Characterization of the Smoothness and Curvature of a Marginal Function for a Trust-Region Problem

David A. Andrews
Luís Nunes Vicente

May 1996
(revised July 1997)

TR96-13

Characterization of the Smoothness and Curvature of a Marginal Function for a Trust-Region Problem

David A. Andrews * Luís N. Vicente †

Abstract

This paper studies the smoothness and curvature of a marginal function for a trust-region problem. In this problem, a quadratic function is minimized over an ellipsoid. The marginal function considered is obtained by perturbing the trust radius, i.e., by changing the size of the ellipsoidal constraint. The values of the marginal function and of its first and second derivatives are explicitly calculated in all possible scenarios. A complete study of the smoothness and curvature of this marginal function is given. The main motivation for this work arises from an application in Statistics.

Keywords. Marginal or value function, perturbation or sensitivity analysis, trust regions

AMS subject classifications. 65U05, 90C20, 90C30, 90C31

1 Introduction

Consider the following minimization problem:

$$\begin{array}{ll} \text{minimize} & q(s) \;\equiv\; g^\top s + \frac{1}{2} s^\top H s \\ \text{subject to} & \|s\| \le \Delta\,, \end{array} \tag{1.1}$$

where $\Delta \in \mathbb{R}^+$, s, $g \in \mathbb{R}^n$, $H \in \mathbb{R}^{n\times n}$, $H = H^\top$, and n is a positive integer. The function $\|\cdot\|$ denotes the Euclidean ℓ_2 norm. The marginal function for this problem that we consider in this paper is defined by

$$v \,:\, [0, +\infty) \longrightarrow \mathbb{R}\,, \quad v(\Delta) = \min\,\{\, q(s) \,:\, \|s\| \le \Delta \,\}\,. \tag{1.2}$$

The marginal function is known also in the literature as the value function, the extreme-value function, the perturbation function, or the optimal value function. See the papers [2], [3], [4], [7], [8], [17], [18], [19] and the references therein. In nonlinear optimization the problem (1.1) appears in the globalization of Newton and quasi-Newton algorithms, and it is usually called the trust-region subproblem [13].

*Department of Statistics, Rice University, Houston, Texas 77005–1892, USA. E-Mail: andrewsd@stat.rice.edu.

†Departamento de Matemática, Universidade de Coimbra, 3000 Coimbra, Portugal. E-Mail: lvicente@mat.uc.pt.
Support of this author has been provided by INVOTAN (NATO scholarship), CCLA (Fulbright scholarship), FLAD, and NSF cooperative agreement CCR–9120008.

We give a complete characterization of the smoothness and curvature of this marginal function. We are motivated by an interesting application in Statistics (see Section 5), where the solution of an equation involving $v(\Delta)$ is required. In order to apply Newton's method to solve such an equation, we need to know the value of the first derivative $v'(\Delta)$. But the mathematical questions we answer in this paper go far beyond this point. We show that the marginal function is continuously differentiable in its domain. There are two possible scenarios in which the first derivative is not differentiable at specific points. If we exclude these undesirable points, then the marginal function has infinite continuous derivatives. We also show that the marginal function can be either convex, concave, or both. We feel that these results are quite surprising and confirm the elegance of trust regions.

This paper is structured as follows. We start in Section 2 by applying the sensitivity theory developed for nonlinear programming. However, this is clearly not enough to answer our mathematical questions. So we explicitly calculate formulae for the marginal function and its first and second derivatives. These calculations are described in great detail in Section 3, where we provide a full characterization of the smoothness of the marginal function. Section 4 characterizes the curvature of the marginal function. In Section 5 we discuss the application in Statistics. The reader may skip Section 2 and start with the analysis of Section 3.

2 Applying sensitivity theory for nonlinear programming

The marginal function in nonlinear programming has been studied by, among others, Gauvin [2], Gauvin and Dubeau [3], Gauvin and Tolle [4], Hogan [7], Janin [8], Rockafellar [17], Seeger [18], and Shapiro [19].

In [3], the authors consider marginal functions arising from perturbations of the data of the left and right sides of the constraints and of the objective function. The work in [2], [4] applies directly to our context since the marginal function depends only on perturbations of the right side of the constraints. The nonlinear program considered in [2], [4] is of the form:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq y_i \, , \; i = 1, \ldots, n_1 \, , \\ & h_j(x) = y_j \, , \; j = n_1 + 1, \ldots, n_1 + n_2 \, , \end{array}$$

where the y's are perturbation parameters.

A direct application of Theorem 2 and Corollary 1 in [2] yields the following result. We point out that the regularity conditions R1 and R2 considered in [2] are trivially satisfied by problem (1.1).

Theorem 2.1 *The marginal function $v(\Delta)$ defined in (1.2) is Lipschitz continuous in $(0, +\infty)$. Furthermore, the marginal function has left and right derivatives in $(0, +\infty)$.*

The following result is an application of the Corollary of Theorem 4 in [17]. (Again, we point out that the conditions and constraint qualifications required to apply this result are trivially satisfied by problem (1.1).) This result provides a formula for the derivative of the marginal function $v(\Delta)$ in the case where problem (1.1) has a unique solution and a corresponding unique Lagrange multiplier.

Theorem 2.2 *Assume for a given Δ, that problem (1.1) has a unique optimal solution $s(\Delta)$ with a unique Lagrange multiplier $\lambda(\Delta)$. Then the marginal function $v(\Delta)$ defined in (1.2) is differentiable at Δ and*

$$v'(\Delta) = -\lambda(\Delta)\Delta.$$

From the results in [8], we know also that the marginal function $v(\Delta)$ given by (1.2) has a right derivative at $\Delta = 0$. Thus this marginal function is Lipschitz continuous in $[0, +\infty)$.

A general characterization of the second-order directional derivatives of the marginal function in nonlinear programming is given in the papers [18], [19]. These results have some implications for problem (1.1), but we will omit them since they are too general and they do not cover all possible scenarios in (1.1).

3 Characterization of the smoothness of the marginal function

In this section we give a complete characterization of the smoothness of the marginal function (1.2). We also provide formulae for the first and second derivatives. The analysis uses the properties of the trust-region problem given by the following two propositions.

