Inconsistent Signal Feasibility Problems: Least-Squares Solutions in a Product Space

Patrick L. Combettes, *Member, IEEE*

***Abstract*—In this paper, we present parallel projection methods to find least-squares solutions to inconsistent convex set theoretic signal synthesis problems. The problem of finding a signal that minimizes a weighted average of the squares of the distances to constraint sets is reformulated in a product space, where it is equivalent to that of finding a point that lies in a particular subspace and at minimum distance from the Cartesian product of the original sets. A solution is obtained in the product space via methods of alternating projections which naturally lead to methods of parallel projections in the original space. The convergence properties of the proposed methods are analyzed and signal synthesis applications are demonstrated.**

I. Introduction

THE goal of a set theoretic signal synthesis (estimation or design) problem is to produce a signal a^* that satisfies simultaneously a collection $(\Psi_i)_{1\leq i\leq m}$ of constraints. In a suitable signal space Ξ, each constraint Ψ_i is associated with a property set $S_i = \{a \in \Xi \mid a \text{ satisfies } \Psi_i\}$ and the feasibility problem is then stated as

$$\text{Find } a^* \in \bigcap_{i=1}^{m} S_i. \tag{1}$$

The pair $(\Xi, (S_i)_{1\leq i\leq m})$ is called a set theoretic formulation. In this work, it will be assumed that Ξ is a Hilbert space and that the S_is are closed and convex. This convex set theoretic feasibility framework has been applied to a wide range of signal processing problems including signal deconvolution, tomographic reconstruction, image restoration, band-limited extrapolation, and image synthesis. The reader is referred to [9] for an in-depth account of the field and an extensive survey of applications. In set theoretic signal estimation problems, e.g., reconstruction [19], enhancement [25], restoration [33], recovery from bispectrum [5] or phase [22] information, the goal is to produce an estimate of the true signal which is consistent with all *a priori* knowledge and the data. $(\Psi_i)_{1\leq i\leq m}$ then represents the available information for the problem. On the other hand, in signal design problems, e.g., pulse shape design [8], image construction [24], filter design [26], the goal is to produce a signal that meets a collection of requirements and $(\Psi_i)_{1\leq i\leq m}$ then represents a collection of design specifications.

Manuscript received May 27, 1993; revised March 4, 1994. This work was partially supported by the National Science Foundation under grant MIP-9308609. The associate editor coordinating the review of this paper and approving it for publication was Prof. Stanley J. Reeves.

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

IEEE Log Number 9404802.

There may be cases when (1) has no solution because incompatible constraints are present, which forces $\bigcap_{i=1}^{m} S_i = \emptyset$. In estimation problems, this situation may be due to inaccurate or imprecise information. For instance, consider the sets proposed in [36]. Most of them depend on attributes of the original signal, e.g., amplitude bounds, region of support, band-limitedness, energy, that may not be known exactly. The same remark also applies to the attributes of the noise that are required to construct the sets described in [12], e.g., bounds, moments, spectral properties, distribution. In addition, such sets based on stochastic information take the form of confidence regions and their construction depends on the specification of a confidence level. If the confidence level is unrealistically low, the sets may not intersect. Inconsistencies may also be due to inadequate data modeling, for instance when random variations in the impulse response of a system [11] or noise perturbations in the measurements [19], [32] are not taken into account. In design problems, inconsistencies are even more likely, as the specifications of the constraints are left to the user. An example is the data window design problem of [17], which involves conflicting time and frequency domain constraints.

Because of the possibility of encountering inconsistent set theoretic formulations, one must be aware of the convergence behavior of existing solution algorithms when they are applied to them. In almost all of the fields of application of set theoretic methods, the so-called POCS (projections onto convex sets) algorithm has prevailed to solve (1). Let P_i be the projection operator onto S_i. In the general form of POCS, an initial estimate a_0 is projected sequentially onto the sets in a cyclic manner according to the iterations [3], [18]

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n + \lambda_n (P_{n(\text{modulo } m)+1}(a_n) - a_n) \tag{2}$$

where

$$(\lambda_n)_{n\geq 0} \subset [\varepsilon, 2-\varepsilon] \quad \text{with} \quad 0 < \varepsilon < 1. \tag{3}$$

If $\bigcap_{i=1}^{m} S_i \neq \emptyset$, any sequence $(a_n)_{n\geq 0}$ generated by POCS converges weakly to a signal in $\bigcap_{i=1}^{m} S_i$. The convergence properties of the unrelaxed version of POCS, that is

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = P_{n(\text{modulo } m)+1}(a_n) \tag{4}$$

in inconsistent problems were studied in [18]. It was shown there that, if one of the sets is bounded, there exist points $(\overline{a}_i)_{1\leq i\leq m}$ such that $P_1(\overline{a}_m) = \overline{a}_1$ and $P_i(\overline{a}_{i-1}) = \overline{a}_i$ for every i in $\{2, \cdots, m\}$. Moreover, the cyclic subsequence $(a_{mn+i})_{n\geq 0}$ converges weakly to such a point $\overline{a}_i$ in S_i. In

the particular case when $m = 2$, this result simply states that the sequence $(a_{2n+1})_{n\geq 0}$ converges weakly to a point $\bar{a}_1$ in S_1 such that $P_1(P_2(\bar{a}_1)) = \bar{a}_1$, i.e., to a signal that satisfies property Ψ_1 and which is closest to satisfying Ψ_2 (this result is also discussed in [6], [17], and [35]). Beyond two sets, however, the above result has no useful interpretation and little practical value. It merely indicates that the limit signal $\bar{a}_i$ lies in S_i and, thereby, satisfies Ψ_i. Aside from Ψ_i, however, the properties of $\bar{a}_i$ are totally unknown and there is no guarantee that any of the remaining constraints will be satisfied, even in an approximate sense. Such a solution clearly constitutes a poor approximation of a feasible signal. Thus, the convergence behavior of POCS in the inconsistent case is generally unsatisfactory.

In inconsistent problems, there exists no signal possessing exactly all the properties $(\Psi_i)_{1\leq i\leq m}$ but one can look for a signal that satisfies them in some approximate sense. Let us consider the basic feasibility problem of solving a system of m linear equations in $\mathbb{R}^k$. If the system is overdetermined, it is customary to look for a least-squares solutions. In set theoretic terms, if $(S_i)_{1\leq i\leq m}$ represents the family of hyperplanes of $\mathbb{R}^k$ associated with the equations, this is equivalent to looking for a point a^* which minimizes $\sum_{i=1}^m d(a, S_i)^2$, the sum of the squares of the distances to the S_is. Along the same lines, the exact feasibility problem (1) can be replaced by the weighted least-squares feasibility problem

$$\text{Minimize } \Phi(a) = \frac{1}{2}\sum_{i=1}^m w_i d(a, S_i)^2 \text{ over } \Xi \tag{5}$$

where $(w_i)_{1\leq i\leq m}$ are strictly convex weights, that is

$$\sum_{i=1}^m w_i = 1 \qquad \text{and} \qquad (\forall i \in \{1, \cdots, m\})\ w_i > 0. \tag{6}$$

We shall call Φ the proximity function. The smaller $\Phi(a)$, the closer the signal a to satisfying all the properties in the above weighted least-squares sense. To that extent, $\Phi(a)$ measures the degree of unfeasibility of a. Thus, in instances when consistency is not guaranteed, the goal will be to find a minimum of the proximity function Φ, i.e., to solve (5) or, equivalently

$$\text{Find } a^* \in G = \{a \in \Xi \mid (\forall b \in \Xi) \quad \Phi(a) \leq \Phi(b)\}. \tag{7}$$

Of course, if $\bigcap_{i=1}^m S_i \neq \emptyset$, the minimum value of the proximity function is 0 which is attained only on $G = \bigcap_{i=1}^m S_i$, so that (1) and (7) coincide. In general, (7) can be viewed as an extension of (1) and G is the set of least-squares solutions of the (possibly inconsistent) signal feasibility problem.

In finite dimensional spaces, and under certain conditions on the problem, it is known that (2) can solve (7) if the sequence of relaxation parameters $(\lambda_n)_{n\geq 0}$ approaches zero [4], [29]. Experimental evidence first suggested this property in the inconsistent tomographic reconstruction problems of [20], where POCS was reported to provide better results with strong underrelaxations than without relaxations, as in (4). From a practical viewpoint, however, strong underrelaxations are not desirable as they impose very small step sizes $(\|a_{n+1} - a_n\|)_{n\geq 0}$ and, overall, excessively slow convergence.

In this paper, we propose simultaneous projections algorithms to solve the least-squares feasibility problem (7). Instead of approaching (7) directly in the original signal space Ξ, we shall reformulate it in the m-fold Cartesian product space $\boldsymbol{\Xi} = \Xi^m$, where it will be posed as a simple 2-set problem and solved via alternating projection schemes. These schemes will in turn directly lead to parallel projection methods once transposed back in Ξ. The paper is organized as follows. In Section II, we review relevant material and establish preliminary theoretical results. In Section III, we recast (7) in the product space $\boldsymbol{\Xi}$ and develop the convergence analysis in this simpler setting. This analysis is brought back to the original space Ξ in Section IV, where the convergence properties of the parallel projection methods are presented. Section V is devoted to applications to set theoretic signal synthesis. The paper is concluded in Section VI.

II. Theory

In this section, we lay the theoretical foundation of the paper. Most developments require only a modest background of standard hilbertian analysis, as can be found in [15] and [31]. To keep the focus on the results, the proofs of this section have been placed in the Appendix.

A. Notations and Assumptions

$\mathbb{N}$ is the set of nonnegative integers and $\mathbb{R}_+$ the set of nonnegative reals. All the signals to be considered belong to a real Hilbert space Ξ with scalar product $\langle \cdot \mid \cdot \rangle$, norm $\|\cdot\|$, and distance d. A sequence of signals $(a_n)_{n\geq 0}$ is said to converge to a signal a strongly if $(\|a_n - a\|)_{n\geq 0}$ converges to 0, and weakly if $(\langle a_n - a \mid b\rangle)_{n\geq 0}$ converges to 0, for every b in Ξ. In Ξ, the constraints $(\Psi_i)_{1\leq i\leq m}$ defining the signal to be synthesized are represented by closed and convex sets $(S_i)_{1\leq i\leq m}$. The distance from a point a in Ξ to the set S_i is $d(a, S_i) = \inf\{d(a, b) \mid b \in S_i\}$. The proximity function Φ for the problem is defined by

$$\begin{aligned} \Phi\colon \Xi &\to \mathbb{R}_+ \\ a &\mapsto \frac{1}{2}\sum_{i=1}^m w_i d(a, S_i)^2 \end{aligned} \tag{8}$$

where $(w_i)_{1\leq i\leq m}$ are strictly convex weights, as in (6). Moreover, G denotes the set of minimizers of Φ, i.e., the set of weighted least-squares-feasible signals.

