Journal of Global Optimization **9**: 141–151, 1996.

Second-Order Sufficient Optimality Conditions for Local and Global Nonlinear Programming

ARNOLD NEUMAIER
AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974-0636, U.S.A.
(email: neum@cma.univie.ac.at)

(Received: 12 October 1994; accepted: 29 December 1995)

Abstract. This paper presents a new approach to the sufficient conditions of nonlinear programming. Main result is a sufficient condition for the global optimality of a Kuhn–Tucker point. This condition can be verified constructively, using a novel convexity test based on interval analysis, and is guaranteed to prove global optimality of strong local minimizers for sufficiently narrow bounds. Hence it is expected to be a useful tool within branch and bound algorithms for global optimization.

Key words: Global optimization, optimality condition, second-order sufficient condition, verification of convexity, interval analysis.

Mathematics Subject Classifications (1991). primary 90C30, secondary 65G10.

In this note we consider smooth nonlinear programs of the form

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\quad u \le x \le v\,, \quad F(x) = 0\,, \end{aligned} \tag{1}$$

where $u, v \in \mathbb{R}^n (u < v)$ define the *box*

$$[u, v] := \{x \in \mathbb{R}^n \mid u \le x \le v\}$$

with nonempty interior, and $f : [u, v] \to \mathbb{R}$, $F : [u, v] \to \mathbb{R}^r$ are continuously differentiable.

Here (as always in the following) all inequalities involving vectors are interpreted componentwise. We shall allow some of the bounds to be infinite, $u_k = -\infty$ or $v_k = \infty$, to allow for one-sided bounds and unbounded variables; but in practice it is advisable to add sensible bounds to all variables. By the introduction of slack variables, arbitrary nonlinear programs can be written in this form.

In the following, $g = \nabla f$ denotes the gradient of f. If K is a set of indices, x_K denotes the subvector of a vector x obtained by discarding rows not indexed by K, $A_{\bullet K}$ (or $A_{K\bullet}$) is the submatrix of a matrix A obtained by discarding columns (or rows) not indexed by K, and A_{KK} is the square submatrix of a square matrix A obtained by discarding rows and columns not indexed by K. $A_{\bullet k}$ denotes the kth column of a matrix A.

Section 1 gives our main result in Theorem 1.2, and some additional remarks. In Section 2 we show that the sufficient global optimality condition derived in

Theorem 1.2 is strong enough to deduce from it the standard second order sufficient conditions for local minimizers (Theorem 2.3) when the box constraint is narrowed to a small neighbourhood of a Kuhn–Tucker point. At the same time, our new proof of these conditions gives important hints for a constructively verifiable version of the global optimality condition that reduces the global optimality proof to one of verifying convexity of the generalized augmented Lagrangian. Section 3 shows that this can be done by means of interval analytic techniques.

1. Global Optimality of Kuhn–Tucker Points

For the nonlinear program (1), the most general necessary conditions for optimality, the Karush–John conditions, can be formulated as follows.

THEOREM 1.1. *(First order optimality conditions for nonlinear programs with two-sided bounds). Let* $\hat{x} \in \mathbb{R}^n$ *be a solution of the nonlinear program* (1) *with* $u < v$*. Then there are multipliers* $\kappa \geq 0$ *and* $\hat{z} \in \mathbb{R}^r$*, not both zero, such that the vector*

$$\hat{y} := \kappa g(\hat{x}) - F'(\hat{x})^T \hat{z} \tag{2}$$

satisfies the two-sided complementarity condition

$$\left.\begin{array}{ll} \hat{y}_k \geq 0 & \text{if } \hat{x}_k = u_k\,, \\ \hat{y}_k \leq 0 & \text{if } \hat{x}_k = v_k\,, \\ \hat{y}_k = 0 & \text{otherwise.} \end{array}\right\} \tag{3}$$

Moreover, if the constraint qualification

$$\operatorname{rk} F'(\hat{x})_{\cdot K_0} = r\,, \qquad \text{where } K_0 = \{k \mid u_k < \hat{x}_k < v_k\} \tag{4}$$

(or one of a number of similar constraint qualifications) holds, we can take $\kappa = 1$.

