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

David A. Andrews
Luís Nunes Vicente

CRPC-TR96649
May 1996

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

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

David A. Andrews
Luís Nunes Vicente

May 1996

TR96-13

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

David A. Andrews *   Luís N. Vicente †

May 16, 1996

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:

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

where $\Delta \in \mathbb{R}^+$, $s, g \in \mathbb{R}^n$, $H \in \mathbb{R}^{n\times n}$, $H = H^T$, 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 as

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

The marginal function is known also in the literature as value function, extremed–value function, perturbation function, or optimal value function. See the papers [2], [3], [4], [7], [8], [14], [15], [16] and the references therein.

*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. This work was developed while the author was a graduate student at the Department of Computational and Applied Mathematics of Rice University. 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 5.1), 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 values of the first derivative $v'(\Delta)$. But the mathematical questions we raise and answer in this paper go far beyond this point. We show that the marginal function is continuously differentiable in its domain. However, 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. To us, these results are quite surprising and confirm the elegance of trust regions.

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 [11]. The real number Δ is called the trust radius. For simplicity consider the unconstrained minimization problem

$$\text{minimize} \quad f(x)\,, \tag{1.3}$$

where f is a twice continuously differentiable function mapping $\mathbb{R}^n$ into $\mathbb{R}$. If a globalization using trust regions is considered, a step s_k for a quasi–Newton algorithm is an approximate solution of the trust–region subproblem

$$\begin{array}{ll} \text{minimize} & g_k^T s + \frac{1}{2} s^T H_k s \\ \text{subject to} & \|s\| \le \Delta_k\,, \end{array} \tag{1.4}$$

where $g_k \simeq \nabla f_k$ and $H_k \simeq \nabla^2 f_k$. In a situation where s_k cannot be accepted as a step, there is the need to solve a new trust–region subproblem that differs from the previous one only in the value of the trust radius. See Section 5.2 for more details.

The numerical experimentations reported in this paper consisted of using `MATLAB` to solve a sequence of trust–region problems. To solve each trust-region problem, the `Fortran 77` subroutine `dgqt.f` of `Minpack 2` is 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`.

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 applications in Statistics and Optimization.

2 Applying sensitivity theory for nonlinear programming

The marginal function in nonlinear programming has been studied among others by Gauvin [2], Gauvin and Dubeau [3], Gauvin and Tolle [4], Hogan [7], Janin [8], Rockafellar [14], Seeger [15], and Shapiro [16].

In [3], the authors consider perturbations in the left and right sides of the constraints as well as in 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 programming 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 for 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 [14]. (Again, we point out that the conditions and constraint qualifications required to apply this result are trivially satisfied for problem (1.1).) This results provides a formula for the derivative of the marginal function $v(\Delta)$ in the case where problem (1.1) has a unique solution and unique corresponding Lagrange multiplier.

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

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

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 [15], [16]. 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 [13].

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*

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

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

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

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

For a proof of this result see Gay [5], or Sorensen [17]. 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

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

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,

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

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

$$\bar{g} = Q^T g\,.$$

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

$$\left.\begin{array}{rcl} v(\Delta) & = & -\frac{1}{2}g^T H^{-1} g\,, \\ v'(\Delta) & = & 0\,, \\ v''(\Delta) & = & 0\,, \end{array}\right\} \text{ for } \Delta > \Delta^*.$$

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

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

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

$$v(\Delta) = -\frac{1}{2}g^T\left(H + \lambda(\Delta)I\right)^{-1} g - \frac{1}{2}\lambda(\Delta)\Delta^2\,. \tag{3.6}$$

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

$$g^T\left(H + \lambda(\Delta)I\right)^{-2} g = \Delta^2\,. \tag{3.7}$$

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

$$\frac{d\lambda(\Delta)}{d\Delta} = -\frac{\Delta}{g^T\left(H + \lambda(\Delta)I\right)^{-3} g}\,. \tag{3.8}$$

