Consiglio Nazionale delle Ricerche

ISTITUTO DI ELABORAZIONE
DELLA INFORMAZIONE

PISA

ON SUBMODULARITY IN CONTINUOUS OPTIMIZATION

Fabio Tardella

Nota interna B4 - 57

Dicembre 1990

On submodularity in continuous optimization

Fabio Tardella

Istituto di Elaborazione della Informazione
Via S. Maria, 46
56100 Pisa, Italy

1. Introduction and notations

Submodularity is a well-known notion in combinatorial optimization. Some of its properties and applications where systematically analyzed for the first time by Edmonds in connection with the theory of matroids [3] . After Edmonds' work, many other authors have contributed to the development of submodularity theory and its applications to combinatorial optimization. Good surveys on this topic may be found in [8,9] and a collection of papers on submodular optimization has recently appeared in [2] .

An important motivation of the interest for submodular functions is the possibility of minimizing in polynomial time any submodular function on the 0-1 hypercube $B^n = \{0,1\}^n$ (see [5]) . This result, which is based on the ellipsoid algorithm, has been extended in various directions: in [6] it was proved that a submodular function can be minimized in polynomial time on any crossing or intersecting family contained in B^n, while in [4] it was shown that the problem of minimizing a submodular integrally convex function on any discrete rectangle in $\mathbf{Z}^n$ can also be solved in polynomial time.

To our knowledge, no efficient algorithm has been developed until now for the minimization of a submodular function in the continuous case. Lovasz remarked that submodularity may, in a sense, be viewed as a discrete version of the *continuous* notion of convexity. Somewhat in contrast with this analogy, we believe that submodularity is an independent property which may be fruitfully exploited also in continuous optimization, possibly in conjunction with convexity. We intend to support this viewpoint in the following sections.

In section 2 we analyze some properties of submodular functions on $\mathbf{R}^n$. The problem of finding a stationary point for the minimization of a submodular function on a rectangle in $\mathbf{R}^n$ may be viewed as a special type of variational inequality problem and, in this more general setting, is addressed in section 3. Finally, in section 4 we present two algorithms for minimizing a convex submodular function in $\mathbf{R}^n$ with lower and/or upper bounds on the variables.

Let us now introduce some definitions and notations that will be used in the sequel. Given two points $x,y \in \mathbf{R}^n$, the points $x \wedge y$ and $x \vee y$ are defined by $(x \wedge y)_i = \min \{x_i, y_i\}$ and $(x \vee y)_i = \max \{x_i, y_i\}$, for $i=1, \ldots ,n$. Furthermore, the inequalities $x \leq y$ and $x < y$ mean $x_i \leq y_i$ and $x_i < y_i$ for $i=1, \ldots ,n$, respectively. A function $f : X \subset \mathbf{R}^n \to \mathbf{R}^m$ is called isotone (antitone), if

$f(x) \leq f(y)$ ($f(y) \leq f(x)$) whenever $x \leq y$. Given $x \in \mathbf{R}^n$ and $I \subset \{1, \ldots, n\}$, x_I denotes the subvector of x formed by the components with index in I. A point $z \in \mathbf{R}^n$ is a lower (upper) bound for a set $X \subset \mathbf{R}^n$ if $z \leq x$ ($x \leq z$) for every $x \in \mathbf{R}^n$. If z is a lower bound for X and $x \in X$, then z is called the least (greatest) element of X. It is trivial to check that the least (greatest) element of a set, if it exists, is unique. A set $X \subset \mathbf{R}^n$ is a sublattice of $\mathbf{R}^n$ if $x \wedge y \in \mathbf{R}^n$ and $x \vee y \in \mathbf{R}^n$ for every $x, y \in \mathbf{R}^n$. A rectangle in $\mathbf{R}^n$ is a set of the form $X = \{x \in \mathbf{R}^n : \ell \leq x \leq u\}$, where $\ell_i \geq -\infty$ and $u_i \leq +\infty$ for $i=1, \ldots, n$. X is sometimes also denoted $[\ell, u]$. Clearly, a rectangle is a particular type of sublattice of $\mathbf{R}^n$. Given a function $f: X \subset \mathbf{R}^n \rightarrow \mathbf{R}$, a (strict) global minimum point for f on X is a point $x^* \in X$ such that $f(x^*) \leq f(x)$ ($f(x^*) < f(x)$) for every $x \in X$ with $x \neq x^*$. A point x' is called a stationary point (or, equivalently, is said to satisfy the first order conditions) for the problem of minimizing a function f on the rectangle $[\ell, u]$, if and only if $x' \in [\ell, u]$ and, for every $i=1, \ldots, n$, either $f_{x_i}(x')=0$ or $x'_i = \ell_i$ and $f_{x_i}(x') \geq 0$ or $x'_i = u_i$ and $f_{x_i}(x') \leq 0$. Finally, e_i denotes the i^{th} unit vector in $\mathbf{R}^n$ for $i=1, \ldots, n$.

2. Submodular functions on $\mathbf{R}^n$

Definition 2.1

Let X be a subset of $\mathbf{R}^n$. A function $f: X \rightarrow \mathbf{R}$ is submodular on X if

$$f(x \wedge y) + f(x \vee y) \leq f(x) + f(x),$$

for every couple of points $x, y \in \mathbf{R}^n$ such that $x \wedge y \in \mathbf{R}^n$ and $x \vee y \in \mathbf{R}^n$.

