DAMTP 2011/NA16

The Lagrange method and SAO with bounds on the dual variables[1]

M.J.D. Powell

Abstract: We consider the minimization of $f_0(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, subject to $f_j(\underline{x})=0$, $1\le j\le m'$, $f_j(\underline{x})\le 0$, $m'+1\le j\le m$ and $\underline{x}\in\mathcal{X}$, where $\mathcal{X}$ is compact. For any $\underline{\lambda}\in\mathcal{R}^m$, let $\underline{x}(\underline{\lambda})$ be a minimizer of the Lagrange function $L(\underline{x},\underline{\lambda})=f_0(\underline{x})+\Sigma_{j=1}^m\lambda_j f_j(\underline{x})$, $\underline{x}\in\mathcal{X}$, and let ϕ be the dual function $\phi(\underline{\lambda})=L(\underline{x}(\underline{\lambda}),\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$. Assuming only that the functions f_j are continuous, it has been proved that, if $\underline{x}(\underline{\lambda})$ is unique, then ϕ has the derivatives $d\phi(\underline{\lambda})/d\lambda_j=f_j(\underline{x}(\underline{\lambda}))$, $1\le j\le m$. Thus we deduce that, if $\phi(\underline{\lambda}^*)$ is the greatest value of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, subject to $\lambda_j\ge 0$, $m'+1\le j\le m$, and if $\underline{x}(\underline{\lambda}^*)$ is unique, then $\underline{x}=\underline{x}(\underline{\lambda}^*)$ is the solution of the given problem. These properties are illustrated by an example with $n=2$ and $m'=m=1$. The given problem may have no feasible point, however, and then ϕ may not be bounded above. Therefore the bounded dual method adds the condition $\|\underline{\lambda}\|_\infty\le\Lambda$ for some prescribed $\Lambda>0$, and we let $\underline{\lambda}^*$ be the new maximizer of ϕ. We find that, if $\underline{x}(\underline{\lambda}^*)$ is unique, then now it minimizes the function $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, which is $f_0(\underline{x})$ plus Λ times the sum of moduli of constraint violations at $\underline{x}$. The term SAO stands for Sequential Approximate Optimization. An outermost iteration makes quadratic approximations to the functions f_j, $0\le j\le m$, with first order accuracy at $\underline{x}^{(k)}$, say, where k is the iteration number. Then, using the approximations instead of the original functions, the bounded dual method is applied, giving a new $\underline{\lambda}^*$ and a unique $\underline{x}(\underline{\lambda}^*)$. The choice of $\underline{x}^{(k+1)}$ can depend on $\Psi(\underline{x}^{(k)})$ and on the new $\Psi(\underline{x}(\underline{\lambda}^*))$. Thus our theory suggests some useful developments of SAO.

Keywords: Bounded dual method; Constrained optimization; Derivatives of dual function; L_1 penalty terms; Lagrange method; Sequential approximate optimization.

Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge CB3 0WA,
England.

March, 2012 (revised).

[1]Presented at the 8th International Conference on Numerical Optimization and Numerical Linear Algebra (Xiamen, China, 2011).

1. Introduction

We consider the minimization of $f_0(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, subject to the constraints

$$\left.\begin{array}{ll} f_j(\underline{x})=0, & j=1,2,\ldots,m', \\ f_j(\underline{x})\leq 0, & j=m'+1,\ldots,m, \\ \text{and} & \underline{x}\in\mathcal{X}. \end{array}\right\} \tag{1.1}$$

We call this problem NLP. The integers m' and m–m' are nonnegative, the number of equality and inequality constraints being zero in the cases $m'=0$ and $m'=m$, respectively. The objective and constraint functions are assumed to be continuous. The set $\mathcal{X}$ is required to be bounded and closed. Thus, for every vector $\underline{\lambda}$ in $\mathcal{R}^m$, the Lagrange function

$$L(\underline{x},\underline{\lambda}) \;=\; f_0(\underline{x})+\textstyle\sum_{j=1}^m \lambda_j\, f_j(\underline{x}), \qquad \underline{x}\in\mathcal{X}, \tag{1.2}$$

attains its least value at a point in $\mathcal{X}$, say $\underline{x}=\underline{x}(\underline{\lambda})$, which need not be unique. We employ the notation

$$\phi(\underline{\lambda}) \;=\; L(\underline{x}(\underline{\lambda}),\underline{\lambda}) \;=\; \min\{L(\underline{x},\underline{\lambda}):\underline{x}\in\mathcal{X}\}, \qquad \underline{\lambda}\in\mathcal{R}^m, \tag{1.3}$$

for the "dual function" of NLP.

Let $\underline{x}^*$ be a solution of NLP, and let $\mathcal{J}^*$ be the set of indices of the constraints that are active at $\underline{x}^*$, which means that the integer j from $[1,m]$ is in $\mathcal{J}^*$ if and only if $f_j(\underline{x}^*)$ is zero. It is well known that, if $\underline{x}^*$ is an interior point of $\mathcal{X}$, if the given functions are continuously differentiable, and if the gradients $\underline{\nabla} f_j(\underline{x}^*)$, $j\in\mathcal{J}^*$, are linearly independent, then there exist unique multipliers λ_j^*, $j\in\mathcal{J}^*$, such that the equation

$$\underline{\nabla} f_0(\underline{x}^*)+\textstyle\sum_{j\in\mathcal{J}^*}\lambda_j^*\,\underline{\nabla} f_j(\underline{x}^*) \;=\; \underline{\nabla} f_0(\underline{x}^*)+\sum_{j=1}^m\lambda_j^*\,\underline{\nabla} f_j(\underline{x}^*) \;=\; 0 \tag{1.4}$$

holds, the numbers λ_j^*, $j\in[1,m]\backslash\mathcal{J}^*$ being set to zero. Hence the definition (1.2) implies that $L(\underline{x},\underline{\lambda}^*)$, $\underline{x}\in\mathcal{X}$, is stationary at $\underline{x}=\underline{x}^*$. Furthermore, if the given functions are twice continuously differentiable, and if conditions for second order sufficiency are satisfied at $\underline{x}^*$, then $\underline{x}^*$ can become a strict local minimum of $L(\underline{x},\underline{\lambda}^*)$, $\underline{x}\in\mathcal{X}$, by modifying f_0 if necessary. The modification adds sufficiently large multiples of squares of the functions f_j, $j\in\mathcal{J}^*$, to f_0. Hence, as described in Fletcher (1987) and in Nocedal and Wright (1999), for instance, the augmented Lagrangian method attempts to calculate a solution of NLP by minimizing the function $L(\underline{x},\underline{\lambda})$, $\underline{x}\in\mathcal{X}$, for a sequence of vectors $\underline{\lambda}\in\mathcal{R}^m$ that is designed to converge to $\underline{\lambda}^*$.

The convergence of $\underline{\lambda}$ to $\underline{\lambda}^*$ may be achieved by seeking the greatest value of the dual function (1.3). Justifying this statement is straightforward in the case $m'=m$, after the augmentation of the previous paragraph has provided the property $\underline{x}^*=\underline{x}(\underline{\lambda}^*)$. Indeed, the definition (1.3) supplies the upper bound

$$\phi(\underline{\lambda}) \;=\; L(\underline{x}(\underline{\lambda}),\underline{\lambda}) \;\leq\; L(\underline{x}^*,\underline{\lambda}) \;=\; L(\underline{x}^*,\underline{\lambda}^*) \;=\; \phi(\underline{\lambda}^*), \qquad \underline{\lambda}\in\mathcal{R}^m, \tag{1.5}$$

on the dual function, the equation $L(\underline{x}^*, \underline{\lambda}) = L(\underline{x}^*, \underline{\lambda}^*)$ being due to $f_j(\underline{x}^*) = 0$, $j = 1, 2, \ldots, m$. The concavity of $\phi(\underline{\lambda})$, $\underline{\lambda} \in \mathcal{R}^m$, is established in Bazaraa, Sherali and Shetty (2006). It is also usual to seek the maximum of $\phi(\underline{\lambda})$ in the case $m' < m$, but now the multipliers λ_j, $m'+1 \leq j \leq m$, of the inequality constraints should be nonnegative, in accordance with first order conditions for optimality.

SAO methods employ the functions (1.2) and (1.3), the name being an acronym for Sequential Approximate Optimization. They generate a sequence of NLPs, where an outermost iteration adjusts the functions f_j, $j = 0, 1, \ldots, m$, and where an optimization algorithm tries to find the solution of each NLP. Further, as in Groenwold and Etman (2008) and in Groenwold, Etman and Wood (2010), a technique crucial to efficiency is to require every f_j, $j = 0, 1, \ldots, m$, to have the separable form

$$f_j(\underline{x}) \;=\; \textstyle\sum_{i=1}^{n} g_{ji}(x_i), \qquad \underline{x} \in \mathcal{R}^n, \tag{1.6}$$

each g_{ji} being a function of only one variable. Also the condition $\underline{x} \in \mathcal{X}$ is restricted to the simple bounds

$$\ell_i \;\leq\; x_i \;\leq\; u_i, \qquad i = 1, 2, \ldots, n, \tag{1.7}$$

the values of ℓ_i and u_i being prescribed. Then $\underline{x}(\underline{\lambda})$ minimizes the Lagrange function (1.2) if and only if its components $x_i(\underline{\lambda})$, $i = 1, 2, \ldots, n$, minimize the functions of one variable

$$g_{0i}(x) \;+\; \textstyle\sum_{j=1}^{m} \lambda_j \, g_{ji}(x), \qquad \ell_i \leq x \leq u_i, \qquad i = 1, 2, \ldots, n. \tag{1.8}$$

Thus the calculation of $\underline{x}(\underline{\lambda})$ becomes much easier than usual, which allows n to be very large. A recent review of these techniques is given by Etman, Groenwold and Rooda (2011).

Let a search for the maximum of $\phi(\underline{\lambda})$ yield a sequence of vectors $\underline{\lambda}$, either in the augmented Lagrangian or in an SAO method. If the corresponding vectors $\underline{x}(\underline{\lambda})$ fail to have a limit that satisfies the constraints of NLP, then the augmented Lagrangian method adds more multiples of squares of constraint functions to f_0, which may be combined with the adjustment of $\underline{\lambda}$. In the SAO method mentioned above, however, such additions are unacceptable, because f_0 would lose the separability property (1.6). Instead the outermost iteration makes suitable changes to the functions f_j, $j = 0, 1, \ldots, m$. For example, when these functions are quadratic with diagonal second derivative matrices, the diagonal elements of $\nabla^2 f_0$ may be increased.

