Research Article

Local and Global Optima: Sheaves and Polynomial Approximations

Fernando Tohmé[1,2]

1. Department of Economy, Universidad Nacional del Sur, Argentina; 2. Instituto de Matemática de Bahía Blanca, CONICET

In this paper, we analyze how global optima of an agent's preferences can be reconstructed from the solutions found for local problems. A sheaf-theoretic analysis provides an abstract characterization of the global solution, and polynomial approximations are obtained when only a few local instances are available.

Corresponding author: Fernando, Tohmé, ftohme@criba.edu.ar

1. Introduction

An appropriate description of the behavior of economic agents is *as if* they run algorithms that translate given external states of the world into specific actions. Most of the economic literature roughly assumes that for each agent: (i) there exists a real-valued function U over outcomes ranking the well-being attributed by the agent to each of them; (ii) that the agent has the subjectively perceived goal of achieving the maximum of U; (iii) the agent consistently and systematically selects the maximum of U. In stronger versions the agent is assumed, (iv) to be able to specify the set of feasible alternatives, given by the objective conditions in which he has to make a decision. Furthermore, (v) the individual is able to find the actions that maximize U under those constraints.

From a logical point of view, these propositions can be considered separately. Different specifications of the decision processes can result from considering combinations of yes-no answers to $(i), \dots, (v)$, from the strictly defined *rational economic agent* (**REA**) to what we could be called a "*Dostoievskian*" *agent*.[1] We will focus on a particular instance, in which the agent tends towards the maximum of U, by solving constrained versions of the choice problem and using the solutions to build up closer approximations to U.

As a point of departure for our analysis, notice that the maximum of U was obtained as a solution to a global optimization problem. But the practicing economist (or the theorist, when building a specific model) represents the agent in a *partial* setting, where it is assumed that the solution does not result as a constrained version of a pre-defined global solution. Furthermore, the solution to the partial problem is seen as independent from the choices the agent could make in other contexts.[2]

It is rather easy to think of conditions under which several partial problems should be solved together, without allowing a decomposition in independent parts. The reason why this is not the usual case in the economic literature is straightforward, at least from a heuristic point of view: trying to represent and solve the global maximization problem of an agent is likely to result in an intractable problem, and the gains in analytical power may not even be large.

The standard practice of describing *model-specific* optimizations could be associated with a pragmatic approach, quite different from maximizing U. This procedure starts with the identification (by the agent) of a problem. Then, a problem-specific function u arises, focused on the arguments relevant to the question at hand. This u is maximized, yielding the problem-specific solution.

The model-specific scenario leads to a broader approach to agent behavior. Here the agent is not cognizant of his whole set of choices but acts by identifying specific problems and solving them one by one. The solutions of the specific problems may remain as independent pieces of information. In this paper we will focus on how they may lead to a learning process, in which the agent gathers information, first solving particular problems and from then *abducing*[3] hypotheses about the shape of the unknown U. This approximation may yield a function V, whose maximum can be contrasted, in a new problem with the solution of the maximization of the specific u. This allows to correct the specification of V. Repeated instances of this procedure yield, in the limit, a V that can be identified with U.

It is easy to think of stories supporting this approach. Individuals living in a steady environment tend to choose time-revered solutions, but once their world is unsettled by, say, an external shock, they will try to extrapolate them to the new situation. But only those agents that learn how to adapt this generalization to the new circumstances will be able to survive.

An evolutionary approach is not the focus of this paper. Instead, we will seek to characterize the process by which an agent may approximate the solution of the *global* optimization problem up from a set of *local* solutions (each for a different problem). A simple criterion of minimality, very much like *Occam's Razor* will help to keep the simplest hypotheses. While the informal description is intuitively clear, the features

that ensure a good approximation are rather involved. In technical terms, we are concerned with a *sheaf construction* under which a global solution under U is obtained from the solutions corresponding to a class of model-specific u functions. We will be concerned, in particular, with polynomial approximations, which can be patched up with relative ease.

The rest of this paper is organized as follows. Section 2 introduces a formal description of the problem. Section 3 provides a sheaf construction that allows to obtain global solutions up from a family of local utility functions. Section 4 considers the particular case of concave utility functions while section 5 presents the general case, analyzed through polynomial approximations. Section 6 discusses the evolution of global solutions. Section 7 concludes.

2. Decision-making: local vs. global

The traditional characterization of decision-making by an individual is as follows. Let $\mathcal{L}_i$ be a space of possible **alternatives** that an agent i may select.[4] Each $x \in \mathcal{L}_i$ is evaluated by means of a *utility* function, $U_i : \mathcal{L}_i \to \mathfrak{R}$. Given a family of constraints that limit the set of options open to the agent to $\hat{L} \subseteq \mathcal{L}_i$, the goal of the agent is to find a $\hat{\mathbf{x}}$ that maximizes U_i over $\hat{L}$.

To proceed, we fix the agent i and drop the corresponding subindex in $\mathcal{L}_i$ and U_i just denoting the space as $\mathcal{L}$ and the utility function U. We assume that the space of options $\mathcal{L}$ is a (real) Hilbert space. That is, it is a complete metric space with an inner product. On the other hand, to ensure the existence of a maximal $\hat{\mathbf{x}}$ we will further assume that $\hat{L}$ is a compact subset of $\mathcal{L}$, while U is a continuous function. Let $\mathbf{x}^*$ be a solution to the problem of optimizing U over $\hat{L}$.