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

$$\begin{array}{rcl} v'(\Delta) & = & \frac{1}{2}g^T\left(H + \lambda(\Delta)I\right)^{-2} g\frac{d\lambda(\Delta)}{d\Delta} - \lambda(\Delta)\Delta - \frac{1}{2}\frac{d\lambda(\Delta)}{d\Delta}\Delta^2 \\ & = & -\lambda(\Delta)\Delta\,. \end{array}$$

Using similar procedures we can calculate the second derivative of $v(\Delta)$ for $0 < \Delta < \Delta^*$:

$$\left.\begin{array}{rcl} v(\Delta) & = & -\frac{1}{2}g^T\left(H + \lambda(\Delta)I\right)^{-1} g - \frac{1}{2}\lambda(\Delta)\Delta^2\,, \\ v'(\Delta) & = & -\lambda(\Delta)\Delta\,, \\ v''(\Delta) & = & -\lambda(\Delta) + \frac{\Delta^2}{g^T\left(H+\lambda(\Delta)I\right)^{-3} g}\,, \end{array}\right\} \text{ for } 0 < \Delta < \Delta^*. \tag{3.9}$$

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 $\Delta = 0$ are $\Delta = \Delta^*$ are considered in Sections 3.2 and 3.3, respectively.

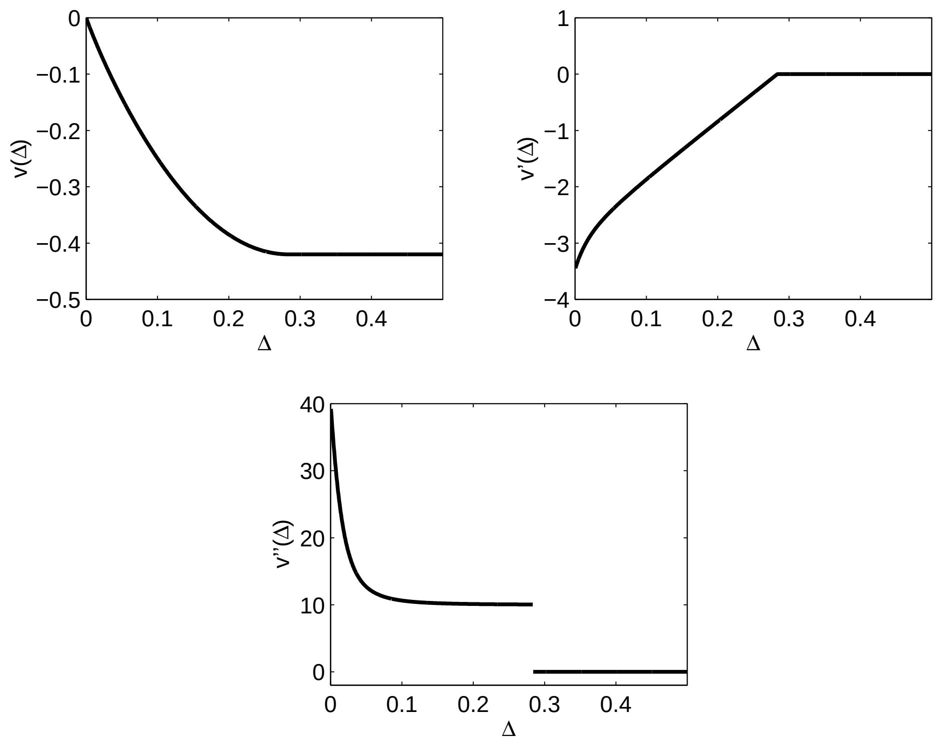

Figure 1: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([\,10\ 10\ 100\,])$, and $g = [\,2\ 2\ 2\,]'$. Here H is positive definite.

