Suboptimal solutions to network team optimization problems

Giorgio Gnecco[⋆,∗] Marcello Sanguineti[∗]

[∗]*Department of Communications, Computer, and System Sciences (DIST) - University of Genova*
Via Opera Pia 13, 16145 Genova, Italy
giorgio.gnecco@dist.unige.it, marcello@dist.unige.it

[⋆]*Department of Mathematics (DIMA) - University of Genova*
Via Dodecaneso, 35, 16146 Genova, Italy

Abstract

Smoothness of the solutions to network team optimization problems with statistical information structure is investigated. Suboptimal solutions expressed as linear combinations of elements from sets of basis functions containing adjustable parameters are considered. Estimates of their accuracy are derived, for basis functions represented by sinusoids with variable frequencies and phases and Gaussians with variable centers and widths.

Keywords: *Statistical information structure, social utility, value of the team, suboptimal solutions, nonlinear approximation schemes.*

1 Introduction

Agents cooperating to achieve a common goal model a variety of problems in engineering, economic systems, management science and operations research, in which distributed optimization processes have to be performed. In this paper, we consider team decision problems defined on a *network of decision makers*, where each *decision maker* (*DM*) cooperates to maximize a *team utility function* (also called "social utility"). In the model that we adopt, each DM has at its disposal a probabilistic *information* (obtained, e.g., by measurement devices) and various possibilities of *actions*. Decisions are generated by the DMs via *strategies*, on the basis of the information available to each of them and in the presence of uncertainties in the external world, not controlled by the DMs.

Sometimes, a team utility function is present from the beginning in the formulation of the network optimization problem. For example, a situation with a natural team formulation is represented by routing in packet-switching telecommunication networks. According to the model proposed in [28] and used in subsequent works (see, e.g., [2, 4, 14]), the DMs are the routers at each node, which choose the actions on the basis of their routing strategies. The DMs do not share a common information on the state of the network (represented, e.g., by the length of the packet queues at the nodes and the delays in the links), but share the common goal of minimizing the total time spent by the messages at the nodes and on the communication links.

In other cases, *a priori* each DM has an individual utility function, so, if one assumes that each DM aims to maximize its own utility with a "selfish behavior", the natural framework is provided by noncooperative game theory and Nash equilibria [22]. Instead, in a team reformulation of the problem, each DM cooperates to maximize a common goal given, e.g., by the sum of the individual utilities. In this context, studying a team problem instead of a noncooperative one may be motivated by considering the *price of anarchy* [23]. Roughly speaking, the price of anarchy measures the loss in the social utility when the DMs play the "worst" Nash equilibrium, compared to the case in which they maximize a common utility, i.e., when they play as members of a team. In some cases, the price of anarchy is very large, especially in the presence of smooth individual utility functions [19, Lecture 3]. For example, this may happen when the individual utilities are related to the congestion of a communication link shared among some DMs. Another reason for which a team reformulation of a problem is sometimes well motivated arises from the so-called "Braess's paradox" [19, Lecture 3] in game theory. Braess's paradox shows that adding a resource (e.g., one communication link) to a network of DMs playing a Nash equilibrium, may even decrease the social utility. This does not happen if the DMs behave as members of a team.

In general, one centralized DM that, relying on the whole available information, maximizes the given common goal, provides a better performance than a set of decentralized DMs, each of them having partial information. However, often centralization is not feasible for various reasons. For example, this happens when the DMs have access to local information and they cannot exchange instantaneously such information. Moreover, in some situations the cost of making the whole information available to one single DM is too high with respect to having several DMs with different available information. This framework models, e.g., communication and computer networks extending in large geographical areas, production plants, energy distribution, traffic systems in large metropolitan areas divided into sectors, freeway systems, etc.

In the networks of DMs that we address, the information of each DM depends on the state of the world but is independent of the actions of the other DMs. This is called a *static team*, in contrast to a *dynamic team*, where each DM's information is affected by the actions of the other members. Static teams were first investigated by Marschak and Radner [20, 21, 26], who derived closed-form solutions for some cases of interests. Then, dynamic teams were studied [6].

Unfortunately, closed-form solutions can be derived only under quite strong assumptions on the team utility function and on the way in which each DM's information is influenced by the state of the world (and, in the case of dynamic teams, by the actions previously taken by the other DMs). In particular, most results hold under the so-called *LQG hypotheses* (i.e., linear information structure, concave quadratic utility, and Gaussian random variables) and with *partially nested information* (i.e, when each DM can reconstruct all the information available to the DMs that affect its own information). In general, one has to search for suboptimal solutions.

In this paper, for static team problems we derive a-priori estimates of the accuracy of suboptimal solutions having the form of linear combinations of k elements from a set of basis functions, containing some adjustable parameters to be optimized. Such approximation scheme, known as *variable-basis approximation* [15, 16], includes free-node splines, trigonometric polynomials with free frequencies and phases, radial-basis-function networks with adjustable centers and widths, and feedforward neural networks. The numerical results in [2, 3, 4] show that these approximators are able to find accurate suboptimal solutions to team optimization problems with high-dimensional states. The present work complements such experimental outcomes providing theoretical results supporting the use of variable-basis approximation schemes in team optimization problems. The analysis is made for static team problems, however, in [31] it was shown that many dynamic team optimization problems can be reformulated in terms of equivalent static ones.

