Journal of Convex Analysis
Volume 5 (1998), No. 2, 269–278

On some Properties of Paramonotone Operators

Alfredo N. Iusem*
IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, CEP 22460-320, Brazil.
e-mail: iusp@impa.br

Received March 19, 1997
Revised manuscript received March 27, 1998

An operator $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is paramonotone iff it is monotone and $\langle T(x) - T(y), x - y\rangle = 0$ implies $T(x) = T(y)$. This definition can be extended to operators defined in a convex set whose values are subsets of $\mathbb{R}^n$. The notion of paramonotonicity is required to ensure convergence of several interior point methods for variational inequalities. In this paper we establish several properties of paramonotone operators. In particular, we give sufficient conditions for paramonotonicity in the differentiable case. We prove that if the symmetric part of the Jacobian matrix is positive semidefinite and its rank is greater than or equal to the rank of the Jacobian matrix at all points then the operator is paramonotone.

Keywords: variational inequalities, monotone operators, convex programming

1991 Mathematics Subject Classification: 90C25, 90C30

1. Introduction

The notion of paramonotocity of operators, which is in between monotonicity and strict monotonicity, was first presented in [2], where it was not given a name. The term *paramonotonicity* was introduced in [5]. We remind that an operator $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is monotone iff $\langle T(x) - T(y), x - y\rangle \geq 0$ for all x, y, and is strictly monotone if additionally $\langle T(x) - T(y), x - y\rangle = 0$ implies $x = y$. Paramonotonicity requires that $\langle T(x) - T(y), x - y\rangle = 0$ imply only $T(x) = T(y)$ (later on we will extend this definition to point-to-set operators). The main motivation behind the introduction of this class of operators lies in the analysis of interior point methods for variational inequalities, so we recall briefly some facts on this type of problems, which appear in many areas of application (see e.g. [7]).

Given a monotone operator $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and a closed convex set $C \subset \mathbb{R}^n$, the variational inequality problem, which we denote by VIP (T, C), consists of finding $z \in C$ such that $\langle T(z), x - z\rangle \geq 0$ for all $x \in C$.

When $T(x) = \nabla f(x)$ for some $f\colon \mathbb{R}^n \to \mathbb{R}$, VIP$(T, C)$ is equivalent to min $f(x)$ s.t. $x \in C$. When $C = \{x \in \mathbb{R}^n : x \geq 0\}$ then VIP(T, C) reduces to the nonlinear complementarity problem, which consists of finding $z \geq 0$ such that $T(z) \geq 0$ and $\langle z, T(z)\rangle = 0$.

The relevance of paramonotonicity stems from the following fact. Consider a property $P(x, y)$ defined on pairs of elements of some set S, and suppose that the problem to be

*Research for this paper was partially supported by CNPq grant no. 301280/86.

ISSN 0944-6532 / $ 2.50

solved consists of finding $z \in S$ such that $P(z,x)$ holds for all $x \in S$. A convenient feature of such a problem is the following: if a point y is such that $P(y,z)$ holds for some solution z of the problem, then y is also a solution of the problem. In other words, given a solution z, it can be verified whether another point y is also a solution by looking only at y and z, rather than at y and all points $x \in S$. This happens for instance in the case of minimization of a function f on S, in which case $P(x,y)$ is just $f(x) \leq f(y)$, because if z is a minimizer and $f(y) \leq f(z)$ then y is also a minimizer. This feature is useful in the analysis of algorithms which generate a sequence $\{x^k\}$ expected to converge to a solution of the problem; in many cases one is able to prove only that the limit (or a cluster point) $\overline{x}$ of $\{x^k\}$ satisfies $P(\overline{x},z)$ for a solution z. When this feature is present one can then conclude that $\overline{x}$ is also a solution.

Unfortunately, VIP(T,C) does not have this property for a general monotone operator T, as we show next. For VIP(T,C), $P(x,y)$ is $\langle T(x), y-x\rangle \geq 0$. Observe that when $C = \mathbb{R}^n$, z solves VIP(T,C) iff z is a zero of T, i.e. $T(z)=0$. Consider the case of $C=\mathbb{R}^2$ and $T(x_1,x_2)=(x_2,-x_1)$. It is easy to check that T is monotone (in fact $\langle T(x)-T(y), x-y\rangle = 0$ for all $x,y\in\mathbb{R}^n$) and that its only zero is $z=(0,0)$. However, for $y=(1,0)$ we have $\langle T(y), z-y\rangle = 0$ (i.e. $P(y,z)$ holds) but nevertheless y is not a solution. The idea is to impose further conditions on T, beyond monotonicity, so that VIP(T,C) has this feature, i.e. we want that whenever $\langle T(y), z-y\rangle \geq 0$ with $y\in C$ and z in the solution set S^* of VIP(T,C), it holds that y also belongs to S^*. Strict monotonicity does the job, because if $0 \leq \langle T(y), z-y\rangle$ and $z\in S^*$ then we have $0 \leq \langle T(y), z-y\rangle \leq \langle T(z), z-y\rangle \leq 0$ so that $\langle T(z)-T(y), z-y\rangle = 0$ and therefore $z=y$, i.e. y is a solution, but the computation just performed also shows that for a strictly monotone operator T, VIP(T,C) has at most one solution, which is a quite restrictive situation.