As discussed in the Introduction, the problem of finding a minimum of the proximity function will be recast in another real Hilbert space $\boldsymbol{\Xi}$ (the m-fold Cartesian product of Ξ). Its scalar product will be denoted by $\langle\langle \cdot \mid \cdot \rangle\rangle$, its norm by $|||\cdot|||$, and its distance by $\mathbf{d}$. Υ will denote the identity operator on $\boldsymbol{\Xi}$. A fixed point of an operator $T\colon \boldsymbol{\Xi} \to \boldsymbol{\Xi}$ is any point $\mathbf{a}$ such that $T(\mathbf{a}) = \mathbf{a}$.

B. Nonexpansive Iterations

Let $(\mathbf{a}, \mathbf{b})$ be an arbitrary pair in $\boldsymbol{\Xi}^2$. An operator $T\colon \boldsymbol{\Xi} \to \boldsymbol{\Xi}$ is said to be nonexpansive if

$$|||T(\mathbf{a}) - T(\mathbf{b})||| \leq |||\mathbf{a} - \mathbf{b}||| \tag{9}$$

and firmly nonexpansive if

$$\langle\langle \mathbf{a}-\mathbf{b} \mid T(\mathbf{a})-T(\mathbf{b})\rangle\rangle \geq |||T(\mathbf{a})-T(\mathbf{b})|||^2. \quad (10)$$

It is easy to see that (10) $\Rightarrow$ (9). If T is nonexpansive, then its set of fixed points is closed and convex. Moreover, if T_1 and T_2 are nonexpansive operators, so are the composition $T_1 \circ T_2$ and the convex combination $(1-\alpha)T_1+\alpha T_2$, where $\alpha \in [0,1]$.

Proposition 1: Let T: $\Xi \rightarrow \Xi$ be a firmly nonexpansive operator with nonempty set of fixed points $\mathbf{F}$, let $(\lambda_n)_{n\geq 0}$ be as in (3), and let $\mathbf{a}_0$ be an arbitrary point in Ξ. Then the iterations

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_{n+1} = \mathbf{a}_n + \lambda_n(T(\mathbf{a}_n)-\mathbf{a}_n) \quad (11)$$

converge weakly to a point in $\mathbf{F}$. □

Proposition 2: [23] Let $T : \Xi \rightarrow \Xi$ be a nonexpansive operator with nonempty set of fixed points $\mathbf{F}$, let $\mathbf{a}_0$ be an arbitrary point in Ξ, and take any sequence $(\alpha_n)_{n\geq 0} \subset [0,1[$ such that

$$\begin{cases} \lim_{n\rightarrow+\infty} \alpha_n = 1, \\ \sum_{n\geq 0}(1-\alpha_n) = +\infty. \\ \lim_{n\rightarrow+\infty}(\alpha_{n+1}-\alpha_n)(1-\alpha_{n+1})^{-2} = 0. \end{cases} \quad (12)$$

Then the iterations

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_{n+1} = (1-\alpha_n)\mathbf{a}_0 + \alpha_n T(\mathbf{a}_n) \quad (13)$$

converge strongly to the point in $\mathbf{F}$ which is at minimum distance from $\mathbf{a}_0$. □

Note that an example of sequence $(\alpha_n)_{n\geq 0}$ that satisfies (12) is

$$(\forall n \in \mathbb{N}) \quad \alpha_n = 1-(n+1)^{-\kappa}, \text{ where } 0<\kappa<1. \quad (14)$$

C. Projections

A detailed account of the properties of projections onto convex subsets of Hilbert spaces can be found in [1] and [37]. We review here a few basic facts.

Let $\mathbf{D}$ and $\mathbf{S}$ be nonempty closed and convex subsets of Ξ. The distance between $\mathbf{D}$ and $\mathbf{S}$ is defined to be $\mathbf{d}(\mathbf{D},\mathbf{S}) = \inf\{\mathbf{d}(\mathbf{a},\mathbf{S}) \mid \mathbf{a} \in \mathbf{D}\}$. The function $\mathbf{d}(\cdot,\mathbf{S})$ is continuous and convex on Ξ. Every $\mathbf{a}$ in Ξ admits a unique projection onto $\mathbf{S}$, i.e., there exists a unique point $P_{\mathbf{S}}(\mathbf{a})$ in $\mathbf{S}$ such that $\mathbf{d}(\mathbf{a},P_{\mathbf{S}}(\mathbf{a})) = \mathbf{d}(\mathbf{a},\mathbf{S})$. The projection operator $P_{\mathbf{S}}$ is characterized by the variational inequality

$$(\forall (\mathbf{a},\mathbf{b}) \in \Xi^2) \quad \langle\langle \mathbf{a}-P_{\mathbf{S}}(\mathbf{a}) \mid P_{\mathbf{S}}(\mathbf{b})-P_{\mathbf{S}}(\mathbf{a})\rangle\rangle \leq 0 \quad (15)$$

which implies that $P_{\mathbf{S}}$ is firmly nonexpansive. If $\mathbf{D}$ is a closed vector subspace of Ξ, $P_{\mathbf{D}}$ is linear and (15) yields

$$(\forall \mathbf{a} \in \Xi)(\forall \mathbf{b} \in \mathbf{D}) \quad \langle\langle \mathbf{a} \mid \mathbf{b}\rangle\rangle = \langle\langle P_{\mathbf{D}}(\mathbf{a}) \mid \mathbf{b}\rangle\rangle. \quad (16)$$

Proposition 3: Let $\mathbf{D} \subset \Xi$ be a closed vector space and $\mathbf{S} \subset \Xi$ a nonempty closed and convex set. Then

i) $P_{\mathbf{D}} \circ P_{\mathbf{S}}$ is firmly nonexpansive on $\mathbf{D}$;
ii) $(\forall \lambda \in [0,2]) \quad \Upsilon + \lambda(P_{\mathbf{S}}-\Upsilon)$ is nonexpansive. □

$R_{\mathbf{S}} = 2P_{\mathbf{S}} - \Upsilon$ is the operator of reflection with respect to $\mathbf{S}$. If ∇ denotes the gradient operator in Ξ, then $(\forall \mathbf{a} \in \Xi)$ $\nabla \mathbf{d}(\mathbf{a},\mathbf{S})^2 = 2(\mathbf{a}-P_{\mathbf{S}}(\mathbf{a}))$.

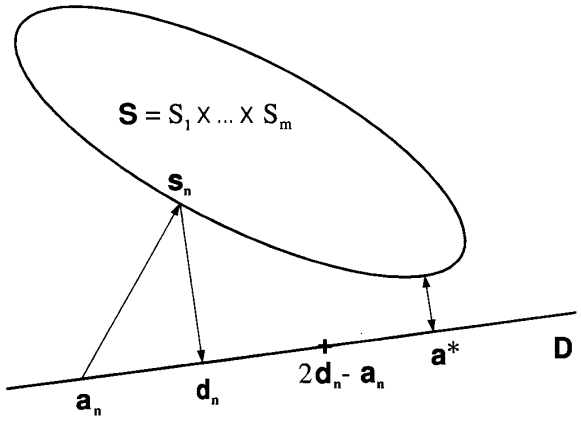

Fig. 1. Iteration in the product space.

D. Alternating Projections

In this section, $\mathbf{D} \subset \Xi$ is a closed vector space and $\mathbf{S} \subset \Xi$ a nonempty closed and convex set. Their projection operators are denoted by $P_{\mathbf{D}}$ and $P_{\mathbf{S}}$, respectively. Let $\mathbf{G}$ be the (possibly empty) set of fixed point of $P_{\mathbf{D}} \circ P_{\mathbf{S}}$, i.e.

$$\mathbf{G} = \{\mathbf{a} \in \mathbf{D} \mid P_{\mathbf{D}}(P_{\mathbf{S}}(\mathbf{a})) = \mathbf{a}\}. \quad (17)$$

Proposition 4: [6] $\mathbf{G} = \{\mathbf{a} \in \mathbf{D} \mid \mathbf{d}(\mathbf{a},\mathbf{S}) = \mathbf{d}(\mathbf{D},\mathbf{S})\}$. □

In words, Proposition 4 states that the fixed points of $P_{\mathbf{D}} \circ P_{\mathbf{S}}$ coincide with the minimizers of the functional

$$\Phi\colon \mathbf{D} \rightarrow \mathbb{R}_+, \quad \mathbf{a} \mapsto \frac{1}{2}\mathbf{d}(\mathbf{a},\mathbf{S})^2. \quad (18)$$

Theorem 1: Suppose that $\mathbf{G} \neq \emptyset$. Then, for any $\mathbf{a}_0$ in $\mathbf{D}$, every sequence of iterates $(\mathbf{a}_n)_{n\geq 0}$ defined by

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_{n+1} = \mathbf{a}_n + \lambda_n(P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n) \quad (19)$$

where $(\lambda_n)_{n\geq 0}$ is as in (3), converges weakly to a point in $\mathbf{G}$. □

A pictorial description of (19) is given in Fig. 1: $\mathbf{s}_n = P_{\mathbf{S}}(\mathbf{a}_n)$ and $\mathbf{d}_n = P_{\mathbf{D}}(\mathbf{s}_n) = P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)$ are first computed and $\mathbf{a}_{n+1}$ is then positioned on the segment between $\mathbf{a}_n$ and $\mathbf{d}_n$ or between $\mathbf{d}_n$ and $2\mathbf{d}_n - \mathbf{a}_n$ according as $\varepsilon \leq \lambda \leq 1$ or $1 \leq \lambda_n \leq 2-\varepsilon$. We shall say that iteration n is underrelaxed if $\lambda_n \leq 1$, unrelaxed if $\lambda_n = 1$, and overrelaxed if $\lambda_n \geq 1$. As discussed in the Introduction, in the unrelaxed case, Theorem 1 was proved in [18].