In our model, we have a *statistical information structure*, i.e., the information available to each DM is expressed via a probability density function. We provide conditions guaranteeing that optimal solutions exist and have a certain degree of smoothness. Exploiting these results, we investigate suboptimal solutions obtained by restricting the search to k-term variable-basis approximators with sinusoidal or Gaussian basis functions. We estimate the accuracy of such suboptimal solutions in terms of the difference between the *value of the team*, i.e., the value of the team utility function when optimal strategies are used, and its value when the strategies used are the optimal ones among restricted families of variable-basis approximators with k basis functions. The upper bounds that we derive are proportional to $k^{-1/2}$.

The paper is organized as follows. Section 2 introduces definitions and assumptions and formulates the team optimization problem (Problem P). Section 3 investigates smoothness properties of optimal strategies for Problem P. Section 4 describes nonlinear approximation of the optimal strategies from sets made up of variable-basis functions and estimates the accuracy of such suboptimal solutions. Section 5 contains some conclusions and discusses possible extensions. Due to lack of space, some proofs are only sketched; the details can be found in [10, 12].

2 Problem formulation

The context in which we shall formalize the optimization problem and derive our results is the following.

- *Team* of n *decision makers* (*DMs*), $i = 1, \dots, n$.
- $x \in X \subseteq \mathbb{R}^d$: random variable, called *state of the world*, with a probability density $p : X \to \mathbb{R}$ describing a stochastic environment. The vector x models the uncertainties in the external world, which are not controlled by the DMs.
- $y_i = f_i(x) \in Y_i \subseteq \mathbb{R}^{d_i}$: *information* that the DM i has about x, which is a given function of the state of the world.
- $s_i : Y_i \to A_i \subseteq \mathbb{R}$: measurable *strategy* of the i-th DM.
- $a_i = s_i(y_i)$: *action* that the DM i chooses on the basis of the information y_i.
- $u : X \times \Pi_{i=1}^n Y_i \times \Pi_{i=1}^n A_i \to \mathbb{R}$: real-valued *team utility* function, to be jointly maximized by the team.
- The information that the n DMs have on the state of the world x is modelled by an n-tuple of random variables $y_1, \dots, y_n$, i.e., one has a *statistical information structure* [7] represented by a probability density $q(x, y_1, \dots, y_n)$ on the set $X \times \Pi_{i=1}^n Y_i$, whose marginal density on X is equal to the density $p(x)$ of the state of the world.

The family of static team optimization problems that we consider is formalized as follows.
Problem P (Static Team Optimization with Statistical Information). *Given the statistical information structure $q(x, y_1, \dots, y_n)$ with marginal density on X equal to $p(x)$, the utility function $u(x, y_1, \dots, y_n, s_1(y_1), \dots, s_n(y_n))$, $u : \mathbb{R}^N \to \mathbb{R}$, where $N = d + \sum_{i=1}^n d_i + n$, find*

$$\sup_{s_1, \dots s_n} v(s_1, \dots, s_n),$$

where

$$v(s_1, \dots, s_n) := \mathop{E}_{x, y_1, \dots, y_n} \{u(x, \{y_i\}_{i=1}^n, \{s_i(y_i)\}_{i=1}^n)\}.$$

The quantity $\sup_{s_1, \dots s_n} v(s_1, \dots, s_n)$ is also called the *value of the team*.

By $\mathcal{C}(\Omega)$ we denote the space of continuous functions on Ω; for a positive integer $m > 0$, by $\mathcal{C}^m(\Omega)$ and $\mathcal{C}^\infty(\Omega)$ we denote the spaces of functions on Ω, which are continuous together with their partial derivatives up to the order m and up to every order, respectively. $\mathcal{C}_0(\Omega)$ and $\mathcal{C}_0^\infty(\Omega)$ are the spaces of those functions in $\mathcal{C}(\Omega)$ and $\mathcal{C}^\infty(\Omega)$, respectively, which have compact support in Ω [1, p. 9].

Throughout the paper, we make the following assumptions A1 and A2.
A1 *The set X of the states of the world is compact, $Y_1, \dots, Y_n$ are compact and convex, $A_1, \dots, A_n$ are bounded closed intervals. For an integer $m \geq 2$, the utility u is of class $\mathcal{C}^m$ on an open set containing $X \times \Pi_{i=1}^n Y_i \times \Pi_{i=1}^n A_i$, and q a (strictly) positive probability density on $X \times \Pi_{i=1}^n Y_i$, which can be extended to a function of class $\mathcal{C}^m$ on an open set containing $X \times \Pi_{i=1}^n Y_i$.*

A concave function f defined on a convex set Ω has *concavity at least* $\tau > 0$ if for all $u, v \in \Omega$ and every supergradient[1] a_u of f at u one has $f(v) - f(u) \leq a_u \cdot (v - u) - \tau \|v - u\|^2$.