We want a condition on T which encompasses at least the convex optimization case (i.e. $T=\nabla f$ for a convex f) and such that VIP(T,C) has the feature discussed above. As we will prove in Proposition 2.3, paramonotonicity is appropriate in this sense. Thus, paramonotonicity is the condition imposed upon T in order to ensure convergence of several interior point methods for VIP(T,C) which use generalized distances, e.g. a Korpelevich-type method with Bregman distances [5], a perturbation method for saddle point problems [6], and also proximal point methods with either Bregman distances [3] or φ-divergences ([1] and [4]). We remark that though monotonicity is enough for the convergence analysis of other interior point methods for variational inequalities (e.g. [8], [9]), the algorithms mentioned above, in which the constraint set appears in the algorithm only through an interior barrier function whose gradient diverges at its boundary, are such that paramonotonicity seems an essential feature in their convergence analysis, in the sense that no convergence proof is known for monotone operators which fail to be paramonotone.

In this paper we bring together several results on paramonotone operators. In Section 2 we show that paramonotonicity enjoys the two properties just considered: subdifferentials of convex functions are paramonotone and VIP(T,C) with paramonotone T has the feature discussed above. This result was proved in [5] for point-to-point operators and we extend it here to the case of point-to-set ones.

In Section 3 we discuss monotonicity, paramonotonicity and strict monotonicity of affine operators (i.e. of the form $T(x)=Ax+b$) in terms of properties of the symmetric part $\tilde{A}$ of A ($\tilde{A}$ is defined as $\frac{1}{2}(A+A^t)$).

The results in Section 3 are rather elementary, but are needed in the study of continuously differentiable operators, presented in Section 4. If T is of class $\mathcal{C}^1$ in C, and we denote by $J_T(x)$ the Jacobian matrix of T at x, then we can consider the linear approximation of T at x, i.e. the operator $L_{T,x}$ defined as $L_{T,x}(y) = J_T(x)(y-x)$.

It is easy to check that monotonicity of T in C is equivalent to monotonicity of $L_{T,x}$ for all $x \in C$, but this is not true for either paramonotonicity or strict monotonicity: the "para" or "strict" properties may be due to higher order terms of T, beyond the linear approximation. On the other hand, if the linear approximations $L_{T,x}$ are all paramonotone or strictly monotone, then the same properties hold for the operator T.

Our main result is the following: If T is a differentiable operator such that the symmetric part of its Jacobian at any point x is positive semidefinite (which is equivalent to monotonicity) and its rank equals the rank of the Jacobian at x itself, then T is paramonotone. This sufficient condition is as easily checkable as monotonicity in terms of matrix analysis in the differentiable case; it says in fact that in order to check that a monotone and differentiable operator is paramonotone it suffices to verify that its Jacobian does not lose rank when it is symmetrized.

Throughout the paper $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$, ∇ is the gradient and ∇^2 the Hessian of a function, and superindex t indicates transpose.

2. Paramonotone operators

Let $C \subset \mathbb{R}^n$ be a convex set with nonempty interior and T an operator defined on $\mathbb{R}^n$ whose values are subsets of $\mathbb{R}^n$ (i.e. elements of $\mathcal{P}(\mathbb{R}^n)$).

Definition 2.1. T is said to be

(i) monotone in C iff $\langle u-v, x-y\rangle \geq 0$ for all $x, y \in C$ and all $u \in T(x)$, $v \in T(y)$,
(ii) paramonotone in C iff it is monotone in C and $\langle u-v, x-y\rangle = 0$ with $x, y \in C$, $u \in T(x)$, $v \in T(y)$ implies $u \in T(y)$ and $v \in T(x)$,
(iii) strictly monotone in C iff it is monotone in C and $\langle u-v, x-y = 0\rangle$ with $x, y \in C$, $u \in T(x)$, $v \in T(y)$ implies $x = y$.

It follows from Definition 2.1 that strict monotonicity in C implies paramonotonicity in C which in turn implies monotonicity in C.

The result of the next proposition was stated without proof in [2]. For the sake of completeness, we present a proof, taken from [5]. $\partial f(x)$ will denote the subdifferential of f at x.