The relaxation parameters in (19) can vary in the interval $[\varepsilon, 2-\varepsilon]$ at each iteration. Let us consider the problem of finding the optimal relaxation parameter at iteration n in terms of bringing $\mathbf{a}_{n+1}$ as close as possible to an arbitrary point $\mathbf{a}^*$ in $\mathbf{G}$. The optimal value of λ_n should therefore minimize $|||\mathbf{a}_{n+1}-\mathbf{a}^*|||$. We have

$$\begin{aligned} |||\mathbf{a}_{n+1}-\mathbf{a}^*|||^2 &= |||\mathbf{a}_{n+1}-\mathbf{a}_n|||^2 \\ &\quad + 2\langle\langle \mathbf{a}_{n+1}-\mathbf{a}_n \mid \mathbf{a}_n-\mathbf{a}^*\rangle\rangle \\ &\quad + |||\mathbf{a}_n-\mathbf{a}^*|||^2 \\ &= \lambda_n^2 |||\mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \end{aligned}$$

$$- 2\lambda_n \langle\langle \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \mid \mathbf{a}_n - \mathbf{a}^* \rangle\rangle + |||\mathbf{a}_n - \mathbf{a}^*|||^2. \quad (20)$$

This quadratic form in λ_n is minimized for

$$\lambda_n^* = \frac{\langle\langle \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \mid \mathbf{a}_n - \mathbf{a}^* \rangle\rangle}{|||\mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)|||^2}. \quad (21)$$

We observe that the optimal relaxation parameter depends on the solution point $\mathbf{a}^*$, which of course is not known. Hence, optimal relaxation cannot be achieved. However, the range of the relaxation parameters can be narrowed down by virtue of the following proposition.

Proposition 5: The optimal one-step relaxations in Theorem 1 are overrelaxations: $(\forall n \in \mathbb{N}) \quad \lambda_n^* \geq 1$. □

Let us now state additional properties of the alternating projection method (19).

Proposition 6: Let $(\mathbf{a}_n)_{n\geq 0}$ be any sequence of iterates in Theorem 1. Then $(\mathbf{\Phi}(\mathbf{a}_n))_{n\geq 0}$ decreases until convergence and

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_{n+1} = \mathbf{a}_n - \lambda_n \nabla_{\mathbf{D}} \mathbf{\Phi}(\mathbf{a}_n) \quad (22)$$

where $\nabla_{\mathbf{D}}$ is the gradient operator in the Hilbert space $\mathbf{D}$. □

E. Strong Convergence

Strong convergence of the unrelaxed version of (19) can be proved if one makes additional assumptions on $\mathbf{S}$, such as compactness [6], finite dimensionality [6], or uniform convexity [18]. The next theorem presents a strong convergence result for a variant of (19) which does not require special conditions.

Theorem 2: Suppose that $\mathbf{G} \neq \emptyset$. Then, for any $\mathbf{a}_0$ in $\mathbf{D}$, every sequence of iterates $(\mathbf{a}_n)_{n\geq 0}$ defined by

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_{n+1} = (1 - \alpha_n)\mathbf{a}_0 + \alpha_n(\lambda P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) + (1 - \lambda)\mathbf{a}_n) \quad (23)$$

where $(\alpha_n)_{n\geq 0}$ is as in (12) and $0 < \lambda \leq 2$, converges strongly to $P_{\mathbf{G}}(\mathbf{a}_0)$. □

Note that, as n increases, (23) tends to behave like a constant-relaxation version of (19).

F. Set of Least-Squares-Feasible Solutions

The properties of the set G of weighted least-squares solutions in (7) are given in the following proposition.

Proposition 7: Suppose that one of the S_is is bounded. Then G is nonempty, closed, convex, and bounded. □

It is noted that in the consistent case $G = \bigcap_{i=1}^m S_i$ and Proposition 7 becomes trivial.

III. Formalization in a Product Space

Let $\mathbf{\Xi} = \Xi^m$ be the m-fold Cartesian product of the original signal space Ξ. We shall denote by $\mathbf{a} = (a^{(1)}, \cdots, a^{(m)})$ an m-tuple of signals in $\mathbf{\Xi}$. $\mathbf{\Xi}$ can be made into a Hilbert space by endowing it with the scalar product $\langle\langle \mathbf{a} \mid \mathbf{b} \rangle\rangle = \sum_{i=1}^m w_i \langle a^{(i)} \mid b^{(i)} \rangle$. The corresponding norm and distance are given by

$$\begin{cases} |||\mathbf{a}||| = (\sum_{i=1}^m w_i \|a^{(i)}\|^2)^{1/2}. \\ \mathbf{d}(\mathbf{a}, \mathbf{b}) = (\sum_{i=1}^m w_i d(a^{(i)}, b^{(i)})^2)^{1/2}. \end{cases} \quad (24)$$

Let $\mathbf{S}$ be the Cartesian product of the sets $(S_i)_{1\leq i\leq m}$, i.e.

$$\mathbf{S} = \mathsf{X}_{i=1}^m S_i = \{\mathbf{a} \in \mathbf{\Xi} \mid (\forall i \in \{1, \cdots, m\}) \quad a^{(i)} \in S_i\} \quad (25)$$

and $\mathbf{D}$ be the diagonal vector subspace, i.e.

$$\mathbf{D} = \{(a, \cdots, a) \in \mathbf{\Xi} \mid a \in \Xi\}. \quad (26)$$

Henceforth, if $\mathbf{a}$ is a point in $\mathbf{D}$, we shall write it as $\mathbf{a} = (a, \cdots, a)$. It is noted that $\mathbf{S}$ and $\mathbf{D}$ are closed and convex subsets of Ξ. Moreover, it follows immediately from (25) and (26) that

$$\mathbf{S} \bigcap \mathbf{D} = \{(a, \cdots, a) \in \mathbf{\Xi} \mid (\forall i \in \{1, \cdots, m\}) \quad a \in S_i\}. \quad (27)$$

Therefore, in the product space $\mathbf{\Xi}$, we can reformulate the feasibility problem (1) as

$$\text{Find } \mathbf{a}^* \in \mathbf{S} \bigcap \mathbf{D}. \quad (28)$$

This product space characterization of (1) was first given in [27] (see also [28]). Note that in the inconsistent case $\mathbf{S} \bigcap \mathbf{D} = \emptyset$.

Proposition 8: [28] We have

$$\begin{cases} (\forall \mathbf{a} \in \mathbf{D}) \quad P_{\mathbf{S}}(\mathbf{a}) = (P_1(a), \cdots, P_m(a)). \\ (\forall \mathbf{a} \in \mathbf{S}) \quad P_{\mathbf{D}}(\mathbf{a}) = (\sum_{i=1}^m w_i a^{(i)}, \cdots, \sum_{i=1}^m w_i a^{(i)}). \end{cases} \quad (29)$$

□

We shall now proceed to derive a product space formulation of the weighted least-squares feasibility problem (7). Let us define $\mathbf{G}$ as in (17) and $\mathbf{\Phi}$ as in (18), and recall that G is the set of minimizers of the proximity function Φ in (8).

Theorem 3: In the product space $\mathbf{\Xi}$, the weighted least-squares problem (7) is equivalent to minimizing $\mathbf{\Phi}$, i.e., to solve

$$\text{Find } \mathbf{a}^* \in \mathbf{G}. \quad (30)$$

□

Proof: Let $\mathbf{a} = (a, \cdots, a)$ be any point in $\mathbf{D}$. We have

$$\begin{aligned} \Phi(a) &= \sum_{i=1}^m w_i d(a, S_i)^2/2 \\ &= \sum_{i=1}^m w_i \|a - P_i(a)\|^2/2 \\ &= |||(a, \cdots, a) - (P_1(a), \cdots, P_m(a))|||^2/2 \\ &= |||\mathbf{a} - P_{\mathbf{S}}(\mathbf{a})|||^2/2 \\ &= \mathbf{d}(\mathbf{a}, \mathbf{S})^2/2 \\ &= \mathbf{\Phi}(\mathbf{a}). \end{aligned} \quad (31)$$

Therefore, if Φ is minimized for a^*, then $\mathbf{\Phi}$ is minimized for $\mathbf{a}^*$, and vice-versa. But, according to Proposition 4, the set of minimizers of $\mathbf{\Phi}$ is $\mathbf{G}$. Hence, we obtain $\mathbf{G} = \{(a, \cdots, a) \in \mathbf{\Xi} \mid a \in G\}$. ◇

Thus, just as (28) was equivalent to (1), the more general problem (30) is equivalent to (7). From this, the advantage of the product space formalization becomes immediately apparent: it allows us to reduce a problem with m convex sets to

one with only two, one of which is a simple vector subspace. Practically speaking, Theorem 3 can be interpreted as follows: the problem of finding a signal that minimizes the weighted average of the squares of the distances to the property sets is equivalent to that of finding an m-tuple of identical signals that lies at minimum distance from the Cartesian product of the property sets. Mathematically, this is simply equivalent to finding a point in $\mathbf{D}$ which is at minimum distance from $\mathbf{S}$, i.e., a point in $\mathbf{G}$. According to Theorems 1 and 2 and Proposition 7, as long as one of the property sets is bounded, such a point exists and can be obtained iteratively via (19) or (23).

IV. Simultaneous Projections Methods

In the previous section we have solved the least-squares feasibility problem (7) in the product space $\boldsymbol{\Xi}$. It remains to reformulate the solution methods in the original signal space Ξ, where they will actually be employed.

A. Theoretical Convergence Results

We can now state our main result, namely the convergence of parallel projection methods to a weighted least-squares solution of the feasibility problem. The practical significance of this result is that, although the constraints may be inconsistent, they will be approximately satisfied, in a weighted least-squares sense, by the signal generated by the simultaneous projections algorithms. From a signal processing point of view, such a solution is clearly more acceptable and useful than that generated by POCS, whose properties are generally elusive.