Proof. This follows easily from the corresponding conditions in the standard form with general inequalities (Karush [5], John [4]) by eliminating the multipliers corresponding to the bound constraints. □

A feasible point x such that (2) and (3) hold for $\kappa = 1$ and suitable $\hat{z}$ is generally referred to as a *Kuhn–Tucker point* (after Kuhn & Tucker [6]), with associated *Lagrange multiplier* $\hat{z}$ (after Lagrange [7], Part II, Ch. XI, Sec. 58). Except for convex programs, the first order conditions are usually not sufficient for optimality, and Kuhn–Tucker points need not be local minima.

However, we shall show that it is possible to modify the original objective function in such a way that convexity of the modification still provides optimality for the original problem. This modification can often be achieved locally, and suffices to derive a semilocal version of well-known sufficient conditions for local optimality. Here the adjective *semilocal* refers to the fact that the new sufficient condition actually verifies not only local optimality but in fact global optimality within a *pre-specified* local region around the local minimum.

This fact allows our new condition to be employed as a natural stopping criterion guaranteeing finite termination in branch and bound algorithms for global constrained optimization, at least in the absence of degeneracy. Indeed, this can be achieved in a similar way as a more specialized sufficient condition for concave optimization is used in Horst & Tuy [2] (Theorem V.1). But algorithmic aspects will not be considered in the present paper.

THEOREM 1.2. *Let $\hat{x}$ be a Kuhn–Tucker point for the nonlinear program* (1), *with associated multiplier $\hat{z}$, and let*

$$\hat{y} := g(\hat{x}) - F'(\hat{x})^T \hat{z}, \tag{5}$$

$$D = \operatorname{Diag}\left(\sqrt{\frac{2|\hat{y}_1|}{v_1 - u_1}}, \ldots, \sqrt{\frac{2|\hat{y}_n|}{v_n - u_n}}\right). \tag{6}$$

If, for some continuously differentiable function $\varphi : \mathbb{R}^m \to \mathbb{R}$ with

$$\varphi(0) = 0, \ \ \varphi'(0) = \hat{z}^T, \tag{7}$$

the generalized augmented Lagrangian

$$\hat{L}(x) := f(x) - \varphi(F(x)) + \frac{1}{2}\|D(x - \hat{x})\|_2^2 \tag{8}$$

is convex in $[u, v]$, then $\hat{x}$ is a global solution of (1). *Moreover, if $\hat{L}(x)$ is strictly convex in $[u, v]$, this solution is unique.*

Proof. Clearly, $\hat{L}(\hat{x}) = f(\hat{x}) - \varphi(0)$, and since

$$\hat{L}'(x) = g(x)^T - \varphi'(F(x))F'(x) + (x - \hat{x})^T D^T D$$

we have

$$\hat{L}'(\hat{x}) = g(\hat{x})^T - \varphi'(0)F'(\hat{x}) = g(\hat{x})^T - \hat{z}^T F'(\hat{x}) = \hat{y}^T.$$

Therefore, for any $x \in [u, v]$, convexity yields

$$\hat{L}(x) \geq \hat{L}(\hat{x}) + \hat{L}'(\hat{x})(x - \hat{x}) = f(\hat{x}) + \hat{y}^T(x - \hat{x}).$$

If x is feasible then $F(x) = 0$, so that

$$f(x) = \hat{L}(x) + \varphi(0) - \frac{1}{2}\|D(x-\hat{x})\|_2^2 \geq f(\hat{x}) + \hat{y}^T(x-\hat{x}) - \frac{1}{2}\|D(x-\hat{x})\|_2^2. \tag{9}$$

Since D is diagonal, this implies that

$$f(x) \geq f(\hat{x}) + \frac{1}{2}\sum_k \left(2\hat{y}_k(x_k - \hat{x}_k) - D_{kk}^2(x_k - \hat{x}_k)^2\right). \tag{10}$$

