Exact D-Optimal Design of Experiments Using Simulated Annealing

University of Victoria

Benjamin Frizzell - Department of Mathematics and Statistics

Introduction

Exact D-optimal designs for various linear and nonlinear models have been obtained using simulated annealing and compared to Disciplined Convex Programming (DCP) methods provided by the optimization software known as CVX. Optimal experimental designs are very useful in practice due to their flexibility in model specification and domain constraints, as well as their ability to provide experimental designs with minimized parameter variance and experimental cost. However, exact designs are difficult to find when models are nonlinear and/or have high-dimensional parameter spaces. We propose the simulated annealing algorithm as a viable alternative to obtaining optimal designs in such cases; often simulated annealing outperforms CVX or provides a good approximation of the design in significantly reduced runtime.

Design of Experiments

Consider a linear regression model with responses y_i and sample size N:

$$y_i = \boldsymbol{\theta}^T \boldsymbol{F}(\boldsymbol{x_i}) + \epsilon_i, \qquad i = 1, \dots N$$

Where $\boldsymbol{\theta} = (\theta_1, \dots \theta_P)$ is the parameter vector to be estimated by the experimenter. In the context of experimental design, we denote $\boldsymbol{X} = (\boldsymbol{x_1}, \dots \boldsymbol{x_k})$ as a vector of *design points* $\boldsymbol{x_i} \in \mathbb{R}^d$, and $F(\boldsymbol{x_i}) = \left(f_1(\boldsymbol{x_i}) \dots f_p(\boldsymbol{x_i})\right)^T$ as the regression functions of the model. Furthermore, we will denote $\chi \in \mathbb{R}^d$ as the *design space*, the set of all possible design points for the model.

In a random experiment, the experimenter may choose a set of distinct design points and measure the response some number of times at each. We summarize this by defining an *exact design* ξ as:

$$\xi = \begin{pmatrix} \boldsymbol{x_1} & \dots & \boldsymbol{x_m} \\ \frac{n_1}{N} & \dots & \frac{n_m}{N} \end{pmatrix}$$

Where n_i denotes the number of times the design point x_i appears in the experiment.

Optimal Design

In practice, an experimenter will not have complete freedom in designing an experiment due to budget constraints, practical, or ethical issues. Therefore, it is prudent to optimize an experiment by strategically selecting a configuration of design points under such constraints so that parameter variance is minimized.

To do so, consider the **Fischer Information Matrix,** which quantifies the 'information' that can be obtained about the true value of $\boldsymbol{\theta}$ given $\boldsymbol{X}$**.** For the linear model described above, this is given as:

$$M(\xi) = \sum_{i=1}^{m} \frac{n_i}{N} \boldsymbol{F}(\boldsymbol{x_i}) \boldsymbol{F}^{\mathrm{T}}(\boldsymbol{x_i})$$

For a linear model, the information matrix is not dependent on the parameter values themselves. This is not true in general, and a pilot study to initialize $\boldsymbol{\theta}$ may be required. For example, a logistic regression model has:

$$M(\xi, \theta) = \sum_{i=1}^{m} \frac{n_i}{N} \frac{\exp(\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{x_i}))}{\left(1 + \exp(\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{x_i}))\right)^2} \boldsymbol{F}(\boldsymbol{x_i}) \boldsymbol{F}^{\mathrm{T}}(\boldsymbol{x_i})$$

Next, we must select an *optimality criterion* as an objective function to compute the optimal design. Many such criteria exist, however one of the most widely used is *D-optimality*:

$$\phi_D(\xi) = -\log(\det M(\xi))$$

Finally, we compute the *exact optimal design* ξ^* by solving the constrained optimization problem:

$$\xi^* = \underset{\xi}{\operatorname{argmin}}\, \phi_D(\xi), \qquad \text{subject to } \sum_{i=1}^{m} n_i = N$$

Note that minimizing $\phi_D(\xi)$ acts as a proxy for **minimizing the variance of** $\boldsymbol{\theta}$. Additional constraints may be present, depending on the context of the experiment.

Computing Optimal Designs

Determining optimal experimental designs so can become particularly challenging with:

- Usage of GLMs or NLMs
- High-dimensional design spaces which must be approximated as grid points

In addition, the discrete nature of exact designs can introduce further complications. Therefore, many opt for a simpler approach by determining an *approximate optimal design:*

$$\xi = \begin{pmatrix} \boldsymbol{x_1} & \dots & \boldsymbol{x_m} \\ \omega_1 & \dots & \omega_m \end{pmatrix}, \qquad \sum_{i=1}^{m} \omega_i = 1$$

Where design points are instead given real-valued weights ω_i which are independent of the prescribed sample size.

