Geometric Methods for Singular C-optimal Designs

Tech Report #629

By Shenghua Fan and Kathryn Chaloner
University of Minnesota

Geometric Methods for Singular C-optimal Designs

By Shenghua Fan and Kathryn Chaloner
University of Minnesota

Abstract

A geometric method for finding c-optimal designs for the univariate linear model was introduced by Elfving (1952). Elfving's Theorem is sometimes difficult to apply. In this case, if a candidate design has a non-singular information matrix, the General Equivalence Theorem can be used to examine straightforwardly whether or not this design is optimal. If the candidate c-optimal design has a singular information matrix its optimality can be verified by a different version of the General Equivalence Theorem which however requires the use of an unknown generalized inverse of the information matrix. This paper provides a geometric approach for finding a generalized inverse which can be used.

1 Introduction

The goal of an experiment is often to estimate a certain function of the parameters. An optimal design is one which minimizes the standard error of the estimate of this function.

Given a model and a parameter vector θ, let $I(\theta, x)$ be the Fisher information matrix of a single observation at control vector x. Let n be the sample size. If a design η puts n_i observations at $x = x_i$ then the information matrix of design η is $\sum_i n_i I(\theta, x_i)$. Define $\eta_i = n_i/n$ then the information matrix can be rewritten as $nM(\theta, \eta)$ where $M(\theta, \eta) = \sum_i \eta_i I(\theta, x_i)$. For mathematical convenience, from now on, a design is a probability measure on the domain of x, denoted by $\mathcal{X}$, and $M(\theta, \eta)$ is the information matrix for design η.

Let $\widehat{\theta}$ be the maximum likelihood estimate (MLE) of θ. Let $c^T\theta$ be the quantity of interest. Under regularity conditions, the asymptotic variance of $c^T\widehat{\theta}$ is $c^T M(\theta, \eta)^{-1} c$. Denote $M(\theta, \eta)$ as M. The following could be used as a definition of c-optimality:

$$\phi_c(M) = \begin{cases} -c^T M^{-1} c & \text{if } M \text{ is nonsingular} \\ -\infty & \text{otherwise} \end{cases}$$

A design η is c-optimal if it maximizes ϕ_c over all possible designs.

This criterion is, however, not continuous so the maximum might not exist. An alternative criterion was therefore introduced, see Silvey (1980, Chapter 5[8]).

Definition. $c^T\theta$ is estimable for an information matrix M if there exists a vector A such that $c = MA$.

The alternative criterion is defined as:

Definition. (c-optimality) Let $c^T\theta$ be the quantity of interest then the criterion of c-optimality is

$$\phi_c(M) = \begin{cases} -c^T M^{-1} c & \text{if } M \text{ is nonsingular} \\ -c^T M^{-} c & \text{if } M \text{ is singular but allows estimation of } c^T\theta \\ -\infty & \text{otherwise} \end{cases}$$

A design η is c-optimal if it maximizes ϕ_c over all possible designs.

This criterion is now concave and continuous so the maximum exists and any local maximum is the global maximum.

Consider an univariate response y and a linear model $y_i = \theta^T x_i + e_i$, $i = 1, 2, ..., n$ where $e_i \sim N(0, \sigma^2), i = 1, 2, ..., n$ are independent and identically distributed. Elfving (1952) gave a geometric method to find all c-optimal designs. Let $\mathcal{X}$ be the domain of all possible predictor vectors x and denote $-\mathcal{X}$ to be the set of points of $-x$.

Theorem. *(Elfving's Theorem) Let $c^T\theta$ be the quantity of interest. Extend the ray from the origin to c, the location z where the ray penetrates the convex hull of $\mathcal{X} \cup -\mathcal{X}$. Then if $z = \sum_i \eta_i x_i$ where $\eta_i > 0, \sum_i \eta_i = 1$ and $x_i \in \mathcal{X}$ or $-\mathcal{X}$ then the c-optimal design puts proportion η_i of observations at x_i or $-x_i$, whichever belongs to $\mathcal{X}$.*

When Elfving's Theorem is too difficult to apply (the convex hull of $\mathcal{X} \cup -\mathcal{X}$ can be very complicated) a candidate c-optimal design can be examined by a version of General Equivalence Theorem.

Definition. The directional derivative of criterion ϕ at matrix M_1 in the direction of matrix M_2 is

$$F_\phi(M_1, M_2) = \lim_{\epsilon \to 0+} \frac{1}{\epsilon}[\phi\{(1-\epsilon)M_1 + \epsilon M_2\} - \phi\{M_1\}]$$

Details are in Silvey (1980, Chapter 3[8]).

Theorem. *(General Equivalence Theorem) For a fixed θ, if a criterion ϕ is concave and differentiable at $M(\theta, \eta)$ the following conditions on a design measure η are equivalent :*

1. *η is ϕ-optimal,*
2. *$\sup_{x \in \mathcal{X}} F_\phi(M(\theta, \eta), I(\theta, x)) = 0$, where $\mathcal{X}$ is the domain of x.*

If the value of criterion ϕ for some singular design is finite an optimal design may be singular and a different version of the theorem can be used.

