

A Geometric Method for Singular c-optimal Designs

BY SHENGHUA KELLY FAN

*Department of Statistics and Applied Probability, National University of
Singapore, Singapore 117543, Singapore*

AND KATHRYN CHALONER

School of Statistics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

SUMMARY

If a candidate c-optimal design gives a singular information matrix, an equivalence theorem from Silvey (1978) can be used to verify optimality. The theorem is difficult to use however, as it requires a generalized inverse of the information matrix but not all generalized inverses can be used. Silvey identified finding such an inverse as an open problem. A characterization of all generalized inverses that can be used is presented here, building on Elfving's (1952) geometric method for finding c-optimal designs and using the Elfving set.

AMS Subject Classifications: 62K05.

Key words: Elfving's Theorem, Equivalence Theorem.

1 Introduction

Consider a linear model for a univariate response, $y_i = \theta^T x_i + e_i, i = 1, \dots, n$, where θ denotes a $p \times 1$ vector of unknown parameters and e_i are independent $N(0, \sigma^2)$. The domain of x , denoted \mathcal{X} , is assumed to be p -dimensional and compact. Let \mathcal{H} be the class of probability measures on the Borel sets of \mathcal{X} . An element η in \mathcal{H} is called a design. The information matrix of a design η putting mass α_i at $x = x_i$ is $M(\eta) = \sum_i \alpha_i x_i x_i^T$. Let \mathcal{M} denote the set of $M(\eta)$ for η in \mathcal{H} , and for notational simplicity $M(\eta)$ will be denoted M .

Suppose the goal of an experiment is to estimate a linear function of the parameters, $c^T \theta$. A c -optimal design is one which minimizes the variance of the estimate of $c^T \theta$. For a design with information matrix M , $c^T \theta$ is estimable if there exists a vector v such that $c = Mv$. Let \mathcal{M}_c be the subset of information matrices M for which $c^T \theta$ is estimable. For singular M , let M^- denote any generalized inverse of M . The criterion of c -optimality is

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

A design η is c -optimal if it maximizes ϕ_c over all M in \mathcal{M} . This criterion is concave and continuous so the maximum exists and any local maximum is the global maximum (see Silvey, 1980). If M is singular the criterion is invariant to which generalized inverse is used.

Elfving (1952) gave a geometric method for finding the set of all c -optimal designs. Let $-\mathcal{X}$ be the set of points $-x$ for $x \in \mathcal{X}$ and denote the convex hull of $\mathcal{X} \cup -\mathcal{X}$ to be $CH(\mathcal{X} \cup -\mathcal{X})$, also called the Elfving set.

Elfving's Theorem

Define a to be the point where the ray from the origin to the point c penetrates $CH(\mathcal{X} \cup -\mathcal{X})$. If a can be expressed as $\sum_i \alpha_i x_i$ where $\alpha_i > 0$, $\sum_i \alpha_i = 1$ and x_i is in either \mathcal{X} or $-\mathcal{X}$, then a corresponding c -optimal design puts proportion α_i of observations at x_i or $-x_i$, whichever belongs to \mathcal{X} . All c -optimal designs can be identified in this way.

Elfving's theorem can be difficult to apply, either in high dimensions or when $CH(\mathcal{X} \cup -\mathcal{X})$ is complicated and hard to visualize. A candidate c -optimal design may be suggested, however, by the geometry and its optimality can be examined using directional derivatives. The directional derivative of the criterion ϕ at M in the direction of M_1 is

$$F_\phi(M, M_1) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} [\phi\{(1 - \epsilon)M + \epsilon M_1\} - \phi\{M\}].$$

If M is nonsingular the usual equivalence theorem can be easily applied (see Silvey, 1980, Chapter 3). The derivative for ϕ_c at a nonsingular M in the direction of M_1 , is

$$F_{\phi_c}(M, M_1) = c^T M^{-1} M_1 M^{-1} c - c^T M^{-1} c.$$

If a candidate M is singular a version of the equivalence theorem given in Silvey

(1980, Chapter 3) is appropriate, but is difficult to use. Let $I(x) = xx^T$ denote the information matrix of a design putting point mass at x .