Proposition 2.2. *If T is the subdifferential of a convex function f on C (i.e. $T(x)$ is the set of subgradients of f at x) then T is paramonotone in C.*

Proof. Monotonicity of T is well known. Assume that $\langle u-v, x-y\rangle = 0$ with $x \in C$, $y \in C$, $u \in T(x)$, $v \in T(y)$, and define $\overline{f}\colon C \to \mathbb{R}$ as $\overline{f}(z) = f(z) + \langle u, x-z\rangle$. Then $\overline{f}$ is convex and $\partial\overline{f}(z) = \partial f(z) - u = \{w - u : w \in \partial f(z)\}$. Taking $w = u$, we get that $0 \in \partial\overline{f}(x)$ and so x is an unrestricted minimizer of $\overline{f}$. By hypothesis and definition of subgradients, $f(x) - f(y) \leq \langle u, x-y\rangle = \langle v, x-y\rangle \leq f(x) - f(y)$, implying $f(x) = f(y) + \langle u, x-y\rangle = \overline{f}(y)$. Since $\overline{f}(x) = f(x)$ by definition of $\overline{f}$, we conclude that $\overline{f}(x) = \overline{f}(y)$, so that y is also an unrestricted minimizer of $\overline{f}$, i.e. $0 \in \partial\overline{f}(y)$. Therefore

$0 = w - u$ for some $w \in \partial f(y)$, which is equivalent to $u \in \partial f(y) = T(y)$. Reversing the roles of $(x, u), (y, v)$ the same argument proves that $v \in T(x)$ and the result is established in view of Definition 2.1(ii). □

Next we show that if T is paramonotone in C, then we can decide whether a point $y \in C$ solves VIP(T, C) just by looking at y and a given solution z of VIP(T, C).

We recall that for a point-to-set operator T, a vector $z \in C$ is a solution of VIP(T, C) iff there exists $u \in T(z)$ such that $\langle u, x - z\rangle \geq 0$ for all $x \in C$.

Proposition 2.3. *Assume that T is paramonotone in C and let z be a solution of* VIP(T, C). *Then $y \in C$ is a solution of* VIP(T, C) *if and only if there exists $v \in T(y)$ such that $\langle v, z - y\rangle \geq 0$.*

Proof. The "only if" part is immediate. We prove the "if" part. Assume that $\langle v, z-y\rangle \geq 0$ for some $v \in T(y)$ and some solution z of VIP(T, C). Since z is a solution of VIP(T, C) there exists $u \in T(z)$ such that

$$\langle u, x - z\rangle \geq 0 \tag{2.1}$$

for all $x \in C$. Then by monotonicity of $T, 0 \leq \langle v, z - y\rangle \leq \langle u, z - y\rangle \leq 0$, implying that

$$\langle v, z - y\rangle = \langle u, z - y\rangle = 0. \tag{2.2}$$

From (2.2), $\langle v - u, z - y\rangle = 0$ and by paramonotonicity of T we get

$$u \in T(y). \tag{2.3}$$

Then, for all $x \in C$

$$\langle u, x - y\rangle = \langle u, x - z\rangle + \langle u, z - y\rangle = \langle u, x - z\rangle \geq 0 \tag{2.4}$$

using (2.2) in the second equation and (2.1) in the inequality. It follows from (2.3) and (2.4) that y is a solution of VIP(T, C). □

3. The case of affine operators

In this section we consider operators of the form $T(x) = Ax + b$ with $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$. For a matrix $A \in \mathbb{R}^{n\times n}$ we will denote by $\tilde{A}$ its symmetric part, i.e. $\tilde{A} = \frac{1}{2}(A + A^t)$. It is immediate that $x^tAx = x^t\tilde{A}x$ for all $x \in \mathbb{R}^n$ and all $A \in \mathbb{R}^{n\times n}$. We will use repeatedly the following well known fact: if A is symmetric positive semidefinite and $x^tAx = 0$ then $Ax = 0$. This follows from the fact that A can be written as $A = B^tB$ for some $B \in \mathbb{R}^{n\times n}$ (e.g. the Cholesky factorization) and then $0 = x^tAx$ implies $0 = x^tB^tBx = \|Bx\|^2$, so that $0 = Bx$ and therefore $0 = B^tBx = Ax$. We will also need the following elementary result.

Proposition 3.1. *If $\tilde{A}$ is positive semidefinite then the following statements are equivalent.*

(i) $\ker(\tilde{A}) \subset \ker(A)$.

(ii) *There exists $V \in \mathbb{R}^{n\times n}$ such that $A = V\tilde{A}$.*

(iii) *If* $U \in \mathbb{R}^{n\times n}$ *is such that* $UU^t = I$ *and* $U^t\tilde{A}U$ *is in Jordan normal form, i.e.* $U^t\tilde{A}U = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ *with* $I \in \mathbb{R}^{p\times p}$, *then* $U^tAU = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}$ *with* $M \in \mathbb{R}^{p\times p}$.

