Towards a New Conceptualization of the Optimization Process: Actor-Network Theory Approach

Loris Serafino

Abstract—One of the most important stages in many areas of engineering and applied sciences is modeling and the use of optimization techniques to increase the quality and performance of products or processes. This paper is an attempt to conceptualize the optimization process at a more abstract level with the aspiration to capture the multidimensional interdependencies and to question issues of cooperation and conflict. It is argued that actor-network theory approach, a discussed line of thought in science studies, offers a new and radical stance for understanding the optimization process and that it is able to grasp its ontological complexity. The starting point is to look at the optimization process as a complex socio-technical practice where material things, human beings, theories as well as representation intermingle in a complex becoming of negotiations and mutual reproduction.

Index Terms—Actor-network theory, evolutionary computing, global optimization, no free lunch theorems.

I. INTRODUCTION

Generally in literature, with the term "optimization" is related to (the output of) a mathematical procedure or algorithm used to identify the extreme value of an arbitrary objective function through the manipulation of a known set of variables and subject to a list of constraining relationships.

More technically, a generic maximization problem with an explicit objective can in general be expressed in the following generic form: to determine the maximum of a function that in what follows will be referred to as the fitness function:

\[ \max_{x \in \Theta} f(x) \] (1)

where \( x \) is a give vector in a generic multidimensional space \( \Theta \) called problem or search space and \( f : \Theta \rightarrow \mathbb{R} \) is a scalar function of the vector \( x \) and \( \Theta \subseteq \mathbb{R}^n \) is a (discrete or continuous) subset of the multidimensional real Euclidean space. Of course it is always possible to transform a minimization problem into a maximization one trough the transformation \( g(x) = -f(x) \). Even if there is no explicit mention of constrains here the formulation is nonetheless enough general since they can be incorporated through an appropriate definition of the search space \( \Theta \).

Global optimization refers to finding the extreme value of a given (however complex but in general non-convex) function in a certain feasible region. The algorithmic solving techniques of (1.1) are the focus of the majority of optimization literature. In this paper we want to look at this issue from a more abstract stance by drawing upon some recent elaborations in the field of Science Studies named actor-network theory. The aim is to conceptualize "optimization" language as both system and process. In this context "optimization" also has a more general usage referring to a system of physical objects (computers, machines) abstract ideas and people combined for discovering and verifying the best implementable solution to high-dimensional design problems in a broad range of engineering and sciences applications. In this more general sense, optimization can be conceptualized as a (complex) system composed with different sub-systems is a multiphase process, in which applying a particular algorithm or technique is just one step. The success or the failure of the optimization process can depend on a single step (right implementation of a algorithm e. g. the “maximization-algorithm application step”) or due to a particular constellation of the different parts at different levels of organization (fig. 1).

The diffusion of new generations of high processing power hardware and its influence in engineering design and optimization continues to expand. The use of computational simulations as an integral part of the modeling process is very common nowadays and the outputs often are visual and highly interactive. The optimization phase is strictly embedded with model simulations that assist in evaluation of the design’s real-time behavior. In recent years there has been a growing interest in looking at the engineering design process not only from a technical prospective but from a broader standpoint as a complex social process [1]. The design process cannot be considered as the activity of an isolated individual but a very large and important multi-disciplinary area strictly connected with computational modeling and optimization [2].

Optimization, as a crucial step in engineering design process, involves the coordination of resources, goals and various forms of knowledge like mathematics, computer science, operational research and statistics. In this sense it is better to talk about an “optimization process” as a practice that can be characterized as a complex, multi-actor with an emergent nature set of procedures. In order to get new insight about how the optimization practice works successfully or fails, how it is performed, and how it is intergraded in broader organizational processes both in industry as well as in the...
academia, it can be fruitful to invest in crossing theoretical borders and to capitalize on progress made in other disciplines and fields of studies [3]-[6].

In order to enable the development of new direction and new fields of research, this paper tries to develop a new attitude on the optimization practice by applying actor-network theory to this field.