Proposition 3.1 *The problem (1.1) has no solutions at the boundary $\{s : \|s\| = \Delta\}$ if and only if H is positive definite and $\|H^{-1}g\| < \Delta$.*

A proof of this simple fact can be found in [15].

Proposition 3.2 *The point $s(\Delta)$ is an optimal solution of the problem (1.1) if and only if $\|s(\Delta)\| \leq \Delta$ and there exists $\lambda(\Delta) \geq 0$ such that*

$$H + \lambda(\Delta)I \quad \textit{is positive semi-definite}, \tag{3.1}$$

$$\Big(H + \lambda(\Delta)I\Big)s(\Delta) = -g\,, \quad \textit{and} \tag{3.2}$$

$$\lambda(\Delta)\,(\Delta - \|s(\Delta)\|) = 0\,. \tag{3.3}$$

The optimal solution $s(\Delta)$ is unique if $H + \lambda(\Delta)I$ is positive definite.

The sufficient part of these conditions was independently discovered by Gay [5] and Sorensen [20]. The necessary part of these conditions is just an application of the first-order and second-order necessary optimality conditions for nonlinear programming. These conditions were independently discovered by Karush [9] and Kuhn and Tucker [10] and are usually called the Karush-Kuhn-Tucker (KKT) conditions. The parameter $\lambda(\Delta)$ is the Lagrange multiplier associated with the trust-region constraint $\|s\| \leq \Delta$.

From the KKT condition (3.2) we can write

$$\begin{aligned} v(\Delta) &= g^\top s(\Delta) + \frac{1}{2}s(\Delta)^\top H s(\Delta) \\ &= -s(\Delta)^\top \Big(H + \lambda(\Delta)I\Big)s(\Delta) + \frac{1}{2}s(\Delta)^\top H s(\Delta) \\ &= -\frac{1}{2}s(\Delta)^\top \Big(H + \lambda(\Delta)I\Big)s(\Delta) - \frac{1}{2}\lambda(\Delta)\|s(\Delta)\|^2\,. \end{aligned} \tag{3.4}$$

3.1 Computing the derivatives

We consider three cases separately corresponding to three different situations: (i) H is positive definite; (ii) H is not positive definite but the hard case does not occur; (iii) H is not positive definite and the hard case occurs.

For the analysis, we consider the eigenvalue decomposition of H,

$$H = Q \Lambda Q^\top ,$$

where Λ is the diagonal matrix formed by the eigenvalues $\lambda_1, \ldots, \lambda_n$ of H, and Q is orthogonal and contains the corresponding eigenvectors. Let λ_1 be the smallest eigenvalue of H and assume that it has multiplicity m_1:

$$\lambda_1 = \cdots = \lambda_{m_1} < \lambda_{m_1+1} \leq \cdots \leq \lambda_n .$$

The subspace $E(\lambda_1) = \{s : Hs = \lambda_1 s\}$ is the eigenspace corresponding to the smallest eigenvalue λ_1. Also let

$$\bar{g} = Q^\top g .$$

Case I. H is positive definite

Let us define Δ^* as

$$\Delta^* = \|H^{-1} g\|.$$

Since in this case $q(s)$ is strictly convex, $-H^{-1}g$ is its unconstrained minimizer. Thus, if $\Delta > \Delta^*$, then $s(\Delta) = -H^{-1}g$ is the optimal solution of problem (1.1). In this case we have

$$\left.\begin{array}{rcl} v(\Delta) & = & -\frac{1}{2} g^\top H^{-1} g \, , \\ v'(\Delta) & = & 0 \, , \\ v''(\Delta) & = & 0 \, , \end{array}\right\} \quad \text{for } \Delta > \Delta^* .$$

If $0 < \Delta < \Delta^*$, then we know from Propositions 3.1 and 3.2 that the optimal solution $s(\Delta)$ of problem (1.1) satisfies

$$\begin{array}{l} s(\Delta) = -\Big(H + \lambda(\Delta) I\Big)^{-1} g \, , \\ \|s(\Delta)\| = \Delta \, . \end{array} \tag{3.5}$$

Using this and the expression (3.4) for $v(\Delta)$ we obtain

$$v(\Delta) = -\frac{1}{2} g^\top \Big(H + \lambda(\Delta) I\Big)^{-1} g - \frac{1}{2} \lambda(\Delta) \Delta^2 . \tag{3.6}$$

To calculate $v'(\Delta)$, we note that (3.5) implies

$$g^\top \Big(H + \lambda(\Delta) I\Big)^{-2} g = \Delta^2 . \tag{3.7}$$

From the eigenvalue decomposition of H, this equation is equivalent to

$$\sum_{i=1}^{n} \frac{\bar{g}_i^2}{(\lambda_i + \lambda(\Delta))^2} = \Delta^2 .$$

By taking derivatives in both sides of this equation we obtain

$$\lambda'(\Delta) = -\frac{\Delta}{g^\top \Big(H + \lambda(\Delta)I\Big)^{-3} g}. \tag{3.8}$$

Now we use (3.6) and (3.7) to calculate the first derivative of $v(\Delta)$:

$$\begin{aligned} v'(\Delta) &= \tfrac{1}{2} g^\top \Big(H + \lambda(\Delta)I\Big)^{-2} g\lambda'(\Delta) - \lambda(\Delta)\Delta - \tfrac{1}{2}\lambda'(\Delta)\Delta^2 \\ &= -\lambda(\Delta)\Delta\,. \end{aligned}$$

Using (3.7) and (3.8), we can calculate the second derivative of $v(\Delta)$ for $0 < \Delta < \Delta^*$:

$$\left.\begin{aligned} v(\Delta) &= -\tfrac{1}{2} g^\top \Big(H + \lambda(\Delta)I\Big)^{-1} g - \tfrac{1}{2}\lambda(\Delta)\Delta^2\,, \\ v'(\Delta) &= -\lambda(\Delta)\Delta\,, \\ v''(\Delta) &= -\lambda(\Delta) + \frac{\Delta^2}{g^\top \Big(H+\lambda(\Delta)I\Big)^{-3} g}\,, \end{aligned}\right\} \text{ for } \ 0 < \Delta < \Delta^*. \tag{3.9}$$

The functions $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ are plotted in Figure 1 for an example where H is positive definite. The special cases $\Delta = 0$ and $\Delta = \Delta^*$ are considered in Sections 3.2 and 3.3, respectively. The plots in Figures 1 through 4 were obtained using MATLAB 5.0. To solve the trust-region problem, the `Fortran 77` subroutine `dgqt.f` of `Minpack-2` was called through a `MEX` interface for MATLAB. This `Minpack-2` subroutine is available by anonymous ftp in `info.mcs.anl.gov` under the directory `/pub/MINPACK-2/gqt`.