The complementarity condition together with the definition of D now implies that each term in the sum is nonnegative: If $\hat{y}_k = 0$ then $D_{kk} = 0$ and $D_{kk}^2(x_k - \hat{x}_k)^2 = 0 = 2\hat{y}_k(x_k - \hat{x}_k)$. If $\hat{y}_k > 0$ then $\hat{x}_k = u_k$ and $D_{kk}^2(x_k - \hat{x}_k)^2 \leq D_{kk}^2(v_k - u_k)(x_k - \hat{x}_k) = 2\hat{y}_k(x_k - \hat{x}_k)$. And if $\hat{y}_k < 0$ then $\hat{x}_k = v_k$ and $D_{kk}^2(x_k - \hat{x}_k)^2 = D_{kk}^2(\hat{x}_k - x_k)^2 \leq D_{kk}^2(v_k - u_k)(\hat{x}_k - x_k) = 2|\hat{y}_k|(\hat{x}_k - x_k) = 2\hat{y}_k(x_k - \hat{x}_k)$. Therefore, (10) implies that $f(x) \geq f(\hat{x})$ for all $x \in [u, v]$ with $F(x) = 0$. Thus, $\hat{x}$ is a global solution of (1). Uniqueness follows along the same line by observing that strict convexity excludes equality except for $x = \hat{x}$. □

REMARK 1.3. The same result holds if we use in place of (6) an arbitrary (not necessarily diagonal) matrix D with

$$\|D_{\bullet k}\|_2^2 \leq 2\hat{y}_k^2/\sigma \quad (k = 1, \ldots, n), \tag{11}$$

where

$$\sigma = \sup\left\{\hat{y}^T(x - \hat{x}) \mid u \leq x \leq v, F(x) = 0\right\}$$

(or any computable upper bound). Indeed, in this case (9) implies

$$f(x) \geq f(\hat{x}) + \frac{1}{2}\min\left\{2|\hat{y}|^T s - \|Cs\|_2^2 \mid s \geq 0, |\hat{y}|^T s \leq \sigma\right\}, \tag{12}$$

where $s = \Sigma(x - \hat{x}), C = D\Sigma$ with $\Sigma = \mathrm{Diag}(\operatorname{sgn}\hat{y}_1, \ldots, \operatorname{sgn}\hat{y}_n)$. But (11) implies that the minimum in (12) is zero, as a consequence of the following result (similar to Proposition 1 of Neumaier [9], used to derive sufficient conditions for global quadratic programs), applied for $a = 2|\hat{y}|, b = |\hat{y}|/\sigma$:

PROPOSITION 1.4. *Let $a, b \in \mathbb{R}^n$ be nonnegative, and suppose that $C \in \mathbb{R}^{m\times n}$ satisfies*

$$\|C_{\bullet k}\|_2^2 \leq a_k b_k \quad (k = 1, \ldots, n).$$

Then

$$\|Cx\|_2^2 \leq a^T x \textit{ for all } x \geq 0 \textit{ with } b^T x \leq 1. \tag{13}$$

Proof. We may assume that all components b_k are positive since the general case follows from this by continuity. The function f defined by $f(x) := a^T x - \|Cx\|_2^2$ is concave, hence assumes its minimum over the simplex $\{x \geq 0 \mid b^T x \leq 1\}$ at a vertex. Thus the minimum is either at $x = 0$ (where $f = 0$), or at a scaled unit vector $x = b_k^{-1}e^{(k)}$ (where $f = a_k b_k^{-1} - \|C_{\bullet k}\|_2^2 b_k^{-2} \geq 0$ by assumption). Thus $f(x) \geq 0$ on the simplex, giving the desired bound (13). □

2. Local Second Order Optimality Conditions

To discuss second-order optimality conditions we shall assume for the following that f and F are twice continuously differentiable in the box $[u, v]$.

