A fresh geometrical look at the general S-procedure

Michel De Lara [1] and Jean-Baptiste Hiriart-Urruty [2]

Abstract. We revisit the S-procedure for general functions with "geometrical glasses". We thus delineate a necessary condition, and almost a sufficient condition, to have the S-procedure valid. Everything is expressed in terms of convexity of augmented sets (*i.e.*, via convex hulls, conical hulls) of images built from the data functions.

Keywords. S-lemma, Convexity of image sets, Separation of convex sets, Theorem of alternatives.

Introduction

The so-called *S-procedure* takes roots in Automatic Control Theory; an excellent survey-paper on its origin and developments in that area is [4]. In the field of Optimization, the subject has also been studied thoroughly, beginning with the quadratic data and further with general functions. As a result, papers concerning the S-procedure abound. Fortunately, there are from time to time survey-papers which allow to take stock of what has been done and what needs to be done; two such examples are [1] and [6]. With that in mind, for the convenience of the reader who is not necessarily "immersed" in the subject, we recall in Section 1 some of the main known results.

The keypoint of the message conveyed in our note is the following: the essential is not the convexity of the image set of the vector-valued mapping obtained from all the involved real-valued functions ; it is rather the convexity of an enlarged version of this image (via operations like adding the positive orthant $(\mathbb{R}_+)^q$, or taking the conical hull). This assumption clearly is weaker than the mere convexity of the image itself.

The S-procedure is intimately linked with the validity of a duality result in a certain mathematical optimization problem (see a recent overview of that in [9]). This was already the main motivation in Fradkov's paper ([3]). But this aspect is not broached here.

Our approach is essentially *geometrical*; the validity of the necessary/sufficient conditions that we develop are expressed in terms of convexity of sets. As expected in such contexts, the main used mathematical tool is the separation of convex sets by hyperplanes (in finite-dimensional vector spaces). Our main results (Theorem 2, Theorem 3) have similarities with some in Fradkov's old paper [3]; they could have been there, as much as the method as indications around some remarks led to them. To a certain extent, our note is a revisit and an extension of Section 1 in [3].

[1] CERMICS, Ecole des Ponts, Marne-la-Vallée, France.
E-mail:michel.delara@enpc.fr
[2] Institut de Mathématiques, Université Paul Sabatier, Toulouse, France.
E-mail: jbhu@math.univ-toulouse.fr

1. The S-procedure for quadratic functions

We recall here some basic results on the S-procedure when only quadratic functions are involved.

Let $Q_0, Q_1, ..., Q_p$ be $1+p$ real $n \times n$ symmetric matrices, let $c_0, c_1, ..., c_p \in \mathbb{R}^n$, let $d_0, d_1, ..., d_p \in \mathbb{R}$, let $q_i(.)$ be the associated *quadratic functions*

$$x \in \mathbb{R}^n \mapsto q_i(x) = \frac{1}{2} \langle Q_i x, x \rangle + \langle c_i, x \rangle + d_i.$$

Here and below, $\langle ., . \rangle$ stands for the usual inner-product in $\mathbb{R}^n$.

When $c_i = 0$ and $d_i = 0$, one speaks of *quadratic form* q_i instead of quadratic function. When $Q_i = 0$, one speaks of *linear (*or *affine) function*, and of *linear form* when moreover $d_i = 0$.

What is called *S-procedure* in Automatic Control Theory is the relationship between

$$(\mathcal{I}) \qquad (q_i(x) \geqslant 0 \text{ for all } i = 1, 2, ..., p) \Rightarrow (\ q_0(x) \geqslant 0)$$

and

$$(\mathcal{C}) \qquad \begin{cases} \text{There exist } \alpha_1 \geqslant 0, ..., \alpha_p \geqslant 0 \text{ such that} \\ q_0(x) - \sum_{i=1}^{p} \alpha_i q_i(x) \geqslant 0 \text{ for all } x \in \mathbb{R}^n. \end{cases}$$

The implication $[(\mathcal{C}) \Rightarrow (\mathcal{I})]$ is trivial. The issue is therefore the converse implication. We say that the *S-procedure is valid (*or *favorable,* or *lossless)* when this converse $[(\mathcal{I}) \Rightarrow (\mathcal{C})]$ holds true, that is to say the equivalence between the two statements $(\mathcal{I})$ and $(\mathcal{C})$. The equivalence may be used in its negative form, *i.e.* $[(not\ \mathcal{I}) \Leftrightarrow (not\ \mathcal{C})]$, whose essential content is $[(not\ \mathcal{C}) \Rightarrow (not\ \mathcal{I})]$.