[1] Recall that, for $\Omega \subseteq \mathbb{R}^d$ convex and $f : \Omega \to \mathbb{R}$ concave, $a_u \in \mathbb{R}^d$ is a *supergradient* of f at $u \in \Omega$ if for every $v \in \Omega$ it satisfies $f(v) - f(u) \leq a_u \cdot (v - u)$.

A2 *There exists $\tau > 0$ such that the team utility function $u : \mathbb{R}^N \to \mathbb{R}$ is separately concave in each of the decision variables, with concavity at least τ.*

Concavity of the team utility function is motivated by tractability reasons and is often encountered in practice, too. For example, in economic problems it is motivated by the "law of diminishing returns", i.e., the fact that the marginal productivity of an input usually diminishes as the amount of output increases [21, p. 99 and p. 110].

3 Smooth optimal strategies

In this section, we give conditions that guarantee existence and smoothness of optimal strategies for Problem P. We shall exploit them in Section 4, to estimate the accuracy of certain suboptimal solutions.

The next theorem (which extends, to a higher degree of smoothness, a result from [17]) gives conditions guaranteeing that Problem P has a solution made of an n-tuple of strategies that are Lipschitz continuous[2] together with their partial derivatives up to a certain order.

Theorem 3.1 *Let Assumptions A1 and A2 hold. If for every n-tuple $\{s_1, \dots, s_n\}$ of strategies, the strategies defined as*

$$\hat{s}_1(y_1) := \underset{a_1}{\operatorname{argmax}} \underset{x, y_2, \dots, y_n \,|y_1}{E} \{u(x, \{y_i\}_{i=1}^n, a_1, \{s_i(y_i)\}_{i=2}^n)\},$$
$$\dots$$
$$\hat{s}_n(y_n) := \underset{a_n}{\operatorname{argmax}} \underset{x, y_1, \dots, y_{n-1} \,|y_n}{E} \{u(x, \{y_i\}_{i=1}^n, \{s_i(y_i)\}_{i=1}^{n-1}, a_n)\}$$

do not lie on the boundaries of $A_1, \dots, A_n$, respectively, then Problem P admits $\mathcal{C}^{m-2}$ optimal strategies $(s_1^o, \dots, s_n^o)$ with partial derivatives that are Lipschitz up to the order $m-2$.

Note that the condition that $\hat{s}_1(y_1), \dots, \hat{s}_n(y_n)$ do not lie on the boundaries of $A_1, \dots, A_n$, respectively, can be imposed *a priori* by strongly penalizing the utility on the boundary. For limitations of space, we sketch the proof of Theorem 3.1 for $n = 2$. For detailed arguments, see [10, 12].

Sketch of proof. Consider a sequence $\{s_1^j, s_2^j\}$ of pairs of strategies, indexed by $j \in \mathbb{N}_+$, such that $\lim_{j\to\infty} v(s_1^j, s_2^j) = \sup_{s_1, s_2} v(s_1, s_2)$ (such a sequence exists by the definition of supremum). From this sequence, we generate another sequence $\{\hat{s}_1^j, \hat{s}_2^j\}$ defined as $\hat{s}_1^j(y_1) = \operatorname{argmax}_{a_1} E_{x, y_2 \,|y_1} \{u(x, y_1, y_2, a_1, s_2^j(y_2)\}$, $\hat{s}_2^j(y_2) = \operatorname{argmax}_{a_2} E_{x, y_1 \,|y_2} \{u(x, y_1, y_2, \hat{s}_1^j(y_1), a_2)\}$. By some technical steps, it can be proved that that for every $j \in \mathbb{N}_+$, $\hat{s}_1^j$ and $\hat{s}_2^j$ are well-defined (i.e., the argmax are singletons) and measurable so it makes sense to evaluate $v(\hat{s}_1^j, \hat{s}_2^j)$ (by construction, $v(\hat{s}_1^j, \hat{s}_2^j) \geq v(s_1^j, s_2^j)$), and Lipschitz with a constant independent of j, equibounded and uniformly equicontinuous.

Let us focus on the strategy of the first DM. The next step consists in showing that there exists a subsequence of $\{\hat{s}_1^j\}$ that converges uniformly to a strategy $s_1^o \in \mathcal{C}^{m-2}(Y_1)$ with Lipschitz $(m-2)$-order partial derivatives. Let $M_1^j(y_1, a_1) = E_{x, y_2 | y_1} \{u(x, y_1, y_2, a_1, s_2^j(y_2)\}$. As by hypothesis $\hat{s}_1^j(y_1)$ is interior, for every $y_1 \in Y_1$ we have $\frac{\partial M_1^j}{\partial a_1}\Big|_{a_1 = \hat{s}_1^j(y_1)} = 0$. Then, by the Implicit Function Theorem, for every $k = 1, \dots, d_1$ we get $\frac{\partial \hat{s}_1^j}{\partial y_{1,k}} = -\left(\frac{\partial^2 M_1^j}{\partial {\hat{s}_1^j}^2}\right)^{-1} \frac{\partial^2 M_1^j}{\partial \hat{s}_1^j \partial y_{1,k}}$, where one can show that $\left(\frac{\partial^2 M_1^j}{\partial {\hat{s}_1^j}^2}\right) \leq -\tau < 0$.

