A- and D-optimal designs based on the second-order least squares estimator

Julie Zhou

Department of Mathematics and Statistics
University of Victoria

Joint work with
Lucy Gao (University of Washington, PhD student)
Yue Yin (University of Victoria, PhD, 2017)

2017

Overview

Consider a regression model

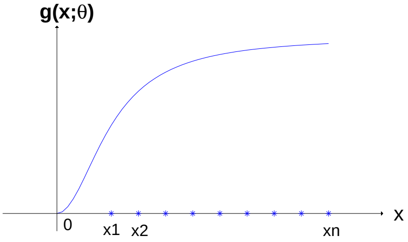
$$y_i = g(\mathbf{x}_i; \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \cdots, n, \tag{1}$$

where

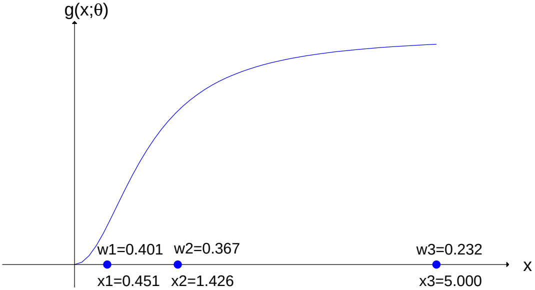
- y is a response variable, $\mathbf{x} \in R^p$ is a vector of independent (design) variables,
- $g(\mathbf{x}_i; \boldsymbol{\theta})$ can be linear or nonlinear function of $\boldsymbol{\theta} \in R^q$,
- the error ϵ_i's are i.i.d with mean 0 and variance σ^2.

Design problem: how do we choose the "optimal" design points $x_1, \cdots, x_n$ from a design space to observe y?

e.g. Emax model: $y_i = \frac{\alpha x_i^{\beta_2}}{\beta_1 + x_i^{\beta_2}} + \epsilon_i, \quad \boldsymbol{\theta} = (\alpha, \beta_1, \beta_2)^\top$



A-optimal design when $\boldsymbol{\theta} = (1, 1, 2)^\top$



Optimal designs depend on:

- regression model and its assumptions
- design space
- regression estimator
- optimality criterion

- Ordinary least squares estimator (OLSE): minimizing $\sum_{i=1}^{n}(y_i - g(\mathbf{x}_i; \boldsymbol{\theta}))^2$.
- OLSE: BLUE
- Second-order least squares estimator (SLSE) in Wang and Leblanc (2008) is more efficient than OLSE, when the error distribution is asymmetric, i.e. $E[\epsilon^3|\mathbf{x}] \neq 0$.

The SLSE $\hat{\gamma}_{SLSE}$ of $\boldsymbol{\gamma} = (\boldsymbol{\theta}^\top, \sigma^2)^\top$ minimizes

$$Q(\gamma) = \sum_{i=1}^{n} \rho_i^\top(\gamma)\tilde{\mathbf{W}}_i \rho_i(\gamma),$$

where vector $\rho_i(\gamma) = (y_i - g(\mathbf{x}_i; \boldsymbol{\theta}), y_i^2 - g^2(\mathbf{x}_i; \boldsymbol{\theta}) - \sigma^2)^\top$ and $\tilde{\mathbf{W}}_i = \tilde{\mathbf{W}}(\mathbf{x}_i)$ is a 2×2 positive semidefinite matrix that may depend on $\mathbf{x}_i$.

Suppose $\boldsymbol{\theta}_{\mathbf{0}}$ and σ_0^2 are the true values of $\boldsymbol{\theta}$ and σ^2, respectively. Let $\mu_3 = E(\epsilon_1^3 \mid \mathbf{x})$, $\mu_4 = E(\epsilon_1^4 \mid \mathbf{x})$, and $t = \mu_3^2/(\sigma_0^2(\mu_4 - \sigma_0^4))$. Define

$$\mathbf{g}_1 = E\left[\frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathbf{0}}}\right],$$

$$\mathbf{G}_2 = E\left[\frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathbf{0}}}\right].$$

The expectation is taken with respect to the distribution of $\mathbf{x}$, $\xi(\mathbf{x})$.