For the simplest choice conforming with (7), namely $\varphi(s) = \hat{z}^T s$, Theorem 1.2 implies that $\hat{x}$ is a local solution of (1) if the Hessian

$$\hat{G} := L''(\hat{x}) = f''(\hat{x}) - \sum_j \hat{z}_j F_j''(\hat{x}) \tag{14}$$

at $\hat{x}$ of the *Lagrangian*

$$L(x) := f(x) - \hat{z}^T F(x) \tag{15}$$

is positive definite. Indeed, if $\hat{G} = L''(\hat{x})$ is positive definite then so is $\hat{L}''(\hat{x}) = L''(\hat{x}) + D^T D$. Therefore $\hat{L}(x)$ is strictly convex in a neighborhood of $\hat{x}$, and we can apply Theorem 1.2 in this neighborhood. Hence $\hat{x}$ is a local solution.

However, the simple univariate example

$$\min -x^2 \quad \text{s.t.} \quad -1 \leq x \leq 1$$

shows that, at a local solution, the Hessian of the Lagrangian need not even be positive semidefinite. Thus the definiteness condition on the Hessian is too restrictive; necessary conditions only provide semidefiniteness of $\hat{G}$ on a subspace:

THEOREM 2.1. *(Second order necessary optimality conditions for nonlinear programs with two-sided bounds). Let $\hat{x}$ be a local solution of* (1), *and let*

$$J_0 = \{k \mid \hat{x}_k = u_k \text{ or } \hat{x}_k = v_k\}. \tag{16}$$

With the above notation, if the constraint qualification (4) *is satisfied then the following, equivalent conditions hold:*

(i) $F'(\hat{x})s = 0, \quad s_{J_0} = 0 \Rightarrow s^T \hat{G} s \geq 0.$

(ii) For some matrix (and hence all matrices) Z_0 whose columns form a basis of the subspace defined by the left-hand side of (i), $Z_0^T \hat{G} Z_0$ is positive semidefinite.

Proof. This is well-known but for the sake of completeness we give a short proof.

(i) The constraint qualification (4) and the implicit function theorem imply that in a neighborhood of $\hat{x}$, the manifold of solutions of $F(x) = 0, x_{J_0} = \hat{x}_{J_0}$ can be parametrized as $\{x(s) \mid s \in S\}$, where

$$S = \{s \in \mathbb{R}^n \mid F'(\hat{x})s = 0, \;\; s_{J_0} = 0\},$$

in such a way that $x(s) = \hat{x} + s + o(\|s\|)$. Hence, for sufficiently small $s \in S$, we have $0 \leq f(x(s)) - f(\hat{x}) = L(x(s)) - L(\hat{x}) = L'(\hat{x})(x(s) - \hat{x}) + \frac{1}{2}(x(s) - \hat{x})^T L''(\hat{x})(x(s) - \hat{x}) + o((\|s\|^2) = s^T \hat{G} s + o((\|s\|^2)$. If we now replace s by ϵs, divide by ϵ^2, and take the limit $\epsilon \to 0$, we arrive at $s^T \hat{G} s \geq 0$.

(i) $\Leftrightarrow$ (ii): If (i) holds then $s = Z_0 p$ satisfies the hypothesis of (i) so that $0 \leq s^T \hat{G} s = p^T Z_0^T \hat{G} Z_0 p$. Thus $Z_0^T \hat{G} Z_0$ is positive semidefinite, and (ii) holds. Conversely, any s satisfying the hypothesis of (i) can be written as $s := Z_0 p$. Hence, if $Z_0^T \hat{G} Z_0$ is positive semidefinite, $s^T \hat{G} s = p^T Z_0^T \hat{G} Z_0 p \geq 0$, and (i) holds. $\square$

REMARKS 2.2. (i) Generally, the Hessian of the Lagrangian agrees with the Hessian of the objective function only when all constraints are linear ($F_i''(x) = 0$). Thus the curvature of the constraints enters the second order conditions in an essential way.

(ii) A slightly more careful argument shows that one can strengthen the necessary condition in Theorem 2.1(i) to

(i') If $F'(\hat{x})s = 0$ and

$$\begin{array}{lll} s_k = 0 & \text{if} & \hat{y}_k \neq 0, \\ s_k \geq 0 & \text{if} & \hat{y}_k = 0,\ \hat{x}_k = u_k, \\ s_k \leq 0 & \text{if} & \hat{y}_k = 0,\ \hat{x}_k = v_k, \end{array}$$

then $s^T\hat{G}s \geq 0$.

