The Convex Feasibility Problem in Image Recovery*

P. L. Combettes

Department of Electrical Engineering
City College and Graduate School
City University of New York
New York, NY 10031, USA

*Vol. 95 of: ADVANCES IN IMAGING AND ELECTRON PHYSICS. ACADEMIC PRESS, NEW YORK, 1996.
*This work was supported by the National Science Foundation under grant MIP-9308609.

Contents

1 Introduction

1.1 The Image Recovery Problem

Image recovery is a broad discipline that encompasses the large body of inverse problems in which an image h is to be inferred from the observation of data x consisting of signals physically or mathematically related to it. The importance of image recovery stems from the growing need for visual information in a wide spectrum of environmental, medical, military, industrial, and artistic fields. More specifically, we can mention scientific applications in astronomy, bioengineering, electron microscopy, interferometry, ultrasonic imaging, flow imaging, radiology, surveillance, nondestructive testing, seismology, and satellite imaging. General references on image recovery and its applications are [5], [53], [98], [156], and [159].

Image restoration and image reconstruction are the two main sub-branches of image recovery. The term image restoration usually applies to the problem of estimating the original form h of a degraded image x. Hence, in image restoration the data consist of measurements taken directly on the image to be estimated, x being a blurred and noise-corrupted version of h. The blurring operation can be induced by the image transmission medium, e.g., the atmosphere in astronomy, or by the recording device, e.g., an out-of-focus or moving camera. On the other hand, image reconstruction refers to problems in which the data x are indirectly related to the form of the original image h. For example, the term reconstruction would apply to the problem of estimating an image given measurement of its line integrals in tomography or given partial diffraction data in extrapolation problems.

Four basic elements are required to solve an image recovery problem.

1. A data formation model.
2. *A priori* information.
3. A recovery criterion.
4. A solution method.

The data formation model is essentially a model of the imaging system, i.e., a mathematical description of the relation between the original image h and the recorded data x. One of the most common data formation models in image restoration is

$$x = T(h) + u, \tag{1.1}$$

where the operator T represents the blurring process and u an additive noise component. Within this generic model, various subcategories can be distinguished, according as T is

linear or nonlinear, deterministic or stochastic, or according as the noise depends on $T(h)$ or not, etc. Different models can also be considered to reflect situations when the noise is multiplicative, or when several noise sources are present, etc. The basic model (1.1) is also appropriate in a number of image reconstruction problems. For instance, $T(h)$ will stand for a low-passed Fourier transform in band-limited extrapolation and a Radon transform in tomography.

A data formation model is always accompanied by some *a priori* knowledge. Thus, in (1.1), information may be available to describe the original image h, the operator T, or the noise u. As emphasized in [170], *a priori* information is an essential ingredient in recovery problems, even if it is often exploited only partially.

The recovery criterion defines the class of images that are acceptable as solutions to the problem. It is chosen by the user on grounds that may include experience, compatibility with the available *a priori* knowledge, personal convictions on the best way to solve the problem, and ease of implementation. The traditional approach has been to use a criterion of optimality, which usually leads to a single "best" solution. An alternative approach is to use a criterion of feasibility, in which consistency with all prior information and the data defines a set of equally acceptable solutions. This will be the framework discussed in this survey.

The solution method is a numerical algorithm that will produce a solution to the recovery problem, i.e., an image that satisfies the recovery criterion. This computational aspect of image recovery is critical, as it restricts the choice of recovery criteria. Indeed, a physically founded criterion may yield a numerical problem for which no solution technique is available and it can therefore not be adopted.

A conceptual formulation of recovery problems in a Hilbert image space Ξ is

$$\min_{a\in\Xi} \Theta(a) \quad \text{subject to constraints} \quad (\Psi_i)_{i\in I}, \tag{1.2}$$

where the functional Θ represents the cost to be minimized[1] and where the constraints $(\Psi_i)_{i\in I}$ arise from *a priori* knowledge and the observed data. A collection of property sets can be defined in Ξ by

$$(\forall i \in I) \;\; S_i = \{a \in \Xi \mid a \text{ satisfies } \Psi_i\}. \tag{1.3}$$

The feasibility set for the problem is the class of all images that are consistent with all the constraints, that is

$$S = \bigcap_{i\in I} S_i = \{a \in \Xi \mid (\forall i \in I) \;\; a \text{ satisfies } \Psi_i\}. \tag{1.4}$$

[1] If a cost Θ is to be maximized, we shall simply minimize $-\Theta$.

Therefore, (1.2) takes the form

$$\min_{a\in\Xi} \Theta(a) \quad \text{subject to} \quad a \in \bigcap_{i\in I} S_i. \tag{1.5}$$

This quite general constrained programming problem can usually not be solved and it must therefore be modified. Modification can be made in two directions: in the conventional image recovery framework, one seeks to preserve the notion of an optimal solution whereas in the set theoretic framework the emphasis is placed on feasibility.

1.2 Optimal Solutions and Point Estimates

In most enginering problems the criterion of optimality with respect to a unimodal cost function Θ has been used to define unique solutions. The systematic quest for optimal solutions, which is now well rooted in the scientific culture, originated in the late 1940's. It has been fueled to a large extent by the conjunction of technological advances in computing machinery as well as progress in branches of applied mathematics such as optimization theory, numerical analysis, and statistics.

Naturally, optimal estimators have also ruled in image recovery and there is no shortage of definitions for optimality. Thus, researchers have proposed criteria such as minimum cross-entropy [23], regularized least-squares residual [53], maximum likelihood [70], [104], [142], least-squares error [5], maximum *a posteriori* [166], [167], and other Bayesian techniques [71], [87], [90], maximum entropy [70], [119], [177], and maximum power [168].

Optimal procedures have undoubtedly provided satisfactory solutions in numerous applications. However, certain reservations can be formulated vis-à-vis such approaches. First, the criterion of optimality is inherently subjective and different criteria may yield different solutions. Thus, some will argue that a maximum likelihood estimate is desirable while others will discount it on account of its many pathologies. Others will argue that the Bayesian framework is better suited to incorporate *a priori* information. However, it requires a probabilistic model for the original image, a highly debatable issue. Moreover, not all *a priori* information can be easily described in probabilistic terms and the resulting prior distribution is usually too complex to yield a tractable minimization of the resulting conditional expectation. In fact, such pathologies exist for almost every type of estimation procedure and have given rise to many controversies [52], [63], [64], [80], [86], [186]. A second concern with optimal formulations is computational tractability, which requires that (1.5) be simplified by choosing a workable cost function Θ and getting rid of some, if not all, of the constraints $(\Psi_i)_{i\in I}$. For that reason, one tends to select Θ on grounds which are seldom related to rational and practical goals reflecting the specificities of the problem

at hand. For instance, the least-squares error criterion, which usually yields tractable problems, has been used in countless recovery algorithms although its inadequacy in imaging sciences has long been recognized [7]. In addition, the necessity of ignoring constraints leads to solutions which violate known facts about the original image.

In short, optimal procedures often amount in practice to finding an image which is optimal with respect to a standard cost function and likely to be outside of the feasibility set S.

1.3 Feasible Solutions and Set Theoretic Estimates

The set theoretic approach in estimation is governed by the notion of feasibility [38]. In other words, one recognizes the importance of the constraints in (1.5) and, at the same time, the inherent arbitrariness that surrounds the choice of a relevant cost function Θ. As a result, the recovery problem is posed as a feasibility problem, namely

$$\text{Find} \quad a^\star \in S = \bigcap_{i \in I} S_i. \tag{1.6}$$

The restoration criterion thus defined is clear: any image which is consistent with all the information available about the problem and the data is acceptable. The solution to the problem is therefore the set S of feasible images.

The main asset of the set theoretic approach is to allow the incorporation of a broad range of statistical as well as nonstatistical information in the definition of a solution. In the engineering literature, this approach seems to have been first applied to systems theory as a nonstatistical way to incorporate uncertainty in modeling, analysis, estimation, and control problems [38]. In this context, the basic idea of an estimation scheme which yields a set based on available information, rather than a single point, can be traced back to [150]. To this day, image recovery remains the most active field of application of set theoretic estimation. This popularity can be explained by two main factors. First, image recovery problems are typically accompanied by a great deal of qualitative information about the original image that is not easily expressed in purely statistical terms, which is the only form that conventional estimation methods can exploit. The second factor is that in most cases, a human observer will judge the quality of the recovered image. Since the human eye is not sensitive to standard mathematical goodness measures, the importance of an optimal recovery, in one sense or another, is significantly diminished.

Set theoretic image recovery departs radically from the conventional framework of Section 1.2, in which the primary criterion of acceptability of a solution was its optimality with respect to some cost and where feasibility was of secondary importance. In this regard, a common criticism against the set theoretic approach is that it does not produce a unique solution. First, as we have just seen, although it may be gratifying to have obtained

the "best" image, optimality claims often have little practical value. At best, if an optimal solution does land in the feasibility set, it can be regarded as a qualitative selection of a feasible solution. Moreover, from a philosophical standpoint, demanding that one, and only one, image be acceptable as a solution in problems which are notoriously affected by uncontrollable factors (e.g., noise, uncertain image formation models) may appear somewhat unwise. Finally, it should be noted that methods which yield unique solutions are usually iterative and their solution depends on a stopping rule. Since there is a whole collection of images that satisfy any given stopping rule, a set of solutions is thus implicitly defined, not a single point. All in all, uniqueness of a solution is merely a conservative postulate in the tradition of a certain scientific culture, not a universal, philosophically correct, and rational requirement.

1.4 The Convex Feasibility Problem

So far, we have not put any restrictions on the set theoretic recovery problem (1.6). However, due to the lack of numerical methods for solving feasibility problems in their full generality, we must restrict ourselves to problems yielding closed and convex sets in the Hilbert space Ξ. In this case (1.6) is called a convex feasibility problem and efficient techniques are available to solve it.

Requiring convexity is certainly a limitation since, as will be seen in Section 3.4, important constraints are not convex in the selected solution space. Fortunately, in many problems, convex constraints will suffice to define meaningful feasibility sets. For instance, all linear and affine constraints lead to convex sets as well as linear inequality constraints. In addition, a large corpus of nonlinear constraints are of the convex type.

A convex set theoretic image recovery problem involves three steps.