Several results concerning the minimization of a submodular function on a lattice (not necessarily contained in $\mathbf{R}^n$) and some applications of this problem, together with an economic interpretation of submodularity, may be found in [15]. In this section, we recall a characterization of submodular functions on $\mathbf{R}^n$ in terms of their first and second derivatives and we establish a result which may help to restrict the region where the global minimum points of a submodular function are to be sought.

Definition 2.2

Let X be a rectangle in $\mathbf{R}^n$. A function $F : X \to \mathbf{R}^n$ is called off-diagonally antitone on X if $x,y \in X$ and $x \leq y$ imply $F_i(x) \geq F_i(y)$ for every i such that $x_i = y_i$.

It is easy to verify that F is off-diagonally antitone if and only if the functions $g_{ij}(t) = F_i(x + t e_j)$ are antitone for every $i \neq j$. Off-diagonally antitone functions are a nonlinear generalization of Z-matrices and have been studied in connection with nonlinear equations and nonlinear complementarity problems (see [10-13]). The relations among submodularity, off-diagonal antitonicity and Z-matrices are expressed in the following result.

Proposition 2.1 [15]

Let f be a twice differentiable function on a rectangle X in $\mathbf{R}^n$. Then the following conditions are equivalent:

(i) f is submodular on X.

(ii) ∇f is off-diagonally antitone on X.

(iii) $H_f(x)$ is a Z-matrix, i.e. $f_{x_i x_j}(x) \leq 0$ for every $i \neq j$, for every $x \in X$.

Here $\nabla f = (f_{x_1}, \dots, f_{x_n})$ denotes the gradient of f and $H_f = (f_{x_i x_j})_{i,j=1,\dots,n}$ denotes the hessian matrix of f.

Proposition 2.2

Let X be a sublattice of $\mathbf{R}^n$ and define, for every $y \in X$, $X_{\leq}(y) = \{x \in X : x \leq y\}$ and $X_{\geq}(y) = \{x \in X : x \geq y\}$. Given a submodular function $f: X \to \mathbf{R}$, let X^* denote the set of global minimum points for f on X. Then

(i) If x' is a strict global minimum point for f on $X_{\leq}(x')$ ($X_{\geq}(x')$), then x' is a lower (upper) bound for X^*.

(ii) If x' is a global minimum point for f on $X_{\leq}(x')$ ($X_{\geq}(x')$) and $X^* \neq \emptyset$, then $X^* \cap X_{\geq}(x') \neq \emptyset$ ($X^* \cap X_{\leq}(x') \neq \emptyset$).

Proof

The proofs for the cases $X_{\leq}(x')$ and $X_{\geq}(x')$ are symmetric. Hence, we prove only the case $X_{\leq}(x')$.

(i) Let $x \in X$ and assume that $x_i < x'_i$ for at least an index i. Then, from assumption (i) and the submodularity of f we deduce

$$f(x) \geq f(x \wedge x') - f(x') + f(x \vee x') > f(x \vee x') .$$

Hence, x cannot be a global minimum point for f.
(ii) Take any $x^* \in X^*$. Then, trivially, $x^* \vee x' \in X_{\geq}(x')$. Furthermore,

$$f(x^*) \geq f(x^* \wedge x') - f(x') + f(x^* \vee x') \geq f(x^* \vee x') .$$

Hence, $x^* \vee x' \in X^*$. ❑

3. Variational inequalities over rectangles

The Variational Inequality Problem VI(X,F) associated with a set $X \subset \mathbf{R}^n$ and a function $F : X \to \mathbf{R}^n$ is formulated in the following way : find a vector $x^* \in X$ such that

(1) $$F(x^*)^T(y-x^*) \geq 0 , \quad \forall y \in X .$$

The set X^* of all $x^* \in X$ that satisfy (1) is called the solution set of VI(X,F).

Variational inequalities provide a general and unifying setting for many different problems, including nonlinear equations, complementarity problems, equilibrium problems and traffic assignment problems. A recent comprehensive survey on theory, algorithms and applications of variational inequality problems may be found in [7] .

In this section we establish an existence result and propose a solution method for VI(X,F) in the case where X is a rectangle and F is off-diagonally antitone. This problem contains as a special case the problem of finding the points satisfying the first order conditions for minimizing a differentiable submodular function f on X. In fact, it may be easily verified that a point $x \in X$ satisfies the first order conditions for minimizing f on X if and only if x solves $VI(X,\nabla f)$ (see [7]).

Note also that, even when X is restricted to be a rectangle, VI(X,F) generalizes the well-known Complementarity Problem CP(F) of finding a vector $x^* \in \mathbf{R}^n_+$ such that $F(x^*) \geq 0$ and $F(x^*)^T x^* = 0$. In fact, CP(F) coincides with $VI(\mathbf{R}^n_+ , F)$. Complementarity problems with off-diagonally antitone functions F have been studied in [10,11] .

In the remaining part of this section X will denote the rectangle defined by $X = \{x \in \mathbf{R}^n : \ell \le x \le u\}$, where $\ell_i \ge -\infty$ and $u_i \le +\infty$. Furthermore, we will use the following sets :

(2) $$S = \{x \in X : (u - x)^T F(x) \ge 0\} , \quad T = \{x \in X : (x - \ell)^T F(x) \le 0\} .$$

If $x \in S$ and $x_i = u_i$ or $x \in T$ and $x_i = \ell_i$, then clearly there is no restriction on the sign of $F_i(x)$. On the other hand, if $u_i = +\infty$, then $F_i(x) \ge 0$ for every $x \in S$ and if $\ell_i = -\infty$, then $F_i(x) \le 0$ for every $x \in T$. It may be easily verified that $S \cap T$ coincides with the solution set X^* of VI(X,F). Hence, trivially, $X^* \subset S$ and $X^* \subset T$. The set S has been introduced by Pang who established the following result.