Good reasons for studying the dual function (1.3) have been mentioned, but there may be no multipliers $\underline{\lambda}^* \in \mathcal{R}^m$ such that $\underline{x}(\underline{\lambda}^*)$ is a solution of NLP. Because a search for the greatest value of $\phi(\underline{\lambda})$ may cause $\underline{\lambda}$ to become unbounded, the "bounded dual" method of Wood, Groenwold and Etman (2011) confines $\underline{\lambda}$ to the set $\mathcal{L}$, where $\underline{\lambda} \in \mathcal{R}^m$ is in $\mathcal{L}$ if and only if its components satisfy the conditions

$$\left.\begin{aligned} -\Lambda \leq \lambda_j \leq \Lambda, \qquad & j = 1, 2, \ldots, m', \\ 0 \leq \lambda_j \leq \Lambda, \qquad & j = m'+1, \ldots, m, \end{aligned}\right\} \tag{1.9}$$

the number Λ being a prescribed positive constant. Usually the maximization of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, supplies a vector $\underline{x}(\underline{\lambda})$ that is useful for the adjustment of the functions $f_j,\ j=0,1,\ldots,m$, on each outermost iteration of SAO. Therefore the author investigated properties of $\underline{x}(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, and $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, during a very pleasant visit to the University of Stellenbosch early in 2011, where he was introduced to SAO methods by Albert Groenwold.

In the theoretical analysis at Stellenbosch, the author assumed differentiablity of the functions f_j, $j=0,1,\ldots,m$, but instead he found later that is is sufficient to assume only that these functions are continuous. By removing the unnecessary assumptions, his proofs became much shorter than before. Discussions of this work with some experts in optimization suggested that the shorter proofs may be new, so they were included in a submission to the journal Optimization Methods and Software, which captured the attention of Oleg Burdakov, who recommended a study of Chapter 6 of the book by Bazaara *et al* (2006).

That chapter includes proofs of the concavity and the differentiability of the dual function that are the same as the shorter proofs of the author. The differentiability result is that, if $\underline{\lambda}\in\mathcal{R}^m$ is such that $\underline{x}(\underline{\lambda})$ is unique in equation (1.3), then the dual function has the derivatives

$$d\phi(\underline{\lambda})\,/\,d\lambda_j \;=\; f_j(\underline{x}(\underline{\lambda})), \qquad j=1,2,\ldots,m. \tag{1.10}$$

After addressing the continuity of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, these properties are stated as theorems in Section 2, the author's proofs being replaced by references to Bazaraa *et al* (2006). We note too in Section 2 that, if $\underline{\lambda}^*$ is a $\underline{\lambda}$ that maximizes $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, subject to $\lambda_j\geq 0$, $j=m'+1,\ldots,m$, and if $\underline{x}(\underline{\lambda}^*)$ is unique, then $\underline{x}(\underline{\lambda}^*)$ is the unique solution of NLP. One advantage of the generality of the compact set $\mathcal{X}$ in NLP is that it allows integer constraints on the variables.

The bounds (1.9) provide the setting of Section 3. We let $\underline{\lambda}^\dagger$ be a vector that maximizes $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$. It is proved that, if $\underline{x}(\underline{\lambda}^\dagger)$ is unique, then $\underline{x}(\underline{\lambda}^\dagger)$ is also the unique vector in $\mathcal{X}$ that minimizes the function

$$\Psi(\underline{x}) \;=\; f_0(\underline{x}) \;+\; \Lambda\,\sum_{j=1}^{m'}\,|\,f_j(\underline{x})\,| \;+\; \Lambda\,\sum_{j=m'+1}^{m}\,\max\,[\,0,\,f_j(\underline{x})\,], \qquad \underline{x}\in\mathcal{X}. \tag{1.11}$$

Gill and Robinson (2010) establish a similar relation between the "bounded dual" method and the treatment of general constraints by penalty terms, when the functions f_j, $j=0,1,\ldots,m$, are differentiable. These properties are useful not only when moderate or small values of Λ are of interest but also when the constraints (1.1) of NLP are inconsistent.

Section 4 presents an example, which illustrates the theory of Section 2, and which demonstrates that every $\underline{x}(\underline{\lambda})$ may be far from a solution of NLP. It is explained in Section 5, however, that a solution to the example of Section 4 can be calculated by seeking local minima of $L(\underline{x},\underline{\lambda})$, $\underline{x}\in\mathcal{X}$, instead of global minima. Also in Section 5, we address briefly some details of "bounded dual" SAO methods when first derivatives are available, giving particular attention to the usefulness of the function (1.11).

2. The continuity, concavity and differentiability of the dual function

The continuity, concavity and differentiability of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, are addressed in Theorems 1, 2 and 3 below, without any restrictions on the signs of the components of $\underline{\lambda}$. The constraints $\lambda_j\geq 0$, $m'+1\leq j\leq m$, are required in Corollary 1, however, which establishes the relation, stated in Section 1, between the maximization of $\phi(\underline{\lambda})$ and the solution of NLP. It is assumed that the functions f_j, $j=0,1,\ldots,m$, of NLP are continuous.

Theorem 1 The dual function $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, is continuous.

Proof Let $\underline{\lambda}$ and $\underline{\mu}$ be any two vectors in $\mathcal{R}^m$. The definitions (1.3) and (1.2) supply the inequality

$$\begin{aligned}\phi(\underline{\lambda})-\phi(\underline{\mu}) &= L(\underline{x}(\underline{\lambda}),\underline{\lambda})-L(\underline{x}(\underline{\mu}),\underline{\mu}) \leq L(\underline{x}(\underline{\mu}),\underline{\lambda})-L(\underline{x}(\underline{\mu}),\underline{\mu})\\ &= \textstyle\sum_{j=1}^m (\lambda_j-\mu_j)\, f_j(\underline{x}(\underline{\mu})),\end{aligned} \tag{2.1}$$

which is equivalent to the condition

$$\phi(\underline{\mu})-\phi(\underline{\lambda}) \leq \textstyle\sum_{j=1}^m (\mu_j-\lambda_j)\, f_j(\underline{x}(\underline{\lambda})), \quad \underline{\lambda}\in\mathcal{R}^m, \quad \underline{\mu}\in\mathcal{R}^m. \tag{2.2}$$

Furthermore, we let $c_{\max}$ be the least constant that satisfies the bound

$$|\, f_j(\underline{x})\,| \leq c_{\max}, \quad \underline{x}\in\mathcal{X}, \quad j=1,2,\ldots,m, \tag{2.3}$$

which exists because of the compactness of $\mathcal{X}$ and the continuity of the functions of NLP. Thus we deduce the property

$$|\,\phi(\underline{\lambda})-\phi(\underline{\mu})\,| \leq c_{\max}\,\|\underline{\lambda}-\underline{\mu}\|_1, \quad \underline{\lambda}\in\mathcal{R}^m, \quad \underline{\mu}\in\mathcal{R}^m, \tag{2.4}$$

which confirms that ϕ is continuous. QED

Theorem 2 The dual function $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, is concave.

Proof See Theorem 6.3.1 of Bazaraa *et al* (2006).

Theorem 3 Let $\underline{\lambda}_\infty$ be any vector in $\mathcal{R}^m$ such that only one vector $\underline{x}(\underline{\lambda}_\infty)$ minimizes $L(\underline{x},\underline{\lambda}_\infty)$, $\underline{x}\in\mathcal{X}$. Then the dual function (1.3) is differentiable at $\underline{\lambda}=\underline{\lambda}_\infty$, its gradient having the components

$$d\phi(\underline{\lambda}_\infty)\,/\,d\lambda_j = f_j(\underline{x}(\underline{\lambda}_\infty)), \qquad j=1,2,\ldots,m. \tag{2.5}$$

Proof See Lemma 6.3.2 and Theorem 6.3.3 of Bazaraa *et al* (2006).

Corollary 1 Let $\underline{\lambda}_\infty$ be a vector $\underline{\lambda}$ that maximizes the dual function (1.3), subject to $\lambda_j\geq 0$, $m'+1\leq j\leq m$. If $\underline{x}(\underline{\lambda}_\infty)$ is unique, as in Theorem 3, then $\underline{x}(\underline{\lambda}_\infty)$ is the unique solution of NLP.

Proof We write $\underline{x}_\infty$ instead of $\underline{x}(\underline{\lambda}_\infty)$. Theorem 3 supplies the derivatives (2.5), and the definition of a derivative provides the property

$$\phi(\underline{\lambda}_\infty + \alpha\,\underline{e}_j) \;=\; \phi(\underline{\lambda}_\infty) \;+\; \alpha\, f_j(\underline{x}_\infty) + o(\alpha), \qquad j = 1, 2, \ldots, m, \tag{2.6}$$

where $\underline{e}_j$ is the j-th coordinate vector in $\mathcal{R}^n$, and where $o(\alpha)$ denotes a term that tends to zero faster than α as $\alpha \to 0$. The given constraints on $\underline{\lambda}$ allow changes to $(\underline{\lambda}_\infty)_j$ of either sign if $j \in [1, m']$, and they allow increases in $(\underline{\lambda}_\infty)_j$ if $j \in [m'+1, m]$, where $(\underline{\lambda}_\infty)_j$ is the j-th component of $\underline{\lambda}_\infty$. Thus, because $\phi(\underline{\lambda}_\infty)$ is the greatest value of $\phi(\underline{\lambda})$ subject to $\lambda_j \geq 0$, $m'+1 \leq j \leq m$, we deduce $f_j(\underline{x}_\infty) = 0$ or $f_j(\underline{x}_\infty) \leq 0$ when j is the index of an equality or inequality constraint, respectively. In other words, the constraints of NLP are satisfied at $\underline{x} = \underline{x}_\infty$, the property $\underline{x}_\infty \in \mathcal{X}$ being part of the definition (1.3) of $\underline{x}(\underline{\lambda}_\infty)$.

