2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Sparse Set Membership identification of nonlinear functions and application to control of power kites for wind energy conversion

C. Novara, L. Fagiano, M. Milanese

***Abstract*— A sparse approximation of a function is an approximation given by a linear combination of "many" basis functions, where the vector of linear combination coefficients is sparse, i.e. it has only "a few" non-zero elements. Identifying a sparse approximation of an unkown function from a set of data can be useful for many applications in the automatic control field: system identification, basis function selection, regressor selection, nonlinear internal model control, nonlinear feed-forward control, direct inverse control, predictive control, fast online applications. In this paper, a combined ℓ_1-relaxed-greedy algorithm for sparse identification is proposed and a Set Membership optimality analysis is carried out. Assuming that the noise affecting the data is bounded in norm and that the unknown function satisfies a mild regularity condition, it is shown that the algorithm provides an almost-optimal (in a worst-case sense) approximation of the unknown function. A simulation example is shown, related to direct-inverse control of a power kite used for high altitude wind energy conversion.**

I. Introduction

Sparse approximation consists in approximating a function using "a few" basis functions properly selected within a "large" set. More precisely, a sparse approximation is a linear combination of "many" basis functions, but the vector of linear combination coefficients is sparse, i.e. it has only "a few" non-zero elements. Deriving a sparse approximation of an unknown function from a set of its values (possibly corrupted by noise) is here called sparse identification.

Sparsification methods are relevant in many applications: compressive sensing [1], [2], [3], bioinformatics [4], computer vision [5], signal processing [6], [7], [8], source separation [9], denoising [10], linear regression [11], and regularization [12]. Analogies between sparse approximation and support vector machines have been shown in [13].

Recently, sparsification methods have been introduced in the automatic control field [14], [15], [16], with promising results. In this field, sparsification methods might be effective for both system identification and control design. In system identification, applications might include regularization, basis function selection, regressor selection, input selection. In control design, applications might include nonlinear internal model control, nonlinear feed-forward control, direct inverse control (see Section VI), predictive control, fast online applications.

The sparsity of a vector is typically measured by the ℓ_0 quasi-norm, defined as the number of its non-zero elements. Sparse identification can be performed by looking for a coefficient vector of the basis function linear combination with a "small" ℓ_0 quasi-norm. However, the ℓ_0 quasi-norm is a non-convex function and its minimization is in general an NP-hard problem. Two main approaches are commonly adopted to deal with this issue: convex relaxation and greedy algorithms, [17], [18], [19], [20]. In convex relaxation, a suitable convex function, e.g. the ℓ_1 norm, is minimized instead of the ℓ_0 quasi-norm, [18], [19], [20]. In greedy algorithms, the sparse solution is obtained iteratively, [17]. An interesting feature of these approaches is that, under certain conditions, they provide sparsest solutions, i.e. solutions which also minimize the ℓ_0 quasi-norm, [17], [18], [19], [21].

In this paper, a combined ℓ_1-relaxed-greedy algorithm is proposed for sparse identification. A condition under which the algorithm provides a sparsest solution is given in [22]. Then, a Set Membership optimality analysis is performed in order to assess the accuracy of the approximation obtained by the ℓ_1-relaxed-greedy algorithm. It is assumed that the noise affecting the data is bounded in norm and that the unknown function satisfies a mild regularity condition. It is shown that the ℓ_1-relaxed-greedy algorithm provides an almost-optimal approximation of the unknown function (i.e. an approximation whose worst-case identification error is not larger that twice the minimum achievable one). Note that the optimality analysis carried out here is completely different from the one in [22], since in [22], the unknown function to approximate is assumed to be of a given parametric form.

A simulation example is finally presented, related to direct-inverse control of a power kite used for high altitude wind energy conversion.

II. Notation and basic notions

A column vector is indicated by $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^{n\times 1}$, a row vector by $a^T = [a_1, a_2, \ldots, a_n] \in \mathbb{R}^{1\times n}$. For a matrix/vector $A \in \mathbb{R}^{K\times n}$, $K \in \{1, 2, \ldots\}$, and a set of indices $\lambda = \{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, n\}$, let us introduce the notation

$$A_\lambda \doteq [A_{i_1}, A_{i_2}, \ldots, A_{im}]$$

where A_j are the columns/elements of A. For a partially ordered set $A = \{A_1, A_2, \ldots, A_N\}$, and a set of indices $\lambda = \{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, N\}$, let us introduce a similar notation:

$$A_\lambda \doteq \{A_{i_1}, A_{i_2}, \ldots, A_{i_m}\}.$$

C. Novara is with Dip. di Automatica e Informatica, Politecnico di Torino, Italy. E-mail: carlo.novara@polito.it. L. Fagiano is with Dip. di Automatica e Informatica, Politecnico di Torino, Italy, and Dept. of Mechanical Engineering, University of California at Santa Barbara, USA. E-mail: lorenzo.fagiano@polito.it, lorenzo.fagiano@engineering.ucsb.edu. M. Milanese is with Kitenergy s.r.l. E-mail: mario.milanese@kitenergy.net.