(iv) $\operatorname{rank}(A) \le \operatorname{rank}(\tilde{A})$.

Proof. The fact that (i) implies (ii) is immediate. For (ii) $\Rightarrow$ (iii), note that $A = V\tilde{A}$ implies $U^tAU = U^tV\tilde{A}U = (U^tVU)(U^t\tilde{A}U) = (U^tVU)\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}$ where M is the upper left $p \times p$ submatrix of U^tVU. (iii) implies (iv) because $\operatorname{rank}(\tilde{A}) = p$ and $\operatorname{rank}(A) = \operatorname{rank}(M) \le p$. We prove that (iv) implies (i). We claim that for $\tilde{A}$ positive semidefinite it always holds that $\ker(A) \subset \ker(\tilde{A})$: if $0 = Ax$ then $0 = x^tAx = x^t\tilde{A}x$ implying $0 = \tilde{A}x$ i.e. $x \in \ker(\tilde{A})$. As a consequence, $\operatorname{rank}(A) \ge \operatorname{rank}(\tilde{A})$. If (iv) holds, then $rank(A) = \operatorname{rank}(\tilde{A})$. Since $\ker(A) \subset \ker(\tilde{A})$, it follows that $\ker(A) = \ker(\tilde{A})$, which clearly implies (i). □

Before establishing the next result, we remark that for point-to-point operators Definition 2.1 reduces to the notions stated in the first paragraph of Section 1.

Proposition 3.2. *Let* $C \subset \mathbb{R}^n$ *be a convex set with nonempty interior and take* $T(x) = Ax + b$. *Then*

(i) *T is monotone in C iff $\tilde{A}$ is positive semidefinite,*

(ii) *T is paramonotone in C iff $\tilde{A}$ is positive semidefinite and* $\ker(\tilde{A}) \subset \ker(A)$,

(iii) *T is strictly monotone in C iff $\tilde{A}$ is positive semidefinite and* $\ker(\tilde{A}) = \{0\}$ *(i.e. $\tilde{A}$ is positive definite).*

Proof. Note first that for $T(x) = Ax+b$ it holds that $\langle T(x)-T(y), x-y\rangle = (x-y)^t\tilde{A}(x-y)$. We prove now the three items.

(i) The "if" part follows from the observation just made. For the "only if" part, given $z \in \mathbb{R}^n$ take any $x \in \operatorname{int}(C)$ and let $y = x - \lambda z$ with $\lambda > 0$ and small enough so that $y \in C$. Then $z^t\tilde{A}z = \lambda^{-2}(x-y)^t\tilde{A}(x-y) = \lambda^{-2}\langle T(x) - T(y), x - y\rangle \ge 0$ by the monotonicity of T in C. Since z is arbitrary, $\tilde{A}$ is positive semidefinite.

(ii) $\Rightarrow$) By (i) $\tilde{A}$ is positive semidefinite. Take $z \in \ker(\tilde{A})$, $x \in \operatorname{int}(C)$ and y, λ as in (i). Since $z \in \ker(\tilde{A})$, $\tilde{A}z = 0$ implying $0 = z^t\tilde{A}z = \lambda^{-2}(x-y)^t\tilde{A}(x-y) = \lambda^{-2}\langle T(x) - T(y), x - y\rangle$. Since $x, y \in C$, the paramonotonicity of T in C implies that $0 = T(x) - T(y) = \lambda Az$, i.e. $z \in \ker(A)$.
$\Leftarrow$) Assume that $0 = \langle T(x)-T(y), x-y\rangle$ with $x, y \in C$. Then $0 = (x-y)^t\tilde{A}(x-y)$ and by the positive semidefiniteness of $\tilde{A}$ we get $\tilde{A}(x-y) = 0$, i.e. $x-y \in \ker(\tilde{A}) \subset \ker(A)$, so that $A(x-y) = 0$ implying $T(x) = Ax + b = Ay + b = T(y)$. So T is paramonotone in C.

(iii) $\Rightarrow$) $\tilde{A}$ is positive semidefinite by (i). Take $z \in \ker(\tilde{A})$ and x, y as in (ii). Then, with the same argument, we get $0 = \langle T(x) - T(y), x - y\rangle$. By strict monotonicity of T in C, $x = y$ so that $z = \lambda(x-y) = 0$, proving that $\ker(\tilde{A}) = 0$.
$\Leftarrow$) If $\langle T(x) - T(y), x - y\rangle = 0$ for $x, y \in C$ then $0 = (x-y)^t\tilde{A}(x-y)$. Since $\tilde{A}$ is positive definite we get $x = y$, establishing strict monotonicity of T in C.

