The Foundations of Set Theoretic Estimation

PATRICK L. COMBETTES, MEMBER, IEEE

The conventional approach to estimation problems has been to optimize an objective function with or without constraints. The solvability of the resulting optimization problem is definitely a central issue and may lead to the selection of an unrealistic objective function and severe limitations in the incorporation of available information. Consequently, the reliability of the solutions becomes questionable, as they may violate known constraints about the problem. Set theoretic estimation is governed by the notion of feasibility and produces solutions whose sole property is to be consistent with all information arising from the observed data and **a priori** *knowledge. Each piece of information is associated with a set in the solution space and the intersection of these sets, the feasibility set, represents the acceptable solutions. The practical use of the set theoretic framework stems from the existence of efficient techniques for finding these solutions. Many scattered problems in systems science and signal processing have been approached in set theoretic terms over the past three decades. The purpose of this paper is to synthesize these various approaches into a single, general framework, to examine its fundamental philosophy, goals, and analytical techniques, and to relate it to conventional methods. Better understanding of the set theoretic approach will result in more applications in sciences and engineering and will stimulate further theoretical research.*

I. Introduction

Most estimation techniques are based upon solving an optimization problem. The signal that achieves the minimum mean-square error, the spectrum with the maximum entropy, the estimate which is the most likely, or the parameter that maximizes the posterior probability density are often regarded as desirable solutions for various problems. These formulations most frequently guarantee a single solution by proper choice of the cost function, e.g., a quadratic form. It is comforting to claim that "the" solution has been found. However, users whose interpretation of the best way to solve the problem differ may obtain different solutions. In addition, even in relatively simple problems, relating a practical aim to a precise mathematical optimization criterion is a difficult task [55], [100]. As pointed out in [222], our insistence on optimal solutions often leads to arbitrary decisions because the selection of a criterion of performance is inherently subjective and solving the problem may require oversimplifications in its formulation. For instance, the Wiener filter that is based on the squared estimation error is often used in image restoration only because of its mathematical simplicity, ignoring the properties of the human eye, which is known not to be an optimal least-squares detector. Moreover, despite a great many controversies they create among statisticians, the philosophical and theoretical problems associated with conventional estimation techniques, their performance measures, and their interpretation are usually ignored. From a practical standpoint, because of the uncertainty that surrounds the specifications of most problems, providing a region of acceptability for a solution rather than a single point seems more realistic. Of course, the question arises, "What is acceptable?"

An objective judgement on the acceptability of a solution must be based on the observed data sample as well as on all *a priori* knowledge about the problem. Following the definition Kant gave in 1781 in his *Kritik der reinen Vernunft*, *a priori* knowledge is knowledge that is independent of experience, i.e., knowledge that does not arise from the particular body of observed data currently being analyzed. Most estimation problems are accompanied with some *a priori* knowledge. Each piece of *a priori* information can only reduce our ignorance about the object to be estimated and is therefore valuable in increasing objectively the precision of the estimate. To a large degree, the amount of *a priori* knowledge available depends on our ingenuity and the extent of our theoretical and practical understanding of the physical system under study. The wide range of *a priori* knowledge frequently encountered in engineering applications includes information on the object to be estimated such as nonnegativity in image processing, properties pertaining to the system that generated the data such as stability of the system that produced speech samples, or information relative to external elements such as probabilistic attributes of the measurement noise.

The most straightforward way to obtain acceptable solutions is to incorporate all available information in the problem formulation. Conceptually, many conventional estimation techniques are capable of incorporating various types of information [152], [162]. However, the resulting constrained optimization problem may not be solvable by any known method. For instance, a major problem in any

Manuscript received November 14, 1991; revised January 6, 1993. The work in this paper was partially supported by the Université de Paris-Sud, Orsay, France.

The author is with the Department of Electrical Engineering, City College and Graduate School, City University of New York, New York, NY 10031.

IEEE Log Number 9208500.

Bayesian analysis is specifying a prior distribution that is sophisticated enough to incorporate all *a priori* knowledge but simple enough to make the problem algebraically tractable. Generally, computational considerations dominate the formulation of conventional estimation problems and little regard is paid to the rational selection of a cost function and the incorporation of *a priori* information, especially when it is nonstatistical. This may lead to estimates which are not consistent with *a priori* knowledge and whose reliability can therefore be questioned.

A close look at the signal processing and system theory literature of the past three decades reveals a number of isolated studies in which the sole property imposed on the solutions was to agree with all available information about the problem, be it arising from *a priori* knowledge or from the observed data. The approaches used in these studies can be labeled as set theoretic because each piece of information is conveniently represented by a set in the solution space and the intersection of such sets constitutes the family of solutions, i.e., the feasibility set. Consequently, the mathematical methods involved in the description, the analysis, and the solution of such problems rest heavily on the formalism of set theory. To capture the essence of these scattered approaches, we can define set theoretic estimation as an estimation framework in which consistency of a solution with the observed data and all *a priori* knowledge serves as the criterion of acceptability. The basic principle that more reliable estimates can be obtained through the incorporation of all available information can actually be implemented in the set theoretic framework for there exist techniques to compute feasible solutions for a large variety of practically important families of sets.

The wide spectrum of problems that have been approached in the set theoretic setting include control, signal restoration, signal reconstruction, image coding, speech processing, system identification, spectral estimation, array, and filter design. To date, however, there has not been any effort to synthesize these various approaches into a single general framework and to examine its fundamental philosophy, goals, and analytical techniques. The purpose of this paper is to address these issues and to establish set theoretic estimation on firm foundations in order to promote its use in proper applications as well as to stimulate further theoretical research. Although most of the discussion will focus on estimation problems, it must be noted that the set theoretic approach is also of great interest in design problems. In this context, each requirement, constraint, or desiratum on the solution is associated with a set in the solution space. The feasibility set is the family of objects that satisfy all the requirements, i.e., the intersection of all the sets. Examples of set theoretic design problems will be given in Section V-D.

The paper is organized as follows. Section II focuses on the general structure and the principles of the set theoretic framework and discusses the methodology involved in the construction of property sets in the solution space from various types of information. Section III addresses in its full generality the problem of the synthesis of a set theoretic solution from a family of property sets. Section IV discusses the connections between the presented framework and other estimation frameworks. Section V is a survey of applications fitting in the set theoretic framework. Further discussions and concluding remarks appear in Section VI. A relatively extensive list of references is included at the end of the paper, some of which are very specialized. Readers interested primarily in material addressing broad aspects of set theoretic estimation are referred to [24], [50], [60], [103], [104], [187], [200], [201], [207], and [221].

The prerequisite for most of the mathematical aspects of this paper is introductory analysis. Readers unfamiliar with the notions of metric space, closedness, compactness, convexity, norm, Hilbert space, and convergence should refer to the Appendix, where basic definitions and notations are provided, or to standard texts such as [68]. In those few places where more advanced notions are needed, the necessary definitions and background will be given in footnotes. Some knowledge of elementary statistics is also assumed.

All the terms that belong to the vocabulary of set theoretic estimation will be underlined the first time they appear in the text. Moreover, the following notations will be used throughout the paper. $\mathbb{N}$ is the set of nonnegative integers, $\mathbb{Z}$ the set of integers, $\mathbb{R}$ the set of real numbers, $\mathbb{C}$ the set of complex numbers, $\mathbb{R}^k$ the set of real k-tuples, and $\mathbb{E}^k$ the k-dimensional euclidean space (i.e., $\mathbb{R}^k$ equipped with the euclidean distance). $\mathfrak{P}(\Xi)$ is the family of all subsets of a space Ξ and $\complement S$ the complement of a set S. In a metric space, S° is the interior of S and $\bar{S}$ its closure. The scalar product of a Hilbert space will be denoted by $\langle\cdot|\cdot\rangle$ and its norm by $\|\cdot\|$.

II. Sets to Define Solutions

A. Classification of Information

In most problems, information to be used in determining the acceptability of a proposed solution can be classified into three groups, namely information about the solution, information about the system, and information about external factors.

Information about the solution represents our direct knowledge about the properties of the result and it explicitly defines the acceptability of a proposed solution. Examples for this type of information include signal intensity ranges in signal processing, nonnegativity of pixels in image processing, region of support in spectral estimation, rank or structure of a matrix in array processing, and stability of a system in system identification. Information about the system is mainly information relative to the properties of the physical system that generated the data and to the data generation model that establishes the relation between the solution and the recorded data. Generally, this type of information is incorporated indirectly in the problem formulation. As an example, if the recorded data x is related to the true solution h by the signal formation operator T, i.e., $x = T(h)$, the corresponding acceptability criterion for a proposed solution a will be $T(a) = x$, incorporating both recorded data and the data generation model. The last group of information pertains to external factors. In many cases, the results of an experiment are affected by unmeasurable factors such as model uncertainty and observation or recording noise. If these factors are totally unknown, it will not be possible to find a scientific estimate for the solution. In general, it is reasonable to assume the existence of some kind of information about various properties of these factors, e.g., bounds or partial statistical description of their stochastic nature. The incorporation of this type of information is mostly accomplished in an indirect form, similar to modeling system-related information. For instance, if the data generation model is $x = T(h) + u$, where u is the noise, an acceptability criterion for a proposed solution a will be that the residual $x - T(a)$ be consistent with the known properties of u.

Besides stochastic information, uncertain or imprecise information that does not have a frequentistic interpretation is common in many practical problems. Consider the following examples. Motion in sequential images is limited by physical constraints but the exact limits are unknown. For a lossless system, the measured energy of the output is close to that of the input. In signal recovery, signals may be described by vague attributes such as impulsiveness, smoothness, high energy, or similarity to a reference signal.

B. Fuzzy Information Modeling

The need for modeling all available information to determine the acceptability of a solution and, on the other hand, the large variety of possible available information necessitate the use of a very flexible information modeling technique. In this respect, the formulation with the most latitude is a list of statements where each statement indicates the acceptability of a proposed solution based on a particular piece of information. Combining each statement with an action in a precedent/antecedent format of a rule-based expert system may be the most straightforward solution approach for finding an acceptable solution. For example, in modeling the nonnegativity constraint on pixel values in image processing, one may impose a rule stating that if the result has negative values, they must be truncated to zero. Although it is possible to construct such expert system-based estimators, there is no general technique to define the actions that lead to a solution consistent with all the rules.

The fuzzy formalism provides a general framework to model precise or imprecise as well as certain or uncertain information.[1] In this framework, a piece of information is represented by a mapping $\Psi_\iota : \Xi \to [0, 1]$, called fuzzy proposition, which assigns to every point a in the solution space Ξ a grade of consistency $\Psi_\iota(a)$. The larger the grade, the stronger the belief that a satisfies the information represented by Ψ_ι. The statement $\Psi_\iota(a) \geq \psi_\iota$ means that the grade of consistency of a with Ψ_ι is at least ψ_ι. The range of a crisp (nonfuzzy) proposition Ψ_ι reduces to $\{0, 1\}$, i.e., $\Psi_\iota(a) = 1$ if Ψ_ι is true for a and $\Psi_\iota(a) = 0$ otherwise. Such propositions are also called Boolean propositions. As an example, let Ξ be a space of signals and let $\|a\|^2$ denote the energy of a signal a in Ξ. Consider the information "The energy of the original signal is 10." This precise information can be associated with the crisp proposition

$$(\forall a \in \Xi) \qquad \Psi_\iota(a) = \begin{cases} 1, & \text{if } \|a\|^2 = 10 \\ 0, & \text{otherwise.} \end{cases} \qquad (1)$$

Now suppose that the information pertaining to $\|h\|^2$ is given as "The energy of the original signal is close to 10." This imprecise information can be associated with the fuzzy proposition

$$(\forall a \in \Xi) \qquad \Psi_\iota(a) = \frac{1}{1 + |10 - \|a\|^2|}. \qquad (2)$$

It is noted that in both (1) and (2) $\Psi_\iota(a) = 1$ only when $\|a\|^2 = 10$. However, when $\|a\|^2 \neq 10$, (1) is identically 0, whereas (2) decreases smoothly to 0 as $\|a\|^2$ moves away from 10.

Let $(\Psi_\iota)_{\iota \in I}$ be the family of fuzzy propositions representing the corpus of information available about the problem. It will be assumed that all the Ψ_ιs are defined on the same solution space Ξ. Our basic objective is to find a point consistent with all the available information. A functional approach to this problem is to combine the Ψ_ιs into a single fuzzy proposition Ψ via some aggregation operator [70] and then find a point which yields the largest value of Ψ. Because the underlying goal is to satisfy all the information simultaneously, a fuzzy intersection operation is suitable. Although conceptually attractive, this strategy would unfortunately face obstacles similar to those encountered in the conventional estimation setting: rational selection of a meaningful intersection operator and computational tractability of the resulting optimization problem.[2]

In the set theoretic approach, each proposition Ψ_ι is associated with a set S_ι in Ξ. The intersection of all the S_ιs is the set of acceptable solutions. The main advantage of this approach stems from the existence of methods for the synthesis of set theoretic estimates, which make the set theoretic framework a very flexible and practical approach. We shall now formalize this framework around the concept of set theoretic formulation.

[1] No particular knowledge of fuzzy set theory will be required. One who wishes to see an explicit development of this matter can turn to [69] or [111].

[2] To illustrate the wide variety of fuzzy intersection operators and the potential complexity of maximizing an intersection of fuzzy propositions, consider the problem of intersecting two fuzzy propositions Ψ_ι and Ψ_κ via the Yager operator. The Yager intersection operation is given by [70]

$$\Psi_w = 1 - \min\{1, ((1 - \Psi_\iota)^w + (1 - \Psi_\kappa)^w)^{1/w}\}$$
$$\text{where } 0 < w < +\infty.$$

The value of $1/w$ determines the strength of the intersection performed. Thus, for $w = 1$, one obtains the aggregation $\Psi_1 = \max\{0, \Psi_\iota + \Psi_\kappa - 1\}$, which corresponds to the highest demand for simultaneous membership. On the other hand, as $w \to +\infty$, one obtains the weakest aggregation, $\Psi_{+\infty} = \min\{\Psi_\iota, \Psi_\kappa\}$, in which the lowest grade dictates the overall grade.

C. Set Theoretic Formulation

Let us consider a general estimation problem where the object to be estimated, h, belongs to a space Ξ. We shall call h the true object or estimandum,[3] Ξ the solution space, and a proposed solution, a, for the problem an estimate of h.

A formal definition of the set theoretic formulation can be given as follows. Let $(\Psi_\iota)_{\iota\in I}$ be the family of fuzzy propositions on Ξ representing all information available about the problem (information arising from the data and *a priori* knowledge) and let $(\psi_\iota)_{\iota\in I}$ be real numbers in $]0, 1]$ representing the strength of the beliefs that the true object satisfies these propositions. Then, a family $(S_\iota)_{\iota\in I}$ of subsets of Ξ can be constructed as follows:

$$(\forall\iota \in I) \quad S_\iota = \{a \in \Xi | \Psi_\iota(a) \geq \psi_\iota\}. \tag{3}$$

Each S_ι will be called a property set. Thus, S_ι is the set of all estimates that are consistent with the information carried by Ψ_ι at level ψ_ι. The pair $(\Xi, (S_\iota)_{\iota\in I})$ will be called a set theoretic formulation of the problem. The subset of Ξ of objects consistent with all available information is the feasibility set

$$S = \bigcap_{\iota\in I} S_\iota. \tag{4}$$

S will be called the solution set. Any point in S will be called a set theoretic estimate. The set theoretic formulation will be said to be consistent if $S \neq \emptyset$, fair if $h \in S$, and ideal if $S = \{h\}$. Fig. 1 depicts various set theoretic formulations.

In the jargon of fuzzy set theory, Ψ_ι is the membership function of a fuzzy set $\tilde{S}_\iota$; $\Psi_\iota(a)$ is the grade of membership of a in $\tilde{S}_\iota$ and the set S_ι in (3) is called the ψ_ι- cut of $\tilde{S}_\iota$ [69], [111]. If Ψ_ι is a crisp proposition, $S_\iota = \{a \in \Xi | \Psi_\iota(a) = 1\}$ and Ψ_ι is simply the indicator function of S_ι, i.e., $\Psi_\iota = 1_{S_\iota}$.

Before closing this section, the reader should be advised that (3) is the formal definition of a property set, not necessarily a constructive one. In many concrete instances, the actual construction of property sets will be done in a more straightforward fashion without explicitly invoking any fuzzy formalism. Examples will be given in Sections II-E and II-F.

D. The Solution Space and the Property Sets

The first fundamental component of a set theoretic formulation, the solution space, can take many forms, e.g., a field of scalars, a space of matrices, functions, or distributions. The primary criterion in selecting the solution space is being able to model all available information easily and accurately. A rule of thumb is to use a solution space that contains those objects directly described by most of the available information. For example, in a digital image restoration problem, if most of the information about the restoration result describes it as a spatial domain sampled and quantized image, the solution space must contain digital images of a given size as its main elements. On the other hand, in an ARMA estimation problem, if most of the available information is on the coefficients of the system, a solution space whose elements are vectors of system coefficients will be more suitable.

[3] In statistics, h is often called the true state of nature.

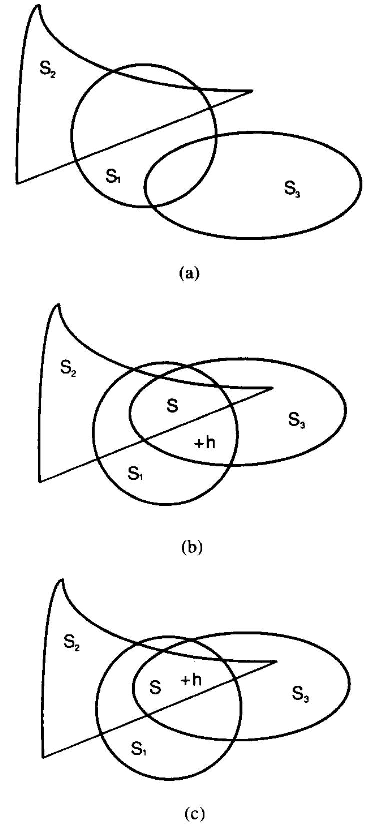


Fig. 1. Set theoretic formulations: (a) inconsistent; (b) unfair; (c) fair.

In many cases, all available information does not describe the solution in the same space. The simplest example for this may be the band-limited extrapolation problem where the available information describes the properties of a signal in both time and frequency domains. In such cases, the information that describes the solution in a space other than the solution space must be formulated so that an equivalent description in the selected solution space is provided. For the above example, if the solution space is selected to be that of discrete-time-domain real signals of length k, $\mathbb{R}^k$, and the frequency domain information states that for $|l| > B, H(l) = 0$, where $H(l)$ is the lth DFT frequency coefficient of h, then the information must be modeled using a set of time-domain signals such as

$$S_\iota = \{a \in \mathbb{R}^k | A(l) = 0 \quad \text{for } |l| > B\}. \tag{5}$$

After determining the type of the basic elements of the solution space, its structure must be defined. In mathematics, there exists a large number of classes of spaces which can be hierarchized according to their structure and their properties, from basic topological spaces to Hilbert spaces [13], [218]. As will be seen in Section III, the algorithms for generating set theoretic solutions that display the best convergence properties require a highly structured space, namely a Hilbert space in which all the property sets are closed and convex. Therefore, although algorithms exist for other cases, it is preferable to construct such set theoretic formulations whenever possible.