The asymptotic covariance matrix of $\hat{\gamma}_{SLSE}$ is

$$Cov(\hat{\gamma}_{SLSE}) = \begin{pmatrix} Cov(\hat{\boldsymbol{\theta}}_{SLSE}) & \frac{\mu_3}{\mu_4-\sigma_0^4} V(\hat{\sigma}^2_{SLSE})\mathbf{G}_2^{-1}\mathbf{g}_1 \\ \frac{\mu_3}{\mu_4-\sigma_0^4} V(\hat{\sigma}^2_{SLSE})\mathbf{g}_1^\top \mathbf{G}_2^{-1} & V(\hat{\sigma}^2_{SLSE}) \end{pmatrix},$$

where

$$Cov(\hat{\boldsymbol{\theta}}_{SLSE}) = (1-t)\,\sigma_0^2 \left(\mathbf{G}_2 - t\mathbf{g}_1\mathbf{g}_1^\top\right)^{-1},$$

$$V(\hat{\sigma}^2_{SLSE}) = \frac{(\mu_4-\sigma_0^4)(1-t)}{1-t\mathbf{g}_1^\top \mathbf{G}_2^{-1}\mathbf{g}_1},$$

The asymptotic covariance matrix of the OLSE, $\hat{\gamma}_{OLSE} = (\hat{\boldsymbol{\theta}}^\top_{OLSE}, \hat{\sigma}^2_{OLSE})^\top$, is

$$Cov(\hat{\gamma}_{OLSE}) = \begin{pmatrix} Cov(\hat{\boldsymbol{\theta}}_{OLSE}) & \mu_3 \mathbf{G}_2^{-1}\mathbf{g}_1 \\ \mu_3 \mathbf{g}_1^\top \mathbf{G}_2^{-1} & V(\hat{\sigma}^2_{OLSE}) \end{pmatrix}$$
$$= \begin{pmatrix} \sigma_0^2 \mathbf{G}_2^{-1} & \mu_3 \mathbf{G}_2^{-1}\mathbf{g}_1 \\ \mu_3 \mathbf{g}_1^\top \mathbf{G}_2^{-1} & \mu_4 - \sigma_0^4 \end{pmatrix}.$$

- We have $0 \leq t < 1$ for any error distribution.
- If the error distribution is symmetric, then $\mu_3 = 0$, $t = 0$, and the covariance matrices for SLSE and OLSE are the same.
- For asymmetric errors, $Cov(\hat{\gamma}_{OLSE}) - Cov(\hat{\gamma}_{SLSE}) \succeq 0$ from Wang and Leblanc (2008), so the SLSE is more efficient than the OLSE.

Design criteria based on the SLSE

- A-optimal design criterion: $\min_{\xi} \text{trace}(Cov(\hat{\theta}_{SLSE}))$
- D-optimal design criterion: $\min_{\xi} \det(Cov(\hat{\theta}_{SLSE}))$

Gao and Zhou (2014): proposed these criteria
Bose and Mukerjee (2015): more results for binary design points
Yin and Zhou (2016), Gao and Zhou (2017): more results for A- and D-optimal designs

On a discrete design space $S_N = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N\}$, let

$$\xi = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \\ w_1 & w_2 & \cdots & w_N \end{pmatrix}.$$

Define $\mathbf{f}(\mathbf{x};\boldsymbol{\theta}) = \partial g(\mathbf{x};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and write $\mathbf{g}_1$ and $\mathbf{G}_2$ as

$$\mathbf{g}_1(\mathbf{w}) = \mathbf{g}_1(\mathbf{w};\boldsymbol{\theta}_0) = \sum_{i=1}^{N} w_i \mathbf{f}(\mathbf{u}_i;\boldsymbol{\theta}_0),$$

$$\mathbf{G}_2(\mathbf{w}) = \mathbf{G}_2(\mathbf{w};\boldsymbol{\theta}_0) = \sum_{i=1}^{N} w_i \; \mathbf{f}(\mathbf{u}_i;\boldsymbol{\theta}_0)\mathbf{f}^\top(\mathbf{u}_i;\boldsymbol{\theta}_0),$$

where weight vector $\mathbf{w} = (w_1, \cdots, w_N)^\top$.

Let $\mathbf{A}(\mathbf{w}) = \mathbf{A}(\mathbf{w}; \theta_0) = \mathbf{G}_2(\mathbf{w}) - t\mathbf{g}_1(\mathbf{w})\mathbf{g}_1^\top(\mathbf{w})$,

then A- and D-optimal designs minimize loss functions

$$\phi_1(\mathbf{w}) = \text{tr}\left((\mathbf{A}(\mathbf{w}))^{-1}\right) \quad \text{and} \quad \phi_2(\mathbf{w}) = \det\left((\mathbf{A}(\mathbf{w}))^{-1}\right)$$

over $\mathbf{w}$, respectively.

