlation structures with a given correlation coefficient. Results in each correlated-errors case are compared with traditional simulated annealing algorithm, that is, the SA algorithm without our improvement. Our improved simulated annealing results had generally higher D-efficiency than traditional simulated annealing algorithm, especially when the correlation parameter was well away from 0.

1. Introduction

D-optimality is a popular criterion for optimal experimental design. The model for polynomial regression can be written as in Zhu et al. [1]:

$$y_i = f_i(x)' \beta + \epsilon_i, \quad (1)$$

where $i = 1 \cdots n$, β is a k -vector of parameters, and $f_i(x) = (f_{1i}(x), f_{2i}(x), \dots, f_{ki}(x))$ is a k vector of polynomial functions of x , and n is the number of observations. Our purpose is to estimate the coefficient vector β or part of the vector β of primary interest.

In some experimental settings, the observations may be correlated according to various structures or patterns. The motivation of the research in optimal designs with correlated observations can be found in Dette et al. [2]. Müller [3] introduced optimal design with correlated observations in detail.

The simulated annealing algorithm is a probabilistic “hill climbing” algorithm for optimization in the absence of an

analytical solution. The application of simulated annealing algorithm for optimal design problem was first proposed by Haines [4]. Lejeune [5] proposed a simulated annealing algorithm for D-optimal design with uncorrelated observations. The simulated annealing algorithm with a reheating process is introduced in Dimitris and Omid [6] and Abdullah et al. [7]. In Zhu [8], Zhu solved the 1-way D-optimal design for polynomial regression with correlated observations using a simulated annealing algorithm. Cheng [9] produced D-optimal designs with block effects, which can be considered as a special case of the D-optimal design problem with correlated observations, since the block effects can be incorporated into the correlation structure.

Most previous work only considered the simplest case, that is, optimal design for 1-way polynomial regression. However, in real world problems, the response variable is usually influenced by multiple effects and their interactions. This kind of problem is more complicated and cannot get satisfactory result by existing algorithms or their generalizations.

In this paper, we propose a modified, improved simulated annealing algorithm to approximately solve for D-optimal design for 2-way and 3-way polynomial regression with correlated observations. This algorithm is applicable to any number of observations, not necessarily a multiple of the dimension of the parameter vector. It conquers the shortcoming of previous work, which mainly concentrated on the case that n (the number of observations) is a multiple of k (the number of the coefficients to be estimated or, equivalently, the dimension of the parameter β). We also provide a reinforced version of our simulated annealing algorithm with a reheating process.

2. Model and Correlation Structures

2.1. Model. The full model for second-order 2-way and 3-way polynomial regression is presented by Boon [10] and Pukelsheim [11]. The model for the second-order 2-way polynomial regression is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + \beta_4 x_{2i}^2 + \beta_5 x_{1i} x_{2i} + \epsilon_i, \quad (2)$$

where $i = 1, 2, \dots, n$, and each of the x_{1i} and x_{2i} are in $[-1, 1]$, where ϵ_i has mean 0 and variance σ^2 but are not necessarily independent.

The design matrix is $X = (x_{ij})_{n \times 6}$. The first column is all 1's, and the other 5 columns correspond to the values of $X_1, X_2, X_1^2, X_2^2, X_1 X_2$, respectively. That is, each column of X corresponds to one design variable (or their square or interaction effect) in the model.

The model for the second-order 3-way polynomial regression is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i}^2 + \beta_5 x_{2i}^2 + \beta_6 x_{3i}^2 + \beta_7 x_{1i} x_{2i} + \beta_8 x_{2i} x_{3i} + \beta_9 x_{1i} x_{3i} + \epsilon_i, \quad (3)$$

where $i = 1, 2, \dots, n$.

The design matrix is $X = (x_{ij})_{n \times 10}$, and the definition is similar with 2-way polynomial regression.

D-optimality aims to maximize of the determinant of the information matrix, where the information matrix for these models is

$$M = X'V^{-1}X, \quad (4)$$

where

$$V = \text{cov}(Y) = \sigma^2(\rho_{ij})_{n \times n} \quad (5)$$

is the variance covariance matrix of the errors. Some common correlation structures for V are introduced below.