Let us recall some important cases when the S-procedure is known to be valid (see [1, 6] and references therein):

- When $p = 1$, provided that there exists x_0 such that $q_1(x_0) > 0$.

- When all the involved functions q_i are linear forms. In that case, this is just the Minkowski-Farkas lemma (in its homogeneous form). Indeed, to have

$$\langle a_0, x \rangle - \sum_{i=1}^{p} \alpha_i \langle a_i, x \rangle \geqslant 0 \text{ for all } x \in \mathbb{R}^n$$

amounts to having $a_0 = \sum_{i=1}^{p} \alpha_i a_i$.

- When all the functions q_i involved are linear functions. In that case, this is again the Minkowski-Farkas lemma (non-homogeneous form). Indeed,

$$(\langle a_i, x \rangle - b_i \geqslant 0 \text{ for all } i = 1, 2, ..., p) \Rightarrow (\ \langle a_0, x \rangle - b_0 \geqslant 0)$$

is equivalent to

$$\begin{cases} \text{There exist } \alpha_1 \geqslant 0, ..., \alpha_p \geqslant 0 \text{ such that} \\ a_0 = \sum_{i=1}^{p} \alpha_i a_i \text{ and } b_0 - \sum_{i=1}^{p} \alpha_i b_i \leqslant 0. \end{cases}$$

2. The S-procedure for general functions

Let $f_0, f_1, ..., f_p : \mathbb{R}^n \to \mathbb{R}$ be $1+p$ (general) functions. For such a collection of functions, we mimic the S-procedure presented for quadratic functions. The objective is to have the equivalence between the two next assertions:

$$(\mathcal{I}) \qquad (f_i(x) \geqslant 0 \text{ for all } i = 1, 2, ..., p) \Rightarrow (\ f_0(x) \geqslant 0)$$

and

$$(\mathcal{C}) \qquad \begin{cases} \text{There exist } \alpha_1 \geqslant 0, ..., \alpha_p \geqslant 0 \text{ such that} \\ f_0(x) - \sum_{i=1}^{p} \alpha_i f_i(x) \geqslant 0 \text{ for all } x \in \mathbb{R}^n. \end{cases}$$

Sometimes, the expected result is written in the following "alternative theorem" form, with

$$(\textit{not } \mathcal{I}) \quad \begin{cases} \text{The system of inequations } (f_i(x) \geqslant 0 \text{ for all } i = 1, 2, ..., \ p) \\ \text{and } (\ f_0(x) < 0) \text{ has a solution } x \in \mathbb{R}^n. \end{cases}$$

The valid S-procedure then reads: exactly one of the two statements (*not* $\mathcal{I}$) and ($\mathcal{C}$) is true.

2.1 First step: when epi-convexity enters into the picture

For real-valued functions $\varphi_1, \varphi_2, ..., \varphi_k$ defined on $\mathbb{R}^n$, we use the standard notation $\text{Im}(\varphi_1, \varphi_2, ..., \varphi_k)$ for the image set $\{(\varphi_1(x), \varphi_2(x), ..., \varphi_k(x)) : x \in \mathbb{R}^n\}$.

The main result in this subsection is as follows:

Theorem 1[3]**.** *Suppose that:*
- There exists x_0 *such that* $f_i(x_0) > 0$ *for all* $i = 1, 2, ..., p$
and
- *The* epi-image *of the mapping* $(f_0, -f_1, -f_2, ..., -f_p)$, *that is*
$\text{Im}\,(f_0, -f_1, -f_2, ..., -f_p) + (\mathbb{R}_+)^{p+1}$, *is convex.*
Then the S-procedure is valid, that is to say: ($\mathcal{I}$) *and* ($\mathcal{C}$) *are equivalent.*

When the epi-image of the mapping $(f_0, -f_1, -f_2, ..., -f_p)$ is convex, we say that the mapping $(f_0, -f_1, -f_2, ..., -f_p)$ is *epi-convex.*

The first assumption: *There exists* x_0 *such that* $f_i(x_0) > 0$ *for all* $i = 1, 2, ..., p$ is common in Optimization; it is a Slater-type assumption. We refer to it hereafter as ($\mathcal{S}$).