However, this condition is much more difficult to check, and is the reason that deciding local optimality (in the absense of strict complementarity) is NP-hard (cf. Pardalos & Schnitger [10] and the recent book by Horst *et al.* [3]).

We show now that, with a more sophisticated choice of φ, we can get sufficient optimality conditions where – as in the unconstrained case – the gap to the necessary conditions is only slight. Of course, these sufficient conditions can also be derived from the traditional sufficient conditions for problems with arbitrary equality and inequality constraints, see e.g., Fletcher [1], Section 9.3. The present proof is given to show that the new global sufficient condition is indeed strong enough to imply the local result. Moreover, details in the proof are needed in Section 3 to motivate a constructively verifiable version of the global optimality condition.

THEOREM 2.3. *(Second order sufficient optimality conditions for nonlinear programs with two-sided bounds). With the above notation, let*

$$J = \{k \mid \hat{y}_k \neq 0\}, \quad K = \{k \mid \hat{y}_k = 0\}.$$

A sufficient condition for a local minimizer is that any of the following, equivalent conditions holds:

(i) $F'(\hat{x})s = 0, \quad s_J = 0 \Rightarrow s^T\hat{G}s > 0$ *or* $s = 0$.

(ii) For some matrix (and hence all matrices) Z whose columns form a basis of the subspace defined by the left-hand side of (i), $Z^T\hat{G}Z$ is positive definite.

(iii) For some matrix (and hence all matrices) A whose rows form a basis of the row space of $F'(\hat{x})_{:K}$,

$$\hat{G}_{KK} + \beta A^T A \quad \textit{is positive definite for some} \quad \beta \geq 0. \tag{17}$$

A Kuhn–Tucker point satisfying the above sufficient conditions for optimality is called a *strong local minimizer* of (1). Both matrices $Z^T\hat{G}Z$ and $Z_0^T\hat{G}Z_0$ are referred to as *reduced Hessians* of the Lagrangian.

If we compare with (16) and (4) we see that $J \subseteq J_0$ and $K \supseteq K_0$ by the complementarity condition (3), hence the condition in Theorem 2.3(i) is a little more restrictive than that in Theorem 2.1(i). However, in practice we usually have *strict complementarity*, i.e., the condition $J = J_0$ (equivalently $K = K_0$) holds at least for *some* choice of the multiplier vector $\hat{z}$, and then the difference between Theorem 2.1(i) and Theorem 2.3(i) is only marginal.

Proof. We first prove that (iii) implies local optimality. The construction of A implies that there is a matrix B such that $A = BF'(\hat{x})_{\bullet K}$. If we define

$$\varphi(s) := \hat{z}^T s - \frac{\beta}{2}\|Bs\|^2 \tag{18}$$

then the generalized augmented Lagrangian (8) satisfies

$$\hat{L}''(\hat{x}) = H + D^T D, \tag{19}$$

where $H = \hat{G} + \beta F'(\hat{x})^T B^T B F'(\hat{x})$. By assumption, the matrix $H_{KK} = \hat{G}_{KK} + \beta A^T A$ is positive definite. Since, for u_k, v_k sufficiently close to $\hat{x}_k$, the diagonal entries D_{kk}^2 ($k \in J$) become arbitrarily large by (6), we conclude from (19) and continuity that $\hat{L}''(x)$ is positive definite in every sufficiently narrow box around $\hat{x}_k$. Local optimality therefore follows from Theorem 1.2. It remains to prove the equivalence of (i) – (iii).

(iii) $\Rightarrow$ (i): If (17) holds then the hypothesis of (i) implies $As_K = 0$, hence $s^T\hat{G}s = s_K^T\hat{G}_{KK}s_K = s_K^T(\hat{G}_{KK} + \beta A^T A)s_K > 0$ unless $s = 0$. Thus (i) is valid.