2.2. Correlation Structures. We define commonly used correlation structures below for a single correlation parameter ρ .

(i) Circulant correlation: see Zhu et al. [1]:

$$\text{cov}(y_i, y_j) = \begin{cases} \sigma^2 & i = j \\ \rho\sigma^2 & |i - j| = 1 \text{ or } |i - j| = n - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

(ii) Nearest neighbor correlation: see Zhu [12]:

$$\text{cov}(y_i, y_j) = \begin{cases} \sigma^2 & i = j \\ \rho\sigma^2 & |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

(iii) Autoregressive correlation: see Dette et al. [2]:

$$\text{cov}(y_i, y_j) = \sigma^2 \rho^{|i-j|}, \quad (8)$$

where $i, j = 1, 2, \dots, n$.

(iv) Completely symmetric block structure: see Cadima et al. [13]:

$$\begin{pmatrix} R & R12 & \dots & R1b \\ R21 & R & \dots & R2b \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & R \end{pmatrix}. \quad (9)$$

Here R is a $k \times k$ matrix with the elements on the main diagonal =1, and all other elements = ρ , (k is the common block size). ρ is the correlation coefficient for the observations in the same block. R_{ij} is a $k \times k$ block with all elements = ρ_{ij} . In this paper we take all of the ρ_{ij} equal to the same coefficient ρ' .

Note that one commonly used block correlation structure is proposed by Atkins and Cheng [14]:

$$\text{cov}(Y) = \sigma^2 (I_b \otimes V) \quad (10)$$

with $V = (1 - \rho)I_k + \rho J_k$. Here J_k is the $k \times k$ matrix with all of the elements =1. This is a special case of (iv) with $R_{ij} = 0$.

3. Improved Simulated Annealing Algorithm for 2-way and 3-way Second-Order Polynomial Regression with Correlated Observations

3.1. The Principle of Simulated Annealing. The simulated annealing (SA) algorithm belongs to a class of heuristic probabilistic hill-climbing algorithms; see Zhu [8] and Lejeune [5]. The SA algorithm attempts to globally maximize an energy function $E(X)$ for X in a specified state space (a design region for our D-optimality problem), by moving about the state space according to a transition mechanism defined by random perturbations of the current solution, X_c , to a new candidate solution, X_n . Letting $dE = E(X_n) - E(X_c)$, if $dE > 0$, accept X_n as the current solution. Otherwise, accept X_n as the current solution with probability $\exp(dE/T_c)$, where

T_c is the current value of a temperature control parameter, T . Thus, there is positive probability that the algorithm will move to a poorer design, which is the key feature of the SA search algorithm, as it provides for the possibility that the algorithm will escape a local maximum. As the algorithm proceeds, the temperature decreases, making it less likely that designs with lower energy will be accepted. Convergence of the SA algorithm to a highly efficient design (a globally optimal solution is never guaranteed to be found) depends on the convergence to a stationary distribution of the underlying Markov chain, which typically requires a large number of iterations as well as a suitably chosen transition scheme over the state space.

3.2. Simulated Annealing Algorithm for D-Optimal Design for 2-Way and 3-Way Polynomial Regression. For 2-way polynomial regression, the $n \times 6$ design matrix is fully determined by the values of X_1 and X_2 , each in $[-1, 1]$. Therefore, at each iteration of our simulated annealing algorithm, a new design matrix is obtained by perturbing the current values of X_1 and X_2 . We denote the current values of X_1 and X_2 by X_{1c} and X_{2c} and new values by X_{1n} and X_{2n} , respectively.

In many applications of simulated annealing, the values of only one current design point are perturbed (by some random mechanism) at each iteration, and typically a systematic pass is made through all design points in this manner, and the process repeated until “convergence” is achieved according to a specified stopping condition. Alternatively, all design points are perturbed simultaneously. However, both of these traditional methods were found to be inefficient for our D-optimal design with correlated errors. Thus, we used a modification that improved convergence and solution quality. Our modification was to divide the design points into three parts, of equal or nearly equal size, and perturb all points in each part in an “inner” loop, while systematically doing this for each of the three parts. This represented a middle ground for the perturbation scheme between the two traditional perturbation methods, one at each extreme, as described above.

Our modified simulated annealing algorithm was as follows.

Step 1. Initialize starting temperature T_0 , finishing temperature T_f , temperature reduction coefficient r , perturbation neighborhood control parameter g_0 , and initial design matrix X_0 . Control parameter g_c is chosen from $[0, 1]$ and is used to adjust the size of the perturbations as the algorithm proceeds. Calculate the energy function of the current design: $E(X_c) = \text{Det}(X'V^{-1}X)$.

Divide the n design points (rows of X) into three parts. If $n = 3k$, for some positive integer k , then each part has $= n/3$ design points. If $n = 3k + 1$, the first two parts have k design points and the third part has $k + 1$. Similarly, if $n = 3k + 2$, the first part has k points, and the other two have $k + 1$ design points.