Theorem 4: Suppose that one of the S_is is bounded. Then, for any a_0 in Ξ, every sequence of iterates $(a_n)_{n\geq 0}$ defined by

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n + \lambda_n \left(\sum_{i=1}^{m} w_i P_i(a_n) - a_n \right) \tag{32}$$

where $(\lambda_n)_{n\geq 0}$ is as in (3), converges weakly to a point in G. □

Proof: It follows from Proposition 7 that $G \neq \emptyset$. In addition, Proposition 8 yields

$$(\forall \mathbf{a} \in \mathbf{D}) \quad P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}) = \left(\sum_{i=1}^{m} w_i P_i(a), \cdots, \sum_{i=1}^{m} w_i P_i(a) \right). \tag{33}$$

Therefore, (19) in $\boldsymbol{\Xi}$ yields (32) in Ξ and Theorem 4 is a corollary of Theorem 1. ◇

Henceforth, we shall refer to the iteration process (32) as the parallel projection method (PPM). Particular cases of PPM have already been studied in the literature via direct approaches in the original space. Thus, Theorem 4 generalizes a result of [13], which was restricted to half-spaces in finite dimensional spaces and could therefore be applied only to linear inequality constraints. It also generalizes a result of [14], which assumed constant relaxations in (32). Other studies have focused on the unrelaxed form of (32), namely

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = \sum_{i=1}^{m} w_i P_i(a_n). \tag{34}$$

In this case, for consistent problems in finite dimensional spaces, Theorem 4 had been established in [2]. In tomographic signal reconstruction, each property set S_i is a hyperplane in $\mathbb{R}^k$ and (34) with $w_i = 1/m$ is known as the simultaneous iterative reconstruction technique (SIRT) [16]. It is reminiscent of Cimmino's method [7] which solves systems of linear equations in $\mathbb{R}^k$ by averaging reflections (rather than projections) with respect to the hyperplanes. Now, define $I_n^- = \{i \in I \mid a_n \notin S_i\}$ and let $\mu(I_n^-)$ be the number of points in I_n^-. In [21], the relaxation scheme

$$(\forall n \in \mathbb{N}) \quad \lambda_n = \begin{cases} 1/\sum_{i\in I_n^-} w_i, & \text{if } \mu(I_n^-) \geq 2 \\ 1, & \text{otherwise} \end{cases} \tag{35}$$

was used in lieu of (3) to study the convergence of (32) in finite dimensional spaces. This scheme was shown to achieve an acceleration effect due to the fact that some relaxations may be larger than 2. However, convergence to a least-squares solution was proved only locally, i.e., for iterations starting at a point a_0 sufficiently close to G. The following proposition describes PPM as a steepest-descent method, which will give more insight about its convergence behavior.

Proposition 9: Let $(a_n)_{n\geq 0}$ be any sequence of iterates in Theorem 4. Then $(\Phi(a_n))_{n\geq 0}$ decreases until convergence and

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n - \lambda_n \nabla \Phi(a_n) \tag{36}$$

where ∇ is the gradient operator in Ξ. □

Proof: It follows directly from (31) and Proposition 6. ◇

As indicated in Section II-E, strong convergence of PPM can be obtained by imposing certain conditions on the sets. We now present a variant of PPM for which strong convergence is guaranteed without further assumptions on the sets.

Theorem 5: Suppose that one of the S_is is bounded. Then, for any a_0 in Ξ, every sequence of iterates $(a_n)_{n\geq 0}$ defined by

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = (1-\alpha_n)a_0 + \alpha_n \left(\lambda \sum_{i=1}^{m} w_i P_i(a_n) + (1-\lambda) a_n \right) \tag{37}$$

where $(\alpha_n)_{n\geq 0}$ is as in (12) and $0 < \lambda \leq 2$, converges strongly to the projection of a_0 onto G. □

Proof: Similar to that of Theorem 4: (23) in $\boldsymbol{\Xi}$ yields (37) in Ξ and Theorem 5 is a corollary of Theorem 2. ◇

It is worth noting that (37) not only converges strongly to a least-squares-feasible solution but also guarantees that this solution is the closest to the initial point a_0. This property is very valuable in certain signal synthesis problems, where one seeks the best feasible approximation of a reference signal a_0 [10]. In particular, if a_0 is the zero signal, (37) yields the least-squares-feasible signal with minimum energy.

B. Numerical Considerations and Implementation

In this section, we discuss the practical issues pertaining to the numerical realization of the proposed methods on a digital computer. In this particular context, the signals are discretized and have finite support so that the underlying Hilbert space is the usual N-dimensional euclidean space. Hence, weak convergence in Theorem 4 becomes strong convergence and

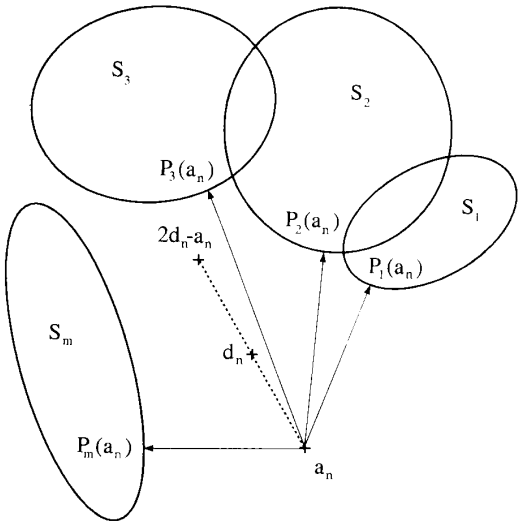

Fig. 2. Iteration in the original space.

the algorithm of interest is PPM (32), which more flexible than (37) for solving (7).

1) Parallel Computing: POCS is a serial method in the sense that only one of the sets can be activated at every iteration. A salient feature of PPM is its parallelism: at every iteration the projections can be computed simultaneously on concurrent processors. Thus, the first phase of an iteration of PPM consists in projecting the current signal a_n onto all the sets, a task which can be distributed among m parallel processors. The second phase is a combination phase in which the projections computed by the m processors are averaged to form $d_n = \sum_{i=1}^{m} w_i P_i(a_n)$. The last phase consists in positioning the new iterate a_{n+1} on the segment between a_n and $2d_n - a_n$. This procedure is illustrated in Fig. 2.

2) Relaxation Parameters and Convergence Behavior: According to Proposition 5, (21), and (33), the value of λ_n that will produce the largest step towards a solution point a^* in G at iteration n is

$$\lambda_n^* = \frac{\langle \sum_{i=1}^{m} w_i P_i(a_n) - a_n \mid a^* - a_n \rangle}{\| \sum_{i=1}^{m} w_i P_i(a_n) - a_n \|^2} \geq 1. \quad (38)$$

In fact, we can deduce from (20) that $\|a_{n+1} - a^*\|$ decreases monotonically as λ_n increases from 0 to λ_n^*, where it attains its minimum, and then starts increasing again. Theoretically, the optimal relaxation λ_n^* could be very large and definitely greater than 2. It should be noted that optimal step sizes at each iteration of an algorithm do not systematically ensure faster convergence of the whole sequence of iterates to a solution. However, we have found this to be the case for PPM, which is in agreement with other studies on parallel projection methods [21], [27], [28], where it was found that overrelaxations had an accelerating effect on the progression of the iterations towards a solution.

In order to find an explicit relaxation rule at each iteration n, let us go back to Proposition 9. Since PPM behaves as a steepest-descent method, we can use the so-called Armijo relaxation scheme which consists in successively reducing the relaxation parameter λ_n until the inequality $\Phi(a_n) - \Phi(a_{n+1}) \geq \alpha\lambda_n \|\nabla\Phi(a_n)\|^2$ is satisfied [30]. In our applications, this adaptation scheme yielded overrelaxations that converged efficiently. It must be noted that PPM will tend to converge slowly when the iterations approach a solution since the gradient will then become small. If the iterations have come sufficiently close to a solution, the relaxation strategy could be switched to the locally convergent scheme (35). If the number of violated constraints is sufficiently small, this may provide larger relaxations than with Armijo's scheme.

3) Weights on the Projections: In the inconsistent case, the values of the weights $(w_i)_{1\leq i\leq m}$ have a direct influence on the solution. More specifically, the larger a particular weight w_i, the closer the solution to the corresponding set S_i. Hence, if some constraints are judged to be more critical than others in defining a least-squares-feasible solution, they should be assigned larger weights. For problems in which no particular group of constraints should be privileged, the weights should be taken to be equal, that is, $w_i = 1/m$.

4) Stopping Rules: The practical goal of PPM is to obtain an approximate minimum of the proximity function Φ in a finite number of steps. According to Proposition 9, the proximity function decreases at every iteration. Hence, the algorithm can be stopped when negligible improvement in the decrease of Φ is observed, i.e., whenever the stopping criterion

$$\Phi(a_n) - \Phi(a_{n+1}) \leq \epsilon \quad (39)$$

is met for a suitably small positive number ϵ. An alternative way of determining the near convergence of the algorithm is to measure the norm of the gradient, which leads to the stopping rule

$$\|\nabla\Phi(a_n)\| = \|a_n - \sum_{i=1}^{m} w_i P_i(a_n)\| \leq \epsilon. \quad (40)$$

5) Proposed Implementation: Based on numerical experience and the recommendations of [30] regarding Armijo's relaxation scheme, we propose the following algorithm as an efficient practical implementation of PPM.

1) Choose an initial guess $a_0 \in \Xi$, strictly convex weights $(w_i)_{1\leq i\leq m}$, and $\epsilon \in]0.+\infty[$. Set $n = 0$.
2) Set $\nabla\Phi(a_n) = a_n - \sum_{i=1}^{m} w_i P_i(a_n)$ and $\lambda_n = 1.999$.
3) Set $a_{n+1} = a_n - \lambda_n \nabla\Phi(a_n)$.
4) If $\Phi(a_n) - \Phi(a_{n+1}) < \lambda_n \|\nabla\Phi(a_n)\|^2/2$, set $\lambda_n = 0.75\lambda_n$, and return to 3.
5) If $\Phi(a_n) - \Phi(a_{n+1}) > \epsilon$, set $n = n+1$, and return to 2.
6) Stop.

V. Applications and Numerical Results

In this section, we investigate the numerical behavior of PPM and demonstrate its application to two standard set theoretic signal synthesis problems. The first problem is signal deconvolution, which is an estimation problem, and the second is pulse shape synthesis, which is a design problem. Unless otherwise stated, PPM is implemented as in Section IV-B-5 with equal weights on all the sets, and POCS as in (4). The underlying Hilbert space is the N-dimensional euclidean space equipped with the usual metric.