1. Selecting a hilbertian solution space Ξ.
2. Selecting the constraints that yield closed and convex property sets $(S_i)_{i\in I}$ in Ξ and constructing these sets.
3. Solving the convex feasibility problem (1.6).

The selection of a solution space is discussed in Section 3, where we provide a general overview of set theoretic image recovery. The construction of convex property sets from various properties of the image to be estimated and of the imaging system is then discussed in Section 4. Section 5 is devoted to the question of solving convex feasibility problems. Numerical simulations are presented in Section 6 to illustrate various theoretical and practical aspects of convex image recovery. The survey is concluded by a brief

summary in Section 7. For the convenience of the reader, we have listed some frequently used acronyms in Appendix A. We shall now start with a review of the necessary mathematical background.

2 Mathematical Foundations

We review here the essential elements of analysis that constitute the mathematical foundation of convex set theoretic image recovery. Notations are definitions used throughout the survey are also introduced.

Complements and background on general mathematical analysis will be found in [57]. More specialized references are: on weak convergence [16] and [189]; on convex analysis [8], [65], [111], and [190]; on projections [8], [16], [17], and [187]; on nonlinear operators [75], [111], and [188].

2.1 General Notations

$\mathbb{C}$ is the set of complex numbers, $\mathbb{R}$ the set of reals, $\mathbb{R}_+$ the set of nonnegative reals, $\mathbb{R}_+^*$ the set of positive reals, $\mathbb{Z}$ the set of integers, $\mathbb{N}$ the set of nonnegative integers, and $\mathbb{N}^*$ the set of positive integers. The complex conjugate of $z \in \mathbb{C}$ is denoted by $\overline{z}$.

The family of all subsets of a set S is denoted by $\mathfrak{P}(S)$. Moreover, the cardinality of S is denoted by card$S$, its complement by $\complement S$, and its indicator function by 1_S, i.e.,

$$1_S(a) = \begin{cases} 1 & \text{if } a \in S \\ 0 & \text{if } a \in \complement S. \end{cases} \tag{2.1}$$

Ξ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$. Its norm is given by $(\forall a \in \Xi) \;\; \|a\| = \sqrt{\langle a \mid a \rangle}$ and its distance by $(\forall (a,b) \in \Xi^2) \;\; d(a,b) = \|a - b\|$. The dimension of Ξ is denoted by $\dim \Xi$, the zero vector in Ξ by 0, and the identity operator on Ξ by Id. The boundary of a set S is denoted by ∂S. If $S \subset \Xi$ is an affine subspace, the vector space $S^\perp$ is its orthogonal complement. Finally, ${}^tM$ denotes the transpose of a matrix M.

2.2 Geometrical Properties of Sets

A vector subspace is any nonempty subset S of Ξ such that

$$(\forall \alpha \in \mathbb{R})(\forall (a,b) \in S^2) \;\; \alpha a + b \in S, \tag{2.2}$$

and an affine subspace is any set $S = \{a + b \mid a \in V\}$, where V is a vector subspace and $b \in \Xi$. Now let b be a nonzero vector in Ξ and (η, κ) a pair of real numbers. The set

$$H = \{a \in \Xi \mid \langle a \mid b \rangle = \kappa\} \tag{2.3}$$

is a (closed) affine hyperplane, the set

$$Q = \{a \in \Xi \mid \langle a \mid b \rangle \leq \kappa\} \tag{2.4}$$

a closed affine half-space, and the set

$$R = \{a \in \Xi \mid \eta \leq \langle a \mid b \rangle \leq \kappa\} \tag{2.5}$$

a closed affine hyperslab. The closed ball of center $r \in \Xi$ and radius $\gamma \in \mathbb{R}_+^*$ is defined as

$$B(r, \gamma) = \{a \in \Xi \mid \|a - r\| \leq \gamma\}. \tag{2.6}$$

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing function that vanishes only at 0. Then S is f-uniformly convex if

$$(\forall (a, b) \in S^2) \;\; B((a+b)/2, f(\|a-b\|)) \subset S, \tag{2.7}$$

which implies that it is bounded, unless $S = \Xi$. All of the above sets are convex, that is

$$(\forall \alpha \in [0,1])(\forall (a,b) \in S^2) \;\; \alpha a + (1-\alpha) b \in S. \tag{2.8}$$

The convex hull of a set S is the smallest convex set containing S. S is called a cone (of vertex 0) if

$$(\forall \alpha \in \mathbb{R}_+^*)(\forall a \in S) \;\; \alpha a \in S. \tag{2.9}$$

A cone S is convex if and only if

$$(\forall (a,b) \in S^2) \;\; a + b \in S. \tag{2.10}$$

One will often have to show that a set is convex. The following proposition gives sufficient conditions for convexity.

Proposition 2.1 [16] *A subset S of Ξ is convex if any of the following conditions holds.*

(i) *S is an arbitrary intersection of convex sets.*

(ii) *$S = \{a + b \mid (a, b) \in C_1 \times C_2\}$, where C_1 and C_2 are convex.*

(iii) *There exists a convex subset C of a vector space Ξ' and*

- *either a linear operator $T : \Xi \to \Xi'$ such that $S = T^{-1}(C) \triangleq \{a \in \Xi \mid T(a) \in C\}$,*
- *or a linear operator $T : \Xi' \to \Xi$ such that $S = T(C) \triangleq \{a \in \Xi \mid (\exists\, a' \in C) \;\; a = T(a')\}$.*

A special case of interest is $\Xi' = \mathbb{R}$, where it is known that the intervals are the only convex sets.

(iv) *There exists a convex functional $g : \Xi \to \mathbb{R}$ (i.e., (2.12) holds) and a real number η such that either $S = g^{-1}(]-\infty, \eta])$ or $S = g^{-1}(]-\infty, \eta[)$.*

2.3 Strong and Weak Topologies

A sequence $(a_n)_{n\geq 0} \subset \Xi$ converges to $a \in \Xi$ 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. We shall use the notations $a_n \overset{n}{\to} a$ and $a_n \overset{n}{\rightharpoonup} a$ to designate respectively the strong and weak convergence of $(a_n)_{n\geq 0}$ to a.

Let S be a subset of Ξ. Then S is (strongly) closed if for every sequence $(a_n)_{n\geq 0} \subset S$, we have $a_n \overset{n}{\to} a \Rightarrow a \in S$. The closure of S is the smallest closed set $\overline{S}$ containing S. S is open if $\complement S$ is closed. The interior of S is the largest open set $\overset{\circ}{S}$ contained in S. The following Proposition gives sufficient conditions for closedness.

Proposition 2.2 [57] *A subset S of Ξ is closed if any of the following conditions holds.*

(i) *S is a finite union or an arbitrary intersection of closed sets.*

(ii) *There exists a continuous functional $g : \Xi \to \mathbb{R}$ and a closed set $C \subset \mathbb{R}$ such that $S = g^{-1}(C)$.*

(iii) *There exists a lower semi-continuous functional $g : \Xi \to \mathbb{R}$ and a real number η such that $S = g^{-1}(]-\infty, \eta])$.*

A point $a \in \Xi$ is a strong 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 strongly to a. $S \subset \Xi$ is compact if every sequence with elements in S admits at least one strong cluster point in S. Every compact set is closed and bounded. S is boundedly compact if its intersection with any closed ball is compact.

Proposition 2.3 [57] *A subset S of Ξ is compact if any of the following conditions holds.*

(i) *S is closed and bounded and* $\dim \Xi < +\infty$.

(ii) *S is a finite union or an arbitrary intersection of compact sets.*

(iii) *S is a closed subset of a compact set.*

(iv) *There exists a compact subset K of a (topological) space Ξ' and a continuous operator $T : \Xi' \to \Xi$ such that $S = T(K)$.*

(v) *$S = \{a + b \mid (a, b) \in C_1 \times C_2\}$, where C_1 and C_2 are compact.*

S is weakly closed if for every sequence $(a_n)_{n\geq 0} \subset S$ we have $a_n \overset{n}{\rightharpoonup} a \Rightarrow a \in S$. Every weakly closed set is closed and every closed and convex set is weakly closed. A point $a \in \Xi$ is called a weak 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 weakly to a.

Proposition 2.4 *Take $(a_n)_{n\geq 0} \subset \Xi$ and $a \in \Xi$. Then the following statements hold.*

(i) *If $a_n \overset{n}{\rightharpoonup} a$, then $(a_n)_{n\geq 0}$ is bounded and $\|a\| \leq \liminf_{n\to+\infty} \|a_n\|$.*

(ii) *If $(a_n)_{n\geq 0}$ is bounded, then it possesses a weak cluster point a.*

(iii) *If $(a_n)_{n\geq 0}$ is bounded and possesses a unique weak cluster point a, then $a_n \overset{n}{\rightharpoonup} a$.*

(iv) *If $a_n \overset{n}{\rightharpoonup} a$ and if $(b_n)_{n\geq 0} \subset \Xi$ satisfies $b_n \overset{n}{\rightharpoonup} b$, then $(\forall \alpha \in \mathbb{R})\ \alpha a_n + b_n \overset{n}{\rightharpoonup} \alpha a + b$.*

(v) $a_n \overset{n}{\to} a \Rightarrow a_n \overset{n}{\rightharpoonup} a$.

