NASA/TM–2018–220117

The Number of Support Constraints for Overlapping Set Optimization with Nested Admissible Sets is Equal to One

John Siratt, Luis Crespo, Anthony Narkawicz, and César Muñoz
NASA Langley Research Center, Hampton, Virginia

October 2018

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

NASA/TM–2018–220117

The Number of Support Constraints for Overlapping Set Optimization with Nested Admissible Sets is Equal to One

John Siratt, Luis Crespo, Anthony Narkawicz, and César Muñoz
NASA Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

October 2018

Abstract

This paper reports on the formalization of a recent result by Crespo, et al., as found in the references. The formalized result bounds the number of support constraints in a particular type of optimization problem. The problem involves discovering an optimal member of a family of sets that overlaps each member of a constraining collection of sets. The particular case addressed here concerns optimizations in which the family of sets is nested. The primary results were formalized in the interactive theorem prover PVS and support the claim that a single support constraint exists in very general circumstances.

1 Introduction

This paper reports on the formalization of a recent result by Crespo, et. al. [3,4] concerning a particular type of optimization that has applications to interval predictor models. Such models are useful in predicting the range of possible output given a body of observational data consisting of input-ouput combinations, as described in [2] and [5]. Specifically, this paper formalizes a proof a theorem in [3, 4] that bounds the number of constraints which support a particular optimal solution.

Given m-dimensional euclidean space $\mathbb{R}^m$, a structured family of subsets $\mathcal{F}$ from $\mathbb{R}^m$, a finite collection $\mathcal{G}$ of arbitrary subsets from $\mathbb{R}^m$, and a function $u : \mathcal{F} \to \mathbb{R}$, what is the u-minimal element of $\mathcal{F}$ which intersects each member of $\mathcal{G}$? The solution (where it exists) is not computable in general, but may be approximated. This is precisely the problem addressed when interval predictor models are optimized. Within this context, the elements of $\mathcal{G}$ are the optimization constraints, and the function u is the objective function.

In this problem domain, the cardinality of $\mathcal{G}$ may be extremely large, requiring extensive computation to obtain the solution (or a reasonable approximation). It is often the case that a strict subset $\mathcal{S} \subset \mathcal{G}$ leads to the same optimal solution. If $\mathcal{S}$ is a minimal such subset – that is, any subset of $\mathcal{S}$ does not share the same optimum – then $\mathcal{S}$ is called a supporting set and its members support constraints. On one hand, the existence of such a subset allows us to simplify future computations by discarding all constraints but those in $\mathcal{S}$. On the other, the members of $\mathcal{S}$ are more likely to be outliers. Removing such outliers may yield a more desirable interval predictor model, and the probability that future observations fall within this new model is related to the number of constraints removed [1].

It would be useful then to have bounds on the cardinality of $\mathcal{S}$. Here we present a formalization of the theorem in [3, 4] stating bounds for arbitrary classes of constraints when the admissible sets in $\mathcal{F}$ are nested. The particular problem was posed by Luis Crespo and addressed in the context of his work in [3,4]. This paper formalizes and extends those results in the interactive theorem prover PVS [6]. The results of the formalization are then used to support a general statement about the problem.

2 Problem

Suppose that a family of sets $\mathcal{F}$ is defined as the image of a function $p : \mathbb{R} \to \mathcal{P}(\mathbb{R}^m)$, where $\mathcal{P}(\mathbb{R}^m)$ is the powerset of $\mathbb{R}^m$. Further, let the image of p – denoted $p[\mathbb{R}]$ – be nested with the property that for any $r, s \in \mathbb{R}$, if $r \leq s$ then $p(r) \subseteq p(s)$. Suppose that $u : \mathcal{F} \to \mathbb{R}$ respects the set inclusion ordering – that is, $A \subseteq B$ implies that $u(A) \leq u(B)$. Due to the nested nature of p in this case, minimality by set inclusion entails u-minimality. Because of this, u-minimality can be disregarded and the problem can be posed as one seeking the smallest member of $p[\mathbb{R}]$ by set inclusion.