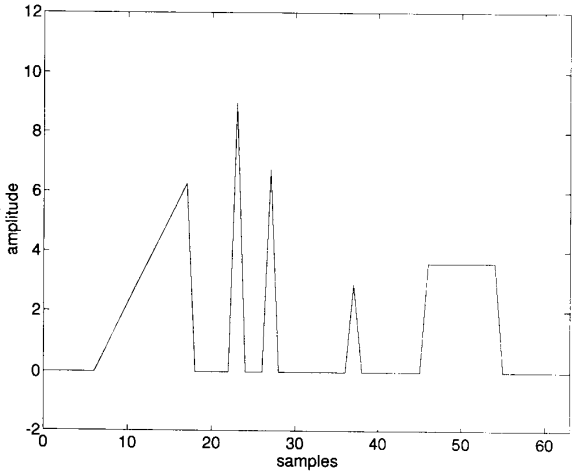

Fig. 3. Original signal.

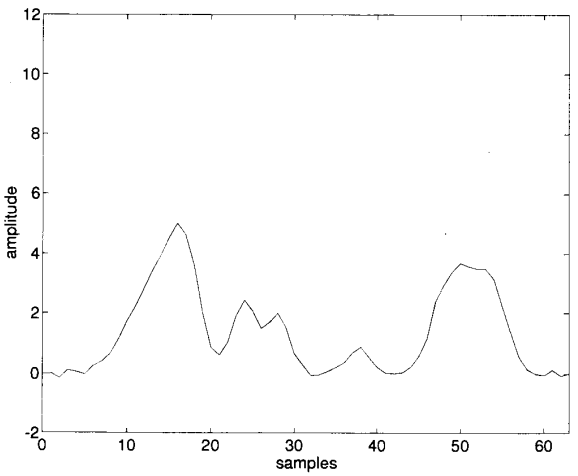

Fig. 4. Degraded signal.

A. Signal Deconvolution

1) Experiment: We consider the problem of deconvolving a noisy discrete-time N-point signal, i.e., of estimating the original form of a signal h which has been passed through a linear shift-invariant system and further degraded by addition of noise. The length of the signals is $N = 64$. The original signal h is shown in Fig. 3. The recorded signal x of Fig. 4 was obtained via the standard linear model $x = Lh + u$, in which the $N \times N$ matrix L models a shift-invariant linear blur and u is a vector of bounded noise samples with $|u_i| \leq 0.15$. The blurring kernel is a Gaussian function with a variance of 2 samples^2.

2) Set Theoretic Formulation: The set theoretic formulation for the problem consists of $m = 66$ closed and convex sets. The sets used here have already been used in various set theoretic signal processing applications, e.g., [12], [22], [33], [36]. The sets $(S_i)_{1\leq i\leq N}$ are based on the knowledge of the blurring operator L and the information that the components of the noise vector u are bounded by $\delta = 0.15$. They take the form of hyperslabs defined by

$$S_i = \{a \in \mathbb{R}^N \mid x_i - \delta \leq \langle L_i \mid a\rangle \leq x_i + \delta\}, \quad \text{for } 1 \leq i \leq N \tag{41}$$

where L_i is the ith row of L. The projection $P_i(a)$ of a signal a onto S_i is given by [33]

$$\begin{cases} a + [(x_i + \delta - \langle L_i \mid a\rangle)/\|L_i\|^2]L_i, & \text{if } \langle L_i \mid a\rangle > x_i + \delta \\ a + [(x_i - \delta - \langle L_i \mid a\rangle)/\|L_i\|^2]L_i, & \text{if } \langle L_i \mid a\rangle < x_i - \delta \\ a, & \text{otherwise.} \end{cases} \tag{42}$$

The next set is constructed by assuming knowledge of the phase of h. If A denotes the discrete Fourier transform of the signal a, this constraint leads to the set

$$S_{m-1} = \{a \in \mathbb{R}^N \mid (\forall k \in \{1, \cdots, N\}) \quad \angle A(k) = \angle H(k)\}. \tag{43}$$

The projection $P_{m-1}(a) = b$ of a signal a onto S_{m-1} is such that for every k in $\{1, \cdots, N\}$ [36] (see (44) at the bottom of this page). The last set arises from the prior knowledge that the components of h are nonnegative and bounded by 12. This leads to the bounded set

$$S_m = \{a \in \mathbb{R}^N \mid (\forall i \in \{1, \cdots, N\})\ 0 \leq a_i \leq 12\}. \tag{45}$$

The projection of a signal a onto S_m is given by $P_m(a) = b$, where for every i in $\{1, \cdots, N\}$

$$b_i = \begin{cases} 0, & \text{if } a_i < 0 \\ 12, & \text{if } a_i > 12 \\ a_i, & \text{otherwise.} \end{cases} \tag{46}$$

3) Result: All algorithms are initialized with the degraded signal, i.e., $a_0 = x$. The feasible signal of Fig. 5 is obtained by POCS. It is seen that most features of h have been fairly well recovered. Next, we introduce inaccuracies in the specifications of the *a priori* information that will induce an inconsistent set theoretic formulation: the variance of the Gaussian impulse response of the system is taken to be 2.5 samples^2 instead of 2, the bound on the noise is taken to be 0.1 instead of 0.15, and the phase of h is recorded in 10 dB of background noise. The limiting signal of the subsequence $(a_{nm})_{n\geq 0}$ generated by POCS in this case is depicted in Fig. 6 and its convergence behavior is shown in Fig. 7, where the values taken by the proximity function $(\Phi(a_{nm}))_{n\geq 0}$ are plotted. As discussed in the Introduction, the only definite property of this signal is to lie in S_m and, thereby, to satisfy

$$B(k) = \begin{cases} |B(k)|\cos(\angle A(k) - \angle H(k))\exp(\imath \angle H(k)), & \text{if } \cos(\angle A(k) - \angle H(k)) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{44}$$

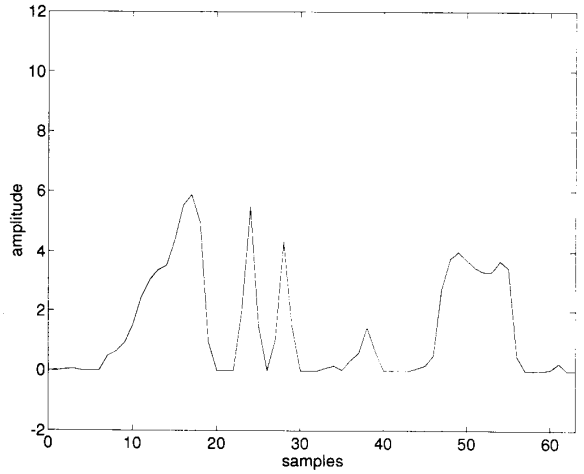

Fig. 5. Consistent case—Deconvolution by POCS.

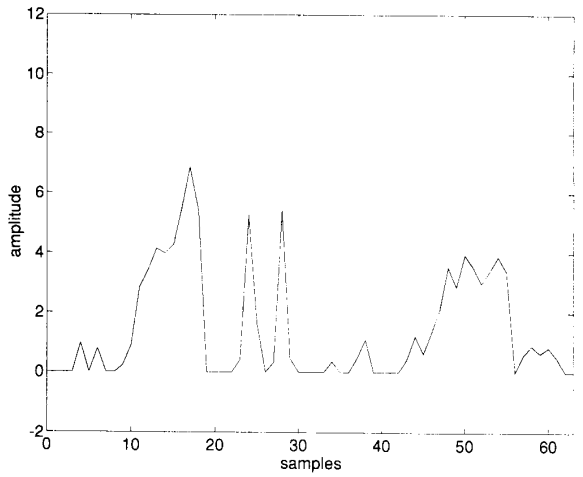

Fig. 6. Inconsistent case—Deconvolution by POCS.

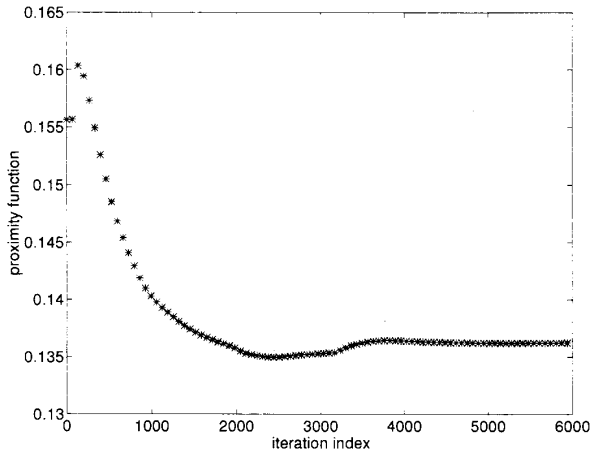

Fig. 7. Inconsistent case—Convergence of POCS.

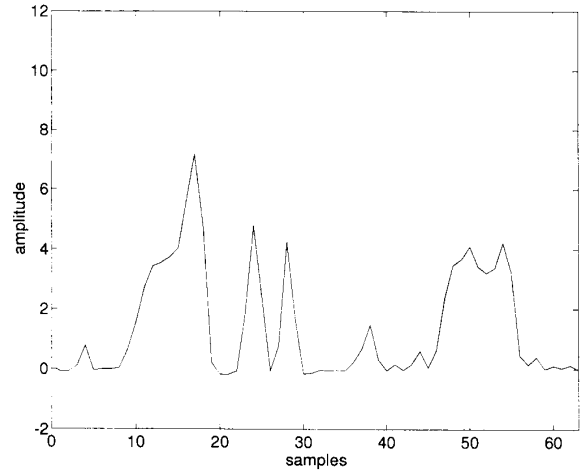

Fig. 8. Inconsistent case—Deconvolution by PPM.