(vi) *If $\dim \Xi < +\infty$, then $a_n \overset{n}{\rightharpoonup} a \Rightarrow a_n \overset{n}{\to} a$.*

(vii) *If $\|a_n\| \overset{n}{\to} \|a\|$, then $a_n \overset{n}{\rightharpoonup} a \Rightarrow a_n \overset{n}{\to} a$.*

(viii) *If $(a_n)_{n\geq 0} \subset S$, where S is boundedly compact, then $a_n \overset{n}{\rightharpoonup} a \Rightarrow a_n \overset{n}{\to} a$.*

(ix) *If $d(a_n, S) \overset{n}{\to} 0$, where S is closed and uniformly convex, and if $a_n \overset{n}{\rightharpoonup} a \in \partial S$ then $a_n \overset{n}{\to} a$.*

Proof. (i)-(vii): see [189]. (viii): According to (i), $(a_n)_{n\geq 0}$ lies in some closed ball B and therefore in the compact set $S \bigcap B$. Therefore, it possesses at least one strong cluster point b, say $a_{n_k} \overset{k}{\to} b$. Then, by (v), $a_{n_k} \overset{k}{\rightharpoonup} b$ and, since $a_{n_k} \overset{k}{\rightharpoonup} a$, we obtain $a = b$. Since $(a_n)_{n\geq 0}$ lies in a compact set and possesses a unique strong cluster point a, we conclude $a_n \overset{n}{\to} a$ [57]. (ix): see [110]. □

2.4 Convex Functionals

A functional on Ξ is an operator $g : \Xi \to \mathbb{R}$.[2] Its sections are the sets

$$(\forall \eta \in \mathbb{R})\ \ S_\eta = g^{-1}(]-\infty, \eta]) = \{a \in \Xi \mid g(a) \leq \eta\}. \tag{2.11}$$

The functional g is convex if

$$(\forall \alpha \in [0,1])(\forall (a,b) \in \Xi^2)\ \ g(\alpha a + (1-\alpha)b) \leq \alpha g(a) + (1-\alpha) g(b). \tag{2.12}$$

If g is convex, then its sections $(S_\eta)_{\eta\in\mathbb{R}}$ are convex sets. If the sections $(S_\eta)_{\eta\in\mathbb{R}}$ are closed, then g is lower semi-continuous (l.s.c.).

[2] As a reminder, this notation means that the domain of g is Ξ.

Proposition 2.5 [8], [65] *Let $g : \Xi \to \mathbb{R}$ be a convex functional. Then g is continuous if either of the following properties holds.*

(i) $\dim \Xi < +\infty$.

(ii) *g is l.s.c.*

In addition, in case (ii)*, g is also weak l.s.c. in the sense that*

$$a_n \overset{n}{\rightharpoonup} a \;\Rightarrow\; g(a) \leq \liminf_{n\to+\infty} g(a_n). \tag{2.13}$$

As a corollary of (i) above, we obtain a useful sufficient condition for closedness and convexity of a set in euclidean (finite dimensional real Hilbert) spaces.

Proposition 2.6 *Let $g : \Xi \to \mathbb{R}$ be a convex functional and suppose that $\dim \Xi < +\infty$. Then, for every $\eta \in \mathbb{R}$, the set $\{a \in \Xi \mid g(a) \leq \eta\}$ is closed and convex.*

We shall say that g is lower semi-boundedly-compact (l.s.b.co.) if for any closed ball B the sets $(\overline{S_\eta \bigcap B})_{\eta\in\mathbb{R}}$ are compact. Now assume that g is convex. The subdifferential of g at a is the set of its subgradients, that is

$$\partial g(a) = \{t \in \Xi \mid (\forall b \in \Xi)\;\; \langle b - a \mid t\rangle \leq g(b) - g(a)\}. \tag{2.14}$$

If g is continuous at a, then it is subdifferentiable at a, i.e., $\partial g(a) \neq \emptyset$. If g is Gâteaux differentiable at a, then there is a unique subgradient, $\nabla g(a)$, called gradient: $\partial g(a) = \{\nabla g(a)\}$.

2.5 Projections

S is a nonempty subset of Ξ.

2.5.1 Distance to a Set

The distance to S is the function $d(\cdot, S)$ defined as

$$(\forall a \in \Xi)\;\; d(a, S) = \inf\{d(a, b) \mid b \in S\}. \tag{2.15}$$

Theorem 2.1 [8], [187] *Suppose that S is closed and convex. Then the functional $d(\cdot, S) : \Xi \to \mathbb{R}_+$ is continuous, convex, and Fréchet differentiable. We have*

$$(\forall a \in \Xi) \;\; \nabla d(a, S)^2 = 2(a - P_S(a)), \tag{2.16}$$

and

$$(\forall a \in \complement S) \;\; \nabla d(a, S) = \frac{a - P_S(a)}{\|a - P_S(a)\|}. \tag{2.17}$$

2.5.2 Projection Operators

The projection operator onto S is the set-valued map

$$\begin{array}{l} \Pi_S : \Xi \to \mathfrak{P}(S) \\ \qquad a \mapsto \{b \in S \mid d(a, b) = d(a, S)\}. \end{array} \tag{2.18}$$

In general $0 \leq \operatorname{card} \Pi_S(a) \leq +\infty$. S is proximinal if $(\forall a \in \Xi) \;\; \Pi_S(a) \neq \emptyset$, i.e., every point admits at least one projection onto S, and it is a Chebyshev set if $(\forall a \in \Xi) \;\; \operatorname{card} \Pi_S(a) = 1$, i.e., every point admits one and only one projection onto S. In the standard euclidean space, such properties were systematically investigated by Bouligand [15], who called points with more than one projection onto a nonempty closed set the multifurcation points of that set. Erdös later showed that the set of multifurcation points of a nonempty closed set of the euclidean space has Lebesgue measure zero [66].

The set S is approximately compact if, for every a in Ξ, every sequence $(b_n)_{n \geq 0} \subset S$ such that $d(a, b_n) \overset{n}{\to} d(a, S)$ possesses a strong cluster point in S.

Proposition 2.7 [17], [49] *Each property in the following list implies the next.*

(i) *S is compact.*

(ii) *S is boundedly compact.*

(iii) *S is approximately compact.*

(iv) *S is proximinal.*

(v) *S is closed.*

In addition, if $\dim \Xi < +\infty$, *properties* (ii) *through* (v) *are equivalent.*

Theorem 2.2 [8], [16] *Suppose that S is closed and convex. Then it is a Chebyshev set: for every $a \in \Xi$ there exists a unique point $P_S(a) \in S$, called projection of a onto S, such that $d(a, P_S(a)) = d(a, S)$. The projection operator P_S is characterized by the variational inequality*

$$(\forall a \in \Xi)(\forall b \in S) \;\; \langle a - P_S(a) \mid b - P_S(a) \rangle \leq 0, \tag{2.19}$$

which becomes

$$(\forall a \in \Xi)(\forall b \in S) \;\; \begin{cases} \langle a - P_S(a) \mid P_S(a) \rangle = 0 \\ \langle a - P_S(a) \mid b \rangle \leq 0, \end{cases} \tag{2.20}$$

if S is a cone, and

$$(\forall a \in \Xi)(\forall b \in S) \;\; \begin{cases} \langle a - P_S(a) \mid b - P_S(a) \rangle = 0 \\ \text{or} \\ \langle a - P_S(a) \mid b \rangle = 0 \end{cases} \tag{2.21}$$

according as S is an affine or a vector subspace.

In euclidean spaces, the class of Chebyshev sets coincides with the class of nonempty closed and convex sets [94]. However, in infinite-dimensional Hilbert spaces, whether every Chebyshev set must be convex is still an open question. A partial answer is that in incomplete pre-Hilbert spaces, Chebyshev sets may not be convex [95], [96].

The projection operators onto the closed and convex sets (2.3)-(2.6) are respectively given by

$$(\forall a \in \Xi) \;\; P_H(a) = a + \frac{\kappa - \langle a \mid b \rangle}{\|b\|^2} b, \tag{2.22}$$

$$(\forall a \in \Xi) \;\; P_Q(a) = \begin{cases} a + \dfrac{\kappa - \langle a \mid b \rangle}{\|b\|^2} b & \text{if } \langle a \mid b \rangle > \kappa \\ a & \text{if } \langle a \mid b \rangle \leq \kappa, \end{cases} \tag{2.23}$$

$$(\forall a \in \Xi) \;\; P_R(a) = \begin{cases} a + \dfrac{\kappa - \langle a \mid b \rangle}{\|b\|^2} b & \text{if } \langle a \mid b \rangle > \kappa \\ a & \text{if } \eta \leq \langle a \mid b \rangle \leq \kappa \\ a + \dfrac{\eta - \langle a \mid b \rangle}{\|b\|^2} b & \text{if } \langle a \mid b \rangle < \eta, \end{cases} \tag{2.24}$$

$$(\forall a \in \Xi) \;\; P_B(a) = \begin{cases} r + \gamma \dfrac{a - r}{\|a - r\|} & \text{if } \|a - r\| > \gamma \\ a & \text{if } \|a - r\| \leq \gamma. \end{cases} \tag{2.25}$$

2.5.3 Relaxed Convex Projections

Let $\lambda \in [0, 2]$ and suppose that S is closed and convex. The relaxed operator of projection onto S is defined as

$$(\forall a \in \Xi) \;\; T_S^\lambda(a) = a + \lambda(P_S(a) - a). \tag{2.26}$$

For $0 \leq \lambda \leq 1$, $T_S^\lambda(a)$ is an underrelaxed projection, or underprojection; for $\lambda = 1$, $T_S^\lambda(a)$ is an unrelaxed projection, or projection; for $1 \leq \lambda \leq 2$, $T_S^\lambda(a)$ is an overrelaxed projection or overprojection; for $\lambda = 2$, $T_S^\lambda(a)$ is the reflection of a with respect to S and is denoted by $R_S(a)$ (see Fig. 1).

2.6 Nonlinear Operators

Let $T : \Xi \to \Xi$ be an operator. The set of fixed points of T is

$$\operatorname{Fix} T = \{a \in \Xi \mid T(a) = a\}. \tag{2.27}$$

T is contractive if

$$(\exists k \in]0, 1[)(\forall (a, b) \in \Xi^2) \;\; \|T(a) - T(b)\| \leq k\|a - b\|, \tag{2.28}$$

nonexpansive if

$$(\forall (a, b) \in \Xi^2) \;\; \|T(a) - T(b)\| \leq \|a - b\|, \tag{2.29}$$

and firmly nonexpansive if

$$(\forall (a, b) \in \Xi^2) \;\; ||T(a) - T(b)||^2 \leq \langle a - b \mid T(a) - T(b)\rangle, \tag{2.30}$$

or, equivalently, if

$$(\forall (a, b) \in \Xi^2) \;\; \|T(a) - T(b)\|^2 \leq \|a - b\|^2 - \|(\mathrm{Id} - T)(a) - (\mathrm{Id} - T)(b)\|^2. \tag{2.31}$$

T is demiclosed if for any sequence $(a_n)_{n\geq 0}$ such that $a_n \overset{n}{\rightharpoonup} a$ and $T(a_n) \overset{n}{\to} b$, we have $T(a) = b$. T is demicompact if any bounded sequence $(a_n)_{n\geq 0}$ admits a strong cluster point whenever the sequence $(T(a_n) - a_n)_{n\geq 0}$ converges strongly.