□

Proposition 3.2 enables us to construct easily paramonotone operators that are not strictly monotone and monotone operators which are not paramonotone; for instance $T(x) = Ax$ with $A = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is paramonotone but not strictly monotone and $T(x) = Ax$ with $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is monotone but not paramonotone.

4. The case of continuously differentiable operators

In this section we assume that T is continuously differentiable in C, i.e. the Jacobian matrix of T, which we denote by $J_T(x)$, is well defined for all $x \in C$ and $J_T(\cdot)$ is continuous in C. For each $x \in C$, consider the operator $L_{T,x}\colon \mathbb{R}^n \to \mathbb{R}^n$ defined as $L_{T,x}(y) = J_T(x)(y-x)$, which is the linear approximation of T at x. The next proposition shows that monotonicity of T in C is equivalent to monotonicity of $L_{T,x}$ for all $x \in C$.

Proposition 4.1. *Let $C \subset \mathbb{R}^n$ be convex with nonempty interior and take $T \in \mathcal{C}^1(C)$. Then the following statements are equivalent:*

(i) *T is monotone in C,*

(ii) *$L_{T,x}$ is monotone for all $x \in C$,*

(iii) *$\tilde{J}_T(x)$ is positive semidefinite for all $x \in C$.*

Proof. (ii) and (iii) are equivalent by Proposition 3.2(i). We prove next that (i) implies (iii). Take any $z \in \mathbb{R}^n$, fix $x \in \text{int}(C)$ and take y and $\lambda \in (0,1)$ as in Proposition 3.2(i). Let $w_\alpha = x + \alpha(y-x)$ with $\alpha \in (0,\lambda) \subset (0,1)$, so that $w_\alpha \in C$, by convexity of C. By monotonicity of T

$$0 \le \langle T(w_\alpha) - T(x), w_\alpha - x\rangle = \alpha\langle T(w_\alpha) - T(x), y - x\rangle. \tag{4.1}$$

Dividing (4.1) by α^2, we get $0 \le \langle \frac{1}{\alpha}[T(w_\alpha) - T(x)], y-x\rangle$ and taking limits as α goes to 0, we obtain $0 \le (x-y)^t J_T(x)(x-y) = (x-y)^t \tilde{J}_T(x)(x-y) = \lambda^2 z^t \tilde{J}_T(x) z$, implying $0 \le z^t \tilde{J}_T(x) z$. Since z is arbitrary, $\tilde{J}_T(x)$ is positive semidefinite for all $x \in \text{int}(C)$ and therefore for all $x \in C$, by the continuity of $J_T(\cdot)$.

Finally we prove that (iii) implies (i). Take $x, y \in C$ and let $\varphi\colon [0,1] \to \mathbb{R}$ be defined as $\varphi(\alpha) = \langle T(x + \alpha(y-x)), x-y\rangle$. Then $\langle T(x) - T(y), x-y\rangle = \varphi(0) - \varphi(1) = \varphi'(\alpha)$ for some $\alpha \in [0,1]$. Let $w = x + \alpha(y-x)$. Then $\varphi'(\alpha) = (x-y)^t J_T(w)(x-y)$ with $w \in C$. So $\langle T(x) - T(y), x-y\rangle = (x-y)^t \tilde{J}_T(w)(x-y) \ge 0$ by the positive semidefiniteness of $\tilde{J}_T(w)$. □

One could expect to obtain similar results for paramonotonicity and strict monotonicity, stating that such properties hold for T if and only if they hold for $L_{T,x}$ for all $x \in C$, but, as we already mentioned in the introduction, this is not true. These properties are not inherited by their linear approximations, as the following examples show.

Take $C = \mathbb{R}$, $T(x) = x^3$. T is strictly monotone because it is the derivative of the strictly convex function $f(x) = \frac{1}{4}x^4$, but $J_T(0) = 0$, so that $L_{T,0}$ is not strictly monotone, by Proposition 3.2(iii). Now consider $C = \mathbb{R}^2$ and $T = T_1 + T_2$, with $T_1(x_1, x_2) = (-x_2, x_1)$ and $T_2(x_1, x_2) = (x_1^3, x_2^3)$. T_1 is monotone, as mentioned in Section 1, and T_2 is strictly monotone, being the gradient of the strictly convex function $f(x_1,x_2)=\frac{1}{4}(x_1^4+x_2^4)$. So T is strictly monotone and therefore paramonotone. However, $J_T(0)=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$ and $\tilde{J}_T(0)=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$, so that $\ker(\tilde{J}_T(0))=\mathbb{R}^2$, which is not contained in $\ker(J_T(0))=\{0\}$, and therefore $L_{T,0}$ is not paramonotone, by Proposition 3.2(ii).

On the other hand, paramonotonicity or strict monotonicity of the linear approximations are enough to ensure that such properties hold for the original operator. We start with paramonotonicity. In the next three propositions it is assumed that $T\in\mathcal{C}^1(C)$ and that C is convex with nonempty interior.