Theoretically, even in abstract estimation problems, finding a hilbertian solution space does not pose a major difficulty since there exist very general Hilbert spaces to which the estimandum can be assumed to belong (examples of a variety of Hilbert spaces can be found in [13] and [218]). For instance, a useful Hilbert space is the space $\mathcal{L}^2(\Omega, \mathcal{F}, \mu)$ of (classes of equivalence of) square-integrable functions on an abstract space Ω with respect to a measure μ;[4] in this space, the metric is given by $d(a,b) = (\int_\Omega |a - b|^2 d\mu)^{1/2}$ [6]. The space ℓ^2 of square-summable infinite complex sequences and the k-dimensional euclidean space $\mathbb{E}^k$ are particular cases of the space $\mathcal{L}^2(\Omega, \mathcal{F}, \mu)$ [6]. Other particular cases of $\mathcal{L}^2(\Omega, \mathcal{F}, \mu)$ of interest are the space of complex matrices equipped with the Frobenius norm, which has been used in signal enhancement problems [24], and the space L_n^2 of Lebesgue square-integrable functions on $\mathbb{R}^n$ commonly used in n-dimensional signal recovery [221]

$$\Xi = L_n^2 = \{a : \mathbb{R}^n \to \mathbb{R} | \cdot \int_{\mathbb{R}^n} |a(x_1, \cdots, x_n)|^2 \, dx_1 \cdots dx_n < +\infty\}. \quad (6)$$

Another important family of Hilbert spaces are Sobolev-type spaces, such as those employed in [186] for restoring two-dimensional vector fields

$$\Xi = \{a = (a_x, a_y) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} | \cdot \int_{\mathbb{R}^2} \left(\frac{\partial a_x}{\partial x}\right)^2 + \left(\frac{\partial a_x}{\partial y}\right)^2 + \left(\frac{\partial a_y}{\partial x}\right)^2 + \left(\frac{\partial a_y}{\partial y}\right)^2 dxdy < +\infty\}. \quad (7)$$

In [130], the Hilbert space

$$\Xi = \{(g_n)_{n \in \mathbb{Z}} | (\forall n \in \mathbb{Z}) \; g_n \in L_1^2 \text{ and } \sum_{n \in \mathbb{Z}} \|g_n\|^2 < +\infty\} \quad (8)$$

was used to reconstruct a signal from wavelet transform information. At any rate, since in applied work many problems are eventually formulated in $\mathbb{R}^k$ via parametrization or discretization, a readily available Hilbert space is the euclidean space $\mathbb{E}^k$ whose metric is given by $d(a,b) = (\sum_{i=1}^k |a_i - b_i|^2)^{1/2}$.

The condition that the sets be closed[5] is not too restrictive since the closure of a set S_ι is given by $\bar{S}_\iota = \{a \in \Xi | d(a, S_\iota) = 0\}$, where d is the underlying metric. In other words, by replacing S_ι by its closure in the set theoretic formulation, one merely adds points that are at distance zero from the points in S_ι, which will have no significant effect on the solution of a practical problem.

On the other hand, convexity may be more difficult to achieve since many constraints lead to nonconvex sets in the natural solution space. For instance, the set of stable autoregressive filters of order greater than two is not convex in the space of regression coefficients [10]; in signal recovery, the sets of signals whose energy is bounded from below [42], that of signals with a given number of levels [42], or that of signals with a prescribed Fourier transform magnitude [126] are not convex in the spatial domain (with the usual vector space structure). In such instances, one may replace all the nonconvex sets by their convex hull. This naive approach is seldom satisfactory, as it often leads to sets that are too large and therefore useless. For example, in discrete signal restoration, the convex hull of the (nonconvex) set of all nonnegative signals possessing no more than a given number of nonzero points is the whole space [49]. A better approach is to seek a new solution space where every piece of information yields a convex set. This strategy was adopted in the signal recovery problems of [34] and [36], where the vector space structure (addition, scalar multiplication) and, consequently, the scalar product of the Hilbert solution space were modified. It must be stressed that, in general, such an option may not be available for there may not exist a workable hilbertian solution space in which all the information can be associated with convex sets. For example, the property of stability for autoregressive filters can be associated with a (convex) hypercube in the space of reflection coefficients [114], but other sets in the formulation that were convex in the regression space (e.g., those discussed in [48] and [50]) may no longer be convex in the reflection space. The problem of maintaining the convexity of convex property sets in a change of solution space is also encountered in the context of two-dimensional phase retrieval [36].

It must be noted at this point that the selection of a solution space can also be considered an implicit way of modeling information. In selecting a space containing the coefficients of an ARMA system for the spectral estimation problem, one is implicitly enforcing a maximum order ARMA model on the signal generation mechanism.

E. Property Sets Based on Crisp Propositions

Some pieces of information can be modeled by crisp propositions, i.e., propositions which are either true or false for every object a in Ξ. From (3), for every ψ_ι in $]0,1]$, the property set S_ι associated with such a piece of information has the conceptual form $S_\iota = \{a \in \Xi | \Psi_\iota(a) = 1\}$. We now give examples of such sets.

As was seen earlier, a commonly used solution space in one-dimensional signal recovery is the space L_1^2 of (6). In this space, the set of nonnegative signals is

$$S_\iota = \{a \in L_1^2 | (\forall x \in \mathbb{R}) \; a(x) \geq 0\}, \quad (9)$$

[4] The definitions of a measure space $(\Omega, \mathcal{F}, \mu)$ and of the integral with respect to a measure μ can be found in [6].

[5] In the engineering literature, proofs of closedness are often unduly complicated. Since sets are often specified in the form $S_\iota = \{a \in \Xi | g_\iota(a) \leq \delta_\iota\} = g_\iota^{-1}(]-\infty, \delta_\iota])$, where $g_\iota : \Xi \to \mathbb{R}$, notice that closedness of S_ι will follow at once (by definition) from the lower semicontinuity (in particular, continuity) of g_ι, which is usually easily verified.

and the set of band-limited signals is

$$S_\iota = \{a \in L_1^2 | (\forall \nu \in \mathbb{R})\, A(\nu) = 0 \text{ if } |\nu| > B\} \qquad (10)$$

where A denotes the Fourier transform of a.

In the context of time series, consider the autoregressive model of order k

$$(\forall i \in \mathbb{Z}) \quad X_i = \sum_{j=1}^{k} h_j X_{i-j} + U_i. \qquad (11)$$

The problem is to estimate $h = (h_1, \cdots, h_k)$. The property set based on the information that the process $(X_i)_{i\in\mathbb{Z}}$ is causal and stationary is [20]

$$S_\iota = \{a \in \mathbb{R}^k | (\forall z \in \mathbb{C}) \quad z^k = \sum_{j=1}^{k} a_j z^{k-j} \Rightarrow |z| < 1\}. \qquad (12)$$

Now suppose that the driving process $(U_i)_{i\in\mathbb{Z}}$ is known to be uniformly bounded, say $(\forall i \in \mathbb{Z})\ |U_i| \leq \lambda$. Then, if $n + k$ data points $(x_i)_{1\leq i\leq n+k}$ of $(X_i)_{i\in\mathbb{Z}}$ have been observed, an estimate a that is consistent with the boundedness information will place the residual samples $(x_{k+i} - \sum_{j=1}^{k} a_j x_{k+i-j})_{1\leq i\leq n}$ in the interval $[-\lambda, \lambda]$. This leads to the property set

$$S_\iota = \bigcap_{i=1}^{n} \{a \in \mathbb{R}^k | \left| x_{k+i} - \sum_{j=1}^{k} a_j x_{k+i-j} \right| \leq \lambda\}. \qquad (13)$$

Let us note that noise boundedness information has been utilized in several set theoretic problems and forms the basis for the broad family of so-called "bounding-ellipsoid" algorithms [175] to be discussed in Section V-A.

F. Property Sets Based on Fuzzy Propositions

Fuzzy propositions are necessary to model pieces of information involving properties that do not change abruptly but according to a continuum of grades over the solution space. Statistical information and vaguely defined deterministic information fall in this category. The use of fuzzy propositions to define the sets in (3) provides a general methodological framework for the construction of property sets in terms of consistency levels with respect to such information.

In fuzzy set theory, there is no general technique for defining membership functions (fuzzy propositions) in an objective and systematic way and many guidelines have been proposed [69]. Ideally, a fuzzy proposition Ψ_ι should be defined so that its value for the true object is close to one and so that its selectivity is high in the sense that Ψ_ι attains large values only over a small region of the space. The higher the selectivity, the smaller the ψ_ι-cut set in (3) for a given ψ_ι. The selectivity of a fuzzy proposition is a function of the precision with which the properties characterizing the underlying information are known.

1) Statistical Information: In general, statistical constraints arise from the knowledge of probabilistic properties of the stochastic processes present in the data formation equation, e.g., noise or model uncertainty. In most cases, the construction of property sets based on such information is handled via statistical confidence theory. In this framework, one constructs a set S_ι of estimates consistent with a given piece of information to within a certain confidence coefficient. For example, let us go back to the autoregressive model (11) and let us now assume that the noise sequence $(U_i)_{i\in\mathbb{Z}}$ is zero mean, white, and Gaussian, with power σ^2. For a proposed solution a, let $m(a)$ be the sample mean of the residual based on the observations $(x_i)_{1\leq i\leq n+k}$ of $(X_i)_{i\in\mathbb{Z}}$, that is

$$m(a) = (1/n) \sum_{i=1}^{n} \left(x_{k+i} - \sum_{j=1}^{k} a_j x_{k+i-j} \right). \qquad (14)$$

The true sample mean obtained for $a = h$ is normal with mean zero and variance $\sigma'^2 = \sigma^2/n$. Thus, the acceptability of an estimate a with respect to the available information on $(U_i)_{i\in\mathbb{Z}}$ can be tested by the value of $|m(a)|$. If a 95% confidence coefficient is chosen, the set of estimates yielding a sample mean consistent with the available information on $(U_i)_{i\in\mathbb{Z}}$ is

$$S_\iota = \{a \in \mathbb{R}^k \,|\, |m(a)| \leq 1.96\sigma'\}. \qquad (15)$$

Similar sets based on noise properties were used in signal restoration in [201]. In [47], this approach was modified to also incorporate probabilistic information relative to random blurring kernels. In [50], it was demonstrated how numerous noise properties can be exploited to construct property sets via statistical confidence theory in a wide class of set theoretic estimation problems. The general form of the resulting property sets is

$$S_\iota = \{a \in \Xi | s(a) \in C_\iota\} \qquad (16)$$

where, for an estimate a, $s(a)$ is the observed value of the statistic of the residual associated with a certain property of the noise, and C_ι the confidence region based on a desired confidence coefficient. This approach was further extended in [44] to also incorporate information relative to model uncertainty.

Although the fuzzy propositions need not be explicitly specified, it is underlying in the construction of such sets based on confidence theory and can be obtained from the distribution of the statistic defining the set. In the above example, suppose that a fuzzy proposition Ψ_ι is constructed via the probability density normalization technique suggested in [69]. Since the true sample mean is normal with mean zero and variance σ'^2, we get

$$\Psi_\iota(a) = \exp(-|m(a)|^2/(2\sigma'^2)). \qquad (17)$$

The property set (15) is now seen to be of the conceptual form of (3) by letting $\psi_\iota = \exp(-1.96^2/2)$. In general, the grade of consistency ψ_ι is directly related to the confidence coefficient. Naturally, as the sample size increases, σ'^2 goes to zero and the support of Ψ_ι undergoes a contraction, which increases selectivity. In the limit, Ψ_ι approaches a crisp proposition. This simply translates the fact that the uncertainty surrounding the value of $m(h)$ has decreased, for the sample mean is computed from a larger sample. In general, the grade of consistency ψ_ι is directly related to the confidence coefficient. Finally, it should be mentioned that other techniques have been developed to specify fuzzy propositions based on probability densities according to certain criteria [41].

2) Vaguely Defined Deterministic Information: As previously discussed, instances of vaguely defined deterministic information are common in estimation problems. In the context of signal restoration, fuzzy propositions associated with various vaguely defined signal attributes were proposed in [42]. For example, a soft upper bound on the energy of a discrete signal in $\mathbb{R}^k$ was modeled with a fuzzy proposition of the form

$$\Psi_\iota(a) = \exp\left(-\alpha \sum_{i=1}^{k} |a_i|^2\right), \quad \alpha > 0. \quad (18)$$

The knowledge that the true signal is in the neighborhood of a prototype signal, r, was modeled with a fuzzy proposition of the form

$$\Psi_\iota(a) = \exp\left(-\sum_{i=1}^{k} \alpha_i |a_i - r_i|^2\right), \quad \alpha_i > 0. \quad (19)$$

Fuzzy propositions modeling smoothness and signals with a finite number of levels were also given in [42]. As another example, consider a problem with a matrix solution space. The information that the true matrix is near-Toeplitz can be modeled by

$$\Psi_\iota(a) = \exp(-\alpha \|a - \mathrm{Toep}(a)\|), \quad \alpha > 0 \quad (20)$$

where $\|\cdot\|$ is a suitable norm and where $\mathrm{Toep}(a)$ is the best Toeplitz approximation of a (obtained by replacing each entry of a by the average value of the entries along its diagonal in the case of the Frobenius norm). While the above propositions are of the exponential type, other forms are possible, e.g., (2).

Of course, in practice, the sets can often be constructed in a more straightforward fashion, which by-passes the specifications of a fuzzy proposition Ψ_ι and of a consistency level ψ_ι. For instance, upon combining (3) and (20), we obtain the set

$$S_\iota = \{a \in \Xi \mid \|a - \mathrm{Toep}(a)\| \leq \delta_\iota\} \quad (21)$$

where $\delta_\iota = -\alpha^{-1} \ln \psi_\iota$. This set could have been obtained directly, by specifying a bound δ_ι in accordance with the user's confidence in the near-Toeplitzness of the estimandum.

G. The Analysis of Set Theoretic Formulations

Once a set theoretic formulation $(\Xi, (S_\iota)_{\iota\in I})$ has been constructed, several questions arise *vis à vis* its properties, its informational content, and its intrinsic value.

In point estimation theory, the estimators are usually accompanied by an estimate of their accuracy. Of course, any criterion of accuracy is arbitrary; a typical one is the mean-square error, which reduces to the variance in the unbiased case. The question of the accuracy of set theoretic estimates was briefly touched upon in [175], where suggested criteria for the ellipsoidal approximation $\{a \in \mathbb{R}^k \mid (a - a_0)^t M^{-1}(a - a_0) \leq 1\}$ of the exact feasibility set are $\mathrm{tr}M$, $\det M$, the largest eigenvalue of M, and $v^t M v, v$ being a direction of interest. In the context of digital signal restoration, the diameter of the solution set is used in [42], [113], and [201] as a criterion of quality. In [9], the accuracy of a set theoretic estimator in $\mathbb{R}^k$ is defined as its measure.

Let h be the estimandum and let $(\Xi, (S_\iota)_{\iota\in I})$ be a fair set theoretic formulation with $S = \cap_{\iota\in I} S_\iota$. Intuitively, the accuracy of $(\Xi, (S_\iota)_{\iota\in I})$ should be evaluated by a monotone function of S, which attains its minimum value when the formulation is ideal, i.e., $S = \{h\}$. These requirements were formalized in [45] by introducing the notion of mensuration. Let $\mathfrak{S}$ be a σ-algebra of subsets of Ξ containing $(S_\iota)_{\iota\in I}$ (e.g., $\mathfrak{S} = \mathfrak{P}(\Xi)$) and let $\mathfrak{S}_h = \{A \in \mathfrak{S} \mid h \in A\}$. Then a mensuration for $(\Xi, (S_\iota)_{\iota\in I})$ is a monotone set function $\nu : \mathfrak{S}_h \to [0, +\infty]$ that does not vanish on any of the S_ιs. The accuracy of $(\Xi, (S_\iota)_{\iota\in I})$ is then defined to be the number $\nu(S)$. Note that the lowest value is achieved in the case of an ideal set theoretic formulation. In this case, the available information describes uniquely the true object. Generally speaking, $\mathfrak{S}_h$ can be thought of as the family of (measurable) sets representing all possible information about h; only the subfamily $(S_\iota)_{\iota\in I}$ of $\mathfrak{S}_h$ is available to characterize the acceptability of an estimate. Assuming that (Ξ, d) is a metric space, specific examples of mensurations of a set A in $\mathfrak{S}_h$ are the diameter of A; the thickness of A (i.e., the diameter of the largest open ball contained in A); $\mu(A)$, $\int_\Xi \delta_A d\mu$, and μ-ess $\sup\{\delta_A(a) \mid a \in \Xi\}$, where μ is a measure on $\mathfrak{S}$ that puts mass on the S_ιs and δ_A a proper deviation function [45]. For instance, in $\mathbb{E}^k$, the diameter, thickness, and Lebesgue measure of an ellipsoid are, respectively, its major and minor axes, and its volume. These basic mensurations are shown in Fig. 2.

Before utilizing a set theoretic formulation, one should remove those property sets that bring little or no original information for they will have little qualitative effect on a solution but will add burden to its computation. In order to identify those sets which carry significant information relative to the other sets present in the formulation, a useful notion is that of contribution. Given a mensuration ν on $\mathfrak{S}_h$, the contribution of $S_\kappa (\kappa \in I)$ to $(\Xi, (S_\iota)_{\iota\in I})$ is defined as [45]

$$\bar{\nu}(S_\kappa) = \nu\left(\bigcap_{\iota\in I-\{\kappa\}} S_\iota\right) - \nu\left(\bigcap_{\iota\in I} S_\iota\right). \quad (22)$$

It can be viewed as the degradation in the accuracy of the set theoretic formulation incurred by the removal of S_κ. The smaller this number, the smaller the contribution. In the extreme, suppose that there exists a set $S_\lambda (\lambda \in I)$ such that

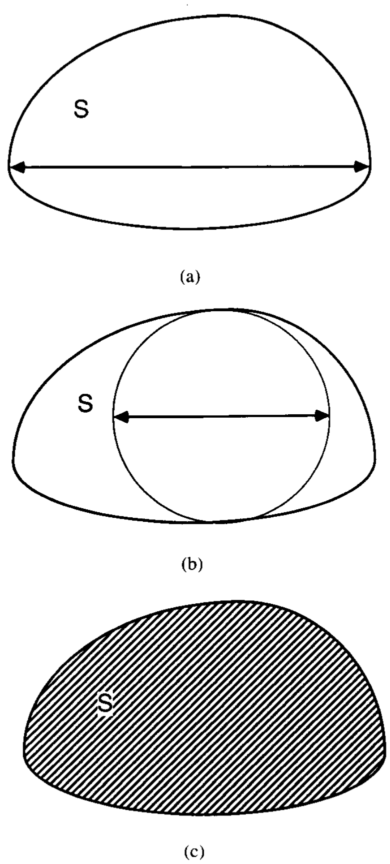


Fig. 2. Mensurations in E^2: (a) diameter; (b) thickness; (c) area.

$S_\lambda \subset S_\kappa$. Then, the information carried by Ψ_κ is redundant in the presence of Ψ_λ and the contribution of S_κ is null: $\bar{\nu}(S_\kappa) = 0$. In some problems, sets with little innovative information may be identifiable directly. For instance, in the bounded noise ARMA estimation problems of [59] and [103], each data sample gives rise to a property set and the resulting feasibility set is approximated by an ellipsoid; a test is developed to discard the sets that do not help and reduce the ellipsoid. Redundancy tests for various noise properties and data models are discussed in [50].

An important characteristic of $(\Xi, (S_\iota)_{\iota \in I})$ is its consistency, i.e., whether or not $S = \cap_{\iota \in I} S_\iota = \emptyset$. Inconsistent set theoretic formulations may arise if one unadvertedly includes mutually exclusive crisp propositions in $(\Psi_\iota)_{\iota \in I}$ or specifies ψ_ι-levels in (3) that are too high. They may also be knowingly constructed when the goal is to obtain an approximate feasible solution (e.g., see [89] and [133]). In Section III–V, algorithms that yield approximate solutions for inconsistent set theoretic formulations will be discussed. In general, consistency is difficult to check analytically and is often revealed by the convergence behavior of the solution algorithm.

There are other properties of set theoretic formulations that may be of interest. For instance, if S is balanced and a is a feasible solution, then so is any down-scaled version αa, where $|\alpha| \leq 1$. If S is convex, averaging feasible solutions will still yield a feasible solution. Finally, let us note that in some problems the property sets or even the feasibility set may be disconnected. A pictorial description of a disconnected property set is given in [50] and problems with disconnected feasibility sets are discussed in [155].

III. Mathematical Methods

The purpose of this section is to address the general problem of the computation of set theoretic estimates, i.e., the problem of finding a point in the solution set (4). The material will be exposed in a relatively detailed and rigorous manner. Several algorithms and results that have been recently proposed in the applied mathematics literature will be made available to the engineering community. In particular, attention will be drawn to parallel methods. Although no mention will be made of any particular application at this point, the reader should constantly bear in mind the geometrical interpretation of the results and should make the natural connection with potential applications. Figures will be provided to visualize some concepts.