Silvey's Theorem

A design η , with M in \mathcal{M}_c and with rank r , $r < p$, is c -optimal if, and only if, there exists a $p \times (p-r)$ matrix H of rank $p-r$ such that $M + HH^T$ is nonsingular and

$$\sup_{x \in \mathcal{X}} F_{\phi_c}(M + HH^T, I(x)) = 0. \quad (1)$$

Identifying such a matrix H is the problem addressed here. Searle (1971, p.22) showed that $(M + HH^T)^{-1}$ must be a generalized inverse of M , but not all such generalized inverses can be used. For concave optimal design criteria, Silvey (1978) showed that if one such matrix H existed then a class of such matrices exist, all of whose columns span the same $p - r$ dimensional subspace of \mathbb{R}^p , but Silvey did not identify the subspace. Pukelsheim (1980) shows that the required generalized inverse should be a contracting g-inverse which can be found by solving a duality problem. Pukelsheim and Titterton (1983) who use both subgradient theory and Lagrangian duality to give an alternative approach of verifying if a design is c -optimal. They do not, however, provide a method for constructing the matrix H (or the required subgradient) for a candidate design.

This paper builds on Elfving's theorem to provide a geometric characterization of all matrices H which can be used.

2 Geometric Interpretation

Recall that if S is a convex set in \mathbb{R}^p then a supporting hyperplane of S at s is a $p - 1$ dimensional hyperplane containing s such that S is all on one side of this hyperplane (see DeGroot, 1970, p.131). The following theorem characterizes all matrices H which can be used in Silvey's theorem.

Theorem

Consider a design η , with singular information matrix of rank r , $r < p$, putting mass α_i at d_i where d_i is in \mathcal{X} and $\alpha_i > 0$, $i = 1, \dots, k$. If η is c -optimal define a to be the vector in Elfving's theorem on the boundary of $CH(\mathcal{X} \cup -\mathcal{X})$ which can be expressed as $\sum \alpha_i x_i$ with $x_i = d_i$ or $-d_i$. Then

- for $k = 1$, a matrix $H = [h_1, h_2, \dots, h_{p-r}]$ satisfies (1) in Silvey's theorem if and only if the set $\{h_1, h_2, \dots, h_{p-r}\}$ is such that the hyperplane

$$\{x | x = a + \sum_{i=1}^{p-r} \lambda_i h_i, \lambda_i \in \mathbb{R}\}$$

is a supporting hyperplane of $CH(\mathcal{X} \cup -\mathcal{X})$ at a .

- for $k > 1$, a matrix $H = [h_1, h_2, \dots, h_{p-r}]$ satisfies (1) in Silvey's theorem if and only if the set $\{x_1 - x_2, x_1 - x_3, \dots, x_1 - x_k, h_1, h_2, \dots, h_{p-r}\}$, denoted $\{z_1, \dots, z_{k-1+p-r}\}$, is such that the hyperplane

$$\{x | x = a + \sum_{i=1}^{k-1+p-r} \lambda_i z_i, \lambda_i \in \mathbb{R}\}$$

is a supporting hyperplane of $CH(\mathcal{X} \cup -\mathcal{X})$ at a .

Essentially, H is defined to have columns which, together with the set of vectors $x_1 - x_i$, $i = 2, \dots, k$, span a supporting hyperplane of $CH(\mathcal{X} \cup -\mathcal{X})$ at a . Let this hyperplane be E_a : then $(M + HH^T)$ can be expressed as $(aa^T + LL^T)$, where the columns of L span E_a . Condition (1) in Silvey's theorem becomes an inequality defining both E_a and a corresponding supporting hyperplane of $CH(\mathcal{X} \cup -\mathcal{X})$ at $-a$. A detailed proof is given in the Appendix.

In practice, if a and a supporting hyperplane, E_a , are easy to identify, Elfving's theorem is straightforward to apply and the use of the equivalence theorem is unnecessary. When a and E_a are hard to visualize however, the theorem provides an alternative method to check optimality: for a candidate $a = \sum \alpha_i x_i$ replace a supporting hyperplane at a in the theorem by a candidate hyperplane containing all x_i 's to calculate H . (Lemma 2, in the Appendix, gives a justification of the replacement). Then $F_{\phi_c}(M + HH^T, I(x))$ can be examined for all x in \mathcal{X} to check that (1) holds. This will be applied in example 1 where it is easy to find an $a = lc$ but it is not clear whether or not a is on the boundary of $CH(\mathcal{X} \cup -\mathcal{X})$.