Proposition 3.1 [7]

Let $F : X \to \mathbf{R}^n$ be continuous and off-diagonally antitone. Then

(a) S is a meet semi-lattice, i.e. $x \wedge y \in S$ whenever $x, y \in S$.

(b) If S is nonempty and bounded below, then S contains a least element x'; moreover x' solves VI(X,F) .

Note that F(x) is off-diagonally antitone on X if and only if $F'(x) = - F(-x)$ is off-diagonally antitone on $X' = \{-x : x \in X\}$. Hence, by applying proposition 3.1 to F' and X', we obtain that the following properties also hold:

(a') T is a join semi-lattice, i.e. $x \vee y \in T$ whenever $x, y \in T$.

(b') If T is nonempty and bounded above, then T contains a greatest element y'; moreover y' solves VI(X,F) .

Observe that if X is bounded below, i.e. $\ell \in \mathbf{R}^n$, then S is clearly bounded below and T is nonempty because $\ell \in T$. Similarly, if $u \in \mathbf{R}^n$, then T is bounded above and S is nonempty. Furthermore, from (b) and (b') we deduce that if X is nonempty and bounded, then there exist $x', y' \in X$ (not necessarily different) such that

(3) $$x' \le x \le y' , \quad \forall x \in X^* .$$

Hence, X^* is nonempty and has a least and a greatest element. We will now establish a more general condition under which S and T are nonempty and bounded below and above respectively. To this end we need to recall the notion of order-coercivity introduced in [13] and exploited by More' [10] to prove the existence of solutions for complementarity problems with off-diagonally antitone functions.

Definition 2.1

A function $F : X \to \mathbf{R}^n$ is upper (lower) order-coercive on X if for each unbounded increasing (decreasing) sequence $\{x_k\} \subset X$, it follows that

$$\lim_{k \to +\infty} F_i(x_k) = +\infty \quad (-\infty)$$

for some index i . If F is both lower and upper order-coercive, then F is called order-coercive.

Note that if F is a linear off-diagonally antitone function, i.e. $F(x) = Ax + b$, where $b \in \mathbf{R}^n$, $A \in \mathbf{R}^{n \times n}$ and $a_{ij} \leq 0$ for $i \neq j$, then F is lower order-coercive if and only if it is upper order-coercive and this happens if and only if A is an M-matrix (see [10]). We will use the following result .

Lemma 3.2 [10]

Let $F : X \to \mathbf{R}^n$ be off-diagonally antitone and lower (upper) order-coercive. Then, for every $x \in X$ and for every i such that $l_i = -\infty$ $(u_i = +\infty)$, we have

$$\lim_{\lambda \to -\infty} F_i(x + \lambda e_i) = -\infty \quad (\text{ resp. } \lim_{\lambda \to +\infty} F_i(x + \lambda e_i) = +\infty) .$$

In view of proposition 3.1, we can prove the existence of a solution for the problem VI(X,F) by showing that S is nonempty and bounded below or T is nonempty and bounded above. A case where these properties hold is illustrated in the next proposition.

Proposition 3.3

Let $F : X \to \mathbf{R}^n$ be a continous off-diagonally antitone function. If F is lower (upper) order-coercive on X, then S (T) is bounded below (above) and T (S) is nonempty.

Proof

We prove only the case where F is lower order-coercive; the other case is analogous. If S is not bounded below, then there exists a sequence $\{y^k\} \subset S$

such that $\lim_{k \to +\infty} y_i^k = -\infty$ for some index i. Since S is a meet semi-lattice, the decreasing sequence $\{x^k\}$ defined by $x^1 = y^1$ and $x^k = x^{k-1} \wedge y^k$ is also contained in S and satisfies $\lim_{k \to +\infty} x_i^k = -\infty$. Hence, by the order-coercivity of F, one has $\lim_{k \to +\infty} F_j(x^k) = -\infty$ for some index j. It follows that $F_j(x^k) < 0$ for every k greater then a given value k'. Since $\{x^k\} \subset S$, this is possible only if $x_j^k = u_j$ for every k. From the off-diagonal antitonicity of F we then deduce the contradiction $F_j(x^k) \leq F_j(x^{k+1})$ for every k. We have thus proved that S is bounded below. In order to prove that T is nonempty, consider the sets $I = \{i : \ell_i \geq -\infty\}$ and $J = \{i : \ell_i = -\infty\}$. By lemma 3.2, given any $y \in X$ and $v \in \mathbf{R}^n$ such that

$$y_I = \ell_I \quad \text{and} \quad F_J(y) \geq v_J ,$$

there exists a $z \in X$ such that

$$z_I = \ell_I , \quad z_J \leq y_J \quad \text{and} \quad F_i\,(y + (z_i - y_i)\,e_i) = v_i , \quad \forall\, i \in J .$$