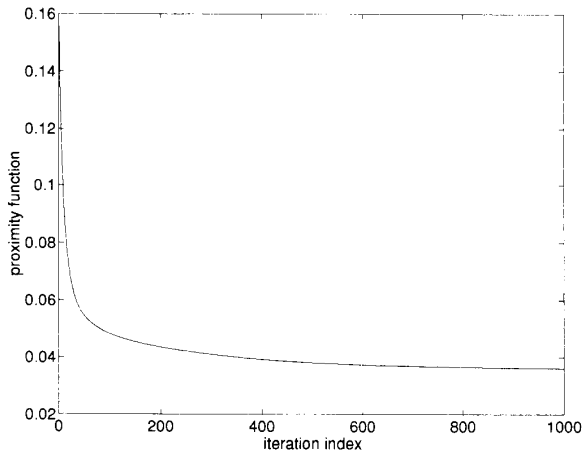

Fig. 9. Inconsistent case—Convergence of PPM.

the amplitude constraints. It is noted that the limiting value of the proximity function (degree of unfeasibility) achieved by POCS is about 0.136. PPM is then employed to produce the restored signal shown in Fig. 8. This least-squares solution to the inconsistent feasibility problem has less artifacts than the solution generated by POCS. The sequence $(\Phi(a_n))_{n\geq 0}$ is shown in Fig. 9. In comparison, PPM achieves a much lower asymptotic degree of unfeasibility of $\Phi(a_\infty) = 0.035$. Finally, Fig. 10 depicts the convergence behavior of PPM subjected to various relaxations schemes. In the underrelaxed case A, $(\lambda_n)_{n\geq 0}$ is generated at random from the interval [0, 1]; in the unrelaxed case B the relaxations are equal to 1; in the overrelaxed case C, $(\lambda_n)_{n\geq 0}$ is generated at random from the interval [1, 2]; in the adapted case D, $(\lambda_n)_{n\geq 0}$ is obtained as in Section IV-B-5. These plots support the claims of Section IV-B-2 that overrelaxations are more effective than underrelaxations and that Armijo's adapted relaxation scheme is preferable. In all cases, $(\Phi(a_n))_{n\geq 0}$ is decreasing, in conformity with Proposition 9.

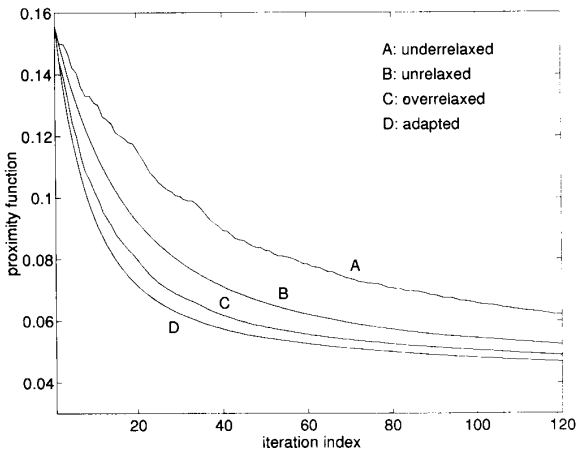

Fig. 10. Convergence of PPM for various relaxations schemes.

B. Pulse Shape Synthesis

1) Experiment: In this second example, PPM is applied to the digital design of pulse shapes for digital data transmission over European 50 Hz power lines. This problem was addressed in [8] with POCS and two design constraints. It amounts to designing an FIR filter under certain constraints. We shall assume that the sampling rate of the system is 2560 Hz and that the length of the FIR filter used to generate the digital pulses is $N = 512$ points. The design specifications are as follows.

a) The lines have a bandwidth of 300 Hz and are contaminated by a dc component and the harmonic noise of the power distribution system. To avoid dc and harmonic noise and be compatible with the available bandwidth, the Fourier transform of the pulse should be zero at the zero frequency, at integer multiples of 50 Hz, and beyond 300 Hz.

b) The pulse should have linear phase and be normalized so that its midpoint has amplitude 1.

c) The energy of the pulse should not exceed $\xi = 4$ to avoid interference with other systems.

d) The actual duration of the pulse should be 50 ms and it should have periodic zero crossings every 3.91 ms to avoid intersymbol interference.

These constraints include all the desirable properties of the pulse shape but are uncompatible.

2) Set Theoretic Formulation: The above constraints give rise to $m = 4$ sets. The projection operators are straightforward and need not be derived here. The spectral constraints a) lead to the set

$$S_1 = \{a \in \mathbb{R}^N \mid (\forall k \in \mathcal{K}) \quad A(k) = 0\} \tag{47}$$

where A designates the discrete Fourier transform of a and $\mathcal{K}$ contains the indices of the discrete frequencies that should be removed. The projection of a pulse shape a onto S_1 is the pulse shape $P_1(a) = b$, where for every k in $\{1, \cdots, N\}$

$$B(k) = \begin{cases} 0, & \text{if } k \in \mathcal{K} \\ A(k), & \text{otherwise.} \end{cases} \tag{48}$$

The constraint b) leads to the set

$$S_2 = \{a \in \mathbb{R}^N \mid a_{N/2} = 1 \text{ and } (\forall i \in \{1, \cdots, N\}) \quad a_i = a_{N+1-i}\}. \tag{49}$$

The projection of a pulse shape a onto S_2 is $P_2(a) = b$, where

$$\begin{cases} (\forall i \in \{1, \cdots, N\}) & b_i = (a_i + a_{N+1-i})/2, \\ b_{N/2} = b_{N/2+1} = 1. \end{cases} \tag{50}$$

The energy constraint c) is associated with the bounded set

$$S_3 = \{a \in \mathbb{R}^N \mid \|a\|^2 \leq \xi\}. \tag{51}$$

The projection of a pulse shape a onto S_3 is

$$P_3(a) = \begin{cases} \sqrt{\xi}a/\|a\|, & \text{if } \|a\|^2 > \xi \\ a, & \text{otherwise.} \end{cases} \tag{52}$$

The last set arises from the time-domain constraints d)

$$S_4 = \{a \in \mathbb{R}^N \mid (\forall i \in \mathcal{I}) \quad a_i = 0\} \tag{53}$$

where $\mathcal{I}$ contains the indices of the components in the zero areas. The projection of a pulse shape a onto S_4 is $b = P_4(a)$, where for every i in $\{1, \cdots, N\}$

$$b_i = \begin{cases} 0, & \text{if } i \in \mathcal{I} \\ a_i, & \text{otherwise.} \end{cases} \tag{54}$$

Since the constraints a)-d) are incompatible, $\bigcap_{i=1}^m S_i = \emptyset$. It is worth noting that in [8] only the sets S_1 and S_4 were used since, as seen in the Introduction, the convergence behavior of POCS when $m > 2$ lacks a meaningful interpretation in the inconsistent case.

3) Result: All the algorithms are initialized with the zero pulse. The solution obtained as the limit of the sequence $(a_{nm})_{n\geq 0}$ generated by POCS is shown in Fig. 11 and the low frequency portion of its spectral density (normalized to a maximum of 0 dB) in Fig. 12. The asymptotic value of the sequence $(\Phi(a_{nm}))_{n\geq 0}$ shown in Fig. 13 is 0.0371. The limit of the sequence $(a_n)_{n\geq 0}$ generated by PPM is displayed in Figs. 14 and 15. It is seen in Fig. 16 that PPM achieves a lower degree of unfeasibility of $\Phi(a_\infty) = 0.0180$. Qualitatively, the solution produced by PPM appears to suppress the undesirable frequencies better than that of POCS and to be more symmetric in the time-domain. Naturally, it does not exactly satisfy the time-duration constraint since it is not exactly in S_4. Fig. 17, which depicts the behavior of PPM under various relaxation strategies, prompts the same conclusions as in Section V-A-3.

VI. Conclusion

A broad class of signal processing problems consist in producing a signal that satisfies a collection of constraints. In certain instances, this feasibility problem may be inconsistent in the sense that some of the constraints are uncompatible.

In this paper, we have considered the problem of finding weighted least-squares solutions to inconsistent signal feasibility problems. Geometrically, this is tantamount to finding a point that minimizes the weighted average of the squares of the distances to m property sets representing the constraints in the signal space. Such solutions cannot be obtained by the

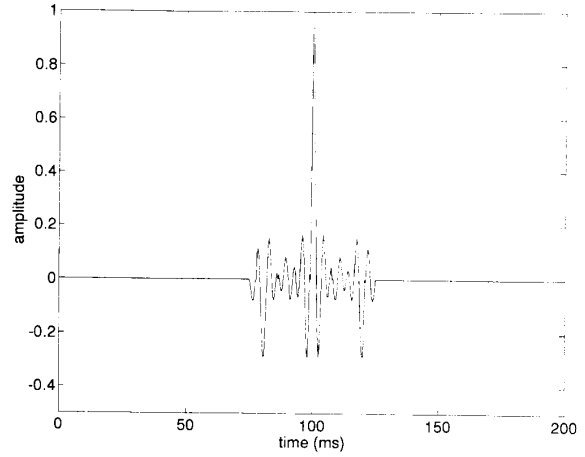

Fig. 11. Pulse generated by POCS.

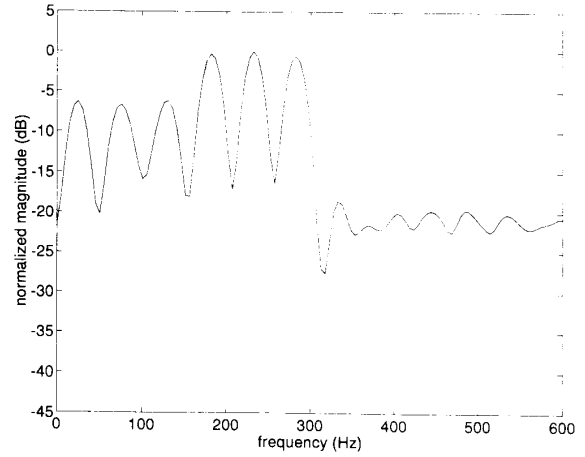

Fig. 12. Normalized spectral density of the pulse generated by POCS.

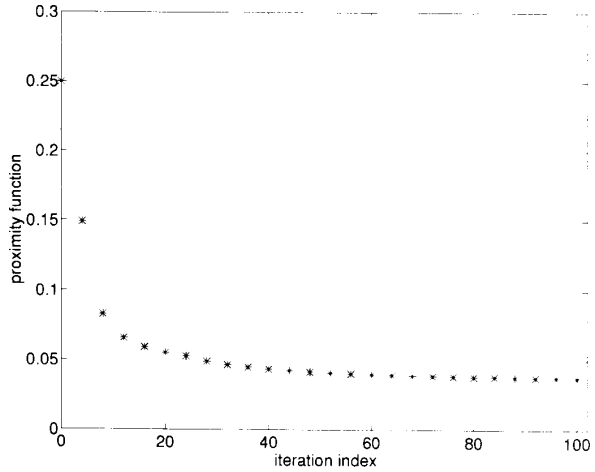

Fig. 13. Convergence of POCS.

method of POCS, which has traditionally prevailed in signal feasibility problems. The problem was recast in the m-fold Cartesian product of the signal space and solved iteratively by alternating projections between the diagonal subspace and the product of the m property sets. This gave rise to simultaneous projections methods in the original signal space that can easily be implemented on parallel processors since all the sets are activated independently at each iteration.