II. HARD GLOBAL OPTIMIZATION PROBLEMS

A. The Technical challenge

Abstraction involves the conceptual isolation of (a partial aspect of) an object. Let’s start considering the standard technical discourse. Usually one of the main reasons of failure in solving global optimization problems is that they can easily be entrapped in local minima and in general. Classical calculus-based optimization techniques have obvious difficulties in these cases and today the spectrum of methods is actually vast and increasing day by day. Just to name a few: genetic algorithms, simulated annealing, taboo search, particle swarm, ant colony optimization, cross-entropy, etc. Optimization solver strategies don’t exist in isolation but they represent one side of a coin where the other side is represented by a system (a machine, an antenna, etc.) whose parameters must be arranged in a way that the system can manifest optimal performance. In turn the fitness can be the regarded as a mathematical model describing the output behavior of the system. Knowledge of the fitness landscape structure - i.e. the (hyper)surface obtained by applying ideally the fitness function to every point in the search space over which search is being executed - is usually the key in developing effective algorithms, and consequently it has been a primary focus in the theoretical analysis [7].

In practice, the complexity of a hard global optimization problem can be ascribed to the following basic causes [8]

High dimensionality. The large number of candidate solutions to an optimization problem makes it computationally very hard to be attacked by evolutionary algorithms because the number of candidate solutions grows exponentially with increasing dimensionality. This fact, which is frequently named the **curse of dimensionality**, is well known by practitioners that have to handle problems with hundreds of variables. This phenomenon can be easily understood by first considering an $n$-dimensional binary search space. Here, adding another dimension to the problem means a doubling of the number of candidate solutions. So in order to obtain reliable optimization result the amount of data required to be extracted from the search space will grow exponentially with the dimensionality. The way to overcome this limitation is one of the most intriguing theoretical and practical areas of research at the intersection of mathematics, statistics and computer science that has already produced a vast literature (for a statistical point of view see [9]).

Very complex or irregular fitness landscape. In landscape surface with weak (low) causality, small changes in the candidate solutions often lead to large changes in the objective values, i.e. ruggedness [10]. Stated informally, a landscape is rugged if there are many local optima of highly varying fitness concentrated in any constrained region of the space. It then becomes harder to decide which region of the problem space $\Theta$ to explore and the optimizer cannot find reliable gradient information to follow. A small modification of a very bad candidate solution may then lead to a new local optimum and the best candidate solution currently known may be surrounded by points that are inferior to all other tested individuals.

Fitness evaluation. Evaluating a solution in the objective space can be the most computationally expensive step of any optimization process for difficult or large-size optimization problems. Finding the optimal solution to complex high dimensional, multimodal problems often requires very expensive fitness function evaluations. For most evolutionary algorithms, a large number of fitness evaluations (performance calculations) are needed before a well acceptable solution can be found [11]. In this case, it may be necessary to forgo an exact evaluation and use an approximated fitness (also called metamodel or surrogate fitness) that is computationally more efficient. Functional surrogate models are in practice algebraic representations of the true problem functions. The most popular ones are polynomials (often known as response surface methodology), Interpolation and regression polynomial techniques can be classified in this category. Other several models are now commonly used for fitness approximation, like the Kriging model, neural networks, including multi-layer perceptrons, radial-basis-function networks and the support vector machines. One can say that functional models are typically based on the following components: a class of basis functions, a procedure for sampling the true functions, a regression or fitting criterion, and some deterministic or stochastic mathematical technique to combine them all. For a comprehensive review see [12] and [13].

These are very difficult challenges for the optimization task, no special tricks exist which can directly mitigate for example the effects of rugged fitness landscapes [10].

We can’t forget that a number of design projects are modeled as multi-objective problems and this makes the optimization task harder. Multi-objective optimization considers vector-valued objective functions, where each component corresponds to one of the different simultaneous and competing objectives (e.g., weight, shape, cost). In his case the solution of the optimization problem is represented by a set whose elements have the following property: moving from one element to another we can't improve on a particular objective without detriment to one or more other objectives. Several algorithms (like the well-known NSGA-II) are available today with the property of determining sets of trade-off solutions that give an overview of the space of possible solutions [14].
In general an important preliminary step is how to get a better understanding of the objective space. There are several traditional ways to categorize the fitness function $f$ according to some properties like continuity, geometry, symmetries, multi-modality, ruggedness, etc. The knowledge of these properties can give insights about the best algorithm to use or appropriately refining the region of interest to be sampled in the search space. In this sense we can picture the optimization process as a self-adjusting mechanism. In a given application context can be useful to start with class of optimization algorithms and testing them with the problem at hand. The information collected about the problem and the response obtained from algorithms can allow updating both the parameters configuration of the single algorithm, the algorithm used and the overall optimization strategy itself.