(i) $\Rightarrow$ (ii): If (i) holds then $p \neq 0$ implies $s := Zp \neq 0$, hence $0 < s^T\hat{G}s = p^T Z^T \hat{G} Z p$. Thus $Z^T\hat{G}Z$ is positive definite, and (ii) holds.

(ii) $\Rightarrow$ (iii): Since the rows of A are linearly independent we can find a matrix P such that $AP = I$. If (ii) holds, the Cholesky factorization LL^T of $Z^T\hat{G}Z$ produces a nonsingular L, and we may define the matrices

$$N := \hat{G}ZL^{-T},$$

$$M := P^T(\hat{G} - NN^T)_{KK}P.$$

Now let $x \in \mathbb{R}^n$ be arbitrary and $w := Ax_K$. The vector s with $s_J = 0$ and $s_K := x_K - Pw$ satisfies $As_K = w - APw = 0$ and hence the hypothesis of (i). Hence $s = Zp$ for some p, and we find

$$\begin{aligned} x_K^T\hat{G}_{KK}x_K &= (s_K + Pw)^T\hat{G}_{KK}(s_K + Pw) \\ &= s_K^T\hat{G}_{KK}s_K + 2(Pw)^T\hat{G}_{KK}s_K + (Pw)^T\hat{G}_{KK}Pw. \end{aligned} \tag{20}$$

Using

$$s_K^T\hat{G}_{KK}s_K = s^T\hat{G}s = p^T Z^T\hat{G}Zp = p^T L^T L p$$

and

$$\hat{G}_{KK}s_K = \hat{G}_{K\bullet}s = \hat{G}_{K\bullet}Zp = N_{K\bullet}L^T p,$$

we can rewrite (20) as

$$\begin{aligned} x_K^T \hat{G}_{KK} x_K &= p^T L^T L p + 2(Pw)^T N_{K\bullet} L^T p + (Pw)^T \hat{G}_{KK} P w \\ &= \|L^T p + N_{K\bullet}^T P w\|_2^2 + w^T M w. \end{aligned} \tag{21}$$

If β is larger than the spectral radius of M (or only of the negative of its smallest eigenvalue) we can deduce that

$$\begin{aligned} x_K^T(\hat{G}_{KK} + \beta A^T A)x_K &= x_K^T \hat{G}_{KK} x_K + \beta w^T w \\ &= \|L^T p + N_{K\bullet}^T P w\|_2^2 + w^T(M + \beta I)w \geq 0. \end{aligned}$$

Equality is possible only when $w = 0$ and the norm vanishes, leading to $L^T p = 0$, $p = 0$, $s = Zp = 0$ and hence to $x_K = 0$. This shows that $\hat{G}_{KK} + \beta A^T A$ is positive definite. □

3. Verifiable Sufficient Conditions for Global Optimality

In the previous section we showed that, in principle, Theorem 1.2 can be used to verify global optimality in any sufficiently narrow box around a Kuhn–Tucker point satisfying the local second-order sufficient conditions. Within a branch and bound framework, this ensures that no infinitely fine subdivision is needed near the global optimizer, thus removing a serious problem from current branch and bound implementations.

However, to be useful for algorithmic applications, all steps in the verification process must be made fully constructive. Standard algorithms for local optimization generally provide, together with a Kuhn–Tucker point, also an associated Lagrange multiplier. Second order methods often provide a basis Z of the relevant null space, too, and a Cholesky factor of the reduced Hessian of the Lagrangian at $\hat{x}$.

To verify the hypothesis of Theorem 1.2, we must first choose a suitable functional φ. The proof of Theorem 2.3 shows that (18) is a sufficiently general choice. Since A is most easily produced as a submatrix of $F'(\hat{x})_{\bullet K}$, B in (18) will be a monomial matrix, with exactly one nonzero per row. A suitable value for β can be found by calculating the smallest eigenvalue of the matrix M defined in the proof. The construction of M given in the proof only uses simple linear algebra; the right inverse P of A needed is typically available implicitly through the factorization which also provides Z; explicit Z and P are not needed when the computation is suitably arranged.