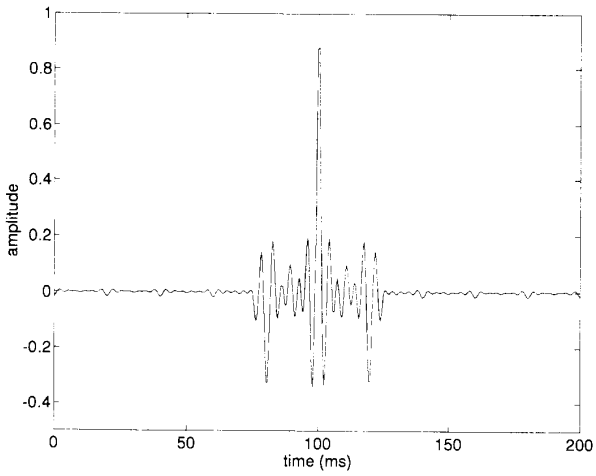

Fig. 14. Pulse generated by PPM.

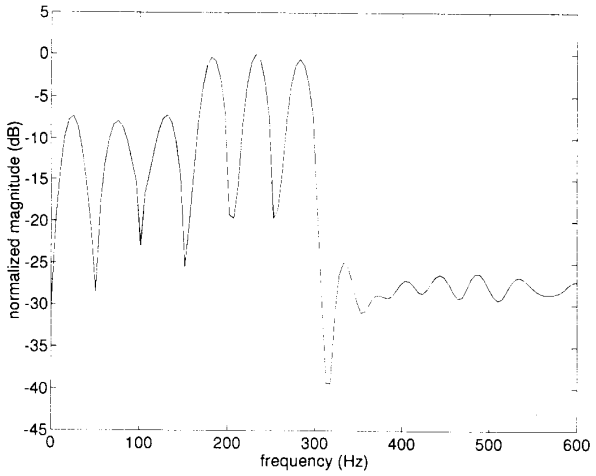

Fig. 15. Normalized spectral density of the pulse generated by PPM.

The main advantage of the product space formalism introduced in this paper is to reduce an m-set problem in the original signal space to a 2-set problem in the corresponding m-fold Cartesian product space, which has proven very useful to treat the question of inconsistent set theoretic formulations. We believe that this formalism will also prove useful in connection with other aspects of set theoretic signal synthesis.

APPENDIX A
PROOFS OF SECTION II

Proof of Proposition 1: Fix $\mathbf{f}$ arbitrarily in $\mathbf{F}$. Then (3), (10), and (11) yield [34, (7)]

$$(\forall n \in \mathbb{N}) \quad |||\mathbf{a}_{n+1} - \mathbf{f}||| \leq |||\mathbf{a}_n - \mathbf{f}|||. \tag{A1}$$

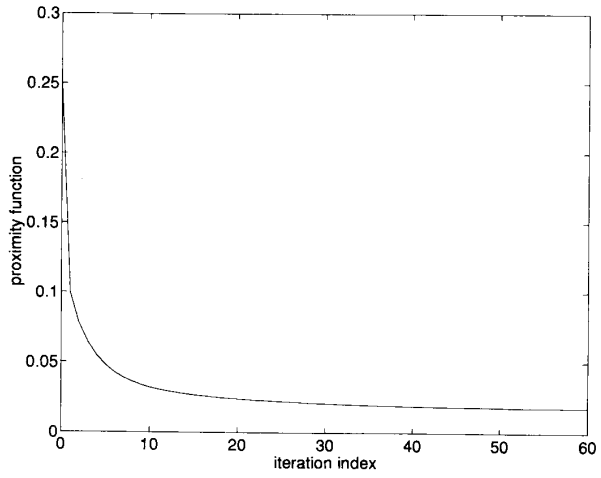

Fig. 16. Convergence of PPM.

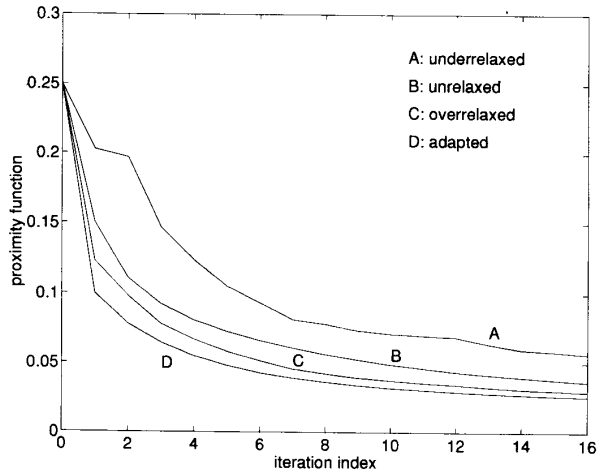

Fig. 17. Convergence of PPM for various relaxations schemes.

Hence $(\mathbf{a}_n)_{n\geq 0} \subset \{\mathbf{b} \in \boldsymbol{\Xi} \mid |||\mathbf{b} - \mathbf{f}||| \leq |||\mathbf{a}_0 - \mathbf{f}|||\} \equiv \mathbf{B}$. But $\mathbf{B}$ is bounded, closed, and convex. Hence $(\mathbf{a}_n)_{n\geq 0}$ possesses a weak cluster point $\mathbf{a}$ and, since $T : \mathbf{B} \to \boldsymbol{\Xi}$ is nonexpansive, $\Upsilon - T$ is demiclosed on $\mathbf{B}$ [38]. Therefore, since $(\|(\Upsilon - T)(\mathbf{a}_n)\|)_{n\geq 0}$ converges to zero [34, (6)], $(\Upsilon - T)(\mathbf{a}) = \mathbf{0}$. Whence $\mathbf{a} \in \mathbf{F}$. But since (A1) implies that $(\mathbf{a}_n)_{n\geq 0}$ cannot have more than one weak cluster point in $\mathbf{F}$, e.g. [3], we conclude that the whole sequence $(\mathbf{a}_n)_{n\geq 0}$ converges weakly to $\mathbf{a}$.[1] ◇

Proof of Proposition 3: i) Let $(\mathbf{a}, \mathbf{b})$ be an arbitrary pair in $\mathbf{D}^2$ and $T = P_{\mathbf{D}} \circ P_{\mathbf{S}}$. Then the firm nonexpansivity of T follows from the relationships

$$\begin{aligned} |||T(\mathbf{a}) - T(\mathbf{b})|||^2 &\leq |||P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b})|||^2 \\ &\leq \langle\langle \mathbf{a} - \mathbf{b} \mid P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b})\rangle\rangle \\ &= \langle\langle \mathbf{a} - \mathbf{b} \mid P_{\mathbf{D}}(P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b}))\rangle\rangle \\ &= \langle\langle \mathbf{a} - \mathbf{b} \mid T(\mathbf{a}) - T(\mathbf{b})\rangle\rangle \qquad \text{(A2)} \end{aligned}$$

[1] Note that we cannot use [34, Theorem 2] directly as it merely states that there is a single weak cluster point in $\mathbf{F}$ and does not exclude the existence of weak cluster points outside of $\mathbf{F}$.

where we have used successively the nonexpansivity of $P_{\mathbf{D}}$, then the firm nonexpansivity of $P_{\mathbf{S}}$, and (16) since $P_{\mathbf{D}}$ is linear and $\mathbf{a} - \mathbf{b} \in \mathbf{D}$; ii) Let $T = \Upsilon + \lambda(P_{\mathbf{S}} - \Upsilon)$ and $\alpha = \lambda/2$. The reflection operator $R_{\mathbf{S}} = 2P_{\mathbf{S}} - \Upsilon$ with respect to $\mathbf{S}$ is nonexpansive [37]. We can write $T = \Upsilon + 2\alpha(P_{\mathbf{S}} - \Upsilon) = (1 - \alpha)\Upsilon + \alpha R_{\mathbf{S}}$, where $\alpha \in [0, 1]$. Since both Υ and $R_{\mathbf{S}}$ are nonexpansive, so is T. ◇

Proof of Theorem 1: By Proposition 3i), $T = P_{\mathbf{D}} \circ P_{\mathbf{S}}$ is firmly nonexpansive. Hence the result is a direct consequence of Proposition 1. ◇

Proof of Proposition 5: If $\mathbf{a}_n \in \mathbf{G}$, there is nothing to prove. Otherwise, let $A = \langle\langle \mathbf{a}_n - T(\mathbf{a}_n) \mid T(\mathbf{a}_n) - \mathbf{a}^*\rangle\rangle$, where $T = P_{\mathbf{D}} \circ P_{\mathbf{S}}$. We have

$$\langle\langle \mathbf{a}_n - T(\mathbf{a}_n) \mid \mathbf{a}_n - \mathbf{a}^*\rangle\rangle = |||\mathbf{a}_n - T(\mathbf{a}_n)|||^2 + A. \qquad \text{(A3)}$$

Therefore, $\lambda_n^* = 1 + A/|||\mathbf{a}_n - T(\mathbf{a}_n)|||^2$. But, since $\mathbf{a}^* \in \mathbf{G}$, Proposition 3i) entails

$$\begin{aligned} |||T(\mathbf{a}_n) - \mathbf{a}^*|||^2 &= |||T(\mathbf{a}_n) - T(\mathbf{a}^*)|||^2 \\ &\leq \langle\langle \mathbf{a}_n - \mathbf{a}^* \mid T(\mathbf{a}_n) - \mathbf{a}^*\rangle\rangle. \qquad \text{(A4)} \end{aligned}$$