Theorem. *(General Equivalence Theorem for singular optimal designs) Suppose $\theta \in \mathbb{R}^p$ and criterion ϕ is concave and differentiable at $M(\theta, \eta)$. Let the design measure η be singular, that is, the rank of $M(\theta, \eta)$ is r, less than p. A sufficient condition that a design measure η is (locally) ϕ-optimal is that there exists a $p \times (p - r)$ of rank $p - r$ matrix H such that $M(\theta, \eta) + HH^T$ is nonsingular and $\sup_{x \in \mathcal{X}} F_\phi(M(\theta, \eta) + HH^T, I(\theta, x)) = 0$, where $\mathcal{X}$ is the domain of x.*

A much more difficult question is whether the sufficient condition is also necessary. Pukelsheim (1981) proved that this condition is necessary for the criterion of c-optimality. Silvey (1978) identified an open problem of how to determine an H when one exists (Silvey, 1980, Chapter 3[8]). Searle (1971, p.22) showed that $(M(\theta, \eta) + HH^T)^{-1}$ is a generalized inverse (g-inverse) of $M(\theta, \eta)$. Not all g-inverses of $M(\theta, \eta)$ which can be written in forms of $(M(\theta, \eta) + HH^T)^{-1}$ can be used in the above theorem to verify whether the design η is optimal. This paper provides a geometric approach to find a suitable matrix H for c-optimality.

2 Geometric View of the General Equivalence Theorem for Singular C-optimal Designs

Let S be a convex set in $\mathbb{R}^p$ and x belong to S. A separating hyperplane of S at x is a $p - 1$ dimension hyperplane containing x such that S is all in one side of this plane.
A tangent hyperplane of S at x is a separating hyperplane of S at x.

Lemma 2.1. *In $\mathbb{R}^p$ consider a nonzero vector a and a $p \times q$ matrix $L = [l_1, l_2, ..., l_q]$ of rank $p-1$ where $l_i, i = 1, 2, ..., q$ are its column vectors. The vector a and $\{l_1, l_2, ..., l_q\}$ are linearly independent if and only if $(aa^T + LL^T)$ is nonsingular, and the hyperplane $a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$ is spanned by $\{l_1, l_2, ..., l_q\}$ and goes through the point a.*

Proof. See Appendix A. □

Remark. The origin is an interior point of the convex hull of $\mathcal{X} \cup -\mathcal{X}$.

Let a be λc, the boundary point of the convex hull of $\mathcal{X} \cup -\mathcal{X}$ reached by extending the ray from the origin to c. Denote a separating hyperplane of the convex hull of $\mathcal{X} \cup -\mathcal{X}$ at boundary point x as E_x.

Lemma 2.2. *Let a be a boundary point of the convex hull of $\mathcal{X} \cup -\mathcal{X}$. If $a = \sum_i \alpha_i x_i$, where $\alpha_i > 0$, $x_i \in$ the convex hull of $\mathcal{X} \cup -\mathcal{X}$ for $i = 1, 2, ..., k$ and $\sum_i \alpha_i = 1$ then*

1. *Every separating hyperplane at a, E_a, contains all x_i for $i = 1, 2, ..., k$.*
2. *The set S_a, separating hyperplanes of the convex hull of $\mathcal{X} \cup -\mathcal{X}$ at point a, is identical to the set S_x, separating hyperplanes of the convex hull of $\mathcal{X} \cup -\mathcal{X}$ containing all x_i for $i = 1, 2, ..., k$.*
3. *x_i is linearly independent of $\{x_i - x_j, i \neq j\}$, for every $i = 1, 2, ..., k$.*

Proof. See Appendix B. □

Theorem 2.3. *Consider a singular design η which puts mass α_i at design point d_i, $i = 1, 2, ..., k$. Let x_i be either d_i or $-d_i$ for $i = 1, 2, ..., k$ such that the angle between vector $\sum_i \alpha_i x_i$ and vector c is minimized. Suppose that the dimension of the space spanned by $\{x_1 - x_2, x_1 - x_3, ..., x_1 - x_k\}$ is t. The design η is c-optimal if and only if*

1. *x_1 is linearly independent of $\{x_1 - x_2, x_1 - x_3, ..., x_1 - x_k\}$.*
2. *there exists a $p \times p - t - 1$ of full rank matrix H, $[h_1, h_2, ..., h_{p-t-1}]$, such that $\{h_1, h_2, ..., h_{p-t-1}, x_1 - x_2, x_1 - x_3, ..., x_1 - x_k\}$ forms a basis of a separating hyperplane E_a and furthermore, $\sup_{x \in \mathcal{X}} F_{\phi_c}(M(\theta, \eta) + HH^T, I(\theta, x)) = 0$*

Proof. It is known that for such a linear model $M(\theta, \eta) = \sum \alpha_i d_i d_i^T$. Since $x_i = d_i$ or $-d_i$ $M(\theta, \eta) = \sum \alpha_i x_i x_i^T$.

The proof of "if" part is straightforward. Suppose condition 1 and 2 hold. Since $M(\theta, \eta) = \sum \alpha_i x_i x_i^T = AA^T$ where $A = [\sqrt{\alpha_1}x_1, \sqrt{\alpha_2}x_2, ..., \sqrt{\alpha_k}x_k]$, the rank of $M(\theta,\eta)$ is equal to the rank of A which is equal to the rank of $[x_1, x_1 - x_2, x_1 - x_3, ..., x_1 - x_k] = 1 + t$ by the first condition. Also because the second condition is satisfied the design η is c-optimal by the General Equivalence Theorem for singular optimal designs.