A. Feasibility Problem

Let $(\Xi, (S_\iota)_{\iota \in I})$ be a set theoretic formulation where I is finite, say $I = \{1, \cdots, m\}$.[6] It will be assumed that Ξ is a metric space with distance d and, unless otherwise stated, that the set theoretic formulation is consistent. Generating a set theoretic estimate is tantamount to solving the feasibility problem

$$\text{Find } a \in S = \bigcap_{\iota \in I} S_\iota. \tag{23}$$

This basic problem has a long history and various solution methods have been proposed, which depend on the metric and geometrical properties of the sets $(S_\iota)_{\iota \in I}$ and the structure of the underlying space Ξ. Because (23) can usually not be solved in one step, most feasibility algorithms are iterative and consist of building a sequence $(a_n)_{n \geq 0}$ converging in some sense to a point in S according to the general recursion

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} \in R_n(a_n) \tag{24}$$

where $(R_n)_{n \geq 0}$ is a sequence of set-valued operators from Ξ into $\mathfrak{P}(\Xi)$ and a_0 a point in Ξ. In words, the update is performed by selecting any point in the set $R_n(a_n)$ that is computed in terms of the current iterate a_n. If, for every a in Ξ and every n in $\mathbb{N}$, $R_n(a_n)$ reduces to a singleton, then (24) takes the form

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = R_n(a_n). \tag{25}$$

The following convergence properties of (24) will be of interest:

$\mathcal{P}1$): For every a_0 in Ξ, $(a_n)_{n \geq 0}$ converges to the projection of a_0 onto S.

[6] Some of the results shall remain true with an infinite number of sets.

- $\mathcal{P}2$): For every a_0 in Ξ, $(a_n)_{n\geq 0}$ converges to a point in S.
- $\mathcal{P}3$): For every a_0 in Ξ, $(a_n)_{n\geq 0}$ converges weakly to a point in S.
- $\mathcal{P}4$): For every a_0 in Ξ, $(a_n)_{n\geq 0}$ possesses at least one cluster point and all of its cluster points are in S.

In general, we have $(\mathcal{P}_1) \Rightarrow (\mathcal{P}_2) \Rightarrow (\mathcal{P}_4)$. If Ξ is a normed vector space, the convergence in $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ is understood to be strong convergence and $(\mathcal{P}_2) \Rightarrow (\mathcal{P}_3)$; moreover, if its dimension is finite (e.g., $\Xi = \mathbb{E}^k$), then $(\mathcal{P}_2) \Leftrightarrow (\mathcal{P}_3)$.

At each iteration, (24) uses the constraints associated with the sets to form the update. The algorithm is said to be serial if only one set is activated at each iteration, i.e.,

$$(\forall n \in \mathbb{N}) \qquad R_n = R_{\iota_n} \tag{26}$$

where $(\iota_n)_{n\geq 0}$ is a sequence of indices in I, called control sequence. The control sequence dictates the order in which the property sets are activated. The most remote set control scheme consists of letting ι_n be the index of the set the farthest from a_n in terms of the metric of Ξ. If $(\forall n \in \mathbb{N})\ \iota_n = n \text{ (modulo } m) + 1$, the control is said to be cyclic. It is said to be chaotic if each index ι in I appears infinitely often in $(\iota_n)_{n\geq 0}$. On the other hand, the algorithm is said to be parallel if several sets are activated simultaneously, i.e.,

$$(\forall n \in \mathbb{N}) \qquad R_n = R_{I_n}, \qquad \text{where} \quad \emptyset \neq I_n \subset I. \tag{27}$$

The control sequence $(I_n)_{n\geq 0}$ is said to be static if, for every n in $\mathbb{N}$, $I_n = I$, almost cyclic if there exists a positive integer M such that, for every integer n, $I \subset \cup_{k=0}^{M-1} I_{n+k}$, and chaotic if each index ι in I is contained in infinitely many I_ns.

We now proceed to introduce feasibility algorithms for increasingly complex set theoretic formulations. It is recalled that in a metric space (Ξ, d) the distance from a point a to a nonempty subset S_ι is defined as $d(a, S_\iota) = \inf\{d(a,b) | b \in S_\iota\}$ and that a projection of a onto S_ι is any point b in S_ι such that $d(a, S_\iota) = d(a, b)$. Such a point is also called a best approximation of a by a point in S_ι. If Ξ is a Hilbert space and if S_ι is closed and convex, every point a admits a unique projection onto S_ι [13], [128] that will be denoted by $P_\iota(a)$; the point $2P_\iota(a) - a$ will be called the reflection of a with respect to S_ι.

B. Sets Defined by Affine Subspaces

In this section, the focus is placed on set theoretic formulations consisting of closed convex sets defined by affine subspaces in a Hilbert space. Let $T_\iota : \Xi \to \mathbb{R}$ be a nonzero continuous linear functional and $\gamma_\iota, \delta_\iota$ real numbers. It is recalled that the sets $\{a \in \Xi \,|\, T_\iota(a) = \delta_\iota\}$, $\{a \in \Xi | T_\iota(a) \leq \delta_\iota\}$, and $\{a \in \Xi | \gamma_\iota \leq T_\iota(a) \leq \delta_\iota\}$ are called a closed affine hyperplane, half-space, and hyperslab, respectively. First of all, suppose that the set theoretic formulation consists of affine hyperplanes in $\mathbb{E}^k$, namely

$$(\forall \iota \in I) \qquad S_\iota = \{a \in \mathbb{R}^k | \langle a | b_\iota \rangle = \delta_\iota\} \tag{28}$$

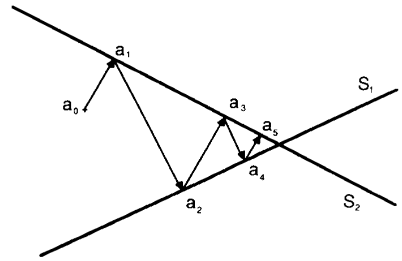


Fig. 3. Kaczmarz's algorithm.

where b_ι is a nonzero vector in $\mathbb{R}^k$, and δ_ι is a real number. The first feasibility algorithms based on projections were proposed by Kaczmarz in 1937 [109] and Cimmino in 1938 [39] for solving systems of linear equations.[7] These fundamental algorithms are of great importance, for all the projection methods to be discussed hereafter can be viewed as extensions of one or the other. Kaczmarz's method is serial and satisfies $(\mathcal{P}_1)$. It proceeds by cyclic projections onto each hyperplane as follows:

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = P_{\iota_n}(a_n), \quad \text{where} \quad \iota_n = n(\text{modulo}\, m) + 1. \tag{29}$$

On the other hand, Cimmino's method is parallel and satisfies $(\mathcal{P}_2)$. It takes as the next iterate the average of the reflections of the current iterate with respect to all the hyperplanes, i.e.,

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = (2/m) \sum_{\iota \in I} P_\iota(a_n) - a_n. \tag{30}$$

A pertinent discussion of Kaczmarz's and Cimmino's methods can be found in [82] and a pictorial description of their operation is shown in Figs. 3 and 4. Halperin [93] has shown that (29) also satisfies $(\mathcal{P}_1)$ in the general case where the S_ιs are arbitrary closed affine subspaces and Ξ any Hilbert space.[8] In this case, it is established in [160] that $(\mathcal{P}_1)$ is also satisfied by the parallel algorithm

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = \sum_{\iota \in I} w_\iota P_\iota(a_n) \tag{31}$$

where the weights on the projections satisfy

$$\sum_{\iota \in I} w_\iota = 1, \qquad \text{and} \quad (\forall \iota \in I)\ w_\iota > 0. \tag{32}$$

Kaczmarz's method is further investigated in [195] and a generalization of Cimmino's method to solve integral equations of the first kind in $L^2[a,b]$ is given in [110]. Cimmino-like algorithms have also been shown to be related to the Landweber iteration [202].

[7] Historically, the first alternating projection algorithm seems to have been developed by Schwarz around 1870 in connection with the integration of partial differential equations [172] (see also [58]).

[8] Halperin's proof is given for vector subspaces but it can be extended routinely to affine subspaces. The case $m = 2$ is known as Von Neumann's Alternating Projection Theorem [206] (see also [210]). For generalizations to nonhilbertian spaces, see [79]. The case of inconsistent formulations is discussed in [117].

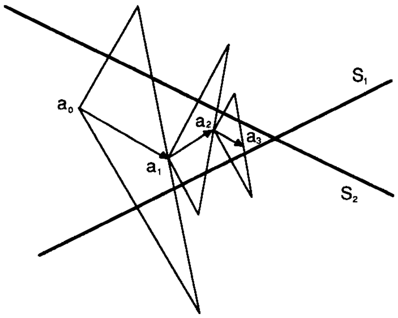


Fig. 4. Cimmino's algorithm.

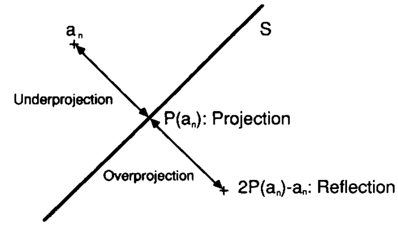


Fig. 5. Relaxed projection onto S.

The relaxation method of Agmon-Motzkin-Schoenberg [2], [141] extends Kaczmarz's method to formulations in which the property sets are closed affine half-spaces, that is

$$(\forall \iota \in I) \qquad S_\iota = \{a \in \mathbb{R}^k \mid \langle a|b_\iota\rangle \leq \delta_\iota\}. \tag{33}$$

In this algorithm, the sets are activated serially according to the iteration

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda(P_{\iota_n}(a_n) - a_n) \tag{34}$$

where the relaxation parameter λ lies in]0, 2]. As illustrated in Fig. 5, the case $\lambda = 2$ corresponds to a reflection, $1 < \lambda < 2$ to an overprojection, $\lambda = 1$ to a projection (unrelaxed iteration), and $0 < \lambda < 1$ to an underprojection. In other words, a_{n+1} is located on the open segment between a_n and its reflection with respect to S_{ι_n}. Algorithm (34) satisfies $(\mathcal{P}_2)$ in the cases of cyclic and most remote set controls. Discussions of its rate of convergence in terms of λ can be found in [87] and [132]. The fact that it satisfies $(\mathcal{P}_2)$ in the general case where the S_ιs are arbitrary closed affine half-spaces and Ξ a general Hilbert space is proved in [92].[9] In [31], a modified least-squares algorithm satisfying $(\mathcal{P}_2)$ is proposed for formulations of type (33). It contains as a special case the parallel projection method

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda_n \left(\sum_{\iota \in I} w_\iota P_\iota(a_n) - a_n \right) \tag{35}$$

where the weights satisfy (32) and where, for every integer n and a fixed number α in]0, 2[, the relaxation parameter is given by

$$\lambda_n = \begin{cases} \alpha / \sum_{\iota \in I_n} w_\iota, & \text{if } \mu(I_n) \geq 2 \\ \alpha, & \text{otherwise} \end{cases}$$
$$\text{with} \qquad I_n = \{\iota \in I \mid a_n \notin S_\iota\} \tag{36}$$

where $\mu(I_n)$ is the number of points in I_n. In the method proposed in [215] for $\Xi = \mathbb{E}^k$, the half-spaces are utilized in a different manner. At iteration n, a surrogate constraint is derived from a group of violated constraints and a_{n+1} is obtained by relaxed projection of a_n onto the hyperplane defined by this constraint. Although this method is serial, the determination of the surrogate constraint can be performed in parallel.

Set theoretic formulations involving closed affine hyperslabs can be handled by the previous algorithms since such sets can be written as the intersection of two affine half-spaces. Alternatively, in $\mathbb{E}^k$ one can employ the serial automatic relaxation algorithm proposed in [28], which takes advantage of the interval structure of the sets.

C. Convex Sets in Hilbert Spaces

It is assumed that Ξ is hilbertian and that all the S_ιs are closed and convex.

1) Projection Methods: Let us first consider the serial projection scheme

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = P_{\iota_n}(a_n). \tag{37}$$

Brègman has established that this algorithm satisfies $(\mathcal{P}_3)$ under most remote set and cyclic controls [17].[10] Moreover, it satisfies $(\mathcal{P}_2)$ if the control is cyclic and one of the sets is boundedly compact[11] [192]. $(\mathcal{P}_2)$ also holds in the case of chaotic control provided that one of the sets is compact [21]. A more general serial projection method can be obtained by introducing relaxation parameters

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda_n(P_{\iota_n}(a_n) - a_n) \tag{38}$$

where the sequence $(\lambda_n)_{n\geq 0}$ lies in an arbitrary but fixed interval $[\epsilon_1, \epsilon_2] \subset$]0, 2[. Moreover, it is assumed that the sets are activated under most remote set or cyclic control. Algorithm (38) was first proposed in $\mathbb{E}^k$ as a direct generalization of the Agmon-Motzkin-Schoenberg relaxation method and shown to satisfy $(\mathcal{P}_2)$ in [74]. In arbitrary Hilbert spaces, it is known to satisfy $(\mathcal{P}_3)$ [92], [221]. It also satisfies $(\mathcal{P}_2)$ if any of the following conditions holds [92]:

1) The dimension of Ξ is finite.
2) All, with the possible exception of one, of the S_ιs are δ-uniformly convex: there exists a nondecreasing

[9] In (34), the direction of movement towards the half-space S_ι is orthogonal to S_ι. Let us mention that there exist so-called ellipsoidal methods where the direction of movement to the next iterate is determined by a variable metric matrix updated at each step. The convergence properties of such schemes are discussed in [88].

[10] Brègman also gives a version of this result for arbitrary topological vector spaces in [18] by introducing so-called D-projections that generalize metric projections.

[11] A set in (Ξ, d) is said to be boundedly compact if its intersection with every closed ball is compact [16]. It is noted that every closed subset of $\mathbb{E}^k$ is boundedly compact.

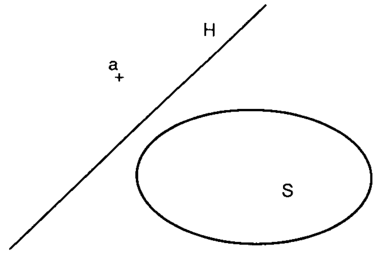


Fig. 6. Hyperplane H separating the point a from the convex set S.

function $\delta : \mathbb{R} \to [0, +\infty[$ that vanishes only at 0 such that $(\forall c \in \Xi)(\forall (a,b) \in S_\iota^2)\|c - (a+b)/2\| \leq \delta(\|a-b\|) \Rightarrow c \in S_\iota$.
3) $(\cap_{\iota \in I} S_\iota)^\circ \neq \emptyset$.
4) All the S_ιs are closed affine half-spaces.

In the literature, the cyclic version of (38) is often referred to as POCS, for projections onto convex sets. When direct projections onto some property sets are difficult to perform, an alternative scheme is to replace P_{ι_n} by the projection onto a hyperplane separating a_n from S_{ι_n} in (38) whenever $a_n \notin S_{\iota_n}$ (Fig. 6 illustrates the concept of separating hyperplane). This approach was adopted in [3], where convergence property $(\mathcal{P}_2)$ was established for chaotic control in $\mathbb{E}^k$ under the provision that, for every ι in I, a point in $S_\iota^\circ \neq \emptyset$ be known.

The basic parallel algorithm for convex sets in $\mathbb{E}^k$ consisting in averaging the projections onto all the sets, that is

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = (1/m) \sum_{\iota \in I} P_\iota(a_n) \tag{39}$$

was shown to satisfy $(\mathcal{P}_2)$ in [7]. In [65], it was shown that the more general algorithm

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda \left(\sum_{\iota \in I} w_\iota P_\iota(a_n) - a_n \right) \quad \text{with } 0 < \lambda < 2 \tag{40}$$

and where the weights conform to (32), satisfies $(\mathcal{P}_3)$ in general and $(\mathcal{P}_2)$ if one of the S_ιs is compact or if all of the S_ιs are closed affine half-spaces.[12] This algorithm can be further extended by allowing varying relaxation coefficients, namely

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda_n \left(\sum_{\iota \in I} w_\iota P_\iota(a_n) - a_n \right). \tag{41}$$

In $\mathbb{E}^k$, if the sequence $(\lambda_n)_{n \geq 0}$ lies in $[\epsilon_1, \epsilon_2] \subset]0, 2[$, then (41) satisfies $(\mathcal{P}_2)$; in addition, if each λ_n is determined in terms of a_n as in (36) with $\alpha = 1$, the algorithm still satisfies $(\mathcal{P}_2)$ and accelerated convergence is achieved since the λ_ns can now attain large values [105]. In [154] (see also [153]), Pierra establishes an interesting connection between cyclic and parallel projection algorithms in Hilbert spaces by showing that some parallel algorithms in Ξ yield simple cyclic algorithms in the cartesian product space Ξ^m. This formalism leads to extrapolated relaxation coefficients with which (41) is reported to converge efficiently. In this case, (41) satisfies $(\mathcal{P}_3)$ in general and $(\mathcal{P}_2)$ if any of conditions 1)–3) holds [154]. Naturally, the static recursion (41) can be extended to a more flexible parallel algorithm in which the subfamily of property sets to be acted upon varies at each iteration. This leads to the general recursion

$$(\forall n \in \mathbb{N}) \qquad a_{n+1} = a_n + \lambda_n \left(\sum_{\iota \in I_n} w_{\iota,n} P_\iota(a_n) - a_n \right) \tag{42}$$

where $(I_n)_{n \geq 0}$ is a sequence of nonempty subsets of I and $((w_{\iota,n})_{\iota \in I_n})_{n \geq 0}$ a sequence of weight vectors such that for every n in $\mathbb{N}$

$$\sum_{\iota \in I_n} w_{\iota,n} = 1, \qquad \text{and} \qquad (\forall \iota \in I_n)\; w_{\iota,n} > 0. \tag{43}$$

It is clear that this recursion encompasses all the previous ones. For instance, letting $I_n = \{\iota_n\}$ in (42) yields (38) whereas letting $I_n = I$ yields (41). Now assume that $(\lambda_n)_{n \geq 0} \subset [\epsilon_1, \epsilon_2] \subset]0, 2[$. It is proven in [4] that (42)/(43) satisfies $(\mathcal{P}_2)$ in $\mathbb{E}^k$ under chaotic control provided that

$$(\forall \iota \in I) \qquad \sum_{n \geq 0} w_{\iota,n} = +\infty. \tag{44}$$

Under similar assumptions, $(\mathcal{P}_2)$ still holds in $\mathbb{E}^k$ if P_ι is replaced by the projection onto a hyperplane separating a_n from S_ι in (42) whenever $a_n \notin S_\iota$ [76]. This generalization simplifies computations whenever direct projections onto the S_ιs are not easily obtained. In arbitrary Hilbert spaces, (42)/(43) satisfies $(\mathcal{P}_3)$ under almost cyclic control provided that the $w_{\iota,n}$s stay bounded away from 0 [46]. In [147], general conditions on the set theoretic formulation [in particular 1)–4)] are given for strong convergence of the method under several control schemes (almost cyclic, most-remote set, chaotic). This study also considers extrapolated relaxation coefficients, as in [154]. An interesting geometric interpretation of (42)/(43) with nonnegative relaxation coefficients is that the search direction at iterate a_n belongs to the convex cone of vertex a_n generated by the points $(P_\iota(a_n))_{\iota \in I_n}$. This is illustrated in Fig. 7.

2) Projection Methods for Closest Feasible Solution: It was seen in Section III-B that in the case of affine subspaces the algorithms (29) and (31)/(32) satisfy $(\mathcal{P}_1)$. For general convex sets, the algorithms of Section III-C1) are guaranteed to satisfy only $(\mathcal{P}_2)$, not necessarily $(\mathcal{P}_1)$. For instance, Fig. 8 depicts a simple situation when POCS does not yield the closest feasible point. This is not a problem since in the set theoretic framework any feasible point is an acceptable solution. Nonetheless, in some applications, a bound δ on the variations of the true object h from some reference point r may arise from physical considerations. This constraint confines estimates to lie in the ball of center r and radius δ. If a useful value of δ cannot be determined reliably, one

[12] Particular cases of this scheme are also discussed in [54], [133], and [160].

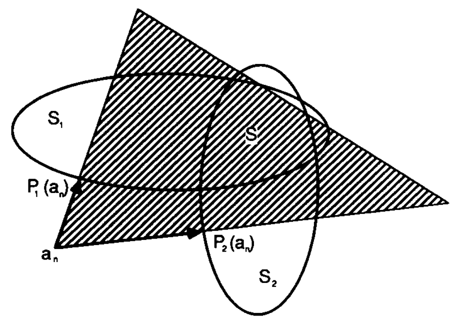


Fig. 7. Update region for a general convex projection algorithm.

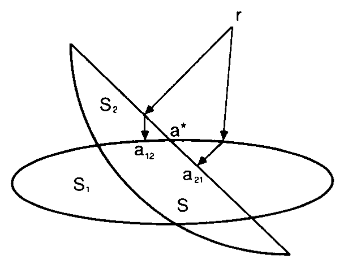


Fig. 8. The method of cyclic projections converges to either a_{12} or a_{21}. The closest feasible point is a^*.