Theorem 2.3 [188] *If T is contractive, it admits one and only one fixed point. Now let C be a nonempty, closed, bounded, and convex subset of Ξ and suppose that $T : C \to C$ is nonexpansive. Then $\operatorname{Fix} T$ is nonempty, closed, and convex.*

Proposition 2.8 *Consider the properties:*

(a) T *is the operator of projection onto a nonempty closed and convex subset* F *of* Ξ.

(b) T *is firmly nonexpansive.*

(c) T *is nonexpansive.*

(d) $\mathrm{Id} - T$ *is demiclosed.*

Then:

(i) (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d).

(ii) *Suppose that* $F = \mathrm{Fix}\, T \neq \emptyset$. *Then* (b) *implies*

$$(\forall a \in \Xi)\ \ T(a) \in B\left(\frac{a + P_F(a)}{2}, \frac{\|a - P_F(a)\|}{2}\right). \tag{2.32}$$

In addition, (b) $\Rightarrow$ (a) *if and only if* $(\forall a \in \Xi)\ \ T(a) \in F$.

(iii) *Suppose that* $F = \mathrm{Fix}\, T \neq \emptyset$. *Then* (c) *implies*

$$(\forall a \in \Xi)\ \ T(a) \in B(P_F(a), \|a - P_F(a)\|). \tag{2.33}$$

(iv) (b) *holds if and only if* $T = (T' + \mathrm{Id})/2$, *where* $T' : \Xi \to \Xi$ *is nonexpansive.*

Proof. (i): (a) $\Rightarrow$ (b) follows from (2.19) (e.g., [187]), (b) $\Rightarrow$ (c) follows directly from (2.31), and (c) $\Rightarrow$ (d) is proved in [19]. (ii): Take any $a \in \Xi$ and let $b = P_F(a)$. Then $T(b) = b$ and (2.30) gives

$$\langle a - P_F(a) \mid T(a) - P_F(a) \rangle \geq \|T(a) - P_F(a)\|^2, \tag{2.34}$$

so that we obtain $\langle a - T(a) \mid T(a) - P_F(a) \rangle \geq 0$. Hence, $\langle T(a) - a \mid a - P_F(a) \rangle \leq -\|a - T(a)\|^2$. Therefore

$$\begin{aligned} \left\| T(a) - \frac{a + P_F(a)}{2} \right\|^2 &= \|T(a) - a\|^2 + \langle T(a) - a \mid a - P_F(a) \rangle \\ &\quad + \left\| \frac{a - P_F(a)}{2} \right\|^2 \qquad (2.35) \\ &\leq \left\| \frac{a - P_F(a)}{2} \right\|^2, \qquad (2.36) \end{aligned}$$

which proves (2.32). To prove the second assertion, note that necessity is obvious. As to sufficiency, take any $a \in \Xi$, suppose that $T(a) \in F$, and put $b = T(a)$ in (2.19). Then we

get $\langle a - P_F(a) \mid T(a) - P_F(a)\rangle \leq 0$ which, in view of (2.34), implies $T(a) = P_F(a)$. (iii): Take any $a \in \Xi$. Then $\|T(a) - P_F(a)\| = \|T(a) - T(P_F(a))\| \leq \|a - P_F(a)\|$. (iv): see [141] or [187]. □

Proposition 2.9 *Let S be a nonempty, closed, and convex subset of Ξ. Then for any $\lambda \in [0, 2]$ the relaxed projection operator $T_S^\lambda = \mathrm{Id} + \lambda(P_S - \mathrm{Id})$ is nonexpansive.*

Proof. Let $\alpha = \lambda/2 \in [0, 1]$. Then $T_S^\lambda = (1 - \alpha)\mathrm{Id} + \alpha(2P_S - \mathrm{Id})$. According to Proposition 2.8(i)+(iv), $R_S = 2P_S - \mathrm{Id}$ is nonexpansive. Therefore T_S^λ is nonexpansive, as a convex combination of the two nonexpansive operators Id and R_S. □

Proposition 2.10 *Let P_S be the operator of projection onto a nonempty, boundedly compact, and convex subset S of Ξ. Then P_S is demicompact.*

Proof. Let $(a_n)_{n\geq 0}$ be a bounded sequence. Then, thanks to Proposition 2.4(ii), it admits a weak cluster point a, say $a_{n_k} \overset{k}{\rightharpoonup} a$. Now suppose $P_S(a_n) - a_n \overset{n}{\to} a' \in \Xi$. Then $P_S(a_{n_k}) - a_{n_k} \overset{k}{\rightharpoonup} a'$ and, thanks to Proposition 2.4(iv), $P_S(a_{n_k}) \overset{k}{\rightharpoonup} a + a'$. But $(P_S(a_{n_k}))_{k\geq 0} \subset S$. Therefore Proposition 2.4(viii) implies that $P_S(a_{n_k}) \overset{k}{\to} a + a'$ and, since $P_S(a_{n_k}) - a_{n_k} \overset{k}{\to} a'$, it follows that $a_{n_k} \overset{k}{\to} a$. In words, $(a_n)_{n\geq 0}$ admits a strong cluster point. □

2.7 Fejér-Monotone Sequences

Let S be a nonempty, closed, and convex subset of Ξ. A sequence $(a_n)_{n\geq 0}$ is Fejér-monotone with respect to S if

$$(\forall n \in \mathbb{N})(\forall a \in S) \quad \|a_{n+1} - a\| \leq \|a_n - a\|. \tag{2.37}$$

Proposition 2.11 [11], [19] *Suppose that $(a_n)_{n\geq 0}$ is Fejér-monotone with respect to S. Then the following properties hold.*

(i) *$(a_n)_{n\geq 0}$ is bounded and admits at least one weak cluster point.*

(ii) *If all the weak cluster points of $(a_n)_{n\geq 0}$ lie in S, then $(\exists\, a \in S)$ $a_n \overset{n}{\rightharpoonup} a$.*

(iii) *If $(a_n)_{n\geq 0}$ admits a strong cluster point a in S, then $a_n \overset{n}{\to} a$.*

(iv) *If $\overset{\circ}{S} \neq \emptyset$, then $(a_n)_{n\geq 0}$ converges strongly.*

2.8 Convex Feasibility in a Product Space

Consider the convex feasibility problem (1.6), and assume that the number of sets is finite, say card $I = m$. Take a real m-tuple $(w_i)_{i\in I}$ such that

$$\sum_{i\in I} w_i = 1 \qquad \text{and} \qquad (\forall i \in I)\ \ w_i > 0, \tag{2.38}$$

and let

$$\mathbf{\Xi} = \underbrace{\Xi \times \cdots \times \Xi}_{m \text{ times}} \tag{2.39}$$

be the m-fold cartesian product of the Hilbert space Ξ. We shall denote by $\mathbf{a} = (a_1, \cdots, a_m) = (a_i)_{i\in I}$ an m-tuple in $\mathbf{\Xi}$. $\mathbf{\Xi}$ can be made into a Hilbert space by endowing it with the scalar product

$$(\forall (\mathbf{a}, \mathbf{b}) \in \mathbf{\Xi}^2)\ \ \langle\langle \mathbf{a} \mid \mathbf{b} \rangle\rangle = \sum_{i\in I} w_i \langle a_i \mid b_i \rangle. \tag{2.40}$$

The associated norm and distance are given by

$$(\forall (\mathbf{a}, \mathbf{b}) \in \mathbf{\Xi}^2)\ \ \begin{cases} |||\mathbf{a}||| = (\sum_{i\in I} w_i \|a_i\|^2)^{1/2} \\ \mathbf{d}(\mathbf{a}, \mathbf{b}) = (\sum_{i\in I} w_i d(a_i, b_i)^2)^{1/2}. \end{cases} \tag{2.41}$$

Let $\mathbf{S}$ be the cartesian product of the sets $(S_i)_{i\in I}$, i.e., the closed and convex set

$$\mathbf{S} = \bigtimes_{i\in I} S_i = \{\mathbf{a} \in \mathbf{\Xi} \mid (\forall i \in I)\ \ a_i \in S_i\}, \tag{2.42}$$

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

$$\mathbf{D} = \{(a, \cdots, a) \in \mathbf{\Xi} \mid a \in \Xi\}. \tag{2.43}$$

Hence, to every point $a \in \Xi$ there corresponds a unique point $\mathbf{a} = (a, \cdots, a) \in \mathbf{D}$ and vice-versa. With these notations, observe that

$$(\forall (\mathbf{a}, \mathbf{b}) \in \mathbf{D}^2)\ \ \langle\langle \mathbf{a} \mid \mathbf{b} \rangle\rangle = \langle a \mid b \rangle \quad \text{and} \quad |||\mathbf{a}||| = \|a\|. \tag{2.44}$$

Whence, we obtain immediately the following result.

Proposition 2.12 *Take $(\mathbf{a}_n)_{n\geq 0} \subset \mathbf{D}$ and $\mathbf{a} \in \mathbf{D}$, in correspondence with $(a_n)_{n\geq 0} \subset \Xi$ and $a \in \Xi$. Then*

(i) $\mathbf{a}_n \overset{n}{\to} \mathbf{a} \Leftrightarrow a_n \overset{n}{\to} a$.

(ii) $\mathbf{a}_n \overset{n}{\rightharpoonup} \mathbf{a} \Leftrightarrow a_n \overset{n}{\rightharpoonup} a.$

It is also clear that (2.42) and (2.43) imply

$$\mathbf{S} \bigcap \mathbf{D} = \{(a, \cdots, a) \in \mathbf{\Xi} \mid (\forall i \in I) \;\; a \in S_i\} \tag{2.45}$$

$$= \{(a, \cdots, a) \in \mathbf{\Xi} \mid a \in \bigcap_{i \in I} S_i\}. \tag{2.46}$$

Therefore, in the product space $\mathbf{\Xi}$, we can reformulate the feasibility problem (1.6) as

$$\text{Find} \;\; \mathbf{a}^\star \in \mathbf{S} \bigcap \mathbf{D}. \tag{2.47}$$

This product space characterization of (1.6) was developed by Pierra in [132]. It reduces the m-set problem (1.6) to the simpler problem (2.47), which involves only a vector subspace and a convex set.