3 Examples

Example 1 is of estimating the turning point in a quadratic regression and so the function of interest is a nonlinear function of θ , $f(\theta)$. In this case the gradient vector $\nabla f(\theta)$, or equivalently $\nabla\{-f(\theta)\}$, evaluated at a particular value of θ , serves as the vector c in the c-optimality criterion and in the corresponding equivalence

theorem (see Silvey, 1980). Example 2 is a trivial example from linear regression which illustrates two subtleties of the equivalence theorem: first that the zeros of the directional derivative at an optimal design do not necessarily correspond to support points and second that not all generalized inverses of M , of the form $(M + HH^T)^{-1}$ can be used in (1). Example 3 is an example of the c-optimality criterion applied to generalized linear models: see Wu (1988), Ford, Torsney and Wu (1992), and Haines (1995).

Example 1. Consider the quadratic regression model with mean response $\theta_0 + \theta_1 u + \theta_2 u^2$, with $\mathcal{X} = \{(1, u, u^2)^T \mid -1 \leq u \leq 1\}$. Suppose that θ_2 is known to be positive and $f(\theta) = f = -\theta_1/(2\theta_2)$ is to be estimated. This f is the value of u where the expected value of y is minimized and so let $c = \nabla(-f) = (0, 1, 2f)$. \mathcal{X} and $-\mathcal{X}$ are shown in Figure 1 together with c for $f = 1/3$. The set $CH(\mathcal{X} \cup -\mathcal{X})$ is not easy to express algebraically and Elfving's theorem is hard to apply directly (but see Buonaccorsi and Iyer, 1986, and Chaloner, 1987). A candidate c-optimal design can however be generated from the geometry.

Suppose $0 \leq f \leq 1/2$ and define $x_1 = (1, 1, 1)^T$, and $x_2 = (-1, -(2f - 1), -(2f - 1)^2)^T$. Then

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 = (1 - f)c,$$

suggesting that the design, η , putting mass $\frac{1}{2}$ at each of $u = 1$ and $u = 2f - 1$ could be c-optimal with $a = (1 - f)c$. The geometry leads to a conjecture that the line joining x_1 and x_2 is on the boundary of $CH(\mathcal{X} \cup -\mathcal{X})$. If so, $x_1 - x_2$ and the tangent

vector to the curve $\{(-1, -u, -u^2)^T | u \in \mathbb{R}\}$ at x_2 , namely $(0, -1, -2(2f - 1))^T$, span the supporting hyperplane containing x_1 and x_2 . Let H , therefore, be the vector $(0, -1, -2(2f - 1))^T$. With this H the directional derivative (1) for ϕ_c in the direction of an information matrix for a one point design at $x = (1, u, u^2)^T$ is then

$$F_{\phi_c}(M + HH^T, I(x)) = \frac{(u - 1)(u - 4f + 3)(u - 2f + 1)^2}{4(f - 1)^6}.$$

Since $0 \leq f \leq 1/2$, this is nonpositive for $-1 \leq u \leq 1$. The conjecture has been confirmed and the design is indeed optimal by Silvey's theorem.

Example 2. Consider the simple linear regression model with mean response $\theta_0 + \theta_1 u$, and $\mathcal{X} = \{(1, u)^T | -1 \leq u \leq 1\}$. Suppose that for some w between 0 and 1 the quantity of interest is $\theta_0 + \theta_1 w$: then $c = (1, w)^T$. $CH(\mathcal{X} \cup -\mathcal{X})$ is a square and Elfving's theorem is very easy to apply. It is clear that there is a convex set of c -optimal designs for any w , one of which is singular and puts all mass at $u = w$. It is also clear that for $0 \leq w < 1$, $a = c$ and the tangent vector to E_a must be $(0, 1)^T$ and so, for the singular design with all mass at $u = w$, $H = (0, 1)^T$. The derivative function is easily calculated and is observed to be identically zero for all x in \mathcal{X} . This example illustrates that the zeros of the derivative function for the singular optimal design are not necessarily support points of the optimal design (although each support point must be a zero). This example also demonstrates that not all generalized inverses of M which can be written as $(M + HH^T)^{-1}$ can be used in (1) to verify optimality: the Moore-Penrose generalized inverse of M is

given by $H = \frac{w\sqrt{2+w^2}}{1+w^2}(1, w)^T$ but substituting this into (1) with $x = (1, u)^T$ gives

$$\frac{1 + wu}{1 + w^2} - 1$$

which is strictly positive for $u > w$.