Case II. H not positive definite — easy case

We consider now the case where g is not orthogonal to $E(\lambda_1)$. In this case, it follows from the geometry of the rational function

$$h(\lambda) \;=\; g^\top \Big(H + \lambda I\Big)^{-2} g \;=\; \sum_{i=1}^{n} \frac{\bar{g}_i^2}{(\lambda_i + \lambda)^2}$$

that $-\lambda_1$ is its rightmost pole. For any value of $\Delta > 0$, there is always a $\lambda(\Delta) > -\lambda_1$ such that $h(\lambda)$ intersects Δ^2 (see Figure 2). Thus $s(\Delta) = -\Big(H + \lambda(\Delta)I\Big)^{-1} g$ and $\lambda(\Delta)$ satisfy all the necessary and sufficient conditions given in Proposition 3.2. Thus, this case reduces to the case where H is positive definite and $\Delta < \Delta^*$. For any $\Delta \in (0, +\infty)$, the expressions for $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ are given as in (3.9). See Figure 3 for plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ corresponding to an example in this case. The case $\Delta = 0$ is analyzed separately in Section 3.2.

Case III. H not positive definite — hard case

It remains to consider the case where H is not positive definite and g is orthogonal to $E(\lambda_1)$. In this case the rational function $h(\lambda)$ has the form

$$h(\lambda) \;=\; g^\top \Big(H + \lambda I\Big)^{-2} g \;=\; \sum_{i=m_1+1}^{n} \frac{\bar{g}_i^2}{(\lambda_i + \lambda)^2}\,,$$

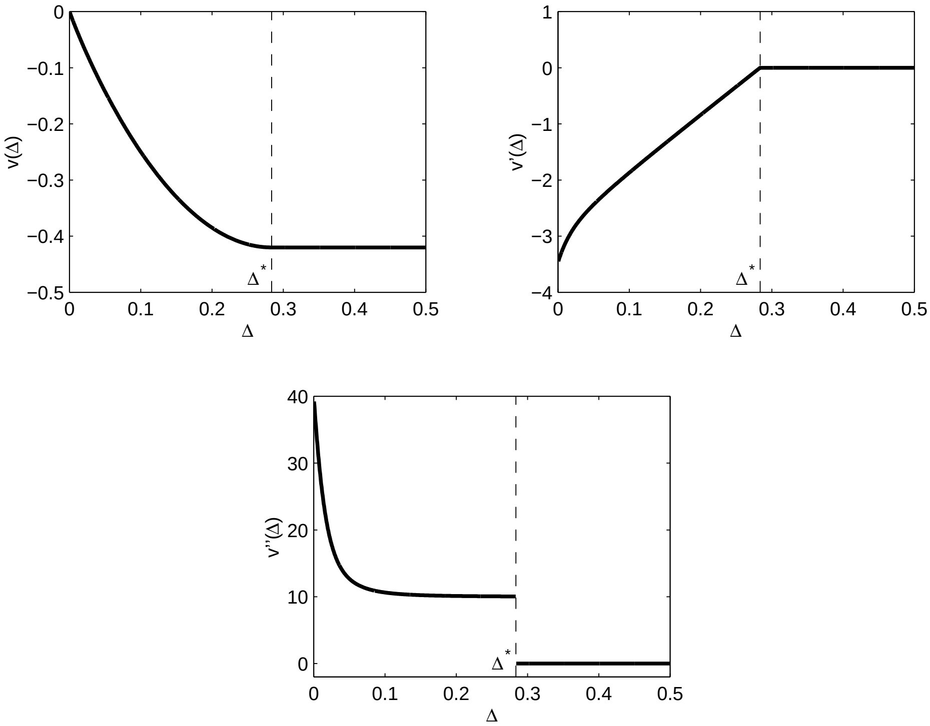

Figure 1: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([\,10 \;\; 10 \;\; 100\,])$, and $g = [\,2\;2\;2\,]'$. Here H is positive definite.

and $-\lambda_1$ is no longer a pole for $h(\lambda)$. Two situations can happen here depending on the value of Δ.

Let $\Delta^{\#}$ be such that $-\lambda_1$ is the rightmost solution for $h(\lambda) = (\Delta^{\#})^2$. In other words let,

$$\Delta^{\#} = \left(\sum_{i=m_1+1}^{n} \frac{\bar{g}_i^2}{(\lambda_i - \lambda_1)^2} \right)^{\frac{1}{2}} .$$

If $0 < \Delta < \Delta^{\#}$, then the rightmost solution $\lambda(\Delta)$ of $h(\lambda) = \Delta^2$, is such that $\lambda(\Delta) > -\lambda_1$ and $H + \lambda(\Delta)I$ is positive definite. Hence this situation is identical to the previous one. The conclusion is that $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for $\Delta < \Delta^{\#}$ are given as in (3.9).

If $\Delta > \Delta^{\#}$, then any solution for $h(\lambda) = \Delta^2$ is such that $H + \lambda I$ is not positive definite. As result we cannot calculate $\lambda(\Delta)$ by solving $h(\lambda) = \Delta^2$. It is shown in [20] that in this case a solution for the problem (1.1) can be calculated by finding the $\tau(\Delta)$ such that

$$\|p + \tau(\Delta)q\| \;=\; \Delta \,,$$

where p solves $\Big(H - \lambda_1 I\Big)p = -g$ and q is any nonzero vector in $E(\lambda_1)$. The solution $s(\Delta)$ is given by $p + \tau(\Delta)q$ and $\lambda(\Delta) = -\lambda_1$.