Furthermore, if $(\underline{\lambda}_\infty)_j$ is positive, then the constraints on $\underline{\lambda}$ allow sufficiently small changes to $(\underline{\lambda}_\infty)_j$ of either sign, which, with the conclusion of the previous paragraph, gives the complementarity conditions

$$(\underline{\lambda}_\infty)_j\, f_j(\underline{x}_\infty) \;=\; 0, \qquad j = 1, 2, \ldots, m. \tag{2.7}$$

Let $\underline{x} = \underline{x}^*$ be any vector in $\mathcal{R}^n$ that satisfies the constraints (1.1). Then all the products $(\underline{\lambda}_\infty)_j\, f_j(\underline{x}^*)$, $j = 1, 2, \ldots, m$, are zero or negative, so the definition (1.2) of the Lagrange function gives $L(\underline{x}^*, \underline{\lambda}_\infty) \leq f_0(\underline{x}^*)$, while the conditions (2.7) imply $L(\underline{x}_\infty, \underline{\lambda}_\infty) = f_0(\underline{x}_\infty)$. Moreover, $L(\underline{x}_\infty, \underline{\lambda}_\infty)$ is the least value of $L(\underline{x}, \underline{\lambda}_\infty)$, $\underline{x} \in \mathcal{X}$. These remarks yield the relations

$$f_0(\underline{x}_\infty) \;=\; L(\underline{x}_\infty, \underline{\lambda}_\infty) \;\leq\; L(\underline{x}^*, \underline{\lambda}_\infty) \;\leq\; f_0(\underline{x}^*). \tag{2.8}$$

Because $\underline{x}^*$ is any feasible point, it follows that $\underline{x}_\infty$ is a solution of NLP.

If $\underline{x}^*$ is also a solution of NLP, then the inequalities of expression (2.8) become equations. In particular, $L(\underline{x}_\infty, \underline{\lambda}_\infty) = L(\underline{x}^*, \underline{\lambda}_\infty)$ holds. Therefore, because $\underline{x}_\infty$ is the unique minimizer of $L(\underline{x}, \underline{\lambda}_\infty)$, $\underline{x} \in \mathcal{X}$, we have $\underline{x}^* = \underline{x}_\infty$. The proof of the corollary is complete. QED

3. A property of the bounded dual method

Let $\mathcal{L} \subset \mathcal{R}^m$ contain the vectors $\underline{\lambda}$ whose components satisfy the conditions (1.9), where Λ is a prescribed positive constant, and where the integers m and m', introduced in the constraints (1.1), are taken from NLP. We recall from Section 1 that the bounded dual method seeks the greatest value of the function

$$\phi(\underline{\lambda}) \;=\; L(\underline{x}(\underline{\lambda}), \underline{\lambda}) \;=\; \min\{L(\underline{x}, \underline{\lambda}) : \underline{x} \in \mathcal{X}\}, \qquad \underline{\lambda} \in \mathcal{L}, \tag{3.1}$$

where L is the Lagrange function (1.2). The advantage of confining $\underline{\lambda}$ to the compact set $\mathcal{L}$ is that there exists $\underline{\lambda}^\dagger$ in $\mathcal{L}$ such that $\phi(\underline{\lambda}^\dagger)$ is the maximum value

of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, the continuity of ϕ being proved in Theorem 1. We are interested in relations between the vector $\underline{x}(\underline{\lambda}^\dagger)$ in $\mathcal{X}$ that minimizes $L(\underline{x},\underline{\lambda}^\dagger)$, $\underline{x}\in\mathcal{X}$, and the solution of NLP, but the relations are hardly ever helpful unless $\underline{x}(\underline{\lambda}^\dagger)$ is unique. We assume uniqueness throughout this section, which may require a modification to NLP, such as a reduction in the range of $\mathcal{X}$. The case when the constraints (1.1) do not have a feasible point is particularly relevant to SAO methods. Then, instead of trying to solve NLP, one can seek the least value of the function

$$\Psi(\underline{x}) \;=\; f_0(\underline{x}) \;+\; \Lambda \sum_{j=1}^{m'} |\, f_j(\underline{x})\,| \;+\; \Lambda \sum_{j=m'+1}^{m} \max[\,0,\, f_j(\underline{x})\,], \qquad \underline{x}\in\mathcal{X}, \tag{3.2}$$

for a sufficiently large value of the constant Λ, as mentioned in the penultimate paragraph of Section 1. Several implementations of this technique have been proposed, including Fletcher (1981). The following theorem explains why Λ in expression (3.2) is taken from the summary of the bounded dual method given above. Its proof has much in common with the proof of Corollary 1.

Theorem 4 Let $\underline{\lambda}^\dagger\in\mathcal{L}$ be a vector that maximizes $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, in the bounded dual method. If there is only one vector $\underline{x}(\underline{\lambda}^\dagger)$ in $\mathcal{X}$ that minimizes $L(\underline{x},\underline{\lambda}^\dagger)$, $\underline{x}\in\mathcal{X}$, then $\underline{x}(\underline{\lambda}^\dagger)$ is also the unique vector that minimizes the function (3.2).

Proof We write $\underline{x}^\dagger$ instead of $\underline{x}(\underline{\lambda}^\dagger)$. Theorem 3 states that the dual function has the gradient $\underline{\nabla}\phi(\underline{\lambda}^\dagger)=\underline{f}(\underline{x}^\dagger)$. Therefore, because $\phi(\underline{\lambda}^\dagger)$ is the greatest value of $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, the conditions (1.9) on $\mathcal{L}$ imply that the components of $\underline{\lambda}^\dagger$ have the properties

$$\left.\begin{array}{rcll} \lambda_j^\dagger &=& -\Lambda \quad \text{if} \quad f_j(\underline{x}^\dagger)<0 \quad \text{and} \quad j\in[1,m'], \\ \lambda_j^\dagger &=& 0 \quad \text{if} \quad f_j(\underline{x}^\dagger)<0 \quad \text{and} \quad j\in[m'+1,m], \\ \lambda_j^\dagger &=& \Lambda \quad \text{if} \quad f_j(\underline{x}^\dagger)>0. \end{array}\right\} \tag{3.3}$$

They provide the equations

$$\left.\begin{array}{rcll} \lambda_j^\dagger\, f_j(\underline{x}^\dagger) &=& \Lambda\,|f_j(\underline{x}^\dagger)| & \text{if} \quad j\in[1,m'], \\ \lambda_j^\dagger\, f_j(\underline{x}^\dagger) &=& \Lambda\,\max[\,0,\, f_j(\underline{x}^\dagger)\,] & \text{if} \quad j\in[m'+1,m], \end{array}\right\} \tag{3.4}$$

including the case when $f_j(\underline{x}^\dagger)$ is zero. Thus the definitions (1.2) and (3.2) supply the identity

$$L(\underline{x}^\dagger,\underline{\lambda}^\dagger) \;=\; \Psi(\underline{x}^\dagger). \tag{3.5}$$

For general $\underline{x}\in\mathcal{X}$, however, the condition $\underline{\lambda}^\dagger\in\mathcal{L}$ gives the bounds

$$\left.\begin{array}{rcll} \lambda_j^\dagger\, f_j(\underline{x}) &\leq& \Lambda\,|f_j(\underline{x})| & \text{if} \quad j\in[1,m'], \\ \lambda_j^\dagger\, f_j(\underline{x}) &\leq& \Lambda\,\max[\,0,\, f_j(\underline{x})\,] & \text{if} \quad j\in[m'+1,m], \end{array}\right\} \tag{3.6}$$

which imply $L(\underline{x},\underline{\lambda}^\dagger)\leq\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$. Moreover, equation (1.3) and $\underline{x}^\dagger=\underline{x}(\underline{\lambda}^\dagger)$ show that $L(\underline{x}^\dagger,\underline{\lambda}^\dagger)$ is the least value of $L(\underline{x},\underline{\lambda}^\dagger)$, $\underline{x}\in\mathcal{X}$. These remarks, including

equation (3.5), yield the inequality

$$\Psi(\underline{x}^\dagger) \;=\; L(\underline{x}^\dagger,\underline{\lambda}^\dagger) \;\leq\; L(\underline{x},\underline{\lambda}^\dagger) \;\leq\; \Psi(\underline{x}), \qquad \underline{x}\in\mathcal{X}. \tag{3.7}$$

We see that $\Psi(\underline{x}^\dagger)$ is the least value of $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, as required.

Let $\underline{x}_\infty$ be any vector in $\mathcal{X}$ that minimizes $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, which is the condition $\Psi(\underline{x}_\infty)=\Psi(\underline{x}^\dagger)$, due to the conclusion of the previous paragraph. Then the choice $\underline{x}=\underline{x}_\infty$ in the relation (3.7) provides the equation $L(\underline{x}^\dagger,\underline{\lambda}^\dagger)=L(\underline{x}_\infty,\underline{\lambda}^\dagger)$. It follows from the uniqueness of $\underline{x}^\dagger=\underline{x}(\underline{\lambda}^\dagger)$ that $\underline{x}^\dagger$ and $\underline{x}_\infty$ are the same, which completes the proof. QED

4. An example

In an attempt to illustrate some features of the theory by a simple numerical example, we consider the minimization of the distance to the origin squared in only two dimensions, when the vector of variables is subject to a quadratic equality constraint. Letting x and y be the components of $\underline{x}\in\mathcal{R}^2$, we pick $f_0(\underline{x})=x^2+y^2$, $\underline{x}\in\mathcal{R}^2$, we pick $m'=m=1$, and we try to let f_1 be a quadratic such that the technique of maximizing the dual function (1.3) exposes some behaviour that is interesting and typical. The equation $f_1(\underline{x})=0$ defines a curve in $\mathcal{R}^2$, which has one or more closest points to the origin. We let there be two such points, $\widehat{\underline{x}}$ and $\check{\underline{x}}$, say, and we let $\mathcal{X}$ have the form $\{\underline{x}:\|\underline{x}\|\leq M\}$, where the constant M is so large that $\widehat{\underline{x}}$ and $\check{\underline{x}}$ are well inside $\mathcal{X}$. By construction $\|\widehat{\underline{x}}\|=\|\check{\underline{x}}\|$ and $f_1(\widehat{\underline{x}})=f_1(\check{\underline{x}})=0$ hold, so, if the choice of $\lambda\in\mathcal{R}$ causes $L(\widehat{\underline{x}},\lambda)$ to be the least value of the Lagrange function

$$L(\underline{x},\lambda) \;=\; x^2+y^2+\lambda\,f_1(\underline{x}), \qquad \underline{x}\in\mathcal{R}^2, \tag{4.1}$$

then $\check{\underline{x}}$ minimizes this function too. It follows that L is a quadratic that satisfies $\underline{\nabla}L(\widehat{\underline{x}})=\underline{\nabla}L(\check{\underline{x}})=0$, which implies that the second derivative matrix ∇^2L is singular. Thus examples of this kind are not simple, due to degeneracies in the Lagrange function.