A general remark. Suppose that $\text{Im}\,(g_0, g_1, g_2, ..., g_p)$ is convex. Then $\text{Im}\,(g_0, -g_1, -g_2, ..., -g_p)$ is also convex (as the image of the previous set under the linear mapping $(u_0, u_1, u_2, ..., u_p) \mapsto (u_0, -u_1, -u_2, ..., -u_p)$. Hence, $\text{Im}\,(g_0, -g_1, -g_2, ..., -g_p) + (\mathbb{R}_+)^{p+1}$, sum of two convex sets, is convex.

Example 1. Suppose that $g_0, g_1, g_2, ..., g_p$ are all convex functions. Then $\text{Im}\,(g_0, g_1, g_2, ..., g_p)$ is not necessarily convex but $\text{Im}\,(g_0, g_1, g_2, ..., g_p) + (\mathbb{R}_+)^{p+1}$ is convex, as this is easily seen from the basic definition of convexity of the g_i's. As a result, it comes from the main theorem above that the S-procedure is

[3]From J.-B. Hiriart-Urruty, *A remark on the general S-procedure.* Unpublished technical note (2020).

valid whenever f_0 is convex and the $f_1, f_2, ..., f_p$ are concave. We thus recover a classical result in convex minimization (with convex inequalities).

Example 2 (from [7, Example 3.1]). *Epi-convex mapping but with non-convex images.*

Let q_0 and q_1 be defined on $\mathbb{R}^2$ as follows:

$$q_0(x,y) = 2x^2 - y^2,\ q_1(x,y) = x + y.$$

Then, $\mathrm{Im}(q_0, q_1) = \left\{(u,v) \in \mathbb{R}^2 : u \geqslant -2v^2\right\}$ is not convex. However, the epi-image $\mathcal{F} = \mathrm{Im}(q_0, -q_1) + (\mathbb{R}_+)^2 = \mathbb{R}^2$ is convex.

Example 3. Indeed a lot of effort has been made by authors to detect (rather strong) assumptions ensuring that an image set like $\mathrm{Im}\,(g_0, g_1, g_2, ..., g_p)$ is convex, especially with quadratic g_i's (see [2, 5, 7, 8]). In addition to that, it has recently been proved that $\mathrm{Im}(q_1, q_2) + (\mathbb{R}_+)^2$ is convex for any pair of quadratic functions (q_1, q_2) ([2, assertion (b) in Theorem 4.19]). The question remains posed for a collection of three or more quadratic functions. We conjecture that the evoked convexity result does not hold true, but do not have any counterexample to offer.

2.2 A further step, via geometrical interpretations of $(\mathcal{I})$ **and** $(\mathcal{C})$

In this subsection, we intend to provide *a geometrical exact characterization of the statement* $(\mathcal{I})$ *and a "close to exact" geometrical characterization of the statement* $(\mathcal{C})$. For that purpose, we posit:

$$\begin{array}{rcl} -\left(\ \mathbb{R}_+^* \times (\mathbb{R}_+)^p\right) & = & \{(\beta_0, \beta_1, ..., \beta_p) : \beta_0 < 0 \text{ and } \beta_i \leqslant 0 \text{ for all } i\} \\ & = & \mathcal{K} \text{ (a polyhedral convex cone in } \mathbb{R}^{p+1}); \\ \mathrm{Im}\,(f_0, -f_1, -f_2, ..., -f_p) & = & \mathcal{F} \text{ (an image set in } \mathbb{R}^{p+1}, \text{ from the data}). \end{array}$$

Given a set S, we denote by co$S$ its convex hull, and by cone$S$ its convex conical hull, that is to say $\left\{\sum_{i=1}^k \lambda_i u_i : k \text{ positive integer, } \lambda_i > 0 \text{ and } u_i \in S \text{ for all } i\right\}$. To link the two definitions, we clearly have that $\mathrm{cone}S = \mathbb{R}_+^*(\mathrm{co}S) = \mathrm{co}\left(\mathbb{R}_+^* S\right)$.