For the proof of "only if" part suppose the design η is c-optimal. Then $\sum \alpha_i x_i$ is equal to a by Elfving's Theorem. Also $F_{\phi_c}(M(\theta,\eta)+HH^T, I(\theta,x)) = c^T(M(\theta,\eta)+HH^T)^{-1}xx^T(M(\theta,\eta)+HH^T)^{-1}c - c^T(M(\theta,\eta)+HH^T)^{-1}$ (Silvey, 1980, Chapter 3 [8]). Hence the second condition can be rewritten as below: for any x in $\mathcal{X}$,

$$\begin{aligned}
& c^T(M(\theta,\eta) + HH^T)^{-1}xx^T(M(\theta,\eta) + HH^T)^{-1}c \\
& -c^T(M(\theta,\eta) + HH^T)^{-1}c \le 0 \\
\Leftrightarrow\ & \frac{1}{\lambda^2}a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}xx^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a \\
& -\frac{1}{\lambda^2}a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a \le 0 \\
\Leftrightarrow\ & [a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x]^2 - a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a \le 0 \\
\Leftrightarrow\ & a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x - \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} \le 0 \text{ and} \\
& a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x + \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} \ge 0
\end{aligned}$$

Since $a = \sum \alpha_i x_i$,

$$\begin{aligned}
aa^T &= \sum \alpha_i^2 x_i x_i^T + \sum_{i \ne j} \alpha_i \alpha_j x_i x_j^T \\
&= \sum \alpha_i x_i x_i^T - \sum \alpha_i(1-\alpha_i)x_i x_i^T + \sum_{i \ne j} \alpha_i \alpha_j x_i x_j^T
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum \alpha_i x_i x_i^T + HH^T &= aa^T + \sum \alpha_i(1-\alpha_i)x_i x_i^T - \sum_{i \ne j} \alpha_i \alpha_j x_i x_j^T + HH^T \\
&= aa^T + D
\end{aligned}$$

where D is $\sum \alpha_i(1-\alpha_i)x_i x_i^T - \sum_{i \ne j} \alpha_i \alpha_j x_i x_j^T + HH^T$.

The condition $a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x + \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} = 0$ is therefore equivalent to $a^T(aa^T + D)^{-1}x + \sqrt{a^T(aa^T + D)^{-1}a} = 0$. It can be easily verified that

$$\begin{aligned} D &= \sum \alpha_i(1-\alpha_i)x_i x_i^T - \sum_{i\neq j} \alpha_i\alpha_j x_i x_j^T + HH^T \\ &= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} \alpha_i\alpha_j(x_i-x_j)(x_i-x_j)^T + HH^T \\ &= \sum_{i<j} \alpha_i\alpha_j(x_i-x_j)(x_i-x_j)^T + HH^T \\ &= LL^T \end{aligned}$$

where $L = [\sqrt{\alpha_1\alpha_2}(x_1-x_2), ..., \sqrt{\alpha_i\alpha_j}(x_i-x_j)_{i<j}, ..., \sqrt{\alpha_{k-1}\alpha_k}(x_{k-1}-x_k), H]$. By Lemma 2.2 the first condition, x_1 is linearly independent of $\{x_1-x_2, x_1-x_3, ..., x_1-x_k\}$, is satisfied. Also $x_i, i = 1, 2, ..., k$ are all on a separating hyperplane E_a. Hence $x_i - x_j$ for $i, j = 1, 2, ..., k$ are vectors on E_a. Let E be the space spanned by $\{x_1-x_2, x_1-x_3, ..., x_1-x_k\}$ and let $h_1, h_2, ..., h_{p-t-1}$ form a linearly independent basis of the quotient space $E_a \backslash E$. Define H to be $[h_1, h_2, ..., h_{p-t-1}]$. It is obvious that the space E is also spanned by $\{x_i - x_j, i < j\}$. Hence the columns of L form a basis of this separating hyperplane E_a. It has been shown in Lemma 2.2 that vector a is linearly independent of any basis of a separating hyperplane E_a. Hence by Lemma 2.1 $a^T(aa^T+LL^T)^{-1}x - \sqrt{a^T(aa^T+LL^T)^{-1}a} = 0$ is this separating hyperplane E_a and so is $a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x - \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} = 0$. Similarly, $a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x + \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} = 0$ is a separating hyperplane at point $-a$. Furthermore, by Equation 1 $a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a = a^T(aa^T + LL^T)^{-1}a = 1$.

By the definition of separating planes of a convex set again all points in this set will be at the same side of its separating plane. The origin, o, belongs to the convex hull of $\mathcal{X} \cup -\mathcal{X}$. Hence

$$\begin{aligned} & a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}o - \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} \\ &= -\sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} = -1 < 0 \end{aligned}$$

implies that for all $x \in \mathcal{X}$

$$a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}x - \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1}a} \leq 0$$

Similarly

$$\begin{aligned}
&a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1} o + \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1} a} \\
&= \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1} a} = 1 > 0
\end{aligned}$$

implies that for all $x \in \mathcal{X}$

$$a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1} x + \sqrt{a^T(\sum \alpha_i x_i x_i^T + HH^T)^{-1} a} \geq 0$$

The second condition therefore holds. The proof is now complete. □

By Lemma 2.2 a separating hyperplane E_a is a separating hyperplane containing all x_i, $i = 1, 2, ..., k$ so a practical corollary follows.