Consider now a family $\{L^k\}_{k=0}^{\kappa}$ of linear subspaces of $\mathcal{L}$, each one the domain of a *local* problem. For each $k = 0, \ldots, \kappa$ we define a projection:

$$\text{Proj}_k : \mathcal{L} \to \bigcup_{k=0}^{\kappa} L^k$$

$\text{Proj}_k(x) = x^k \in L^k$, where x^k is the *projection* of $x \in \mathcal{L}$ on L^k.[5]

A compact subset $\hat{L}^k \subseteq L^k$ is the domain of a *local* maximization problem. The projection of a global solution $\mathbf{x}^*$ onto $\hat{L}^k$ will return the points that are the closest to $\mathbf{x}^*$. If the projection does not return a local solution, it can be generalized to a map $\Gamma_k : \hat{L} \to \hat{L}^k$:

$$\Gamma_k(x) = \{x^k \in \hat{L}^k : x^k \in \text{argmin}_{y \in \hat{L}^k} |y - \text{Proj}_k(x)|\}.$$

That is, $\Gamma_k(x)$ yields points in $\hat{L}^k$ that are the closest to the orthogonal projection of x on the subspace L^k of $\mathcal{L}$. Then, given any solution $\hat{\mathbf{x}}^k$ of the local maximization problem over $\hat{L}^k$, we want to ensure that $\hat{\mathbf{x}}^k \in \Gamma_k(\mathbf{x}^k)$.

More importantly, if the global solution is not given, we seek to find it by gluing together local ones, using the inverse of Γ_k. To formalize this we will introduce a category of local problems. This means that, given a family of local solutions $\{\hat{\mathbf{x}}^k\}_{k=0}^{\kappa}$, we seek a $\mathbf{x} \in \mathcal{L}$ such that $\Gamma_k(\mathbf{x}) = \hat{\mathbf{x}}^k$ for each k.

We will start solving this problem by providing a sheaf-theoretic characterization of the abduction.

3. The global solution: a sheaf-theoretic result

To obtain $\hat{\mathbf{x}}$ (the "global" solution) by patching up the "local" maxima, $\{\hat{\mathbf{x}}^k\}_{k=0}^{\kappa}$ we resort to some previous definitions:[6]

Definition 1. *Let $\mathcal{P}\mathbf{R}$ to be the category of local problems, where*

- *The objects are local problems s^k of the form $\langle \hat{L}^k, u^k, \hat{\mathbf{X}}^k \rangle$ where a continuous local utility function u^k is maximized over a compact set $\hat{L}^k \subset L^k$, a linear subspace of $\mathcal{L}$, yielding a family of solutions $\hat{\mathbf{X}}^k$.*

- *A morphism $\rho_{kj} : s^k \to s^j$ is defined as $\hat{L}^k \subseteq \hat{L}^j$, $u^k = u^j|_{L^k}$ and $\dim(L^k) \leq \dim(L^j)$.[7] It follows from this definition that an identity morphism $\rho_{kk} : s^k \to s^k$ trivially exists for every object s^k. Furthermore, given two morphisms $\rho_{kj} : s^k \to s^j$ and $\rho_{jl} : s^j \to s^l$ there exists their composition $\rho_{jl} \circ \rho_{kl} = \rho_{kl}$, since $\hat{L}^k \subseteq \hat{L}^j \subseteq \hat{L}^l$, $\dim(L^k) \leq \dim(L^j) \leq \dim(L^l)$ and by transitivity of the restrictions $u^k = u^j|_{L^k}$ and $u^j = u^l|_{L^j}$ we have $u^k = u^l|_{L^k}$.*

- *Given two local problems s^k and s^j, if $\hat{L}^k \cap \hat{L}^j \neq \emptyset$, a problem $s^{k \wedge j}$ with that domain has, by the properties of morphisms, has to verify that $u^{k \wedge j}$ is such that*

$$u^k|_{L^k \cap L^j} = u^{k \wedge j} = u^j|_{L^k \cap L^j}$$

We can also define $\mathcal{P}(\mathcal{L})$ as the category in which the objects are subsets of $\mathcal{L}$ and a morphism between two objects $f_{AB} : A \to B$ is defined as $A \subseteq B$.

Now we define a functor $F : \mathcal{P}\mathbf{R}^{op} \to \mathcal{P}(\mathcal{L})$, where $\mathcal{P}\mathbf{R}^{op}$ is the dual category of $\mathcal{P}\mathbf{R}$.[8]

Definition 2. *$F : \mathcal{P}\mathbf{R}^{op} \to \mathcal{P}(\mathcal{L})$ is such that:*

- *For each object s^k in $\mathcal{P}\mathbf{R}^{op}$, $F(s^k) = \{\mathbf{x} \in \mathcal{L} : \mathbf{x} \in \Gamma_k^{-1}(\hat{\mathbf{X}}^k)\}$.*

- *$F(\rho_{kj})$ is defined for any morphism $\rho_{kj}^{op} : s^j \to s^k$ in $\mathcal{P}\mathbf{R}^{op}$ (that is, $\rho_{kj} : s^k \to s^j$ in $\mathcal{P}\mathbf{R}$) as the inclusion $F(s^j) \subseteq F(s^k)$. It follows trivially that F preserves identities and compositions of morphisms in $\mathcal{P}\mathbf{R}^{op}$.*

The following result ensures that F is well-defined:

Proposition 1. *Given two local problems s^k and s^j, such that there exists a morphism $\rho_{kj} : s^k \to s^j$ in $\mathcal{P}\mathbf{R}$,*

$$\{\mathbf{x} \in \mathcal{L} : \mathbf{x} \in \Gamma_j^{-1}(\hat{\mathbf{X}}^j)\} \subseteq \{\mathbf{x} \in \mathcal{L} : \mathbf{x} \in \Gamma_k^{-1}(\hat{\mathbf{X}}^k)\}.$$

Proof: The existence of $\rho_{kj} : s^k \to s^j$ means, in particular, that (1) $\hat{L}^k \subseteq \hat{L}^j$ and (2) $u^k = u^j|_{L^k}$.

By (1), $\mathcal{L} \setminus \hat{L}^j \subseteq \mathcal{L} \setminus \hat{L}^k$. On the other hand, by (2) $\hat{\mathbf{x}}^k$ optimizes $u^j|_{L^k}$ and $\hat{\mathbf{x}}^j$ maximizes u^j. Then, for any $\mathbf{x}$ such that $\Gamma_j(\mathbf{x}) = \hat{\mathbf{x}}_j$, it follows that $\Gamma_k(\mathbf{x}) = \hat{\mathbf{x}}_k$. Then $\{\mathcal{L} : \mathbf{x} \in \Gamma_j^{-1}(\hat{\mathbf{X}}^j)\} \subseteq \{\mathcal{L} : \mathbf{x} \in \Gamma_k^{-1}(\hat{\mathbf{X}}^k)\}$.