3 Overview of Convex Set Theoretic Image Recovery

In this section, we provide a general overview of convex set theoretic image recovery. We discuss the mathematical formalization and the history of the field, as well as specific applications. Finally, we discuss nonconvex problems.

3.1 Theoretical Framework

3.1.1 Basic Assumptions

Throughout this survey, the image space is a real Hilbert space Ξ with scalar product $\langle \cdot \mid \cdot \rangle$, norm $\| \cdot \|$, and distance d. The original image h is described by a family of constraints $(\Psi_i)_{i \in I}$, where $\emptyset \neq I \subset \mathbb{N}$. A family $(S_i)_{i \in I}$ of property sets is constructed in Ξ via (1.3). Their intersection S is nonempty, unless otherwise stated.

3.1.2 The Image Space

3.1.2.1 General Model

Let $(\Upsilon, \mathcal{A}, \mu)$ be a measure space. For most of our purposes, it will be sufficient to take Ξ as the Hilbert space $\mathcal{L}^2(\Upsilon, \mathcal{A}, \mu)$ of (classes of equivalence of) square μ-integrable real-valued functions of two variables on the domain Υ [59], [149]. In Ξ, the scalar product is defined as

$$(\forall (a, b) \in \Xi^2) \;\; \langle a \mid b \rangle = \int_{\Upsilon} a(\varsigma) b(\varsigma) \mu(d\varsigma). \tag{3.1}$$

As we shall see, this representation has the advantage of encompassing analog, discrete, and digital image models.

3.1.2.2 Analog Model

Here, $\Upsilon = \mathbb{R}^2$, $\mathcal{A}$ is the associated Borel σ-algebra, and μ is the two-dimensional Lebesgue measure. Ξ then becomes the usual space L^2 with scalar product

$$(\forall (a, b) \in L^2 \times L^2) \;\; \langle a \mid b \rangle = \iint_{\mathbb{R}^2} a(x, y) b(x, y) dx dy. \tag{3.2}$$

In L^2, the Fourier transform operator $\mathfrak{F} : a \mapsto \widehat{a}$ is defined by

$$\begin{array}{rccl} \widehat{a}: & \mathbb{R}^2 & \to & \mathbb{C} \\ & (\nu_1, \nu_2) & \mapsto & \displaystyle\iint_{\mathbb{R}^2} a(x,y) \exp(-\imath 2\pi(x\nu_1 + y\nu_2))dxdy. \end{array} \tag{3.3}$$

3.1.2.3 Discrete Model

Here, $\Upsilon = \mathbb{Z}^2$, $\mathcal{A} = \mathfrak{P}(\Upsilon)$, and μ is the counting measure ($\mu : A \mapsto \text{card}\, A$). Ξ then becomes the usual space ℓ^2 with scalar product

$$(\forall (a,b) \in \ell^2 \times \ell^2) \;\; \langle a \mid b \rangle = \sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} a(m,n)b(m,n). \tag{3.4}$$

In ℓ^2, the Fourier transform operator $\mathfrak{F} : a \mapsto \widehat{a}$ is defined by

$$\begin{array}{rccl} \widehat{a}: [-1/2, 1/2]^2 & & \to & \mathbb{C} \\ & (\nu_1, \nu_2) & \mapsto & \displaystyle\sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} a(m,n) \exp(-\imath 2\pi(m\nu_1 + n\nu_2)). \end{array} \tag{3.5}$$

3.1.2.4 Digital Model

In digital image processing applications, we are dealing with finite extent, $N \times N$ discretized images [138]. Such an image can be represented by an $N \times N$ matrix $[a^{(m,n)}]_{0\leq m,n\leq N-1}$ whose entries are called pixels. The value of a pixel is called a gray level and represents the brightness of the image at that point. It is usually more convenient to represent an $N \times N$ image by the N^2-dimensional vector a obtained by stacking the rows of the image matrix $[a^{(m,n)}]_{0\leq m,n\leq N-1}$ on top of each other [138]. In other words, the ith component of the vector a is the pixel $a^{(m,n)}$, where $i = mN + n$. Consequently, Ξ can be taken to be the standard N^2-dimensional euclidean space $\mathbb{E}^{N^2}$, which is obtained by taking $\Upsilon = \{0, \cdots, N^2 - 1\}$, $\mathcal{A} = \mathfrak{P}(\Upsilon)$, and μ as the counting measure in $\mathcal{L}^2(\Upsilon, \mathcal{A}, \mu)$.

The Fourier transform $\mathfrak{F}(a) = \widehat{a}$ of a stacked image $a \in \mathbb{E}^{N^2}$ is its two-dimensional discrete Fourier transform (DFT), i.e.,

$$\begin{array}{rccl} \widehat{a}: \{0, \cdots, N-1\}^2 & & \to & \mathbb{C} \\ & (k,l) & \mapsto & \displaystyle\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a^{(mN+n)} \exp(-\imath \frac{2\pi}{N}(mk + nl)). \end{array} \tag{3.6}$$

3.1.3 Set Theoretic Formulation

A set theoretic image recovery problem is entirely specified by its set theoretic formulation, i.e., the pair $(\Xi, (S_i)_{i\in I})$. The solution, or feasibility, set is $S = \bigcap_{i\in I} S_i$. All the images in S are equally acceptable solutions to the problem. The set theoretic formulation is said to be finite if $\operatorname{card} I < +\infty$ and countable if $\operatorname{card} I = +\infty$ (recall that $I \subset \mathbb{N}$). It is said to be ideal if $S = \{h\}$, meaning that the constraints uniquely define h; unfair if $h \notin S$, meaning that h fails to satisfy at least one of the specified constraints; inconsistent if $S = \emptyset$, meaning that at least two of the constraints are incompatible [38] (see Figs. 2-5).

Unfair formulations and, *a fortiori*, inconsistent ones arise when inaccurate or imprecise constraints are present. For instance, most of the sets that will be described in Section 4.2 depend on attributes of the original image that may not be known exactly. The same remark also applies to the attributes of the uncertainty process that will be required to construct the sets of Section 4.3. In addition, such sets based on stochastic information will be seen to be confidence regions whose 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 point spread function of an imaging system [48] or noise perturbations in the data [84], [152] are not taken into account. A method for obtaining meaningful solutions to inconsistent problems will be discussed in Section 5.3.

The degree of unfeasibility of an image $a \in \Xi$ will be quantified via the proximity function

$$\begin{aligned} \Phi : \Xi &\to \mathbb{R}_+ \\ a &\mapsto \frac{1}{2}\sum_{i\in I} w_i d(a, S_i)^2, \end{aligned} \tag{3.7}$$

where the weights $(w_i)_{i\in I}$ are strictly convex, i.e.

$$\sum_{i\in I} w_i = 1 \qquad \text{and} \qquad (\forall i \in I)\ \ w_i > 0. \tag{3.8}$$

In other words, the smaller $\Phi(a)$, the more feasible a. Note that $\Phi(a) = 0 \ \Leftrightarrow\ a \in S$.

3.2 Historical Developments

It is assumed here that the set theoretic formulation is finite and comprises m sets.

3.2.1 Computerized Tomography

The field of computerized tomography can be regarded as the starting point of the set theoretic approach in image recovery in the early 1970s. In computerized tomography, measurements are made of the line integrals of a property of the cross-section of an object (e.g., X-ray attenuation) along various straight lines by varying lateral displacements at a given angle. The problem is then to reconstruct the image of the cross-section from these measurements taken at various angles [85]. This problem is fundamental in diagnostic medicine but also in an increasing number of nonmedical applications [28].

With proper discretization, the original image can be represented by a vector in $\mathbb{E}^{N^2}$ and the reconstruction problem can be written as a system of m linear equations of the type $\langle a \mid b_i \rangle = \delta_i$, for $1 \leq i \leq m$. From a set theoretic standpoint, each of these constraints restricts estimates to a hyperplane

$$S_i = \{a \in \mathbb{E}^{N^2} \mid \langle a \mid b_i \rangle = \delta_i\}, \tag{3.9}$$

and the problem is then to find a point in their intersection S. In [78], a so-called algebraic reconstruction technique (ART) was proposed to this end. It employs the periodic recursion

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = P_{i(n)}(a_n) \quad \text{with} \quad i(n) = n \,(\text{modulo}\, m) + 1 \tag{3.10}$$

to generate a feasible solution. In fact, this mathematical method was developed by Kaczmarz in 1937 [97] to solve systems of linear equations. An alternative projection method was then proposed in [74] under the name simultaneous iterative reconstruction technique (SIRT). In this parallel method, the projections onto all the sets are averaged to form the update, namely

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = \frac{1}{m} \sum_{i \in I} P_i(a_n). \tag{3.11}$$

SIRT is similar to the algorithm devised by Cimmino in 1938 [35] to solve linear systems of equations by successive averaging of reflections onto the sets.

A problem with the set theoretic formulation (3.9), is that noise and other uncertainty sources are ignored. As a result, it may be unfair or even inconsistent. In order to incorporate these disturbances, the hyperplanes were replaced in [84] by the hyperslabs

$$S_i = \{a \in \mathbb{E}^{N^2} \mid \delta_i - \epsilon_i \leq \langle a \mid b_i \rangle \leq \delta_i + \epsilon_i\}, \tag{3.12}$$

where ϵ_i is a tolerance factor. This feasibility problem was solved by the Agmon-Motzkin-Schoenberg algorithm for affine inequalities [1], [122]

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda(P_{i(n)}(a_n) - a_n) \quad \text{with} \quad \begin{cases} i(n) = n \,(\text{modulo}\, m) + 1 \\ 0 < \lambda < 2. \end{cases} \tag{3.13}$$

Simply stated, this algorithm proceeds as follows: starting with an initial estimate a_0, a sequence is generated, where the new iterate a_{n+1} lies on the segment between the current iterate a_n and its reflection $2P_{i(n)}(a_n) - a_n$ with respect to the set $S_{i(n)}$. The position of a_{n+1} on this segment depends on the value of the relaxation parameter λ, which determines the step size $\|a_{n+1} - a_n\|$. This framework is discussed furtherin [88].