Corollary. *In the second condition of Theorem 2.3 E_a, a separating hyperplane at a, can be replaced by a separating hyperplane at any x_j, which contains all $x_i, i = 1, 2, ..., k$.*

3 Examples

Example 1. $p = 2, k = 1$ case. Consider the model $y = \theta_0 + \theta_1 x + e$, e is $N(0, \sigma^2)$ and $-1 \leq x \leq 1$. Suppose that for some u between 0 and 1 the quantity of interest is $\theta_0 + \theta_1 u$. Then $\mathcal{X} = \{(1, x)^T | -1 \leq x \leq 1\}$ and $c = (1, u)^T$. The convex hull of $\mathcal{X} \cup -\mathcal{X}$ is the rectangle shown in Figure 1. The ray from the origin to c for $u = 0.5$ case is also drawn in Figure 1. It is clear by Elfving's Theorem that there are many c-optimal designs, one of which puts all mass at $x = u$. Its optimality can be also verified by General Equivalence Theorem as the following. Let the tangent hyperplane of the convex hull at $a = (1, k)^T$ be the separating hyperplane E_a in Theorem 2.3 and the matrix H is therefore equal to the tangent vector of the convex hull at a, which is clearly $(0, 1)^T$. Then the directional derivative of this singular design with $H = [0, 1]^T$, $f(x)$, is

$$\begin{aligned}
f(x) &= c^T(M(\theta, \eta) + HH^T)^{-1} x x^T (M(\theta, \eta) + HH^T)^{-1} c \\
&\quad - c^T(M(\theta, \eta) + HH^T)^{-1} c \\
&= \left\{ \begin{bmatrix} 1 & u \end{bmatrix} \left(\begin{bmatrix} 1 \\ u \end{bmatrix} \begin{bmatrix} 1 & u \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ x \end{bmatrix} \right\}^2 \\
&\quad - \begin{bmatrix} 1 & u \end{bmatrix} \left(\begin{bmatrix} 1 \\ u \end{bmatrix} \begin{bmatrix} 1 & u \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ u \end{bmatrix} \\
&= 0
\end{aligned}$$

So $f(x)$ is obviously nonpositive for all $-1 \le x \le 1$.

An interesting point is that the zeros of the directional derivative of a singular optimal design are not necessarily all design points, though each design point is one of those zeros (see Silvey, 1980, Chapter 3[8]). In addition, not all g-inverses of $M(\theta, \eta)$ which can be written in form of $(M(\theta,\eta)+HH^T)^{-1}$ for some matrix H can be used to verify the optimality. For instance, let M^+ be the Moore-Penrose g-inverse of $M(\theta, \eta)$:

$$M^+ = \frac{1}{(1+u^2)^2} \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix}$$

and let H be $\frac{u\sqrt{2+u^2}}{1+u^2}(1,u)^T$ then it is easy to verify that M^+ is a g-inverse of $M(\theta,\eta)$ and can be written as $(M(\theta,\eta)+HH^T)^{-1}$. Using this H

$$\begin{aligned} f(x) &= c^T M^+ x x^T M^+ c - c^T M^+ c \\ &= \frac{1+ux}{1+u^2} - 1. \end{aligned}$$

No matter what the value of u is, $f(1) = \frac{u(1-u)}{1+u^2}$ is always positive since $0 < u < 1$. This matrix H is therefore not appropriate for proving the optimality of this c-optimal design in this example.

Theorem 2.3 can be also applied to higher dimensions as Example 2 shows.

Example 2. $p = 3, k = 2$ case. Consider the model $y = \theta_0 + \theta_1 x + \theta_2 x^2 + e$, e is $N(0,\sigma^2)$ and $-1 \le x \le 1$. Suppose that θ_2 is known to be positive and $g = -\theta_1/(2\theta_2)$ is what needs to be estimated (the expected value of y is minimized at $x = g$). Hence a vector c can be defined as $c = \nabla(-g) = (0, 1, 2g)$. Since $\mathcal{X} = \{(1, x, x^2)^T | -1 \le x \le 1\}$ the convex hull of $\mathcal{X} \cup -\mathcal{X}$ is the domain shown in Figure 3. The formula of this domain is not easy to write down and it is hard to apply Elfving's Theorem directly in this case. A candidate c-optimal design can be generated from Elfving's idea and then verified to be optimal by Theorem 2.3 using the tangent vectors of the convex hull of $\mathcal{X} \cup -\mathcal{X}$.