We can use this expression for $s(\Delta)$ and the form (3.4) of the marginal function to calculate a

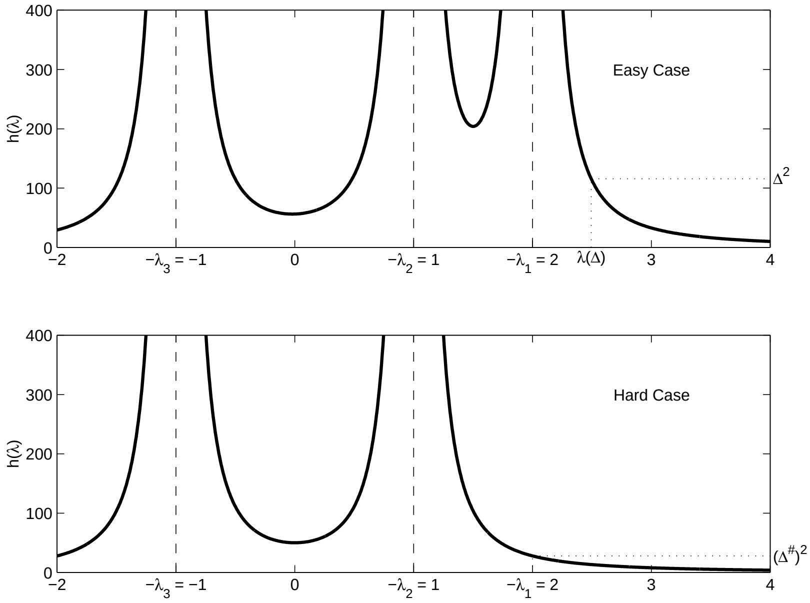

Figure 2: Plot of $h(\lambda)$ for the case $n = 3$, $H = diag([\,-2\ -1\ 1\,])$, $g = [\,5\ 5\ 5\,]'$ (Easy Case), and $g = [\,0\ 5\ 5\,]'$ (Hard Case).

formula for $v(\Delta)$ and its derivatives. In fact we have,

$$\begin{aligned} v(\Delta) &= -\tfrac{1}{2}(p+\tau(\Delta)q)^\top\Big(H-\lambda_1 I\Big)(p+\tau(\Delta)q)+\tfrac{1}{2}\lambda_1\Delta^2 \\ &= -\tfrac{1}{2}p^\top\Big(H-\lambda_1 I\Big)p+\tfrac{1}{2}\lambda_1\Delta^2\,. \end{aligned}$$

Thus

$$\left.\begin{aligned} v(\Delta) &= -\tfrac{1}{2}p^\top\Big(H-\lambda_1 I\Big)p+\tfrac{1}{2}\lambda_1\Delta^2\,, \\ v'(\Delta) &= \lambda_1\Delta\,, \\ v''(\Delta) &= \lambda_1\,, \end{aligned}\right\}\quad \text{for } \Delta > \Delta^{\#}\,.$$

The functions $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ are depicted in Figure 4 for an example in this case. Again, the special cases $\Delta = 0$ are $\Delta = \Delta^{\#}$ are considered in Sections 3.2 and 3.3, respectively.

The next theorem characterizes the smoothness of the marginal function $v(\Delta)$ in most of its domain. Its proof is a direct consequence of the analysis carried out for the last three cases. This analysis is summarized in Table 1.

Theorem 3.1

If H is positive definite, the marginal function $v(\Delta)$ defined in (1.2) is C^∞ in $(0,+\infty)\setminus\{\Delta^*\}$.

If H is not positive definite and g is not orthogonal to $E(\lambda_1)$, the marginal function is C^∞ in $(0,+\infty)$.

If H is not positive definite, but g is orthogonal to $E(\lambda_1)$, the marginal function is C^∞ in $(0,+\infty)\setminus\{\Delta^{\#}\}$.

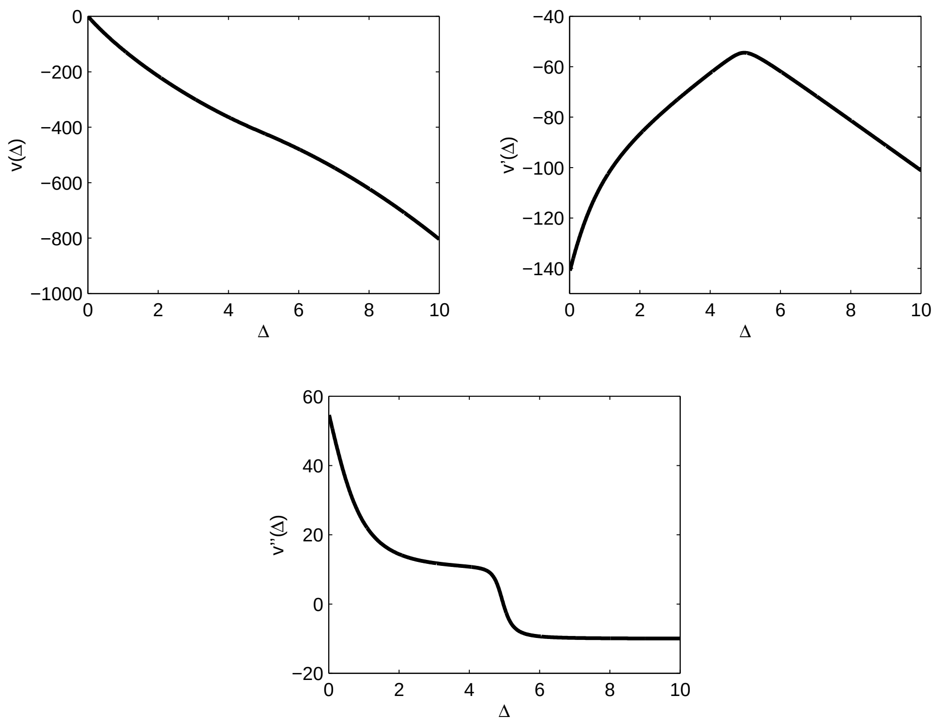

Figure 3: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([-10\ 10\ 100])$, and $g = [\,1\ 100\ 100\,]'$. Here H is not positive definite but we have the easy case.