All this means that we have defined a *contravariant functor* $F : \mathcal{P}\mathbf{R} \to \mathcal{P}(\mathcal{L}_i)$. In category-theoretical terms this means that F is a $\mathcal{P}(\mathcal{L}_i)$-valued *presheaf* on $\mathcal{P}\mathbf{R}$. If some additional properties are satisfied, F would also be a *sheaf*. The importance of this is that it would mean that there exists an object s^* in $\mathcal{P}\mathbf{R}$ such that $F(s^*)$ "glues" together all the objects $F(s^k)$ in $\mathcal{P}(\mathcal{L})$. To see that this is indeed the case and understand exactly what this means, we need some previous considerations.

The category $\mathcal{P}(\mathcal{L})$ includes products of families of objects. That is, given $A_1, A_2, \ldots, A_n$ objects in $\mathcal{P}(\mathcal{L})$, it exists an object $\prod_j A_j$ such that given projections $p_j : \prod_j A_j \to A_j$, if there exists another object C and morphisms $f_j : C \to A_j$ there exists a single $p : C \to \prod_j A_j$ such that $p_j \circ p = f_j$ for every j. But recall that in $\mathcal{P}(\mathcal{L})$ morphisms are inclusions of sets. Then $\prod_j A_j$ is the largest subset common to every A_j, that is, $\prod_j A_j = \bigcap_j A_j$.

In $\mathcal{P}\mathbf{R}$, the problem $s^{j \wedge l}$ is, as indicated in Definition 1, interpreted as the local problem in which the set of alternatives is $\hat{L}^j \cap \hat{L}^l$. The utility function $u^{j \wedge l}$ is the restriction of the continuous functions u^j and u^l over this domain, and is thus also continuous. The corresponding morphisms $\rho_{\{j,l\}j} : s^{j \wedge l} \to s^j$ and $\rho_{\{j,l\}j} : s^{j \wedge l} \to s^j$, which in $\mathcal{P}(\mathcal{L})$ induce, in turn, the morphisms $c_{j,l} : F(s^j) \to F(s^{j \wedge l})$ and $d_{j,l} : F(s^l) \to F(s^{j \wedge l})$. Varying j and l we obtain two morphisms c and d such that:

$$\prod_k F(s^k) \xrightarrow{c} \prod_{j,l} F(s^{j \wedge l}) \quad \text{and} \quad \prod_k F(s^k) \xrightarrow{d} \prod_{j,l} F(s^{j \wedge l})$$

Finally, in $\mathcal{P}\mathbf{R}$ an object $s^* = \max_k(s^k)$ can be defined, in which the set of alternatives is $\hat{L}^* = \bigcup_k \hat{L}^k$ and u^* which is a continuous function that satisfies that for each compact $\hat{L}^{k'} \subseteq \hat{L}^*$ corresponding to a problem $s^{k'}$, $u^{k'} = u^*|_{\hat{L}^{k'}}$. It is a well-defined object of $\mathcal{P}\mathbf{R}$ that is such that for every k there exists a morphism $\rho_{*k} : s^k \to s^*$ and a corresponding $\varepsilon_k \equiv F(s^k, s^*) : F(s^*) \to F(s^k)$. The latter induces in $\mathcal{P}(\mathcal{L}_i)$ the morphism $\varepsilon : F(s^*) \to \prod_k F(s^k)$.

Then, consider the following definition:

Definition 3 (Barr & Wells 1999). *$F : \mathcal{P}\mathbf{R}^{op} \to \mathcal{P}(\mathcal{L})$ is a sheaf if given s^* in $\mathcal{P}\mathbf{R}$, ε is the equalizer of c and d.*

We have:

Proposition 2. *$F : \mathcal{P}\mathbf{R}^{op} \to \mathcal{P}(\mathcal{L})$ is a sheaf.*

Proof: *It suffices to show that ϵ is an equalizer of c and d. This means that $(1) c \circ \epsilon = d \circ \epsilon$ and (2) for any other $f : A \to \prod_k F(s^k)$ for a $\mathcal{P}(\mathcal{L})$ object A, such that $c \circ f = d \circ f$ there exists a unique $g : A \to F(s^*)$ verifying that $\epsilon \circ g = f$. But in $\mathcal{P}(\mathcal{L})$ the morphisms are set inclusions. By the characterization of products, it means that $F(s^*) \subseteq \bigcap_k F(s^k) \subseteq \bigcap_{j,l} F(s^{j \wedge l})$, satisfying (1).*

Suppose that $A \subseteq \bigcap_k F(s^k) \subseteq \bigcap_{j,l} F(s^{j \wedge l})$, such that $A \nsubseteq F(s^*)$. This means that there exists an option $\mathbf{x} \in A \subseteq \bigcap_k F(s^k)$ such that $\mathbf{x} \notin F(s^*)$. This means that $\mathbf{x} \in \Gamma_k^{-1}(\hat{\mathbf{X}}^k)$ for every k but $\Gamma_k(\mathbf{x})$ does not maximize $\max_k(u^k)$ over $\bigcup_k \hat{L}^k$. But then, $\Gamma_k(\mathbf{x})$ maximizes each u^k when restricted to $\hat{L}^k$. Since for every compact set $\hat{L}^{k^*}$, $u^k = u^{k^*}|_{\hat{L}^k}$, we have that $u^{k^*}[\Gamma_k(\mathbf{x})] = u^k[\Gamma_k(\mathbf{x})]$. Contradiction. Then (2) is proven, showing that ϵ is the equalizer of c and d.

Example 1. Consider $\mathcal{L} = \mathbb{R}^2$ and two problems:

- s^k: $\hat{L}^k = [0,1] \times \{0\}$ and $u^k : x$. Thus $\mathbf{X}^k = \{(1,0)\}$.
- s^j: $\hat{L}^k = \{(x,y) : 0 \le x + y \le 1\}$ and $u^k : x + 2y$. Thus $\mathbf{X}^k = \{(0,1)\}$.