The ray from the origin to c is also drawn in Figure 3. Suppose $0 \le g \le 1/2$ then

$$0.5\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 0.5\begin{pmatrix} -1 \\ -(2g-1) \\ -(2g-1)^2 \end{pmatrix} = (1-g)\begin{pmatrix} 0 \\ 1 \\ 2g \end{pmatrix} = (1-g)c.$$

Let a be this boundary point, $a = (1-g)(0,1,2g)^T$, x_1 be the point $(1,1,1)^T$, and x_2 be the point $(-1,-(2g-1),-(2g-1)^2)^T$, which are all shown in Figure 3. Elfving's Theorem suggests that the design putting mass 0.5 at 1 and $2g-1$ each might be c-optimal. Its optimality will be verified by the General Equivalence Theorem. It is clear that the tangent hyperplane of the convex hull at x_2 is a separating hyperplane containing x_1 and x_2. According to Corollary this tangent hyperplane can be used to replace the separating hyperplane E_a in Theorem 2.3. From Figure 3 the tangent vector of curve: $\{(-1,t,-t^2)^T| -1 \le t \le 1\}$ at x_2 is also the tangent vector of the convex hull at x_2. This tangent vector can be easily found to be $(0,1,-2(-(2g-1)))^T = (0,1,2(2g-1))^T$. Let this vector be the 3×1 matrix H in Theorem 2.3. Then the directional derivative, $f(x)$, of this singular design with $H = [0,1,2(2g-1)]^T$ is therefore

$$\begin{aligned}
f(x) &= c^T(M(\theta,\eta)+HH^T)^{-1}\begin{bmatrix}1\\x\\x^2\end{bmatrix}\begin{bmatrix}1 & x & x^2\end{bmatrix}(M(\theta,\eta)+HH^T)^{-1}c - \\
&\quad c^T(M(\theta,\eta)+HH^T)^{-1}c \\
&= c^T(\sum x_i x_i^T + HH^T)^{-1}\begin{bmatrix}1\\x\\x^2\end{bmatrix}\begin{bmatrix}1 & x & x^2\end{bmatrix}(\sum x_i x_i^T + HH^T)^{-1}c - \\
&\quad c^T(\sum x_i x_i^T + HH^T)^{-1}c \\
&= \frac{(x-1)(x-4g+3)(x-2g+1)^2}{4(g-1)^6}
\end{aligned}$$

Since $0 \le g \le 1/2$ implies $1 \le -4g+3 \le 3$, $(x-1)(x-4g+3)$ is nonpositive for $-1 \le x \le 1$. $f(x)$ is therefore nonpositive for $-1 \le x \le 1$. Figure 2 shows this claim, $f(x) \le 0$, for $g = 1/3$ case.

Theorem 2.3 can be also applied to nonlinear regression models and generalized linear models as long as the information matrix of such a model can be written as $\sum_i g(x_i)g(x_i)^T$ for some function $g(x)$(Ford, Torsney and Wu, 1992[2]). The proof of Theorem 2.3 only involves the information matrices not the model. The value of $g(x)$ can be thought of as new "design point" as in Example 3.

Example 3. Consider the logistic regression of a binary response y on a predictor x. The probability of "success" is $p(x)$ and $\log \frac{p(x)}{1-p(x)} = \theta_0 + \theta_1 x$.

The information matrix for a single observation at x is:

$$\begin{aligned} I(\theta,x) &= \frac{e^{\theta_0+\theta_1 x}}{(1+e^{\theta_0+\theta_1 x})^2}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} \\ &= g(x)g(x)^T \end{aligned}$$

where $g(x) = (g_1(x), g_2(x))^T = \{\exp[(\theta_0+\theta_1 x)/2]/[1+\exp(\theta_0+\theta_1 x)]\}(1,x)^T$. The point $g(x)$ can be thought of as a design point, so $\mathcal{X} = \{g(x)|x \in \mathbb{R}\}$. It is clear that $\mathcal{X}$ depends on the values of θ_0 and θ_1. For local optimality assume $\theta_0 = 0$ and $\theta_1 = 1$. The plot of $\mathcal{X}$ and $-\mathcal{X}$ is shown in Figure 4. Suppose $\theta_0 + \theta_1$ is the quantity of interest so $c = (1,1)^T$. The ray from the origin to c is shown in Figure 4. The boundary point, a, reached by extending this ray is $g(1) = (0.4434, 0.4434)^T$ (If $g_1(x) = g_2(x)$ then $x = 1$). This point a is also shown in Figure 4. Elfving's Theorem is actually quite straightforward in this case. It says that the design putting all mass at $x = 1$ is c-optimal. Its optimality can be also verified by General Equivalence Theorem for illustration. The tangent hyperplane at point a is a separating hyperplane E_a. Also the tangent vector of the convex hull of $\mathcal{X} \cup -\mathcal{X}$ at point a is:

$$\begin{pmatrix} \frac{dg_1(x)}{dx} \\ \frac{dg_2(x)}{dx} \end{pmatrix}_{x=1} = \begin{pmatrix} -0.1025 \\ 0.3410 \end{pmatrix}$$

Hence this vector, $(-0.1025, 0.3410)^T$, can be assigned to be the matrix H in Theorem 2.3. Let M be the information matrix of this c-optimal design. Then

$$\begin{aligned} f(x) &= [c^T(M+HH^T)^{-1}g(x)]^2 - c^T(M+HH^T)^{-1}c \\ &= -5.0862 + \left(\frac{3.911e^{0.5x}}{1+e^x} + \frac{1.1752xe^{0.5x}}{1+e^x}\right)^2 \end{aligned}$$

$f(x)$ is nonpositive supported by the plot of $f(x)$ in Figure 5. The design putting mass one at $x = 1$ is therefore c-optimal by the General Equivalence Theorem.