3.2 Continuity of the derivatives at zero

We know from the analysis presented in Section 2 that $v(\Delta)$ is continuous at $\Delta = 0$ and $v'_+(0)$ exists. One can also show that the right derivatives $v'_+(\Delta)$ and $v''_+(\Delta)$ exist and are continuous at $\Delta = 0$. (The knowledge of the value of $v''_+(0)$ is important for the study of the curvature of $v(\Delta)$.) In fact, if Δ is sufficiently small, $v(\Delta)$ has the same expression (given by (3.9)) for each of the three cases analyzed in Section 3.1, and the same happens for $v'(\Delta)$. From the asymptotic behavior of $h(\lambda)$, we know that $\lambda \to +\infty$ as $\Delta \to 0^+$. Thus, from (3.6), (3.7), and $v(0) = 0$, we obtain

$$\begin{aligned}
v'_+(0) &= \lim_{\Delta\to 0^+} \frac{v(\Delta) - v(0)}{\Delta} \\
&= \lim_{\Delta\to 0^+} \frac{-\frac{1}{2}g^\top(H+\lambda(\Delta)I)^{-1}g - \frac{1}{2}\lambda(\Delta)\Delta^2}{\Delta} \\
&= \lim_{\lambda\to +\infty} \frac{-\frac{1}{2}g^\top(H+\lambda I)^{-1}g - \frac{1}{2}\lambda g^\top(H+\lambda I)^{-2}g}{(g^\top(H+\lambda I)^{-2}g)^{\frac{1}{2}}} \\
&= \lim_{\lambda\to +\infty} \frac{-\frac{1}{2}\lambda g^\top(H+\lambda I)^{-1}g - \frac{1}{2}\lambda^2 g^\top(H+\lambda I)^{-2}g}{(\lambda^2 g^\top(H+\lambda I)^{-2}g)^{\frac{1}{2}}}
\end{aligned}$$

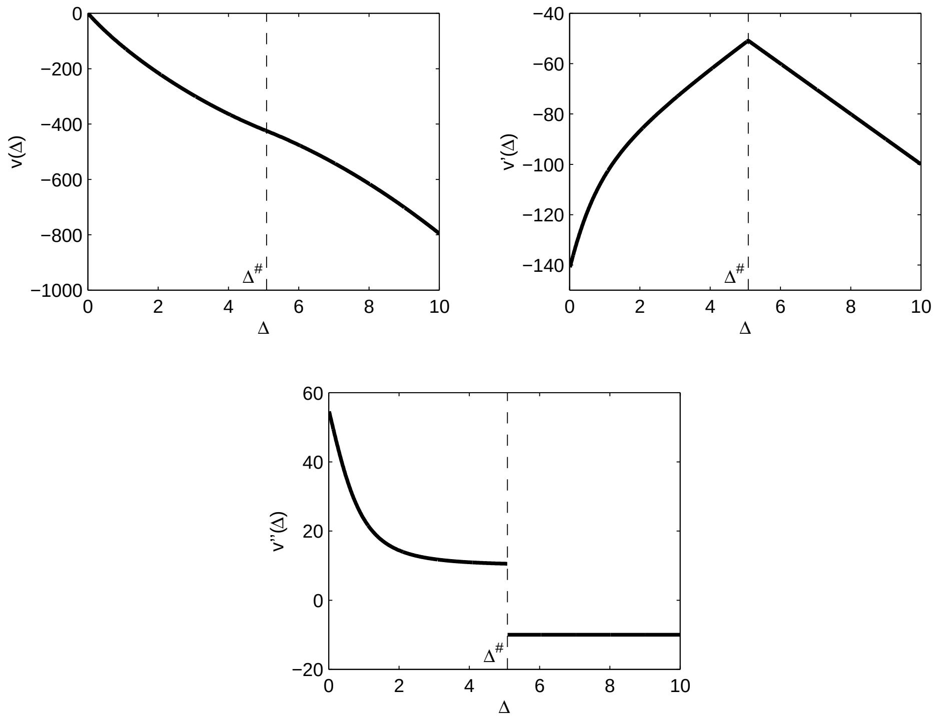

Figure 4: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([-10\ 10\ 100])$, and $g = [\,0\ 100\ 100\,]'$. Here H is not positive definite and the hard case occurs.

$$= \lim_{\lambda\to+\infty} \frac{-\frac{1}{2}\lambda\sum_{i=1}^{n}\frac{\bar{g}_i^2}{\lambda_i+\lambda} - \frac{1}{2}\lambda^2\sum_{i=1}^{n}\frac{\bar{g}_i^2}{(\lambda_i+\lambda)^2}}{\left(\lambda^2\sum_{i=1}^{n}\frac{\bar{g}_i^2}{(\lambda_i+\lambda)^2}\right)^{\frac{1}{2}}}$$

$$= -\left(\sum_{i=1}^{n}\bar{g}_i^2\right)^{\frac{1}{2}},$$

and

$$\lim_{\Delta\to0^+} v'_+(\Delta) = \lim_{\Delta\to0^+} -\lambda(\Delta)\Delta = \lim_{\lambda\to+\infty} -\lambda\left(g^\top\left(H+\lambda I\right)^{-2}g\right)^{\frac{1}{2}} = -\left(\sum_{i=1}^{n}\bar{g}_i^2\right)^{\frac{1}{2}}.$$

In conclusion,

$$v'_+(0) = \lim_{\Delta\to0^+} v'_+(\Delta) = -\left(\sum_{i=1}^{n}\bar{g}_i^2\right)^{\frac{1}{2}} = -\|g\|.$$

Using similar arguments one can show that

$$v''_+(0) = \lim_{\Delta\to0^+} v''_+(\Delta) = \frac{\sum_{i=1}^{n}\bar{g}_i^2\lambda_i}{\sum_{i=1}^{n}\bar{g}_i^2} = \frac{g^\top Hg}{g^\top g}. \tag{3.10}$$