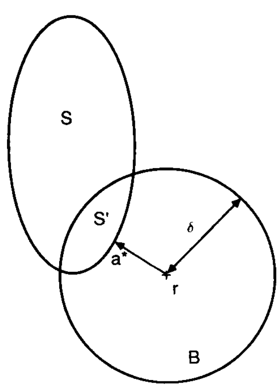


Fig. 9. The best feasible approximation a^* of r is always in $S' = S \cap B$, where B is any ball centered at r and intersecting with S.

can still exploit the constraint by choosing as a solution the feasible point that lies nearest r, i.e., the projection of r onto S. Indeed, such a point will be guaranteed to lie in S and in any ball centered at r and intersecting with S (see Fig. 9).

Direct projection onto an intersection of closed and convex sets can be achieved via the serial algorithms of [15], [80], and [94] and the parallel algorithm of [106], which all satisfy ($\mathcal{P}_1$). These algorithms are closely related to the feasibility algorithms of Section III-C1), to which they add little computational complexity. In [43], they are further discussed and applied to set theoretic signal recovery by best feasible approximation of a reference signal. The problem can also be approached via the extrapolated parallel method of [153].

3) Other Schemes: A shortcoming of projection methods is the numerical tedium sometimes involved in computing the projections at each iteration. A projection of a point a in (Ξ, d) onto S_ι is a point in S_ι that yields a global minimum the function $d(a, \cdot)$. It is obtained by solving

$$\min d(a, b) \text{ subject to } b \in S_\iota. \tag{45}$$

In some cases, this constrained minimization problem is easily solved. For instance, the euclidean projection operator onto the hyperplane (28) is simply given by [82]

$$(\forall a \in \mathbb{R}^k) \qquad P_\iota(a) = a + \frac{\delta_\iota - \langle a | b_\iota \rangle}{\langle b_\iota | b_\iota \rangle} b_\iota. \tag{46}$$

Algorithms are also available for special cases such as polyhedrons [5], cones [67], or polytopes [214]. Now, suppose that $\Xi = \mathbb{E}^k$ and that an equation is known for the boundary of S_ι, e.g., $\partial S_\iota = \{a \in \mathbb{R}^k | g_\iota(a) = 0\}$. Since the projection of a point a in $\complement S_\iota$ onto S_ι belongs to ∂S_ι, (45) can be put in the form

$$\min \|a - b\|^2 \text{ subject to } g_\iota(b) = 0. \tag{47}$$

This problem can be approached via the method of Lagrange multipliers [128]. Oftentimes, (47) must be solved iteratively (e.g., [52], [185], and [201]). An alternative iterative scheme to compute approximate projections is to define a sequence of simpler sets $(S_{\iota,k})_{k \geq 0}$ converging to S_ι and such that each projection $P_{\iota,k}(a)$ is easily determined. Then, under certain condition, $(P_{\iota,k}(a))_{k \geq 0}$ will converge to $P_\iota(a)$ [171].

As was seen in Section III-C1) the problem of computing exact projections can be circumvented in some cases by projecting serially [3] or simultaneously [76] onto separating hyperplanes rather than directly onto the sets. In addition, methods that do not require projections exist for set theoretic formulations of the form

$$(\forall \iota \in I) \qquad S_\iota = \{a \in \mathbb{R}^k | g_\iota(a) \leq 0\} \tag{48}$$

where $(g_\iota)_{\iota \in I}$ are convex functions on $\mathbb{R}^k$. A method consists of moving in the direction of the subgradient[13] of each g_ι evaluated at the current iterate, in a cyclic manner [33]. A modified version of this method that finds a solution in a finite number of steps is proposed in [66] and its parallel counterpart, in which the update consists of a convex linear combination of the subgradients of each g_ι, is proposed in [107]. Other schemes exist for set theoretic formulations of type (48), as discussed in [27].

[13] A vector t in $\mathbb{R}^k$ is said to be a subgradient of a function $g_\iota : \mathbb{R}^k \to \mathbb{R}$ at a point a if

$$(\forall b \in \mathbb{R}^k) \ \langle t | b - a \rangle \leq g_\iota(b) - g_\iota(a).$$

If g_ι is differentiable, its gradient $\nabla g_\iota(a)$ is the only subgradient at a.

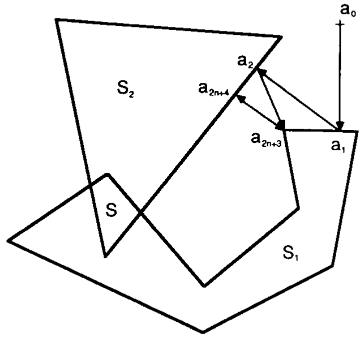


Fig. 10. Cyclic projections may fail to converge to a feasible point in the presence of nonconvex sets. Here, the subsequences $(a_{2n+3})_{n\geq 0}$ and $(a_{2n+4})_{n\geq 0}$ are constant and do not belong to S.

D. Towards Arbitrary Sets

The feasibility algorithms described above apply only to set theoretic formulations such that Ξ is a Hilbert space in which all the S_ιs are closed and convex. Feasibility algorithms applicable to more general set theoretic formulations are needed, since, as was seen in Section II-D, there may not always exist a hilbertian solution space in which all the information can be associated with closed and convex sets.

1) Projection Methods: In the presence of nonconvex sets and/or nonhilbertian spaces, the methods of Section III-C are no longer guaranteed to satisfy $(\mathcal{P}_2)$ or $(\mathcal{P}_3)$ for any initial point a_0. For instance, Fig. 10 shows a simple example in $\mathbb{E}^2$ when the basic cyclic projection scheme (29) fails to produce a feasible point for a particular choice of a_0 (it is easily seen that other schemes in Section III-C would face the same problem).

The informal use of cyclic projections in the presence of nonconvex sets in $\mathbb{E}^k$ is reported in several places in the literature, e.g., [23], [96], [102], [126], [164], and [189], and a local convergence statement of type $(\mathcal{P}_4)$ appears in [24]. A rigorous convergence analysis of the cyclic projection algorithm in arbitrary metric spaces is provided in [49].[14]

When dealing with projections in a metric space (Ξ, d), it must be borne in mind that a point in Ξ may admit several or no projections onto a given set. By the projection map onto a subset S_ι of Ξ, we mean the set-valued map

$$\begin{aligned} \Pi_\iota : \Xi &\to \mathfrak{P}(\Xi) \\ a &\mapsto \{b \in S_\iota | d(a,b) = d(a, S_\iota)\}. \end{aligned} \quad (49)$$

Thus, the set of projections of a onto S_ι is $\Pi_\iota(a)$.[15]

In [49], it is shown that if the sets $(S_\iota)_{\iota \in I}$ are approximately compact,[16] if one of them, say S_1, is boundedly compact, and if S is nonempty and bounded, then the serial algorithm

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad & a_{n+1} \in \Pi_{\iota_n}(a_n), \\ \text{where} \quad & \iota_n = n\,(\text{modulo}\, m) + 1 \end{aligned} \quad (50)$$

satisfies $(\mathcal{P}_2)$ or $(\mathcal{P}_4)$ locally, i.e., if a_0 lies in a certain region of the space; in addition, if $\Xi = \mathbb{E}^k$, the above conditions are simply equivalent to all the S_ιs being closed and S bounded. The above recursion can be regarded as a set-valued version of (29). Similar local $(\mathcal{P}_2)/(\mathcal{P}_4)$ convergence properties can be established for the set-valued versions of some of the parallel projection algorithms of Section III-C1), which take the general form (24) where R_n is an affine function of the Π_ιs [46].

For particular set theoretic formulations in which all the sets have the form $S_\iota = \{a \in \mathbb{R}^k | g_\iota(a) = 0\}$, the locally convergent algorithms discussed in [136] and in the references therein are of interest.

2) Random Search Method: A method for generating set theoretic estimates in $\mathbb{R}^k$, which circumvents all the theoretical (e.g., convexity, local convergence) and computational (e.g., computation of projections) limitations of the above techniques, is the adapted method of random search proposed in [51]. In this Monte Carlo method, points are generated at random over an adaptively reduced search region until a feasible point is found. The main advantage of this approach is its basic simplicity and the lack of assumptions about the geometrical complexity of the sets. It should however be borne in mind that, because of computational cost, this method is not suited for large-scale applications.

3) Approximation of the Feasibility Set: Set theoretic estimation methods have been developed, in which one does not seek to obtain a feasible point but an approximation of the exact feasibility set S by a simpler set, e.g., an ellipsoid or a hyperparallelepiped. It will be seen in Section V-A that this approach has become extremely popular in the area of systems theory. In the case of the "bounding-ellipsoid" algorithm, the approximating set is an ellipsoid containing S whose center is usually chosen as a solution. It should be stressed that, theoretically, such a solution is not guaranteed to be feasible, especially if the ellipsoid does not tightly approximate S. Besides the applications of Section V-A, the bounding-ellipsoid approach has also been utilized in signal restoration [113].

[14] A review of [24] and [49] is proposed in [104].

[15] Let us clarify some points about existence and uniqueness of projections that are often misunderstood in the engineering literature. A nonempty subset S_ι of (Ξ, d) is said to be proximinal if $(\forall a \in \Xi)(\exists b \in S_\iota) d(a,b) = d(a, S_\iota)$ [i.e., $\Pi_\iota(a) \neq \emptyset$] and is said to be a Chebyshev set if $(\forall a \in \Xi)(\exists! b \in S_\iota) d(a,b) = d(a, S_\iota)$ [i.e., $\Pi_\iota(a)$ is a singleton]. S_ι is necessarily closed to be proximinal since no point in $\bar{S}_\iota \cap \complement S_\iota$ admits a projection onto S_ι. In finite-dimensional normed vector spaces, closedness is also sufficient to ensure proximinality. This is no longer true in infinite dimension (see [173] for an example of nonempty, bounded, and closed set that is not proximinal). As was seen earlier, in Hilbert spaces, nonempty closed and convex sets are Chebyshev sets. In $\mathbb{E}^k$, the converse is true. However, in general, Chebyshev sets need not be convex, even in a pre-Hilbert space [108]. In a nonhilbertian space, closed and convex sets need not be Chebyshev sets: in the Banach space obtained by endowing $\mathbb{R}^2$ with the norm $\|a\| = |a_1| + |a_2|$, take $S_\iota = B[0,1]$ and $a = (1,1)$. Then $\Pi_\iota(a) = \{(\alpha, 1-\alpha) | 0 \leq \alpha \leq 1\}$. For more details on proximinal and Chebyshev sets, see [16] and [49].

[16] A subset S_ι of (Ξ, d) is said to be approximately compact if, for every a in Ξ, every sequence $(b_n)_{n\geq 0}$ of points in S_ι such that $(d(a,b_n))_{n\geq 0}$ converges to $d(a, S_\iota)$ possesses a cluster point in S_ι. For nonempty sets, compactness $\Rightarrow$ bounded compactness $\Rightarrow$ approximative compactness $\Rightarrow$ proximinality $\Rightarrow$ closedness [16], [49].

E. Inconsistent Set Theoretic Formulations

It was remarked in Section II-G that in some problems the set theoretic formulation may be inconsistent. It is therefore of interest to know which of the above feasibility algorithms converge in such instances, and, if they do, what are the properties of the limit point. As we shall see, some serial algorithms converge cyclically to a point in one of the sets while some parallel algorithms converge to a weighted least-squares solution. Throughout this section, Ξ is a Hilbert space, and the S_ιs are closed and convex.

Let G be the set of global minimizers of the functional

$$\begin{aligned} \Phi : \Xi &\to [0, +\infty[\\ a &\mapsto \sum_{\iota \in I} w_\iota d(a, S_\iota)^2 \end{aligned} \tag{51}$$

where the weights $(w_\iota)_{\iota \in I}$ satisfy (32). The set G is closed, convex, and possibly empty. It is clear that if $S \neq \emptyset$, then $G = S$ since

$$(\forall a \in \Xi) \qquad \Phi(a) = 0 \Leftrightarrow (\forall \iota \in I)\, a \in S_\iota. \tag{52}$$

On the other hand, if $S = \emptyset, G$ can be viewed as the set of weighted least-squares solutions of the inconsistent feasibility problem. For convenience, we introduce three additional convergence properties.

- $\mathcal{P}5$): For every a_0 in $\Xi, (a_n)_{n \geq 0}$ converges to the projection of a_0 onto G.
- $\mathcal{P}6$): For every a_0 in $\Xi, (a_n)_{n \geq 0}$ converges to a point in G.
- $\mathcal{P}7$): For every a_0 in $\Xi, (a_n)_{n \geq 0}$ converges weakly to a point in G.

Thus, $(\mathcal{P}_5)$–$(\mathcal{P}_7)$ coincide with $(\mathcal{P}_1)$–$(\mathcal{P}_3)$ in the consistent case and generalize them otherwise.

1) Serial Methods: A serial algorithm such as (38) cannot converge if $S = \emptyset$ since it will oscillate indefinitely. However, in the case of set theoretic formulations of type (28), for every ι in I, the subsequence $(a_{mn+\iota})_{n \geq 0}$ of $(a_n)_{n \geq 0}$ generated by (38) satisfies $(\mathcal{P}_6)$ (with $w_\iota = 1/m$) provided that λ_n goes to zero [30]. Now assume that $(a_n)_{n \geq 0}$ is a sequence generated by (37) under cyclic control with arbitrary convex sets. If $m = 2$ and S_2 is either compact or finite dimensional, the subsequence $(a_{2n})_{n \geq 0}$ of $(a_n)_{n \geq 0}$ converges (strongly) to a point that minimizes $d(\cdot, S_1)$ over S_2 [37].[17] For a generalization of this result, we now follow [92]. Suppose that one of the S_ιs is bounded. Then there exists at least one m-tuple $(\bar{a}_\iota)_{\iota \in I}$ in Ξ^m such that $P_1(\bar{a}_m) = \bar{a}_1$ and, for every ι in $\{2, \cdots, m\}, P_\iota(\bar{a}_{\iota-1}) = \bar{a}_\iota$. The $\bar{a}_\iota$s are called stationary points. Moreover, for every ι in I, the subsequence $(a_{mn+\iota})_{n \geq 0}$ of $(a_n)_{n \geq 0}$ converges weakly to $\bar{a}_\iota$; the convergence is strong is any of the conditions 1)–4) of Section III-C1) holds. This result is illustrated in Fig. 11(a).

2) Parallel Methods: For set theoretic formulations of type (33), the parallel algorithm (35) with (32) and $(\lambda_n)_{n \geq 0}$ in $[\epsilon_1, \epsilon_2] \subset]0, 2[$ satisfies $(\mathcal{P}_6)$ [64]. For arbitrary convex

[17]This problem is also considered in [89] and [220].

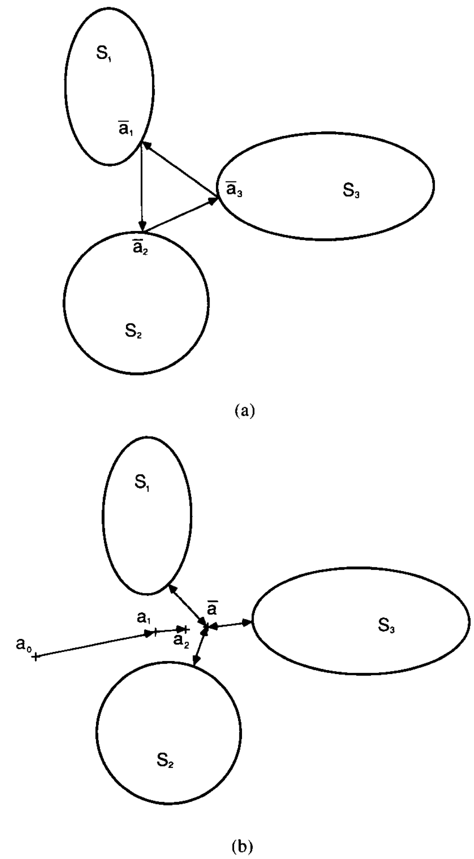


Fig. 11. Convergence with inconsistent set theoretic formulations: (a) serial algorithm: the subsequence $(a_{3n+\iota})_{n \geq 0}$ of the sequence of cyclic projections $(a_n)_{n \geq 0}$ converges to the stationary point $\bar{a}_\iota, 1 \leq \iota \leq 3$; (b) parallel algorithm: the sequence of averaged projections $(a_n)_{n \geq 0}$ converges to a point $\bar{a}$, which minimizes the average of the squares of the distances to the sets.

sets, (40) with (32) satisfies $(\mathcal{P}_7)$ if one of the S_ιs is bounded; in addition, it satisfies $(\mathcal{P}_6)$ if one of the S_ιs is compact or if all of the S_ιs are closed affine half-spaces [65]. This convergence behavior is depicted in Fig. 11(b). For $\alpha = 1$ and $\Xi = \mathsf{E}^k$, the parallel algorithm (35)/(36) satisfies $(\mathcal{P}_6)$ locally [105]. Finally, $(\mathcal{P}_5)$ is satisfied by the parallel algorithm of [106] discussed in Section III-C2).

F. Selection of an Algorithm

Given the relatively large number of methods relevant to the synthesis of set theoretic estimates, it is necessary to establish some guidelines for the selection of a feasibility algorithm.

The first step in the selection of an algorithm is to determine whether or not the set theoretic formulation consists of closed convex sets in a Hilbert space, say E^k in practice. If so, the algorithms of Section III-C (and of Section III-B if the sets are defined by affine subspaces) are guaranteed to produce a feasible solution. If seeking a direct projection onto the feasibility set is justified, one should turn to the algorithms of Section III-C2). The methods of

Section III-C3) do not rely on projections and are of interest in instances when (45) cannot be solved efficiently for all the sets.

In choosing a serial—as opposed to parallel—algorithm, several factors need be considered. As regards to convergence rate, studies pertaining to specific algorithms are available [87], [92], [99], [105], [115], [132], [154]. The unrelaxed POCS algorithm (37) is probably not the best choice although, to date, it has been the most widely used in applications. Despite the fact that, in certain specific problems, faster convergence of serial methods can be achieved by introducing relaxations, as in (38), parallel methods such as (42) are inherently more versatile and have a higher potential for fast convergence if the relaxation parameters are properly adapted. Moreover, as was seen in Section III-E2), in the case of inconsistent set theoretic formulations, they possess the remarkable property of converging to an approximate solution which satisfies all the constraints in a least-squares sense. This behavior is much more satisfactory than that of serial methods which, at best, converge to a point that is guaranteed to lie only in one of the sets [see Section III-E1)]. Another advantage of parallel methods is that they lend themselves naturally to implementations on concurrent processors. They are however more demanding than serial methods in terms of storage requirements, which may constitute a drawback in some applications. Recent developments in the parallel implementation of some of the algorithms of Section III-B with applications in medical imaging can be found in [29].

If the set theoretic formulation does not meet the above requirements (e.g., at least one of the sets is not convex), one should employ one of the methods of Section III-D. It should be noted that the projection methods of Section III-D1) guarantee only local convergence. Thus, they may not produce a feasible point if the initial estimate where the iterations are started is not sufficiently close to S. If the number of parameters to be estimated is not too large, this limitation can be circumvented by using the method of random search of Section III-D2).

IV. Connections with Other Estimation Procedures

Generally speaking, an estimation procedure can be viewed as a mapping $\mathcal{T}$ which assigns to the observed data x a subset $\mathcal{T}(x)$ of the solution space Ξ, that is,

$$\begin{aligned} \mathcal{T} : \Delta &\rightarrow \mathfrak{P}(\Xi) \\ x &\mapsto \mathcal{T}(x) \end{aligned} \tag{53}$$

where Δ is the observation space, i.e., the space to which the mathematical object representing the observed data belongs.[18] The set of points $\mathcal{T}(x)$ represents our guess of the value of the true state of nature h given the data x and some *a priori* knowledge. In the set theoretic estimation framework proposed in this paper, $\mathcal{T}(x) = S$ is defined by (4). We shall now examine how other estimation methods proceed in the determination of $\mathcal{T}(x)$ and how they relate to the present framework.

A. Point Estimation

The result of a point estimation procedure is a point $a(x)$ rather than a set. This point is typically obtained by optimizing some preset objective function. In most cases, the optimum is unique and, therefore, the set of solutions is a singleton $\mathcal{T}(x) = \{a(x)\}$. Various optimality criteria have been proposed for point estimates, e.g., maximum likelihood, minimax, maximum entropy, maximum *a posteriori* and other Bayesian criteria [9], [11], [112], [168].