Appendix A: Proof of Lemma 2.1

Since L is of rank $p-1$ there exists an orthonormal $p \times p$ matrix Γ such that the last $q - p + 1$ columns and the last row of ΓL are all zero vectors, say,

$\Gamma L = \begin{bmatrix} L_s & 0 \\ 0 & 0 \end{bmatrix}$ where L_s is a full rank $p-1 \times p-1$ matrix. An orthonormal transformation is only a rotation so it will not change the geometry.

$$a^T\Gamma^T(\Gamma aa^T\Gamma^T + \Gamma LL^T\Gamma^T)^{-1}\Gamma x - \sqrt{a^T\Gamma^T(\Gamma aa^T\Gamma^T + \Gamma LL^T\Gamma^T)^{-1}\Gamma a}$$
$$= a^T\Gamma^T\Gamma(aa^T + LL^T)^{-1}\Gamma^T\Gamma x - \sqrt{a^T\Gamma^T\Gamma(aa^T + LL^T)^{-1}\Gamma^T\Gamma a}$$
$$= a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a}$$

Without loss of generality therefore consider $L = \begin{bmatrix} L_s & 0 \\ 0 & 0 \end{bmatrix}$.

The proof of "if" part:

If the vector a is not linearly independent of $\{l_1, l_2, ..., l_q\}$, a can be written as a linear combination of $\{l_1, l_2, ..., l_q\}$, $a = Lv$ where v is a $q \times 1$ vector. Since l_i for $i > p-1$ are zero vector, v can be written as $\begin{pmatrix} v_s \\ 0 \end{pmatrix}$ where v_s is a $p-1 \times 1$ vector. Let I_t be the $t \times t$ identity matrix then

$$\begin{aligned} aa^T + LL^T &= Lvv^TL^T + LL^T \\ &= L(vv^T + I_q)L^T \\ &= \begin{bmatrix} L_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_sv_s^T + I_{p-1} & 0 \\ 0 & I_{q-p+1} \end{bmatrix} \begin{bmatrix} L_s^T & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} L_s(v_sv_s^T + I_{p-1})L_s^T & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

The matrix $aa^T + LL^T$ is therefore singular. This contradicts the nonsingularity of $aa^T + LL^T$. Hence vector a and $\{l_1, l_2, ..., l_q\}$should be linearly independent.

The proof of "only if" part:

Partition the vector a into two parts: $a^T = [a_s^T, a_p]$ such that a_s is $p-1 \times 1$ vector. The matrix L_s is of rank $p-1$ so the columns of L form a basis of the subspace: $\{x = (x_1, x_2, ..., x_p)^T | x_p = 0\}$. Hence a_p is nonzero because of the linear independence between a and the columns of L. Let W be $L_sL_s^T$. Using the following formula (page 69, Morrison)[3]

$$(A + cbb^T)^{-1} = A^{-1} - \frac{c}{1 + cb^TA^{-1}b}A^{-1}bb^TA^{-1}$$

the inverse matrix of $aa^T + LL^T$ exists and can be written as

$$\begin{bmatrix} W^{-1} & -\frac{W^{-1}a_s}{a_p} \\ -\frac{a_s^TW^{-1}}{a_p} & \frac{1+a_s^TW^{-1}a_s}{a_p^2} \end{bmatrix}$$

Then

$$\begin{aligned} & a^T(aa^T + LL^T)^{-1} \\ &= [a_s^T, a_p] \begin{bmatrix} W^{-1} & -\frac{W^{-1}a_s}{a_p} \\ -\frac{a_s^T W^{-1}}{a_p} & \frac{1+a_s^T W^{-1} a_s}{a_p^2} \end{bmatrix} \\ &= [0, 1/a_p] \end{aligned}$$

It is therefore easy to see that

$$a^T(aa^T + LL^T)^{-1}a = [0, 1/a_p] \begin{bmatrix} a_s \\ a_p \end{bmatrix} = 1 \tag{1}$$

and $a^T(aa^T + LL^T)^{-1}l_i = [0, 1/a_p]l_i = 0$ for all $i = 1, 2, ..., q$. Hence, for $x = a, a + l_1, a + l_2, ..., a + l_q$

$$a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$$

So $a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$ is the hyperplane spanned by $\{l_1, l_2, ..., l_q\}$, and going through point a.

Let p be the minimum dimension of all Euclidean spaces including the convex hull of $\mathcal{X} \cup -\mathcal{X}$. After a suitable orthonormal transformation $\mathcal{X}$ can be embedded into $\mathbb{R}^p$. From now on consider the domain of x, $\mathcal{X}$, in such a $\mathbb{R}^p$.