Cases	Formulae for $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$
H PD $\Delta > \Delta^*$	$v(\Delta) = -\frac{1}{2}g^\top H^{-1}g$ $v'(\Delta) = 0$ $v''(\Delta) = 0$
H PD $0 < \Delta < \Delta^*$ H not PD Easy Case H not PD Hard Case $0 < \Delta < \Delta^\#$	$v(\Delta) = -\frac{1}{2}g^\top \Big(H + \lambda(\Delta)I\Big)^{-1} g - \frac{1}{2}\lambda(\Delta)\Delta^2$ $v'(\Delta) = -\lambda(\Delta)\Delta$ $v''(\Delta) = -\lambda(\Delta) + \frac{\Delta^2}{g^\top \left(H+\lambda(\Delta)I\right)^{-3} g}$
H not PD Hard Case $\Delta > \Delta^\#$	$v(\Delta) = -\frac{1}{2}p^\top \Big(H - \lambda_1 I\Big)p + \frac{1}{2}\lambda_1\Delta^2$ $v'(\Delta) = \lambda_1\Delta$ $v''(\Delta) = \lambda_1$

Table 1: Formulae for $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$.

3.3 Non-differentiability of the first derivative

The only points where $v(\Delta)$ and $v'(\Delta)$ can be non-differentiable, or differentiable but not continuously differentiable, are Δ^* in Case I and $\Delta^\#$ in Case III. So, we need to compute the left and right derivatives of $v(\Delta)$ and $v'(\Delta)$ at these points. In fact it can be shown that

$$\begin{aligned} v'_-(\Delta^*) &= v'_+(\Delta^*) = 0 \quad \text{and} \\ v'_-(\Delta^\#) &= v'_+(\Delta^\#) = \lambda_1\Delta^\# = \lambda_1 \left(\sum_{i=m_1+1}^{n} \frac{\bar{g}_i^2}{(\lambda_i - \lambda_1)^2} \right)^{\frac{1}{2}} . \end{aligned}$$

As result of this we have the following theorem.

Theorem 3.2 *The marginal function $v(\Delta)$ defined in (1.2) is continuously differentiable in* $[0, +\infty)$.

Similar calculations in the respective cases lead to:

$$
\begin{array}{rclcl}
v''_-(\Delta^*) & = & \lim_{\Delta\to\Delta^{*-}} v''(\Delta) & = & \frac{\sum_{i=1}^n \bar{g}_i^2/\lambda_i^2}{\sum_{i=1}^n \bar{g}_i^2/\lambda_i^3}\,, \\
v''_+(\Delta^*) & = & \lim_{\Delta\to\Delta^{*+}} v''(\Delta) & = & 0\,, \\
v''_-(\Delta^\#) & = & \lim_{\Delta\to\Delta^{\#-}} v''(\Delta) & = & \lambda_1 + \frac{\sum_{i=m_1+1}^n \bar{g}_i^2/(\lambda_i-\lambda_1)^2}{\sum_{i=m_1+1}^n \bar{g}_i^2/(\lambda_i-\lambda_1)^3}\,, \\
v''_+(\Delta^\#) & = & \lim_{\Delta\to\Delta^{\#+}} v''(\Delta) & = & \lambda_1\,.
\end{array}
\tag{3.11}
$$

Theorem 3.3

If H is positive definite, the marginal function $v(\Delta)$ defined in (1.2) is not twice differentiable at Δ^*.

If H is not positive definite, but g is orthogonal to $E(\lambda_1)$, the marginal function $v(\Delta)$ defined in (1.2) is not twice differentiable at $\Delta^\#$.

Although in these cases $v(\Delta)$ is not twice differentiable, it has a Schwartz second derivative (see [6, Section 5.3]) since $v''_-(\Delta^*) + v''_+(\Delta^*)$ and $v''_-(\Delta^\#) + v''_+(\Delta^\#)$ are finite.

4 Curvature of the marginal function

To study the curvature of the marginal function $v(\Delta)$ we look at the sign of the second derivative $v''(\Delta)$ in its domain. We already know what values $v''_+(\Delta)$ takes when $\Delta = 0$ (see (3.10)). We show next that $v''(\Delta)$ is monotone decreasing in its domain.

A simple derivation using (3.8) yields

$$
\left(-\lambda(\Delta) + \frac{\Delta^2}{g^\top\left(H+\lambda(\Delta)I\right)^{-3}g}\right)' \;=\; \frac{3\Delta\left(g^\top\left(H+\lambda(\Delta)I\right)^{-3}g\right)^2 - 3\Delta^3\left(g^\top\left(H+\lambda(\Delta)I\right)^{-4}g\right)}{\left(g^\top\left(H+\lambda(\Delta)I\right)^{-3}g\right)^3}\,.
$$

Since by the Cauchy-Schwartz inequality

$$
\begin{array}{rcl}
\left(g^\top\left(H+\lambda(\Delta)I\right)^{-3}g\right)^2 & = & \left[\left((H+\lambda(\Delta)I)^{-1}g\right)^\top\left((H+\lambda(\Delta)I)^{-2}g\right)\right]^2 \\
& \leq & \left(g^\top\left(H+\lambda(\Delta)I\right)^{-2}g\right)\left(g^\top\left(H+\lambda(\Delta)I\right)^{-4}g\right)\,,
\end{array}
$$

we have that

$$
\left(-\lambda(\Delta) + \frac{\Delta^2}{g^\top\left(H+\lambda(\Delta)I\right)^{-3}g}\right)' \;\leq\; 0\,.
$$

From this, from the expressions for $v''(\Delta)$ given in Table 1, and from the values for $v''_-(\Delta)$ and $v''_+(\Delta)$ given in equations (3.10) and (3.11), we immediately conclude that $v''(\Delta)$ is monotone decreasing in its domain.

The curvature of $v(\Delta)$ depends then on the signs of $v''_+(0)$ and $\lim_{\Delta\to+\infty} v''(\Delta)$. The value for $v''_+(0)$ is given in (3.10). To find $\lim_{\Delta\to+\infty} v''(\Delta)$ for all possible cases, all we need is the following calculation:

$$\lim_{\Delta\to+\infty}\left(-\lambda(\Delta)+\frac{\Delta^2}{g^\top\left(H+\lambda(\Delta)I\right)^{-3}g}\right) \;=\; \lambda_1\,,$$

for the case where H is not positive definite and the easy case occurs. This limit can be easily proved since from the asymptotic behavior of $h(\lambda)$, $\lambda \longrightarrow -\lambda_1$ when $\Delta \longrightarrow +\infty$. From Table 1, the conclusion is that

$$\lim_{\Delta\to+\infty} v''(\Delta) \;=\; 0\,,$$

if H is positive definite, and

$$\lim_{\Delta\to+\infty} v''(\Delta) \;=\; \lambda_1\,,$$

if H is not positive definite.