Additionally, a helpful scale-invariance property between the design space and the information matrix exists[2]. Consider a design space scaled by a matrix $\boldsymbol{V} = \mathrm{diag}(s1, \dots s_d), s_i > 0\ (j = 1, \dots d)$:

$$\chi_V = \{\boldsymbol{V x_1}, \dots, \boldsymbol{V x_n}\}$$

Theorem 1. *Suppose* $\xi^* = \{(\boldsymbol{x_i^*}, \omega_i^*), i = 1, \dots, m\}$ *is a D-optimal design for a model with true parameter* $\boldsymbol{\theta}$ *on design space* $\boldsymbol{\chi} = \{\boldsymbol{x_1}, \dots, \boldsymbol{x_N}\}$*. If there exists a non-singular matrix* $\boldsymbol{T}$ *(independent of* $\boldsymbol{\theta}$ *and* $\boldsymbol{x}$*), and parameter value* $\boldsymbol{\theta^V}$ *such that*

$$M(\xi, \boldsymbol{\theta^V}) = \boldsymbol{T} M(\xi, \theta) \boldsymbol{T}^{\mathrm{T}}, \qquad \forall \boldsymbol{x} \in \boldsymbol{\chi}$$

Then $\tilde{\xi}^* = \{(\boldsymbol{V x_i^*}, \omega_i^*), i = 1, \dots, m\}$ *is a D-optimal design for the model with parameter* $\boldsymbol{\theta^V}$ *on the scaled design space* χ_V*.*

We can use this property to determine an optimal design on a 'small' design space and scale our solution to the design space of interest in the experiment.

Many algorithms to compute optimal designs exist. *CVX-based numerical algorithms* are some of the most common methods, due to their accessibility and flexibility. CVX makes use of Newton's method and interior point methods, which have been shown to be viable methods of computing some optimal designs over traditional methods but can be slow or fail to provide solutions in the cases described above[2].

Simulated Annealing

As an alternative to CVX, we propose the usage of *simulated annealing*, which has been shown to be effective for optimal designs of various models in the past[3]. With this approach, we allow for exploration of the design space by **randomly substituting points in the current design** with **new points in their neighbourhood** iteratively. Potentially sub-optimal solutions can be accepted before the algorithm terminates, to prevent the algorithm from 'settling' into a locally minimal solution. Important parameters and aspects of the algorithm include:

Temperature ($\boldsymbol{T}$): This controls the probability that a new design is accepted over the current optimal design, if the new design is sub-optimal. A higher temperature corresponds to a greater probability that the new design is accepted. After some specified number of iterations $\boldsymbol{n_T}$, temperature decays according to a parameter $\boldsymbol{\alpha < 1}$, to allow for greater freedom of exploration at earlier iterations, and tighter constraints at the end to focus on optimization. The algorithm continues until some minimum temperature $\boldsymbol{T_{min}}$ is reached.

$\boldsymbol{C_0}$: The maximum number of design points that may be replaced. For each iteration, 1 to C_0 points will be replaced randomly.

Search Radius ($\boldsymbol{\delta}$): This controls the radius of the neighbourhood in which new design points may be selected; if a design point $\boldsymbol{x_i}$ is removed from the design, it may be randomly replaced by any design point of Euclidean distance δ from $\boldsymbol{x_i}$.

Acceptance Probability ($\boldsymbol{P}$): The probability that a new design is accepted after the i^{th} iteration is usually given by an exponential function of the following form:

$$P(\xi_i, \tilde{\xi}, T_i) = \begin{cases} \exp -\dfrac{(\phi_D(\xi_i) - \phi_D(\tilde{\xi}))}{T_i}, & \phi_D(\xi_i) > \phi_D(\tilde{\xi}) \\ 1, & \phi_D(\xi_i) \leq \phi_D(\tilde{\xi}) \end{cases}$$

Where T_i and ξ_i are the temperature and new design at the i^{th} iteration, and $\tilde{\xi}$ is the current 'candidate' for the optimal design.

Algorithm

- Initialize $T, n_T, \alpha, C_0, \delta, T_{min}, N, \xi_0$
- Calculate $M(\xi_0), \phi_D(\xi_0)$

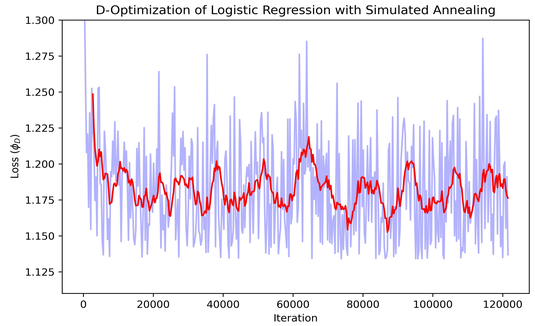
$\xi^* \leftarrow \xi_0$

while $(T > T_{min})$ do:
  for $j = 1$ to n_T do:
    - Remove 1 to C_0 design points from $\tilde{\xi}$
    - Replace design points randomly with points in neighbourhood of previous
    - Store new design in ξ_{new}
    - Calculate $M(\xi_{new}), \phi_D(\xi_{new})$

    if $P(\xi_{new}, \xi^*, T) >$ Random(0,1) then:
      $\xi^* \leftarrow \xi_{new}$