Definition. *We say that $X \in p[\mathbb{R}]$ satisfies a constraint G if $X \cap G$ is nonempty. We abbreviate this with the predicate term $\mathcal{T}_G(X)$. Further, we say that $X \in p[\mathbb{R}]$ satisfies a collection of constraints $\mathcal{G}$ if, for all $G \in \mathcal{G}$, X satisfies G. We likewise abbreviate this as $\mathcal{T}_{\mathcal{G}}(X)$.*

Definition. *We say that X is a minimal p-set for G if $\mathcal{T}_G(X)$ and, for all $Y \in p[\mathbb{R}]$, if $\mathcal{T}_G(Y)$, then $X \subseteq Y$.*

Definition. *Similarly, we say that X is a solution for a set of constraints $\mathcal{G}$ if $\mathcal{T}_{\mathcal{G}}(X)$ and, for all $Y \in p[\mathbb{R}]$, if $\mathcal{T}_{\mathcal{G}}(Y)$, then $X \subseteq Y$.*

These definitions can be easily modified for use in cases where u-minimality is the target. This would only be a matter of replacing the subset condition $X \subseteq Y$ in the above definitions with the condition $u(X) \leq u(Y)$. We can see that in the case of inclusion, minimal p-sets and solutions are unique. This does not typically hold for u-minimality.

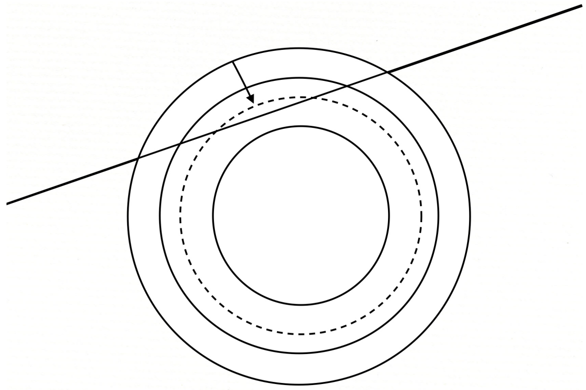

Figure 1: A minimal set may not exist due to the nature of the family of sets.

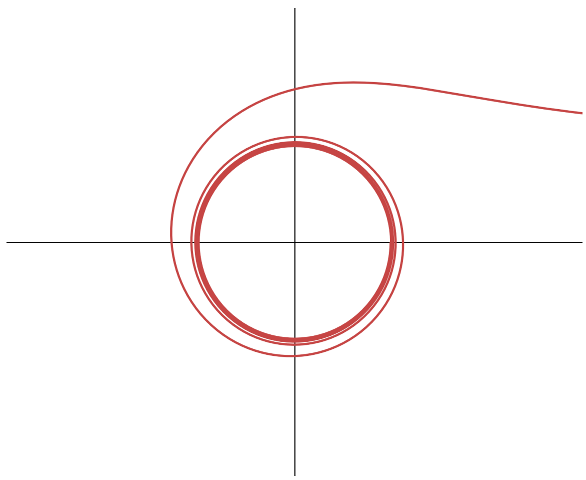

Figure 2: A minimal set may not exist due to the nature of the constraint.

One question that must be addressed is whether a minimal set (with respect to any form of minimality) actually exists for an arbitrary constraint or set of constraints. In general, the answer to this is no, and a few simple counterexamples are provided for inclusion minimality.

Suppose $\mathcal{F}$ contains all circles centered at some fixed point x with radius > 1, and all circles centered at x with radius $\leq 1/2$. If we have, for example, a linear constraint with a minimum distance to x between 1 and 1/2, then we will not have a minimal set which satisfies the constraint. This is illustrated in figure 1.

Suppose that $\mathcal{F}$ contains all circles centered at the origin. Consider a constraint defined by the polar equation $r(\theta) = 1 + \frac{1}{\theta}$, for $\theta \in (0, \infty)$. This constraint approaches, but never obtains, the unit circle. So in this case the lack of a minimal p-set is not due to some pathology of the family $\mathcal{F}$, but due to the constraint. This is illustrated in figure 2.

To addresses the behavior in figure 2, this section will deal only with constraints which form closed sets, which will be called closed constraints. A formalization of the solution is presented under the assumption that, for all nonempty subsets $G \subset \mathbb{R}^m$ closed in the standard topology, there exists a minimal p-set.