Instead we let f_1 be a piecewise quadratic function with continuous first derivatives. Specifically, we consider the minimization of $f_0(\underline{x})=x^2+y^2$, $\underline{x}\in\mathcal{R}^2$, subject to the constraint $f_1(\underline{x})=0$, where f_1 is the function

$$f_1(\underline{x}) \;=\; \begin{cases} -\frac{3}{4}\,x^2-\frac{1}{4}\,xy-\frac{1}{3}\,y^2-\frac{1}{2}\,x+\frac{4}{3}\,y+\frac{8}{3}, & y\leq-\frac{1}{2}, \\ -\frac{3}{4}\,x^2-\frac{1}{4}\,xy-\frac{4}{3}\,y^2-\frac{1}{2}\,x+\frac{1}{3}\,y+\frac{29}{12}, & |y|\leq\frac{1}{2}, \\ -\frac{3}{4}\,x^2-\frac{1}{4}\,xy-y^2-\frac{1}{2}\,x+\frac{5}{2}, & y\geq\frac{1}{2}, \end{cases} \qquad \underline{x}\in\mathcal{R}^2. \tag{4.2}$$

The closed curve in Figure 1 is the graph $f_1(\underline{x})=0$, $\underline{x}\in\mathcal{R}^2$. It includes the points $\widehat{\underline{x}}=(1,-1)$ and $\check{\underline{x}}=(1,1)$, which are the only points on the closed curve that are nearest to the origin. Let M be a positive constant such that all of Figure 1 is inside the set $\mathcal{X}^\pm=\{\underline{x}:\|\underline{x}\|\leq M\}$, and let $\mathcal{X}^-$ and $\mathcal{X}^+$ contain the points of

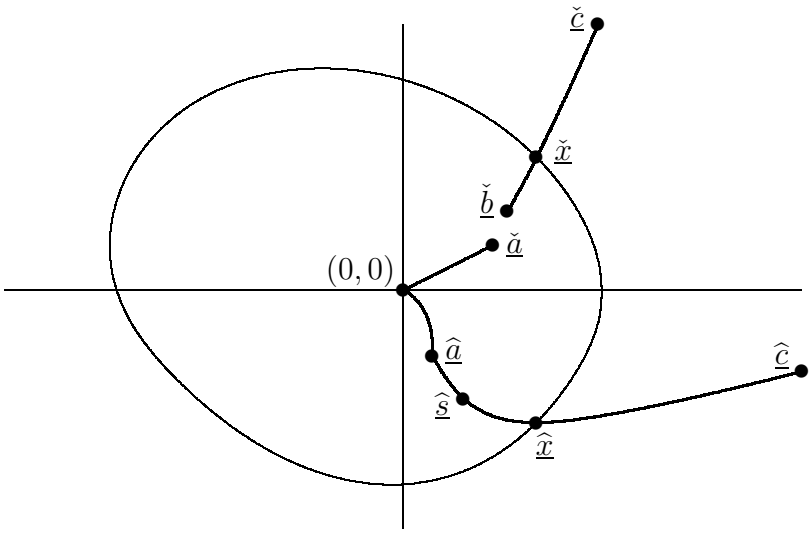

Figure 1: The paths of $\underline{x}(\lambda)$ in the example of Section 4

$\mathcal{X}^{\pm}$ that satisfy $y \leq \frac{1}{2}x$ and $y \geq \frac{1}{2}x$, respectively. For each of the choices $\mathcal{X} = \mathcal{X}^{-}$, $\mathcal{X} = \mathcal{X}^{+}$ and $\mathcal{X} = \mathcal{X}^{\pm}$, we seek the greatest value of the dual function

$$\phi(\lambda) \;=\; L(\underline{x}(\lambda), \lambda) \;=\; \min\{L(\underline{x}, \lambda) : \underline{x} \in \mathcal{X}\}, \qquad \lambda \in \mathcal{R}, \tag{4.3}$$

where L is the function (4.1). The sign of f_1 has been chosen so that the optimal value of λ is positive for each of the sets $\mathcal{X}^{-}$, $\mathcal{X}^{+}$ and $\mathcal{X}^{\pm}$.

When $\mathcal{X} = \mathcal{X}^{-}$, the value of λ that maximizes $\phi(\lambda)$ in this problem is $\lambda = 8/7$, and then $\underline{x}(\lambda)$ is the vector $\widehat{\underline{x}} = (1, -1)$ as required, which is shown in Figure 1. The thick curve in the lower half of this figure is the path of $\underline{x}(\lambda)$ as λ increases from zero until $\underline{x}(\lambda)$ leaves the figure at $\widehat{c}$. The middle line of expression (4.2) is relevant for sufficiently small λ. Then the components of $\underline{x}(\lambda)$ take the values

$$x(\lambda) \;=\; \frac{48\,\lambda - 68\,\lambda^2}{192 - 400\,\lambda + 189\,\lambda^2} \quad \text{and} \quad y(\lambda) \;=\; \frac{-32\,\lambda + 30\,\lambda^2}{192 - 400\,\lambda + 189\,\lambda^2}, \tag{4.4}$$

because they provide the unconstrained minimum of the quadratic function

$$x^2 + y^2 + \lambda\,\{\,-\tfrac{3}{4}\,x^2 - \tfrac{1}{4}\,xy - \tfrac{4}{3}\,y^2 - \tfrac{1}{2}\,x + \tfrac{1}{3}\,y + \tfrac{29}{12}\,\}, \qquad \underline{x} \in \mathcal{R}^2. \tag{4.5}$$

This minimum is in $\mathcal{X}^{-}$ as required, and also it is in the strip $\mathcal{S} = \{\underline{x} : |y| \leq \frac{1}{2}\}$ until λ reaches the value $\widehat{\lambda} = 0.620229$. The points $\underline{x}(\lambda)$, $0 \leq \lambda \leq \widehat{\lambda}$, define the path from $(0, 0)$ to $\widehat{a}$ in Figure 1, $\widehat{a} = \underline{x}(\widehat{\lambda})$ being on the boundary of $\mathcal{S}$. For $\lambda = \widehat{\lambda}$, the zero gradient of the quadratic (4.5) at $\underline{x} = \widehat{a}$ is inherited by the function

$$x^2 + y^2 + \lambda\,\{\,-\tfrac{3}{4}\,x^2 - \tfrac{1}{4}\,xy - \tfrac{1}{3}\,y^2 - \tfrac{1}{2}\,x + \tfrac{4}{3}\,y + \tfrac{8}{3}\,\}, \qquad \underline{x} \in \mathcal{R}^2, \tag{4.6}$$

because expression (4.2) has continuous first derivatives. Moreover, the quadratic (4.6) is strictly convex for every λ that is relevant to Figure 1. Thus its minimum

in the case $\lambda = \widehat{\lambda}$ also occurs at $\widehat{\underline{a}}$. Further, for $\lambda \geq \widehat{\lambda}$, the minimizer of the quadratic (4.6) has the components

$$x(\lambda) \;=\; \frac{48\,\lambda - 32\,\lambda^2}{192 - 208\,\lambda + 45\,\lambda^2} \quad \text{and} \quad y(\lambda) \;=\; \frac{-128\,\lambda + 102\,\lambda^2}{192 - 208\,\lambda + 45\,\lambda^2}. \tag{4.7}$$

They provide the thick curve from $\widehat{\underline{a}}$ to $\widehat{\underline{c}}$ in the figure, including $\underline{x}(\lambda) = \widehat{\underline{x}} = (1, -1)$ at $\lambda = 8/7$. The value of λ at $\widehat{\underline{c}}$ is defined by $x(\lambda) = 3$, which gives $\lambda = 1.238060$.

When $\mathcal{X}$ is $\mathcal{X}^+$ and when $\lambda > 0$ is sufficiently small, the Lagrange function is also the quadratic (4.5), but the components (4.4) fail to satisfy the condition $y \geq \frac{1}{2}x$ that defines points of $\mathcal{X}^+$. Instead $\underline{x}(\lambda)$ is the point on the boundary of $\mathcal{X}^+$ where the quadratic (4.5) is least, which has the components

$$x(\lambda) \;=\; \frac{4\,\lambda}{30 - 29\,\lambda} \qquad \text{and} \qquad y(\lambda) \;=\; \frac{2\,\lambda}{30 - 29\,\lambda}. \tag{4.8}$$

The gradient of $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{R}^2$, at this point is orthogonal to the line $y = \frac{1}{2}x$, being the vector $(-1, 2)$ multiplied by $(7\lambda - 8\lambda^2)/(30 - 29\lambda)$. This factor has to be nonnegative, because otherwise a move from $\underline{x}(\lambda)$ into the interior of $\mathcal{X} = \mathcal{X}^+$ would supply a first order reduction in $L(\cdot, \lambda)$. It follows that, as λ increases from zero, the vector $\underline{x}(\lambda)$ that minimizes $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{X}^+$, leaves the boundary of $\mathcal{X}^+$ at $\lambda = \check{\lambda}$, say, where $\check{\lambda}$ is at most $7/8$. The path $\underline{x}(\lambda)$, $0 < \lambda < \check{\lambda}$, is the straight line from $(0, 0)$ to $\check{\underline{a}}$ in Figure 1. Because formula (4.8) provides $y(7/8) = 14/37$, the point $\check{\underline{a}}$ is well inside the strip $\mathcal{S}$. We find $\check{\lambda} = 0.858835$ later, so $\check{\underline{a}}$ has the coordinates $(0.67442, 0.33721)$.