If $\mathbf{A}(\mathbf{w})$ is singular, $\phi_1(\mathbf{w})$ and $\phi_2(\mathbf{w})$ are defined to be $+\infty$.

The A- and D-optimal designs are denoted by $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$, respectively.

Define

$$\mathbf{B}(\mathbf{w}) = \begin{pmatrix} 1 & \sqrt{t}\ \mathbf{g}_1^\top(\mathbf{w}) \\ \sqrt{t}\ \mathbf{g}_1(\mathbf{w}) & \mathbf{G}_2(\mathbf{w}) \end{pmatrix}$$

The D-optimal design based on the SLSE minimizes $1/\det(\mathbf{B}(\mathbf{w}))$.

It is clear that both $-log(\det(\mathbf{B}(\mathbf{w})))$ and $-(\det(\mathbf{B}(\mathbf{w})))^{1/(q+1)}$ are convex functions of $\mathbf{w}$.

Characterization of the A-optimal design problem:

Theorem

If $\mathbf{G}_2(\mathbf{w})$ *is nonsingular, then*
$\phi_1(\mathbf{w}) = tr\left((\mathbf{A}(\mathbf{w}))^{-1}\right) = tr\left(\mathbf{C}(\mathbf{B}(\mathbf{w}))^{-1}\right)$, *where* $\mathbf{C} = 0 \oplus \mathbf{I}_q$ *is a* $(q+1) \times (q+1)$ *matrix.*

– Main result to develop effective and efficient algorithms for finding A-optimal designs under SLSE.

Invariance properties of D-optimal designs:
symmetry of D-optimal designs
shift invariance
scale invariance

Invariance properties of A-optimal designs:
symmetry of A-optimal designs

Since $w_N = 1 - \sum_{i=1}^{N-1} w_i$,

define $\tilde{\mathbf{w}} = \left(w_1, w_2, \cdots, w_{N-1}, 1 - \sum_{i=1}^{N-1} w_i \right)^\top$.

Let $\mathbf{D}(\tilde{\mathbf{w}}) = \text{diag}\left(w_1, w_2, \cdots, w_{N-1}, 1 - \sum_{i=1}^{N-1} w_i \right)$ be a diagonal matrix.

Thus, the A- and D-optimal design problems become, respectively,

$$\begin{cases} \min_{\tilde{\mathbf{w}}} \phi_1(\tilde{\mathbf{w}}), \\ \text{subject to: } \ \mathbf{D}(\tilde{\mathbf{w}}) \succeq 0, \end{cases} \tag{2}$$

$$\begin{cases} \min_{\tilde{\mathbf{w}}} \log(\phi_2(\tilde{\mathbf{w}})), \\ \text{subject to: } \ \mathbf{D}(\tilde{\mathbf{w}}) \succeq 0. \end{cases} \tag{3}$$

Both $\phi_1(\tilde{\mathbf{w}})$ and $\log(\phi_2(\tilde{\mathbf{w}}))$ (or $(\phi_2(\tilde{\mathbf{w}}))^{1/q}$) are convex functions of $\tilde{\mathbf{w}}$.

In order to use CVX program for finding D-optimal designs, we need to use matrix $\mathbf{B}(\mathbf{w})$ in $\phi_2(\tilde{\mathbf{w}})$,

$$\mathbf{B}(\mathbf{w}) = \begin{pmatrix} 1 & \sqrt{t}\ \mathbf{g}_1^\top(\mathbf{w}) \\ \sqrt{t}\ \mathbf{g}_1(\mathbf{w}) & \mathbf{G}_2(\mathbf{w}) \end{pmatrix}$$

Matlab code :

```
cvx_begin
    variable w(N);
    expression B(q+1,q+1); %target function
⋮
    minimize (-det_rootn(B))
    subject to
        sum(w)==1;
        w>=0;
cvx_end
```

To compute A-optimal designs, we use SeDuMi program in Matlab.

We need to transform the design problem into a semi-definite programming (SDP) problem:

– objective function: linear
– constraint: linear matrix being positive semi-definite

Boyd, S. and Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press, New York.