Example 3. Consider the logistic regression of a binary response y on a predictor x . The probability of response is $p(x)$ and $\log[p(x)/\{1 - p(x)\}] = \theta_0 + \theta_1 x$. Let $g(x) = (g_1(x), g_2(x))^T = [\exp\{(\theta_0 + \theta_1 x)/2\}/\{1 + \exp(\theta_0 + \theta_1 x)\}](1, x)^T$. Then the information matrix for a single observation at x is:

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

The point $g(x)$ can be thought of as a design point (as in for example Haines, 1995), so let $\mathcal{X} = \{g(x) | x \in \mathbb{R}\}$. Clearly \mathcal{X} depends on the values of θ_0 and θ_1 : for illustration assume $\theta_0 = 0$ and $\theta_1 = 1$. The plot of \mathcal{X} and $-\mathcal{X}$ is shown in Figure 2. Suppose $\theta_0 + \theta_1$ is of interest and so $c = (1, 1)^T$. The ray from the origin to c is shown in Figure 2. The boundary point, a , is $g(1) = (0.4434, 0.4434)^T$. Elfving's Theorem is straightforward in this case: the design putting all mass at $x = 1$ is c -optimal.

The tangent vector of $CH(\mathcal{X} \cup -\mathcal{X})$ at a is:

$$\left(\begin{array}{c} \frac{dg_1(x)}{dx} \\ \frac{dg_2(x)}{dx} \end{array} \right)_{x=1} = \left(\begin{array}{c} -0.1025 \\ 0.3410 \end{array} \right)$$

Hence this vector, $(-0.1025, 0.3410)^T$, can be used as the matrix H in Silvey's Theorem. Let M be the information matrix of this design. Then

$$F_{\phi_c}(M + HH^T, I(x)) = -5.0862 + \left(\frac{3.911e^{0.5x}}{1 + e^x} + \frac{1.1752xe^{0.5x}}{1 + e^x} \right)^2.$$

This function can be shown to be non-positive by plotting it.

4 Discussion

Dette (1993, 1996), Haines (1993, 1995) and Studden (1971) give related geometric methods for optimal design that also use the Elfving set. In this paper a geometric solution has been provided to specify generalized inverses for the elegant equivalence theorem of Silvey (1978). The solution has been programmed into software for finding optimal designs for binary response models (see Smith and Ridout, 2000). Silvey's theorem complements the powerful theorem of Elfving, which remains the best way of characterizing all c-optimal designs for a particular problem.

In problems where $p > 3$ the Elfving set is often difficult to visualize and so it may be difficult to use Elfving's theorem or to generate candidate optimal designs. The geometric method given here for implementing Silvey's theorem provides an additional tool, but higher dimensional problems may remain challenging.

Acknowledgements

This research was supported in part by a grant from the National Security Agency.

We are very grateful to the referees for helpful comments and also to M.L. Eaton,

F. Pukelsheim and L. Tierney.

Appendix

The following two lemmas are both standard algebraic results. Lemma 1 is required in the proof of the Theorem and Lemma 2 is required at the end of Section 2 in interpreting the Theorem. The proofs are straightforward and are given in Fan (1999).