The second derivative matrix of the quadratic (4.5) is positive definite as λ increases from zero until it becomes singular at $\lambda = 0.735840$. Hence, for $\lambda = \check{\lambda}$, the Lagrange function $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{X}^+ \cap \mathcal{S}$, has a direction of negative curvature. Therefore, when $\underline{x}(\lambda)$ leaves the boundary of $\mathcal{X}^+$ at $\lambda = \check{\lambda}$, as mentioned above, it jumps to a point $\check{\underline{b}}$, say, that is on or beyond the $y = \frac{1}{2}$ boundary of $\mathcal{S}$. We deduce from the continuity of expression (4.2) that the function $L(\cdot, \lambda)$ at $\check{\underline{b}}$ is the quadratic

$$x^2 + y^2 + \lambda\,\{\,-\tfrac{3}{4}\,x^2 - \tfrac{1}{4}\,xy - y^2 - \tfrac{1}{2}\,x + \tfrac{5}{2}\,\}, \qquad \underline{x} \in \mathcal{R}^2, \tag{4.9}$$

which is strictly convex for $0 < \lambda < 0.950772$. Its unconstrained minimizer has the components

$$x(\lambda) \;=\; \frac{16\,\lambda - 16\,\lambda^2}{64 - 112\,\lambda + 47\,\lambda^2} \quad \text{and} \quad y(\lambda) \;=\; \frac{2\,\lambda^2}{64 - 112\,\lambda + 47\,\lambda^2}. \tag{4.10}$$

They have the properties $x(\lambda) > 0$ and $y(\lambda) \geq \frac{1}{2}$ for $0.846607 \leq \lambda < 0.950772$. In particular, the components of $\check{\underline{b}}$ take the values $x(\check{\lambda}) = 0.782945$ and $y(\check{\lambda}) = 0.595420$. The thick curve from $\check{\underline{b}}$ to $\check{\underline{c}}$ in Figure 1 is the path (4.10) as λ increases from $\check{\lambda}$ to $\lambda = 0.916165$, the last value being defined by $y(\lambda) = 2$. We see that the path includes the solution $\check{\underline{x}}$ of NLP, which occurs at $\lambda = 8/9$.

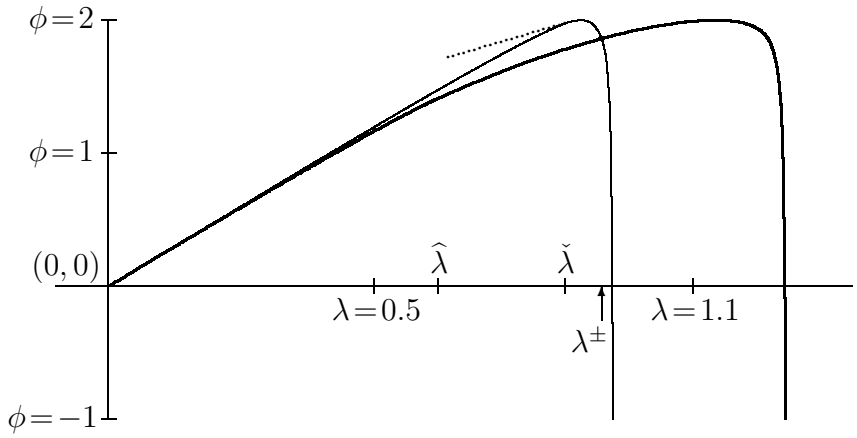

Figure 2: Graphs of $\phi(\lambda)$, $\lambda\in\mathcal{R}$, when $\mathcal{X}=\mathcal{X}^-$ and $\mathcal{X}=\mathcal{X}^+$

When $\mathcal{X}=\mathcal{X}^-$, the value of $\phi(\lambda)$ is the function (4.5) with the components (4.4) or the function (4.6) with the components (4.7), which gives the formula

$$\phi(\lambda)=\begin{cases}(1856\lambda-3936\lambda^2+1915\lambda^3)\,/\,(768-1600\lambda+756\lambda^2), & 0\le\lambda\le\widehat{\lambda},\\ (512\lambda-652\lambda^2+196\lambda^3)\,/\,(192-208\lambda+45\lambda^2), & \widehat{\lambda}\le\lambda\le\widehat{\mu},\end{cases} \tag{4.11}$$

where $\widehat{\mu}$ is set to 1.271843, because this value provides $\phi(\widehat{\mu})=-1$. Similarly, when $\mathcal{X}=\mathcal{X}^+$, expressions (4.5), (4.8), (4.9) and (4.10) supply the formula

$$\phi(\lambda)=\begin{cases}(290\lambda-283\lambda^2)\,/\,(120-116\lambda), & 0\le\lambda\le\check{\lambda},\\ (320\lambda-568\lambda^2+243\lambda^3)\,/\,(128-224\lambda+94\lambda^2), & \check{\lambda}\le\lambda\le\check{\mu},\end{cases} \tag{4.12}$$

where $\check{\mu}=0.948348$ is defined by $\phi(\check{\mu})=-1$. Plots of the functions (4.11) and (4.12) are displayed in Figure 2, the thicker and thinner curves being for the cases $\mathcal{X}=\mathcal{X}^-$ and $\mathcal{X}=\mathcal{X}^+$, respectively. Both functions have the derivative $\phi'(0)=29/12$, which agrees with Theorem 3, because equation (4.1) shows that the least value of $L(\underline{x},0)$, $\underline{x}\in\mathcal{R}^2$, is at the origin, and the constraint function (4.2) takes the value $f_1(0)=29/12$. The thinner curve in the figure is above the thicker one for sufficiently small λ. Indeed, for $0\le\lambda\le\check{\lambda}$, the value (4.11) is the minimum of $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{X}^-$, but the value (4.12) is the minimum of $L(\underline{x},\lambda)$ on the boundary of $\mathcal{X}^-$, the boundaries of $\mathcal{X}^-$ and $\mathcal{X}^+$ being the same. We see that both functions in Figure 2 are continuous and concave, as stated in Theorems 1 and 2. Further, although the direction of the lower path in Figure 1 is discontinuous at $\widehat{\underline{a}}=\underline{x}(\widehat{\lambda})$, the thick curve of Figure 2 has a continuous first derivative at $\widehat{\lambda}$, in accordance with Theorem 3. The jump from $\check{\underline{a}}$ to $\check{\underline{b}}$ in Figure 1, however, reduces the gradient of the thin curve of Figure 2 from 1.6423 to 1.1777 at $\lambda=\check{\lambda}$. The dotted line in the figure helps to expose this reduction, because it is the tangent to the thin curve $\phi(\lambda)$, $\lambda\ge\check{\lambda}$, at $\lambda=\check{\lambda}$. We find the value $\check{\lambda}=0.858853$ by solving a quartic polynomial equation, given by the remark that $\check{\lambda}$ is the unique number in the interval $(0,\check{\mu})$ such that the function (4.12) is continuous.

The arrow in Figure 2 indicates the value of λ where the thick and the thin curves cross. This value, $\lambda=\lambda^{\pm}$ say, is defined by the property that the second parts of expressions (4.11) and (4.12) are the same, so the solution of another polynomial equation gives $\lambda^{\pm}=0.927829$. Moreover, when $\mathcal{X}=\mathcal{X}^{\pm}=\mathcal{X}^{-}\cup\mathcal{X}^{+}$, equation (4.3) can be written in the form

$$\phi(\lambda) \;=\; \min\left\{\min\{L(\underline{x},\lambda):\underline{x}\in\mathcal{X}^{-}\},\;\min\{L(\underline{x},\lambda):\underline{x}\in\mathcal{X}^{+}\}\right\},\quad \lambda\in\mathcal{R}. \tag{4.13}$$

It follows that the graph of this function in Figure 2 is the thick curve or the thin curve when λ is in the interval $[0,\lambda^{\pm}]$ or $[\lambda^{\pm},\check{\mu}]$, respectively. We see that this function is continuous and concave, with a first derivative discontinuity at $\lambda=\lambda^{\pm}$. Further, every point $\underline{x}(\lambda)$, $0\leq\lambda\leq\check{\mu}$, is unique, except that, when $\lambda=\lambda^{\pm}$, equation (4.13) allows $\underline{x}(\lambda)$ to be taken from either $\mathcal{X}^{-}$ or $\mathcal{X}^{+}$. We call these two points $\widehat{\underline{s}}\in\mathcal{X}^{+}$ and $\check{\underline{s}}\in\mathcal{X}^{-}$. They have the coordinates $\widehat{\underline{s}}=(0.450,-0.820)$ and $\check{\underline{s}}=(1.970,3.166)$, which are derived by substituting $\lambda=\lambda^{\pm}=0.927829$ into expressions (4.7) and (4.10). Hence the path $\underline{x}(\lambda)$, $0\leq\lambda\leq\lambda^{\pm}$ in Figure 1 is now the thick line from $(0,0)$ to $\widehat{\underline{s}}$ through $\widehat{\underline{a}}$, but at $\lambda=\lambda^{\pm}$ the path jumps to $\check{\underline{s}}$, which is outside the figure beyond $\check{\underline{c}}$. We see that, when $\mathcal{X}=\mathcal{X}^{\pm}$, none of the points $\underline{x}(\lambda)$ comes close to the locus of the equality constraint. On the other hand, in the cases $\mathcal{X}=\mathcal{X}^{-}$ and $\mathcal{X}=\mathcal{X}^{+}$, the maximization of the concave function ϕ provides $\phi'(\lambda)=f_1(\underline{x}(\lambda))=0$, so $\underline{x}(\lambda)$ satisfies the constraints of NLP. Further, the minimization of the Lagrange function (4.1) takes up the remaining freedom in $\underline{x}(\lambda)$ automatically by minimizing the objective function of NLP.

5. Final remarks

We have assumed so far that, for each choice of $\underline{\lambda}\in\mathcal{R}^m$, a vector $\underline{x}(\underline{\lambda})$ that minimizes $L(\underline{x},\underline{\lambda})$, $\underline{x}\in\mathcal{X}$, can be calculated, as shown in equation (1.3). We are going to address now, however, the more usual situation when an algorithm for local minimization is applied to $L(\underline{x},\underline{\lambda})$, $\underline{x}\in\mathcal{X}$. We let $\underline{x}(\underline{\lambda}_{\infty})$ be generated in this way for some $\underline{\lambda}_{\infty}$ in $\mathcal{R}^m$, this notation being used for the dual variables in order to apply Theorem 3. The point $\underline{x}(\underline{\lambda}_{\infty})$ is a strict local minimizer of $L(\cdot,\underline{\lambda}_{\infty})$ if there exists a neighbourhood $\mathcal{N}$ of $\underline{x}(\underline{\lambda}_{\infty})$ such that $\underline{x}(\underline{\lambda}_{\infty})$ is the unique global minimizer of the function $L(\underline{x},\underline{\lambda}_{\infty})$, $\underline{x}\in\mathcal{X}\cap\mathcal{N}$. Then we replace $\mathcal{X}$ by $\mathcal{X}\cap\mathcal{N}$ temporarily, in order to satisfy the conditions of Theorem 3. Thus $\underline{\nabla}\phi(\underline{\lambda})$ at $\underline{\lambda}=\underline{\lambda}_{\infty}$ has the components (2.5) for the temporary choice of $\mathcal{X}$. Therefore, because we are trying to maximize the dual function $\phi(\underline{\lambda})=L(\underline{x}(\underline{\lambda}),\underline{\lambda})$, $\underline{\lambda}\in\mathcal{R}^m$, it is suitable to give the next trial vector of dual variables, $\underline{\lambda}_{\rm new}$, say, the ascent property