This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. PIOF-GA-2009-252284 - Marie Curie project "Innovative Control, Identification and Estimation Methodologies for Sustainable Energy Technologies."

The ℓ_q norm of a vector a is defined as

$$\|a\|_q \doteq \left(\sum_{i=1}^{n} |a_i|^q\right)^{\frac{1}{q}}, \ q \in [1,\infty),$$
$$\|a\|_\infty \doteq \max_{i=1,..,n} |a_i|.$$

The ℓ_0 quasi-norm of a vector $a \in \mathbb{R}^n$ is defined as the number of its elements which are not null:

$$\|a\|_0 \doteq \mathrm{card}\left(\mathrm{supp}\left(a\right)\right) \tag{1}$$

where $\mathrm{card}\left(\cdot\right)$ is the set cardinality, and $\mathrm{supp}\left(a\right)$ is the support of a, defined as the set of indices at which a is not null:

$$\mathrm{supp}\left(a\right) \doteq \left\{i \in \{1,2,\ldots,n\} : a_i \neq 0\right\}.$$

The ℓ_0 quasi-norm is commonly used to measure the *sparsity* of a vector: the smaller is the ℓ_0 quasi-norm, the sparser is the vector. The complement of $\mathrm{supp}\left(a\right)$, i.e. the set of indices at which a is null, is denoted by

$$\begin{aligned}\overline{\mathrm{supp}}\left(a\right) &\doteq \left\{i \in \{1,2,\ldots,n\} : a_i = 0\right\} \\ &= \{1,2,\ldots,n\} \setminus \mathrm{supp}\left(a\right).\end{aligned}$$

The L_p norm of a function $f : X \to Y$, where $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}$, is defined as

$$\|f\|_p \equiv \|f\left(\cdot\right)\|_p \doteq \left[\int_X |f\left(x\right)|^p dx\right]^{\frac{1}{p}}, \ p \in [1,\infty),$$
$$\|f\|_\infty \equiv \|f\left(\cdot\right)\|_\infty \doteq \mathrm{ess\,sup}_{x\in X} |f\left(x\right)|.$$

Consider a generic feasible optimization problem

$$\begin{aligned}&a = \arg\min_a J\left(a\right) \\ &\text{subject to} \quad g\left(a\right) \leq 0.\end{aligned}$$

If this problem admits a set of solutions, then a indicates one of these solutions. Otherwise, a is the unique solution.

III. Problem setting

Consider a nonlinear function f_0 defined by

$$y = f_0\left(x\right) \tag{2}$$

where $x \in X \subset \mathbb{R}^{n_x}$, $y \in Y \subset \mathbb{R}$, and X and Y are compact sets. Suppose that f_0 is not known but a set of noise-corrupted data $D = \{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L}$ is available, described by

$$\widetilde{y}_k = f_0\left(\widetilde{x}_k\right) + d_k, \quad k = 1,2,\ldots,L \tag{3}$$

where d_k is noise.

Problem 1: Identify a sparse approximation of f_0 from the data set D. That is, identify from the data set D a function of the following parameterized form:

$$f_a\left(x,a\right) = \sum_{i=1}^{n} a_i \phi_i\left(x\right) = \phi\left(x\right) a \tag{4}$$

where $\phi\left(x\right) = \left[\phi_1\left(x\right), \phi_2\left(x\right), \ldots, \phi_n\left(x\right)\right]$, $\phi_i : X \to Y$ are basis functions, and $a = \left(a_1, a_2, \ldots, a_n\right) \in \mathbb{R}^n$ is a coefficient vector such that:
(i) a is "sparse";
(ii) the identification error

$$e\left(f_a\right) \doteq \|f_0 - f_a\|_p \tag{5}$$

is "small". ■

In this paper, following a Set Membership framework [23], [24], [25], [26], [27], the noise sequence $d = \left(d_1, d_2, \ldots, d_L\right)$ is assumed to be unknown but bounded:

$$\|d\|_2 \leq \mu \tag{6}$$

for some $\mu \geq 0$. No statistical assumptions on the noise d_k such as stationarity, uncorrelation, type of distribution, etc. are made. Indeed, such assumptions may be hard to be verified in several situations. Moreover, they are not suitable to evaluate the identification error for finite number of data L. A Set Membership optimality analysis for the case of ℓ_∞-norm-bounded noise is at present under investigation. As well known, the ℓ_∞ norm is more suitable for amplitude bounded noises, while the ℓ_2 norm for energy bounded noises.