Theorem 2. *We have the following:*

$$\begin{array}{rcll} (\mathcal{I}) \textit{ holds true} & \Leftrightarrow & \mathcal{F} \cap \mathcal{K} = \emptyset \ \Leftrightarrow\ \mathbb{R}_+^* \mathcal{F} \cap \mathcal{K} = \emptyset & (1) \\ (\mathcal{C}) \textit{ holds true} & \Rightarrow & \mathrm{co}\mathcal{F} \cap \mathcal{K} = \emptyset \Leftrightarrow\ \mathrm{cone}\mathcal{F} \cap \mathcal{K} = \emptyset & (2) \\ (\mathrm{cone}\mathcal{F} \cap \mathcal{K} = \emptyset \textit{ and } (\mathcal{S})) & \Leftrightarrow & (\mathrm{co}\mathcal{F} \cap \mathcal{K} = \emptyset \textit{ and } (\mathcal{S})) \Rightarrow (\mathcal{C}) \textit{ holds true} & (3) \end{array}$$

In short:

- A geometrical equivalent form of $(\mathcal{I})$ is $\mathcal{F} \cap \mathcal{K} = \emptyset$ or $\mathbb{R}_+^* \mathcal{F} \cap \mathcal{K} = \emptyset$.
- Provided the (slight) Slater-type assumption $(\mathcal{S})$ is satisfied on the f_i's, a geometrical equivalent form of $(\mathcal{C})$ is either $\mathrm{co}\mathcal{F} \cap \mathcal{K} = \emptyset$ or $\mathrm{cone}\mathcal{F} \cap \mathcal{K} = \emptyset$.

Proof of Theorem 2 **-** For the first equivalence in (1), maybe it is easier to consider (*not* $\mathcal{I}$). To have (*not* $\mathcal{I}$) means that there exists $x \in \mathbb{R}^n$ such

that: $f_i(x) \geqslant 0$ for all $i = 1, ..., p$, and $f_0(x) > 0$. This exactly expresses that $\mathcal{F} \cap \mathcal{K} \neq \emptyset$.

The second equivalence in (1) is clear from the relation $\mathbb{R}_+^* \mathcal{K} = \mathcal{K}$.

- We intend to prove the first implication in (2). We use the notation $\langle ., . \rangle$ for the usual inner product in $\mathbb{R}^{p+1} = \mathbb{R} \times \mathbb{R}^p$; thus $\langle \alpha, z \rangle = \alpha_0 z_0 + \alpha_1 z_1 + ... + \alpha_p z_p$ whenever $\alpha = (\alpha_0, \alpha_1, ..., \alpha_p) \in \mathbb{R} \times \mathbb{R}^p$ and $z = (z_0, z_1, ..., z_p) \in \mathbb{R} \times \mathbb{R}^p$.

By definition of the statement $(\mathcal{C})$ itself, there exists $\alpha = (\alpha_0, \alpha_1, ..., \alpha_p) \in -\mathcal{K}$ (that is to say $\alpha_0 > 0$ and $\alpha_i \geqslant 0$ for all $i = 1, ..., p$) such that

$$\langle \alpha, z \rangle \geqslant 0 \text{ for all } z = (z_0, z_1, ..., z_p) \in \mathcal{F}.$$

Clearly, this is equivalent to

$$\langle \alpha, z \rangle \geqslant 0 \text{ for all } z = (z_0, z_1, ..., z_p) \in co\mathcal{F}. \tag{4}$$

We prove by contradiction that $\text{co}\mathcal{F} \cap \mathcal{K}$ is empty. Therefore, suppose there exists some $\beta = (\beta_0, \beta_1, ..., \beta_p)$ lying in $\text{co}\mathcal{F} \cap \mathcal{K}$. Then, according to the inequality (4) just above,

$$\langle \alpha, \beta \rangle = \alpha_0 \beta_0 + \sum\nolimits_{i=1}^{p} \alpha_i \beta_i \geqslant 0. \tag{5}$$

But, by definition of $\mathcal{K}$, we have $\beta_0 < 0$ and $\beta_i \leqslant 0$ for all $i = 1, ..., p$. Thus, recalling the signs of the α_i's, one gets at $\langle \alpha, \beta \rangle < 0$, which contradicts (5).