The “optimization-as-adaptive-process” framework means basically that maximization operations and knowledge about a give problem must precede hand-in-hand, like two sides of a coin, and that the information from one side can help improvements in the other.

III. OPTIMIZATION AS ACTOR-NETWORK

The basic tenets of Actor-Network Theory

In the rest of this paper I will divert from these technical issues and I will try grasp insights about the optimization process at a higher abstract and organizational level using Actor-Network Theory (ANT) and to point out relatively unexplored research potential. ANT should be viewed as part of a wider “practice turn” within different disciplines. Practice-based approaches to the optimization process place an emphasis on understanding the practices through which the process emerges out of the unpredictable, embodied, and materially mediated, lifeworlds, of practitioners themselves, together with representation of “best practice” ideals, abstractions, theorems and disciplinary authority.

The push to gain conceptual distance from an image of the optimization process as an automatic machine-like algorithm application and to put value on the unpredictable emergence of social and material interactions that influence outcomes. Emerging during the mid-1980s, ANT was a body of knowledge situated within the sociology of science and technology. The literature on ANT contains many concepts but their interpretation and application vary and reviews of ANT are numerous. Canonical starting points can be [15, 16, 17]. Here we will give a brief elementary introduction to the basic elements of the framework. The starting point of ANT conceptualization is the semiotic insight that ‘entities take their form and acquire their attributes as a result of their relations with other entities’ [4]. In semiotics, those entities are signs, but in ANT they can be anything as we will see. That wider set of relations is called a ‘network’ (translation of the original French word réseau). A network is often described in terms of nodes and links. In ANT the nodes are called actants, any entity that interacts with other actors or serves as an intermediary between actors and that performs moulding locally and/or globally the configuration of the network.

In ANT actants can be real, concrete entities but also abstract objects like ideas, representation etc. Summarizing actants can be: i) texts, theories, ideas: especially how network associations are strengthened through citations and references; ii) technical artifacts; iii) human beings: and the relationships derived from position, skills, authority and responsibility. In the vocabulary of ANT, they are ‘enrolled’ into a network through a process called “translation”. Translation involves bringing together seemingly quite different entities, for example, Darwin’s theory of evolution and optimization techniques.

Let’s consider the example of the evolutionary computing to clarify this point. The basic idea of the evolution theory is that the individuals with a greater “fitness to the environment” have a greater probability of surviving and a greater probability of winning the fights for mating. In such a way the genetic content of the best individuals will be more and more present in the following generations, since it will be transmitted by the offspring. This has inspired the development of genetic algorithms, proposed by Holland in the 1960s. Nowadays Biology-inspired algorithms play an important part of computational sciences since they are applicable to a wide variety of optimization problems: discrete, continuous, or even mixed parameters without any a priori assumptions about their continuity and differentiability [18].

In terms of ANT this translation of a domain into another is resulted in a specific assemblage of networks evolution that was not only unpredictable, but truly creative, with the appearance of a new form of dialog between the technologists and life scientists that ultimately transformed (and will continue to affect) the optimization practice. From an ANT standpoint, the wide diffusion of this or that optimization methods at the expense of others is not only and necessarily a matter of their intrinsic properties, but the results of their subsequent translations and incorporations in powerful external networks. Following the main actors, and tracking associations and translations it is possible to decrypt how ‘artifacts’ – materials as well as pieces of embodied knowledge – are translated into ‘facts’ and successively into ‘black boxes’, which, at least for a while, ‘act as one’ [3].

From an established terminology, biology-inspired algorithms and the more general nature-inspired algorithms (NIAs) are classes of heuristic algorithms based on populations of objects. Even if they still lack a strong theoretical foundation able to understand or predict its behavior other than a superficial level, NIAs are conceptually simple and it is relatively easy to develop program code for any kind of application. This can partly explain the speed of practical and empirical diffusion of these techniques virtually in every area of applied sciences far outside the field of computer sciences. At the same time in the academic environment the “publish or perish” mantra seems accountable for the proliferation of new variants of NIAs based on the phenomenology of some natural phenomena at the point that nowadays new papers on the topic are countless. How the actor-network temporary coagulate and rationalized this social fact into a meaningful mainframe will be briefly considered in the following section.
IV. PRACTICES AND CONFLICTS IN OPTIMIZATION