Proposition 4.2. *If $\tilde{J}_T(x)$ is positive semidefinite for all $x\in C$ and any of the following statements hold*

(i) $\ker(\tilde{J}_T(x))\subset\ker(J_T(x))$ *for all* $x\in C$,
(ii) $\operatorname{rank}(J_T(x))\le\operatorname{rank}(\tilde{J}_T(x))$ *for all* $x\in C$,
(iii) *there exists* $V(x)\in\mathbb{R}^{n\times n}$ *such that* $J_T(x)=V(x)\tilde{J}_T(x)$ *for all* $x\in C$,
(iv) $L_{T,x}$ *is paramonotone for all* $x\in C$,

then T is paramonotone in C.

Proof. First observe that (i), (ii) and (iii) are equivalent by Proposition 3.1 and that (i) and (iv) are equivalent by Proposition 3.2(ii), so that it suffices to establish the result under e.g. (i). We proceed to do so. By Proposition 4.1 T is monotone in C. Assume that $0=\langle T(x)-T(y),x-y\rangle$ with $x,y\in C$. For $\alpha\in(0,1)$, let $w_\alpha=y+\alpha(x-y)$. Then $w_\alpha\in C$ for all $\alpha\in(0,1)$ by convexity of C and

$$\begin{aligned}0&=\langle T(x)-T(w_\alpha),x-y\rangle+\langle T(w_\alpha)-T(y),x-y\rangle\\ &=\frac{1}{1-\alpha}\langle T(x)-T(w_\alpha),x-w_\alpha\rangle+\frac{1}{\alpha}\langle T(w_\alpha)-T(y),w_\alpha-y\rangle.\end{aligned}\tag{4.2}$$

Both terms in the rightmost expression of (4.2) are nonnegative by the monotonicity of T in C, and therefore both vanish. In particular

$$0=\frac{1}{\alpha}\langle T(w_\alpha)-T(y),w_\alpha-y\rangle=\langle T(w_\alpha)-T(y),x-y\rangle\tag{4.3}$$

for all $\alpha\in(0,1)$. Take now $\beta\in(0,1)$ and $\gamma\in(0,1-\beta)$, so that $\beta+\gamma<1$. By (4.3), $0=\langle T(w_\beta)-T(y),x-y\rangle=\langle T(w_{\beta+\gamma})-T(y),x-y\rangle$ implying $0=\langle T(w_{\beta+\gamma})-T(w_\beta),x-y\rangle$, so that, for all $\gamma\in(0,1-\beta)$

$$0=\langle\frac{1}{\gamma}[T(w_{\beta+\gamma})-T(w_\beta)],x-y\rangle.\tag{4.4}$$

Taking limits in (4.4) as γ goes to 0, we get $0=(x-y)^tJ_T(w_\beta)(x-y)=(x-y)^t\tilde{J}_T(w_\beta)(x-y)$ for all $\beta\in(0,1)$, and therefore for all $\beta\in[0,1]$, by the continuity of $J_T(\cdot)$. Since $w_\beta\in C$, $\tilde{J}_T(w_\beta)$ is positive semidefinite, and we can conclude that $x-y\in\ker(\tilde{J}_T(w_\beta))\subset\ker(J_T(w_\beta))$, i.e. that

$$J_T(w_\beta)(x-y)=0\tag{4.5}$$

for all $\beta \in [0,1]$. Let now T_j $(1 \le j \le n)$ denote the components of T (i.e. $T_j(x) = T(x)_j$). Then

$$T_j(y) = T_j(x) + \nabla T_j(w_{\beta_j})^t (y-x) \tag{4.6}$$

for some $\beta_j \in [0,1]$. Note that $\nabla T_j(w_{\beta_j})$ is the j-th row of $J_T(w_{\beta_j})$ so that $\nabla T_j(w_{\beta_j})^t(x-y) = 0$ by (4.5) and then $T_j(y) = T_j(x)$ by (4.6). It follows that $T(x) = T(y)$. $\square$

Before establishing a similar proposition for strict monotonicity, we present an intermediate result, which is of some interest on its own.