Since F is off-diagonally antitone, we then have

$$F_I(z) \geq F_I(y) \quad \text{and} \quad F_J(z) \geq v_J .$$

Hence, taking any $x^0 \in X$ such that $x^0_I = \ell_I$ and choosing $v = 0 \wedge F(x^0)$, it is possible to find a sequence $\{x^k\}$ such that for every $k \geq 0$ one has

$$x^{k+1} \leq x^k , \qquad F_I(x^{k+1}) \geq F_I(x^k) , \qquad F_J(x^k) \geq v_J \qquad \text{and}$$

(4) $$F_i\,(x^k + (x_i^{k+1} - x_i^k)\,e_i) = v_i , \qquad \forall\, i \in J.$$

Since F is lower order-coercive on X, $\{x^k\}$ is bounded and hence converges to a point $x^* \in X$. Clearly one has $x^*_I = \ell_I$. Furthermore, from (4) we deduce that $F_J(x^*) = v_J$. Since $v_J \leq 0$ by definition, we then have $x^* \in T$. ❑

Corollary 3.3.1
Under the assumptions of proposition 3.3, if F is order-coercive then there exist $x', y' \in X$ such that (3) holds.

We will now describe some further properties of S and a method for finding its least element x', which is also the least solution of VI(X,F), whenever it exists. It should be clear that analogous results hold for T, although we do not state them explicitly.

Proposition 3.4
Let $F : X \to \mathbf{R}^n$ be off-diagonally antitone and let $x^1, x^2 \in X$ be such that $x^1 \leq x^2$ and $F(x) < 0$ for every $x \neq x^2$ satisfying $x^1 \leq x \leq x^2$. Then, if $y \in S$ and $y \geq x^1$, it follows $y \geq x^2$.
Proof
Assume on the contrary that there exists a point $y \in S$ such that $y \geq x^1$ and $y_i < x_i^2$ for some index i. Then clearly $y_i < u_i$ and hence $F_i(y) \geq 0$. On the other hand, from the off-diagonal antitonicity and the sign assumption on F we derive the contradiction $F_i(y) \leq F_i(x^1 + (y_i - x_i^1) e_i) < 0$. ❑

Corollary 3.4.1
Let $F : X \to \mathbf{R}^n$ be off-diagonally antitone. Then
(i) If $z \in X$ and $F_i(z + \lambda e_i) < 0$ for some index i and for every $\lambda > 0$, then there is no $x \in S$ satisfying $x \geq z$.
(ii) If y is a lower bound for S and there exist $\lambda_1, \ldots, \lambda_n \in \mathbf{R}^n$ such that $F_i(y + \lambda e_i) < 0$ for every $\lambda \geq \lambda_i$ and for every i, then $y' = y + \Sigma_{i=1,n} \lambda_i e_i$ is an upper bound for S.
Proof
(i) By proposition 3.4, if $x \in S$ and $x \geq z$, then $x \geq z + \lambda e_i$ for every $\lambda > 0$, which is clearly impossible.
(ii) If $x \in S$ and $x_i \geq y_i + \lambda_i e_i$ for some index i, then $x \geq y + \lambda_i e_i$ because y is a lower bound for S. Hence, if we take $z = y + \lambda_i e_i$, a contradiction follows from part (i) of this corollary. ❑

Corollary 3.4.2
Let $F : X \to \mathbf{R}^n$ be off-diagonally antitone. If there exists a $z \in X$ such that $z_i = \ell_i$ for every $i \in I = \{i : \ell_i > -\infty\}$ and $F_i(x) < 0$ for every $i \notin I$ and for every $x \in X$

such that $x \leq z$, then $z \leq x$ for every $x \in S$.

Proof

It is a straightforward consequence of proposition 3.4 taking $x^1 = x \wedge z$ and $x^2 = z$. ❑

When F is a continuous off-diagonally antitone function and S is nonempty and has a lower bound z, proposition 3.1 guarantees the existence of a least element x' of S which is also a solution of VI(X,F) . We will now describe a method for finding x'. Consider a sequence $\{x^k\} \subset X$ defined as follows :

$$(5) \qquad \begin{cases} x^0 = z \\ x^k = x^{k-1} + \lambda_k e_j , \quad \text{for} \quad k \geq 1 , \end{cases}$$

where $j \equiv k \pmod n$, $\lambda_k = 0$ if $F_j(x^{k-1}) \geq 0$ and λ_k is any scalar such that $x_i^{k+1} + \lambda_k \leq u_i$ and $F_j(x^{k-1} + \lambda e_j) < 0$ for every $0 \leq \lambda < \lambda_k$, if $F_j(x^{k-1}) < 0$. Clearly $\{x^k\}$ is monotone increasing . Furthermore, one has the following result.

Proposition 3.5

Let $F : X \to \mathbf{R}^n$ be continuous and off-diagonally antitone. Assume that S is nonempty and has a lower bound $z \in X$. If $\{x^k\}$ is defined by (5) then :

(i) x^k is a lower bound for S for every k and hence $\{x^k\}$ converges to a point $x^* \in X$ such that $x^* \leq x'$.

(ii) $x^* = x'$ if and only if for every index j either $x_j^* = u_j$ or $F_j(x^{k-1}) \geq 0$ for every $k \equiv j \pmod n$ or $F_j(x^k) \to 0$ when $k \equiv j \pmod n$ and $k \to +\infty$.