The standpoint trough which we can to grab hold of a better understanding of the optimization process from an ANT perspective stems from some initial premise about the ontological status of an actor-network. Actor and networks mutually produce each other’s: that’s why we can’t separate them and we talk about an irreducible actor-network. Objects and entities – any actant we can conceive - in themselves possess no essential meaning. Meaning that springs out is always an effect or an outcome of the wider system of relations of which an entity is part. An example can be fruitful to explore this; a particular a multiobjective genetic algorithm code on one given problem can be used for determine a Pareto optimal set or as a pre-optimization tool to explore the surface response of a given mono-objective fitness function. The role of the actant here (for example an implementation of the NSGA-II algorithm) in the network can acquire very different meanings, be more or less useful. So different set of parameters can be used and the role and performance of the algorithm will be different. In the optimization actor-network a variety of entities are enrolled as, and mobilized as, actants into the optimization process: algorithms, theoretical approaches, computer experts, statistical packages, etc. (fig. 2).

Even the apparently abstract fact of framing a problem in optimization terms is not innocent at all. The kind of entities and actors that are enrolled or not enrolled into the network and how they are enrolled is a characteristic of the context, where “context” here must be interpreted in a broader way as the situated character of social reality, involving coexistence, connections and ‘togetherness’ as a series of associations and entanglements in time–space (for a general discussion see [19]). For example in the case of industrial product design optimization the reason for choosing a given class of optimization strategies (a genetic algorithm instead of particle swarm a Monte Carlo based method for example) can be connected to the local availability of expert knowledge, constrained to the utilization of a given machine or proprietary software or imposed from some bias internal to a given epistemic community.

Nowadays there are a number of resources freely available on Internet and it is common nowadays to adapt pre-existing codes instead of creating them ex-novo for time saving reasons and this in turn is connected to higher level organizational settings (availability of time and resources) that together can affect the final outcome. It is also important actor-networks themselves, e.g. in the shape of project management practice, external companies pressure, etc. organizing their own heterogeneous complexity. There will therefore be an unavoidable tension in the interactions between the actor-networks.

As it has been told, in industrial design the optimization activity is just a part, a sub-module, of a broad modeling process that is in itself a sub-module of the product design process. In this sense more actor-networks enter in a complex process of interaction/negotiations. An artifact like computer design software is produced from a certain set of ideas of how the design process is organized inside the company and can include or not an optimization module that can constrain to a given extent the possibility of the users in a way that can be unpredictable and contingent. Here, the notion of contingency is central to ANT. The accomplishment of a certain actor-network in a given configuration for a given time is always just one among (infinitely) many possible outcomes. Contingency then means that actor-networks emerge out of delocalized choices, there is no higher-level plan prescribing the mobilization of the network and there is no platform for making these choices rationally, because the networks self-establish its own schema of rationality.

Social practice, identities and non-human actants

An ANT stance forces a kind of analysis that is performance centered. Performance is looking at reality “in the making”. It is connected to the coming true of social meaningful routines goings-on through active human and non-human expression. Methodologically, field observations, ethno-methodological and text analytical approaches are commonly employed in the decryption of the actor-network constellation. Performance not only involves a process of the relational construction of identity; the term is widened to include all sorts of practices that are involved in the human project of (re)creating optimization actions in conditions that can be defined as negotiations. Identity, e.g. the institutional role of human actor involved, is one important facet in this relational interaction.

For example the depiction of the optimization task from the point of view of an engineer can be different from the point of view of a computer scientist or a mathematician. The former can be more interested on the computational efficiency response whereas the others on a deeper understanding of the gap between theoretical underpinnings and actual outcomes of a given optimization approach. A statistician would assume a different approach too, focusing primary on how to acquire most information possible from sampled data and only subsequently thinking about the optimization problem. Those representations are continuously negotiated and a deep analysis of this can shed light on the successes or failures of the overall process. At the same time abstract entities can exert their agency too in a way that can remodel widely the actor-network.

According to the most common understanding of NFLT there is no optimization method superior to others for all possible optimization problems. Moreover, an algorithm that performs well on one class of problems must perform worse than random search on all remaining problems. Running an algorithm on a small number of problems with a small range of parameter settings will not be a good predictor of that
algorithm’s performance on other problems, especially problems of a different typology. This theoretical result represents now an obligatory passage point [6] for every expert in the field that has to cope with and the interpretation of them and they have pervasive impact on routinely performance of the optimization process. According to some influential authors it is possible to derive the practical implication that produce actively and diffusely a schema of rationality for the daily work of optimization. For example Yang ([21], p. 26) wrote:

“Even though, the NFL theorems are valid mathematically, their influence on parameter search and optimization is limited. For any specific set of functions, some algorithms do perform much better than others. In fact, for any specific problem with specific functions, there usually exist some algorithms that are more efficient than others if we do not need to measure their average performance. The main problem is probably how to find these better algorithms for a given particular type of problem1. […] The knowledge about the particular problem of interest is always helpful for the appropriate choice of the best or most efficient methods for the optimization procedure.”