A solution to Problem 1 can thus be found by looking for a function $f_a\left(x,a\right)$ of the form (4) such that 1) a is sparse; 2) $f_a\left(\widetilde{x},a\right)$ is consistent with the measured data and the prior assumptions on noise. Since the measured data are described by (3) and the prior assumptions on noise are given by (6), we have consistency if the following inequality holds: $\|\widetilde{y} - f_a\left(\widetilde{x},a\right)\|_2 = \|\widetilde{y} - \Phi a\|_2 \leq \mu$, where

$$\begin{aligned}\widetilde{y} &\doteq \left(\widetilde{y}_1, \ldots, \widetilde{y}_L\right) \\ \Phi &\doteq \begin{bmatrix} \phi_1\left(\widetilde{x}_1\right) & \cdots & \phi_n\left(\widetilde{x}_1\right) \\ \vdots & \ddots & \vdots \\ \phi_1\left(\widetilde{x}_L\right) & \cdots & \phi_n\left(\widetilde{x}_L\right) \end{bmatrix} \\ &= \begin{bmatrix} \phi_1\left(\widetilde{x}\right) & \cdots & \phi_n\left(\widetilde{x}\right) \end{bmatrix}.\end{aligned}$$

From (1), minimizing the ℓ_0 quasi-norm of a vector corresponds to minimizing the number of its non-zero elements, i.e. to maximizing its sparsity. Thus, the sparsest approximation of f_0 consistent with the measured data and the prior assumption is a function $f_a\left(x,a^0\right)$ of the form (4), where a^0 a solution of the following optimization problem:

$$\begin{aligned}&a^0 = \arg\min_{a\in\mathbb{R}^n} \|a\|_0 \\ &\text{subject to} \quad \|\widetilde{y} - \Phi a\|_2 \leq \mu.\end{aligned} \tag{7}$$

Definition 1: For given $\widetilde{y}$, Φ and μ, a coefficient vector is said *maximally sparse* if it is solution of the optimization problem (7). ■

Unfortunately, the optimization problem (7) cannot be solved in general, since the ℓ_0 quasi-norm is a non-convex function and its minimization is an NP-hard problem. Two main approaches are commonly adopted to deal with this issue: convex relaxation and greedy algorithms, [17], [18], [19], [20]. In convex relaxation, an optimization problem similar to (7) is solved, where the ℓ_0 quasi-norm is replaced by a suitable convex function. The ℓ_1 norm is often used, since this norm is the *convex envelope* of the ℓ_0 quasi-norm, [18], [19], [20]. In greedy algorithms, a sparse solution is obtained iteratively, by successively individuating the "most important" vector elements, [17]. A fundamental feature of these approaches is that, under certain conditions, they provide sparsest solutions, [17], [18], [19], [21]. However, the verification of these conditions is in general hard from

a computational standpoint, and can actually be performed only for particular types of basis functions.

In [22], a condition easily verifiable for any kind of basis functions is given, under which a vector is maximally sparse. Nevertheless, in situations where this condition is not satisfied, it is useful to consider a notion of maximum sparsity weaker than the one introduced in Definition 1.

Definition 2: For given $\widetilde{y}$, Φ and μ, a coefficient vector is said *maximally ζ-sparse* if it is solution of the optimization problem (7) subject to the additional constraints $a_i = 0$, $i \notin \zeta$, where $\zeta \subset \{1, 2, \ldots, n\}$. ■

According to Definitions 1 and 2, a coefficient vector which is maximally ζ-sparse for a set of indices $\zeta \subset \{1, 2, \ldots, n\}$, may not be maximally sparse, but it is anyway the sparsest one among all the vectors with $a_i = 0$, $i \notin \zeta$. This property is important since it ensures that no other elements a_i of a except those with index $i \notin \zeta$ can be set to zero without yielding $\|\widetilde{y} - \Phi a\|_2 > \mu$.