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

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

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

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

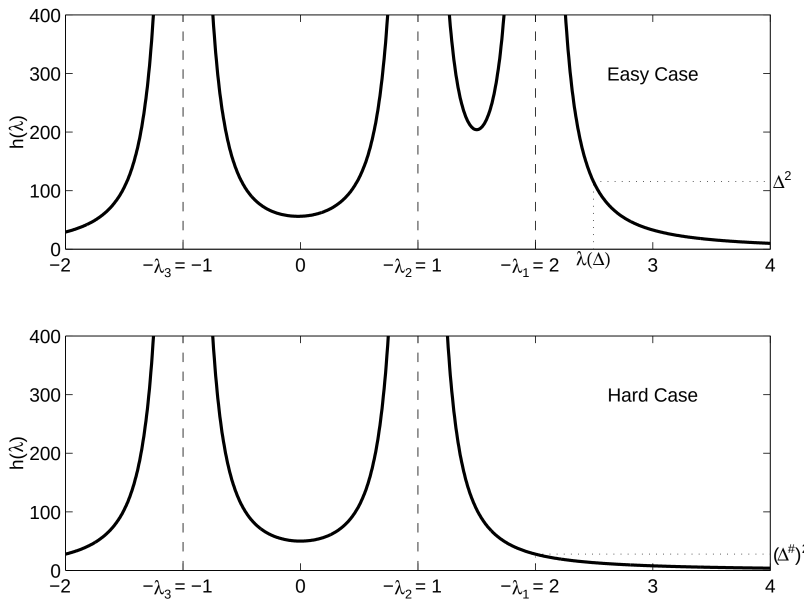

Figure 2: Plot of $h(\lambda)$ for the case $n = 3$, $H = diag([\,-2\ -1\ 1\,])$, $g = [\,5\ 5\ 5\,]'$ (Easy Case), and $g = [\,0\ 5\ 5\,]'$ (Hard Case).

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 $\Delta < \Delta^{\#}$, 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 [17] that in such 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 formula for $v(\lambda)$ and its derivatives. In fact we have,

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

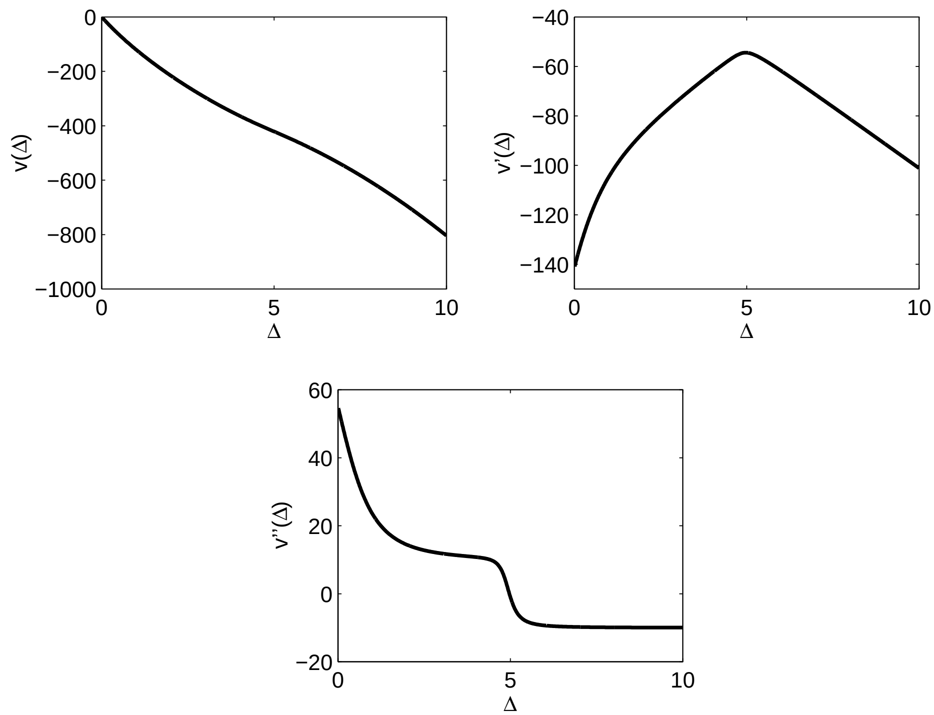

Figure 3: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([-5\ 1\ 15\ ])$, and $g = [\,1\ 1\ 1\,]'$. Here H is not positive definite but we have the easy case.