As for the equivalence in the second part of (2), it is clear from the following observations: $\text{cone}S = \mathbb{R}_+^*(\text{co}S)$ and $\mathbb{R}_+^* \mathcal{K} = \mathcal{K}$.

- We now are going to prove that ($\text{co}\mathcal{F} \cap \mathcal{K} = \emptyset$ and $(\mathcal{S})$) $\Rightarrow$ $(\mathcal{C})$.

As expected in such a context, the proof is based on a separation theorem on convex sets. Because the two convex sets $\text{co}\mathcal{F}$ and $\mathcal{K}$ in $\mathbb{R}^{p+1}$ do not intersect, one can separate them properly: there exists $\alpha^* = (\alpha_0^*, \alpha_1^*, \alpha_2^*, ..., \alpha_p^*) \neq 0$ in $\mathbb{R} \times \mathbb{R}^p = \mathbb{R}^{p+1}$ such that

$$\sup_{b \in \mathcal{K}} \langle \alpha^*, b \rangle \leqslant \inf_{z \in \mathcal{F}} \langle \alpha^*, z \rangle = \inf_{z \in \text{co}\mathcal{F}} \langle \alpha^*, z \rangle, \tag{6}$$

$$\inf_{b \in \mathcal{K}} \langle \alpha^*, b \rangle < \sup_{z \in \mathcal{F}} \langle \alpha^*, z \rangle. \tag{7}$$

The second property (7) is useless here, due the nonemptiness of the interior of $\mathcal{K}$.

Due to the specific structure of $\mathcal{K}$, we deduce from (6) that $\alpha_i^* \geqslant 0$ for all $i = 0, 1, 2, ..., p$ and, further, $\sup_{b \in \mathcal{K}} \langle \alpha^*, b \rangle = 0$. Now, what is in the right-hand side of (6) is just $\inf_{x \in \mathbb{R}^n} \left[\alpha_0^* f_0(x) - \sum\nolimits_{i=1}^{p} \alpha_i^* f_i(x) \right]$. We therefore have proved that

$$\alpha_0^* f_0(x) - \sum\nolimits_{i=1}^{p} \alpha_i^* f_i(x) \geqslant 0 \text{ for all } x \in \mathbb{R}^n. \tag{8}$$

We claim that $\alpha_0^* > 0$. If not, we would have

$$-\sum\nolimits_{i=1}^{p} \alpha_i^* f_i(x_0) \geqslant 0,$$

which comes into contradiction with our SLATER-type assumption $(\mathcal{S})$: $f_i(x_0) > 0$ for all $i = 1, 2, ..., p$, and $\alpha_i^* \geqslant 0$ for all $i = 1, 2, ..., p$ (and one of them is > 0).

It now remains to divide (8) by $\alpha_0^* > 0$ to get at the desired result. □

Now, we are at the point for providing a rather general geometrical condition ensuring the validity of the S-procedure.

Theorem 3. *Assume the* SLATER*-type condition $(\mathcal{S})$, and suppose there exists a set $\mathcal{Z} \subset (\mathbb{R}_+)^{p+1}$ containing 0 such that $\mathbb{R}_+(\mathcal{F} + \mathcal{Z})$ is convex. Then the S-procedure is valid, that is to say: $(\mathcal{I})$ implies $(\mathcal{C})$ (hence $(\mathcal{I})$ and $(\mathcal{C})$ are equivalent).*

The set $\mathcal{Z}$ plays the role of a "convexifier" of the extended image-set $\mathbb{R}_+\mathcal{F}$. Let us see how the made assumption covers the three following known cases:

- (The most stringent one). When the image set $\mathcal{F}$ itself is convex; the assumed condition simply is satisfied with $\mathcal{Z} = \{0\}$.