Let $\mathbf{e}_i$ be the ith unit vector in R^{q+1}, $i = 1, \cdots, q+1$, $\mathbf{v} = (v_2, \cdots, v_{q+1})^\top$, and

$$\mathbf{B}_i = \begin{pmatrix} \mathbf{B}(\tilde{\mathbf{w}}) & \mathbf{e}_i \\ \mathbf{e}_i^\top & v_i \end{pmatrix}, \quad \text{for} \quad i = 2, \cdots, q+1,$$
$$\mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v}) = \mathbf{B}_2 \oplus \cdots \oplus \mathbf{B}_{q+1} \oplus \mathbf{D}(\tilde{\mathbf{w}}). \qquad (4)$$

Since $\mathbf{B}(\tilde{\mathbf{w}})$ and $\mathbf{D}(\tilde{\mathbf{w}})$ are linear matrices in $\tilde{\mathbf{w}}$, $\mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v})$ is a linear matrix in $\tilde{\mathbf{w}}$ and $\mathbf{v}$. Then $\xi^A(\mathbf{x})$ can be solved through

$$\begin{cases} \min_{\tilde{\mathbf{w}}, \mathbf{v}} \sum_{i=2}^{q+1} v_i, \\ \text{subject to: } \ \mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v}) \succeq 0, \end{cases} \qquad (5)$$

Example 1: $y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 + \epsilon$

The design spaces:

$S_{9,1} =$
$\{(1,0),(-1,0),(0,1),(0,-1),(1,1),(-1,1),(1,-1),(-1,-1),(0,0)\},$

$S_{9,2} =$
$\{(\sqrt{2},0),(-\sqrt{2},0),(0,\sqrt{2}),(0,-\sqrt{2}),(1,1),(-1,1),(1,-1),$
$(-1,-1),(0,0)\}.$

Table: A- and D-optimal weights, w_1^A, w_5^A, w_9^A, w_1^D, w_5^D, w_9^D

t	w_1^A	w_5^A	w_9^A	w_1^D	w_5^D	w_9^D
Design space $S_{9,1}$						
0	0.131	0.119	0.000	0.071	0.179	0.000
0.3	0.130	0.120	0.000	0.072	0.178	0.000
0.5	0.128	0.122	0.000	0.074	0.176	0.000
0.9	0.118	0.121	0.044	0.088	0.162	0.000
Design space $S_{9,2}$						
0	0.104	0.146	0.000	0.125	0.125	0.000
0.3	0.104	0.146	0.000	0.125	0.125	0.000
0.5	0.104	0.146	0.000	0.125	0.125	0.000
0.9	0.088	0.125	0.148	0.116	0.116	0.072

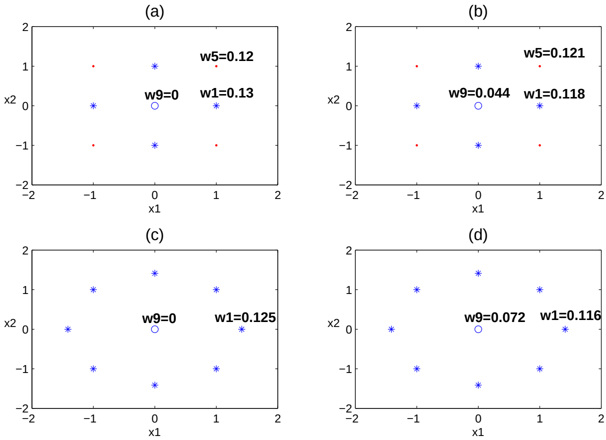


Figure: A- and D-Optimal designs for Example 1

Optimal design results for Example 1:

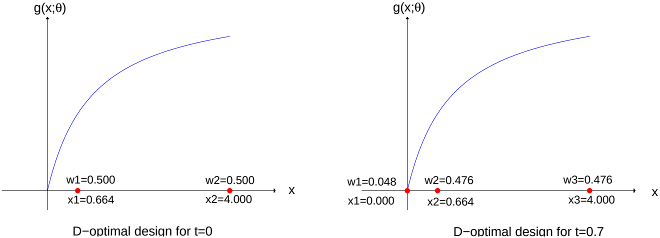
(a) A-optimal design on design space $S_{9,1}$ for $t = 0.3$,
(b) A-optimal design on design space $S_{9,1}$ for $t = 0.9$,
(c) D-optimal design on design space $S_{9,2}$ for $t = 0.3$,
(d) D-optimal design on design space $S_{9,2}$ for $t = 0.9$.