Proof

(i) is a straightforward consequence of proposition 3.4 . In order to prove (ii), observe that, by (i) , $x^* = x'$ if and only if $x^* \in S$. Hence, $x^* = x'$ if and only if for every j such that $x_j^* < u_j$ one has $F_j(x^*) \geq 0$. Furthermore, form the off-diagonal antitonicity of F and the definition of x^k we derive that $F_j(x^*) \geq 0$ if and only if $F_j(x^{k-1}) \geq 0$ for every $k \equiv j \pmod n$ or $F_j(x^k) \to 0$ when $k \equiv j \pmod n$ and $k \to +\infty$. ❑

Remark 3.1

By (ii) of the above proposition, the sequence defined by (5) converges to the least point x' of S if there exists an index $\bar{k}$ such that for every $k \geq \bar{k}$ and for every j such that $F_j(x^{k-1}) < 0$, λ_k in (5) is chosen as the unique value of λ for which $F_j(x^{k-1} + \mu e_j) < 0$ for every $0 \leq \mu < \lambda_k$ and $F_j(x^k) = F_j(x^{k-1} + \lambda_k e_j) = 0$. Note that, by corollary 3.4.1, these two conditions are certainly satisfied by some $\lambda_k \geq 0$ if S is nonempty.

4. Minimizing a convex submodular function on a rectangle

Consider the problem

$$(6) \qquad \min\ f(x) \qquad \text{s.t.} \qquad x \in X = \{x \in \mathbf{R}^n : \ell \leq x \leq u\} ,$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $\ell_i \geq -\infty$ and $u_i \leq +\infty$ for $i = 1, \ldots, n$. In the previous section we observed that if f is submodular and continuously differentiable, then the stationary points for (6) are the solutions of the problem VI(X,∇f) . Hence, if f is also convex, a point $x^* \in X$ is a solution of (6) if and only if x^* solves VI(X,∇f) . In this section we assume that f is convex, submodular and continuously differentiable and we will present a method for solving (6) under these assumptions.

First of all we need to introuce some notations. Given a rectangle $Y = \{ x \in \mathbf{R}^n : s \leq x \leq t \}$, we define $S(Y) = \{x \in Y : (t - x)^T \nabla f(x) \geq 0 \}$ and $T(Y) = \{x \in Y : (x - s)^T \nabla f(x) \leq 0 \}$. Furthermore, given $z \in \mathbf{R}^n$ and $I \subset N = \{1, \ldots, n\}$ we define $\mathbf{R}^n_I(z) = \{ x \in \mathbf{R}^n : x_I = z_I \}$.

Note that, by proposition 3.1, if $S(X)$ is bounded below, then it is nonempty if and only if (6) has a solution. More precisely, S(X) is nonempty if and only if its least element is the least global minimum point for f on X. Hence, if z is a lower bound for S(X), then z is a lower bound for the set of solutions of (6) . Clearly, similar properties hold for T(X) and its upper bounds.

Proposition 4.1

Let X be defined as in (6). If there exists an $\varepsilon > 0$ such that $z \in X$ ($w \in X$) is a strict global minimum point for f on $[z - \varepsilon e^I, z]$ ($[w, w + \varepsilon e^I]$), where $e^I = \Sigma_{i \in I} e_i$ and $I = \{i : z_i > \ell_i\}$ ($I = \{i : w_i < u_i\}$), then z is a lower bound for $S(X)$ (w is an upper bound for $T(X)$).

Proof.

Note that if z is a strict global minimum point for f on $[z - \varepsilon e^I, z]$ for some $\varepsilon > 0$, then the same property holds for every $\varepsilon > 0$ by the convexity of f. Take any point $x \in S(X)$ and let $K = \{i : x_i < z_i\}$. Trivially, we have $x_i < u_i$ for every $i \in K$. Hence, by the off-diagonal antitonicity of f, $f_{x_i}(x \wedge z) \geq f_{x_i}(x) \geq 0$ for every $i \in K$. Thus, the convexity of f implies that $x \wedge z$ is a global minimum point for f on $[x \wedge z, z]$. Since $x \wedge z \in [z - \varepsilon e^I, z]$ for some $\varepsilon > 0$, we conclude that $x \wedge z = z$. Hence, $z \leq x$. The proof of the fact that w is an upper bound for $T(X)$ is analogous. ❑

Corollary 4.1.1

If $z \in X$ and $f_{x_i}(z) < 0$ for every $i \in I = \{i : z_i > \ell_i\}$, then z is a lower bound for $S(X)$.

Given $\ell \in \mathbf{R}^n$, we now describe an algorithm for solving the problem :

(7) $$\min\ f(x) \qquad \text{s.t.} \qquad x \in X = \{x \in \mathbf{R}^n : x \geq \ell\}.$$

Algorithm **Convsubmod1**

Step 0 : Set $x^0 = \ell$ and $k = 0$.

Step 1 : Compute $I_k = \{i : f_{x_i}(x^k) \geq 0\}$ and $J_k = \{i : x_i^k = \ell_i\}$. If $J_k \subset I_k$, then stop ; x^k is the least global minimum point for f on X.

Step 2 : If f does not attain its minimum on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$, then stop; problem (7) has no solution . Otherwise, set x^{k+1} equal to the least global minimum point for f on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$, $k = k+1$ and go to step 1.

In order to prove the correctness of **Convsubmod1** and to analyze its complexity we need to establish some lemmas.

Lemma 4.2

Let X be defined as in (7). If $S(X) \neq \emptyset$, then $S(\mathbf{R}^n_I(\ell)) \neq \emptyset$ for every $I \subset N$.