IV. Sparse identification of nonlinear functions

In this section, a combined ℓ_1-relaxed-greedy algorithm, completely based on convex optimization, is proposed for solving the sparse identification Problem 1.

Algorithm 1

1) Solve the optimization problem

$$a^1 = \arg\min_{a \in \mathbb{R}^n} \|a\|_1 \tag{8}$$

$$\text{subject to} \quad \|\widetilde{y} - \Phi a\|_2 \leq \mu \tag{9}$$

and define the following set of indices:

$$r\left(a^1\right) \doteq \left\{i_1, \ldots, i_j : 0 < \left|a^1_{i_1}\right| \leq \ldots \leq \left|a^1_{i_j}\right|\right\}$$

2) Compute the coefficient vector a^* as follows:

$$\begin{aligned}
&\text{for } k = 1 : \operatorname{card}\left(r\left(a^1\right)\right)\\
&\quad c^k = \arg\min_{a \in \mathbb{R}^n} \|\widetilde{y} - \Phi a\|_2 \qquad (10)\\
&\quad \text{subject to } a_i = 0,\\
&\qquad i \in \overline{\operatorname{supp}}\left(a^1\right) \cup r_\lambda\left(a^1\right)\\
&\qquad \lambda = \{1, \ldots, k\}\\
&\quad \text{if } \left\|\widetilde{y} - \Phi c^k\right\|_2 \leq \mu \qquad (11)\\
&\qquad a^* = c^k\\
&\quad \text{else}\\
&\qquad \text{break}\\
&\quad \text{end}\\
&\text{end} \quad ■
\end{aligned}$$

Algorithm 1 provides an estimate a^* of a^0, where a^0 is a maximally sparse coefficient vector, solution of the non-convex optimization problem (7). In [22], a condition is provided, ensuring that a^* has the same support as a^0, and is thus a maximally sparse vector as well. If a^* turns out to be not maximally sparse, this condition can be applied to the reduced vector a^*_ζ and matrix Φ_ζ, where ζ is the set of indices at which a^* is not null, in order to check if a^* is maximally ζ-sparse (see Definition 2).

V. Set Membership optimality analysis

In Section IV, an ℓ_1-relaxed-greedy algorithm is presented, able to derive a "sparse" approximation of the function f_0, thus allowing the accomplishment of the requirement (i) of the identification Problem 1. In this section, considering a Set Membership framework [23], [24], [25], [26], [27], this approximation is shown to have "small" identification error, thus allowing us to satisfy also the requirement (ii) of Problem 1.

In order to ensure a bounded identification error, some assumptions have to be made on the noise affecting the data and on the unknown function f_0. In this paper, the noise sequence $d = (d_1, d_2, \ldots, d_L)$ in (3) is assumed to be bounded according to (6). A regularity assumption is made on f_0, not requiring any knowledge on its functional form. Consider the following approximation of the function f_0:

$$f^*(x) = \sum_{i=1}^{n} a_i^* \phi_i(x) \tag{12}$$

where $\phi_1(x), \phi_2(x), \ldots, \phi_n(x)$ are known basis functions and a^* is the parameter vector identified using Algorithm IV from the data set D described in (3). Define the *residue function* Δ as

$$\Delta(x) \doteq f_0(x) - f^*(x).$$

In this subsection, it is assumed that f_0 is a function whose residue Δ is Lipschitz continuous over the compact set X:

$$\Delta \in \mathcal{F}(\gamma) \doteq \{f : |f(x) - f(\widehat{x})| \leq \gamma \|x - \widehat{x}\|_2, \forall x, \widehat{x} \in X\} \tag{13}$$

where the Lipschitz constant γ can be chosen by means of the procedure presented in [27]. Note that this assumption is not restrictive. Indeed, it is certainly satisfied for some $\gamma < \infty$, if f_0 and f^* are Lipschitz continuous. This kind of assumption has been introduced in [27], and allows us to analyze the optimality properties of an approximation when no information is available on the parametric form of f_0. Under the assumptions (13) and (6), we have that $f_0 \in FFS$, where FFS is called the Feasible Function Set.

Definition 3: The *Feasible Function Set* is

$$FFS \doteq \{f : f = f^* + \Delta,\ \Delta \in \mathcal{F}(\gamma), \|\widetilde{y} - f(\widetilde{x})\|_2 \leq \mu\}$$

where $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_L)$ and $f(\widetilde{x}) \doteq (f(\widetilde{x}_1), \ldots, f(\widetilde{x}_L))$. ■

According to this definition, FFS is the set of all functions consistent with prior assumptions and data. As in Subsection V, the prior assumptions are considered validated if at least an estimate consistent with these assumptions and the data exists, i.e. if FFS is not empty, see also [24], [26], [27].