Proposition 4.3. *If T is paramonotone in C and $J_T(x)$ is nonsingular for all $x \in C$ then T is strictly monotone in C.*

Proof. Assume that $\langle T(x)-T(y), x-y\rangle = 0$ with $x, y \in C$. As in the proof of Proposition 4.2, we get that $0 = \langle T(w_\alpha) - T(y), w_\alpha - y\rangle$ for all $\alpha \in (0,1)$. Since $w_\alpha \in C$, it follows, by the paramonotonicity of T in C, that $T(w_\alpha) = T(y)$, i.e.

$$0 = \frac{1}{\alpha}[T(w_\alpha) - T(y)] \tag{4.7}$$

for all $\alpha \in (0,1)$. Taking limits in (4.7) as α goes to 0, we get $0 = J_T(y)(x-y)$, so that $x = y$ by nonsingularity of $J_T(y)$. $\square$

Note that Proposition 4.3 does not hold if we assume that T is just monotone; the operator $T(x_1, x_2) = (x_2, -x_1)$ has nonsingular Jacobian matrix equal to $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ for all x, but it is not strictly monotone. In order to ensure strict monotonicity we need in general nonsingularity of $\tilde{J}_T(x)$, rather than of $J_T(x)$, as the next proposition shows.

Proposition 4.4. *If $\tilde{J}_T(x)$ is positive definite for all $x \in C$ then T is strictly monotone in C.*

Proof. Since $\ker(\tilde{J}_T(x)) = \{0\}$, we have $\ker(\tilde{J}_T(x)) \subset \ker(J_T(x))$ and we are within the hypotheses of Proposition 4.2, so that T is paramonotone in C. Since $\ker(J_T(x)) \subset \ker(\tilde{J}_T(x))$, as shown in the proof of Proposition 3.1, we have that $\ker(J_T(x)) = \{0\}$, i.e. $J_T(x)$ is nonsingular. By Proposition 4.3, T is strictly monotone. $\square$

We present now the specific form of Proposition 4.2 for an important family of monotone operators, namely those which arise from constrained saddle point problems. Given $K\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, convex in its first argument and concave in the second one, and closed convex sets $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, the constrained saddle point problem $CSP(K,X,Y)$ consists of finding $(\overline{x}, \overline{y}) \in X \times Y$ such that $K(\overline{x}, y) \le K(\overline{x}, \overline{y}) \le K(x, \overline{y})$ for all $(x,y) \in X \times Y$. If $K \in \mathcal{C}^2(X \times Y)$ then $CSP(K,X,Y)$ is equivalent to VIP(T,C) with $T(x,y) = (\nabla_x K(x,y), -\nabla_y K(x,y))$ and $C = X \times Y$ ($\nabla_x K$ and $\nabla_y K$ denote the gradient of K with respect to its first and second argument respectively). It is easy to check that this T is monotone, and we will establish sufficient conditions for paramonotonicity for the case in which $T \in \mathcal{C}^1(C)$, i.e. $K \in \mathcal{C}^2(X \times Y)$.

Let $\nabla^2_{xx} K \in \mathbb{R}^{n\times n}$, $\nabla^2_{yy} K \in \mathbb{R}^{m \times m}$ and $\nabla^2_{xy} K \in \mathbb{R}^{n \times m}$ be the matrices of second derivatives of K (e.g $(\nabla^2_{xy} K(x,y))_{ij} = \frac{\partial^2}{\partial x_j \partial y_i} K(x,y)$). Then Proposition 4.2(iii) takes the following form.

Proposition 4.5. *If for all $(x, y) \in X \times Y$ there exist matrices $G(x, y), H(x, y) \in \mathbb{R}^{n\times m}$ such that*

$$\nabla^2_{xy} K(x, y) = \nabla^2_{xx} K(x, y) G(x, y) = H(x, y) \nabla^2_{yy} K(x, y) \tag{4.8}$$

then $T = (\nabla_x K, -\nabla_y K)$ is paramonotone in $X \times Y$.

Proof. For $T = (\nabla_x K, -\nabla_y K)$ we have $J_T(x, y) = \begin{bmatrix} D & B \\ -B^t & E \end{bmatrix}$, $\tilde{J}_T(x, y) = \begin{bmatrix} D & 0 \\ 0 & E \end{bmatrix}$ with $D = \nabla^2_{xx} K(x, y)$, $E = -\nabla^2_{yy} K(x, y)$, $B = \nabla^2_{xy} K(x, y)$. By (4.8), $B = DG = HE$. Take $V(x, y) = \begin{bmatrix} I & H \\ -G^t & I \end{bmatrix}$, so that $J_T(x, y) = V(x, y)\tilde{J}_T(x, y)$ and apply Proposition 4.2 under hypothesis (iii). □

Note that nonsingularity of $\nabla^2_{xx} K(x, y)$ (respectively $\nabla^2_{yy} K(x, y)$) implies existence of $G(x, y)$ (respectively $H(x, y)$). The result of Proposition 4.5 can be slightly improved: if $K(\cdot, y)$ is strictly convex for all $y \in Y$ (respectively $-K(x, \cdot)$ is strictly convex for all $x \in X$) then the existence of $H(x, y)$ (respectively $G(x, y)$) as in Proposition 4.5 is enough to ensure paramonotonicity of T. This can be proved by reworking the proof of Proposition 4.2 for the case of $T = (\nabla_x K, -\nabla_y K)$.