We can see that $\min(s^k, s^j) = s^k$ and since $F(s^j) \subseteq F(s^k)$, a straightforward application of Proposition 2 we have that $s^* = s^j$ and thus s^* is such that $\hat{\mathbf{X}}^* = \{(0,1)\}$ with $\Gamma_j[(0,1)] = (0,1)$ and $\Gamma_k[(0,1)] = (1,0)$.

The importance of Proposition 2 is that any $\mathbf{x} \in \hat{\mathbf{X}}^*$, when restricted on each $\hat{L}^k$ yields an element $\Gamma_k(\mathbf{x}) \in \hat{\mathbf{X}}^k$. That is, up from local solutions we have obtained a global one. To see the features of such global solutions we can consider two cases: one when each u_i^k is concave and a general one, which may be obtained via polynomial approximations.

4. Concave utility functions

Following the tradition in Economics we consider the case of a concave $\max_k(u^k)$, which by simplicity we will identify with U. Furthermore, $\hat{L} = \bigcup_k \hat{L}^k$. This means that each u^k is the continuous and concave restriction of U over L^k. It follows that:

Proposition 3. *If U is strictly concave and $u^k \equiv U|_{L^k}$ then $\hat{\mathbf{x}}^k = \Gamma_k(\hat{\mathbf{x}})$.*

Proof: *Suppose that $\hat{\mathbf{x}}^k \ne \Gamma(\hat{\mathbf{x}})$. Then,*

$$\operatorname{argmax}_{x \in \hat{L}^k} u^k(x) \neq \operatorname{argmax}_{x \in \hat{L}^k} U|_{L^k}(x).$$

But this is absurd, since $u^k \equiv U|_{L^k}$, $\hat{L}^k = \hat{L}|_{L^k}$, and U and u^k, being continuous and strictly concave, yield unique solutions.

Then, the solution of a sequence of problems $\{s^k\}_{k=0}^{\kappa}$, yields a family $\{\hat{\mathbf{x}}^k\}_{k=0}^{\kappa}$, such that for each s^k, $\hat{\mathbf{x}}^k = \Gamma(\hat{\mathbf{x}})$.

Then, we have that:

Proposition 4. *Assume that $\{s^k\}_{k=0}^{\kappa}$ is a family of problems in which each u^k is the continuous and strictly concave restriction of a continuous and strictly concave function U over a compact $\hat{L} \subseteq \mathcal{L}$. If $\{d_k\}_{k=0}^{\kappa}$ is the set such that for each k, $d_k = \dim(L_k)$ and the solutions $\{\mathbf{x}^k\}_{k=0}^{\kappa}$ are linearly independent and span $\mathcal{L}$, a unique approximation $\tilde{\mathbf{x}}$ to the global maximum $\hat{\mathbf{x}}$ can be uniquely determined if $\sum_{k=0}^{\kappa} d_k = \dim(\mathcal{L})$.*

Proof: *For each $\hat{\mathbf{x}}^k$, the unique solution of the maximization of the concave function u^k, consider $\operatorname{Proj}_k^{-1}(\hat{\mathbf{x}}^k) \subseteq \Gamma_k^{-1}(\hat{\mathbf{x}}^k)$. For $k = 0, \ldots, \kappa$, we have that $\operatorname{Proj}_k^{-1}(\hat{\mathbf{x}}^k)$ yields an affine subspace of $\mathcal{L}$, say l_k^{-1}. Then, by the linear independence of $\{\hat{\mathbf{x}}^k\}_{k=0}^{\kappa}$, $\cap_{k=0}^{\kappa} l_k^{-1}$ is an affine subspace of $\mathcal{L}$. Since $\{\hat{\mathbf{x}}^k\}_{k=0}^{\kappa}$ is a basis of $\mathcal{L}$, $\hat{\mathbf{x}} \in \cap_{k=0}^{\kappa} l_k^{-1} \neq \emptyset$ and since $\sum_{k=0}^{\kappa} = \dim(\mathcal{L})$, there exists at most one point in that intersection, which we denote as $\tilde{\mathbf{x}}$. Notice that, in general, $\tilde{\mathbf{x}} \neq \hat{\mathbf{x}}$. But, by construction, $\tilde{\mathbf{x}}$ is the unique element in $\mathcal{L}$ such that, for each k minimizes $|\operatorname{Proj}_k(\tilde{\mathbf{x}}) - \operatorname{Proj}_k(\hat{\mathbf{x}})|$.*

While the conditions of linear independence of local maxima and the dimensionality of the "covering" of the global problem by the local ones may seem too demanding, they indicate that, to recover the global maximum, the family of local problems should be *informative*. That is, we can approximate the global solution by putting together the scattered pieces of information provided by the local solutions.

As it is well known, a sufficient condition for $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$ is that each s^k must be such that:

- $\hat{L}$ can be written as $\hat{L} = \hat{L}^k \times \hat{L}^{k\perp}$, where $\hat{L}^{k\perp}$ is a compact set in which its elements are orthogonal to those in $\hat{L}^k$.

- $0 \in \hat{L}^{k\perp}$, for $0 \in L^{k\perp}$, the subspace of $\mathcal{L}$ of which $\hat{L}^{k\perp}$ is a subset.

- $U = u^k + u^{k\perp}$, such that $U(\hat{\mathbf{x}}) = u^k(\hat{\mathbf{x}}^k) + u^{k\perp}(\hat{\mathbf{x}}^{k\perp})$, where U, u^k and $u^{k\perp}$ are continuous strictly concave functions.