Proof

Pick any $x \in S(X)$ and set $y_i = \ell_i$ if $i \in I$ and $y_i = x_i$ if $i \notin I$. Then $y \leq x$ and $y_I = \ell_I$. Hence, $y \in \mathbf{R}^n_I(\ell)$ and $f_{x_i}(y) \geq f_{x_i}(x) \geq 0$ for every $i \notin I$ by the off-diagonal antitonicity of f. Therefore, $y \in S(\mathbf{R}^n_I(\ell))$. ❑

Lemma 4.3

Let X be defined as in (6) and let $x^0, x^k \in X$, with $x^0 \leq x^k$ ($x^k \leq x^0$). If x^0 is a lower (upper) bound for S(X) (for T(X)) and x^k is a strict global minimum point for f on $[x^0, x^k]$ ($[x^k, x^0]$), then x^k is a lower bound for S(X) (x^k is an upper bound for T(X)) .

Proof

It is a straightforward consequence of propositions 2.2 and 3.1 and of the convexity of f . ❑

Lemma 4.4

Let X be defined as in (7) and, given $x^k \in X$, let I_k and J_k denote the sets introduced in step 1 of algorithm **Convsubmod1**. Assume that x^k is a strict global minimum point for f on $[\ell, x^k]$, $J_k \not\subset I_k$ and $S(\mathbf{R}^n_{I_k \cap J_k}(\ell)) \neq \emptyset$. Then f has a least global minimum point x^{k+1} on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$. Furthermore, $x^{k+1} \geq x^k$ and $J_{k+1} = J_k \cap I_k$.

Proof

Let $e^k = \Sigma_{i\in J_k \setminus I_k} e_i$. Since f is continuously differentiable it is possible to find an $\varepsilon > 0$ such that $f_{x_i}(x) < 0$ for every $i \in J_k \setminus I_k$ and $x \in [x^k, x^k+\varepsilon\, e^k]$. In view of proposition 2.2 and of the assumptions on x^k it follows that $x^k+\varepsilon\, e^k$ is a strict global minimum point for f on $[\ell, x^k+\varepsilon\, e^k]$. Hence, by proposition 4.1 , $x^k+\varepsilon\, e^k$ is a lower bound for $S(\mathbf{R}^n_{I_k \cap J_k}(\ell))$. Since $S(\mathbf{R}^n_{I_k \cap J_k}(\ell))$ is nonempty, it has a least element $x^{k+1} \geq x^k+\varepsilon\, e^k$ by proposition 3.1. Furthermore, the convexity of f implies that x^{k+1} is also the least global minimum point for f on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$. The relations $x^{k+1} \geq x^k$ and $J_{k+1} = J_k \cap I_k$ trivially follow from the inequality $x^{k+1} \geq x^k+\varepsilon\, e^k$. ❑

Proposition 3.5

Algorithm **Convsubmod1** finds the least solution of (7), or indicates that (7) has no solution, in at most n iterations.

Proof

Note that $x^0 = \ell$ is trivially a strict global minimum point for f on $[\ell, x^0]$. Furthermore, if at the k^{th} iteration of the algorithm one has $J_k \not\subset I_k$ and f attains its minimum on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$, then clearly $S(\mathbf{R}^n_{I_k \cap J_k}(\ell)) \neq \emptyset$ and lemma 4.4 guarantees that x^{k+1} exists and satisfies $x^{k+1} \geq x^k$ and $J_{k+1} = J_k \cap I_k$. Hence, at every iteration of the algorithm we have that x^k is a strict global minimum point for f on $[\ell, x^k]$, $x^{k+1} \in X$ and $|J_k| = |J_{k-1}| - 1$. Therefore, after at most n iterations either $J_k \subset I_k$ or f does not attain its minimum on $\mathbf{R}^n_{I_k \cap J_k}(\ell)$. By definition of x^k, we have $f_{x_i}(x^k) = 0$ for every $i \notin J_k$. Hence, $J_k \subset I_k$ implies $x^k \in S(X)$. Furthermore, from lemma 4.3 we deduce that x^k is a lower bound for $S(X)$. Hence, if $J_k \subset I_k$, x^k is the least element of $S(X)$. Thus, by proposition 3.1 and the convexity of f , x^k is the least global minimum point for f on X . In order to complete the proof observe that if f attains its minimum on X, then $S(X) \neq \emptyset$. Hence, $S(\mathbf{R}^n_{I_k \cap J_k}(\ell)) \neq \emptyset$ for every k by lemma 4.2 . Therefore, if f attains its minimum on X, the algorithm must terminate with $J_k \subset I_k$. ❑