Conceptually, the Bayesian approach shares with the present set theoretic framework its attempt to exploit the observed data as well as *a priori* knowledge to carry out the decision process. Indeed, the posterior distribution involved in the computation of a Bayesian estimate comprises a prior distribution, which reflects the *a priori* information on the true state of nature h, and a likelihood function, which reflects sample information. The introduction of a prior distribution on Ξ, which is the basis for the Bayesian viewpoint, has caused extreme disagreements among statisticians and the reader is referred to [11], [72] and the references therein for recent developments on the controversy. In our opinion, a major concern *vis à vis* the incorporation of *a priori* knowledge in the Bayesian approach is that h is regarded as a realization of a random element and, therefore, a probability theoretic modeling of prior information is required. Unfortunately, not all *a priori* information can be conveniently described in probabilistic terms. Moreover, the resulting prior distribution is usually too complex to yield a tractable optimization of the resulting conditional expectation.

The point estimation approach differs fundamentally from that of the present set theoretic framework since a point, rather than a set, is selected as the solution to the problem. Theoretically, the former can nevertheless be regarded as set theoretic, the set of solutions being that of all points that satisfy the chosen optimality criterion. From a more practical viewpoint, the implicit set theoretic nature of point estimates can be brought to light by noting that, except in the rare cases when a closed-form solution is available, point estimates are computed iteratively. Unless convergence can be achieved in finite number of steps, a stopping rule $\mathcal{R}$ must be specified to terminate the iterations. This procedure leads implicitly to the set of solutions

$$S = \{a \in \Xi | a \text{ satisfies } \mathcal{R}\}. \tag{54}$$

It is noted that if $\mathcal{R}$ is a purely analytical criterion this set, and hence the properties of a solution, may not be well defined;[19] on the other hand, stopping rules based on physical constraints can lead to a well defined set [127], [194], [199], [205].

[18] For completeness, it should be noted that in statistics the data model is probabilistic. If $(\Omega, \mathcal{A}, P)$ denotes the underlying probability space, the observed data are regarded as a realization $x = X(\omega)$ of a random element $X : \Omega \rightarrow \Delta$ and a set theoretic estimator is a measurable map $\mathcal{M} = \mathcal{T} \circ X : \Omega \rightarrow \mathfrak{S}$, where $\mathfrak{S}$ is a σ-algebra of subsets of Ξ [9].

[19] For instance, a widespread stopping rule for an algorithm generating a sequence $(a_n)_{n \geq 0}$ in a metric space (Ξ, d) is $d(a_n, a_{n+1}) \leq \epsilon$.

B. Confidence Regions

In statistics, confidence regions constitute a well-established set theoretic method of estimation [9], [19], [168]. Let $(\Omega, \mathcal{A}, (\mathrm{P}_h)_{h\in\Xi})$ be the underlying statistical model and let ω be the elementary event giving rise to the observation $X(\omega) = x$ of the data process. A confidence region is a subset $\mathcal{T}(x) = \mathcal{T}(X(\omega))$ of Ξ such that[20]

$$(\forall h \in \Xi) \qquad \mathrm{P}_h\{\omega \in \Omega | h \in \mathcal{T}(X(\omega))\} \geq \alpha. \tag{55}$$

The theory of confidence regions provides a means to construct property sets, $S_\iota = \mathcal{T}(x)$, based on probabilistic information. Therefore, although the set theoretic estimation framework is nonstatistical in the sense that is does not require a probabilistic data model, it can effectively utilize statistical theory if such a model is assumed.

Besides the observed data, the construction of a confidence region requires some *a priori* knowledge. The most common procedures exploit the distribution of a point estimator of the estimandum [19]. For sake of simplicity, consider the problem of estimating a real parameter h (a similar procedure applies to the multidimensional case). Let $a(x)$ be the observed value of an asymptotically normal and relatively unbiased estimator of h with variance σ^2 (under certain conditions, the maximum likelihood estimator will satisfy these assumptions). Then, for large sample sizes, a confidence region is $\mathcal{T}(x) = [a(x) - \gamma\sigma, a(x) + \gamma\sigma]$, where γ can be obtained from the tables of the normal distribution in terms of α [19]. This example also demonstrates that statistical point estimators can play an active rôle in set theoretic estimation by providing sets to be added to the set theoretic formulation. Confidence regions need not be based only on point estimators of the estimandum. They can be based on more general statistics, as was done in Section II-F1) to construct property sets from stochastic information.

C. Set-Valued Bayesian Estimation

A notorious shortcoming of conventional Bayesian estimation is the lack of robustness of the end result with respect to the specifications of the problem, i.e., the prior distribution and the loss function [11]. Thus, the specification of a prior constitutes a critical step in which one is required to choose a single distribution that will properly model all *a priori* knowledge. A way to relax this requirement while remaining faithful to the Bayesian philosophy is to consider a set of prior distributions, each of which is an equally acceptable candidate to model the uncertainty surrounding the true state of nature. An estimate is obtained for each prior distribution and $\mathcal{T}(x)$ is therefore a set. Such a generalization of classical Bayesian inference has been proposed in the 1960s (e.g., see [63]). It is also discussed in [193] and applied to Bayesian filtering in [140], where a set-valued Kalman filter based on convex sets of distributions is developed.

[20] Let us stress that the inequality in (55) concerns the probability that the random set $\mathcal{T}(X)$ contains a given point h.

V. Applications

The set theoretic framework has been applied, in various forms, to a vast number of engineering problems. The early applications are found in the area of control in the 1960s [175] and then in image reconstruction in the 1970s [99]. From the mid-70s to the mid-80s, applications in other areas of signal recovery were reported [187]. Since the mid-80s, one witnesses a steady increase in the number of applications in fields as diverse as filter design [1], array signal processing [24], electron microscopy [26], speckle interferometry [71], antenna array design [73], topography [137], spectral estimation [148], neural networks [181], and color systems [200].

A large number of problems treated within the set theoretic framework is surveyed in this section. The emphasis is placed on showing the great versatility and the generality of set theoretic formulations rather than on the technicalities regarding particular implementations.[21] For a detailed account of the latter, the reader is referred to the cited references.

A. Systems Theory

In the engineering literature, the set theoretic approach seems to have been first applied to systems theory as a nonstatistical way to incorporate uncertainty in modeling, analysis, estimation, and control problems. In this context, the basic idea of an estimation scheme that yields a set based on available information, rather than a single point, can be traced back to [174], and related concepts can be found in [176] and [213].

In the state estimation problem of [174], the main information used in constructing the feasibility set is that the noise, or more generally the various disturbances on the system, is bounded in amplitude. This work initiated a series of bounded-noise set theoretic methods that have been applied to various state estimation [169], [175], control [86], [116], [119], filtering [197], and identification [78], [103], [144], [159] problems. These methods are time recursive in that the feasibility set is updated with every new observed data sample, thus giving an iterative set theoretic algorithm. As regards to implementation, so called "bounding-ellipsoid" algorithms have been proposed to avoid the time-consuming computation of the exact feasibility set at each iteration by approximating it by the ellipsoidal superset

$$S_n = \{a \in \mathbb{R}^k | (a - a_n)^t R_n^{-1} (a - a_n) \leq 1\} \tag{56}$$

characterized by its center a_n and the positive definite matrix R_n [175]. The attractive features of these algorithms are their simplicity, their recursive nature, and their selective update strategy, which uses only data samples with sufficiently innovative information [57], [59], [78], [103]. However, the ellipsoidal approximation being somewhat loose, tighter approximations have been proposed for

[21] In passing, however, it should be pointed out that there are a few instances when POCS is wrongly invoked to justify the convergence of algorithms in that either the underlying norm is not hilbertian [(A10) is not satisfied] or the operators involved are not projections.

specific problems involving linear models, e.g., a hyper-parallelepiped approximation in [138] and [150], and an exact polyhedral approximation in [208]. Since the mid-80s, bounded noise set theoretic identification has become a major area of research in system identification. A detailed account of recent developments can be found in [62] and [207]. A tutorial review of some aspects of the field is proposed in [60].

Because they employ very specific information, namely noise boundedness, these set theoretic methods lead to very simple set theoretic formulations where exact or approximative descriptions of the feasibility set are available and can be efficiently updated with new data samples. This feature makes them applicable to "on-line" problems such as those encountered in system identification. Nonetheless, although the noise boundedness assumptions is quite reasonable in many applications, a tight bound may not be available. In these instances, given that the methods rely primarily on this piece of information, the solution set will not be very restrictive, especially if the data record is short.

Time-recursive set theoretic algorithms similar to those discussed above have also been developed from an energy—rather than amplitude—constraint on the noise [12], [77]. In [140], time-recursive set-valued filtering and smoothing is developed in a generalized Kalman filtering context based on the framework of Section IV-C.

B. Spectral Estimation and Related Fields

The spectral estimation problem is to estimate the spectral distribution of a stochastic process $(X_i)_{i\in\mathbb{Z}}$ from a finite number of observations. This basic problem and its offsprings are of great importance in many branches of statistical sciences [20] and engineering [114]. As will be seen shortly, the introduction of the set theoretic formalism in spectral estimation is relatively recent.

In a variety of applications, the data are modeled as a sum of unknown sinusoids in additive noise, namely

$$(\forall i \in \mathbb{Z}) \quad X_i = \sum_{j=1}^{p} h_{1,j} \sin(2\pi h_{2,j} i + h_{3,j}) + U_i. \quad (57)$$

In [51], a set theoretic formulation for estimating the vector $h = (h_{1,1}, \cdots, h_{3,p})$ of parameters (amplitudes, frequencies, phases) is proposed. The solution space is $\mathbb{R}^{3p}$ and the set theoretic formulation comprises sets based on parameter bounds as well as on noise properties. Because the analytical complexity of the sets precludes the use of projection techniques, the method of random search of Section III-D2) is utilized to generate a set theoretic estimate.

A different set theoretic approach to the problem of recovering the frequencies of the p sinusoids in (57) is proposed in [24]. This approach exploits the fact that, ideally, an appropriately formed data matrix should satisfy certain constraints (Toeplitz, Hankel, maximum rank). Property sets based on these constraints are constructed in a matrix solution space. A set theoretic estimate is found by applying (50)[22] with the noisy data matrix as an initial estimate and the frequencies are obtained by Fourier transforming the resulting feasible data. In [24], set theoretic formulations involving matrix solution spaces and similar matrix constraint sets (Toeplitz, Hermitian, maximum rank) are also shown to be effective for applications involving exponential modeling of data as well as in array signal processing.[23] In the latter case, a sample covariance matrix consistent with rank q (q being the number of sources impinging on the array) and Hermitian-Toeplitz constraints is obtained via (50) and the MUSIC bearing estimation method [170] is then applied to it.

A parametric model of great interest in spectral estimation is the autoregressive model (11). It reduces the estimation of the spectral distribution of the underlying process to that of the vector $h = (h_1, \cdots, h_k)$ in $\mathbb{R}^k$. Several standard methods proceed by minimization of various estimates of the prediction error power in the regression space (autocorrelation, covariance, and modified covariance methods) or in the reflection space (Burg's method) [114]. In the time-recursive set theoretic approach of [103], an optimal bounding ellipsoid algorithm is developed in the space of regression coefficients under the assumption that a uniform bound on the driving process $(U_i)_{i\in\mathbb{Z}}$ is available. In [48], the set theoretic formulation incorporates more information. It consists of (12) and of sets constructed from various properties assumed to be known *a priori* about $(U_i)_{i\in\mathbb{Z}}$, e.g., mean, power, whiteness. The more general problem of autoregressive moving average spectral estimation is formulated in a matrix solution space in [148], where sets based on various structural and spectral matrix properties are utilized.

C. Signal Recovery

Generally speaking, the term *signal recovery* refers to a large class of inverse problems where an original signal is to be estimated from data consisting of one or more signals physically related to it. To date, signal recovery is undeniably the field that has seen the largest number of applications of set theoretic estimation. This can be explained by the fact that, because most recovery problems are ill-posed, the incorporation of available information will greatly improve their solutions. Signal recovery problems fall into two main categories: signal restoration and signal reconstruction. The goal of signal restoration is to estimate an original signal from measurements of the signal obtained by some sensor, the measurements being taken directly on the signal to be restored. On the other hand, in the case of signal reconstruction, the data is indirectly related to the form of the signal. For example, the term *restoration* would apply to the case of estimating an original image from measurements of a blurred and noisy version of it; the term *reconstruction* would apply to estimating an original image from measurements taken of its Fourier transform.

[22] The set of matrices of rank q or less is not convex.

[23] A detailed review of such applications can be found in [104].

Signal recovery problems are particularly amenable to the set theoretic approach because there is a great deal of qualitative information about the original signal that is not easily expressed in purely statistical terms, which is the only form conventional estimation methods can exploit. For instance, suppose that a portion of an image to be recovered is known. While it is possible to construct conditional probability distributions that include this knowledge, such a strategy is not practical for images of realistic sizes [204]. On the other hand, the corresponding property set is easily constructed as

$$S_\iota = \{a \in \Xi | a1_K = h1_K\} \quad (58)$$

where 1_K is the indicator function of the region K over which the true image h is known. Likewise, subjective qualities such as smoothness can be included by using a bound on the spatial derivative of the image. Finding an appropriate bound for such a set can be based on edge information [182] or statistical deviation from a smooth prototype [184]. Impulsiveness, as might be found in astronomical images or X-ray fluorescence spectra, can be defined in a set theoretic sense by limiting the number of nonzero values within a given area. The constraints on the original signal that are often encountered in recovery problems include band limitedness, space limitedness, intensity range, energy boundedness, nonnegativity, sparsity, piecewise constancy, and partial knowledge of the Fourier transform (examples of other properties of physical significance and their associated sets can be found in [180], [185], [187], and [221]).

In n-dimensional signal recovery, the natural solution space is L_n^2 for continuous models, e.g., [221] and $\mathbb{R}^k$ for discrete models (by representing a tensor as a vector via stacked vector notations), e.g., [201]. The structure of the natural solution space can be modified in order to render some property sets convex. This is done in [36] where the set of discrete signals with a prescribed Fourier transform magnitude, which is not convex in ℓ^2, is made convex in a new sequence space, $\ell^\star$, obtained by redefining both addition and scalar multiplication. Likewise, the set of sequences with prescribed bispectrum value at a given frequency pair, which is not convex in ℓ^2, can be shown to be convex in $\ell^\star$ [34]. A set theoretic formulation can also be posed in a solution space different from the natural signal space by using alternative signal representations. For instance, in [130], the reconstruction of an L_1^2-signal from the zero crossings of its wavelet transform is posed in the solution space (8). Other example are found in [166], where an image is restored in a singular value solution space, and in [211], where an L_1^2-signal is reconstructed from its general bilinear time-frequency representation in L_2^2.

Comparative studies of conventional and set theoretic signal recovery methods can be found in [146] and [196].

1) Reconstruction of Fourier Transform Pairs: Let $h : x \mapsto h(x)$ denote a signal and $H : \nu \mapsto H(\nu)$ its Fourier transform.[24] A common problem in many fields is the reconstruction of a Fourier transform pair (h, H) from partial information on either or both functions [75], [95]. A set theoretic interpretation of this problem is to find a point in the intersection of the sets representing the temporal (or spatial) and spectral information.

Several methods have been proposed in the literature to recover h iteratively by enforcing time and Fourier domain constraints in an alternating fashion, one of the earliest being [122]. In [83], the underlying physically problem is to reconstruct an object from intensity measurements in the image plane (spatial information) and the diffraction plane (spectral information). The proposed reconstruction method is to alternate resubstitutions of the known magnitude data in both domains. Set theoretically, this can be interpreted as a cyclic projection algorithm onto the sets

$$S_1 = \{a \in L_1^2 | (\forall x \in \mathbf{R})\, |a(x)| = g(x)\} \quad (59)$$

and

$$S_2 = \{a \in L_1^2 | (\forall \nu \in \mathbf{R})\, |A(\nu)| = G(\nu)\} \quad (60)$$

where g and G are known functions. Almost at the same time, a method of alternating projections onto the property sets S_1 and

$$S_3 = \{a \in L_1^2 | (\forall \nu \in \mathbf{R})\, |A(\nu)| = G1_K(\nu)\} \quad (61)$$

(where K is a known frequency support, and G is a known constant) was applied to kinoform design [81]. According to the results of Section III-D1), convergence of such schemes is guaranteed only locally since the set theoretic formulations are not convex. In [84], a finite extent object is extrapolated from limited diffraction data (i.e., a prescribed portion of the Fourier transform) via what can be interpreted as an application of unrelaxed POCS (37) to the affine subspaces

$$S_4 = \{a \in L_1^2 | (\forall x \in \mathbf{R})\, a(x) = 0 \text{ if } |x| \geq b\} \quad (62)$$

and

$$S_5 = \{a \in L_1^2 | (\forall \nu \in \mathbf{R})\, A(\nu) = G(\nu) \text{ if } |\nu| \leq B\} \quad (63)$$

where G is a prescribed function. The same method is applied to the dual problem of extrapolating a band-limited signal known on a compact support in the time domain in [149]. A more general signal recovery framework involving alternating projections onto affine subspaces of Hilbert spaces is proposed in [219]. In that abstract approach, the recovery problem is posed as that of recovering a signal h knowing that h belongs to a closed subspace of Ξ and that the observed data consist of the projection of h onto another closed subspace of Ξ. This framework found further extensions and additional applications in set theoretic signal recovery from spatial and spectral information [188], [221].

In [125], the relaxed POCS method (38) was used to reconstruct a time-limited signal from a prescribed Fourier phase with the sets S_4 and

$$S_6 = \{a \in L_1^2 | (\forall \nu \in \mathbf{R})\, \angle A(\nu) = \phi(\nu)\} \quad (64)$$

where ϕ is a prescribed function. A relaxed version of (50) was employed in [126] to reconstruct a time-limited

[24] For convenience, the notations refer to the one-dimensional case.

signal from a prescribed Fourier magnitude, leading to a nonconvex set theoretic formulation consisting of S_2 and S_4. Such phase retrieval problems were approached in [8] with parallel projections techniques similar to (41). In [36], a Hilbert solution space was constructed by modifying the natural vector space structure in order to render S_2 convex. This made possible the use of POCS to reconstruct a minimum phase signal from the knowledge of its Fourier magnitude and phase at a finite number of frequencies.

2) Image Reconstruction from Projections: The problem of image reconstruction from projections[25] is to estimate a multidimensional function from recorded values of its line integrals, usually obtained by passing energy rays through an object. This problem arises in a large number of fields, e.g., nondestructive testing, seismology, satellite remote sensing, and, most notably, diagnostic medicine, where cross-sectional images of the human body are reconstructed from measurements of the attenuation of X-rays along lines through the cross section [99].

With proper discretization, the reconstruction problem can be posed as a system of linear equations and gives rise to an affine set theoretic formulation of type (28). The original approach based on this formulation is the so-called algebraic reconstruction technique (ART) of [90] that uses Kaczmarz's method (29) (see also [91]). An alternative method to solve this formulation [the Cimmino-like parallel algorithm of (39)] is proposed in [85] under the name simultaneous iterative reconstruction technique (SIRT) and reported to perform better the original ART of [90] in noisy environments. Because a linear set theoretic formulation ignores noise and other uncertainty sources, it may be unfair or even inconsistent. A more realistic approach is discussed in [98] in which the property sets are hyperslabs of the form $\{a \in \mathbb{R}^k | \delta_i - \epsilon_i \leq \langle a|b_i\rangle \leq \delta_i + \epsilon_i\}$, where ϵ_i is a tolerance factor. This leads to a set theoretic formulation of type (33) solved by the Agmon-Motzkin-Schoenberg algorithm (34) ([98], [99], [101]).

Reconstructions must often be performed with limited view data, i.e., with inaccurately measured projections and/or an insufficient number of projections, which will typically result in severe artifacts such as streaking and geometric distortion [158]. In such instances, the set theoretic approach has proven particularly well suited to incorporate *a priori* knowledge and thereby improve the reconstruction. Thus, a convex set theoretic formulation is used to extrapolate tomographic images reconstructed from a limited range of views in [124] and [177]. In [178], the formulation of [177] is modified to account for noisy data. In [179], POCS is combined with the method of direct Fourier tomography to reconstruct an image from limited-view projection data. Strictly speaking, these approaches are not set theoretic reconstruction methods *per se* but, rather, syntheses of a reconstruction method and a set theoretic restoration method. In that sense, they should not be regarded as extensions of ART (or SIRT). In ART, the property sets simply translate the requirement that the reconstruction be consistent with the observed projections. In [145], a more sophisticated convex set theoretic formulation was developed by incorporating additional constraints such as known object support and energy boundedness. Other types of constraints can also be imposed, such as consistency of the error between the recorded projection data and the data obtained by reprojecting the reconstructed image with the uncertainty caused by the numerical approximations of the reprojection method [203].