Definition 4: Prior assumptions are validated if $FFS \neq \varnothing$. ■

The following theorem gives a necessary and sufficient condition for the validation of prior assumptions.

Theorem 1: $FFS \neq \varnothing$ if and only if the optimization problem (8) is feasible.

Proof. See [28] ■

Now assume that $FFS \neq \varnothing$. For a given approximation $\widehat{f}$ of f_0, a tight bound on the identification error $e\left(\widehat{f}\right)$ defined in (5) is given by the following worst-case error.

Definition 5: Worst-case identification error of an approximation $\widehat{f}$:

$$EN\left(\widehat{f}\right) \doteq \sup_{f \in FFS} \left\| f - \widehat{f} \right\|_p$$

where $\|\cdot\|_p$ is the functional L_p norm. ■

An optimal approximation is thus defined as a function f_{op} which minimizes the worst-case identification error.

Definition 6: An approximation f_{op} is *optimal* if

$$EN\left(f_{op}\right) = \inf_{\widehat{f}} EN\left(\widehat{f}\right). \quad ■$$

However, finding optimal approximations may be hard or not convenient, and sub-optimal solutions are looked for. In particular, approximations called interpolatory are often considered in the literature.

Definition 7: An approximation f_I is *interpolatory* if

$$f_I \in FFS. \quad ■$$

A fundamental property of an interpolatory approximation is that it guarantees a "small" worst-case error. Indeed, the degradation of an interpolatory approximation with respect to an optimal approximation is of at most 2: $EN\left(f_I\right) \leq 2 \inf_{\widehat{f}} EN\left(\widehat{f}\right)$, see [29], [24]. An approximation with this property is called *almost-optimal*. Note that any interpolatory approximation is almost-optimal, but not necessarily an almost-optimal approximation is interpolatory.

The following theorem shows that the approximation f^* defined in (12), where a^* is identified by Algorithm 1 from the data set D described in (3) is interpolatory (and thus almost-optimal). The theorem also provides an explicit expression of the worst-case identification error. Let us define the following functions:

$$\overline{\Delta}\left(x, \widehat{\varepsilon}\right) \doteq \min_{k=1,\ldots,L} \left(\delta_k\left(a^*\right) + \widehat{\varepsilon}_k + \gamma \left\| x - \widetilde{x}_k \right\|_2\right)$$
$$\underline{\Delta}\left(x, \widehat{\varepsilon}\right) \doteq \max_{k=1,\ldots,L} \left(\delta_k\left(a^*\right) - \widehat{\varepsilon}_k - \gamma \left\| x - \widetilde{x}_k \right\|_2\right)$$

where $\delta_k\left(a^*\right) = \widetilde{y}_k - f^*\left(\widetilde{x}_k\right)$ and $\widehat{\varepsilon}_k \geq 0$, $k = 1, \ldots, L$.

Theorem 2: For any functional L_p norm, with $p \in [1, \infty]$:
(i) The approximation f^* is interpolatory (and thus almost-optimal).
(ii) The worst-case identification error of f^* is bounded as

$$EN\left(f^*\right) \leq \max_{\substack{\|\widehat{\varepsilon}\|_2 \leq \mu \\ \|\widehat{\epsilon}\|_2 \leq \mu}} \left\| \overline{\Delta}\left(\cdot, \widehat{\varepsilon}\right) - \underline{\Delta}\left(\cdot, \widehat{\epsilon}\right) \right\|_p. \qquad (14)$$

Proof. See [28]. ■

Note that both the approximations provided by step 1 and 2 of Algorithm 1 are sparse. The one derived in step 2 is in general sparser than the one given by step 1. Both the approximations are almost-optimal if the required assumptions are satisfied, but none of them is guaranteed to be optimal. Indeed, deriving an optimal sparse approximation is at present an open problem.