By some technical steps and exploiting the Ascoli-Arzelà Theorem[3] [1, Theorem 1.30, p. 10] one can show that for every $(i_1, \dots, i_{d_1})$ such that $i_1 + \dots + i_{d_1} = m-2$, the sequence $\left\{ \frac{\partial^{m-2} \hat{s}_1^j}{\partial y_{1,1}^{i_1}, \dots, \partial y_{1,d_1}^{i_{d_1}}} \right\}$ admits a subsequence that converges uniformly to a Lipschitz function on Y_1.

By integrating $m-2$ times, we conclude that also the integrals of these subsequences converge uniformly to the integrals of the limit functions. Therefore, there exists a subsequence of $\{\hat{s}_1^j\}$ that converges

[2]Recall that a function $f : \Omega \to \mathbb{R}$ is said to be *Lipschitz continuous on* Ω if there exists a constant $M > 0$ such that for all $u, v \in \Omega$, $|f(u) - f(v)| \leq M \, \|u - v\|$.

[3]For a compact set $\Omega \subset \mathbb{R}^d$, Ascoli-Arzelà's theorem states that if a set $\mathcal{F} \subset \mathcal{C}(\Omega)$ is closed, equibounded, and uniformly equicontinuous, then it is a compact subset of $\mathcal{C}(\Omega)$.

uniformly to a strategy $s_1^o \in \mathcal{C}^{m-2}(Y_1)$ with Lipschitz $(m-2)$-order partial derivatives. Similarly, one proves that there exists a subsequence of $\{\hat{s}_2^j\}$ that converges uniformly to $s_2^o \in \mathcal{C}^{m-2}(Y_1)$ with Lipschitz $(m-2)$-order partial derivatives.

By the continuity of the functional $v(s_1, s_2)$ on $\mathcal{C}(Y_1) \times \mathcal{C}(Y_2)$ with the respective maximum norms, finally we obtain $v(s_1^o, s_2^o) = \lim_{j\to\infty} v(\hat{s}_1^j, \hat{s}_2^j) = \sup_{s_1,s_2} v(s_1, s_2)$. ■

4 Accuracy of suboptimal solutions to Problem P

Problem P admits a closed-form solution when, e.g., the team utility function u is a second-degree polynomial with positive definite quadratic term [17, Theorem 4, p. 68]. However, in more general situations only suboptimal solutions can be obtained. One possible way to find them consists in restricting the search to strategies having a simple form.

An approximation scheme that has received considerable attention in recent years consists in using linear combinations $\psi(\cdot, w_1), \ldots, \psi(\cdot, w_k)$ of functions obtained from a "mother function" ψ dependent on vectors $w_1, \ldots, w_k$ of adjustable parameters, to be optimized together with the coefficients $c_1, \ldots c_k$ of the linear combination

$$\sum_{i=1}^{k} c_i \, \psi(x, w_i) \,. \tag{1}$$

In general, the presence of the "inner" parameters $w_1, \ldots, w_k$ "destroys" linearity of the approximation scheme, so (1) is a nonlinear approximation scheme, which belongs to the class of *variable-basis approximation schemes* [15, 16]. With suitable choices of the function ψ, (1) models a variety of approximating families used in applications, such as free-node splines, trigonometric polynomials with free frequencies and phases, radial-basis-function networks with adjustable centers and widths, and feedforward neural networks. Advantages of certain variable-basis approximation schemes of the form (1) over classical linear ones were investigated, e.g., in [5, 15] for function approximation and in [9, 30] for functional optimization. Roughly speaking, for a desired accuracy of approximation, variable-basis approximation schemes may require much less parameters to be optimized than linear ones [9, 15, 16, 30].

In the following, we shall exploit the smoothness results obtained in Section 3 and certain properties of variable-basis approximation schemes to derive upper bounds on the distance between the value of the team, i.e., the quantity $\sup_{s_1,\ldots,s_n} v(s_1, \ldots, s_n)$, and the suboptimal value when the supremum is performed over a subset of strategies $\tilde{s}_1, \ldots, \tilde{s}_n$ having the form (1) with a fixed k. In particular, we shall estimate the number k of basis functions sufficient to guarantee a desired approximation accuracy $\varepsilon > 0$ for the value of the team.

As a first step, the following proposition allows one to reduce the optimization Problem P to a function approximation problem.

Proposition 4.1 *Let $u(x, y_1, \ldots, y_n, a_1, \ldots, a_n)$ be Lipschitz with respect to $(a_1, \ldots, a_n)$ with Lipschitz constant L and suppose that Problem P has a solution $(s_1^o, \ldots, s_n^o)$. Then for every positive integer n and every n-tuple $(s_1, \ldots, s_n)$ of strategies,*

$$v(s_1^o, \ldots, s_n^o) - v(s_1, \ldots, s_n) \le L \sum_{i=1}^{n} \sqrt{\mathop{E}_{y_i}\{(s_i^o(y_i) - s_i(y_i))^2\}} \,.$$

Sketch of proof. The proof follows by using the Jensen and Minkowski Inequalities, together with the Lipschitz continuity of u. Details can be found in [10, 12]. ■

According to Proposition 4.1, to obtain a "good approximation" of $\sup_{s_1,\ldots,s_n} v(s_1, \ldots, s_n)$, it is sufficient to have a "good approximation" of $(s_1^o, \ldots, s_n^o)$.