Step 2. Outer loop: cycle through each of the 3 parts of X systematically, repeating the following inner loop.

Inner loop:

- (i) let Z_1 and Z_2 be $n \times 1$ vectors with each element of Z_i ($i = 1, 2$) sampled at random from $[-1, 1]$ for those design points belonging to the current part of X under consideration; all remaining elements of Z_i are set equal to 0;
- (ii) generate new candidate design points $X_{1n} = X_{1c} + g_c Z_1$ and $X_{2n} = X_{2c} + g_c Z_2$; if any element of X_{1n} or X_{2n} falls outside $[-1, 1]$, set the value to the closest boundary value of the design region;
- (iii) determine $E(X_n)$;
- (iv) if $dE = E(X_n) - E(X_c) > 0$, accept the new design by setting $X_c = X_n$; otherwise, compare $\exp(dE/T)$ with a random number chosen uniformly from $[0, 1]$ multiplied by a coefficient 1.01^c ; if $\exp(dE/T)$ is greater than this number, we set $X_c = X_n$; if not, keep the X_c unchanged.

Step 3. If $T_c < T_f$, stop. Otherwise, increase the counter c to $c + 1$, set $T_c = rT_{c-1}$, $g_c = rg_{c-1}$, $c = c + 1$, and return to Step 2.

Reduction Control Parameter r . This tuning parameter is chosen by the user but is often set about 0.98-0.99 for geometric rate of reduction in the temperature.

Perturbation Control Parameter g . Typically, g_0 is set close to 1, allowing large perturbations in design points at early iterations. As solution quality improves and the temperature decreases, g_c also decreases, localizing perturbations to a smaller neighborhood of the current design which is more likely to be close to a global optimum when iteration counter c is large.

Reheating. The annealing algorithm can be reinforced by using “reheating.” Specifically, after the usual stopping condition based on the temperature is reached in Step 3, the process is repeated, often several times, by reheating to the original starting temperature, and continuing at Step 2. Since the reheating process is much more time consuming than nonreheating process, so we have to weigh and balance between the computing time and effect of computation. In this paper, we mainly run the improved simulated annealing algorithm without reheating process. In Table 6, we present results of the algorithm for $n = 12$ and three correlation structures without and with reheating.

For 3-way polynomial regression, the only difference of the algorithm is in the inner circulation; we do all of the operation on 3 vectors: X_1 , X_2 , and X_3 .

3.3. Improvements from This Algorithm Compared with the Traditional Simulated Annealing Algorithm. (1) There are 2 vectors, X_{1c} and X_{2c} , to be changed (for 3-way polynomial regression, there are 3 vectors to be changed). In this case, the traditional simulated annealing algorithm, which treats the perturbation vector Z as a whole, does not produce satisfactory results. In our modified algorithm, we divide the

Z (and consequently the perturbation process) into 3 parts and make perturbations part by part. This method ensures that we do not miss any corner of the design region and is more precise than the usual annealing method. Additionally, this part-by-part perturbation scheme allows the number of observations to be any number, not necessarily to be multiple of the number of coefficients. This makes our algorithm more flexible since it can be applied to experiments with any number of observations.

(2) We shrink the search neighborhood and increase the threshold for accepting a perturbation each time we lower the temperature. That is, when the temperature is high, we search in a wide neighborhood and are more likely to jump out of the local optimum. At each time we lower the temperature, we make the perturbation neighborhood smaller and make the acceptance threshold higher so it becomes harder to leave a local optimum. We implement this approach by multiplying the scale number g by the reduction coefficient r and multiplying the random number to be compared with dE by a coefficient, 1.01^c , at each time we decrease the temperature. Here c initially is 0 and will increase by 1 each time we decrease the temperature.

This approach is in accordance with the idea of simulated annealing; that is, when the temperature becomes lower, the “molecules” are less active and tend to an equilibrium stabilization. This modification resulted in improved relative efficiency of the final design.

(3) In each part of Step 3, we repeat the iterations until the improvement is less than a small threshold value multiple times. This guarantees we go to the next step only when the improvement is negligible and none in the current step. In other words, we do not miss any valuable improvement. We take the threshold as $0.02 \times$ determinant of the current information matrix as the threshold value.

4. Results and Comparison with Traditional Simulated Annealing Algorithm

In this part, we compare the results from our improved simulated annealing algorithm with that of traditional simulated annealing algorithm, that is, the SA algorithm without our improvement. Since the most often used correlation parameters are 0.1 and 0.4, in Tables 1, 2, 3, 4, 5, 6, and 7, we mainly use these 2 parameters in the computation and comparison. The result of other correlation parameters can be gotten by the same algorithm by a simple adjustment of the parameters.