This leaves the convexity check of the generalized augmented Lagrangian as the only nontrivial part of the global optimality verification procedure. And indeed, this causes some problems, since $\varphi(x)$ as given by (18) is concave so that convex and concave parts in (8) may partially cancel. A general method for checking convexity of arbitrary C^1 functions in a box using finitely many function values is impossible since convexity over a given box is no longer a local property. Therefore we can only hope to get a working method by restricting the kind of objective functions and constraints admitted.

For quadratic programs, the generalized augmented Lagrangian is a quadratic function, and convexity is simply checked by attempting a Cholesky factorization of its Hessian. (This gives a much simpler sufficient condition than my previous result in [9].) For nonquadratic programs, we must use more sophisticated techniques.

It turns out that a useful black box convexity check is possible with methods of interval analysis, at least in the case where all nonlinearities are given by arithmetic expressions and the boxes are not too wide. The basic paradigm is that with interval arithmetic and automatic differentiation, it is possible to calculate intervals containing the ranges of a function and its partial derivatives when the argument ranges over a box. The computed intervals usually overestimate the precise ranges; however, under very mild conditions on the form of the expressions, the width of the computed interval is of the order $O(r)$ when the width of all components of the box is of order $O(r)$. See, e.g., Neumaier [8], where the relevant background can be found.

In particular, if $\mathbf{G}$ is a matrix of intervals (usually simply called an *interval matrix*), calculated as an enclosure of $\hat{L}''(x)$ for $x \in [u, v]$, then, with $r = \max\{v_k - u_k \mid k = 1, \dots, n\}$, we generally have

$$|\mathbf{G} - \hat{G}| = O(r).$$

Such a statement implies a corresponding statement for all individual matrices $\tilde{G} \in \mathbf{G}$, with absolute values taken component-wise. In particular, if $\hat{G}$ is positive definite then all matrices in $\mathbf{G}$ are definite, too, provided the underlying box is not too wide. This was precisely the kind of argument used in the proof of Theorem 2.3, and it shows that the information provided by interval analysis is precisely what is needed to make the present results fully constructive.

The only task remaining is to give a constructive criterion for simultaneously checking the definiteness of all members of an interval matrix. We give just one such result, showing that it can be done; various refinements are possible, using more of the machinery of interval analysis. However, since the methods are so different, details will be discussed elsewhere.

THEOREM 3.1. *(Sufficient conditions for convexity). Let $f : [u, v] \to \mathbb{R}$ be twice continuously differentiable on the compact box $[u, v]$, and suppose that $\mathbf{G}$ is a symmetric interval matrix such that*

$$f''(x) \in \mathbf{G} \quad \textit{for all} \quad x \in [u, v]. \tag{22}$$

(i) If some symmetric matrix $G_0 \in \mathbf{G}$ is positive definite and all symmetric matrices in $\mathbf{G}$ are nonsingular then they are all positive definite, and f is uniformly convex in $[u, v]$.

(ii) In particular, this holds if the midpoint matrix

$$\check{G} = (\sup \mathbf{G} + \inf \mathbf{G})/2$$

is positive definite with inverse C, and the preconditioned radius matrix

$$\Delta = |C| \operatorname{rad} \mathbf{G},$$

where

$$\operatorname{rad} \mathbf{G} = (\sup \mathbf{G} - \inf \mathbf{G})/2,$$

satisfies the condition

$$\|\Delta\| < 1 \tag{23}$$

(in an arbitrary norm).

Proof. (i) Since the eigenvalues are continuous functions of the matrix entries and the product of the eigenvalues (the determinant) cannot vanish, no eigenvalue changes sign. Hence the eigenvalues of all matrices in $\mathbf{G}$ are positive, since this is the case for the positive definite member. Thus all matrices in $\mathbf{G}$ are positive definite. By well-known results, uniform convexity of f now follows from (22).