We shall need the following definitions. Given a Lebesgue-measurable set $\Omega \subseteq \mathbb{R}^d$, by $\mathcal{L}_p(\Omega)$ and $\|\cdot\|_{\mathcal{L}_p(\Omega)}$, for $1 \le p \le \infty$, we denote the corresponding Lebesgue space and norm, respectively, where integration is performed with respect to the Lebesgue measure. For $\Omega \subseteq \mathbb{R}^d$ open, a positive integer m and $1 \le p \le \infty$, we denote by $\mathcal{W}^{m,p}(\Omega)$ the Sobolev space of functions whose weak partial derivatives up to the order m are in $\mathcal{L}_p(\Omega)$. Finally, for an open set $\Omega \subseteq \mathbb{R}^d$, $\mathcal{W}_0^{m,p}(\Omega)$ is the closure of $\mathcal{C}_0^\infty(\Omega)$ in the Sobolev space $\mathcal{W}^{m,p}(\Omega)$ (see [1, pp. 44-45]).

The next proposition estimates the accuracy of approximation of the optimal strategies in terms of sinusoids with variable frequencies and phases used in (1). We denote by Prj_{A_i} the projection on A_i (recall that every admissible strategy s_i takes its values on A_i).

Proposition 4.2 *Let the assumptions of Theorem 3.1 hold with* $m > \frac{\max_i\{d_i\}}{2} + 1$. *Then there exists a positive constant* C *(dependent on* $(s_1^o, \ldots, s_n^o)$*), such that for every positive integer* k *there exists an* n*-tuple of strategies* $(\tilde{s}_1^k, \ldots, \tilde{s}_n^k)$ *such that*

$$v(s_1^o, \ldots, s_n^o) - v(\tilde{s}_1^k, \ldots \tilde{s}_n^k) \le \frac{C}{\sqrt{k}},$$

where

$$\tilde{s}_i^k(y_i) = \sum_{j=1}^k c_{ij} \mathrm{Prj}_{A_i}(g_{ij}(y_i)), \; g_{ij} \in \mathcal{G}_i,$$

$$g_{ij} \in G_i,$$

$$G_i := \left\{ g_i : Y_i \to \mathbb{R} \;|\; g_i(y_i) = \Pi_{k=1}^{d_i} \cos(\omega_{i,k} y_{i,k} + \theta_{i,k}), \omega_{i,k} = \frac{2\pi h}{y_{i,k}^u - y_{i,k}^l}, \; h \in \mathbb{N}, \; \theta_{i,k} \in [0, 2\pi) \right\},$$

$$\sum_{j=1}^k |c_{ij}| \le \sum_{j_1, \ldots, j_{d_i}=0}^{\infty} |A_i^{j_1, \ldots, j_{d_i}}|,$$

and

$$\{A_i^{j_1, \ldots, j_{d_i}}\}$$

are the coefficients of the Fourier series expansion of a suitable extension of s_i^o *on a set*

$$(y_{i,1}^l, y_{i,1}^u) \times \ldots \times (y_{i,d_i}^l, y_{i,d_i}^u) \supset Y_i.$$

Proof. Let $1 \le p < \infty$. By Theorem 3.1, $s_i^o \in \mathcal{W}^{m-1,\infty}(\mathrm{int}(Y_i)) \subset \mathcal{W}^{m-1,p}(\mathrm{int}(Y_i))$. As every Y_i is a bounded convex set, by Sobolev's extension theorem (see [29, Theorem 5, p. 181] and [29, Example 2, p. 189]), s_i^o can be extended to a function $s_i^{o,\mathrm{ext},p} \in \mathcal{W}^{m-1,p}(\mathbb{R}^{d_i})$.

Fix a set $Y_i^{\mathrm{per}} := (y_{i,1}^l, y_{i,1}^u) \times \ldots \times (y_{i,d_i}^l, y_{i,d_i}^u) \supset Y_i$ and consider a function $\psi_i \in \mathcal{C}_0^\infty(Y_i^{\mathrm{per}})$ with Lipschitz $(m-2)$-order partial derivatives, such that $\psi_i(y_i) = 1, \forall y_i \in Y_i$. By using [1, Theorem 3.18, p. 54] one can prove that $s_i^{o,\mathrm{per},p} := s_i^{o,\mathrm{ext},p} \cdot \psi_i \in \mathcal{W}_0^{m-1,p}(Y_i^{\mathrm{per}})$ if $1 \le p < \infty$. By the Sobolev Embedding Theorem [1, Theorem 5.4, Part III, Case C', pp. 97-98], if $d_i < \bar{p} < \infty$, then $\mathcal{W}_0^{m-1,\bar{p}}(Y_i^{\mathrm{per}}) \subset \mathcal{C}^{m-2}([y_{i,1}^l, y_{i,1}^u] \times \ldots \times [y_{i,d_i}^l, y_{i,d_i}^u])$. This, together with the boundary conditions on $s_i^{o,\mathrm{per},\bar{p}}$, allows one to apply a result[4] from [24, pp. 81-82], according to which the Fourier series coefficients of $s_i^{o,\mathrm{per},\bar{p}}$ satisfy

$$K_i := \sum_{j_1, \ldots, j_{d_i}=0}^{\infty} |A_i^{j_1, \ldots, j_{d_i}}| < \infty,$$

(i.e., using the notations of [24, pp. 81-82], $s_i^{o,\mathrm{per},\bar{p}} \in A(1)$).