VI. Example: direct-inverse control of a power kite for high altitude wind energy conversion

The Kitenergy technology aims to harvest High Altitude Wind Energy (HAWE) by using tethered flexible wings (power kites), connected to the ground by means of two lines, made of strong composite fibre and wound around two drums, kept at ground level and linked to reversible electric motors. The system composed by the kite, the lines, the onboard sensors, the drums, the generators and the control hardware is named Kite Steering Unit (KSU). The KSU can be employed in different configurations to generate energy (see e.g. [30] for details). In the so-called KE-yoyo configuration, the KSU is fixed with respect to the ground and energy is generated by continuously repeating a two-phase cycle, in which the lines are unrolled under high pulling forces, thus generating power, and then rolled back under low pulling forces. In the KE-carousel configuration, the line length is kept constant and energy is produced by exploiting the motion of the KSU along a fixed path on the ground, towed by the kite. Whatever configuration is used, the kite has to be controlled to fly on figure-eight paths in crosswind conditions (see Figure 2), which maximize the pulling forces on the line and hence the generated electrical power. However, such paths are unstable and thus cannot be tracked without some feedback control, see e.g. [31]. Therefore, one of the key components of the Kitenergy system is the control system, which should guide the kite in order to generate the maximum amount of power, while at the same time satisfying operational constraints, since the wing has to be kept above a minimal height from the ground and line wrapping has to be avoided. The design of the kite control system has been carried out in [31], [30] by applying nonlinear model predictive control techniques with quite good results, but they rely on an accurate model of the system, which in this case is hard to derive. A Sparse Direct Inverse Control (SDIC) approach is employed here instead.

The model described in [30] is used as the "real" system, with the parameters' values indicated in Table I. In this model, a Cartesian coordinate system (X, Y, Z) is considered, with X axis aligned with the nominal wind speed

TABLE I
Model parameters

m	5	Kite mass (kg)
A	10	Characteristic area (m^2)
d_l	0.0035	Diameter of a single line (m)
ρ_l	970	Line density (kg/m^3)
$C_{D,l}$	1	Line drag coefficient
α_0	3.5	Base angle of attack (°)
ρ	1.2	Air density (kg/m^3)
r	50	Line length (m)
d	5	Distance between the lines' hang points on the kite (m)
Δ_t	0.1	Sample time (s)

vector direction. Wind speed vector is represented as $\vec{W}_l = \vec{W}_0 + \vec{W}_t$, where $\vec{W}_0$ is the nominal wind, supposed to be known and expressed in (X, Y, Z) as:

$$\vec{W}_0 = \begin{bmatrix} W_x(Z) \\ 0 \\ 0 \end{bmatrix} \tag{15}$$

$W_x(Z)$ is a known function which describes the variation of wind speed with respect to the altitude Z (see e.g. [32]). The term $\vec{W}_t$ may have components in all directions and is not supposed to be known, accounting for wind unmeasured turbulence. In system (X, Y, Z), the kite position can be expressed as a function of its distance r from the origin and of the two angles θ and φ. The variable r is also the length of the lines, supposed to be straight. For simplicity, a fixed value of r is used here, but the SDIC approach proposed here can be used without significant modifications also with a variable line length r. The system model is described by a set of differential equations (see [30]):

$$\dot{z}(t) = h(z(t), u(t), W_t(t)) \tag{16}$$

where $z(t) = (\theta(t), \dot{\theta}(t), \varphi(t), \dot{\varphi}(t))$ is the state of the system, W is the wind speed, $u(t) = \arcsin(\Delta l/d)$ is the command input of the system, Δl is the difference between the lengths of the two lines, d is the distance between the attachment points of the two lines on the kite. The model (16) is supposed to be unknown here, but a finite set of noise-corrupted data can be acquired through preliminary experiments. In order to collect these data, a real-time simulator of the model (16) has been developed, where the input $u(t)$ can be chosen by means of a joystick. Using this simulator and considering a sampling period $T_s = 0.5$ s, a set of $L = 2000$ measurements $\widetilde{u}_k = u(T_s k)$, $\widetilde{z}_k = z(T_s k)$, $k = 1, 2, \ldots, L$, has been generated. In this simulation, the kite has been "manually" piloted by means of the joystick in such a way to cover the region $\varphi \in [-1, 1]$ rad, $\theta \in [0.5, 1.5]$ rad, without falling down. The following wind shear model (see (15)) has been used

$$W_x(Z) = \frac{\log\left(\frac{Z}{0.1}\right)}{\log\left(\frac{50}{0.1}\right)} 6.7 \tag{17}$$

Nominal wind speed is about 6 m/s at 30 m of altitude. Wind turbulence has been simulated by adding to the nominal wind $\vec{W}_0$ a random Gaussian vector of zero mean and standard deviation $\mathrm{std}(w(t)) = (1.5, 1.5, 1.5)\ m/s$. The measurements of each component of the state $\widetilde{z}_k$ have been corrupted by a white gaussian noise with a noise to signal standard deviation ratio of 3%.