Lemma 1. *In \mathbb{R}^p consider a nonzero vector b and a $p \times q$ matrix $L = [l_1, l_2, \dots, l_q]$ of rank $p - 1$, where l_i , $i = 1, 2, \dots, q$ are column vectors. If $(bb^T + LL^T)$ is nonsingular, then the hyperplane $\{x | x = b + \sum_{i=1}^q \lambda_i l_i, \lambda_i \in \mathbb{R}\}$ can also be expressed as $\{x | b^T (bb^T + LL^T)^{-1} x - \sqrt{b^T (bb^T + LL^T)^{-1} b} = 0\}$.*

Lemma 2. *Let b be a boundary point of $CH(\mathcal{X} \cup -\mathcal{X})$. If $b = \sum_i \alpha_i x_i$, where $\alpha_i > 0$, $\sum_i \alpha_i = 1$, and $x_i \in CH(\mathcal{X} \cup -\mathcal{X})$ for $i = 1, 2, \dots, k$ then*

1. *All x_i for $i = 1, 2, \dots, k$ are on the boundary of $CH(\mathcal{X} \cup -\mathcal{X})$.*
2. *The set S_b , of all supporting hyperplanes of $CH(\mathcal{X} \cup -\mathcal{X})$ at the point b , is identical to the set S_x , of all supporting hyperplanes of $CH(\mathcal{X} \cup -\mathcal{X})$ containing all points x_i for $i = 1, 2, \dots, k$.*

Proof of Theorem

If the design η is c -optimal, Elfving's theorem guarantees the existence of a point $a = \sum \alpha_i x_i$, with $x_i = d_i$ or $-d_i$, which equals to lc and is on the boundary of $CH(\mathcal{X} \cup -\mathcal{X})$. There must therefore be at least one supporting hyperplane at a : denote such a hyperplane as E_a .

Let H be a $p \times (p-r)$ matrix of rank $p-r$. Define L as: $L = H$ for $k = 1$ and $L = [\sqrt{\alpha_1 \alpha_2}(x_1 - x_2), \dots, \sqrt{\alpha_i \alpha_j}(x_i - x_j)_{i < j}, \dots, \sqrt{\alpha_{k-1} \alpha_k}(x_{k-1} - x_k), h_1, h_2, \dots, h_{p-r}]$ for $k > 1$. It is easy to check that

$$M + HH^T = \sum \alpha_i x_i x_i^T + HH^T = aa^T + LL^T.$$

Silvey's condition (1), $\sup_{x \in \mathcal{X}} F_\phi(M + HH^T, I(x)) = 0$, is therefore equivalent to

$$c^T(aa^T + LL^T)^{-1}xx^T(aa^T + LL^T)^{-1}c - c^T(aa^T + LL^T)^{-1}c \leq 0,$$

for all x in \mathcal{X} . This is, in turn, equivalent to

$$\begin{aligned} a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} &\leq 0 \\ \text{and } a^T(aa^T + LL^T)^{-1}x + \sqrt{a^T(aa^T + LL^T)^{-1}a} &\geq 0, \end{aligned}$$

for all x in \mathcal{X} .

If H satisfies the condition of the theorem, then the matrix $aa^T + LL^T$ is nonsingular and so by Lemma 1, the equality $a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$ defines an E_a . Similarly the equality $a^T(aa^T + LL^T)^{-1}x + \sqrt{a^T(aa^T + LL^T)^{-1}a} = 0$ defines a supporting hyperplane at $-a$. Silvey's condition (1) is therefore satisfied for this H .

If H does not satisfy the condition of the theorem, then let $\{l_1, \dots, l_q\}$ be the columns of L . The hyperplane $\{x | x = a + \sum_{i=1}^q \lambda_i l_i, \lambda_i \in \mathbb{R}\}$ cannot be a supporting hyperplane of $CH(\mathcal{X} \cup -\mathcal{X})$. If $aa^T + LL^T$ is singular, it is obvious that this H cannot satisfy Silvey's condition (1). Otherwise, if $aa^T + LL^T$ is nonsingular, by Lemma 1 again, there is therefore at least one x in \mathcal{X} such that either $a^T(aa^T + LL^T)^{-1}x - \sqrt{a^T(aa^T + LL^T)^{-1}a} > 0$ or $a^T(aa^T + LL^T)^{-1}x + \sqrt{a^T(aa^T + LL^T)^{-1}a} < 0$. This implies that this H cannot satisfy Silvey's condition (1).