Thus

$$\left.\begin{array}{rcl} v(\Delta) & = & -\frac{1}{2}p^T\Big(H - \lambda_1 I\Big)p + \frac{1}{2}\lambda_1\Delta^2\,, \\ v'(\Delta) & = & \lambda_1\Delta\,, \\ v''(\Delta) & = & \lambda_1\,, \end{array}\right\} \text{ for } \Delta > \Delta^{\#}\,.$$

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 in $C^\infty(0, +\infty)$.

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

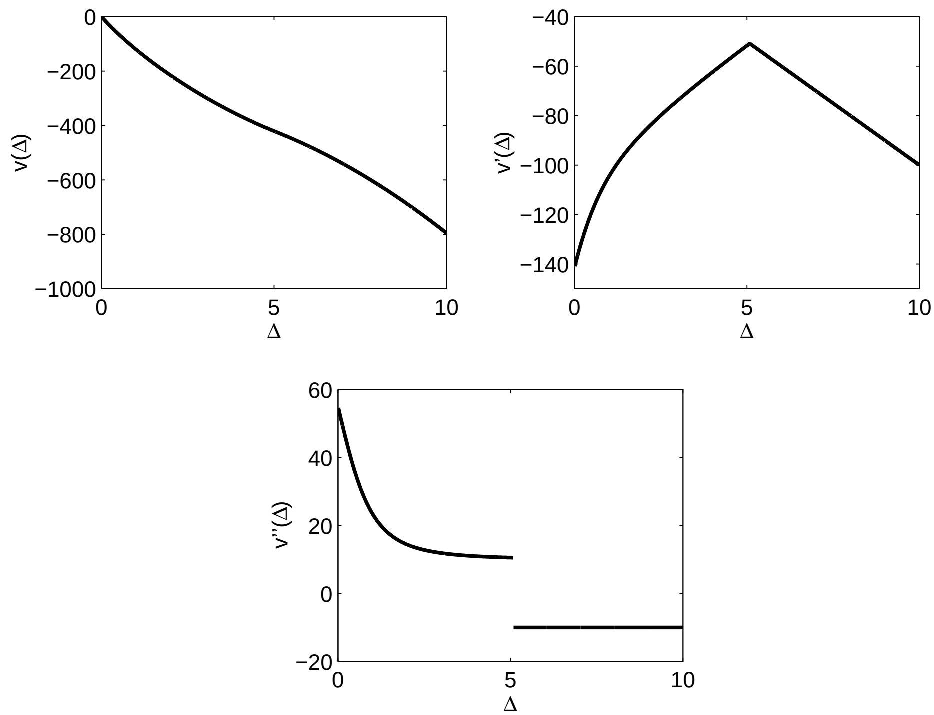

Figure 4: Plots of $v(\Delta)$, $v'(\Delta)$, and $v''(\Delta)$ for the case $n = 3$, $H = diag([-10\ 10\ 100])$, and $g = [\,0\ 100\ 100\,]'$. Here H is not positive definite and the hard case occurs.

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, we have the same expressions for $v(\Delta)$ and $v'(\Delta)$, no matter what case is being considered. From the asymptotic behavior of $h(\lambda)$, we know that $\lambda \to +\infty$ as $\Delta \to 0^+$. Thus, from (3.7) and $v(0) = 0$,