A discrete-time inverse model of (16) has been identified from the generated data. This inverse model is given by

$$u_k = f^*(z_{k+1}, z_k)$$

where $k = 0, 1, \ldots$ is the discrete time, $u_k = u(T_s k)$, $z_k = z(T_s k)$, and f^* is an interpolatory sparse function of the form (12), identified by means of Algorithm 1 from the data $D = \{(\widetilde{z}_{k+1}, \widetilde{z}_k), \widetilde{u}_k\}_{k=1}^{L}$. A set of $n = L = 2000$ Gaussian basis functions of the form

$$\phi_i(x) = e^{-\|Q(x - \widetilde{x}_i)\|^2} \tag{18}$$

has been used, where $Q \in \mathbb{R}^{n_x \times n_x}$ is a diagonal matrix whose kth element is proportional to the width of the function along the dimension k. An optimal choice of the diagonal elements of Q has been performed by means of Lemma 2 in [27]. Among this set of functions, Algorithm 1 selected 47 basis functions (11 iterations have been performed in step 2 of the algorithm). The optimization problems in Algorithm 1 have been solved using the CVX package [33], [34].

The direct-inverse controller has been obtained as

$$u_k = f^*(r_{k+1}, z_k)$$

where r_k is a reference signal. Note that, since f^* is sparse (only 47 basis functions are used), the evaluation of this controller is very "fast" and can be easily performed online. Based on this controller, the control system depicted in Figure 1 has been implemented in Simulink®.

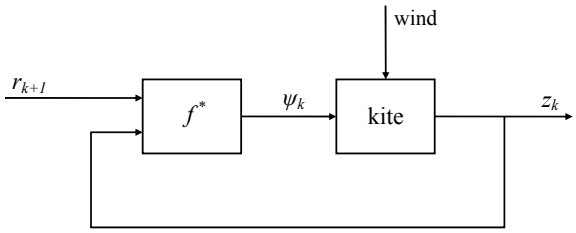

Fig. 1. Kite control system.

The control system has been tested using a reference signal corresponding to a periodic orbit having, in the (φ, θ)-plane, a figure-eight shape, see Figure 2. As previously discussed, this kind of orbit is optimal in terms of traction force maximization, but it is unstable without feedback control.

In Table II, the Root Mean Square tracking Errors are reported for two levels of turbulence: weak turbulence ($\mathrm{std}(w(t)) = (0.5, 0.5, 0.5)\ m/s$) and strong turbulence ($\mathrm{std}(w(t)) = (2, 2, 2)\ m/s$). These errors have been computed as $RMSE^i = \sqrt{\frac{1}{200}\sum_{k=1}^{200}\left(r_k^i - z_k^i\right)^2}$, where the superscript indicates the vector component. In Figure 2, the kite orbit is compared in the (φ, θ)-plane to the reference for the two wind strength levels. From these results, it can be concluded that the direct-inverse controller is able to yield an accurate tracking, even for strong wind turbulence.

Note that the situation simulated in this example is quite realistic: in a first phase, the kite is "manually" piloted in order to generate data; in a second phase, the data are used for model identification and/or control design; in a third phase, the kite is automatically piloted by the designed controller.

VII. Conclusions

A combined ℓ_1-relaxed-greedy algorithm of nonlinear functions has been proposed. A Set Membership optimality

wind std	$RMSE^1$	$RMSE^2$	$RMSE^3$	$RMSE^4$
1	0.0079	0.0086	0.0131	0.0082
5	0.0288	0.0202	0.0521	0.0316

TABLE II

ROOT MEAN SQUARE TRACKING ERRORS.

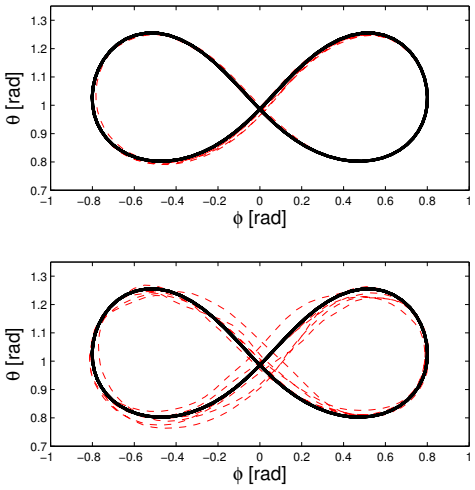

Fig. 2. Kite orbit. Above: weak wind. Below: strong wind. Bold (black) line: reference. Dashed (red) line: kite trajectory.

analysis has been performed in order to assess the identification accuracy of the approximations provided by the algorithm. It has been shown that the algorithm allows us to derive interpolatory (and thus almost-optimal) approximations of the unknown function to be identified. A simulation example has been presented, related to direct-inverse control of a power kite used for high altitude wind energy conversion. This example shows that the proposed algorithm can be used with satisfactory results in quite difficult problems, involving complex nonlinear systems.