The computational effort of algorithm **Convsubmod1** is mostly concentrated in step 2 : finding the least global minimum point for f on $R^n_{I_k \cap J_k}(\ell)$. This problem is in general more difficult then the problem of finding any global minimum point for f on $R^n_{I_k \cap J_k}(\ell)$, which is essentially an unconstrained convex minimization problem and can be solved efficiently in many cases. However, these two problems coincide when f has at most one global minimum point on $R^n_{I_k \cap J_k}(\ell)$. This unicity property clearly holds if f is strictly convex on R^n. We will now show that it holds also when f is a quadratic function. Let $f(x) = 1/2\, x^T A x + b^T x + c$, where A is a positive semidefinite $n \times n$ matrix and $b,c,x \in R^n$. Then, the global minimum points for f on $R^n_{I_k \cap J_k}(\ell)$ are the solutions of the linear system of equations:

$$(8) \qquad \begin{cases} x_L = \ell_L \\ A_{MM}\, x_L = \ell_L^T A_{ML} - b_M \end{cases}$$

where $L = I_k \cap J_k$, $M = N \backslash L$ and A_{IJ} denotes the submatrix of A formed by the rows indexed by I and the columns indexed by J . By proposition 4.5, if (8) has a solution, then it must have a least solution. However, since the solutions of (8) form a linear subspace of R^n , there exists a least solution of (8) if and only if (8) has a unique solution. Note also that (8) can be solved in $O(n^3)$ time in the general case and even more efficiently if A has some special structure. Hence, when f is a general convex and submodular quadratic function algorithm **Convsubmod1** runs in $O(n^4)$ time .

Let us now describe an algorithm for solving (6) in the case where $\ell, u \in R^n$

.Algorithm **Convsubmod2**

Step 0 : Set $x^0 = u$ and $k = 0$.

Step 1 : Compute $I_k = \{i : (x_i^k - \ell_i) f_{x_i}(x^k) \leq 0\}$ and $J_k = \{i : x_i^k = u_i\}$. If $J_k \subset I_k$, then stop; x^k is the greatest global minimum point for f on X .

Step 2 : Set x^{k+1} equal to the greatest global minimum point for f on $Y_k(\ell, u) = \{x \in \mathbf{R}^n_{I_k \cap J_k}(u) : x \geq \ell\}$, $k = k+1$ and go to step 1.

Note that (6) always has a solution if X is bounded. Furthermore, in this case, algorithm **Convsubmod2** computes the greatest solution of (6) . The proof of the correctness of algorithm **Convsubmod2** is analogous to that of algorithm **Convsubmod1** and is based on the following lemma, whose proof is omitted in view of its similarity with lemma 4.4.

Lemma 4.6

Let X be defined as in (6) with $\ell, u \in \mathbf{R}^n$ and, given $x^k \in X$, let I_k and J_k denote the sets introduced in step 1 of algorithm **Convsubmod2**. Assume that x^k is a strict global minimum point for f on $[x^k, u]$ and $J_k \not\subset I_k$. Then f has a greatest global minimum point x^{k+1} on $Y_k(\ell, u)$. Furthermore, $x^{k+1} \leq x^k$ and $J_{k+1} = J_k \cap I_k$.

Proposition 4.7

If $\ell, u \in \mathbf{R}^n$, then algorithm **Convsubmod2** computes the greatest solution of (6) in at most n iterations .

Proof

Note that $x^0 = u$ is trivially a strict global minimum point for f on $[x^0, u]$. Furthermore, if, at the k^{th} iteration, x^k is a strict global minimum point for f on $[x^k, u]$, $x^k \leq u$ and $J_k \not\subset I_k$, then lemma 4.6 guarantees that x^{k+1} exists and satisfies $x^{k+1} \leq u$ and $J_{k+1} = J_k \cap I_k$. Hence, at every iteration of the algorithm we have that $x^k \in X$, x^k is a strict global minimum point for f on $[x^k, u]$ and $|J_k| \leq |J_{k-1}| - 1$. Therefore, after at most n iterations we must have $J_k \subset I_k$. From the definition of x^k it follows that $f_{x_i}(x^k) = 0$ whenever $x_i^k > \ell_i$ and $i \notin J_k$. Hence $J_k \subset I_k$ implies $x_i^k \in T(X)$. Furthermore, by lemma 4.3, x^k is an upper bound for T(X). Thus, if $J_k \subset I_k$, x^k is the greatest element of T(X) and hence, by proposition 2.1 and the convexity of f , x^k is the greatest global minimum point for f on X . ❑

Like in the case of algorithm **Convsubmod1**, the computational burden of **Convsubmod2** is carried mostly by step 2 . Note that the problem of minimizing f on $Y_k(\ell, u)$ may be viewed as a problem of type (7). Hence, if f has a unique minimizer on $Y_k(\ell, u)$, then step 2 of algorithm **Convsubmod2** may be executed by means of algorithm **Convsubmod1**.

The algorithms presented in this section may be applied to solve the problem of minimizing f with finite lower and/or upper bounds on the variables. The next simple result allows us to deal with the case where some of the bounds are infinite.

Proposition 4.8

Let X be defined as in (6) . Assume that $z \in X$ is a lower bound for S(X) , $w \in X$ is an upper bound for T(X) and denote by X^*, $X^*_{\geq}(z)$, $X^*_{\leq}(w)$ and $X^*(z,w)$ the sets of global minimum points for f on X , $X_{\geq}(z) = \{x \in X : x \geq z\}$, $X^*_{\leq}(w) = \{x \in X : x \leq w\}$ and $X^*(z,w) = \{x \in X : z \leq x \leq w\}$ respectively. Then $X^* = X^*_{\geq}(z) = X^*_{\leq}(w)$ and, if $X^* \neq \varnothing$, $X^* = X^*(z,w)$.