In Table 1 through Table 4, we present the comparisons of our improved simulated annealing results and the traditional simulated annealing algorithm when observations number n is a multiple of 6 using each of the autoregressive, circulant, nearest neighbor, and block correlation structures. We also get the ratio of the results of the two algorithms, which will tell us how much our improved simulated annealing algorithm is better than the traditional simulated annealing algorithm.

Tables 1–3 present the comparison of the results of the two SA algorithms for the autoregressive, circulant, and nearest neighbor structure for designs of sizes 6, 12, and 18 and

correlation parameters of 0.1 and 0.4, and Table 4 presents the comparison of the results of the two SA algorithms for the block structure for designs of size 12 and correlation parameters of 0.1 and 0.4.

From these tables, we see that under all of the cases, the determinants obtained by our improved simulated annealing algorithm are much higher than that of the traditional simulated annealing algorithm. When ρ changes from 0.1 to 0.4, and when n (the number of observations) gets larger, the ratio of the determinants of our improved simulated annealing algorithm and that of the traditional simulated annealing algorithm increase rapidly. So the D-efficiency of our improved simulated annealing algorithm is much better than that of the traditional simulated annealing algorithm, especially when ρ and n get even larger.

For the case that the observations number n is not a multiple of the dimension of the parameter vector, the traditional simulated annealing algorithm does not work. However, our improved simulated annealing algorithm is a powerful algorithm to get the determinant for any number of n . We list the results of our improved simulated annealing algorithm with various ρ and n for circulant correlation structure in Table 5.

Table 6 provides the comparison of the result of our algorithm without and with reheating process. From this table, we can see that, with the addition of the reheating process, the results are much better than the nonreheating process. However, the reheating process is much more time consuming than nonreheating process.

Table 7 provides the comparison of our improved simulated annealing results with the traditional simulated annealing algorithm results for 3-way polynomial regression with $n = 10$.

From Table 7, we can see that the results of our improved simulated annealing algorithm are much higher than that of the traditional simulated annealing algorithm for all of the 3 correlation structures for 3-way polynomial regression. When ρ gets larger, the ratio of the determinants of our improved simulated annealing and that of the traditional simulated annealing algorithm increase rapidly. All of the tables point out that our improved simulated annealing is much more powerful than traditional simulated annealing algorithm.

5. Discussions

This paper demonstrates that an improved simulated annealing algorithm can successfully determine highly efficient D-optimal designs for second-order polynomial regression on $[-1, 1]^2$ and third order polynomial regression on $[-1, 1]^3$ for a variety of correlated error structures and with the design size, n , not limited to a multiple of the number of regression parameters. The combination of (i) a “part-by-part” perturbation scheme, (ii) the use of a parameter that controls the size of the neighborhood for the perturbations, and (iii) increase of the threshold for accepting a perturbation each time we lower the temperature lead to designs that—while not likely globally optimal—are better than those

TABLE 1: 2-way polynomial regression with autoregressive correlation.

n	ρ	Traditional annealing determinant	Improved annealing determinant	Ratio
6	0.1	231.4	281.2	1.2152
6	0.4	578.2	751.8	1.3002
12	0.1	13582	17769	1.3083
12	0.4	31721	45108	1.4220
18	0.1	195260	272620	1.3962
18	0.4	416720	889690	2.1350

TABLE 2: 2-way polynomial regression with circulant correlation.

n	ρ	Traditional annealing determinant	Improved annealing determinant	Ratio
6	0.1	218.4	279	1.2775
6	0.4	712.5	1047	1.4695
12	0.1	12492	17815	1.4261
12	0.4	43962	65894	1.4989
18	0.1	143620	206010	1.4344
18	0.4	623820	1091400	1.7495

TABLE 3: 2-way polynomial regression with nearest neighbor correlation.

n	ρ	Traditional annealing determinant	Improved annealing determinant	Ratio
6	0.1	212.4	279.1	1.3140
6	0.4	534.2	742.5	1.3899
12	0.1	23982	32901	1.3719
12	0.4	45842	74276	1.6203
18	0.1	136922	206010	1.5046
18	0.4	639520	1175800	1.8386

TABLE 4: 2-way polynomial regression with block correlation.

n	ρ	Traditional annealing determinant	Improved annealing determinant	Ratio
12	0.1	18204	25088	1.3782
12	0.4	22912	39870	1.7401

TABLE 5: Circulant correlation structure with various ρ and n .