Example 2. *Consider $U(x, y) = 3 - x^2 - y^2 + 2y + x$, defined over the compact set $\{(x, y) : 0 \leq x, y \leq 1\} \subset \mathbb{R}^2$. U is concave. The global problem of the maximization of U has a unique solution: $\hat{\mathbf{x}} = (\frac{1}{2}, 1)$. Then, consider two local problems:*

- s^1: *the maximization of the restriction of U over the compact segment $[0, 1]$ of the 1-dimensional subspace consisting of the x-axis of $\mathbb{R}^2$. The restriction is $u(x) = 3 - x^2 + x$, which yields the unique solution $\hat{\mathbf{x}}^1 = \frac{1}{2}$.*

- s^2: *the maximization of the restriction of U over the compact segment $[0, 1]$ of the 1-dimensional subspace consisting of the y-axis of $\mathbb{R}^2$. The restriction is now $u(y) = 3 - y^2 + 2y$, which yields a solution $\hat{\mathbf{x}}^2 = 1$.*

Since in $\mathbb{R}^2$, the solutions of the local problems are the linearly independent vectors $(\frac{1}{2}, 0)$ and $(0, 1)$, and the sum of the dimensions of the local subspaces is 2, the intersection $\mathrm{Proj}_1(\frac{1}{2}) = \{(\frac{1}{2}, y)\}$ and $\mathrm{Proj}_2(1) = \{(x, 1)\}$ yields the global solution $(\frac{1}{2}, 1)$.

5. Local-to-global via polynomial approximations

Here we will consider the class of solutions to a family of problems $\{s^k\}_{k=0}^m$ such that $L^* \subseteq \bigcup_{k=0}^m L^k = \mathcal{L}$, is the domain of a problem s^* with utility u^*. We study the problem of finding a polynomial approximation to u^*, yielding a function V with the same maxima. That is, given a family of problems in $\mathcal{PR}$ we seek to find the simplest[9] continuous function $V : \mathcal{L} \to \mathbb{R}$ such that the Γ-projections of its maxima in L^* are in $\hat{\mathbf{X}}^k$, for $k \in \{0, \ldots, m\}$.[10]

The subspaces underlying $\{s^k\}_{k=0}^m$, $\{\langle L^k \rangle\}_{k=0}^m$ determine a partition of the entire space $\mathcal{L}$. The classes in the partition are of three types:

- Sets of the form $L^k \setminus \bigcup_{j \neq k; j \in \{0, \ldots, m\}} L^j$. That is, the subsets of the local spaces that do not overlap with any other of the given subspaces.

- Sets of the form $\bigcap_{k \in \gamma \subseteq \{0, \ldots, m\}} L^k$, such that there is no $\gamma' \subseteq \{0, \ldots, m\}$ with $\bigcap_{j \in \gamma'} L^j \subseteq \bigcap_{k \in \gamma} L^k$. This means that each intersection among subspaces is decomposed in minimal intersections, each one being a class in the partition.

- $\mathcal{L} \setminus \bigcup_{k=0}^m L^k$. That is, the entire space without the subspaces corresponding to the local problems $s^0, \ldots, s^m$.

Notice that sets in these classes are not necessarily compact nor convex. Nevertheless we can define $\mathbf{F}_m^r$ to be the class of continuous piecewise multidimensional polynomials with real coefficients of degree at most r, such that each $f \in \mathbf{F}_m^r$ is $f : \mathcal{L} \to \mathbb{R}$, defined on the aforementioned partition of $\mathcal{L}$:

$$f(x) = \begin{cases} p^k(x), & \text{if } x \in L^k \setminus \bigcup_{j \neq k} L^j \\ p^\gamma(x), & \text{if } x \in \bigcap_{k \in \gamma} L^k, \text{ and } x \notin \bigcap_{j \in \gamma'} L^k, \text{ for } \gamma' \neq \gamma \\ p^\alpha(x), & \text{if } x \in \mathcal{L} \setminus \bigcup_{k=0}^m L^k. \end{cases}$$

where each $p^k(x)$ or $p^\gamma(x)$ is a polynomial defined on $x_1, \ldots, x_\lambda$, such that $x = \sum_{j=1}^\lambda x_j e_j$ where $\{e_1, \ldots, e_\lambda\}$ is the orthonormal basis of $\mathcal{L}$. They have the lowest degree such that $\hat{\mathbf{x}}^k$, expressed in terms of the e_js is a maximum of u^k over $\hat{L}^k \subseteq L^k$ (if it does not belong to any other subspace L^j) or over $\bigcap_{k \in \gamma} L^k$. The polynomial p^α, instead, describes the value of f over all the possible local problems not yet tested.

If $\mathcal{B}(S)$ is the *boundary* of a subset $S \subseteq \mathcal{L}$, the continuity of each f is ensured by requiring that:[11]

- if $x \in \mathcal{B}(L^k \setminus \bigcup_{j \neq k} L^j) \cap \mathcal{B}(\bigcap_{k \in \gamma} L^k)$, then $p^k(x) = p^\gamma(x)$,

- if $x \in \mathcal{B}(L^k) \cap \mathcal{B}(\mathcal{L} \setminus \bigcup_{k=0}^m L^k)$, then $p^k(x) = p^\alpha(x)$,

- if $x \in \mathcal{B}(\bigcap_{k \in \gamma} L^k) \cap \mathcal{B}(\mathcal{L} \setminus \bigcup_{k=0}^m L^k)$, then $p^\gamma(x) = p^\alpha(x)$.

The conditions on the boundaries ensure that for the characterization of f, each class in the partition of $\mathcal{L}_i$ is compact. This is particularly important for the existence of f:

Lemma 1. *f is well-defined.*

Proof: *It is easy to see that p^k and p^γ, satisfying the boundary conditions, are well-defined (Propositions 3 and 4 elaborate this point). The definition of p^α is more involved. First, let us focus on the components of $\bigcup_{k=0}^m L_i^k$. Nothing ensures that for any pair $x, x' \in \bigcup_{k=0}^m L_i^k$ there exists a curve $[x, x']$, joining them, entirely contained in $\bigcup_{k=0}^m L_i^k$. If such curve exists, we take $\mathcal{B}(\bigcup_{k=0}^m L_i^k) \cap \mathcal{B}(\mathcal{L}_i \setminus \bigcup_{k=0}^m L_i^k)$, which by definition admits a continuous function (the polynomials of the forms p^k and p^γ on it). Then, a straightforward application of the Brouwer-Urysohn-Tietze Extension Lemma ([11]) allows to find a continuous function over the entire space, coinciding with the continuous function on the boundary (we just need to consider the constraint on $\mathcal{L}_i \setminus \bigcup_{k=0}^m L_i^k$). Then, the Stone-Weierstrass Approximation Theorem ensures the existence of p^α arbitrarily close to the extended continuous function ([2]).*