Proof

By the convexity of f we have $X^* = S(X) \cap T(X)$. Hence, $X^* \subset X^*_{\geq}(z)$, $X^* \subset X^*_{\leq}(w)$ and $X^* \subset X^*(z,w)$. If $X^* \neq \varnothing$, then the above inclusions imply $X^* = X^*_{\geq}(z) = X^*_{\leq}(w) = X^*(z,w)$. On the other hand, if $X^*_{\geq}(z) \neq \varnothing$

$(X^*_{\leq}(w) \neq \varnothing)$, then $S(X_{\geq}(z)) \neq \varnothing$ $(T(X_{\leq}(w)) \neq \varnothing)$. Therefore, since $S(X_{\geq}(z)) = S(X)$ $(T(X_{\leq}(w)) = T(X))$, we have $X^* \neq \varnothing$ by proposition 3.1 . Hence, the equalities $X^* = X^*_{\geq}(z) = X^*_{\leq}(w)$ are always satisfied. ❑

Note that if z is a lower bound for S(X) and w is an upper bound for T(X) , then algorithm **Convsubmod2** may be applied to compute a global minimum point x* for f on X(z,w) . However, if X is not bounded, proposition 4.8 does not guaratee that x* is a global minimum point for f on X . This should be verified, e.g., by checking that x* satisfies the first order conditions for minimality. If x* is not a global minimum point for f on X, then proposition 4.8 allows us to conclude that f does not attain its minimum on X .

It may be easily verified that the convexity assumption on f has been exploited in this section only to guarantee that a stationary point for f on a set Z is a global minimum point for f on Z. Therefore we may slightly weaken this assumption by requiring that f is only semistrictly quasiconvex on X, i.e. that if $x,y \in X$, $\alpha \in (0,1)$ and $f(y) > f(x)$ then $f(\alpha x + (1- \alpha)y) > f(x)$. For a detailed description of the properties of semistrict quasiconvex functions and for others generalizations of the notion of convexity see [1].

5. Final remarks

We have seen in the previous section that the problem of minimizing a convex submodular function f on a rectangle in $\mathbf{R}^n$ can be solved by performing at most n^2 unconstrained minimizations of f. Hence, when the latter problem can be solved in polynomial time, like in the case where f is quadratic, the former problem can also be solved in polynomial time.

Another case where a submodular function f can be minimized in polynomial time over a rectangle X in $\mathbf{R}^n$ is when f is concave with respect to each variable (i.e. $f_{x_i x_i}(x) \leq 0$ for every $x \in X$, if f is twice differentiable) . In fact, in this case, at least one global point for f lies at a vertex of X and hence the problem of minimizing f on X can be transformed into the problem of minimizing a submodular function on the 0-1 hypercube B^n (see [14]).

The problem of minimizing a submodular function on a rectangle in $\mathbf{R}^n$ may be viewed as the continuous version of the problem of minimizing a submodular function on B^n. In analogy with the discrete case, it would be interesting to investigate the problem of minimizing a submodular function on a more general sublattice of $\mathbf{R}^n$, e.g. on a polyhedral sublattice.

[1] M. Avriel, W.E. Diewert, S. Schaible and I. Zang , "*Generalized Concavity*" (Plenum Press, New York, 1988) .

[2] W.H. Cunningham (ed.) , "*Submodular Optimization*" Mathematical Programming series B, 22 (1988) .

[3] J. Edmonds, "*Submodular functions, matroids and certain polyhedra*", in: R. Guy, H. Hanans, N. Sauer and J. Schönheim, eds., Combinatorial Structures and their Applications (Gordon and Breach, New York 1970).

[4] P. Favati and F. Tardella, "*Convexity in nonlinear integer programming*", Ricerca Operativa 53 (1990) 3-44 .

[5] M. Grötschel, L. Lovász and A. Schrijver, "*The ellipsoid method and its consequences in combinatorial optimization*", Combinatorica 1 (1981) 169-187 .

[6] M. Grötschel, L. Lovász and A. Schrijver, "*Geometric algorithms and combinatorial optimization*" (Springer-Verlag, Berlin, 1988) .

[7] P.T. Harker and J.S. Pang, "*Finite dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications*" Mathematical Programming 48 (1990) 161-220 .

[8] E.L. Lawler, "*Submodular functions and polymatroid optimization*" in: M. O'hEigeartaigh, J.K. Lenstra and A.H.G. Rinnooy Kan, eds., Combinatorial Optimization - Annotated Bibligraphies (John Wiley & Sons, New York 1985) .

[9] L. Lovász, "*Submodular functions and convexity*", in: A. Bachem, M. Grötschel and B. Korte, eds., Mathematical Programming: the state of the art - Bonn 1982 (Springer-Verlag, Berlin 1983) .

[10] J.J. More' , "*Coercivity conditions in nonlinear complementarity problems*", SIAM Review 16 (1974) 1-16 .

[11] J.J. More' , "*Classes of functions and feasibility conditions in nonlinear complementarity problems*", Mathematical Programming 6 (1974) 327-338 .

[12] J.J. More' and W.C. Rheinboldt, "*On P- and S-functions and related classes of n-dimensional nonlinear mappings*" Linear Algebra and its Applications 6 (1973) 45-68 .

[13] W.C. Rheinboldt, "*On M-functions and their application to nonlinear Gauss-Seidel iterations and to network flows*", Journal of Mathematical Analysis and Applications 32 (1970) 274-307 .

[14] F. Tardella, "*On the equivalence between some discrete and continuous optimization problems*", Annals of Operations Research 25 (1990) 291-300 .

[15] D.M. Topkis, "*Minimizing a submodular function on a lattice*" Operations Research 26 (1978) 305-321 .