$$(\underline{\lambda}_{\rm new}-\underline{\lambda}_{\infty})^T\,\underline{\nabla}\phi(\underline{\lambda}_{\infty}) \;=\; (\underline{\lambda}_{\rm new}-\underline{\lambda}_{\infty})^T\,\underline{f}(\underline{x}(\underline{\lambda}_{\infty})) \;>\; 0, \tag{5.1}$$

where $\underline{f}(\underline{x}(\underline{\lambda}_{\infty}))$ has the components $f_j(\underline{x}(\underline{\lambda}_{\infty}))$, $j=1,2,\ldots,m$. We also require the dual variables of any inequality constraints to be nonnegative, this restriction being imposed on both $\underline{\lambda}_{\infty}$ and $\underline{\lambda}_{\rm new}$. It is straightforward to satisfy all these

conditions on $\underline{\lambda}_{\rm new}$ if one (or more) of the f_j constraints in expression (1.1) is violated at $\underline{x}=\underline{x}_\infty$. Thus Theorem 3 may be useful to the adjustment of the dual variables when each $\underline{x}(\underline{\lambda})$ is found by an algorithm for local minimization. Furthermore, if $\|\underline{\lambda}_{\rm new}-\underline{\lambda}_\infty\|$ is not too large, the vector $\underline{x}(\underline{\lambda}_\infty)$ is likely to be a good starting point for the local minimization of $L(\underline{x},\underline{\lambda}_{\rm new})$, $\underline{x}\in\mathcal{X}$.

We apply this procedure to the example of Section 4. Again we write λ instead of $\underline{\lambda}$, because there is only one f_j constraint in NLP, namely $f_1(\underline{x})=0$. The ascent condition (5.1) states that $\lambda_{\rm new}-\lambda_\infty$ shall be positive or negative if $f_1(\underline{x}(\lambda_\infty))$ is positive or negative, respectively. In other words, the current value of the dual variable is increased or decreased if $\underline{x}(\lambda_\infty)$ is inside or outside the closed curve of Figure 1. We consider each of the three choices of $\mathcal{X}$ in Section 4. We set λ_∞ to zero initially, which gives $\underline{x}(\lambda_\infty)=(0,0)$. The dual variable λ is increased until the calculated local minimizer $\underline{x}(\lambda)$ reaches or goes beyond the boundary of the constraint $f_1(\underline{x})=0$.

In the case $\mathcal{X}=\mathcal{X}^-$, when λ is increased gradually from $\lambda=0$, the vectors $\underline{x}(\lambda)$ follow the path in Figure 1 from $(0,0)$ through $\widehat{\underline{a}}$ and $\widehat{\underline{s}}$ to $\widehat{\underline{x}}=\underline{x}(8/7)$, because each calculated local minimum is also the global minimum of $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{X}$. The situation $\lambda_\infty<8/7$ and $\lambda_{\rm new}>8/7$ may occur, and then the sign of $f_1(\underline{x}(\lambda))$ changes from positive to negative, which indicates that the required value of λ is in the interval $(\lambda_\infty,\lambda_{\rm new})$. Thus the updating of λ can be regarded as solving the nonlinear equation $f_1(\underline{x}(\lambda))=0$.

We show next that the case $\mathcal{X}=\mathcal{X}^\pm$ is the same as the case $\mathcal{X}=\mathcal{X}^-$ when local minima of $L(\cdot,\lambda)$ are calculated. The points $\underline{x}(\lambda)$ on the path from $(0,0)$ to $\widehat{\underline{s}}$ in Figure 1 are the unique global minimizers of the function $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{R}^2$, for $0\leq\lambda<\lambda^\pm$, the value $\lambda^\pm=0.927829$ being introduced in Figure 2. Further, the points on the continuation of this path from $\widehat{\underline{s}}$ to $\widehat{\underline{c}}$ are the unique global minimizers of the function $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{X}^-$, as λ increases from $\lambda^\pm$ to 1.238060, and they are interior points of $\mathcal{X}^-$. Thus all of these points are local minimizers of the function $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{X}^\pm$, which allows them to be generated by local minimization when λ is increased gradually. Thus the solution $\widehat{\underline{x}}$ of NLP can be calculated, although it is shown in Section 4 that global minimization causes every vector $\underline{x}(\lambda)$ to be far from the locus of the equality constraint.

In the case $\mathcal{X}=\mathcal{X}^+$, both global and local minimization of $L(\underline{x},\lambda)$, $\underline{x}\in\mathcal{X}$, provide points $\underline{x}(\lambda)$ on the $y=\frac{1}{2}x$ boundary of $\mathcal{X}^+$ for sufficiently small $\lambda\geq 0$. Then, as stated in Section 4, $\underline{x}(\lambda)$ has the components (4.8), and $\underline{\nabla}L(\underline{x}(\lambda),\lambda)$ is the vector $(1,-2)$ multiplied by $(7\lambda-8\lambda^2)/(30-29\lambda)$. This $\underline{x}(\lambda)$ is a local minimizer of $L(\cdot,\lambda)$ when λ satisfies $0\leq\lambda<7/8$, but we have found already that it is a global minimizer only for $0\leq\lambda\leq\check{\lambda}=0.858835$. It follows that, when λ is in the interval $(\check{\lambda},7/8)$, the global minimizers are on the path from $\check{\underline{b}}$ to $\check{\underline{c}}$ in Figure 1, but the local minimizers $\underline{x}(\lambda)$ can remain on the boundary of $\mathcal{X}^+$, between $\check{\underline{a}}=(0.67442,0.33721)$ and $\underline{x}(7/8)=(28/37,14/37)=\underline{x}_0$, say. To discover the fate of the local minimizers for $\lambda>7/8$, we begin a local minimization of the Lagrange function $L(\underline{x},7/8)$, $\underline{x}\in\mathcal{X}^+$, at the point $\underline{x}_0$. We see above that the gradient $\underline{\nabla}L(\underline{x}_0,7/8)$ is zero, and $\lambda=7/8$ gives the Lagrange function (4.5) some negative

curvature. Therefore a single line search from $\underline{x}_0$ for the least value of $L(\cdot, 7/8)$ on the line can cross the $y=\frac{1}{2}$ boundary of the strip $\mathcal{S}$. Thus the Lagrange function becomes the strictly convex quadratic (4.9), whose minimizer is unique and has the components (4.10) with $\lambda = 7/8$. Hence the search for the local minimum of $L(\underline{x}, 7/8)$, $\underline{x} \in \mathcal{X}^+$, supplies the point $(112/127, 98/127)$, which is on the path from $\check{\underline{b}}$ to $\check{\underline{x}}$ in Figure 1. The situation where local minimization is the same as global minimization has been restored, so increases in λ from 7/8 to 8/9, with local minimization of $L(\cdot, \lambda)$, lead to the solution of NLP at $\check{\underline{x}}$.

We give further attention to differences between local and global minimization of $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{X}^+$, in the previous paragraph. Local minimization can provide points $\underline{x}(\lambda)$ that are close to $\underline{x}_0 = (28/37, 14/37)$ or to $(112/127, 98/127) = \underline{x}_1$, say, when λ is slightly less than or slightly greater than 7/8, respectively. Further, the Lagrange function (4.1) takes the values

$$L(\underline{x}_0, 7/8) \;=\; \tfrac{2373}{1184} \;=\; 2.004223 \quad \text{and} \quad L(\underline{x}_1, 7/8) \;=\; \tfrac{4053}{2032} \;=\; 1.994587. \tag{5.2}$$

It follows that local minimization causes a discontinuity in $\phi(\lambda)$ when $\underline{x}(\lambda)$ jumps from the boundary of $\mathcal{X}^+$ to the path between $\check{\underline{b}}$ and $\check{\underline{x}}$ in Figure 1. On the other hand, when each $\underline{x}(\lambda)$ is generated by global minimization, then Theorem 1 states that the dual function (1.3) is continuous.

In the original version of the example of Section 4, the boundary between the regions $\mathcal{X}^-$ and $\mathcal{X}^+$ was the line $y = 0$ instead of the line $y = \frac{1}{2}x$. The change was made because the purpose of the example is to demonstrate some typical properties of the Lagrangian method, but the original version has the following unexpected degeneracy. We retain the equations (4.1) and (4.2) that define $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{R}^2$, $\lambda \in \mathcal{R}$, and we retain the definition $\mathcal{X}^{\pm} = \{\underline{x} : \|\underline{x}\| \leq M\}$, where M is a large positive constant, but now we let $\mathcal{X}^+ \subset \mathcal{R}^2$ be the region $\mathcal{X}^{\pm} \cap \{\underline{x} : y \geq 0\}$. The global minimizer $\underline{x}(\lambda)$ of $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{X}^+$, is on the $y = 0$ boundary of $\mathcal{X}^+$ for sufficiently small $\lambda \geq 0$. Expression (4.5) is still the Lagrange function on the strip $\mathcal{S} = \{\underline{x} : |y| \leq \frac{1}{2}\}$, which reduces to the function of one variable

$$\ell(x, \lambda) \;=\; x^2 + \lambda \left\{ -\tfrac{3}{4}x^2 - \tfrac{1}{2}x + \tfrac{29}{12} \right\}, \qquad x \in \mathcal{R}, \tag{5.3}$$

when y is zero. Hence $\underline{x}(\lambda)$ has the components

$$x(\lambda) \;=\; \lambda \,/\, (4 - 3\,\lambda) \quad \text{and} \quad y(\lambda) \;=\; 0 \tag{5.4}$$

for sufficiently small λ. In particular, at $\lambda = 8/9$ we find $x(\lambda) = 2/3$ and $\phi(\lambda) = \ell(2/3, 8/9) = 2$. Now the global minimizer $\underline{x}(\lambda)$ of $L(\cdot, \lambda)$ jumps from the $y = 0$ boundary of $\mathcal{X}^+$ across the $y = \frac{1}{2}$ boundary of $\mathcal{S}$ when λ increases through the value λ_0, say, this jump being analogous to the one from $\check{\underline{a}}$ to $\check{\underline{b}}$ in Figure 1. As before, $\lambda = \lambda_0$ can be calculated by solving the polynomial equation that makes the value of the function (4.5) at the point (5.4) the same as the value of the function (4.9) at the point (4.10). It has been noted already, however, that at $\lambda = 8/9$ the components (4.10) provide the solution $\check{\underline{x}} = (1, 1)$ of NLP, and there the value of

the function (4.1) is 2, because $f_1(\check{\underline{x}})$ is zero. This value of the Lagrange function is the same as $\ell(2/3, 8/9) = 2$, given above, which implies $\lambda_0 = 8/9$. Therefore the degeneracy of the original version of our example is that, when $\underline{x}(\lambda)$ leaves the $y = 0$ boundary of $\mathcal{X}^+$, which happens at $\underline{x}(\lambda) = (2/3, 0)$, it jumps directly to $\check{\underline{x}}$.