REFERENCES

- Buonaccorsi, J.P. and Iyer, H.K. (1986). Optimal designs for ratios of linear combinations in the general linear model. *Jour.Statist.Plann.Infer.*13, 345-356.
- Chaloner, K. (1987). Using MACSYMA in optimal experimental design. *Proceedings of the 1987 Statistical Computing Section of the American Statistical Association*, 22-30.
- DeGroot, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill.
- Dette, H. (1993). Elfving's theorem for D-optimality. *Ann.Statist.*21, 767-771.
- Dette, H. (1996). A note on Bayesian c - and D -optimal designs in nonlinear regression models. *Ann.Statist.*24, 1225-1234.
- Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann.Math.Statist.*23, 255-262.
- Fan, S. (1999). *Multivariate Optimal Designs*. PhD Thesis, University of Minnesota.
- 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, 569-583.
- Haines, L.M. (1993). Optimal design for nonlinear regression models. *Commun.Statist. - Theory and Methods* 22, 1613-1627.

- Haines, L.M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J.R.Statist.Soc.B* 57, 575-598.
- Pukelsheim, F. (1980). On linear regression designs which maximize information. *J. Statist.Plann.Infer.* 4, 339-364.
- Pukelsheim, F. and Titterington D.M. (1983). General differential and lagrangian theory for optimal experimental design. *Ann.Statist.*11, 1060-1068.
- Searle, S.R. (1971). *Linear Models*. New York, Wiley.
- Silvey, S.D. (1978). Optimal design measures with singular information matrices. *Biometrika* 65, 553-559.
- Silvey, S.D. (1980). *Optimal Design*. London, Chapman and Hall.
- Smith, D. and Ridout, M. (2000). Local and optimal designs for criterion involving log(potency) in comparative bioassays. *Manuscript*.
- Studden, W.J. (1971). Elfving's theorem and optimal designs for quadratic loss. *Ann.Math.Statis.*42, 1613-1621.
- Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. In Y. Dodge, V.V. Fedorov, H.P. Wynn, Eds., *Optimal Design and Analysis of Experiments*, North-Holland, Amsterdam, 213-224.