By a direct application of the Hahn-Banach Theorem we can find hyperplanes in $\mathcal{L} \setminus \bigcup_{k=0}^m L^k$ pairwise separating the components. The intersection of those hyperplanes defines (not necessarily "rectangular") boxes, each one containing a component. On each of these boxes, an application of the Brouwer-Urysohn-Tietze's Lemma yields a continuous extension. On any of the hyperplanes, say $H^{1,2}$, which separates boxes B^1 and B^2, let us consider a point x, and its value under the continuous function found for B^1, $f^1(x)$. Suppose that $f^1(x) \neq f^2(x)$, where f^2 is the continuous function on B^2. Then, we can build a continuous function f, such that $f(x) = f^1(x)$, for every $x \in B^1$. Since $\mathcal{L}$ separates points, we can find a hyperplane $H^{1,2'}$ with the same normal vector as $H^{1,2}$, such that $f(x) = f^2(x)$ on $H^{1,2'} \cap B^2$. For any $x \in B^2$, between $H^{1,2'}$ and $H^{1,2}$, we

can pick $x^1 \in H^{1,2}$ and $x^2 \in H^{1,2'}$ such that $x \in [x^1, x^2]$, a segment collinear with the normal vector of both hyperplanes. Then define $f(x) = \frac{x-x^1}{x^2-x^1} f(x^1) + \frac{x^2-x}{x^2-x^1} f(x^2)..$ Then, f becomes continuous over $B^1 \cup B^2$.

On lower-dimensional intersections between hyperplanes the same procedure can be applied, only now based on defining, instead of segments, higher-dimensional simplexes among points in the parallel hyperplanes and points in the intersections among the original hyperplanes. This yields again a continuous function. By means of this brute force approach, we obtain a continuous function over the entire $\mathcal{L} \setminus \bigcup_{k=0}^{m} L^k$. By the Stone-Weierstrass theorem, there exists a polynomial p^α arbitrarily close to this function.

The properties of the polynomials p^k and p^γ are easy to obtain:

Proposition 5. *If $x \in L^k$ while $x \notin L^j$ for $j \neq k$, $f(x)$ is a polynomial of degree at most 2.*

Proof: *The degree of $p^k(x)$ is 2 if $\hat{\mathbf{x}}^k \in \hat{L}^k$ is an interior point of L^k, since a quadratic expression ensures a single maximum. On the other hand, if $\hat{\mathbf{x}}^k$ is on $\mathcal{B}(L^k)$, either a quadratic or a linear function (if the boundary is zero-dimensional) may yield it as a maximum.*

On the other hand, the degree of $p^\gamma(x)$ depends on the number of local maxima in $\bigcap_{k \in \gamma} L^k$:[12]

Proposition 6.

1. *If there is no $\hat{\mathbf{x}}^k$ in $\bigcap_{k \in \gamma} L^k$, $p^\gamma(x)$ is a constant function for every $x \in \bigcap_{k \in \gamma} L^k$.*

2. *If there is a single $\hat{\mathbf{x}}^k \in \bigcap_{k \in \gamma} L^k$, $p^\gamma(x)$ is at most a quadratic function with a single maximum on $\hat{\mathbf{x}}^k$.*

3. *If there are more than one local maxima in $\bigcap_{k \in \gamma} L^k$, the degree of $p^\gamma(x)$ is at least 3.*

Proof: *Case (1) follows from the fact that $p^\gamma(x)$ has the lowest possible degree, without having to reach a maximum on $\bigcap_{k \in \gamma} L^k$. A constant function satisfies these two conditions.*

Case (2) is identical to $p^k(x)$, since $\bigcap_{k \in \gamma} L^k \subseteq \mathcal{B}(L^k)$.

Finally, case (3) follows immediately from the fact that $p^\gamma(x)$ has more than one maximum and no polynomial function of degree less than 2 yields a multiplicity of local maxima.

Then, it follows that:

Theorem 1. *The highest degree r of $p^\gamma(x)$ is equal to the highest number of local maxima in $\bigcap_{k \in \gamma} L^k$ for any $\gamma \subseteq \{0, \ldots, m\}$.*

Proof: *Immediate from case (3) in Proposition 6 and the continuity of f.*

These results do not yield a single f. Worse yet, the cardinality of $\mathbf{F}_m^r$ is infinite. Fortunately, we can drastically reduce the number of these alternatives. To see this, just consider $\mathbf{P}(\mathcal{L})$, the set of polynomials

that can be defined on $\mathcal{L}$. If the local solutions $\{\langle L^k, \hat{L}^k, \hat{\mathbf{x}}^k \rangle\}_{k=0}^m$ yield a partition of $\mathcal{L}$ in λ classes, we have that $\mathbf{F}_m^r$ is a subset of $\mathbf{P}(\mathcal{L})^\lambda$. On the other hand, if τ is the number of common boundaries between any two partitions, $\mathbf{P}(\mathcal{L})^\tau$ denotes the class of vectors in which the components are polynomials defined over the boundaries. Both $\mathbf{P}(\mathcal{L})^\lambda$ and $\mathbf{P}(\mathcal{L})^\tau$ are *modules*, i.e. generalizations of vector spaces in which the scalars are now elements of the *ring* of polynomials $\mathbf{P}(\mathcal{L})$ [3].

We can define a linear transformation $T : \mathbf{P}(\mathcal{L})^\lambda \to \mathbf{P}(\mathcal{L})^\tau$, where T, being linear, can be given a matrix representation with:

- τ rows,

- λ columns,

- each element jl in the matrix is $0, 1$ or -1 indicating if τ_j is a boundary of λ_l or not,

- each row τ_j contains only two non-zero entries, 1 in jl and the other -1, in jk, corresponding to the two partitions λ_l and λ_k that share the boundary τ_j.