By collecting all possible situations, we obtain a complete characterization of the curvature of the marginal function $v(\Delta)$ defined in (1.2).

Theorem 4.1

If H is positive definite, the marginal function is convex in $[0,+\infty)$.

If H is not positive definite, we have several cases:

If $v''_+(0) \leq 0$, then the marginal function is concave in $[0,+\infty)$.

If $v''_+(0) > 0$, then either $\lambda_1 = 0$ and the marginal function is convex in $[0,+\infty)$, or $\lambda_1 < 0$ and there exists $\bar{\Delta} > 0$ such that the marginal function is convex in $[0,\bar{\Delta}]$ and concave in $[\bar{\Delta},+\infty)$.

5 A Statistical Problem

We describe in this section the application from statistics that motivated this work. The two-group gaussian discriminant analysis problem of statistics[1] assumes that there are two n-dimensional populations with gaussian distributions, for which we wish to find a method to determine and exploit differences of the two groups. In this work, we assume that the mean vectors, $\mu_1, \mu_2 \in \mathbb{R}^n$, and covariance matrices, $\Sigma_1, \Sigma_2 \in \mathbb{R}^{n\times n}$, are known, that $\mu_1 \neq \mu_2$, and that the populations are equally likely. The covariance matrices are assumed to be positive definite.

We will use the following property of the gaussian distribution: Suppose $X_i, i = 1,2$ are gaussian random vectors with the corresponding means and covariances given above. For a given $\Delta > 0$ we define the ellipsoidal sets

$$\mathcal{E}_i(\Delta) = \{x : (x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i) \leq \Delta\}, \quad i = 1,2. \tag{5.1}$$

[1]The books by McLachlan [11] and Ripley [16] have extended definitions and examples of the discriminant problem in statistics.

The gaussian distribution has the property that for any value for the means, covariances and Δ

$$P\left[X_1 \in \mathcal{E}_1(\Delta)\right] = P\left[X_2 \in \mathcal{E}_2(\Delta)\right]. \tag{5.2}$$

That is, the ellipsoids have equal size (in terms of probability under their respective distribution) as long as the ellipsoids are determined by the same Δ. Note that $\mathcal{E}_i(0) = \{\mu_i\}$ and that $\mathcal{E}_i(\Delta)$ grows to contain all of $\mathbb{R}^n$ as $\Delta \to \infty$.

One way to approach the discriminant analysis problem is to examine measures of separation. A standard measure of separation of the groups is given by the squared Mahalanobis distance between the means

$$m_i = (\mu_1 - \mu_2)^\top \Sigma_i^{-1} (\mu_1 - \mu_2). \tag{5.3}$$

If the covariance matrices are equal, then $m_1 = m_2$, and this distance is useful. However, if the covariance matrices are different, then the m_i's are not equal, and using one or the other would not be "fair" since it ignores the structure of the other group. Using property (5.2), a metric of separation can be constructed that takes into account the covariance structure of both groups:

$$\sup\{\Delta \geq 0 : \mathcal{E}_1(\Delta) \cap \mathcal{E}_2(\Delta) = \emptyset\}. \tag{5.4}$$

In words, this means that we are looking for the largest equal-probability ellipsoids that do not overlap. If we define

$$\kappa(\Delta) = \min\{(x - \mu_1)^\top \Sigma_1^{-1} (x - \mu_1) : (x - \mu_2)^\top \Sigma_2^{-1} (x - \mu_2) \leq \Delta\}, \tag{5.5}$$

then the solution to the nonlinear equation

$$\kappa(\Delta) = \Delta \tag{5.6}$$

is also the solution to problem (5.4). Problem (5.5) is easily reparameterized into the form (1.2) (see Section 6). Thus, we can use the results of this paper to calculate the derivative of (5.5) and so solve (5.6) using Newton's method.

This method was implemented in the S-PLUS statistical software package, using the routine `dgqt.f` from `Minpack-2` to solve the trust-region subproblem. Using the quantities defined in equation (5.3), the initial point used for Newton's method is

$$\Delta_0 = \frac{m_1 m_2}{m_1 + m_2}. \tag{5.7}$$

This is obtained from linearly interpolating the function κ at the points $\Delta = 0$ and $\Delta = m_2$, using the fact that $\kappa(0) = m_1$ and $\kappa(m_2) = 0$. This interpolant, $\bar{\kappa}$, is then used to get Δ_0 by solving $\bar{\kappa}(\Delta) = \Delta$. Note that the results of this paper can be used to get derivatives of κ at $\Delta = 0$ and $\Delta = m_2$ in terms of the mean vectors and covariance matrices. Thus higher order interpolants could be used in calculating Δ_0. However, our experiments have shown that Δ_0 given by equation (5.7) is a sufficiently good starting point.

We present results in Table 2 on three datasets obtained from the Machine Learning Data Repository at the University of California at Irvine [12]. We used the maximum likelihood estimates of the mean and covariance as the inputs to the routine. For dataset `iris`, we used the Virginica and Versicolor groups of Fisher's iris data [1], a classic dataset in discriminant analysis. This data consists of four measurements on fifty irises from each of two varieties. The `wdbc` (Wisconsin

Diagnostic Breast Cancer) dataset consists of ten mean measurements of cell nuclei taken from a breast mass that is independently assigned to either a benign or malignant group. There are 357 benign and 212 malignant observations. The `wine` dataset consists of 13 chemical measurements of wine made from two different cultivars of grape grown in the same region of Italy. There are 59 observations in the first group and 71 in the second.