This condition may seem simple, but during formalization the consistency of the requirements on p are a concern. If there is some contradiction inherent within the requirements, then any statement predicated on these will be vacuously true. Thus, it is important to justify the consistency of the above stated requirements on the problem by providing an example that satisfies all of the requirements. The example is given in the justification below.

Justification. *Consider the function $\tilde{p} : \mathbb{R} \to \mathcal{P}(\mathbb{R}^m)$ defined by $\tilde{p}(r) = \mathbb{R}^m$ for all $r \in \mathbb{R}$. We see that this satisfies the nested requirement since, for $r \leq s$, $\tilde{p}(r) = \mathbb{R}^m \subseteq \mathbb{R}^m = \tilde{p}(s)$. It also trivially satisfies the condition on closed constraints since $\mathbb{R}^m$ is the minimal $\tilde{p}$-set for any constraint.*

Perhaps the most obvious nested p is one that produces a family of nested circles centered at some point. By definition, this satisfies the nested condition on p. It may be observed that any closed constraint in $\mathbb{R}^m$ will contain a point, perhaps not unique, which is closest to the origin of the family of circles. Assuming otherwise, a sequence can be constructed within the constraint which converges to a point outside of the constraint, giving a contradiction. The proof of this is beyond the scope of the formalization. The distance to this closest point is then the radius of the minimal radius circle of this family satisfying the constraint.

A key result in the case where p satisfies these conditions may now be stated. This was proven within PVS.

Theorem. *For any finite collection of closed constraints $\mathcal{G}$, there exists $G \in \mathcal{G}$ such that a minimal p-set X_G for G is also a solution for $\mathcal{G}$.*

Proof. Let $\mathcal{G} = \{G_i\}_{i \leq N}$ be a finite collection of closed constraints. For each G_n, by the conditions on p, there exists some $p(r_n)$ which is a minimal supporting p-set for G_n. The set of real numbers $\{r_i\}_{i \leq N}$ is a finite and contains a maximum. Without loss of generality, suppose $r_j \in \{r_i\}$ is the maximum of the set.

Note that, for all $i \leq N$, we have $p(r_i) \subseteq p(r_j)$ since $r_i \leq r_j$. Thus $\mathcal{T}_{\mathcal{G}}(p(r_j))$. Let $Y \in p[\mathbb{R}]$ such that $\mathcal{T}_{\mathcal{G}}(Y)$. This entails $\mathcal{T}_{G_j}(Y)$, but since $p(r_j)$ is a minimal p-set of G_j, we know $p(r_j) \subseteq Y$.

Therefore $p(r_j)$ is a solution for $\mathcal{G}$. □

With this result to perform the heavy lifting, a stronger statement can be made. This was proven within PVS.

Theorem. *For any finite collection of closed constraints $\mathcal{G}$, there is a singleton subset $\mathcal{S} \subset \mathcal{G}$ such that the solution of $\mathcal{S}$ is the solution of $\mathcal{G}$.*

Proof. By the previous theorem, there is some $G_j \in \mathcal{G}$ for which a minimal p-set X_G of G is a solution of $\mathcal{G}$. Since minimal sets and solutions are unique, it follows that X_G is the only solution of $\mathcal{G}$.

Consider the singleton collection of constraints $\mathcal{S} = \{G_j\}$. Since X_G is the unique minimal p-set satisfying G_j, it is apparent that X_G is the solution of $\mathcal{S}$. Therefore $\mathcal{S}$ and $\mathcal{G}$ share the same solution. □

This establishes $|\mathcal{S}| = 1$ in this special case and subsumes the theorem Crespo, et al., state in [4] [3].

3 Formalization

The formally verified version of the previous section was developed within PVS [6]. The above presentation parallels the formal definitions and proof with the exception of some routine steps that would detract from communicating the essential details.