Then, for each $f \in \mathbf{F}_m^r$, Tf yields an algebraic sum of the polynomials that constitute f, evaluated at the common boundaries. These sums can either be of the $p^k - p^\gamma$, $p^k - p^\alpha$ or $p^\gamma - p^\alpha$ types. By definition of f, these are all zero, and therefore $Tf = 0$. This means that $\mathbf{F}_m^r$ is the *kernel* of T.

It is a well known fact (see [4]) that the kernel of the transformation $T : \mathbf{P}(\mathcal{L})^\lambda \to \mathbf{P}(\mathcal{L})^\tau$ is a *submodule* of $\mathbf{P}(\mathcal{L})^\lambda$. Then, Proposition 5.2.7 in [5] ensures the existence of a finite *Gröbner basis* of $\mathbf{F}_m^r$:[13]

Proposition 7. *There exists* $G = \{\bar{f}_j\}_{j \in J} \subset \mathbf{F}_m^r$, *with* $|J| < \infty$, *such that any* $f \in \mathbf{F}_m^r$, *is* $f = \sum_{j \in J} p^j \bar{f}_j$, *where each* $p^j \in \mathbf{P}(\mathcal{L})^\lambda$.

Then, given a family of local solutions $\{\langle L^k, \hat{\mathbf{x}}^k \rangle\}_{k=0}^m$, we can define a function $V^m \in \mathbf{F}_m^r$, $V^m(x) = \sum_{j \in J} \bar{f}_j$, where $\{\bar{f}_j\}_{j \in J}$ is the Gröbner basis of $\subset \mathbf{F}_m^r$.[14] In turn, the maximization over $\mathcal{L}$ of $V^m(x)$ yields a family $\{\hat{\mathbf{x}}_j^{*m}\}_{j=1}^\mu$, which we deem the maximal elements corresponding to $\{s^k\}_{k=0}^m$. Notice that $\mu \leq \lambda$, indicating that the number of maximal elements is at most equal to the number of classes in the partition of $\mathcal{L}$ given the local solutions.

6. Evolving global solutions

While $\hat{\mathbf{x}}^{*m}$ may be a close approximation of the intended global solution, further local solutions will allow its improvement. So, if an additional solution is included, namely a local subspace L^{m+1}, with a corresponding maximum $\hat{\mathbf{x}}^{m+1}$, a new family $\mathbf{F}_{m+1}^{r'}$ can be obtained.

Proposition 8. *r' is either r or $r + 1$.*

Proof: *r', as seen in Theorem 4, equals the multiplicity of local maxima in intersections of the family $\{L^k\}_{k=0}^{m+1}$. Since L^{m+1} may add only one more maximum to the intersections already obtained for $\{L^k\}_{k=0}^{m}$, then the number of local maxima is now either r or $r + 1$ (if the new maxima adds to the intersection where the multiplicity of maxima was exactly r).*

The particular refinement of the abduction performed for m local problems is as follows:

Theorem 2. $\mathbf{F}_{m+1}^{r'} \equiv \mathbf{F}_m^r$ *over* $\bigcup_{k=0}^{m} L^k \setminus L^{m+1}$.

Proof: *Suppose, by contradiction, that there exists $f' \in \mathbf{F}_{m+1}^{r'}$ such that $f' \neq f$ for every $f \in \mathbf{F}_m^r$. That is, for every $f \in \mathbf{F}_m^r$ there exists $x \in \bigcup_{k=0}^{m} L_i^k \setminus L^{m+1}$ such that $f'(x) \neq f(x)$. But then, we have the following taxonomy of cases:*

- *x is in a L^k (without being in L^j for $j \neq k$, $j \in \{0, \ldots, m+1\}$). By definition any p'^k that is maximum over $\hat{\mathbf{x}}^k$ is acceptable, and there exists at least one $f \in \mathbf{F}_m^r$ such that $p^k(x) = p'^k$.*

- *x is in a $\bigcap_{k \in \gamma} L^k$ and $x \notin L^{m+1}$. Again, by definition any p'^γ that ensures continuity with the p'^ks and maximality (if there is a maximum in the intersection) is acceptable and so, there is at least a $f \in \mathbf{F}_m^r$ such that $p^\gamma(x) = p'^\gamma$.*

Therefore, there exists at least one $f \in \mathbf{F}_m^r$ such that $f'(x) = f(x)$ for every $x \in \bigcup_{k=0}^{m} L^k \setminus L^{m+1}$. Contradiction.

Of course, new functions may obtain, by defining polynomials of the lowest degree on $L^{m+1} \cap \bigcap_{k \in \gamma} L^k$ for $\gamma \subseteq \{0, \ldots, m\}$. The fact is that any new problem yields a finer partition of $\mathcal{L}$. On the classes of the new partition that were part of the original partition, the functions remain the same. But some classes become partitioned in finer classes. The only proviso is that the old polynomials should coincide in the new boundaries with the new ones.

By finding the Gröbner basis of $\mathbf{F}_{m+1}^{r'}$, we obtain now $V^{m+1}(x)$ and $\{\hat{\mathbf{x}}_j^{*m+1}\}_{j=1}^{\mu(m+1)}$. The question we want to explore now is whether any of the $\hat{\mathbf{x}}_j^{*m+1}$ comes close to a "true" global maximum when m becomes larger.

To make sense of this question we have to assume that we are back to the conditions in section 3, namely that there exists a global strictly concave utility function $U = u^*$ such that for every k, $u^k \equiv U|_{\hat{L}^k}$. If the local utility functions were not related to U, no relation between V and U could be predicated.