In all of the above studies, the solution space is that of the reconstructed image. In [120], a different set theoretic approach is proposed in which the solution space is the space of Radon transforms of images. A complete set of line integrals consistent with *a priori* knowledge and the measured line integrals is first obtained by POCS and then used to reconstruct the image via ordinary convolution backprojection. In [209], POCS is used to synthesize the projection matrix from noisy measurements made by a moving array of detectors and the image is then reconstructed by filtered backprojection.

3) Signal Restoration: The most common signal restoration problem is to estimate the original form h of a blurred and noise-corrupted signal x. A general degradation model assumes that the blurring opertor T is linear and that the noise u is additive, which yields the data formation equation

$$x = T(h) + u. \qquad (65)$$

Besides the properties of the original signal, the available information in such a problem may consist of information about the blur and the noise. If such information is not known *a priori*, it can often be estimated from the data [183]. In [201], it is demonstrated how a wide variety of convex property sets could be constructed from noise properties. In [42], some of these sets are reexamined in the context of fuzzy set theory. The stochastic nature of some blurring functions such as atmospheric turbulence and camera vibration has also been addressed using set theoretic methods [47]. Set theoretic restoration in the presence of bounded kernel disturbances and noise was considered in [53]. Sets based on locally adaptive constraints [121] as well as on smoothness constraints [190] have also been proposed. In addition, set theoretic restoration has been used with other statistically based methods. Since a Wiener estimate is commonly computed for image restoration problems, the estimate can be used to define a convex set [182], [184]. The bounds for this set can be determined from the standard statistics available for the Wiener filter [184].

In many cases, set theoretic restoration techniques have been demonstrated on one-dimensional signals. These signals usually model some physical process. A typical one-dimensional (1-D) example is data from X-ray fluoroscopy [42], [49], [201]. This signal has many properties that make it an ideal candidate for set theoretic methods; it is sparse, impulsive with only a few nonzero points. In [49], the impulsiveness property is modeled by limiting the number of nonzero values in the signal. This resulted in a nonconvex set and (50) was used to obtain a solution. Other types of signals that have been restored with set theoretic methods,

[25] The term *projection* here refers to a line integral projection, not to be confused with the projection onto a set.

including multiband satellite images [38], character images [123], echographic images [129], diffraction wave fields [139], optical flow fields, and electromagnetic fields [186]. In order to best exploit specific *a priori* information, the set theoretic restoration problem of [166] was posed in a singular value space rather than in the natural image space.

4) Other Recovery Problems: Besides the problems mentioned, set theoretic estimation can also be credited for applications in problems such as signal reconstruction from level crossings [56], [223], signal reconstruction from the zero crossings of the wavelet transform [130], signal reconstruction from multiscale edges [131], signal reconstruction from the bispectrum [34], signal reconstruction from bilinear time-frequency representation [211], signal reconstruction from Q-distributions [212], acoustic signal reconstruction from auditory representations [216], signal reconstruction from nonuniform samples [167], [217], recovery of the angular energy spectrum of an object imaged through a turbulent atmosphere [71], reconstruction of images remotely sensed by image-plane detector arrays [191], image reconstruction from digital holograms [133], image reconstruction in emission computerized tomography [32], signal recovery in electron microscopy [26], and inversion of eddy current data and reconstruction of flaws in composite materials [163].

D. Design Problems

The problems described so far can be labeled as estimation problems for they consist in guessing the value of an object that, for a given a model, actually gave rise to the observed data. The set theoretic framework has also proven very useful in solving design problems. In this context, design constraints or requirements on the object to be synthesized are associated with fuzzy propositions and, via (3), with a set theoretic formulation. It should be noted that inconsistent set theoretic formulations are more frequent in design problems than in estimation problems. Indeed, a design formulation is primarily based on the desirata of the user, which may be conflicting. In this respect, the methods of Section III-E2), which converge to a weighted least-squares solution, are valuable tools to generate a design that best approximates uncompatible constraints.

A common synthesis problem is that of digital filter design and several studies have been devoted to its set theoretic treatment. In the set theoretic design of two-dimensional (2-D) FIR linear phase filters of [1], a family of property sets in the solution space of filter coefficients is constructed by constraining the amplitude of the frequency response at specified frequency points to lie within some neighborhood of the desired response and a solution is obtained via (38). In [35], the same problem is revisited by imposing constraints in both time and frequency domains. Other set theoretic formulations in the FIR coefficient space involving time and frequency domains constraints for specific problems have also been investigated [40], [143], [151]. In [25], a matrix solution space is proposed to recursively approximate an ideal 2-D frequency response. Alternating projections onto property sets of nullity one and block Toeplitz matrices are used to synthesize a feasible excitation-response matrix from which the recursive filter coefficients are computed. Finally, let us note that in the somewhat more recent field of color systems, several approaches for the set theoretic design of scanning filters are proposed in [200].

Various other design problems have been treated in the set theoretic framework. In the area of optics, let us mention cohoe design [102] and phase grating design [189]. A set theoretic design of data windows for spectral estimation in the presence of inconsistent requirements is proposed in [89]. In [142] and [164], set theoretic projection methods are used to design (construct) images in connection with the problem of compensating for various distortion processes. The set theoretic approach has also been employed in various antenna [157] and antena array [22], [23], [73], [156] design problems. The set theoretic synthesis of ambiugity functions is discussed in [212].

E. Miscellaneous

There are several set theoretic studies that do not fit directly in the main categories discussed above. These include [61], where an optimal bounding ellipsoid algorithm (see Section V-A) is developed for linear speech prediction; [134], [135], and [181], where a set theoretic framework and the formalism of projections are used to analyze vrious aspects of the dynamic behavior of neural networks; [200], where set theoretic methods are applied to various problems in color science; [137], where POCS is utilized to interpolate topographic profiles, topographic maps, and physical properties of the earth; [161], where POCS is applied to optical pattern recognition; [165], where image coding is posed as a set theoretic problem; [96], where cyclic projections are used to suppress the artifacts caused by patient motion in magnetic resonance imaging; [97] and [198], where POCS is used in analog-to-digital conversion; and [118], where the estimation of the parameters of a multipath channel is based on a set theoretic formulation.

VI. Further Discussion and Conclusions

In the tradition of recent decades, a good solution to an estimation or design problem has been one that is optimal in some sense. However, as "optimal" estimators for "exact" models can rarely be implemented, some reservations can be expressed *vis à vis* this estimation setting. Of primary concern are the facts that the selection of the objective function is generally driven by computational tractability rather than rational considerations and that limitations are imposed in the incorporation of available information. As a result, the actual properties of such estimates are seldom related to physical realities and are somewhat elusive. Moreover, given the inherently uncertain environments in which most estimation problems are posed, the practical value of optimality claims is questionable.

In this paper, we have presented a synthetic view of set theoretic estimation. In order to lay a secure foundation for further theoretical research and build a common framework for all existing set theoretic approaches, some formalization was necessary. Fuzzy propositions were employed to model the wide range of information encountered in estimation problems. In this context, a property set was defined as the cut of the fuzzy set whose membership function is the fuzzy proposition modeling a particular piece of information. This conceptual definition has the advantage of being quite general and flexible. Nonetheless, it was pointed out that it should not be taken literally as a constructive one since, in practice, the fuzzy formalism is often by-passed in the process of defining property sets. A set theoretic formulation for the problem was defined as the family of all property sets in a given solution space. It provides a complete description of a set theoretic estimation problem and constitutes a valuable tool in connection with various theoretical and practical questions, from the analysis of the problem to the synthesis of a solution.

The basic philosophical motivation for the set theoretic approach is that more reliable solutions can be obtained by exploiting known information rather than imposing an often subjective notion of optimality. Thus, in the set theoretic framework, the emphasis is placed on the feasibility of a solution rather than its optimality, as in done in the conventional approach. The goal is not to produce a "best" solution but one that is consistent with all available information. In set theoretic estimation, all the members of the feasibility set are acceptable solutions. They can be regarded as the objects that, in light of all available information, may have given rise to the observed data. The only way to restrict objectively the feasibility set is to incorporate more information in the formulation. If some of the feasible solutions are not acceptable, then it must be the case that the formulation fails to include some constraint that has not been identified. Once the set based on this constraint is incorporated, any point in the feasibility set should be acceptable; if not the cycle is repeated. Usually, there is more than one solution, which may be counterintuitive from the standpoint of conventional point estimation theory where, to extract a single solution, an objective function with a unique extremum is employed. On the other hand, because of the arbitrariness in the selection of such an objective function, the result is, at best, nothing but a qualitative selection of a feasible solution.

From a practical standpoint, the main asset of set theoretic estimation is the availability of mathematical methods to solve the basic feasibility problem (23). In fact, historically, the level of sophistication of set theoretic formulations, which reflects not only the complexity of the incorporated information but also the refinement of the underlying data model, has always been limited by the availability of feasibility algorithms. This point can be illustrated by considering the evolution of set theoretic formulations in signal recovery. The early set theoretic formulations were limited to linear varieties and solved by Kaczmarz or Cimmino-like algorithms [85], [90]. With the Agmon-Motzkin-Schoenberg algorithm, they evolved to include half-spaces [98]. As the POCS algorithms of Brègman and Gubin *et al.* became known in image processing circles, the restriction to half-spaces disappeared, allowing the use of more general convex set theoretic formulations [124], [221]. More recently, a theoretical analysis of the convergence of cyclic nonconvex and nonhilbertian projections rationalized the inclusion of nonconvex property sets [49]. At present, though, nonconvex set theoretic formulations remain a significant difficulty, and the development of better feasibility algorithms than those discussed in Section III-D remains a critical step towards broadening even further the scope of set theoretic estimation.

At this point, it should be remarked that set theoretic and conventional estimation theory can be used jointly to solve a problem. Indeed, even if it is sometimes at the expense of rationally posing the estimation problem, a definite advantage of some conventional estimation methods is to yield simple problem formulations and, in some cases, closed-form solutions (if necessary, one can always have recourse to standard cost functions and assumptions to simplify the problem). This expedient approach has at least the merit of leading to a solution. On the other hand, although there are methods for computing set theoretic estimates for a wide class of problems, setting up a tractable set theoretic formulation may not always be possible. Even if it is, the use of whatever conventional estimation method seems appropriate should not be precluded, especially if a solution can be computed efficiently. This solution can then be tested for feasibility with respect to available information, which represents a relatively easy and computationally inexpensive task. If it is feasible, it must be accepted; if not, one should expect a set theoretic solution to bring improvement.

Certainly, set theoretic estimation is not immune from criticism. Its principles are often criticized on the grounds that the end result is not a unique object. We believe that enough has been said in this paper to dismiss such a claim. A more serious criticism is that the construction of the property sets is subjective since the choice of the fuzzy propositions $(\Psi_\iota)_{\iota\in I}$ and, more importantly, of the grades of beliefs $(\psi_\iota)_{\iota\in I}$ in (3) are eventually left to the user. First, it should be noted that this problem arises only in the case of information that is not modeled by crisp propositions. In the other cases, there is no doubt that ψ_ι may indeed be interpreted as the user's personal degree of conviction that the estimandum is consistent with the information modeled by Ψ_ι. On that score, our contention is that any estimation procedure that allows the incorporation of information will be exposed to some degree of subjectivism. This inherent subjectivism can, however, be mitigated by using experience as a guide to determine realistic values for the ψ_ιs.

As demonstrated in Section V, set theoretic estimation has been applied successfully to a wide spectrum of problems. Based on the trend of the past twenty years and the on-going research in the field, on both theoretical and applied questions, it can safely be anticipated that the number of applications will keep growing in increasingly varied areas. Nonetheless, set theoretic estimation is still in its infancy and has yet to be accepted in many scientific disciplines. In this regard, it is hoped that this paper will contribute to consolidate its position as a reliable alternative to the conventional framework of optimization that has traditionally ruled over estimation problems in sciences and engineering.

VII. APPENDIX

The purpose of this appendix is to provide the basic definitions of mathematical analysis needed in the paper. Standard references on this topic are Dieudonné [68], Schwartz [173], and Yosida [218]. Readers interested in an authoritative account of set theory, its history, and its rôle as a basic structure in modern mathematics are referred to Bourbaki [14].

A. Set Theory

The quantifiers, $\forall, \exists$, and $\exists!$ mean "for all," "there exists at least one," and "there exists exactly one," respectively. $\emptyset$ denotes the set with no elements (empty set). Let Ξ be a nonempty set called space thereafter. The family of all subsets of Ξ is denoted by $\mathfrak{P}(\Xi)$. The elements of Ξ are called points. The relation $a \in \Xi$ means that a is an element of Ξ. Its negation is written $a \notin \Xi$. The relation $S \subset \Xi$ means that every element of the set S is an element of Ξ; S is then called a subset of Ξ, and Ξ a superset of S. $\{a \in \Xi | a \text{ satisfies } \Psi\}$ is the set of all points a in the space Ξ that satisfy a given property Ψ. Let $I \subset \mathbb{R}$ be an nonempty index set and let $(S_\iota)_{\iota \in I}$ be a family of subsets of Ξ. Set union, intersection, and complementation are, respectively, defined as

$$\begin{cases} \bigcup_{\iota \in I} S_\iota = \{a \in \Xi | (\exists \iota \in I)\, a \in S_\iota\} \\ \bigcap_{\iota \in I} S_\iota = \{a \in \Xi | (\forall \iota \in I)\, a \in S_\iota\} \\ \complement S_\iota = \{a \in \Xi | a \notin S_\iota\} \end{cases} \tag{A1}$$

The indicator function of the set S is the function 1_S, which takes value 1 on S and 0 on $\complement S$. The Cartesian product of two spaces Ξ and Ξ' is $\Xi \times \Xi' = \{(a, a') | a \in \Xi, a' \in \Xi'\}$. The expressions $T : \Xi \to \Xi'$ and $T : a \mapsto T(a)$ mean that T is a mapping from Ξ into Ξ' and that T assigns $T(a)$ to a, respectively. Let $\mathfrak{S}$ denote a family of subsets of Ξ. A function $\nu : \mathfrak{S} \to [-\infty, +\infty]$ is said to be monotone if

$$(\forall (S, S') \in \mathfrak{S}^2) \qquad S \subset S' \Rightarrow \nu(S) \leq \nu(S'). \tag{A2}$$

B. Metric Spaces

A function $d(\cdot, \cdot) : \Xi \times \Xi \to [0, +\infty[$ is called a distance (or metric) on Ξ if

1) $(\forall (a, b) \in \Xi^2)\, d(a, b) = 0 \Leftrightarrow a = b$
2) $(\forall (a, b) \in \Xi^2)\, d(a, b) = d(b, a)$
3) $(\forall (a, b, c) \in \Xi^3)\, d(a, c) \leq d(a, b) + d(b, c)$

The pair (Ξ, d) is called a metric space. The diameter of a nonempty subset S of Ξ is defined as $\delta(S) = \sup\{d(a, b) | a \in S, b \in S\}$. S is said to be bounded if $\delta(S) < +\infty$. The distance from a point a to S is defined as $d(a, S) = \inf\{d(a, b) | b \in S\}$.

Let $c \in \Xi$ and $r \in]0, +\infty[$. The open and closed balls of center c and radius r are, respectively, defined as $B[c, r[= \{a \in \Xi | d(c, a) < r\}$ and $B[c, r] = \{a \in \Xi | d(c, a) \leq r\}$. $S \subset \Xi$ is open if

$$(\forall a \in S)(\exists r \in]0, +\infty[)\ B[a, r[\subset S, \tag{A3}$$

and closed if $\complement S$ is open. Any intersection of closed sets is closed. The interior of S is the largest open set S° contained in S; the closure of S is the smallest closed set $\bar{S}$ containing S.

Let $(a_n)_{n \geq 0}$ and a be points in Ξ. Then $(a_n)_{n \geq 0}$ converges to a if $(d(a_n, a))_{n \geq 0}$ converges to 0, i.e.,

$$(\forall r \in]0, +\infty[)(\exists p \in \mathbb{N})(\forall n \in \mathbb{N} | n \geq p)\, a_n \in B[a, r[. \tag{A4}$$

$S \subset \Xi$ is closed if every convergent sequence with elements in S has its limit in S. It is said that a is a cluster point of $(a_n)_{n \geq 0}$ if there exists a subsequence $(a_{n_k})_{k \geq 0}$ of $(a_n)_{n \geq 0}$ converging to a. $S \subset \Xi$ is compact if every sequence with elements in S admits at least one cluster point in S. We call $(a_n)_{n \geq 0}$ a Cauchy sequence if $(d(a_m, a_n))_{m,n \geq 0}$ converges to 0 as m and n go to $+\infty$, i.e.,

$$(\forall r \in]0, +\infty[)(\exists p \in \mathbb{N})(\forall m \in \mathbb{N} | m \geq p) \cdot (\forall n \in \mathbb{N} | n \geq p)\, d(a_m, a_n) < r. \tag{A5}$$

(Ξ, d) is called complete if every Cauchy sequence in Ξ converges to a point in Ξ.

C. Normed Vector Spaces

A vector space$(\Xi, +, \cdot)$ over a field $\mathbb{K}$ is a space Ξ of object called vectors endowed with an operation $+ : \Xi \times \Xi \to \Xi$ called addition such that

1) $(\forall (a, b) \in \Xi^2)\ a + b = b + a$
2) $(\forall (a, b, c) \in \Xi^3)\ a + (b + c) = (a + b) + c$
3) $(\exists! 0_\Xi \in \Xi)(\forall a \in \Xi)\ a + 0_\Xi = a$
4) $(\forall a \in \Xi)(\exists! (-a) \in \Xi)\ a + (-a) = 0_\Xi$

and an operation $\cdot : \mathbb{K} \times \Xi \to \Xi$ called scalar multiplication such that

5) $(\forall \alpha \in \mathbb{K})(\forall (a, b) \in \Xi^2)\ \alpha \cdot (a + b) = \alpha \cdot a + \alpha \cdot b$
6) $(\forall (\alpha, \beta) \in \mathbb{K}^2)(\forall a \in \Xi)\ (\alpha + \beta) \cdot a = \alpha \cdot a + \beta \cdot a$
7) $(\forall (\alpha, \beta) \in \mathbb{K}^2)(\forall a \in \Xi)\ (\alpha\beta) \cdot a = \alpha \cdot (\beta \cdot a)$
8) $(\forall a \in \Xi)\ 1 \cdot a = a$

where 1 is the unit element of $\mathbb{K}$. From now on, the symbol $\cdot$ will be omitted in scalar multiplications and $\mathbb{K}$ will be $\mathbb{R}$ or $\mathbb{C}$. Let Ξ' be another vector space. An operator $T : \Xi \to \Xi'$ is said to be linear if

$$(\forall \alpha \in \mathbb{K})(\forall (a, b) \in \Xi^2) \qquad T(\alpha a + b) = \alpha T(a) + T(b) \tag{A6}$$

and is called a functional if $\Xi' = \mathbb{K}$. Let S be a nonempty subset of Ξ. S is a vector subspace if

$$(\forall \alpha \in \mathbb{K})(\forall (a, b) \in S^2)\ \alpha a + b \in S \tag{A7}$$

and an affine subspace if $S = \{a + b | a \in V\}$, where V is a vector subspace and b a vector in Ξ. S is balanced if $(\forall \alpha \in \mathbb{K})(\forall a \in S)\ |\alpha| \leq 1 \Rightarrow \alpha a \in S$, and convex if

$$(\forall \alpha \in]0, 1[)(\forall (a, b) \in S^2)\ \alpha a + (1 - \alpha) b \in S. \tag{A8}$$

Any intersection of convex sets is convex. The convex hull of S is the smallest convex set containing S. A norm on Ξ is a function $\|\cdot\| : \Xi \to [0, +\infty[$ such that

1) $(\forall a \in \Xi)\ \|a\| = 0 \Leftrightarrow a = 0_\Xi$
2) $(\forall \alpha \in \mathbb{K})(\forall a \in \Xi)\ \|\alpha a\| = |\alpha| \cdot \|a\|$
3) $(\forall (a,b) \in \Xi^2)\ \|a + b\| \leq \|a\| + \|b\|$

$(\Xi, \|\cdot\|)$ is called a normed vector space (NVS). A norm $\|\cdot\|$ induces a distance via the relation

$$(\forall (a,b) \in \Xi^2)\ d(a,b) = \|a - b\| \tag{A9}$$

Thus, every NVS is a metric vector space. A linear functional $T : \Xi \to \mathbb{K}$ is said to be bounded if $\sup\{|T(a)|\, | a \in B[0,1]\} < +\infty$. In a NVS, $(a_n)_{n\geq 0}$ converges strongly to a if $(\|a_n - a\|)_{n\geq 0}$ converges to 0 and weakly if $(T(a_n - a))_{n\geq 0}$ converges to 0 for every bounded linear functional T on Ξ. If a sequence converges strongly to a point, it converges weakly to that point. In finite dimensional spaces, the converse is also true. A Banach space is a complete NVS. Every finite dimensional NVS is a Banach space. A scalar product on Ξ is a function $\langle\cdot|\cdot\rangle : \Xi \times \Xi \to \mathbb{K}$ that satisfies

1) $(\forall a \in \Xi)\ a \neq 0_\Xi \Rightarrow \langle a|a\rangle > 0$
2) $(\forall \alpha \in \mathbb{K})(\forall (a,b) \in \Xi^2)\ \langle \alpha a|b\rangle = \alpha\langle a|b\rangle$
3) $(\forall (a,b,c) \in \Xi^3)\ \langle a + b|c\rangle = \langle a|c\rangle + \langle b|c\rangle$
4) $(\forall (a,b) \in \Xi^2)\ \langle b|a\rangle = \overline{\langle a|b\rangle}$

where $\bar{z}$ denotes the complex conjugate of z in 4) above. A pre-Hilbert space is a vector space Ξ endowed with a scalar product. In a pre-Hilbert space $(\Xi, \langle\cdot|\cdot\rangle)$, the scalar product induces a norm as follows: $(\forall a \in \Xi)\ \|a\| = \sqrt{\langle a|a\rangle}$. Moreover, the norm is characterized by

$$(\forall (a,b) \in \Xi^2)\ \ \|a + b\|^2 + \|a - b\|^2 = 2(\|a\|^2 + \|b\|^2). \tag{A10}$$

Thus, a pre-Hilbert space is a NVS. Two vectors a and b are said to be orthogonal if $\langle a|b\rangle = 0$. A Hilbert space is a complete pre-Hilbert space. If $(\Xi, \langle\cdot|\cdot\rangle)$ is a Hilbert space, $(a_n)_{n\geq 0}$ converges weakly to a if and only if $(\langle a_n - a|b\rangle)_{n\geq 0}$ converges to 0, for every b in Ξ.