- The epi-convex case (see §2.1): Take $\mathcal{Z} = (\mathbb{R}_+)^{p+1}$ to fulfill the proposed assumption. Here, instead of considering $\mathcal{F}$ solely, one takes its so-called "upper set" $\mathcal{F} + (\mathbb{R}_+)^{p+1}$.

- The "conical convex" case, *i.e.* when $\mathbb{R}_+\mathcal{F}$ is convex; again the considered assumption is verified with $\mathcal{Z} = \{0\}$.

Proof of Theorem 3.

First step. We start from the assumption $(\mathcal{I})$ in its equivalent form $\mathcal{F} \cap \mathcal{K} = \emptyset$ (see (1) in Theorem 2). We make it a bit more general by observing that $(\mathcal{F} + \mathcal{Z}) \cap \mathcal{K} = \emptyset$ for every set $\mathcal{Z}$ contained in $(\mathbb{R}_+)^{p+1}$. This is easy to check, as $\mathcal{Z}$ is contained in a cone placed "oppositely" to $\mathcal{K}$. We even go further by observing that $\mathbb{R}_+^* (\mathcal{F} + \mathcal{Z}) \cap \mathcal{K} = \emptyset$, since $\mathcal{K}$ is a cone. Finally, because $0 \notin \mathcal{K}$, we summarize the result of this first step in:

$$\mathbb{R}_+ (\mathcal{F} + \mathcal{Z}) \cap \mathcal{K} = \emptyset. \tag{9}$$

Second step. Since $0 \in \mathcal{Z}$, we have $\mathcal{F} \subset \mathcal{F} + \mathcal{Z}$; hence $\mathcal{F} \subset \mathbb{R}_+ (\mathcal{F} + \mathcal{Z})$. By the assumed convexity of $\mathbb{R}_+(\mathcal{F} + \mathcal{Z})$, we get at

$$\mathrm{co}\mathcal{F} \subset \mathbb{R}_+ (\mathcal{F} + \mathcal{Z}). \tag{10}$$

Final step. We infer from (9) and (10) that $co\mathcal{F} \cap \mathcal{K} = \emptyset$. It remains to apply the result (3) in Theorem 2 to get at the desired conclusion $(\mathcal{C})$. □

Conclusion

We have expressed all the ingredients of the general S-procedure in purely geometrical forms, as this was initiated in the seminal paper by FRADKOV ([3, pages 248 − 251]. In doing so, we hope to have shed a fresh new light at this kind of results, which could help to explain or to get at new conditions for the S-procedure to be valid.

References

1. K. Derinkuyu and M. C. Pinar, *On the S-procedure and some variants.* Math. Methods Oper. Res. 64, n°1 (2006), 55 – 77.

2. F. Flores-Bazan and F. Opazo, *Characterizing the convexity of joint-range for a pair of inhomogeneous quadratic functions and strong duality.* Minimax Theory and its Applications, Vol. 1, n°2 (2016), 257 – 290.

3. A. L. Fradkov, *Duality theorems for certain nonconvex extremal problems.* Siberian Math. Journal 14 (1973), 247 – 264.

4. S. V. Gusev and A. L. Likhtarnikov, *Kalman-Popov-Yakubovich lemma and the S-procedure: a historical survey.* Automation and Remote Control, Vol. 67, n°11 (2006), 1768 – 1810.

5. J.-B. Hiriart-Urruty and M. Torki, *Permanently going back and forth between the "quadratic world" and the "convexity world" in optimization.* J. of Applied Math. and Optimization 45 (2002), 169 – 184.

6. I. Polik and T. Terlaky, *A survey of the S-lemma.* SIAM Review, Vol. 49, n°3 (2007), 371 – 418.

7. B. T. Polyack, *Convexity of quadratic transformations and its use in control and optimization.* J. of Optimization Theory and Applications, Vol. 99, n°3 (1998), 553 – 583.

8. M. Ramana and A.J. Goldman, *Quadratic maps with convex images.* Rutcor Research Report 36 – 94 (October 1994).

9. M. Teboulle, *Nonconvex quadratic optimization: a guided detour.* Talk in Montpellier (September 2009), and

Hidden convexity in nonconvex quadratic optimization. Talk at One World Optimization Seminar (April 2020).