Then, we have that:

Theorem 3. *If $\mathcal{L} \subseteq \bigcup_{k=0}^{\kappa} L^k$, there exists $\hat{m} > 0$ such that for $m > \hat{m}$ there exist maximal element $\hat{\mathbf{x}}^{*m}$ of $V^m(x)$ such that $|\hat{\mathbf{x}}^{*m} - \hat{\mathbf{x}}| = 0$, where $|\cdot|$ is the norm on $\mathcal{L}$.*

Proof: *Since $\mathcal{L} \subseteq \bigcup_{k=0}^{\kappa} L^k$, consider a closed ball $B(\hat{\mathbf{x}}, \epsilon)$ of radius ϵ, around $\hat{\mathbf{x}}$. $B(\hat{\mathbf{x}}, \epsilon)$ is compact. Then, since $B(\hat{\mathbf{x}}, \epsilon) \subseteq \bigcup_{k=0}^{\kappa} L^k$, there exists a finite family $\{L^k\}_{k=k_1,\ldots,k_K}$ that covers $B(\hat{\mathbf{x}}, \epsilon)$. This means that $\hat{\mathbf{x}} \in \bigcup_{k=k_1}^{k_K} L^k$. By choosing these subspaces and the corresponding local solutions we obtain $\mathbf{F}_{k_K}^r$ and consequently V^{k_K}.*

Furthermore, by a straightforward application of Theorem 2, we have that $V^m(\hat{\mathbf{x}}) = V^{k_K}(\hat{\mathbf{x}})$, for every $m \geq k_K$. On the other hand, from $\hat{\mathbf{x}} \in \bigcup_{k=k_1}^{k_K} L^k$ we have that $\hat{\mathbf{x}}^{k_j} = \hat{\mathbf{x}}$ for some $j = 1, \ldots, K$ (because of the fact that $u^k \equiv U|_{L^k}$). Then, one of the maximal elements of V^m for $m > \hat{m} = k_K$ will be $\hat{\mathbf{x}}$.

7. Conclusions

Theorem 3 shows that even by groping in the dark, the global sheaf-theoretic solution could be found, by selecting problems that cover the whole space of possibilities. Of course, this is not always possible, since the local problems usually present themselves without concern for the global solution. But if the global u^* function is concave, as it is usually assumed in most economic applications, then V will yield the global maximum of u^*.

The results presented here are quite general, keeping to a minimum the requirements on functions and choice sets. From a Machine Learning perspective, this procedure could allow finding solutions without the risk of overfitting, exhibiting a behavior close to what is known for neural networks as *grokking*[6].

Notes

MSC2020: 91B02, 12D10, 90C59, 18F20.

Footnotes

[1] This notion is due to Ana Marostica and Daniel Heymann. The Underground Man in Fyodor Dostoyevsky's *Notes from Underground*, avoids choosing the outcome that maximizes U, just to show that he is free and can overcome the determinism implicit in that choice.

[2] This is the realm of Partial Equilibrium Analysis, which involves "demand" and "supply" functions of certain goods of interest. See [7].

^{3} See [8] for a thorough discussion of the scope of this inference mechanism.

^{4} The meaning of these alternatives depends on the context. If the agent is a consumer in a competitive market with a finite number of goods, she has to choose a vector of those commodities. In a planning problem, she has to select a plan specifying the amounts of resources used or consumed at each period, etc.

^{5} According to the Linear Projection Theorem, $|x - x^k| = \min_{y \in L^k} |x - y|$, where $|\cdot|$ is the norm of $\mathcal{L}$ [9].

^{6} In this section we draw heavily on the literature on category theory, even if the results are elementary. Although almost self-contained, this analysis uses notions that can be found in [10], [11], and [12].

^{7} $\dim(\cdot)$ yields the dimension of a subspace of $\mathcal{L}$.

^{8} The objects of $\mathcal{P}\mathbf{R}^{op}$ are the same of $\mathcal{P}\mathbf{R}$, while for each morphism $\rho_{kj} : s^k \to s^j$ in $\mathcal{P}\mathbf{R}$ there is a $\rho_{kj}^{op} : s^j \to s^k$ in $\mathcal{P}\mathbf{R}^{op}$.

^{9} By *simplest* we mean that it can be defined piecewise by polynomials of the lowest possible degree

^{10} This relates our approach to the literature on *splines*

^{11} From now on, we assume that γ yields a minimal intersection among subspaces.

^{12} Here we lift the restriction of concavity of the utility functions.

^{13} A Gröbner basis is a finite set of elements of the submodule that allows to express any element in it [13].

^{14} This obtains by choosing $p^j(x) = 1$ for every $x \in \mathcal{L}$ and every $j \in J$.

References

1. △Ok E (2007). ***Real Analysis with Economic Applications***. *Princeton University Press, Princeton NJ.*

2. △*Prenter P (1969). "A Weierstrass Theorem for Normed Linear Spaces". Bulletin of the American Mathematical Society. 860–862.*

3. △*Romer S (1992). Advanced Linear Algebra. Springer–Verlag, Berlin.*

4. △*Mac Lane S, Birkhoff G (1999). Algebra. 3rd ed. Providence, RI: American Mathematical Society.*

5. △*Cox D, Little J, O'Shea D (2005). **Using Algebraic Geometry**. Springer–Verlag, Berlin.*

6. △*Mohamadi M, Li Z, Wu L, Sutherland D (2024). "Why do you Grok? A Theoretical Analysis of Grokking Modular Addition". arXiv. arXiv:2407.12332.*

7. △*Mas-Colell A, Whinston M, Green J (1995). Microeconomic Theory. **Oxford University Press**, NY.*

8. ^*Gabbay D, Woods J (2005). The Reach of Abduction. Elsevier, Amsterdam.*

9. ^*Luenberger D. **Projection Pricing**. Journal of Optimization Theory and Applications. **109** (2001): 1-25.*

10. ^*Goldblatt R. (1984). **Topoi. The Categorical Analysis of Logic**. North-Holland, Amsterdam.*

11. ^*Barr M, Wells Ch (1999). **Category Theory. Lecture Notes for ESSLLI**. http://www.ling.ohio-state.edu/.*

12. ^*Adámek J, Herrlich H, Strecker G (2006). Abstract and Concrete Categories. The Joy of Cats. Reprints in The ory and Applications of Categories. **17**.*

13. ^*Sturmfels B (2005). "What is a Gröbner Basis?". Notices of the American Mathematical Society. **52** (10): 2– 3.*

Declarations