ρ	n				
	7	8	9	10	11
0.1	517.3	2523.2	4417.6	6738.3	16975
0.2	1261.1	3666.9	7672.5	16211	21788
0.3	1958.1	5540	16406	31880	42016
0.4	4046.7	13514	52982	61529	64205

TABLE 6: 2-way polynomial regression with $n = 12$; compare reheated simulated annealing with nonreheated simulated annealing.

Correlation structure	ρ	Nonreheated determinant	Reheated determinant	Ratio
Nearest neighbor	0.4	74276	97284	1.3098
Circular	0.4	65894	87291	1.3247
Autoregress	0.4	45234	68548	1.5154

TABLE 7: 3-way polynomial regression with $n = 10$.

Correlation structure	ρ	Traditional annealing determinant	Improved annealing determinant	Ratio
Nearest neighbor	0.1	1.8342×10^6	2.4030×10^6	1.3101
Nearest neighbor	0.4	1.1529×10^7	2.1257×10^7	1.8438
Circular	0.1	1.6215×10^6	2.2202×10^6	1.3692
Circular	0.4	1.4284×10^7	2.3343×10^7	1.6342
Autoregress	0.1	1.8128×10^6	2.3423×10^6	1.2921
Autoregress	0.4	6.2563×10^6	1.08510×10^7	1.7342

obtained by traditional simulated annealing algorithm. In particular, when the true correlation parameter is well away from 0, our improved simulated annealing algorithm has much greater relative efficiency than the traditional simulated annealing algorithm.

The SA algorithm needs only a well-defined energy function to maximize here the determinant of the information matrix. Thus, the same algorithm may be used for other design optimality criteria, for example, A- and E-optimality. In the absence of exact analytic optimal designs when errors are correlated, the SA algorithm is an attractive, easily implemented method to find highly efficient designs. Extensions to higher degree polynomial regression models are immediate, except for the likely need for longer run times and slower reduction of the temperature to allow for more effective searching over a larger design region.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Z. Zhu, D. C. Coster, and L. B. Beasley, "Properties of a covariance matrix with an application to D -optimal design," *Electronic Journal of Linear Algebra*, vol. 10, pp. 65–76, 2003.
- [2] H. Dette, J. Kunert, and A. Pepelyshev, "Exact optimal designs for weighted least squares analysis with correlated errors," *Statistica Sinica*, vol. 18, no. 1, pp. 135–154, 2008.
- [3] W. G. Muller, *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*, Springer, 2001.
- [4] L. M. Haines, "The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models," *Technometrics*, vol. 29, no. 4, pp. 439–447, 1987.
- [5] M. A. Lejeune, "Heuristic optimization of experimental designs," *European Journal of Operational Research*, vol. 147, no. 3, pp. 484–498, 2003.
- [6] B. Dimitris and N. Omid, *Robust Optimization with Simulated Annealing*, Springer Science+Business Media, LLC, 2009.
- [7] S. Abdullah, L. Golafshan, and M. Z. A. Nazri, "Re-heat simulated annealing algorithm for rough set attribute reduction," *International Journal of Physical Sciences*, vol. 6, no. 8, pp. 2083–2089, 2011.
- [8] Z. Zhu, *Application of simulated annealing to d -optimal design for polynomial regression with correlated observations [Ph.D. thesis]*, Department of Mathematics and Statistics, Utah State University, 2004.
- [9] C.-S. Cheng, "Optimal regression designs under random block-effects models," *Statistica Sinica*, vol. 5, no. 2, pp. 485–497, 1995.
- [10] J. E. Boon, "Generating exact d optimal design for polynomial models," in *Proceedings of the Spring Simulation Multiconference (SpringSim '07)*, 2007.
- [11] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50, SIAM, Philadelphia, Pa, USA, 2006.
- [12] Z. Zhu, *Optimal experimental designs with correlated observations [Ph.D. thesis]*, Department of Mathematics and Statistics, Utah State University, 2004.
- [13] J. Cadima, F. L. Calheiros, and I. P. Preto, "The eigenstructure of block-structured correlation matrices and its implications for principal component analysis," *Journal of Applied Statistics*, vol. 37, no. 3-4, pp. 577–589, 2010.
- [14] J. E. Atkins and C.-S. Cheng, "Optimal regression designs in the presence of random block effects," *Journal of Statistical Planning and Inference*, vol. 77, no. 2, pp. 321–335, 1999.