Appendix B: Proof of Lemma 2.2

Since a is a boundary point of the convex hull of $\mathcal{X} \cup -\mathcal{X}$, a is nonzero. Let E_a be arbitrarily given and $\{l_1, l_2, ..., l_{p-1}\}$ be a basis of this separating hyperplane E_a. Then the vector a is linearly independent of $\{l_1, l_2, ..., l_{p-1}\}$. Otherwise, the origin $(= a - a)$ will be on this separating hyperplane E_a. But the origin is clearly an interior point of the convex hull of $\mathcal{X} \cup -\mathcal{X}$, which conflicts with the definition of separating planes of a convex set. In addition, the vector a is therefore linear independent of any sets of vectors on a separating hyperplane E_a. By Lemma 2.1 this separating hyperplane E_a is the hyperplane $a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$ where $L = [l_1, l_2, ..., l_{p-1}]$. If for some j, x_j is not on the separating hyperplane E_a then $a^T(aa^T + LL^T)^{-1}x_j - \sqrt{a^T(aa^T + LL^T)^{-1}a}$ is not equal to 0. By the definition of separating hyperplane all points in the convex hull of $\mathcal{X} \cup -\mathcal{X}$ should be at the same side of its separating plane.

Hence if $a^T(aa^T + LL^T)^{-1}x_j - \sqrt{a^T(aa^T + LL^T)^{-1}a} < 0$ (or > 0) then $a^T(aa^T + LL^T)^{-1}x_i - \sqrt{a^T(aa^T + LL^T)^{-1}a} \leq 0$ (or ≥ 0) for all $i \neq j$. In addition, a is of course on this separating hyperplane E_a. Hence

$$\begin{aligned} 0 &= a^T(aa^T + LL^T)^{-1}a - \sqrt{a^T(aa^T + LL^T)^{-1}a} \\ &= \sum_i \alpha_i a^T(aa^T + LL^T)^{-1}x_i - \sqrt{a^T(aa^T + LL^T)^{-1}a} \\ &< 0 \text{ (or } > 0) \end{aligned}$$

This is impossible. So x_i for $i = 1, 2, ..., k$ are all on this separating hyperplane E_a. Since E_a is arbitrarily given the first claim is proved.

Since every separating hyperplane at a, E_a, contains all x_i, $i = 1, 2, ..., k$ the set S_a is therefore contained in the set S_x. A hyperplane containing all x_i, $i = 1, 2, ..., k$ will also contain a because $a = \sum_i \alpha_i x_i$, a convex combination of $x_i, i = 1, 2, ..., k$. If this hyperplane is a separating hyperplane then it is also a separating plane at a. Hence S_x is contained in S_a, too. The proof of the second claim, $S_a = S_x$, is now complete. Because $x_i - x_j, i \neq j$ are vectors on E_a, which is obviously also a separating hyperplane at x_i, x_i is linearly independent of $\{x_i - x_j, i \neq j\}$. The proof is therefore complete.