[4]That result states that $\mathrm{Lip}(\alpha) \subset A(s)$ when $s > \frac{2d_i}{2\alpha + d_i}$ (see [24, pp. 81-82] for the notations and the precise definitions of $\mathrm{Lip}(\alpha)$ and $A(s)$). By the construction of $s_i^{o,\mathrm{per},\bar{p}}$ and the definition of $\mathrm{Lip}(\alpha)$, one can easily show that $s_i^{o,\mathrm{per},\bar{p}} \in \mathrm{Lip}(m-2)$ if $d_i < \bar{p} < \infty$. Moreover, the condition $s > \frac{2d_i}{2\alpha + d_i}$ is satisfied with $s = 1$ and $\alpha = m-2$, since by assumption $m > \frac{\max_i\{d_i\}}{2} + 1$, then $s_i^{o,\mathrm{per},\bar{p}} \in A(1)$, as desired.

Hence, $s_i^{o,per,\bar{p}}$ belongs to the closure (with respect to the $\mathcal{L}_2((y_{i,1}^l, y_{i,1}^u) \times \ldots \times (y_{i,d_i}^l, y_{i,d_i}^u))$-norm) of the convex hull of the set

$$G_i^{\text{per}} := \Big\{ g_i : (y_{i,1}^l, y_{i,1}^u) \times \ldots \times (y_{i,d_i}^l, y_{i,d_i}^u) \to \mathbb{R} \,|$$

$$g_i(y_i) = b\, \Pi_{k=1}^{d_i} \cos(\omega_{i,k} y_{i,k} + \theta_{i,k}) \,,\, |b| \leq K_i \,, \omega_{i,k} = \frac{2\pi h}{y_{i,k}^u - y_{i,k}^l},\ h \in \mathbb{N},\ \theta_{i,k} \in [0, 2\pi) \Big\} \,.$$

Let $\mathcal{H}_i$ be the Hilbert space of functions $f_i : Y_i \to \mathbb{R}$ such that $E_{y_i}\{|f_i(y_i)|^2\} < \infty$, where the expected value is evaluated on Y_i. As the probability density q is bounded and $s_i^{o,\text{per},\bar{p}}$ is an extension of s_i^o, the latter belongs to the closure (with respect to $\mathcal{H}_i$ norm) of the convex hull of G_i. Moreover, for each $g_i \in G_i$ we have

$$\|g_i\|_{\mathcal{H}_i} = \sqrt{E_{y_i}\{|g_i(y_i)|^2\}} \leq K_i \,.$$

Hence, by the Maurey-Jones-Barron Lemma [8, Lemma 8.1], for every integer $k \geq 1$ and every $C_i > K_i^2 - \|s_i^o\|_{\mathcal{H}_i}^2$ there exists a function $\tilde{s}_i^k$ in the convex hull of k elements of G_i such that

$$\|s_i^o - \tilde{s}_i^k\|_{\mathcal{H}_i}^2 = E_{y_i}\{|s_i^o(y_i) - \tilde{s}_i^k(y_i)|^2\} \leq \frac{C_i}{k} \,. \tag{2}$$

We conclude the proof by taking projections and applying (2) and Proposition 4.1. ■

The next proposition estimates the accuracy of suboptimal solutions when Gaussian basis functions with variable centers and widths are used in (1).

Proposition 4.3 *Let the assumptions of Theorem 3.1 hold with an odd integer $m > \max_i\{d_i\}+1$. Then there exist $K(m, d_i) > 0$, $i = 1, \ldots, n$, and a positive constant C, which depends on $(s_1^o, \ldots, s_n^o)$, such that for every positive integer k there exists an n-tuple of strategies $(\tilde{s}_1^k, \ldots, \tilde{s}_n^k)$ such that*

$$v(s_1^o, \ldots, s_n^o) - v(\tilde{s}_1^k, \ldots \tilde{s}_n^k) \leq \frac{C}{\sqrt{k}} \,,$$

$$\tilde{s}_i^k(y_i) = \sum_{j=1}^k c_{ij} \text{Prj}_{A_i}(g_{ij}(y_i)) \,,$$

$$g_{ij} \in G_i \,,$$

$$G_i := \Big\{ g_i : Y_i \to \mathbb{R} \,|\, g_i(y_i) = e^{-\frac{\|y_i - t_i\|^2}{\delta_i}}, t_i \in \mathbb{R}^{d_i},\ \delta_i > 0 \Big\} \,,$$

$$\sum_{j=1}^k |c_{ij}| \leq K(m, d_i) \|\lambda_i\|_{\mathcal{L}_1(\mathbb{R}^{d_i})} \,,$$

and $\lambda_i \in \mathcal{L}_1(\mathbb{R}^{d_i})$ is such that $s_i^{o,\text{ext},1} = \mathcal{B}_{m-1} * \lambda_i$ for a suitable extension $s_i^{o,\text{ext},1}$ of s_i^o on $\mathbb{R}^{d_i}$.