3.2.2 The Gerchberg-Papoulis Algorithm

The fundamental problem of estimating an image from partial spatial and spectral information has been the focus of a lot of research in various disciplines ranging from crystallography to astronomy. In the absence of any additional information, it can be formalized as the problem of finding an image in $S_1 \bigcap S_2$, where S_1 is the subset of Ξ of all images consistent with the spatial information and S_2 that of all images consistent with the spectral information.

Although not formulated explicitly in set theoretic terms, the idea of constructing a sequence of points that would alternate between S_1 and S_2 in order to converge to their intersection can be found in [102]. In [72], Gerchberg considered the problem of recovering a finite object from limited diffraction data, i.e., of reconstructing a spatially limited image from partial knowledge of its Fourier transform. The proposed reconstruction method was to alternate resubstitutions of the known data in both domains. It can be regarded as a method of alternating projections between the affine subspaces

$$\begin{cases} S_1 = \{a \in \Xi \mid a = 0 \text{ outside } K_1\} \\ S_2 = \{a \in \Xi \mid \widehat{a} = g \text{ on } K_2\}, \end{cases} \tag{3.14}$$

where $\widehat{a}$ is the Fourier transform of a, g a known function, and K_1 and K_2 are neighborhoods of the origin in the spatial and spectral domains, respectively. Almost at the same time, Papoulis [128] proposed the same method to solve a dual problem, namely, to reconstruct a band-limited signal which is partially known in the time-domain. In this case, the affine subspaces are of the form

$$\begin{cases} S_1 = \{a \in \Xi \mid a = g \text{ on } K_1\} \\ S_2 = \{a \in \Xi \mid \widehat{a} = 0 \text{ outside } K_2\}. \end{cases} \tag{3.15}$$

In the early 1980s attempts were made to formalize the Gerchberg-Papoulis algorithm into larger classes of successive approximation methods amenable to the incorporation of certain types of *a priori* knowledge [145], [148], [165].

3.2.3 Affine Constraints

One of the very first abstract set theoretic approaches to image recovery appeared in [181]. The recovery problem considered in this paper was to find an image h known to belong to a closed subspace S_2 of Ξ given that the observed data consist of the projection of h onto another closed subspace S_1 of Ξ. Under certain conditions, a modified sequence of alternated projections onto S_2 and $S_1^{\perp}$ (the orthogonal complement of S_1) was shown to converge to h. This framework encompasses several basic problems including that considered by Gerchberg and Papoulis (as discussed in Section 3.2.2) and the various extensions considered in [160]. Additional affinely constrained problems can be found in [120].

The following theorem provides the mathematical foundation for such methods. It was proved in [81] for vector subspaces but the proof can be extended routinely to affine subspaces. In the case of two subspaces, it is known as the Alternating Projection Theorem and is due to Von Neumann [176].

Theorem 3.1 *Let $(S_i)_{i\in I}$ be a finite family of m closed affine subspaces of Ξ with nonempty intersection S. Then any sequence $(a_n)_{n\geq 0}$ constructed as in (3.10) converges strongly to a point in S.*

3.2.4 Arbitrary Convex Constraints - The POCS Algorithm

Despite the apparent disparity in their original formulation, all the above methods share the common objective of producing a solution consistent with a collection of affine or affine inequality constraints. As a result, the associated set theoretic formulations comprise only affine subspaces or half-spaces. The scope of this framework is limited since many useful constraints encountered in practice are nonaffine, as will be seen in Section 4. Thus, the main motivation for the extension to convex set theoretic formulations is to allow a much larger class of information to be exploited. This extension was made possible by the availability of convex feasibility algorithms.

As appealing as it may seem, the set theoretic approach would be fairly futile if efficient methods were not available to actually solve (1.6). The field of set theoretic image recovery entered a new era when the image processing community became aware of one such method called POCS, for projections onto convex sets. Although POCS had been used in image reconstruction in [106], it is really [184] which popularized the method and established a broad conceptual and computational basis for convex image recovery.

The method of POCS, which extends the Kaczmarz (3.10) and Agmon-Motzkin-Schoenberg

(3.13) algorithms to arbitrary closed convex sets, is defined by the serial algorithm

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda_n (P_{n \,(\text{modulo } m)+1}(a_n) - a_n), \tag{3.16}$$

where the relaxation parameters $(\lambda_n)_{n \geq 0}$ satisfy

$$(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq 2 - \varepsilon \quad \text{with} \quad 0 < \varepsilon < 1. \tag{3.17}$$

The relaxation parameters provide the flexibility of under- or overprojecting at each iteration. Figs 6-8 depict orbits generated by POCS for relaxations $\lambda_n = 1.0$, $\lambda_n = 0.5$, and $\lambda_n = 1.5$, respectively.

Theorem 3.2 *Let $(S_i)_{i \in I}$ be a finite family of m closed and convex subsets of Ξ with nonempty intersection S. Then every orbit $(a_n)_{n \geq 0}$ of POCS converges weakly to a point in S. In addition, the convergence is strong if any of the following conditions holds.*

(i) $(\exists j \in I) \;\; S_j \bigcap (\bigcap_{i \in I \setminus \{j\}} S_i)^\circ \neq \emptyset$.

(ii) *All but possibly one of the sets in $(S_i)_{i \in I}$ are f-uniformly convex.*

(iii) *$(S_i)_{i \in I}$ is a family of closed affine half-spaces.*

(iv) *One of the sets in $(S_i)_{i \in I}$ is boundedly compact.*

Proof. Weak convergence was proved in [18]. Assertions (i)-(iii) were proved in [79] and assertion (iv) in [163].[3] □

3.3 Applications

In this section, we briefly indicate some of the image recovery problems that have been approached within the convex set theoretic framework. We somewhat arbitrarily classify them into restoration problems, tomographic reconstruction problems, and other recovery problems. Some of these studies involve only one-dimensional signals but they can also be applied to images. It should also be mentioned that the majority of these problems have been solved via the unrelaxed version of POCS. Comparative studies of conventional versus set theoretic image recovery in specific problems can be found in [125] and [164].

[3] The results of [18] and [163] pertain only to the unrelaxed model (3.10) but they still hold true for (3.16)-(3.17).

3.3.1 Restoration Problems

The first application of set theoretic methods in image restoration was demonstrated in [171] where various properties of the noise were shown to produce useful convex sets. The stochastic nature of some blurring functions such as atmospheric turbulence and camera vibration has also been addressed using set theoretic methods [48]. Set theoretic restoration in the presence of bounded kernel disturbances and noise was considered in [51]. Sets based on locally adaptive constraints [101] as well as on smoothness constraints [161] have also been proposed. In addition, set theoretic restoration has been used with other statistically based methods, such as Wiener filtering [155], [157]. Other studies have focused on the restoration of specific types of image, e.g., multi-band satellite images [34], character images [103], echographic images [113], diffraction wave fields [120], optical flow fields and electromagnetic fields [158]. In order to best exploit specific *a priori* information, the set theoretic restoration problem of [146] was posed in a singular value space rather than in the natural image space. Restoration in the presence of an inconsistent set of constraints was considered in [42]. Finally, set theoretic approaches to regularized restoration were proposed in [99] and [144].

3.3.2 Tomographic Reconstruction Problems

In the early work discussed in Section 3.2.1, the chief objective was to generate an image consistent with the (possibly noisy) projection data [85]. Various extensions of ART and SIRT relevant to such set theoretic formulations are surveyed in [28] and [26].

More recent work has been geared towards the incorporation of additional constraints relevant to specific situations. For instance, 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 [140]. 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 was used to extrapolate tomographic images reconstructed from a limited range of views in [106] and [151]. In [152], the formulation of [151] was modified to account for noisy data. In [153], POCS was 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), where the property sets simply translate the requirement that the reconstruction be consistent with the observed projections. In [124], a more sophisticated convex set theoretic formulation was developed by incorporating additional constraints such as known object support and en-

ergy 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 [172]. Set theoretic methods have also been used in fan-beam tomography [130].

In the above studies, the solution space is that of the reconstructed image. In [100], a different set theoretic approach was 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 was first obtained by POCS and then used to reconstruct the image via ordinary convolution backprojection. In [178], POCS was used to synthesize the projection matrix from noisy measurements made by a moving array of detectors and the image was then reconstructed by filtered backprojection.

3.3.3 Other Image Recovery Problems

Applications of convex set theoretic image recovery can be found in such fields as electron microscopy [24], speckle interferometry [62], halftone imaging [83], holography [116], and biomagnetic imaging [123]. Other applications include image recovery from multiple frames of sparse data [139], image recovery from nonuniform samples [147], [180], and recovery of images remotely sensed by image-plane detector arrays [162].

3.4 The Issue of Convexity

The basic objective of the set theoretic approach is to provide a flexible framework for the incorporation of a wide range of information in the recovery process. However, our discussion in this survey is confined to problems in which the constraints yield closed and convex sets in some Hilbert space Ξ. As mentioned in Section 1.4, the reason for this is quite simple: there does not exist any method that is guaranteed to produce a point in the intersection of sets when at least one of them is not convex.

The condition that the sets be closed should not cause concern since a property set S_i can always be replaced by its closure

$$\overline{S}_i = \{a \in \Xi \mid d(a, S_i) = 0\}. \tag{3.18}$$

In doing so, one merely adds points which are at distance zero from the points in S_i, which will have no significant effect on the solution of a practical problem. The issue of convexity is a more serious one as many important constraints do not yield convex sets in the desired solution space, which precludes their use. A classical example of nonconvex

set is the set

$$S_i = \{a \in \Xi \mid |\widehat{a}1_K| = |\widehat{h}1_K|\}, \tag{3.19}$$

based on the knowledge of the Fourier magnitude of the original image over some frequency band K. It arises in various problems in which intensity measurements can be made in the diffraction plane. It is in particular found in the Gerchberg-Saxton method [73] as well as in phase recovery problems [91], [109].

We shall now describe the three main approaches that are presently available to deal with nonconvex problems.

3.4.1 Convexification

Convexification is the process of partially enforcing constraints by replacing nonconvex property sets by their convex hulls. This yields a larger set that can still be useful. An example of useful convexification is the set S_p^+ of (4.41). Another example is found in [184], where the set (3.19) is replaced by

$$S_i = \{a \in \Xi \mid |\widehat{a}1_K| \leq |\widehat{h}1_K|\}. \tag{3.20}$$

In somes cases, the convexification process may give trivial results. For instance [49], consider the set of all digital images in $\mathbb{E}^{N^2}$ whose maximum number of nonzero values is known (e.g., star images in astronomy). It turns out that the convex hull of this set is $\mathbb{E}^{N^2}$ itself, which means that the convexification process has eliminated the constraint.