Finally, we mention that an important class of monotone operators fail to be paramonotone. Consider the convex optimization problem $\min f(x)$ s.t. $g_i(x) \le 0$ $(1 \le i \le m)$ with convex $f, g_i \colon \mathbb{R}^n \to \mathbb{R}$. If f, g_i are of class $\mathcal{C}^1$ then under standard regularity conditions (e.g. [10]), this problem is equivalent to $CSP(K, \mathbb{R}^n, \mathbb{R}^m_+)$ where $\mathbb{R}^m_+ = \{y \in \mathbb{R}^m : y_i \ge 0\ (1 \le i \le m)\}$ and K is the Lagrangian function, i.e. $K(x, y) = f(x) + \sum_{i=1}^m y_i g_i(x)$, so that $T(x, y) = (\nabla f(x) + J_g(x)^t y, -g(x))$, where $g(x) = (g_1(x), \dots, g_m(x))$ and $J_g(x)$ is the Jacobian matrix of g at x. Take $x \in \mathbb{R}^n$ such that $J_g(x) \ne 0$ and $v \in \mathbb{R}^m$ such that $J_g(x)^t v \ne 0$ and write $v = y - y'$ with $y, y' \in \mathbb{R}^m_+$. If $z = (x, y)$, $z' = (x, y')$ then $\langle T(z) - T(z'), z - z' \rangle = \langle J_g(x)^t (y - y'), x - x \rangle - \langle g(x) - g(x), y - y' \rangle = 0$ while $T(z) - T(z') = (J_g(x)^t (y - y'), 0) = (J_g(x)^t v, 0) \ne (0, 0)$, i.e. T is not paramonotone in $\mathbb{R}^n \times \mathbb{R}^m_+$ (unless $J_g(x) = 0$ for all x, in which case the convex optimization problem is indeed unconstrained). In term of the conditions of Proposition 4.5, assuming that f, g are of class $\mathcal{C}^2$, we have $\nabla^2_{xx} K(x, y) = \nabla^2 f(x) + \sum_{i=1}^m y_i \nabla^2 g_i(x)$, $\nabla^2_{xy} K(x, y) = J_g(x)^t$ and $\nabla^2_{yy} K(x, y) = 0$, so that (4.8) becomes $J_g(x) = 0$, i.e. the hypothesis of Proposition 4.5 does not hold in the constrained case.

5. Final remarks

Our most relevant result, namely Proposition 4.2, means that in the differentiable case we can detect paramonotonicity by checking properties of some matrices, namely $J_T(x)$, as in the case of convexity for twice differentiable functions.

In connection with convexity, we point out first that if $J_T(x)$ is symmetric for all $x \in C$ then the conditions in Proposition 4.2 are trivially satisfied, because $J_T(x) = \tilde{J}_T(x)$. However, symmetry of $J_T(x)$ in the open set $\mathrm{int}(C)$ implies that T is indeed the gradient of a function f defined in C, so that $J_T(x) = \nabla^2 f(x)$ and Proposition 4.1 reduces to the well known result that positive semidefiniteness of the Hessian matrix ensures convexity of the function.

The really interesting monotone operators are those which are not gradients of convex functions, i.e. those whose Jacobian matrices fail to be symmetric at some points. In this case the conditions in Proposition 4.2 turn out to be the key properties for paramonotonicity, which is itself essential in the analysis of several interior point methods for variational inequalities, as mentioned in Section 1.

References

[1] A. A. Auslender, M. Haddou: An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities, Mathematical Programming 71 (1995) 77–100.

[2] R. E. Bruck: An iterative solution of a variational inequality for certain monotone operators in a Hilbert space, Bulletin of the American Mathematical Society 81 (1975) 890–892.

[3] R. S. Burachik, A. N. Iusem: A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, SIAM Journal on Optimization 8 (1998) 197–216.

[4] R. S. Burachik, A. N. Iusem: A generalized proximal point algorithm for the nonlinear complementarity problem (to be published in RAIRO, Recherche Opérationelle).

[5] Y. Censor, Y., A. N. Iusem, S. Zenios: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators, Mathematical Programming 81 (1998) 373–400.

[6] A. N. Iusem, M. A. Kallio: An interior point method for constrained saddle point problems (to appear).

[7] D. Kinderlehrer, G. Stampacchia: An Introduction to Variational Inequalities and their Applications, Academic Press, New York, 1980.

[8] Y. Nestorov, A. S. Nemirovsky: Interior Point Polynomial Algorithms in Convex Programming, SIAM Publications, Philadelphia, PA, 1994.

[9] F. A. Potra, Y. Ye: Interior point methods for nonlinear complementarity problems, Technical Report 15, Department of Mathematics, University of Iowa, 1991.

[10] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, N.J., 1970.