The points $\underline{x}(\lambda)$ that are calculated by local minimization of $L(\underline{x}, \lambda)$, $\underline{x} \in \mathcal{X}^+$, when $\mathcal{X}^+$ is the set $\mathcal{X}^{\pm} \cap \{\underline{x} : y \geq 0\}$ are interesting too. As before, they have the components (5.4) for sufficiently small $\lambda \geq 0$, and at this point the gradient $\underline{\nabla} L(\underline{x}(\lambda), \lambda)$ is the vector $(0, 1)$ multiplied by $(\frac{4}{3}\lambda - \frac{5}{4}\lambda^2)/(4 - 3\lambda)$. Therefore $\underline{x}(\lambda)$ is a local minimum for $0 \leq \lambda < 16/15$. Then, at $\lambda = 16/15$, the gradient becomes zero, and also the Lagrange function (4.5) has directions of negative curvature. Thus local minimization with $\lambda = 16/15$ reduces $L(\cdot, \lambda)$ by moving from the $y = 0$ boundary of $\mathcal{X}^+$ across the $y = \frac{1}{2}$ boundary of $\mathcal{S}$. The form of $L(\cdot, \lambda)$ changes from the quadratic (4.5) to the function (4.9), which is another quadratic with some negative curvature. It follows that local minimization takes $\underline{x}(16/15)$ to the $\|\underline{x}\| = M$ boundary of the region $\mathcal{X}^+ \cap \{\underline{x} : y \geq \frac{1}{2}\}$. Because this boundary is well outside the $f_1(\underline{x}) = 0$ closed curve in Figure 1, the sign of $f_1(\underline{x}(16/15))$ is negative. Therefore condition (5.1) causes λ to decrease gradually from 16/15, until $\underline{x}(\lambda)$ reaches or crosses the $f_1(\underline{x}) = 0$ curve, each new $\underline{x}(\lambda)$ being calculated by local minimization. The point $\underline{x}(\lambda)$ cannot leave the $\|\underline{x}\| = M$ boundary of $\mathcal{X}^+$ until the second derivative matrix of the Lagrange function (4.9) becomes positive definite, which happens at $\lambda = 0.950772$. Then $\underline{x}(\lambda)$ becomes the minimizer of this function, which has the coordinates (4.10). Thus $\underline{x}(\lambda)$ is on an extension of the path through $\check{\underline{b}}$, $\check{\underline{x}}$ and $\check{\underline{c}}$ in Figure 1, the extension being to the circle $\|\underline{x}\| = M$. Local minimization for smaller values of $\lambda \geq 0.858835$ provides points $\underline{x}(\lambda)$ that are also on this path, the number 0.858835 being the value of λ that supplies $\check{\underline{b}}$. Thus $\underline{x}(\lambda)$ can reach the solution $\check{\underline{x}}$ of NLP at $\lambda = 8/9$.

We now return to SAO algorithms that seek the solutions of a sequence of NLPs. We let the main calculation be the minimization of $f_0^*(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, subject to the constraints

$$\left.\begin{array}{ll} f_j^*(\underline{x}) = 0, & j = 1, 2, \ldots, m' \\ f_j^*(\underline{x}) \leq 0, & j = m'+1, \ldots, m \\ \text{and} & \underline{x} \in \mathcal{X} \end{array}\right\}, \tag{5.5}$$

and we require the gradients $\underline{\nabla} f_j^*(\underline{x})$, $\underline{x} \in \mathcal{X}$, to be available. Each outermost iteration chooses the functions f_j, $j = 0, 1, \ldots, m$, for the next NLP, and then applies the bounded dual version of the Lagrange method, the value of Λ in the conditions (1.9) being a prescribed positive constant. The functions f_j of the k-th outermost iteration are quadratics that satisfy the conditions

$$f_j(\underline{x}^{(k)}) \;=\; f_j^*(\underline{x}^{(k)}) \quad \text{and} \quad \underline{\nabla} f_j(\underline{x}^{(k)}) \;=\; \underline{\nabla} f_j^*(\underline{x}^{(k)}), \quad j = 0, 1, \ldots, m, \tag{5.6}$$

where $\underline{x}^{(k)} \in \mathcal{X}$ is the best vector of variables so far at the beginning of the k-th outermost iteration, $\underline{x}^{(1)}$ being given. The second derivatives of the quadratic functions f_j, $j = 0, 1, \ldots, m$, have to provide the property that the minimizer of $L(\underline{x}, \underline{\lambda})$, $\underline{x} \in \mathcal{X}$, is unique for each $\underline{\lambda}$ that occurs, which can be done by adding a

sufficiently large multiple of $\|\underline{x}-\underline{x}^{(k)}\|^2$, $\underline{x}\in\mathcal{R}^n$, to f_0 if necessary, but we are not assuming at present that the second derivative matrices are diagonal. Let $\underline{\lambda}_\infty^{(k)}$ be the vector that maximizes $\phi(\underline{\lambda})$, $\underline{\lambda}\in\mathcal{L}$, in the k-th outermost iteration, and let $\underline{x}_\infty^{(k)}$ be the minimizer of $L(\underline{x},\underline{\lambda}_\infty^{(k)})$, $\underline{x}\in\mathcal{X}$. If another outermost iteration is required, then $\underline{x}^{(k+1)}$ is set to either $\underline{x}^{(k)}$ or $\underline{x}_\infty^{(k)}$.

Theorem 4 is highly valuable for making the choice between these alternatives. Because we are applying the bounded dual method, it allows us to take the view that the main calculation is the minimization of the function

$$\Psi^*(\underline{x}) \;=\; f_0^*(\underline{x}) \;+\; \Lambda\sum_{j=1}^{m'}|\,f_j^*(\underline{x})\,| \;+\; \Lambda\sum_{j=m'+1}^{m}\max[\,0,\,f_j^*(\underline{x})\,],\qquad \underline{x}\in\mathcal{X}. \quad (5.7)$$

Therefore we set $\underline{x}^{(k+1)}=\underline{x}_\infty^{(k)}$ or $\underline{x}^{(k+1)}=\underline{x}^{(k)}$ if $\Psi^*(\underline{x}_\infty^{(k)})<\Psi^*(\underline{x}^{(k)})$ or $\Psi^*(\underline{x}_\infty^{(k)})\geq\Psi^*(\underline{x}^{(k)})$ holds, respectively. Of course we require the calculation of $\underline{x}_\infty^{(k+1)}$ from $\underline{x}^{(k+1)}$ to be different from the calculation of $\underline{x}_\infty^{(k)}$ from $\underline{x}^{(k)}$ in the case $\underline{x}^{(k+1)}=\underline{x}^{(k)}$.

Trust region methods are helpful for achieving $\Psi^*(\underline{x}_\infty^{(k)})<\Psi^*(\underline{x}^{(k)})$. Specifically, we let the compact set $\mathcal{X}$ during the k-th outermost iteration be the intersection $\mathcal{X}^*\cap\{\underline{x}:\|\underline{x}-\underline{x}^{(k)}\|\leq\rho^{(k)}\}$, where $\mathcal{X}^*$ is the given set $\mathcal{X}$ of the main calculation and where $\rho^{(k)}$ is a positive constant for each k. If $\Psi^*(\underline{x}_\infty^{(k)})\geq\Psi^*(\underline{x}^{(k)})$ occurs, causing $\underline{x}^{(k+1)}$ to be set to $\underline{x}^{(k)}$, then the choice $\rho^{(k+1)}=\frac{1}{2}\|\underline{x}_\infty^{(k)}-\underline{x}^{(k)}\|$ would ensure that $\underline{x}_\infty^{(k+1)}$ is different from $\underline{x}_\infty^{(k)}$. Furthermore, smaller values of $\rho^{(k)}$ assist $\Psi^*(\underline{x}_\infty^{(k)})<\Psi^*(\underline{x}^{(k)})$ to be inherited from the fact that $\Psi(\underline{x}_\infty^{(k)})$ is the least value of $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, the reason being that the definitions (1.11) and (5.7), with the interpolation equations (5.6), imply that the approximation $\Psi(\underline{x})\approx\Psi^*(\underline{x})$ has first order accuracy when $\underline{x}$ is close to $\underline{x}^{(k)}$. If $\mathcal{X}^*$ is given by the simple bounds (1.7), it is convenient to employ the infinity vector norm in the trust region method, because then $\underline{x}\in\mathcal{R}^n$ is in the set $\mathcal{X}^*\cap\{\underline{x}:\|\underline{x}-\underline{x}^{(k)}\|\leq\rho^{(k)}\}$ when its components satisfy the constraints

$$\max[\,\ell_i,\,x_i^{(k)}-\rho^{(k)}\,]\;\leq\;x_i\;\leq\;\min[\,u_i,\,x_i^{(k)}+\rho^{(k)}\,],\qquad i=1,2,\ldots,n. \quad (5.8)$$

We deduce from the last remark that the advantages of separability, mentioned in Section 1, can be enjoyed by trust region methods. We let the second derivative matrices of the quadratic functions f_j, $j=0,1,\ldots,m$, be diagonal, in order that these functions take the form (1.6), we employ the infinity norm in the definition $\{\underline{x}:\|\underline{x}-\underline{x}^{(k)}\|\leq\rho^{(k)}\}$ of the trust region of the k-th outermost iteration, and we require the given compact set $\mathcal{X}^*$ to be defined by the simple bounds (1.7). Then, for every $\underline{\lambda}\in\mathcal{L}$, the components of $\underline{x}(\underline{\lambda})$ minimize the functions of one variable in expression (1.8), after replacing $\ell_i\leq x_i\leq u_i$ by the constraints (5.8). We compare this technique with another idea, which is to calculate $\underline{x}_\infty^{(k)}$ from $\underline{x}^{(k)}$ by minimizing the function (1.11) directly, which eliminates the need for $\underline{\lambda}\in\mathcal{L}$. Unfortunately, however, the separability that allows n to be huge is eliminated too. Therefore it may be advantageous to employ SAO and the dual variables $\underline{\lambda}\in\mathcal{L}$ in several applications that require the vector $\underline{x}$ that minimizes a function of the form (5.7).