REFERENCES

[1] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 –1306, apr. 2006.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489 – 509, feb. 2006.

[3] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406 –5425, dec. 2006.

[4] Z. Liu, S. Lin, and M. Tan, "Sparse support vector machines with l_p penalty for biomarker identification," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 7, no. 1, pp. 100 –107, jan. 2010.

[5] N. Ozay, M. Sznaier, and O. Camps, "Sequential sparsification for change detection," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–6.

[6] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 51, no. 1, pp. 101 – 111, jan. 2003.

[7] P. Frossard, P. Vandergheynst, R. Figueras i Ventura, and M. Kunt, "A posteriori quantization of progressive matching pursuit streams," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 525 – 535, feb. 2004.

[8] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520 –544, jan. 2010.

[9] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computations*, vol. 13, no. 4, pp. 863–882, 2001.

[10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[11] A. Miller, *Subset selection in regression*. London, UK: Chapman and Hall, 2002.

[12] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[13] F. Girosi., "An equivalence between sparse approximation and support vector machines," *Neural Computation*, vol. 10, no. 6, pp. 1455–1480, 1998.

[14] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, "A sparsification approach to set membership identification of a class of affine hybrid systems," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, 2008, pp. 123 –130.

[15] C. Feng, C. M. Lagoa, and M. Sznaier, "Hybrid system identification via sparse polynomial optimization," in *American Control Conference (ACC), 2010*, jun. 2010, pp. 160 –165.

[16] Y. Chen, Y. Gu, and A. Hero, "Sparse lms for system identification," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, apr. 2009, pp. 3125–3128.

[17] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231 – 2242, oct. 2004.

[18] J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, pp. 3601 –3608, oct. 2005.

[19] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1030 –1051, mar. 2006.

[20] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6 – 18, jan. 2006.

[21] R. Gribonval, R. Figueras i Ventura, and P. Vandergheynst, "A simple test to check the optimality of a sparse signal approximation," *Signal processing*, vol. 86, no. 3, pp. 496–510, 2006.

[22] C. Novara, "Sparse identification of nonlinear functions and parametric set membership optimality analysis," in *submitted to American Control Conference*, San Francisco, USA, 2011.

[23] M. Milanese and A. Vicino, "Optimal algorithms estimation theory for dynamic systems with set membership uncertainty: an overview," *Automatica*, vol. 27, pp. 997–1009, 1991.

[24] M. Milanese, J. Norton, H. P. Lahanier, and E. Walter, *Bounding Approaches to System Identification*. Plenum Press, 1996.

[25] A. Kurzhanski and V. Veliov, *Modeling techniques for uncertain systems*. Birkhäuser, 1994.

[26] J. Chen and G. Gu, *Control-Oriented System Identification: An H_∞ Approach*. New York: John Wiley & Sons, 2000.

[27] M. Milanese and C. Novara, "Set membership identification of nonlinear systems," *Automatica*, vol. 40/6, pp. 957–975, 2004.

[28] C. Novara, L. Fagiano, and M. Milanese, "Sparse identification of nonlinear functions and non-parametric Set Membership optimality analysis," *DAUIN internal report, Politecnico di Torino, Italy*, 2011.

[29] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*. Academic Press, Inc., 1988.

[30] L. Fagiano, "Control of tethered airfoils for high–altitude wind energy generation," Ph.D. dissertation, Politecnico di Torino, Italy, February 2009, available on–line: http://lorenzofagiano.altervista.org/docs/PhD_thesis_Fagiano_Final.pdf.

[31] A. Ilzhöfer, B. Houska, and M. Diehl, "Nonlinear MPC of kites under varying wind conditions for a new class of large-scale wind power generators," *International Journal of Robust and Nonlinear Control*, vol. 17, pp. 1590–1599, 2007.

[32] C. L. Archer and M. Z. Jacobson, "Evaluation of global wind power," *J. Geophys. Res.*, vol. 110, D12110, 2005.

[33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Aug. 2010.

[34] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/ boyd/graph_dcp.html.