Proof. Recall that for a positive integer d and $r > 0$, $\mathcal{B}_r : \mathbb{R}^d \to \mathbb{R}$ is the Bessel potential of order r, which is defined [29, Chapter 5, Section 3] as the inverse Fourier transform of $\hat{\mathcal{B}}_r(\nu) := \frac{1}{(1+4\pi^2\|\nu\|^2)^{\frac{r}{2}}}$.

As in the proof of Proposition 4.2, s_i^o can be extended to a function $s_i^{o,\text{ext},1} \in \mathcal{W}^{m-1,1}(\mathbb{R}^{d_i})$. Since $m-1$ is even, we can apply the strict inclusion of Sobolev spaces into Bessel potential spaces stated in [29, Remark 6.6 (b), p. 160]). Then, there exists $\lambda_i \in \mathcal{L}_1(\mathbb{R}^{d_i})$ such that $s_i^{o,\text{ext},1} = \mathcal{B}_{m-1} * \lambda_i$. As $m-1 > d_i$, the statement follows by [11, Corollary 5.2], Proposition 4.1, and the fact that, as Y_i is bounded, $\sqrt{E_{y_i}\{|s_i^o(y_i) - s_i(y_i)|^2\}}$ can be bounded from above by a constant times $\sup_{y_i \in Y_i} |s_i^o(y_i) - s_i(y_i)|$. ■

Propositions 4.2 and 4.3 provide upper bounds on the distance between the value of the team and its suboptimal value obtained in correspondence to suboptimal solutions expressed as k-term linear combinations (see (1)) of sinusoidal or Gaussian basis functions. The bounds are of the form $C\,k^{-1/2}$, where C is independent of k (it depends on $(s_1^o, \dots, s_n^o)$) and in principle can be bounded from above in terms of properties of the probability density q modelling the statistical information and of the other quantities that define the problem. So, to guarantee an approximation accuracy $\varepsilon > 0$ for the value of the team, it is sufficient to use $k \geq C^2\,\varepsilon^{-2}$ basis functions. Hence, the minimum number of sinusoidal or Gaussian basis functions required to guarantee an accuracy ε grows at most quadratically with ε.

It is interesting to compare the degree of smoothness of the team utility function required to apply Propositions 4.2 and 4.3 with the degree of smoothness required to apply the same propositions for a centralized version of the problem, where there is only a single decision maker with information vector $(y_1, \dots, y_n) \in \Pi_{i=1}^n Y_i \subset \mathbb{R}^{\sum_{i=1}^n d_i}$. In this case the value of the one-member team is obviously greater than or equal to the value of the decentralized team, however this centralized version has at least two drawbacks: the cost of making all information available to a single decision maker, and the larger degree of smoothness required.

5 Discussion

Team optimization problems are closely related to *potential games* [27, Section 3], for which there exists an algorithm that finds an ε-Nash equilibrium [27, Section 4]. Under certain conditions, it is known that a Nash equilibrium for a static team problem is team optimal, too [18, Section 3]. In such cases, an ε-Nash equilibrium is a reasonable approximation to a team-optimal set of strategies. In this framework, high-order smoothness properties of the optimal strategies helps in evaluating the high-dimensional integrals involved in a stochastic extension of the problem and of the algorithm in [27, Section 4]. This algorithm is particularly suitable for problems defined on networks of agents, as at each iteration each agent needs to know only the strategies of its neighbors in the network.

As to the results in the first part of the paper (Section 3), we remark that proving a suitable degree of smoothness of the (unknown) optimal strategies for team optimization problems may allow one to exploit in their solution the so-called *blessing of smoothness* [25]. Theorem 3.1 may be applied, e.g., to stochastic versions of the congestion, routing, and bandwidth allocation problems considered in [19, Lectures 3 and 4], which are stated in terms of smooth and concave utility functions.

Searching for the optimal values of all the parameters of the k-term suboptimal solutions (i.e., the coefficients of the linear combinations and the inner parameters of the basis functions) at the same time may result not only in high-dimensional optimization problems with a lot of local minima, but also in numerical instabilities. To cope with this problem, one may resort to the so-called *greedy algorithms*, in which, loosely speaking, the parameters are not optimized at once, but one after the other, hence by solving a sequence of low-dimensional problems. The k-term approximation is obtained inductively as a combination of the $(k-1)$-term one and a new element from the basis set. Depending on the way in which the low-dimensional optimization problems are defined, different kinds of greedy algorithms are obtained, which have shown their effectiveness in many contexts; see, e.g., [13] for variable-basis approximation and [32] for variable-basis convex optimization.