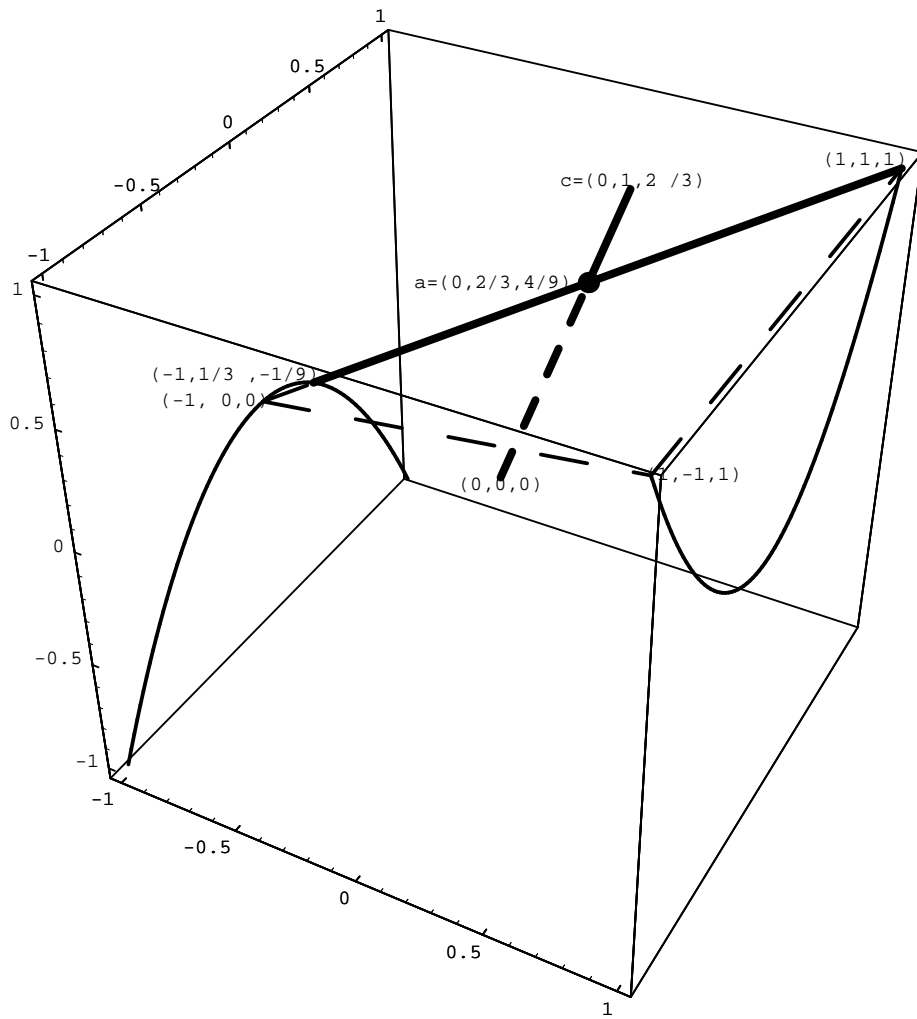


Figure 1: The set \mathcal{X} and the set $-\mathcal{X}$ for Example 1, together with the line joining x_1 to x_2 and the vector c and the point a for $f = \frac{1}{3}$.

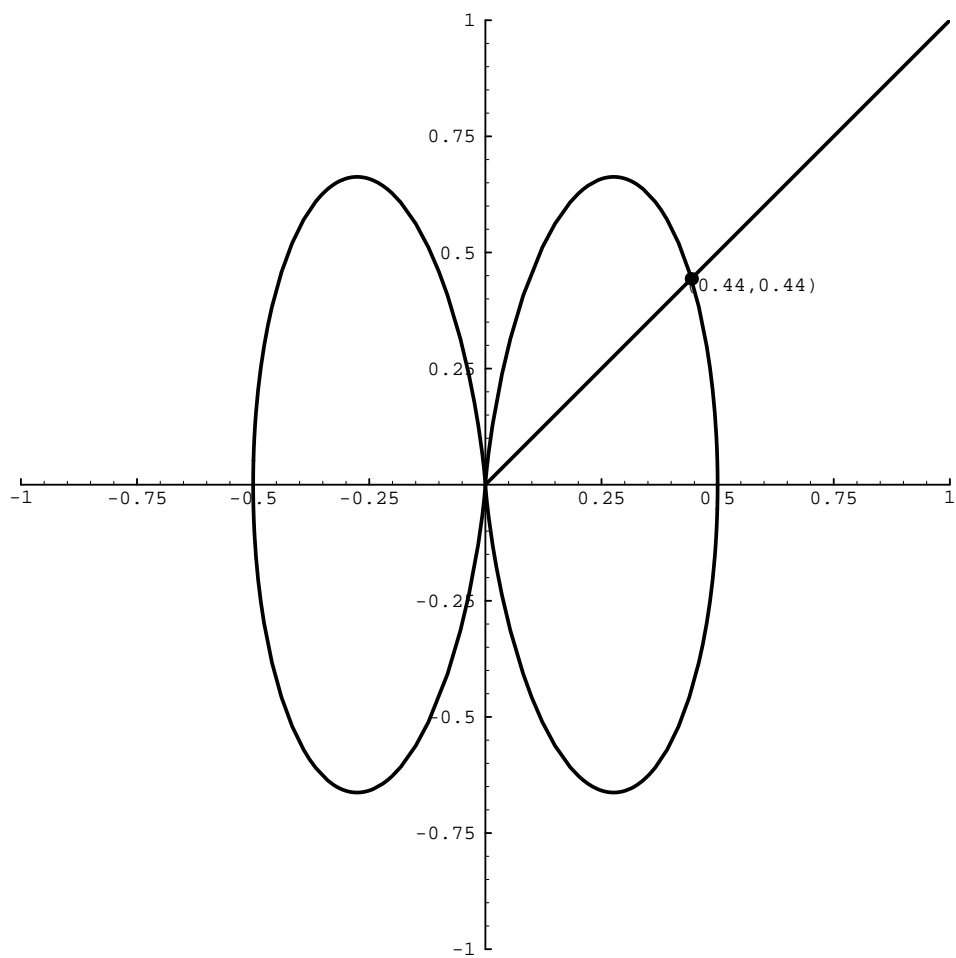


Figure 2: The set \mathcal{X} and the set $-\mathcal{X}$ for Example 3, together with the ray joining 0 to c and the point a .