Dataset	Dimension	No. of Iterations
`iris`	4	7
`wdbc`	10	10
`wine`	13	7

Table 2: Summary of results for application of Newton's method in solving the statistical distance problem (5.5)-(5.6). The dimension of the data is given, along with the number of iterations of Newton's method needed to achieve $|\kappa(\Delta) - \Delta| \leq 10^{-14}$.

If Δ is the solution of the problem (5.5)-(5.6), there is a unique x such that

$$(x - \mu_1)^\top \Sigma_1^{-1}(x - \mu_1) = (x - \mu_2)^\top \Sigma_2^{-1}(x - \mu_2) = \Delta,$$

which is same as saying $\mathcal{E}_1(\Delta)$ and $\mathcal{E}_2(\Delta)$ are tangent at x, and so share a common $(n-1)$-dimensional tangent plane, P. If w is a vector orthogonal to P, we may project the data into $\mathbb{R}$ along w. This projection preserves the separation of the two ellipsoids, and so also preserves the separation of the data. The gaussian distribution is conserved under projections, so that we may also project the data into P and then repeat the entire process in $n-1$ dimensions, obtaining a second projection of the data into $\mathbb{R}$. Combining these two projections we may look at the data in $\mathbb{R}^2$, and the data should be well separated. Thus the method described here is useful for data visualization. Such a procedure was used on the `iris` dataset to generate Figure 5.

6 Final remarks

The results given in this paper can be extended for more general forms of the trust-region problem (1.1). For instance, any trust-region problem of the form

$$\begin{array}{ll} \text{minimize} & g^\top \bar{s} + \frac{1}{2}\bar{s}^\top H \bar{s} \\ \text{subject to} & \|W\bar{s}\| \leq \Delta\,, \end{array} \tag{6.1}$$

where W has n linearly independent columns, is equivalent to (1.1). The change of variables $s = (W^\top W)^{\frac{1}{2}}\bar{s}$ reduces this trust-region problem to (1.1).

Moré [14] considers another generalization of the form

$$\begin{array}{ll} \text{minimize} & g^\top s + \frac{1}{2}s^\top H s \\ \text{subject to} & c(s) \leq \Delta\,, \end{array}$$

where $c(s) = d^\top s + \frac{1}{2}s^\top C s$ is another quadratic function. A characterization of smoothness and curvature similar to the one given in this paper could be easily developed for the marginal function associated with this problem.

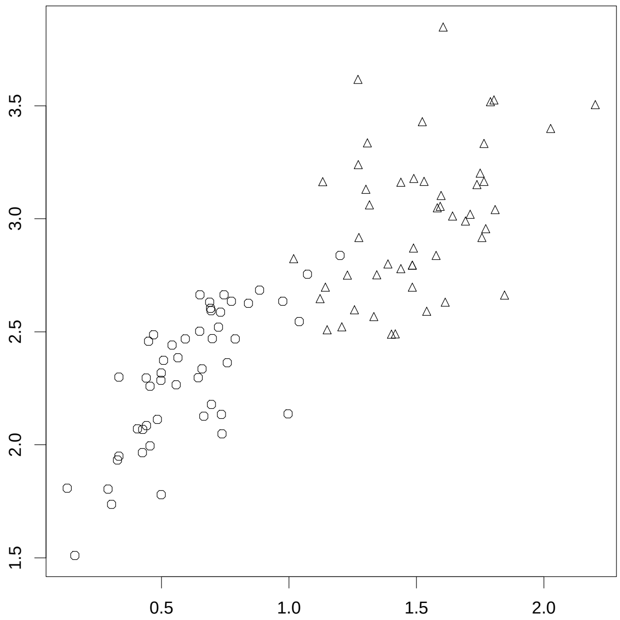

Figure 5: Projection of the `iris` dataset into $\mathbb{R}^2$ using the method described in Section 5. The two varieties are show as triangles and circles respectively.

References

[1] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, Ann. Eugen., 7 (1936), pp. 179–188.

[2] J. Gauvin, *The generalized gradient of a marginal function in mathematical programming*, Math. Oper. Res., 4 (1979), pp. 458–463.

[3] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*, Math. Programming Stud., 19 (1982), pp. 101–119.

[4] J. Gauvin and J. W. Tolle, *Differential stability in nonlinear programming*, SIAM J. Control Optim., 15 (1977), pp. 294–311.

[5] D. M. Gay, *Computing optimal locally constrained steps*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 186–197.

[6] J.-B. Hiriart-Urruty and C. Lamaréchal, *Convex Analysis and Minimization Algorithms I*, Springer–Verlag, Berlin, 1993.

[7] W. W. Hogan, *Directional derivatives for extreme–value functions with applications to the completely convex case*, Oper. Res., 21 (1973), pp. 188–209.

[8] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, Math. Programming Stud., 21 (1984), pp. 110–126.

[9] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Constraints*, master's thesis, Department of Mathematics, University of Chicago, 1939.

[10] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, 1951.

[11] G. J. McLachlan, *Discriminant Analysis and Statisical Pattern Recognition*, John Wiley & Sons, New York, 1992.

[12] C. J. Merz and P. M. Murphy, *U. C. I. repository of machine learning databases*, University of California, Department of Information and Computer Science, Irvine, CA, 1996. [`http://www.ics.uci.edu/~mlearn/MLRepository.html`].

[13] J. J. Moré, *Recent developments in algorithms and software for trust regions methods*, in Mathematical programming. The state of art, A. Bachem, M. Grotschel, and B. Korte, eds., Springer Verlag, New York, 1983, pp. 258–287.

[14] ——, *Generalizations of the trust region problem*, Optimization Methods and Software, 2 (1993), pp. 189–209.

[15] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[16] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, London, New York, 1996.

[17] R. T. Rockafellar, *Directional differentiability of the optimal value function in a nonlinear programming problem*, Math. Programming Stud., 21 (1984), pp. 213–226.

[18] A. Seeger, *Second order directional derivatives in parametric optimization problems*, Math. Oper. Res., 13 (1988), pp. 124–139.

[19] A. Shapiro, *Second–order derivatives of extreme–value functions and optimality conditions for semi–infinite programs*, Math. Oper. Res., 10 (1985), pp. 207–219.

[20] D. C. Sorensen, *Newton's method with a model trust region modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426.