Acknowledgement

The authors were partially supported by a PRIN grant from the Italian Ministry for University and Research (project "Models and Algorithms for Robust Network Optimization").

References

[1] R. A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.

[2] M. Baglietto, T. Parisini, and R. Zoppoli, *Distributed-information neural control: The case of dynamic routing in traffic networks*, IEEE Trans. on Neural Networks **12** (2001), no. 3, 485–502.

[3] M. Baglietto, T. Parisini, and R. Zoppoli, *Numerical solutions to the Witsenhausen counterexample by approximating networks*, IEEE Trans. on Automatic Control **46** (2001), 1471–1477.

[4] M. Baglietto, M. Sanguineti, and R. Zoppoli, *The extended Ritz method for functional optimization: Overview and applications to single-person and team optimal decision problems*, Optimization Methods and Software **24** (2009), 15–43.

[5] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. on Information Theory **39** (1993), 930–945.

[6] T. Basar and R. Bansal, *The theory of teams: A selective annotated bibliography*, Lecture Notes in Control and Information Sciences, vol. 119, Springer, 1989.

[7] D. Blackwell, *Equivalent comparison of experiments*, Annals of Mathematical Statistics **24** (1953), 265–272.

[8] F. Girosi and G. Anzellotti, *Rates of convergence for Radial Basis Functions and neural networks*, Artificial Neural Networks for Speech and Vision (R. J. Mammone, ed.), Chapman & Hall, 1993, pp. 97–113.

[9] S. Giulini and M. Sanguineti, *Approximation schemes for functional optimization problems*, Journal of Optimization Theory and Applications **140** (2009), 33–54.

[10] G. Gnecco, *Functional optimization by variable-basis approximation schemes*, PhD Thesis in Mathematics and Applications, Department of Mathematics, University of Genoa, 2009.

[11] G. Gnecco and M. Sanguineti, *Approximation error bounds via Rademacher's complexity*, Applied Mathematical Sciences **2** (2008), 153–176.

[12] G. Gnecco and M. Sanguineti, *Error estimates for suboptimal solutions to team optimization problems with statistical information structure*, Tech. report, DIST, University of Genova, 2008.

[13] A. Hofinger, *Nonlinear function approximation: Computing smooth solutions with an adaptive greedy algorithm*, J. of Approximation Theory **143** (2006), 159 175.

[14] A. Iftar and E. J. Davison, *A decentralized discrete–time controller for dynamic routing*, Int. J. of Control **69** (1998), 599–632.

[15] V. Kůrková and M. Sanguineti, *Comparison of worst case errors in linear and neural network approximation*, IEEE Trans. on Information Theory **48** (2002), 264–275.

[16] V. Kůrková and M. Sanguineti, *Geometric upper bounds on rates of variable-basis approximation*, IEEE Trans. on Information Theory **54** (2008), 5681–5688.

[17] K. H. Kim and F. W. Roush, *Team theory*, Ellis Horwood Limited, Chichester, 1987.

[18] J. C. Krainak, J. L. Speyer, and S. I. Marcus., *Static team problems - Part 1: Sufficient conditions and the exponential cost criterion*, IEEE Trans. on Automatic Control **27** (1982), 839–848.

[19] Y. Mansour, *Lecture notes for a course on computational game theory*, http://www.cs.tau.ac.il/m̃ansour/course_games/2006/course_games_05_06.htm, 2006.

[20] J. Marschak, *Elements for a theory of teams*, Management Sciences, 1127–1137.

[21] J. Marschak and R. Radner, *Economic theory of teams*, Yale University Press, New Haven, 1972.

[22] J. F. Nash, *Non-cooperative games*, Annals of Mathematics **54** (1951), no. 2, 286–295.

[23] C. H. Papadimitriou, *Algorithms, games, and the Internet*, Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC), 2001, pp. 749–753.

[24] J. R. Patadia, *Local theorems for the absolute convergence of multiple lacunary Fourier series*, Mathematische Annalen, Springer **271** (1985), 81–89.

[25] T. Poggio and S. Smale, *The mathematics of learning: dealing with data*, Notices of the American Mathematical Society **50** (2003), 537–544.

[26] R. Radner, *Team decision problems*, Annals of Mathematical Statistics **33** (1962), 857–881.

[27] A. Rantzer, *Using game theory for distributed control engineering*, Tech. Report ISRN LUTFD2/TFRT--7620--SE, Department of Automatic Control, Lund University, Sweden, July 2008 - Presented at the 3rd World Congress of the Game Theory Society (Games 2008).

[28] A. Segall, *The modeling of adaptive routing in data–communication networks*, IEEE Trans. on Communications **25** (1977), 85–95.

[29] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ, 1970.

[30] V. Kůrková and M. Sanguineti, *Error estimates for approximate optimization by the extended Ritz method*, SIAM J. of Optimization **18, No. 2** (2005), 461–487.

[31] H. S. Witsenhausen, *Equivalent stochastic control problems*, Mathematics of Control, Signals, and Systems **1** (1988), 3–11.

[32] T. Zhang, *Sequential greedy approximation for certain convex optimization problems*, IEEE Trans. on Information Theory **49** (2003), 682–691.