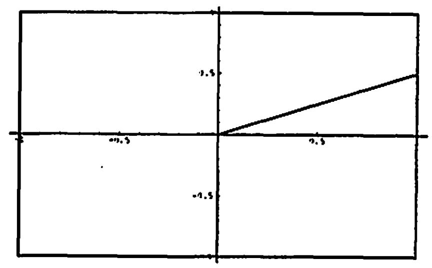

Figure 1: Convex hull of $\mathcal{X} \cup -\mathcal{X}$ in Example 1

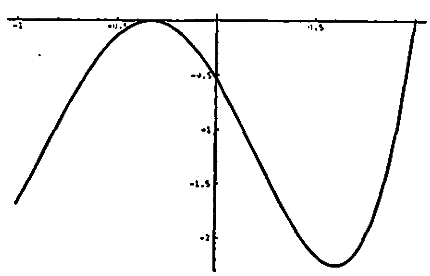

Figure 2: the plot of directional derivative in Example 2 for $g = 1/3$ case

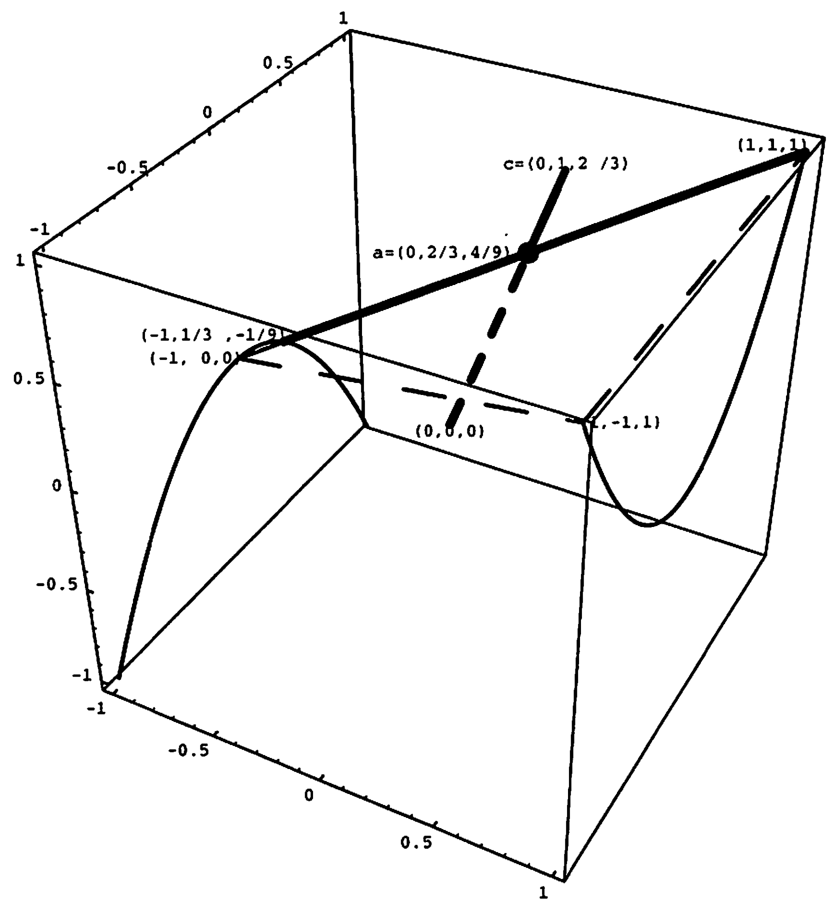

Figure 3: Convex hull of $\mathcal{X} \cup -\mathcal{X}$ in Example 2

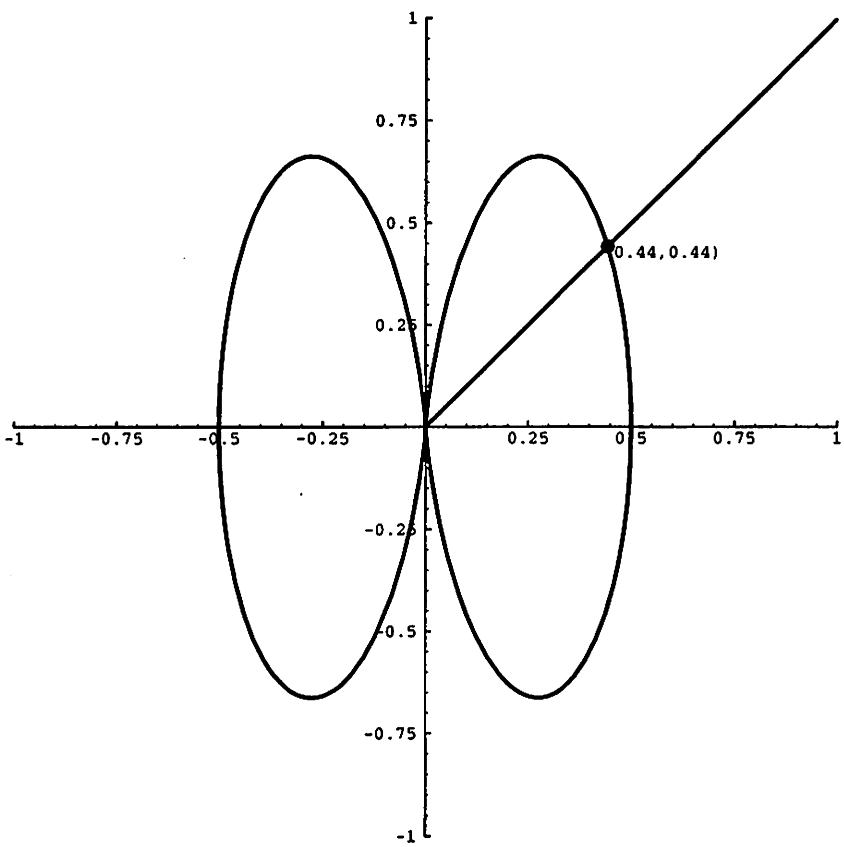

Figure 4: Plot of $\mathcal{X}$ and $-\mathcal{X}$ in Example 3

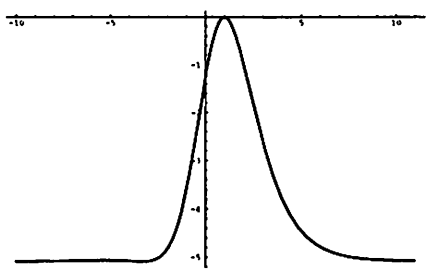

Figure 5: the plot of directional derivative in Example 3 for $\theta_0 = 0$ and $\theta_1 = 1$ case

References

[1] Elfving, G. 1952. Optimum allocation in linear regression theory. *Ann. Math. Statist.*, 23, p255-262.

[2] Ford, I. , Torsney, B. and Wu, C. F. J. 1992. The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J.R. Statist. Soc. B.*, 54, p569-583.

[3] Morrison, Donald. 1976. *Multivariate Statistical Methods.* McGraw-Hill Book Company.

[4] Pukelsheim, F. 1981. On c−optimal design measures. *Mathematische Operationsforschung und Statistik*, 12, p13-20.

[5] Pukelsheim, F. 1981. Linear models and convex geometry: Aspects of non-negative variance estimation. *Mathematische Operationsforschung und Statistik*, 12, p271-286.

[6] Searle, S. R. 1971. *Linear Models.* New York : Wiley.

[7] Silvey, S. D. 1978. Optimal design measures with singular information matrices. *Biometrika*, 65, p553-559.

[8] Silvey, S. D. 1980. *Optimal Design.* Chapman and Hall.

[9] Studden, W. J. 1971. Elfving's theorem and optimal designs for quadratic loss. *Ann. Math. Statis.*, 42, p1613-1621.

[10] White, V. L. 1975. *The Optimal Design of Experiments For Estimation In Nonlinear Models.* PhD Thesis in the University of London.