Similarly Wise wrote ([22], p. 213):

“The No Free Lunch Theorem argues that it is not possible to develop the one optimization algorithm […] this must sound very depressing […] Actually, quite the opposite is the case, at least from the point of view of a researcher. The No Free Lunch Theorem means that there will always be new ideas, new approaches which will lead to better optimization algorithms to solve a given problem. Instead of being doomed to obsolescence, it is far more likely that most of the currently known optimization methods have at least one niche, one area where they are excellent. […] There will always be a chance that an inspiring moment, an observation in nature, for instance, may lead to the invention of a new optimization algorithm which performs better in some problem areas than all currently known ones.”

So according to this interpretation, far from representing a limiting result, the NFL theorems offer one way to explain the adoption or diffusion of some techniques at the expense of other for a given class of problems and this embodies the theoretical rationale that fosters the never-ending research of improvement and new algorithms and whose agency – although with contextual negotiations variants - inform the everyday practice in industry and in the academic research. At the same time from the practitioner point of view it makes the question “which optimization method is the best for our problem?” as the central question in the application context.

V. CONCLUSIONS AND DISCUSSION

This paper was intended to be a cross-disciplinary effort in order to learn more about how optimization practice works from a broader perspective, how it is performed, and how it produces outcomes and spaces. It is just a starting point in order to enable the development of new approaches and enlarge the way in which we usually look at the optimization from a different viewpoint. Scientists, researchers, engineers, and other practitioners, continuously everyday co-operate within communities of people and collection of things, ideas, theorems and techniques, building more or less coherent actor-networks that together re-create the optimization phenomenon and its effectiveness. An effectiveness that even rests, it is worth to remember, on an imponderable, unquantifiable, non-mercantile organizational milieu.

To discover how they face and try to overcome resistance, how they try to conceal, define, hold in place, mobilize, and bring into play the juxtaposed people and things in the optimization practice, the researcher has to penetrate in the web of actions, routines and taken-for-granted assumptions that produce a meaningful whole. As the design process become increasingly complex, understanding critical areas of conflict, tension and discomfort has become more and more important in engineering practice. The process of orchestration between internal and external actor-networks create a contextual environment together with a body of tacit knowledge and a sort of “path dependency” that can make hard to operate organizational changes in the future.

In practice, dealing with complexity of the optimization practice calls for new approaches, which form an important part of the ANT paradigm we have described here.

ANT radicalize the idea of a network ontology of social facts. At the same time, it stresses the ability of each entity, in particular nonhuman ones, to act and interact in a specific way with other humans or nonhumans (like engineers and algorithms, computers, theorems). This provided a theoretical orientation for analyzing how the optimization practice reality is constructed via a dynamic socio-technical-representational process where, given an organizational context, multiple fluid subjective realities are possible. This paradigm should be considered just a starting point to think about an organizational model from the recognition that there are differing contextual actor-network arrangements, interacting in differing ways with varied sets of pre-existing conditions. This implies that organizational strategies adopted will need to be adjusted to the requirements of different context. ANT must be taught as reflexive tool that helps to comprehend the complex dynamics of optimization as embedded in broader design projects and it can enable practitioners to be aware of emergent properties, unintended consequences and unpredictable behavior, through the consideration of networks of non-human and human actors.

REFERENCES


Loris Serafino is Corporation Research Fellow at Kuang-Chi Institute of Advanced Technology in Shenzhen, China. Before reaching China he has had academic experiences as lecturer in Italian, African and Pakistan universities. He was born in November 1971 in Udine, a small town in the North-East of Italy. He holds a Master degree in Mathematics from University of Udine and an interdisciplinary Ph.D. degree from the same University. Her areas of interest are multi-objective optimization, complex systems and applied statistics. In the past he has been involved in several interdisciplinary research projects including the development of simulation tools for the management and environmental impact of bio-fuels and bio-energy and the application of critical social theory approaches in the analysis of the regional planning process. Dr. Serafino is the author of several research works and book chapters.