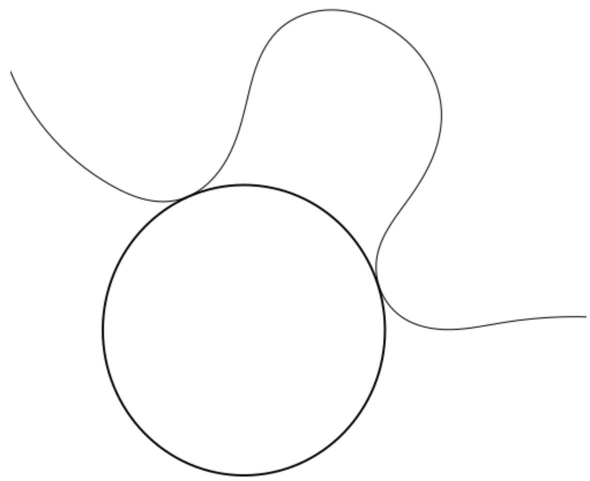

Figure 3: A non-linear constraint intersecting its minimal at two points.

Of note, the formalization defines closed sets in terms of the taxi metric, where given $x, y \in \mathbb{R}^m$ then the distance between the two points is calculated as

$$d(x, y) = \sum_{i \leq m} |x_i - y_i|.$$

Since this induces the standard topology on $\mathbb{R}^m$, it was selected in order to simplify the arithmetic anywhere it was needed. Since the $\tilde{p}$ chosen for the consistency proof did not rely on the fact that the constraints were closed, the choice of metric proved to be unimportant.

Thus this same $\tilde{p}$ establishes the consistency of requiring p to produce a minimal set for all constraints, calling into question the limitation of the problem statement to closed constraints. Observe again from figure 2 that if we require a minimal set for all constraints, then we are now excluding some of the most obvious families – such as nested circles – from our consideration.

Within the original problem domain, the constraints are generated by an equation $M(x, p) = y$, where $p \in \mathbb{R}^m$, and a collection of observations of the form (x, y). For each pair (x_j, y_j), the solution set of $M(x_j, p) = y_j$ forms a constraint. In [4] [3], the authors require that M be continuous with respect to p. As a result, as long as M is consistent with the observations, the constraints imposed will be nonempty closed sets in $\mathbb{R}^m$.

The requirement that M is continuous in p (and incidentally, also in x) is an artifact of the original problem [5]. These conditions ensure that $M(x, P)$ generates an interval predictor model when P is a connected set. Since this formalization abstracted out these details, an assumption of continuity would be meaningless. It was through study of counterexamples such as those shown that the more general requirement of closed constraints was developed.

Another interesting observation regards the temptation to assume that a supporting constraint in this case must have a singleton intersection with the solution. As a modification to figure 2, consider a family $\mathcal{F}$ that obtains the dashed circle. The dashed circle is then the minimal set satisfying the line, but it is obvious that the intersection is not a singleton set. In figure 3 we see an example of a nonlinear constraint and its minimal p-set which have two points in the intersection.

There are ways to exclude or include such examples using notions such as interior and exterior, and requiring some form of continuity or continuous deformation of p. However, doing so would drastically increase the complexity of the formalization. Fundamentally, the work in the previous section relies on the order properties of $p[\mathbb{R}]$, including the existence of minimal p-sets. The requirement concerning closed constraints is sufficient for this, but not strictly necessary.

The proofs do not rely on the assumption that the constraints are closed, but the fact that, for *some* class of constraints – here the nonempty closed sets – there are minimal satisfying elements in $p[\mathbb{R}]$. Restricting finite constraint sets to contain only the nonempty members of an arbitrary class of sets in $\mathbb{R}^m$ yields the same results. The entire development was duplicated as a parameterized PVS theory that takes a positive integer m and a predicate on sets in $\mathbb{R}^m$ to demonstrate this. This generalization was completely proven in PVS, extending every result previously stated to arbitrary classes of (nonempty) constraints.

In this setting the onus is on the user to verify that the choice of class leads to meaningful results. For instance, instantiating the theory with the predicate that always returns true is equivalent to the previously mentioned situation wherein we require all nonempty subsets to have a minimal in $p[\mathbb{R}]$. At the other extreme, the predicate that always returns false gives the empty class and everything is satisfied vacuously.