Consequently

$$A = \langle\langle \mathbf{a}_n - \mathbf{a}^* \mid T(\mathbf{a}_n) - \mathbf{a}^*\rangle\rangle - |||T(\mathbf{a}_n) - \mathbf{a}^*|||^2 \geq 0 \qquad \text{(A5)}$$

which gives $\lambda_n^* \geq 1$. ◇

Proof of Proposition 6: Take any n such that $\mathbf{a}_n \notin \mathbf{G}$. Then, from the linearity of $P_{\mathbf{D}}$ and (16), we get

$$\begin{aligned} &\langle\langle \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \mid \mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n)\rangle\rangle \\ &\quad = \langle\langle \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \mid P_{\mathbf{D}}(\mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n))\rangle\rangle \\ &\quad = |||\mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \qquad \text{(A6)} \\ &\quad > 0. \qquad \text{(A7)} \end{aligned}$$

Using (A6), (A7), and (3), we then get

$$\begin{aligned} \mathbf{d}(\mathbf{a}_{n+1}, \mathbf{S})^2 &\leq |||\mathbf{a}_{n+1} - P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \\ &= |||\mathbf{a}_{n+1} - \mathbf{a}_n|||^2 \\ &\quad + 2\langle\langle \mathbf{a}_{n+1} - \mathbf{a}_n \mid \mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n)\rangle\rangle \\ &\quad + |||\mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \\ &= \lambda_n^2 |||\mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \\ &\quad - 2\lambda_n \langle\langle \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \mid \mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n)\rangle\rangle \\ &\quad + \mathbf{d}(\mathbf{a}_n, \mathbf{S})^2 \\ &= \lambda_n(\lambda_n - 2)|||\mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)|||^2 \\ &\quad + \mathbf{d}(\mathbf{a}_n, \mathbf{S})^2 \\ &< \mathbf{d}(\mathbf{a}_n, \mathbf{S})^2. \qquad \text{(A8)} \end{aligned}$$

Hence, $\boldsymbol{\Phi}(\mathbf{a}_{n+1}) < \boldsymbol{\Phi}(\mathbf{a}_n)$. To prove the second assertion, recall from Section II-C that in $\boldsymbol{\Xi}$ we have $\nabla \mathbf{d}(\mathbf{a}_n, \mathbf{S})^2 = 2(\mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n))$. Therefore

$$\nabla_{\mathbf{D}} \boldsymbol{\Phi}(\mathbf{a}_n) = P_{\mathbf{D}}(\mathbf{a}_n - P_{\mathbf{S}}(\mathbf{a}_n)) = \mathbf{a}_n - P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) \qquad \text{(A9)}$$

and (22) is equivalent to (19). ◇

Proof of Theorem 2: Let $T = P_{\mathbf{D}} \circ (\lambda(P_{\mathbf{S}} - \Upsilon) + \Upsilon)$. Then, by Proposition 3ii), T is nonexpansive since $P_{\mathbf{D}}$ and $\lambda(P_{\mathbf{S}} - \Upsilon) + \Upsilon$ are. Moreover, since $\mathbf{a}_n \in \mathbf{D}$ and $P_{\mathbf{D}}$ is linear, the update equation in (23) can be written as

$$\mathbf{a}_{n+1} = (1 - \alpha_n)\mathbf{a}_0 + \alpha_n P_{\mathbf{D}}(\lambda P_{\mathbf{S}}(\mathbf{a}_n) + (1 - \lambda)\mathbf{a}_n)$$

$$= (1-\alpha_n)\mathbf{a}_0 + \alpha_n P_{\mathbf{D}}(\lambda(P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n) + \mathbf{a}_n)$$
$$= (1-\alpha_n)\mathbf{a}_0 + \alpha_n T(\mathbf{a}_n). \quad \text{(A10)}$$

The result then follows from Proposition 2, since the set of fixed points of T coincides with $\mathbf{G}$. $\diamond$

Proof of Proposition 7: As seen in Section II-C, the functions $(d(\cdot, S_i))_{1\le i\le m}$ are continuous and convex. Since $x \mapsto |x|^2$ is continuous, convex, and increasing on $\mathbb{R}_+$, the functions $(d(\cdot, S_i)^2)_{1\le i\le m}$ are also continuous and convex and so is their convex combination Φ in (8). Now suppose that S_i is bounded. Then, $d(a, S_i) \to +\infty$ as $\|a\| \to +\infty$. Since $(\forall a \in \Xi)\ \Phi(a) \ge w_i d(a, S_i)^2/2$, we get that $\Phi(a) \to +\infty$ as $\|a\| \to +\infty$, i.e., Φ is weakly coercive. Since the set of minimizers of a continuous, convex, and weakly coercive functional on a Hilbert space is nonempty, closed, convex, and bounded [39], the proof is complete. $\diamond$

REFERENCES

[1] J. P. Aubin, *Optima and Equilibria—An Introduction to Nonlinear Analysis*. New York: Springer-Verlag, 1993.

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

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

[4] Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," *Numer. Math.*, vol. 41, no. 1, pp. 83–92, Apr. 1983.

[5] A. E. Cetin, "An iterative algorithm for signal reconstruction from bispectrum," *IEEE Trans. Signal Processing*, vol. 39, no. 12, pp. 2621–2628, Dec. 1991.

[6] W. Cheney and A. A. Goldstein, "Proximity maps for convex sets," in *Proc. Amer. Math. Soc.*, vol. 10, no. 3, pp. 448–450, June 1959.

[7] G. Cimmino, "Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari," *La Ricerca Scientifica (Roma)*, vol. 1, pp. 326–333, 1938.

[8] M. R. Civanlar and R. A. Nobakht, "Optimal pulse shape design using projections onto convex sets," *Proc. IEEE ICASSP Conf.* (New York, NY), Apr. 11–14, 1988, pp. 1874–1877. See also 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.

[9] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.

[10] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 269–271, Apr. 1993.

[11] P. L. Combettes and H. J. Trussell, "Methods for digital restoration of signals degraded by a stochastic impulse response," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, no. 3, pp. 393–401, Mar. 1989.

[12] P. L. Combettes and H. J. Trussell, "The use of noise properties in set theoretic estimation," *IEEE Trans. Signal Processing*, vol. 39, no. 7, pp. 1630–1641, July 1991.

[13] A. R. De Pierro and A. N. Iusem, "A simultaneous projections method for linear inequalities," *Linear Algebra Applicat.*, vol. 64, pp. 243–253, Jan. 1985.

[14] A. R. De Pierro and A. N. Iusem, "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.

[15] J. A. Dieudonné, *Foundations of Modern Analysis*, 2nd ed. New York: Academic, 1969.

[16] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theoret. Biol.*, vol. 36, no. 1, pp. 105–117, July 1972.

[17] 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.

[18] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *USSR Comput. Math., Math. Phys.*, vol. 7, no. 6, pp. 1–24, 1967.

[19] G. T. Herman, "A relaxation method for reconstructing objects from noisy X-rays," *Math. Programming*, vol. 8, no. 1, pp. 1–19, Feb. 1975.

[20] G. T. Herman, *Image Reconstruction from Projections, the Fundamentals of Computerized Tomography*. New York: Academic, 1980.

[21] A. N. Iusem and A. R. De Pierro, "Convergence results for an accelerated nonlinear Cimmino algorithm," *Numer. Math*, vol. 49, no. 4, pp. 367–378, Aug. 1986.

[22] A. Levi and H. Stark, "Signal reconstruction from phase by projection onto convex sets," *J. Opt. Soc. Am. A*, vol. 73, no. 6, pp. 810–822, June 1983.

[23] P. L. Lions, "Approximation de points fixes de contractions," *C. R. Acad. Sci. Paris*, vol. A284, no. 21, pp. 1357–1359, June 1977.

[24] K. M. Nashold and B. E. A. Saleh, "Image construction through diffraction-limited high-contrast imaging systems: An iterative approach," *J. Opt. Soc. Am. A*, vol. 2, no. 5, pp. 635–643, May 1985.

[25] S. Oh, C. Ramon, R. J. Marks II, A. C. Nelson, and M. G. Meyer, "Resolution enhancement of biomagnetic images using the method of alternating projections," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 4, pp. 323–328, Apr. 1993.

[26] 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.

[27] G. Pierra, "Méthodes de projections parallèles extrapolées relatives à une intersection de convexes," Rapport de Recherche, INPG, Grenoble, France, Sept. 1975.

[28] G. Pierra, "Decomposition through formalization in a product space," *Math. Programming*, vol. 28, no. 1, pp. 96–115, Jan. 1984.

[29] S. V. Plotnikov, "Cyclic projection on a system of convex sets with empty intersection," (in Russian) in *Improper Optimization Problems*, I. I. Eremin and V. D. Skarin, Eds. Sverdlovsk: Akademiia Nauk SSSR, 1982, pp. 60–66.

[30] E. Polak, *Computational Methods in Optimization: A Unified Approach*. New York: Academic, 1971.

[31] L. Schwartz, *Analyse Hilbertienne*. Paris: Hermann, 1979.

[32] M. I. Sezan and H. Stark, "Image restoration by convex projections in the presence of noise," *Appl. Optics*, vol. 22, no. 18, pp. 2781–2789, Sept. 1983.

[33] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 2, pp. 201–212, Apr. 1984.

[34] P. Tseng, "On the convergence of products of firmly nonexpansive mappings," *SIAM J. Optim.*, vol. 2, no. 3, pp. 425–434, Aug. 1992.

[35] 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.

[36] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imaging*, vol. 1, no. 2, pp. 81–94, Oct. 1982.

[37] E. H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory," in *Contributions to Nonlinear Functional Analysis*, E. H. Zarantonello, Ed. New York: Academic, 1971, pp. 237–424.

[38] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems*. New York: Springer-Verlag, 1986.

[39] E. Zeidler, *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*. New York: Springer-Verlag, 1985.

Patrick L. Combettes (S'84–M'89) was born on May 24, 1962 in Salon de Provence, France. 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, North Carolina State University, Raleigh. In August 1990, he joined the Department of Electrical Engineering at the City College and the Graduate School of the City University of New York, New York, where he is now an Associate Professor. He has received the IEEE Signal Processing Society 1993 Paper Award (Image and Multidimensional Signal Processing Area) for a paper coauthored by Joel Trussell on set theoretic estimation and has held several Visiting Scientist positions at the CNRS, Laboratoire des Signaux et Systèmes, Gif sur Yvette, France. His current research interests are in signal theory.

Dr. Combettes is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, the American Mathematical Society, and the IEEE Signal Processing and Information Theory Societies.