3.4.2 New Solution Space

If the available information does not yield convex property sets in the selected Hilbert image space Ξ, one may seek a new hilbertian solution space Ξ'. An option is to obtain Ξ' via a (nonlinear) transformation of Ξ. Another option is to redefine the vector space structure of the space. Indeed, recall that the structure of a real vector space V is defined by a so-called addition operation $\oplus : V^2 \to V$ and a so-called scalar multiplication operation $\odot : \mathbb{R} \times V \to V$ which satisfy certain axioms [57]. Hence the definition of the convexity of a set $A \subset V$, i.e.,

$$(\forall \alpha \in]0,1[)(\forall (a,b) \in A^2) \quad (\alpha \odot a) \oplus ((1-\alpha) \odot b) \in A, \tag{3.21}$$

depends on the choice of the operations $\oplus$ and $\odot$. Therefore, by changing the vector space structure, one can render some sets convex. Such a strategy was implemented in [31] in connection with the reconstruction of square-summable discrete signals. The

natural space ℓ^2 with norm $\|a\| = (\sum_{i\in\mathbb{Z}} |a^{(i)}|^2)^{1/2}$ was replaced by the new space $\ell^\star$ of absolutely summable sequences a whose Fourier transform $\mathfrak{F}(a) = \widehat{a}$ satisfies $(\forall \nu \in [-1/2, 1/2])\quad \widehat{a}(\nu) \neq 0$. In $\ell^\star$, the operation $\oplus$ was taken to be convolution and the operation $\odot$ was defined as

$$(\forall (\alpha, a) \in \mathbb{R} \times \ell^\star)\ \ \alpha \odot a = \mathfrak{F}^{-1}(\exp(\alpha \ln(\widehat{a}))). \tag{3.22}$$

In addition, a prehilbertian structure was defined by the scalar product

$$(\forall (a, b) \in \ell^\star \times \ell^\star)\ \ \langle a \mid b \rangle_\star = \int_{-1/2}^{1/2} \ln(\widehat{a}(\nu))\overline{\ln(\widehat{b}(\nu))}\, d\nu. \tag{3.23}$$

It was shown that certain sets, in particular (3.19), that were not convex in ℓ^2 became convex in $\ell^\star$. The space $\ell^\star$ also proved useful for reconstruction from bispectral information [30].

In general, a difficulty that arises in a change of solution space is to render the nonconvex sets convex while preserving the convexity of the other sets in the set theoretic formulation.

3.4.3 Feasibility with Nonconvex Sets

If the two above approaches turn out to be unsatisfactory, a third option is to try and solve the nonconvex feasibility as is. Heuristic attempts have been made to use the periodic projection algorithm (3.10) in the presence of nonconvex sets. This is essentially the approach of [73] and [109]. A formal local convergence result for this method is the following theorem, in which $(\Pi_i)_{i\in I}$ designates the family of set-valued projection operators onto the sets $(S_i)_{i\in I}$, as defined in Section 2.5.2.

Theorem 3.3 [49] *Let $(S_i)_{i\in I}$ be a finite family of m approximately compact subsets of Ξ with nonempty bounded intersection S and suppose that one of them, say S_1, is boundedly compact. Let $(a_n)_{n\geq 0}$ be any sequence constructed according to the algorithm*

$$(\forall n \in \mathbb{N})\ \ a_{n+1} \in \Pi_{i(n)}(a_n) \qquad \textit{with} \qquad i(n) = n \ (\text{modulo}\ m) + 1, \tag{3.24}$$

where a_0 is a point of attraction of $(S_i)_{i\in I}$ in the sense that

$$(\exists \dot{n} \in \mathbb{N})(\forall n \in \mathbb{N})\ \left\{ \begin{array}{l} n \geq \dot{n} \\ a_{mn} \notin S \end{array} \right. \quad \Rightarrow \quad d(a_{m(n+1)}, S) < d(a_{mn}, S). \tag{3.25}$$

Then $(a_{mn})_{n\geq 0}$ admits at least one strong cluster point and all of its cluster points lie in S. In addition, $(a_{mn})_{n\geq 0}$ converges strongly to a point in S if $\sum_{n\geq 0} d(a_{mn}, S) < +\infty$.

An interpretation of the above result is that convergence takes place locally, i.e., when the initial point a_0 is suitably positioned with respect to the property sets. A possible candidate for a starting point a_0 is an image which is feasible with respect to all the convex sets (such an image can be obtained by POCS or by any of the methods described in Section 5). Let us also mention that according to Proposition 2.7, if $\dim \Xi < +\infty$, the conditions on the sets in Theorem 3.3 reduce to closedness of the S_is and boundedness of S. A theorem similar to Theorem 3.3 can also be established for the set-valued version of the SIRT algorithm (3.11).

Besides projection methods, another approach to nonconvex feasibility problems is via the unconstrained minimization of a functional whose set of global minimizers is contained in S, e.g., the proximity function Φ of (3.7). Note that Φ will no longer be convex and that only local convergence results should be expected. In digital image recovery, certain stochastic minimization procedures could be contemplated. However, given their prohibitive computational cost in high dimensional spaces, this approach seems unrealistic at the present time.

4 Construction of Property Sets

4.1 Generalities

In this section we describe how *a priori* knowledge and data can be used to generate constraints on the solution and construct property sets in Ξ. In general, information may be known *a priori* or can be extracted *a posteriori* from the data. An example of possible *a priori* knowledge is the range of intensity values of the original image. In a recovery problem, *a posteriori* information can be obtained in various ways. For instance, if the original scene to be restored contains a point source, the blur function can be estimated from the degraded image; moreover, the statistics of flat regions in the degraded image can be used to estimate pertinent noise properties.

In image recovery problems, the two main sources of constraints are the intrinsic properties of the original image h and the properties of the imaging system. As demonstrated by the following examples, a lot of useful constraints give rise to closed and convex property sets. These few examples are meant only to illustrate some commonly used constraints. By no means do they exhaust the virtually unlimited list of sets that can be created.

4.2 Sets Based on Intrinsic Properties of the Image

The importance of spatial and spectral information in image recovery problems has been recognized in countless studies, e.g., [9], [37], [82], [117], [159], and [160]. As we shall now see, such information is to a large extent straightforward to incorporate in the form of convex sets. Other types of constraints will also be considered.

4.2.1 Spatial Properties

We provide here examples of sets based on attributes describing the original image h itself in the spatial domain. As a first example, suppose that lower and upper bounds on the amplitude of the original image h are known. This knowledge can be associated with the property set

$$S_i = \{a \in \Xi \mid \text{range}(a) \subset [\gamma, \delta]\}. \tag{4.1}$$

Another common assumption is that the image has limited region of support K [72]. The set associated with this information is

$$S_i = \{a \in \Xi \mid a = a1_K\}. \tag{4.2}$$

Next, suppose as in [173] that h is known over some domain K. Then the corresponding property set is

$$S_i = \{a \in \Xi \mid a1_K = h1_K\}. \tag{4.3}$$

If a bound γ^2 is available on the energy $\|h\|^2$ of the original image, one can define the set

$$S_i = \{a \in \Xi \mid ||a|| \leq \gamma\}. \tag{4.4}$$

More generally, when a bound γ is available on the maximum deviation of h from a reference image r, as in [100], [124], and [157], the associated property set is the ball

$$S_i = \{a \in \Xi \mid \|a - r\| \leq \gamma\}. \tag{4.5}$$

A further generalization of this type of closed and convex set is

$$S_i = \{a \in \Xi \mid \|\mathfrak{T}(a) - r\| \leq \gamma\}, \tag{4.6}$$

where $\mathfrak{T} : \Xi \to \Xi$ is a bounded linear operator. For instance, if $r = 0$ and $\mathfrak{T}$ is a differential operator, (4.6) is a set of smooth images; if $r = 0$ and $\mathfrak{T} = \mathrm{Id} - \mathfrak{T}'$, (4.6) is a set of images that are nearly invariant under the operator $\mathfrak{T}'$. Moment constraints have also been employed [154]. They yield property sets in the form of hyperslabs

$$S_i = \{a \in \Xi \mid \gamma \leq \langle a \mid b \rangle \leq \delta\}. \tag{4.7}$$

4.2.2 Spectral Properties

In many problems, certain attributes of the Fourier transform $\mathfrak{F}(a) = \widehat{a}$ of the original image are available. In optical experiments, they arise from partial measurements in the diffraction plane. In the following, $|\widehat{a}|$ is the Fourier magnitude of an image a and $\angle\widehat{a}$ its phase. The Fourier transform operator $\mathfrak{F}$ is defined in accordance with the $\mathfrak{L}^2$ space selected in Section 3.1.2.

A common assumption in image recovery is that the original image is band-limited [184]. If we designate by K the corresponding low frequency band, we obtain the property set

$$S_i = \{a \in \Xi \mid \widehat{a} = \widehat{a}1_K\}. \tag{4.8}$$

In [72] and [151], the stronger hypothesis that $\widehat{h}$ was known over some frequency band K led to the set

$$S_i = \{a \in \Xi \mid \widehat{a}1_K = \widehat{h}1_K\}. \tag{4.9}$$

This constraint can be generalized by considering the set of images that match approximately a reference image r over some frequency band K, i.e., [152]

$$S_i = \{a \in \Xi \mid \|(\widehat{a} - \widehat{r})\, 1_K\| \leq \gamma\}. \tag{4.10}$$

In particular, the set

$$S_i = \{a \in \Xi \mid ||\widehat{a} 1_K|| \leq \gamma\} \tag{4.11}$$

of images whose energy in a certain frequency band K is within some bound γ^2 was proposed in [184]. The same study also proposed the closed and convex cone

$$S_i = \{a \in \Xi \mid \text{range}(\widehat{a}) \subset \mathbb{R}_+\} \tag{4.12}$$

of images with nonnegative Fourier transform. The larger set of images with real Fourier transforms had been used previously in [106]. In [42], [108], and [184] knowledge of the phase of h was assumed to construct the set

$$S_i = \{a \in \Xi \mid \angle \widehat{a} = \angle \widehat{h}\}. \tag{4.13}$$

4.2.3 Other Properties

In the previous section, we have seen how sets could be derived from attributes of the Fourier transform $\mathfrak{F}(h)$ of the original image. Sets can also be constructed from attributes of other transforms $\mathfrak{T}(h)$ of h. For instance, $\mathfrak{T}$ can be the wavelet transform [32], [114], the bispectral transform [30], the singular value decomposition [146], or a differential operator [162].