$$\begin{aligned}
v'_+(0) &= \lim_{\Delta\to 0^+} \frac{v(\Delta)-v(0)}{\Delta} \\
&= \lim_{\Delta\to 0^+} \frac{-\frac{1}{2}g^T\left(H+\lambda(\Delta)I\right)^{-1}g-\frac{1}{2}\lambda(\Delta)\Delta^2}{\Delta} \\
&= \lim_{\lambda\to+\infty} \frac{-\frac{1}{2}g^T\left(H+\lambda I\right)^{-1}g-\frac{1}{2}\lambda g^T\left(H+\lambda I\right)^{-2}g}{\left(g^T\left(H+\lambda I\right)^{-2}g\right)^{\frac{1}{2}}} \\
&= \lim_{\lambda\to+\infty} \frac{-\lambda\frac{1}{2}g^T\left(H+\lambda I\right)^{-1}g-\frac{1}{2}\lambda^2 g^T\left(H+\lambda I\right)^{-2}g}{\left(\lambda^2 g^T\left(H+\lambda I\right)^{-2}g\right)^{\frac{1}{2}}} \\
&= -\left(\sum_{i=1}^n \bar{g}_i^2\right)^{\frac{1}{2}}\ ,
\end{aligned}$$

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

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

and

$$\lim_{\Delta \to 0^+} v'_+(\Delta) \;=\; \lim_{\Delta \to 0^+} -\lambda(\Delta)\Delta \;=\; \lim_{\lambda \to +\infty} -\lambda \left(g^T \Big(H + \lambda I\Big)^{-2} g \right)^{\frac{1}{2}} \;=\; -\left(\sum_{i=1}^{n} \bar{g}_i^2 \right)^{\frac{1}{2}} .$$

In conclusion,

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

Using similar arguments one can show that

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

3.3 Nondifferentiability of the first derivative

The only points where $v(\Delta)$ and $v'(\Delta)$ can be nondifferentiable, 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

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

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:

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

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 yields

$$\left(-\lambda(\Delta) + \frac{\Delta^2}{g^T \left(H + \lambda(\Delta) I \right)^{-3} g} \right)' = \frac{3\Delta \left(g^T \left(H + \lambda(\Delta) I \right)^{-3} g \right)^2 - 3\Delta^3 \left(g^T \left(H + \lambda(\Delta) I \right)^{-4} g \right)}{\left(g^T \left(H + \lambda(\Delta) I \right)^{-3} g \right)^3} .$$

Since

$$\left(g^T \left(H + \lambda(\Delta) I \right)^{-3} g \right)^2 \leq \left(g^T \left(H + \lambda(\Delta) I \right)^{-2} g \right) \left(g^T \left(H + \lambda(\Delta) I \right)^{-4} g \right) ,$$

we have that

$$\left(-\lambda(\Delta) + \frac{\Delta^2}{g^T \left(H + \lambda(\Delta) I \right)^{-3} g} \right)' \leq 0 .$$

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 values 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:

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

From Table 1, the conclusion is that

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

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 $v''_+(0)\le 0$, 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 Applications

5.1 A Statistical Problem

In this section we outline an interesting application of the marginal function that arises in Statistics. This application was the main motivation for the work in this paper and is described in detail in Andrews [1].

A simple description of the statistical problem is as follows: suppose we have two populations that have p-dimensional normal distributions with (positive definite) covariance matrices Σ_1 and Σ_2 and mean vectors $\mu_1 \neq \mu_2$. We would like to find the value of $c^2 > 0$ such that the equi-probable ellipsoids

$$\mathcal{E}_i(c^2) = \{x : (x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) \le c^2\},\quad i = 1,2$$

are tangent, i.e. $\mathcal{E}_1(c^2)\cap\mathcal{E}_2(c^2)$ consists of exactly one point. This gives us a measurement of distance between the groups that takes into account the scaling of both groups. This also leads directly to a projection into a lower dimensional space which optimally maintains the group structure.

We can re-pose this problem by examining the marginal function

$$v(c^2) = \max\{(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) : (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) \le c^2\} \tag{5.1}$$

and then looking for the solution of the nonlinear equation

$$v(c^2) = c^2 \,. \tag{5.2}$$

Problem (5.1) is easily reparameterized into the standard form (1.2), see Section 6. Thus, we can use the results of this paper to calculate the derivative of (5.1) and so solve (5.2) using Newton's method. This approach has been applied to a few statistical data sets with successful numerical results [1].