Example 2. Michaelis-Menten model:
$y = \alpha x/(\beta + x) + \epsilon,\ 0 \le x \le 4,$

Table: A- and D-optimal design points and their weights (in parentheses) for the Michaelis-Menten model with $\alpha = 1$ and $\beta = 1$

$N = 501$				
$t = 0$	ξ^A:	0.504 (0.670)	4.000 (0.330)	
	ξ^D:	0.664 (0.500)	4.000 (0.500)	
$t = 0.3$	ξ^A:	0.536 (0.662)	4.000 (0.338)	
	ξ^D:	0.664 (0.500)	4.000 (0.500)	
$t = 0.7$	ξ^A:	0.632 (0.642)	4.000 (0.358)	
	ξ^D:	0.000 (0.048)	0.664 (0.476)	4.000 (0.476)
$t = 0.9$	ξ^A:	0.000 (0.158)	0.664 (0.536)	4.000 (0.306)
	ξ^D:	0.000 (0.260)	0.664 (0.370)	4.000 (0.370)

Michaelis-Menten model with $\alpha = 1$ and $\beta = 1$



Example 3. Emax model: $y_i = \frac{\alpha x_i^{\beta_2}}{\beta_1 + x_i^{\beta_2}} + \epsilon_i, \quad \boldsymbol{\theta} = (\alpha, \beta_1, \beta_2)^\top$

Table: A-optimal and D-optimal design points and their weights in the brackets for the Emax model with various values of t

$N = 201$					
$t = 0$	ξ^A:	0.451 (0.401)	1.426 (0.367)	5.000 (0.232)	
	ξ^D:	0.551 (0.333)	1.476 (0.334)	5.000 (0.333)	
$t = 0.3$	ξ^A:	0.451 (0.401)	1.426 (0.367)	5.000 (0.232)	
	ξ^D:	0.551 (0.333)	1.476 (0.334)	5.000 (0.333)	
$t = 0.7$	ξ^A:	0.501 (0.376)	1.476 (0.381)	5.000 (0.243)	
	ξ^D:	0.551 (0.333)	1.476 (0.334)	5.000 (0.333)	
$t = 0.9$	ξ^A:	0.001 (0.103)	0.531 (0.328)	1.501 (0.348)	5.000 (0.221)
	ξ^D:	0.001 (0.166)	0.551 (0.278)	1.476 (0.278)	5.000 (0.278)

Number of support points in optimal designs:

For $t = 0$, optimal designs under SLSE and OLSE are the same.

For small t, optimal designs under SLSE and OLSE have the same number of support points from numerical results.

For large t, optimal designs under SLSE usually have one more support points than those under OLSE from numerical results.

We are trying to prove some theoretical results on the number of support points.

CVX and SeDuMi programs in Matlab are very powerful to solve convex optimization problems with constraints.

Many optimal design problems are convex optimization problems. We have successfully applied CVX and SeDuMi programs for finding optimal regression designs for various models and optimality criteria.

– linear/nonlinear regression models, generalized linear models, multi-response regression models

– OLSE, SLSE, MLE, WLSE

– A-, A_s-, c-, D-, D_s-, I-, L-optimality criteria

Wong, W.K., Y. Yin and J. Zhou (2017). "Use SeDuMi to find various optimal designs for regression models", Statistical Papers, to appear.

Numerical algorithms for finding minimax designs?

Objective functions are not convex.

References:

1. Bose, M. and Mukerjee, R. (2015). Optimal design measures under asymmetric errors, with application to binary design points. *Journal of Statistical Planning and Inference*. 159, 28-36.
2. Gao, L.L. and Zhou, J. (2014). New optimal design criteria for regression models with asymmetric errors. *Journal of Statistical Planning and Inference*. 149, 140-151.
3. Wang, L. and Leblanc, A. (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics*. 60, 883-900.
4. Yin, Y. and J. Zhou (2016). Optimal designs for regression models using the second-order least squares estimator. *Statistica Sinica*, to appear.

5. Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, New York.
6. Grant, M.C. and Boyd, S.P. (2013). *The CVX Users Guide*. Release 2.0 (beta), CVX Research, Inc. (http://cvxr.com/cvx/doc/CVX.pdf, October 14, 2013.)