(ii) $G_0 = \check{G}$ belongs to $\mathbf{G}$, and condition (23) implies strong regularity of the interval matrix $\mathbf{G}$ ([8], Section 4.1) and hence nonsingularity of all matrices in $\mathbf{G}$. Thus (i) applies. □

REMARK 3.2. In many cases, the Hessian of the augmented Lagrangian can be shown to have the form

$$f''(x) = \sum u_i A_i \quad \text{with } u_i \in \mathbf{u}_i,$$

for suitable constructively available real matrices A_i and intervals $\mathbf{u}_i$. In this case, the above result can be strengthened (with virtually the same proof) by replacing $\check{G}$ and Δ with

$$\check{G} = \sum \check{u}_i A_i$$

and

$$\Delta' = \sum \operatorname{rad} \mathbf{u}_i |CA_i|,$$

respectively, where

$$\check{u} = (\sup \mathbf{u} + \inf \mathbf{u})/2$$

and

$$\operatorname{rad} \mathbf{u} = (\sup \mathbf{u} - \inf \mathbf{u})/2.$$

Indeed, it is not difficult to see that for $G = \sum \mathbf{u}_i A_i$, we always have $0 \leq \Delta' \leq |\Delta|$, so that the refined test is easier to satisfy.

By applying this convexity test to the generalized augmented Lagrangian in place of f, we have a criterion which verifies the convexity of $\hat{L}(x)$ for $x \in [u, v]$ in arbitrary sufficiently narrow boxes $[u, v]$ around a strong local minimizer $\hat{x}$, and thus proves that $\hat{x}$ is the unique global minimizer within such a box.

As an aid for an actual implementation, we add a few final remarks. The Hessian of the generalized augmented Lagrangian is

$$G(x) = \hat{L}''(x) = f''(x) - \sum_j \varphi'(F_j(x))F_j''(x) + \beta F'(x)^T B^T B F'(x) + D^T D,$$

and reduces to the expression for H given after (19) when $x = \hat{x}$. Thus the interval evaluation $\mathbf{G} := G([u, v])$ gives an interval matrix satisfying (22) with $\hat{L}$ in place of f. Alternative forms for $\hat{L}''(x)$ seem to suffer from a higher degree of overestimation when intervals are inserted.

We construct an appropriate function $\varphi(F(x))$ using (18). Actually, many matrix B will do and define a corresponding A as in the line before (18), but there may be better and worse choices. The assumption in (iii) allows quite a lot of freedom in the choice of A. The natural choice is to use Gaussian elimination with partial pivoting (perhaps after equilibration) to select a subset of rows of F'. In this case, B is simply a $(0, 1)$-projection matrix that picks these rows. β must just be chosen such that (17) holds; and the line after (21) shows how to construct it explicitly.

References

1. Fletcher, R. (1987), *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester.
2. Horst, R. and Tuy, H. (1990), *Global Optimization: Deterministic Approaches*, 2nd ed., Springer, Berlin, 1990.
3. Horst, R. Pardalos, P.M. and Thoai, N.V. (1995), *Introduction to Global Optimization*, Kluwer, Dordrecht 1995.
4. John, F. (1985), pp. 543–560 in: *Fritz John, Collected Papers*, Vol.2, Extremum problems with inequalities as subsidiary conditions (J. Moser, ed.), Birkhäuser, Boston, 543–560.
5. Karush, W. (1939), Minima of Functions of Several Variables with Inequalitis as Side Constraints, M.Sc. Dissertation, Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois.
6. Kuhn, H.W. and Tucker, A.W. (1951) pp. 481–492 in: *Nonlinear Programming*, Proc. 2nd Berkeley Symp. Math. Stat. Prob. (J. Neyman, ed.), Univ. of California Press, Berkeley, California.
7. Lagrange, J.L. (1797), *Théorie des fonctions analytiques*, Impr. de la République, Paris
8. Neumaier, A. (1990), *Interval Methods for Systems of Equations*, Cambridge Univ. Press, Cambridge.
9. Neumaier, A. (1992) An optimality criterion for global quadratic optimization, *J. Global Optimization* **2**, 201–208.
10. Pardalos, P.M. and Schnitger, G. (1988), Checking local optimality in constrained quadratic programming is NP-hard, *Oper. Res. Lett.* **7**, 33–35.