5.2 Reducing the trust radius in optimization

Consider the unconstrained optimization problem (1.3) and assume that a step $s_k = s(\Delta_k)$ is computed by approximately solving the trust–region subproblem (1.4). A typical trust–region algorithm decides to accept or reject s_k depending on the ratio

$$r_k = \frac{f(x_k + s_k) - f(x_k)}{g_k^T s_k + \frac{1}{2} s_k^T H_k s_k} \,.$$

If $r_k < \eta_1$ for some $\eta_1 \in (0,1)$ (a typical value for η_1 is 0.001), the step s_k is usually rejected and the trust–region subproblem (1.4) is solved again for $g_{k+1} = g_k$, $H_{k+1} = H_k$, and a smaller value Δ_{k+1} of the trust radius. Usually Δ_{k+1} is set to $\alpha \Delta_k$ or to $\alpha \|s_k\|$, where $\alpha \in (0,1)$. These updates allow global convergence for the trust–region algorithm. However, the dependence on the value of α is too heuristic. In this section we propose a new framework for reducing the trust radius. (Numerical results will be reported elsewhere.)

It seems reasonable to reduce the trust radius so that in the next iteration the step s_{k+1} is accepted and the trust radius is left unchanged, i.e. $\Delta_{k+1} = \Delta_k$. This is the case for many trust–region algorithms if r_{k+1} lies in $[\eta_2, \eta_3)$, with $\eta_3 > \eta_2 > \eta_1$. For instance, we can have $\eta_2 = 0.1$ and $\eta_3 = 0.75$. Based on this observation, we suggest the following update for Δ_{k+1} :

$$\Delta_{k+1} = \min \left\{ \alpha \|s_k\|, \, \Delta_k + \Delta_k^{\mathsf{N}} \right\} ,$$

where Δ_k^{N} is a Newton step on

$$g(\Delta) \;=\; f(x_k) - f(x_k + s(\Delta)) + \eta \left(g_k^T s(\Delta) + \frac{1}{2} s(\Delta)^T H_k s(\Delta) \right) \;=\; 0 \,,$$

and $\eta > 0$ is the desired target for r_{k+1}. In other words, $\Delta_k^{\mathsf{N}} = -\frac{g(\Delta_k)}{g'(\Delta_k)}$. By taking derivatives we get:

$$\Delta_k^{\mathsf{N}} = - \frac{f(x_k) - f(x_k + s_k) + \eta v(\Delta_k)}{-\nabla f(x_k + s_k)^T \frac{ds}{d\Delta}(\Delta_k) + \eta v'(\Delta_k)} \,.$$

Based on this framework, we can consider other schemes to reduce the trust radius.

The value for the derivative $\frac{ds}{d\Delta}(\Delta_k)$ depends on what case is being considered. In most of the cases this derivative is calculated as:

$$\frac{ds}{d\Delta}(\Delta_k) \;=\; -\Big(H + \lambda(\Delta_k) I\Big)^{-1} \left(H + \frac{d\lambda}{d\Delta}(\Delta_k) I \right) s_k \,,$$

where $\frac{d\lambda}{d\Delta}(\Delta_k)$ is given by (3.8).

If H is positive definite and $\Delta > \Delta^*$, then $\frac{ds}{d\Delta}(\Delta_k) = 0$. If H is not positive definite, the hard case occurs, and $\Delta > \Delta^{\#}$, then

$$\frac{ds}{d\Delta}(\Delta_k) \;=\; \frac{d\tau}{d\Delta}(\Delta_k)\, q \;=\; \left(\frac{\Delta_k}{p^T q + \tau(\Delta_k) q^T q}\right) q \,.$$

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

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

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

Moré [12] considers another generalization of the form

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

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

References

[1] D. A. Andrews, *Optimal projections for normal groups*, tech. rep., Department of Statistics, Rice University, 1996.

[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] 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.

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

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

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

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

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

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