4 Results

It was first shown and verified that, for nested $p[\mathbb{R}]$ in which minimals existed for all closed constraints, any finite collection of closed constraints had a singleton supporting set. Next the extension replacing the class of closed constraints with an arbitrary class was verified.

Now consider an arbitrary $p : \mathbb{R} \to \mathcal{P}(\mathbb{R}^m)$ for which the image is a nested family of sets. Define the predicate φ on $X \subseteq \mathbb{R}^m$ as follows,

$$\varphi(X) \equiv \begin{cases} \top & X \text{ has minimal in } p[\mathbb{R}] \\ \bot & \text{otherwise.} \end{cases}$$

Let $\Phi = \{X \subseteq \mathbb{R}^m \mid \varphi(X)\}$ and define a Φ-constraint as any nonempty member of Φ. Based on the formal extension to all classes of constraints it can then be stated:

For any finite collection of Φ-constraints $\mathcal{G}$, there exists $G \in \mathcal{G}$ such that the minimal p-set X_G for G is also the solution for G.

Thus:

For any nested p and finite collection $\mathcal{G}$ of nonempty sets which have minimal p-sets, there exists $G \in \mathcal{G}$ for which the minimal p-set X_G for G is also the solution for G.

Hence the formal results imply a strong statement about this type of problem.

5 Conclusion

Assuming a family of sets which is both nested and contains a minimal satisfying element for every closed constraint, any finite collection of closed constraints has an optimal which is determined by a single constraint in the collection. This constraint may not be unique.

Of note, the results presented here are generalizations of that work since the continuity condition in the cited work implies that the imposed constraint will be closed. Further, we have presented a formally verified extension for arbitrary classes of constraints which imply that, for an arbitrary nested family $\mathcal{F}$, there will always be a singleton supporting set for finite collections of nonempty constraints which all have minimal satisfying sets in $\mathcal{F}$.

For more complex families of sets, such as hyper-rectangles in $\mathbb{R}^m$, the problem is still open. Proving bounds on the supporting subsets of this type of optimization problem promises to have beneficial applications in the area of control systems from which the cited papers are drawn. The problem domain is richer than the abstraction presented here, so other approaches may provide meaningful insights as well.

References

1. Marco C Campi and Simone Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.

2. Luis G Crespo, Daniel P Giesy, and Sean P Kenny. Interval predictor models with a formal characterization of uncertainty and reliability. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 5991–5996. IEEE, 2014.

3. Luis G Crespo, Daniel P Giesy, Sean P Kenny, and Julio Deride. A scenario optimization approach to system identification with reliability guarantees. submitted to American Control Conferences 2019.

4. Luis G Crespo, Sean P Kenny, Julio Deride, and Daniel P Giesy. A data-based approach to system identification, uncertainty quantification and robust control design. In *pending*.

5. Luis G Crespo, Sean P Kenny, and Daniel P Giesy. Interval predictor models with a linear parameter dependency. *Journal of Verification, Validation and Uncertainty Quantification*, 1(2):021007, 2016.

6. Sam Owre, John M Rushby, and Natarajan Shankar. Pvs: A prototype verification system. In *International Conference on Automated Deduction*, pages 748–752. Springer, 1992.

REPORT DOCUMENTATION PAGE				*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD-MM-YYYY)*
01-10-2018

2. REPORT TYPE
Technical Memorandum

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
The Number of Support Constraints for Overlapping Set Optimization with Nested Admissible Sets is Equal to One

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
John Siratt, Luis Crespo, Anthony Narkawicz, and César Muñoz

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
411931-02-51-07-01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L–20973

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/TM–2018-220117

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 03
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov.

14. ABSTRACT
This paper reports on the formalization of a recent result by Crespo, et al., as found in the references. The formalized result bounds the number of support constraints in a particular type of optimization problem. The problem involves discovering an optimal member of a family of sets that overlaps each member of a constraining collection of sets. The particular case addressed here concerns optimizations in which the family of sets is nested. The primary results were formalized in the interactive theorem prover PVS and support the claim that a single support constraint exists in very general circumstances.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	13	STI Help Desk (email: help@sti.nasa.gov)

19b. TELEPHONE NUMBER ***(Include area code)***
(757) 864-9658