ACKNOWLEDGMENT

The author wishes to acknowledge the contribution of Dr. Reha Civanlar from AT&T Bell Labs who coauthored a paper presented at ICASSP'91 that discussed several of the ideas found in this work.[26] He also wishes to thank Prof. Joel Trussell from North Carolina State University for valuable suggestions regarding several aspects of this paper, and to acknowledge fruitful discussions with Mony from the University of Nancago, and Dr. Pascal Bondon from the Laboratoire des Signaux et Systèmes. Part of this paper was written while the author was visiting the CNRS, laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France, during the summer of 1991 with the support of the Université de Paris-Sud.

[26] P. L. Combettes and M. R. Civanlar, "The foundations of set theoretic estimation," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, May 14–17, 1991, pp. 2921–2924.

REFERENCES

[1] A. Abo-Taleb and M. M. Fahmy, "Design of FIR two-dimensional digital filters by successive projections," *IEEE Trans. Circuits Syst.*, vol. 31, no. 9, pp. 801–805, Sept. 1984.

[2] S. Agmon, "The relaxation method for linear inequalities," *Canadian J. Math.*, vol. 6, no. 3, pp. 382–392, 1954.

[3] R. Aharoni, A. Berman, and Y. Censor, "An interior points algorithm for the convex feasibility problem," *Advances Appl. Math.*, vol. 4, no. 4, pp. 479–489, Dec. 1983.

[4] R. Aharoni and Y. Censor, "Block-iterative methods for parallel computation of solutions to convex feasibility problems," *Linear Algebra and Its Applications*, vol. 120, pp. 165–175, Aug. 1989.

[5] M. Arioli, A. Laratta, and O. Menchi, "A big-M type method for the computation of projections onto polyhedrons," *J. Optimiz. Theory Applications*, vol. 47, no. 1, pp. 17–34, Sept. 1985.

[6] R. B. Ash, *Real Analysis and Probability*. New York: Academic, 1972.

[7] A. Auslender, *Optimisation—Méthodes Numériques*. Paris: Masson, 1976.

[8] R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, band-limited Fourier transform pairs from noisy data. II: The nonlinear problem of phase retrieval," *J. Integral Eqs.*, vol. 9, no. 1 (supplement), pp. 77–125, July 1985.

[9] J. R. Barra, *Notions Fondamentales de Statistique Mathématique*. Paris: Dunod, 1971.

[10] M. Benidir and B. Picinbono, "Nonconvexity of the stability domain of digital filters," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 38, no. 8, pp. 1459–1460, Aug. 1990.

[11] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, second edition. New York: Springer-Verlag, 1985.

[12] D. P. Bertkestas and I. B. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *IEEE Trans. Automat. Contr.*, vol. 16, no. 2, pp. 117–128, Apr. 1971.

[13] N. Bourbaki, *Eléments de Mathématique—Espaces Vectoriels Topologiques*. Paris: Masson, 1981.

[14] N. Bourbaki, *Eléments de Mathématique—Théorie des Ensembles*. Paris: Masson, 1990.

[15] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," *Lecture Notes in Statistics*, vol. 37, pp. 28–47, 1986.

[16] D. Braess, *Nonlinear Approximation Theory*. New York: Springer-Verlag, 1986.

[17] L. M. Brègman, "The method of successive projection for finding a common point of convex sets," *Soviet Mathematics—Doklady*, vol. 6, no. 3, pp. 688–692, May 1965.

[18] ——, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *U.S.S.R. Computational Math. Math. Physics*, vol. 7, no. 3, pp. 200–217, 1967.

[19] L. Breiman, *Statistics with a View towards Applications*. Boston, MA: Houghton Mifflin Co., 1973.

[20] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, second edition. New York: Springer-Verlag, 1991.

[21] R. E. Bruck, "Random products of contractions in metric and Banach spaces," *J. Math. Analysis Applications*, vol. 88, no. 2, pp. 319–332, Aug. 1982.

[22] O. M. Bucci, G. Franceschetti, G. Mazzarella, and G. Panariello, "Intersection approach to array pattern synthesis," *IEE Proc.-H*, vol. 137, no. 6, pp. 349–357, Dec. 1990.

[23] O. M. Bucci, G. Mazzarella, and G. Panariello, "Reconfigurable arrays by phase-only control," *IEEE Trans. Antennas Propagat.*, vol. 39, no. 7, pp. 919–925, July 1991.

[24] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.

[25] J. A. Cadzow and T. C. Chen, "Alebraic approach to two-dimensional recursive digital filter synthesis," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 37, no. 5, pp. 655–664, May 1989.

[26] J. M. Carazo and J. L. Carrascosa, "Information recovery in missing angular data cases: An approach by the convex projections method in three dimensions," *J. Microscopy*, vol. 145, pt. 1, pp. 23–43, Jan. 1987.

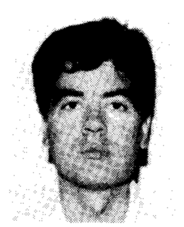
[27] Y. Censor, "Iterative methods for the convex feasibility problem," *Annals of Discrete Math.*, vol. 20, pp. 83–91, 1984.

[28] ——, "An automatic relaxation method for solving interval linear inequalities," *J. Math. Analysis Applications*, vol. 106, no. 1, pp. 19–25, Feb. 1985.
[29] ——, "Parallel application of block-iterative methods in medical imaging and radiation therapy," *Math. Programming*, vol. 42, no. 2, pp. 307–325, 1988.
[30] Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," *Numerische Mathematik*, vol. 41, no. 1, pp. 83–92, Apr. 1983.
[31] Y. Censor and T. Elfving, "New methods for linear inequalities," *Linear Algebra and Its Applications*, vol. 42, pp. 199–211, Feb. 1982.
[32] Y. Censor, D. E. Gustafson, A. Lent, and H. Tuy, "A new approach to the emission computerized tomography problem: Simultaneous calculation of attenuation and activity coefficients," *IEEE Trans. Nuclear Sci.*, vol. NS-26, no. 2, pp. 2775–2779, Apr. 1979.
[33] Y. Censor and A. Lent, "Cyclic subgradient projections," *Math. Programming*, vol. 24, no. 2, pp. 233–235, 1982.
[34] A. E. Çetin, "An iterative algorithm for signal reconstruction from bispectrum," *IEEE Trans. Signal Process.*, vol. 39, no. 12, pp. 2621–2628, Dec. 1991.
[35] A. E. Çetin and R. Ansari, "Iterative procedure for designing two-dimensional FIR filters," *Electronics Letters*, vol. 23, no. 3, pp. 131–133, Jan. 1987.
[36] ——, "Convolution-based framework for signal recovery and applications," *J. Optical Soc. America A*, vol. 5, no. 8, pp. 1193–1200, Aug. 1988.
[37] W. Cheney and A. A. Goldstein, "Proximity maps for convex sets," *Proc. American Math. Soc.*, vol. 10, no. 3, pp. 448–450, June 1959.
[38] R. T. Chin, C. L. Yeh, and W. S. Olson, "Restoration of multichannel microwave radiometric images," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 7, no. 4, pp. 475–484, July 1985.
[39] G. Cimmino, "Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari," *La Ricerca Scientifica (Roma)*, vol. 1, pp. 326–333, 1938.
[40] M. R. Civanlar and R. A. Nobakht, "Optimal pulse shape design using projections onto convex sets," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, New York, NY, Apr. 11–14, 1988, pp. 1874–1877.
[41] M. R. Civanlar and H. J. Trussell, "Constructing membership functions using statistical data," *Fuzzy Sets and Systems*, vol. 18, no. 1, pp. 1–13, Jan. 1986.
[42] ——, "Digital signal restoration using fuzzy sets," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 4, pp. 919–936, Aug. 1986.
[43] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Trans. Image Process.*, vol. 2, no. 2, Apr. 1993.
[44] P. L. Combettes, M. Benidir, and B. Picinbono, "A general framework for the incorporation of uncertainty in set theoretic estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, San Francisco, CA, vol. 3, Mar. 23–26, 1992, pp. 349–352.
[45] P. L. Combettes and T. J. Chaussalet, "Critères de qualité en estimation ensembliste," in *Proc. Thirteenth GRETSI Symp.*, Juan-les-Pins, France, Sept. 16–20, 1991, pp. 249–252.
[46] P. L. Combettes and H. Puh, "Parallel projection methods for set theoretic signal reconstruction and restoration," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Minneapolis, MN, vol. 5, Apr. 27–30, 1993, pp. 297–300.
[47] P. L. Combettes and H. J. Trussell, "Methods for digital restoration of signals degraded by a stochastic impulse response," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 37, no. 3, pp. 393–401, Mar. 1989.
[48] ——, "Set theoretic autoregressive spectral estimation," in *Proc. Fifth ASSP Workshop on Spectrum Estimation and Modeling*, Rochester, NY, Oct. 10–12, 1990, pp. 261–264.
[49] ——, "Method of successive projections for finding a common point of sets in metric spaces," *J. Optimiz. Theory Applications*, vol. 67, no. 3, pp. 487–507, Dec. 1990.
[50] ——, "The use of noise properties in set theoretic estimation," *IEEE Trans. Signal Process.*, vol. 39, no. 7, pp. 1630–1641, July 1991.
[51] ——, "Set theoretic estimation by random search," *IEEE Trans. Signal Process.*, vol. 39, no. 7, pp. 1669–1671, July 1991.
[52] ——, "Best stable and invertible approximations for ARMA systems," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 3066–3069, Dec. 1992.
[53] ——, "Deconvolution with bounded uncertainty," *Int. J. Adaptive Control Signal Process.*, accepted for publication.
[54] G. Crombez, "Image recovery by convex combinations of projections," *J. Math. Analysis Applications*, vol. 155, no. 2, pp. 413–419, Mar. 1991.
[55] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Annals Statistics*, vol. 19, no. 4, pp. 2032–2066, Dec. 1991.
[56] S. R. Curtis, A. V. Oppenheim, and J. S. Lim, "Signal reconstruction from Fourier transform sign information," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 33, no. 3, pp. 643–657, June 1985.
[57] S. Dasgupta and Y. F. Huang, "Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise," *IEEE Trans. Informat. Theory*, vol. 33, no. 3, pp. 383–392, May 1987.
[58] R. Dautray and J. L. Lions, *Analyse Mathématique et Calcul Numérique pour les Sciences et les Techniques.* Paris: Masson, 1984.
[59] J. R. Deller, Jr., "A 'systolic array' formulation of the optimal bounding ellipsoid algorithm," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 37, no. 9, pp. 1432–1436, Sept. 1989.
[60] ——, "Set membership identification in digital signal processing," *IEEE ASSP Mag.*, vol. 6, no. 4, pp. 4–20, Oct. 1989.
[61] J. R. Deller, Jr. and T. C. Luk, "Linear prediction analysis of speech based on set-membership theory," *Computer Speech Language*, vol. 3, no. 4, pp. 301–327, Oct. 1989.
[62] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, accepted for publication.
[63] A. P. Dempster, "A generalization of Bayesian inference (with discussion)," *J. Royal Statistical Soc.*, vol. B30, no. 2, pp. 205–247, 1968.
[64] A. R. De Pierro and A. N. Iusem, "A simultaneous projections method for linear inequalities," *Linear Algebra and Its Applications*, vol. 64, pp. 243–253, Jan. 1985.
[65] ——, "A parallel projection method for finding a common point of a family of convex sets," *Pesquisa Operacional*, vol. 5, no. 1, pp. 1–20, July 1985.
[66] ——, "A finitely convergent 'row-action' method for the convex feasibility problem," *Appl. Math. Optimiz.*, vol. 17, pp. 225–235, 1988.
[67] F. Deutsch, J. H. McCabe, and G. M. Phillips, "Some algorithms for computing best approximations from convex cones," *SIAM J. Numerical Analysis*, vol. 12, no. 3, pp. 390–403, June 1975.
[68] J. Dieudonné, *Foundations of Modern Analysis*, second ed. New York: Academic, 1969.
[69] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic, 1980.
[70] ——, "A review of fuzzy set aggregation connectives," *Informat. Sci.*, vol. 36, no. 1, pp. 85–121, July 1985.
[71] S. Ebstein, "Stellar speckle interferometry energy spectrum recovery by convex projections," *Appl. Optics*, vol. 26, no. 8, pp. 1530–1536, Apr. 1987.
[72] B. Efron, "Why isn't everyone a Bayesian?" *American Statistician*, vol. 40, no. 1, pp. 1–5, Feb. 1986.
[73] H. Elmikati and A. A. Elsohly, "Extension of projection method to nonuniformly linear antenna arrays," *IEEE Trans. Antennas Propagat.*, vol. 32, no. 5, pp. 507–512, May 1984.
[74] I. I. Eremin, "Generalization of the relaxation method of Motzkin-Agmon," *Uspekhi Matematicheskikh Nauk*, vol. 20, no. 2, pp. 183–187, Mar. 1965 (in Russian).
[75] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Optics*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.
[76] S. D. Flåm and J. Zowe, "Relaxed outer projections, weighted averages, and convex feasibility," *BIT*, vol. 30, no. 2, pp. 289–300, 1990.
[77] E. Fogel, "System identification via membership set constraints with energy constrained noise," *IEEE Trans. Automat. Contr.*, vol. 24, no. 5, pp. 752–758, Oct. 1979.
[78] E. Fogel and Y. F. Huang, "On the value of information in system identification—Bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, Mar. 1982.
[79] C. Franchetti and W. Light, "The alternating algorithm in uniformly convex spaces," *J. London Math. Soc.*, vol. 29, pp. 545–555, 1984.

[80] N. Gaffke and R. Mathar, "A cyclic projection algorithm via duality," *Metrika*, vol. 36, no. 1, pp. 29–54, 1989.
[81] N. C. Gallagher and B. Liu, "Method for computing kinoforms that reduces image reconstruction error," *Applied Optics*, vol. 12, no. 10, pp. 2328–2335, Oct. 1973.
[82] N. Gastinel, *Analyse Numérique Linéaire*. Paris: Hermann, 1966.
[83] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–246, Apr. 1972.
[84] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta*, vol. 21, no. 9, pp. 709–720, Sept. 1974.
[85] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theoretical Biol.*, vol. 36, no. 1, pp. 105–117, July 1972.
[86] J. D. Glover and F. C. Schweppe, "Control of linear dynamic systems with set constrained disturbances," *IEEE Trans. Automat. Contr.*, vol. 16, no. 5, pp. 411–423, Oct. 1971.
[87] J. L. Goffin, "The relaxation method for solving systems of linear inequalities," *Math. Operations Res.*, vol. 5, no. 3, pp. 388–414, Aug. 1980.
[88] ——, "Convergence of a cyclic ellipsoid algorithm for systems of linear equalities," *Math. Programming*, vol. 22, no. 3, pp. 239–260, 1982.
[89] M. Goldburg and R. J. Marks II, "Signal synthesis in the presence of an inconsistent set of constraints," *IEEE Trans. Circuits Syst.*, vol. 32, no. 7, pp. 647–663, July 1985.
[90] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theoretical Biology*, vol. 29, no. 3, pp. 471–481, Dec. 1970.
[91] R. Gordon and G. T. Herman, "Three-dimensional reconstruction from projections: A review of algorithms," *Int. Review Cytology*, vol. 38, pp. 111–151, 1974.
[92] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *U.S.S.R. Computational Math. Math. Physics*, vol. 7, no. 6, pp. 1–24, 1967.
[93] I. Halperin, "The product of projection operators," *Acta Scientiarum Mathematicarum (Szeged)*, vol. 23, no. 1, pp. 96–99, 1962.
[94] S. P. Han, "A successive projection method," *Math. Programming*, vol. 40, no. 1, pp. 1–14, Jan. 1988.
[95] M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 30, no. 2, pp. 140–154, Apr. 1982.
[96] M. Hedley, H. Yan, and D. Rosenfeld, "Motion artifact correction in MRI using generalized projections," *IEEE Trans. Medical Imaging*, vol. 10, no. 1, pp. 40–46, Mar. 1991.
[97] S. Hein and A. Zakhor, "Reconstruction of oversampled band-limited signals from $\Sigma\Delta$ encoded binary sequences," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, San Francisco, CA, vol. 4, Mar. 23–26, 1992, pp. 161–164.
[98] G. T. Herman, "A relaxation method for reconstructing objects from noisy X-rays," *Math. Programming*, vol. 8, no. 1, pp. 1–19, Feb. 1975.
[99] G.T. Herman, *Image Reconstruction from Projections, the Fundamentals of Computerized Tomography*. New York: Academic, 1980.
[100] G. T. Herman, "Mathematical optimization versus practical performance: A case study based on the maximum entropy criterion in image reconstruction," *Math. Programming Study*, vol. 20, pp. 96–112, Oct. 1982.
[101] G. T. Herman, A. Lent, and P. H. Lutz, "Relaxation methods for image reconstruction," *Communications ACM*, vol. 21, no. 2, pp. 152–158, Feb. 1978.
[102] R. G. Hoptroff, P. W. McOwan, T. J. Hall, W. J. Hossak, and R. E. Burge, "Two optimization approaches to cohoe design," *Optics Communications*, vol. 73, no. 3, pp. 188–194, Oct. 1989.
[103] Y. F. Huang, "A recursive estimation algorithm using selective updating for spectral analysis and adaptive signal processing," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 5, pp. 1331–1334, Oct. 1986.
[104] N. E. Hurt, "Signal enhancement and the method of successive projections," *Acta Applicandae Mathematicae*, vol. 23, no. 2, pp. 145–162, May 1991.
[105] A. N. Iusem and A. R. De Pierro, "Convergence results for an accelerated nonlinear Cimmino algorithm," *Numerische Mathematik*, vol. 49, no. 4, pp. 367–378, Aug. 1986.
[106] ——, "On the convergence of Han's method for convex programming with quadratic objective," *Math. Programming*, vol. 52, no. 2, pp. 265–284, Aug. 1991.
[107] A. N. Iusem and L. Moledo, "A finitely convergent method of simultaneous subgradient projections for the convex feasibility problem," *Mathemática Aplicada e Computacional*, vol. 5, no. 2, pp. 169–184, 1986.
[108] G. G. Johnson, "A nonconvex set which has the unique nearest point property," *J. Approximation Theory*, vol. 51, no. 4, pp. 289–332, Dec. 1987.
[109] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen," *Bulletin de l'Académie des Sciences de Pologne*, vol. A35, pp. 355–357, 1937.
[110] W. J. Kammerer and M. Z. Nashed, "A generalization of a matrix iterative method of G. Cimmino to best approximate solution of linear integral equations of the first kind," *Atti della Accademia Nazionale dei Lincei—Rendicoti*, vol. 51, pp. 20–25, 1971.
[111] A. Kandel, *Fuzzy Mathematical Techniques with Applications*. Reading, MA: Addison-Wesley, 1986.
[112] J. N. Kapur, "Twenty-five years of maximum-entropy principle," *J. Math. Physical Sci.*, vol. 17, no. 2, pp. 103–156, Apr. 1983.
[113] A. K. Katsaggelos, J. Biemond, R. W. Schafer, and R. M. Mersereau, "A regularized iterative image restoration algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 914–929, Apr. 1991.
[114] S. M. Kay, *Modern Spectrum Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[115] S. Kayalar and H. L. Weinert, "Error bounds for the method of alternating projections," *Math. Control, Signals, Systems*, vol. 1, no. 1, pp. 43–59, 1988.
[116] K. Keesman and G. Van Straten, "Identification and prediction propagation of uncertainty in models with bounded noise," *Int. J. Control*, vol. 49, no. 6, pp. 2259–2269, June 1989.
[117] P. Kosmol and X. L. Zhou, "The product of affine orthogonal projections," *J. Approximation Theory*, vol. 64, no. 3, pp. 351–355, Mar. 1991.
[118] Z. Kostić, M. I. Sezan, and E. L. Titlebaum, "Estimation of the parameters of a multipath channel using set-theoretic deconvolution," *IEEE Trans. Commun.*, vol. 40, no. 6, pp. 1006–1011, June 1992.
[119] R. L. Kosut, "Adaptive control via parameter set estimation," *Int. J. Adaptive Control Signal Process.*, vol. 2, no. 4, pp. 371–399, Dec. 1988.
[120] H. Kudo and T. Saito, "Sinogram recovery with the method of convex projections for limited-data reconstruction in computed tomography," *J. Optical Soc. America A*, vol. 8, no. 7, pp. 1148–1160, July 1991.
[121] S. S. Kuo and R. J. Mammone, "Image restoration by convex projections using adaptive constraints and the L_1 norm," *IEEE Trans. Signal Process.*, vol. 40, no. 1, pp. 159–168, Jan. 1992.
[122] H. J. Landau and W. L. Miranker, "The recovery of distorted band-limited signals," *J. Math. Analysis Applications*, vol. 2, no. 1, pp. 97–104, Feb. 1961.
[123] A. M. Landraud, "Image restoration and enhancement of characters, using convex projection methods," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 1, pp. 85–92, Jan. 1991.
[124] A. Lent and H. Tuy, "An iterative method for the extrapolation of band-limited functions," *J. Math. Analysis Applications*, vol. 83, no. 2, pp. 554–565, Oct. 1981.
[125] A. Levi and H. Stark, "Signal reconstruction from phase by projection onto convex sets," *J. Optical Soc. America*, vol. 73, no. 6, pp. 810–822, June 1983.
[126] ——, "Image restoration by the method of generalized projections with application to restoration from magnitude," *J. Optical Soc. America A*, vol. 1, no. 9, pp. 932–943, Sept. 1984.
[127] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Medical Imaging*, vol. 8, no. 2, pp. 186–193, June 1989.
[128] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[129] G. E. Mailloux, F. Langlois, P. Y. Simard, and M. Bertrand, "Restoration of the velocity field of the heart from two-