The "majorization" technique, which is also known as "conservatism", may be applied in SAO if there are no equality constraints ($m'=0$), and if the given

functions f_j^*, $j=0,1,\ldots,m$, have bounded second derivatives. Then the functions f_j of the k-th outermost iteration, besides satisfying the equations (5.6), are given the property

$$f_j(\underline{x}) \;\geq\; f_j^*(\underline{x}), \qquad \underline{x}\in\mathcal{X}, \qquad j=0,1,\ldots,m, \tag{5.9}$$

which can be done by making the diagonal elements of the second derivative matrices $\nabla^2 f_j$, $j=0,1,\ldots,m$, sufficiently large. Then the definitions (1.11) and (5.7) imply $\Psi(\underline{x}) \geq \Psi^*(\underline{x})$, $\underline{x}\in\mathcal{X}$. Therefore, in the usual case when $\Psi(\underline{x}^{(k)})$ is not the least value of $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, the calculations of the k-th outermost iteration provide the strict inequality $\Psi(\underline{x}^{(k)})>\Psi(\underline{x}_\infty^{(k)})\geq\Psi^*(\underline{x}_\infty^{(k)})$. It follows from $\Psi(\underline{x}^{(k)})=\Psi^*(\underline{x}^{(k)})$ and $\underline{x}^{(k+1)}=\underline{x}_\infty^{(k)}$ that the sequence $\Psi^*(\underline{x}^{(k)})$, $k=1,2,3,\ldots$, decreases strictly monotonically, except that $\underline{x}_\infty^{(k)}=\underline{x}^{(k)}$ may occur, which is a suitable condition for terminating the sequence of outermost iterations.

A disadvantage of majorization is that high curvature in the functions f_j, $j=0,1,\ldots,m$, causes fast second order growth in $\Psi(\underline{x})$, $\underline{x}\in\mathcal{X}$, when $\underline{x}$ moves away from a small neighbourhood of $\underline{x}^{(k)}$, where we are retaining $m'=0$ and the equations (1.11) and (5.6). Hence the vector $\underline{x}_\infty^{(k)}$ that minimizes $\Phi(\cdot)$ is close to $\underline{x}^{(k)}$. Instead of these small distances $\|\underline{x}_\infty^{(k)}-\underline{x}^{(k)}\|$, better efficiency is likely if the chosen second derivatives are designed to make $f_j(\underline{x}_\infty^{(k)})$ close to $f_j^*(\underline{x}_\infty^{(k)})$ for every index j in $[0,m]$. We accept not only violations of the inequalities (5.9) at $\underline{x}=\underline{x}_\infty^{(k)}$ but also the possibility $\Phi^*(\underline{x}_\infty^{(k)})>\Phi(\underline{x}_\infty^{(k)})$, because the k-th outermost iteration is successful if the difference $\Phi^*(\underline{x}^{(k)})-\Phi^*(\underline{x}_\infty^{(k)})$ captures most of the predicted reduction $\Phi(\underline{x}^{(k)})$–$\Phi(\underline{x}_\infty^{(k)})$. Otherwise, if $\Phi^*(\underline{x}^{(k)})$–$\Phi^*(\underline{x}_\infty^{(k)})$ compares unfavourably with $\Phi(\underline{x}^{(k)})-\Phi(\underline{x}_\infty^{(k)})$, we try to improve the accuracy of the approximations $f_j(\underline{x}_\infty^{(k)})\approx f_j^*(\underline{x}_\infty^{(k)})$, when the diagonal elements of the second derivative matrices $\nabla^2 f_j$, $j=0,1,\ldots,m$, are chosen on the next outermost iteration.

The author is not trying to specify a particular version of the SAO method, because he has not run any numerical experiments himself that investigate possible options. Instead, the main purpose of the comments above is to support the claim that Theorem 4 is likely to be very helpful in practice. Groenwold and Etman (2010), however, present numerical results for four implementations of SAO. They investigate not only practical versions of majorization but also the use of the filter method (Fletcher and Leyffer, 2002), that balances reductions in the objective function with gains in feasibility.

Theorem 3 is also helpful, partly because the functions f_j, $j=0,1,\ldots,m$, do not have to be differentiable, and the generality of the compact set $\mathcal{X}$ is welcome too. This theorem shows the importance of the uniqueness of $\underline{x}(\underline{\lambda})$ for the current $\underline{\lambda}$, which is illustrated by the example of Section 4. The uniqueness is unnecessary, however, when $\underline{x}(\underline{\lambda})$ is a strict local minimum that is calculated by local minimization. Then we can apply again the technique that replaces $\mathcal{X}$ temporarily by the intersection $\mathcal{X}\cap\{\underline{x}:\|\underline{x}-\underline{x}(\underline{\lambda})\|\leq\rho\}$, where $\rho>0$ is such that $\underline{x}(\underline{\lambda})$ becomes the global minimizer of $L(\underline{x},\underline{\lambda})$, $\underline{x}\in\mathcal{X}$. Thus Theorem 3 provides the gradient $\underline{\nabla}\phi(\underline{\lambda})$ for the temporary choice of $\mathcal{X}$, which can be regarded as a gradient that is suitable for local calculations, after restoring the original $\mathcal{X}$.

Although local calculations provide solutions to the three test problems of Section 4, the different problems being generated by different choices of $\mathcal{X}$, the Lagrange method with local minimization is inadequate for finding local solutions of many NLPs. In particular, let f_0 be a quadratic function with some negative curvature, let all the constraint functions f_j, $j = 1, 2, \ldots, m$, be linear, and let $\mathcal{X}_0$ be the set of vectors $\underline{x}$ in $\mathcal{R}^n$ that satisfy the f_j constraints of expression (1.1). Further, let $\mathcal{X}_0$ be nonempty and bounded, and let $\mathcal{X}$ be so large that every point of $\mathcal{X}_0$ is an interior point of $\mathcal{X}$. This setting implies that NLP has at least one solution, all of them being in $\mathcal{X}_0$. The linear constraints, however, give the property $\nabla^2 L(\underline{x}, \underline{\lambda}) = \nabla^2 f_0(\underline{x})$ for every $\underline{\lambda} \in \mathcal{R}^m$, so the negative curvature of f_0 causes every $\underline{x}(\underline{\lambda})$ to be on the boundary of $\mathcal{X}$. Thus, although $\mathcal{X}_0$ is nonempty, every calculated $\underline{x}(\underline{\lambda})$ violates at least one constraint by an amount that is bounded away from zero.

A major difficulty of this quadratic programming problem is that, if $\nabla^2 f_0$ has many negative eigenvalues, then the number of local minima can be huge. It is usually possible to calculate one or a few of them after applying a version of the augmented Lagrangian method to modify f_0, as mentioned in Section 1. If the original f_0 is retained, however, then again we may replace $\mathcal{X}$ temporarily by the intersection of the original $\mathcal{X}$ with a trust region. Furthermore, a new violation of a constraint can be avoided by including the constraint explicitly in the definition of the current $\mathcal{X}$. Such changes may also be permanent. For example, the condition $f_m(\underline{x}) \leq 0$ can be removed from expression (1.1) by replacing the original $\mathcal{X}$ by the compact set $\mathcal{X} \cap \{\underline{x} : f_m(\underline{x}) \leq 0\}$.

Another approach is to apply SAO to the original NLP, choosing the quadratic approximations to the original functions within each outermost iteration so that the Lagrange function $L(\underline{x}, \underline{\lambda})$, $\underline{x} \in \mathcal{R}^n$, has positive curvature for every $\underline{\lambda}$. The given theory and remarks are highly relevant to this technique. Therefore they may assist greatly the construction of new algorithms for constrained optimization, especially if bounds are imposed on the dual variables.

Acknowledgements

It is mentioned in Section 1 that I was introduced to SAO methods by Albert Groenwold. I cannot thank him enough for renewing my interest in the use of Lagrange functions. I am also very grateful to the Mathematical Sciences Department of the University of Stellenbosch, where I received much support, encouragement and hospitality for two months while the given work was begun. Helpful comments on a draft of this paper were made by Pascal Etman and by Daniel Robinson. I am indebted to Oleg Burdakov for the reference to Bazaraa *et al* (2006).

References

M.S. Bazaraa, H.D. Sherali and C.M. Shetty (2006), *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons (New York).

L.F.P. Etman, A.A. Groenwold and J.E. Rooda (2011), "First-order sequential convex programming using approximate diagonal QP subproblems", *Struct. Multidisc. Optim.*, Online First, 28 November, 2011.

R. Fletcher (1981), "Numerical experiments with an exact L_1 penalty function method", in *Nonlinear Programming 4*, eds. O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press (New York), pp. 99–129.

R. Fletcher (1987), *Practical Methods of Optimization*, John Wiley & Sons (Chichester).

R. Fletcher and S. Leyffer (2002), "Nonlinear programming without a penalty function", *Math. Programming*, Vol. 91, pp. 141–159.

P.E. Gill and D.P. Robinson (2010), "A primal-dual augmented Lagrangian", *Comput. Optim. Appl.*, Online First, 3 August 2010.

A.A. Groenwold and L.F.P. Etman (2008), "Sequential approximate optimization using dual subproblems based on incomplete series expansions", *Struct. Multidisc. Optim.*, Vol. 36, pp. 547–570.

A.A. Groenwold and L.F.P. Etman (2010), "On the conditional acceptance of iterates in SAO algorithms based on convex separable approximations", *Struct. Multidisc. Optim.*, Vol. 42, pp. 165–178.

A.A. Groenwold, L.F.P. Etman and D.W. Wood (2010), "Approximated approximations for SAO", *Struct. Multidisc. Optim.*, Vol. 41, pp. 39–56.

J. Nocedal and S.J. Wright (1999), *Numerical Optimization*, Springer (New York).

D.W. Wood, A.A. Groenwold and L.F.P. Etman (2011), "Bounding the dual of Falk to circumvent the requirement of relaxation in globally convergent SAO algorithms", Draft Technical Computing Report, Dept. of Mechanical Engineering, University of Stellenbosch, South Africa.