  $T \leftarrow \alpha T$

return ξ^* as the optimal design



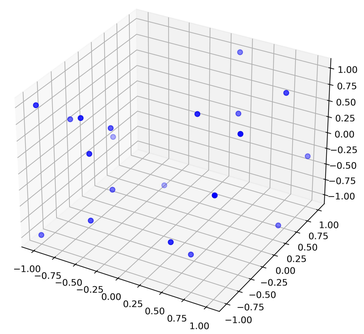


Figure 3-4: Optimization process (top) and D-optimal design (bottom) for a saturated logistic regression model with three regressors, obtained using simulated annealing.

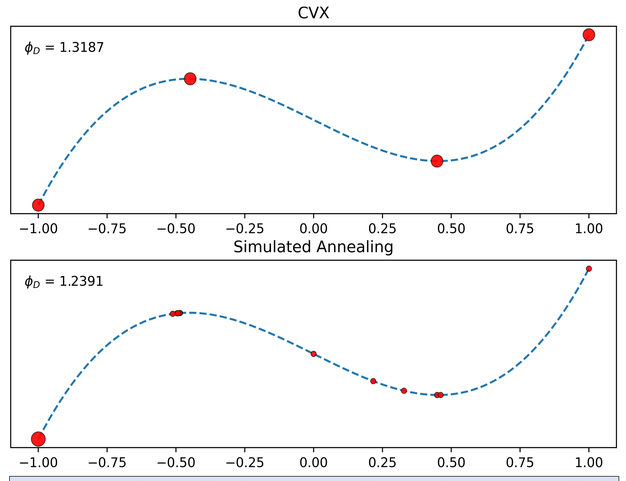


Figure 1: D-Optimal Designs of a cubic regression model over the design space interval [-1,1], optimized using CVX and the simulated annealing algorithm. Larger points correspond to a greater number of samples. ($N = 20$)

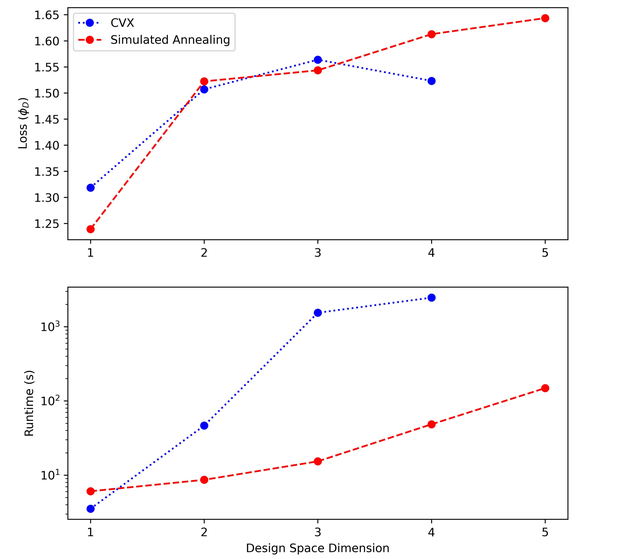


Figure 2: Loss function and runtime are plotted for simulated annealing and CVX algorithms, applied to D-optimal designs of cubic regression models with varying dimension/number of regressors.

Results and Future Work

The simulated annealing algorithm can determine D-optimal designs in significantly reduced runtime with marginal differences in loss in comparison to CVX. Future work may include:

- Determining optimal parameter configurations: currently it is known that initial temperatures on the order of $\mathbf{10^{-5}}$ achieve the best results with this implementation.
- Testing the simulated annealing algorithm with other optimality criterion, and evaluating success
- Compare to other randomized algorithms (for example, genetic algorithms have shown some promise in the application of exact D-optimal designs[4])

References

1.F., B. M. P., & Wong, W. K. (2009). *An introduction to optimal designs for Social and Biomedical Research*. Wiley.
2.Wong, W.K. and Zhou, J. (2019), CVX-based algorithms for constructing various optimal regression designs. Can J Statistics, 47: 374-391. https://doi.org/10.1002/cjs.11499
3.Meyer, R. K., & Nachtsheim, C. J. (1988). Constructing Exact D-Optimal Experimental Designs by Simulated Annealing. *American Journal of Mathematical and Management Sciences*, *8*(3–4), 329–359. https://doi.org/10.1080/01966324.1988.10737244
4.Montepiedra, G. (1998). Application of genetic algorithms to the construction of exact D-optimal designs. *Journal of Applied Statistics*, *25*(6), 817–826. https://doi.org/10.1080/02664769822800