dimensional echograms," *IEEE Trans. Medical Imaging*, vol. 8, no. 2, pp. 143–153, June 1989.
[130] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Trans. Informat. Theory*, vol. 37, no. 4, pp. 1019–1033, July 1991.
[131] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.
[132] J. Mandel, "Convergence of the cyclical relaxation method for linear inequalities," *Math. Programming*, vol. 30, no. 2, pp. 218–228, 1984.
[133] C. P. Mariadassou and B. Yegnanarayana, "Image reconstruction from noisy digital holograms," *IEE Proc.-F*, vol. 137, no. 5, pp. 351–356, Oct. 1990.
[134] R. J. Marks II, "Class of continuous level associative memory neural nets," *Appl. Optics*, vol. 26, no. 10, pp. 2005–2010, May 1987.
[135] R. J. Marks II, S. Oh, and L. E. Atlas, "Alternating projection neural networks," *IEEE Trans. Circuits Syst.*, vol. 36, no. 6, pp. 846–857, June 1989.
[136] J. M. Martinez, "Solution of nonlinear systems of equations by an optimal projection method," *Computing*, vol. 37, no. 1, pp. 59–70, 1986.
[137] W. Menke, "Applications of the POCS inversion method to interpolating topography and other geophysical fields," *Geophysical Res. Letters*, vol. 18, no.3, pp. 435–438, Mar. 1991.
[138] M. Milanese and G. Belaforte, "Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors: Linear families of models and estimators," *IEEE Trans. Automat. Contr.*, vol. 27, no. 2, pp. 408–414, Apr. 1982.
[139] W. D. Montgomery, "Optical applications of Von Neumann's alternating-projection theorem," *Optics Letters*, vol. 7, no. 1, pp. 1–3, Jan 1982.
[140] D. R. Morrell and W. C. Stirling, "Set-valued filtering and smoothing," *IEEE Trans. Systems, Man, Cybernet.*, vol. 21, no. 1, pp. 184–193, Jan. 1991.
[141] T. S. Motzkin and I. J. Schoenberg, "The relaxation method for linear inequalities," *Canadian J. Math.*, vol. 6, no. 3, pp. 393–404, 1954.
[142] K. M. Nashold and B. E. A. Saleh, "Image construction through diffraction-limited high-contrast imaging systems: An iterative approach," *J. Optical Soc. America A*, vol. 2, no. 5, pp. 635–643, May 1985.
[143] R. A. Nobakht, M. R. Civanlar, and S. H. Ardalan, "Comments on 'Design of a class of time-constrained FIR digital filters by successive projections'," *IEEE Trans. Circuits Syst.*, vol. 37, no. 12, p. 1581, Dec. 1990.
[144] J. P. Norton, "Identification and application of bounded-parameter models," *Automatica*, vol. 23, no. 4, pp. 497–507, July 1987.
[145] P. Oskoui-Fard and H. Stark, "Tomographic image reconstruction using the theory of convex projections," *IEEE Trans. Medical Imaging*, vol. 7, no. 1, pp. 45–58, Mar. 1988.
[146] ——, "A comparative study of three reconstruction methods for a limited-view computer tomography problem," *IEEE Trans. Medical Imaging*, vol. 8, no. 1, pp. 43–49, Mar. 1989.
[147] N. Ottavy, "Strong convergence of projection-like methods in Hilbert spaces," *J. Optimiz. Theory Applications*, vol. 56, no. 3, pp. 433–461, Mar. 1988.
[148] R. D. Palmer, J. R. Cruz, and D. S. Zrnić, "Enhanced autoregressive moving average spectral estimation applied to the measurement of Doppler spectral width," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, no. 3, pp. 358–368, May 1991.
[149] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.*, vol. 22, no. 9, pp. 735–742, Sept. 1975.
[150] R. K. Pearson, "Block-sequential algorithms for set-theoretic estimation," *SIAM J. Matrix Analysis Applications*, vol. 9, no. 4, pp. 513–527, Oct. 1988.
[151] S. C. Pei and I. I. Yang, "Design of a class of time-constrained FIR digital filters by successive projections," *IEEE Trans. Circuits Syst.*, vol. 36, no. 1, pp. 164–167, Jan. 1989.
[152] B. Picinbono and M. Bouvet, "Constrained Wiener filtering," *IEEE Trans. Informat. Theory*, vol. 33, no. 1, pp. 160–166, Jan. 1987.
[153] G. Pierra, "Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique," *Lecture Notes Computer Sci.*, vol. 41, pp. 200–218, 1976.
[154] ——, "Decomposition through formalization in a product space," *Math. Programming*, vol. 28, no. 1, pp. 96–115, Jan. 1984.
[155] H. Piet-Lahanier and E. Walter, "Characterization of non-connected parameter uncertainty regions," *Math. Computers in Simulation*, vol. 32, no. 5–6, pp. 553–560, Dec. 1990.
[156] G. T. Poulton, "Antenna power pattern synthesis using method of successive projections," *Electronics Letters*, vol. 22, no. 20, pp. 1042–1043, Sept. 1986.
[157] G. T. Poulton and S. G. Hay, "Efficient design of shaped reflectors using successive projections," *Electronics Letters*, vol. 27, no. 23, pp. 2156–2158, Nov. 1991.
[158] R. Rangayyan, A. P. Dhawan, and R. Gordon, "Algorithms for limited-view computed tomography: An annoted bibliography and a challenge," *Appl. Optics*, vol. 24, no. 23, pp. 4000–4012, Dec. 1985.
[159] A. K. Rao, Y. F. Huang, and S. Dasgupta, "ARMA parameter estimation using a novel recursive estimation algorithm with selective updating," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 38, no. 3, pp. 447–457, Mar. 1990.
[160] S. Reich, "A limit theorem for projections," *Linear Multilinear Algebra*, vol. 13, no. 3, pp. 281–290, 1983.
[161] J. Rosen and J. Shamir, "Application of the projection onto constraint sets algorithm for optical pattern recognition," *Optics Letters*, vol. 16, no. 10, pp. 752–754, May 1991.
[162] T. J. Rothenberg, *Efficient Estimation with a Priori Information*. New Haven: Yale University Press, 1973.
[163] H. A. Sabbagh, L. D. Sabbagh, and T. M. Roberts, "An eddy-current model and algorithm for three-dimensional nondestructive evaluation of advanced composites," *IEEE Trans. Magnetics*, vol. 24, no. 6, pp. 3201–3212, Nov. 1988.
[164] B. E. A. Saleh and K. M. Nashold, "Image construction: Optimum amplitude and phase masks in photolitography," *Appl. Optics*, vol. 24, no. 10, pp. 1432–1437, May 1985.
[165] P. Santago and S. A. Rajala, "Using convex set techniques for combined pixel and frequency domain coding of time-varying images," *IEEE J. Selected Areas Commun.*, vol. 5, no. 7, pp. 1127–1139, Aug. 1987.
[166] O. Sasaki and T. Yamagami, "Image restoration in singular vector space by the method of convex projections," *Appl. Optics*, vol. 26, no. 7, pp. 1216–1221, Apr. 1987.
[167] K. D. Sauer and J. P. Allebach, "Iterative reconstruction of band-limited images from nonuniformly spaced samples," *IEEE Trans. Circuits Syst.*, vol. 34, no. 12, pp. 1497–1506, Dec. 1987.
[168] L. J. Savage, *The Foundations of Statistics*, second ed. New York: Dover, 1972.
[169] F. M. Schlaepfer and F. C. Schweppe, "Continuous-time state estimation under disturbances bounded by convex sets," *IEEE Trans. Automat. Contr.*, vol. 17, no. 2, pp. 197–205, Apr. 1972.
[170] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[171] I. E. Schochetman and R. L. Smith, "Convergence of best approximations from unbounded sets," *J. Math. Analysis Applications*, vol. 166, no. 1, pp. 112–128, May 1992.
[172] H. A. Schwarz, "Grenzübergang durch alternirendes Verfahren," 1870; reprinted in *Gesammelte Mathematische Abhandlungen*. Berlin: Springer-Verlag, vol. 2, 1890, pp. 133–143.
[173] L. Schwartz, *Analyse—Topologie Générale et Analyse Fonctionnelle*, second edition. Paris: Hermann, 1980.
[174] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. Automat. Contr.*, vol. 13, no. 1, pp. 22–28, Feb. 1968.
[175] ——, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
[176] F. C. Schweppe and H. K. Knudsen, "The theory of amorphous cloud trajectory prediction," *IEEE Trans. Informat. Theory*, vol. 14, no. 3, pp. 415–427, May 1968.
[177] M. I. Sezan and H. Stark, "Image restoration by the method of convex projections: Part 2—Applications and numerical results," *IEEE Trans. Medical Imaging*, vol. 1, no. 2, pp. 95–101, Oct. 1982.
[178] ——, "Image restoration by convex projections in the presence of noise," *Appl. Optics*, vol. 22, no. 18, pp. 2781–2789, Sept. 1983.
[179] ——, "Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion," *IEEE Trans. Medical Imaging*, vol. 3, no. 2, pp. 91–98, June

1984.
[180] ——, "Incorporation of *a priori* moment information into signal recovery and synthesis problems," *J. Math. Analysis Applications*, vol. 122, no. 1, pp. 172–186, Feb. 1987.
[181] M. I. Sezan, H. Stark, and S. J. Yeh, "Projection method formulations of Hopfield-type associative memory neural networks," *Appl. Optics*, vol. 29, no. 17, pp. 2616–2622, June 1990.
[182] M. I. Sezan and A. M. Tekalp, "Adaptive image restoration with artifact suppression using the theory of convex projections," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 1, pp. 181–185, Jan. 1990.
[183] ——, "Survey of recent developments in digital image restoration," *Optical Eng.*, vol. 29, no. 5, pp. 393–404, May 1990.
[184] M. I. Sezan and H. J. Trussell, "Prototype image constraints for set-theoretic image restoration," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2275–2285, Oct. 1991.
[185] P. Y. Simard and G. E. Mailloux, "A projection operator for the restoration of divergence-free vector fields," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 10, no. 2, pp. 248–256, Mar. 1988.
[186] ——, "Vector field restoration by the method of convex projections," *Comput. Vision, Graphics, Image Process.*, vol. 52, no. 3, pp. 360–385, Dec. 1990.
[187] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, CA: Academic, 1987.
[188] H. Stark, D. Cahana, and H. Webb, "Restoration of arbitrary finite energy optical objects from limited spatial and spectral information," *J. Optical Soc. America A*, vol. 71, no. 6, pp. 635–642, June 1981.
[189] H. Stark, W. C. Catino, and J. L. LoCicero, "Design of phase gratings by generalized projections," *J. Optical Soc. America A*, vol. 8, no. 3, pp. 566–571, Mar. 1991.
[190] H. Stark and E. T. Olsen, "Projection-based image restoration," *J. Optical Soc. America A*, vol. 9, no. 11, pp. 1914–1919, Nov. 1992.
[191] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Optical Soc. America A*, vol. 6, no. 11, pp. 1715–1726, Nov. 1989.
[192] W. J. Stiles, "Closest point maps and their product II," *Nieuw Archief voor Wiskunde*, vol. 13, no. 3, pp. 212–225, Nov. 1965.
[193] W. C. Stirling and D. R. Morrell, "Convex Bayes decision theory," *IEEE Trans. Systems, Man, Cybernetics*, vol. 21, no. 1, pp. 173–183, Jan. 1991.
[194] B. J. Sullivan and A. Katsaggelos, "New termination rule for linear iterative image restoration algorithms," *Optical Eng.*, vol. 29, no. 5, pp. 471–477, May 1990.
[195] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications," *Numerische Mathematik*, vol. 17, no. 3, pp. 203–214, 1971.
[196] A. M. Tekalp and H. J. Trussell, "Comparative study of some statistical and set-theoretic methods for image restoration," *CVGIP: Graphical Models Image Processing*, vol. 53, no. 2, pp. 108–120, Mar. 1991.
[197] R. Tempo, "Robust estimation and filtering in the presence of bounded noise," *IEEE Trans. Automat. Contr.*, vol. 33, no. 9, pp. 864–867, Sept. 1988.
[198] N. T. Thao and M. Vetterli, "Optimal MSE signal reconstruction in oversampled A/D conversion using convexity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, San Francisco, CA, vol. 4, Mar. 23–26, 1992, pp. 165–168.
[199] H. J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 1, pp. 129–136, Feb. 1983.
[200] ——, "Application of set theoretic methods to color systems," *Color Res. Applications*, vol. 16, no. 1, pp. 31–41, Feb. 1991.
[201] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 201–212, Apr. 1984.
[202] H. J. Trussell and M. R. Civanlar, "The Landweber iteration and projection onto convex sets," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 6, pp. 1632–1634, Dec. 1985.
[203] H. J. Trussell, H. Orun-Ozturk, and M. R. Civanlar, "Errors in reprojection methods in computerized tomography," *IEEE Trans. Medical Imaging*, vol. 6, no. 3, pp. 220–227, Sept. 1987.
[204] H. J. Trussell and P. L. Vora, "Bounds on restoration quality using *a priori* information," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, New York, NY, Apr. 11–14, 1988, pp. 1758–1761.
[205] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Medical Imaging*, vol. 6, no. 4, pp. 313–319, Dec. 1987.
[206] J. Von Neumann, "On rings of operators. Reduction theory," *Annals Math.*, vol. 50, no. 2, pp. 401–485, Apr. 1949.
[207] E. Walter (Guest Editor), "Parameter identifications with error bounds," *Math. Comp. Simulation*, vol. 32, no. 5–6, Dec. 1990.
[208] E. Walter and H. Piet-Lahanier, "Exact recursive polyhedral description of the feasible parameter set for bounded-error models," *IEEE Trans. Automat. Contr.*, vol. 34, no. 8, pp. 911–915, Aug. 1989.
[209] M. N. Wernick and C. T. Chen, "Superresolved tomography by convex projections and detector motion," *J. Optical Soc. America A*, vol. 9, no. 9, pp. 1547–1553, Sept. 1992.
[210] N. Wiener, "On the factorization of matrices," *Commentarii Mathematici Helvetici*, vol. 29, pp. 97–111, 1955.
[211] L. B. White, "Signal synthesis from Cohen's class of bilinear time-frequency representations using convex projections," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toronto, Canada, May 14–17, 1991, pp. 2053–2056.
[212] ——, "The wide-band ambiguity function and Altes' Q-distribution: Constrained synthesis and time-scale filtering," *IEEE Trans. Informat. Theory*, vol. 38, no. 2, pp. 886–892, Mar. 1992.
[213] H. S. Witsenhausen, "Sets of possible states of linear systems given perturbed observations," *IEEE Trans. Automat. Contr.*, vol. 13, no. 5, pp. 556–558, Oct. 1968.
[214] P. Wolfe, "Finding the nearest point in a polytope," *Math. Programming*, vol. 11, no. 2, pp. 128–149, Oct. 1976.
[215] K. Yang and K. G. Murty, "New iterative methods for linear inequalities," *J. Optimiz. Theory Applications*, vol. 72, no. 1, pp. 163–185, Jan. 1992.
[216] X. Yang, K. Wang, and S. A. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Informat. Theory*, vol. 38, no. 2, pp. 824–839, Mar. 1992.
[217] S. J. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *J. Optical Soc. America A*, vol. 7, no. 3, pp. 491–499, Mar. 1990.
[218] K. Yosida, *Functional Analysis*, sixth edition. New York: Springer-Verlag, 1980.
[219] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 694–702, Sept. 1978.
[220] D. C. Youla and V. Velasco, "Extensions of a result on the synthesis of signals in the presence of inconsistent constraints," *IEEE Trans. Circuits Syst.*, vol. 33, no. 4, pp. 465–468, Apr. 1986.
[221] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Medical Imaging*, vol. 1, no. 2, pp. 81–94, Oct. 1982.
[222] L. A. Zadeh, "What is optimal?," *IRE Trans. Informat. Theory*, vol. 4, no. 1, p. 3, Mar. 1958.
[223] A. Zakhor and A. V. Oppenheim, "Reconstruction of two-dimensional signals from level crossings," *Proc. IEEE*, vol. 78, no. 1, pp. 31–55, Jan. 1990.



Patrick L. Combettes (Member, IEEE) was born in Salon de Provence, France, on May 24, 1962. He received the Diplôme d'Ingénieur from l'Institut National des Sciences Appliquées de Lyon, France, and the M.S. and Ph.D. degrees from North Carolina State University, Raleigh, all in electrical engineering, in 1985, 1987, and 1989, respectively.

During the academic year 1989–1990, he was a Visiting Assistant Professor in the Department of Electrical and Computer Engineering at North Carolina State University, Raleigh. In August 1990, he joined the Department of Electrical Engineering at the City University of New York, as an Assistant Professor. His current research interests are in mathematical signal processing.

Dr. Combettes is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, and the American Mathematical Society.