4.3 Sets Based on Properties of the Imaging System

4.3.1 Overview

In this section, we describe how information pertaining to the imaging system can be used to construct property sets. The basic principle is as follows. From the data and the knowledge of the deterministic component of the imaging system, one forms an estimation residual which is then constrained to be consistent with those known probabilistic properties of the uncertain components in the system, i.e., measurement noise and, possibly, model uncertainty. As the estimation residual depends on the estimate, one thus obtains property sets in the solution space. Pieces of information relative to quantities such as range, moments, absolute moments, and second order probabilistic attributes are considered.

In image recovery, the idea of imposing noise-based constraints on the estimation residual was first implemented in the constrained least-squares restoration problem of [89], where the sample second moment of the residual was forced to match that of the noise. This particular constraint has also been employed in other restoration techniques, e.g., [48], [169], [171]. In the set theoretic deconvolution problem posed in [171], new constraints were introduced by considering other pieces of noise information (mean, outliers, spectral density) under the assumption that the noise was white and Gaussian. Work in this direction was pursued by considering random convolution kernels [48] as well as more general hypotheses on the noise and the imaging system [44], [50]. Some of the sets developed in [171] were re-examined via fuzzy set theory in [36]. In [51], the set theoretic deconvolution problem was studied in the context of bounded-error models and the only information available about the noise and the disturbances induced by random kernel perturbations consisted of amplitude bounds. Applications involving residual-based property sets can be found in [42], [49], [99], [127], and [144].

The following presentation is a synthesis of the results of [44], [50], [51], and [171] relevant to the construction of convex property sets.

4.3.2 Data Formation Model

In this section we introduce our mathematical model for the imaging system.

4.3.2.1 Notations

All the random elements are defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. All r.v.s are real-valued. The chi-square distribution with L degrees of freedom and mean L is denoted by χ^2_L. For every $p \in \mathbb{R}^*_+$, $\mathcal{L}^p(\mathsf{P})$ denotes the vector space of r.v.s with finite pth absolute moment. The abbreviations a.s. and i.i.d. stand, respectively, for P-almost surely and independent and identically distributed.

4.3.2.2 General Model

The observed data are discrete and consist of a sequence of r.v.s. $(X_n)_{n\in\mathbb{Z}}$ related to the original image h via the model

$$(\forall n \in \mathbb{Z}) \quad X_n = \mathcal{T}_n(h) + V_n, \tag{4.14}$$

where the random operators $(\mathcal{T}_n)_{n\in\mathbb{Z}}$ represent the imaging system and where the random sequence $(V_n)_{n\in\mathbb{Z}}$ represents measurement noise. Furthermore, these operators are

decomposable as

$$(\forall n \in \mathbb{Z}) \;\; \mathcal{T}_n = \overline{\mathcal{T}}_n + \widetilde{\mathcal{T}}_n, \tag{4.15}$$

where $\overline{\mathcal{T}}_n$ denotes the known, deterministic component of $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_n$ its unknown component, i.e., the component associated with model uncertainty. $(\overline{\mathcal{T}}_n)_{n\in\mathbb{Z}}$ and $(\widetilde{\mathcal{T}}_n)_{n\in\mathbb{Z}}$ are taken to be sequences of a.s. bounded linear random functionals on Ξ.[4] Moreover, the processes $(\widetilde{\mathcal{T}}_n(h))_{n\in\mathbb{Z}}$ and $(V_n)_{n\in\mathbb{Z}}$ are second-order, independent from each other, with mean zero. It follows from the Riesz representation theorem that there exists a sequence of Ξ-valued random elements $(\widetilde{T}_n)_{n\in\mathbb{Z}}$ such that (4.14) can be expressed as

$$(\forall n \in \mathbb{Z}) \;\; X_n = \overline{\mathcal{T}}_n(h) + U_n, \tag{4.16}$$

where

$$(\forall n \in \mathbb{Z}) \;\; U_n = \widetilde{\mathcal{T}}_n(h) + V_n = \langle h \mid \widetilde{T}_n \rangle + V_n. \tag{4.17}$$

The process $(U_n)_{n\in\mathbb{Z}}$ will be called the uncertainty process. It stands for the uncertainty arising from the inaccurate model and the noise.

4.3.2.3 Digital Model

If a digital image model is assumed (see Section 3.1.2.4), then X_n may represent the nth pixel of the degraded image in a restoration problem, or a point in a sinogram in a tomographic reconstruction problem. In addition, the $\widetilde{T}_n$s in (4.17) are simply N^2-dimensional random vectors. Note that if the vectors $(\widetilde{T}_n)_{n\in\mathbb{Z}}$ are identically distributed with uncorrelated components of variance v_1 and if the noise $(V_n)_{n\in\mathbb{Z}}$ is white with power v_2, then the power of the uncertainty process reduces to

$$\mathsf{E}|U_0|^2 = \mathsf{E}|\langle h \mid \widetilde{T}_0 \rangle|^2 + \mathsf{E}|V_0|^2 = \|h\|^2 \mathsf{v}_1 + \mathsf{v}_2. \tag{4.18}$$

4.3.2.4 Remarks

The model (4.16)-(4.17) is far from universal as it covers only situations in which the known component of the system is linear and the uncertainty is additive and affine. Nonetheless, it adequately approximates many physical systems encountered in imaging science and has the advantage of allowing unmodeled dynamics and random perturbations. In general, the operators $(\overline{\mathcal{T}}_n)_{n\in\mathbb{Z}}$ will represent a known mean component of the system and the operators $(\widetilde{\mathcal{T}}_n)_{n\in\mathbb{Z}}$ unknown variations about it. This level of generality is required in various contexts. For instance, in atmospheric, imaging random fluctuations

[4] The $\overline{\mathcal{T}}_n$s are therefore continuous. Linearity of the $\overline{\mathcal{T}}_n$s will guarantee the convexity of the S_is while continuity of the $\overline{\mathcal{T}}_n$s will guarantee closedness of the S_is [50].

of the index of refraction can seriously degrade the image and they must be accounted for. Other examples are found in X-ray imaging, where the image formed by a phosphor screen-film system results from the stochastic amplification and the random scattering of quanta, and in applications where the recording device is subject to random motions. Pertinent statistical descriptions of various imaging systems can be found in [14], [69], [77], and [185]. On the other hand, there also exists a vast body of problems for which the noise-only model

$$(\forall n \in \mathbb{Z}) \;\; X_n = \overline{\mathcal{T}}_n(h) + V_n, \tag{4.19}$$

is adequate. This is actually the model considered in [50], which resulted in a somewhat simpler analysis of the set construction process.

4.3.3 Set Construction Method

Given a proposed estimate a of h, the residual process $(Y_n(a))_{n\in\mathbb{Z}}$ is defined by

$$(\forall n \in \mathbb{Z}) \;\; Y_n(a) = X_n - \overline{\mathcal{T}}_n(a). \tag{4.20}$$

According to (4.16), the processes $(Y_n(h))_{n\in\mathbb{Z}}$ and $(U_n)_{n\in\mathbb{Z}}$ are equivalent, i.e.

$$(\forall n \in \mathbb{Z}) \;\; Y_n(h) = U_n \;\; \text{a.s.} \tag{4.21}$$

Therefore, they share the same probability theoretic property. Consequently, any known probabilistic property Ψ_i of the uncertainty process $(U_n)_{n\in\mathbb{Z}}$ constrains estimates to lie in the random set

$$S_i = \{a \in \Xi \mid (Y_n(a))_{n\in\mathbb{Z}} \text{ satisfies } \Psi_i\}\,. \tag{4.22}$$

Of course, this set cannot be utilized directly since only a finite segment $Y(a) = (X_n - \overline{\mathcal{T}}_n(a))_{1\le n\le L}$ of the residual process is observable in practice. We therefore replace (4.22) by the property set

$$S_i = \{a \in \Xi \mid Y(a) \text{ is consistent with } \Psi_i\}\,. \tag{4.23}$$

The above consistency statement can be formulated explicitly via statistical confidence theory. To this end, Ψ_i is associated with a statistic $Q_i(h)$ of $Y(h)$ whose distribution, exact or asymptotic, is determined. The set (4.23) is then rewritten in the more practical form

$$S_i = \{a \in \Xi \mid Q_i(a) \in r_i\}, \tag{4.24}$$

where the confidence region r_i is based on the distribution of $Q_i(h)$ and some confidence coefficient $1 - \epsilon_i \in]0, 1]$, i.e.,

$$1 - \epsilon_i = \mathsf{P}\{\omega \in \Omega \mid h \in S_i(\omega)\} = \mathsf{P}\{\omega \in \Omega \mid Q_i(h, \omega) \in r_i\}. \tag{4.25}$$

Henceforth, L will designate the length of the sample path.

4.3.4 Sets Based on Range Information

The L property sets arising from bounds on the amplitude range of the random variables $(U_n)_{n\in\mathbb{Z}}$ are

$$S_n = \{a \in \Xi \mid |X_n - \overline{\mathcal{T}}_n(a)| \leq \delta_n\} \qquad \text{for} \qquad 1 \leq n \leq L. \tag{4.26}$$

Proposition 4.1 [50] *The sets $(S_n)_{1\leq n\leq L}$ are closed and convex.*

We shall now see how the parameters $(\delta_n)_{1\leq n\leq L}$ can be determined from information on $(\langle h \mid \widetilde{T}_n\rangle)_{n\in\mathbb{Z}}$ and $(V_n)_{n\in\mathbb{Z}}$. The energy $\|h\|^2$ of the original image is assumed to be known.

4.3.4.1 Bounded Error Model

Suppose that the random sequences $(\|\widetilde{T}_n\|)_{n\in\mathbb{Z}}$ and $(V_n)_{n\in\mathbb{Z}}$ are a.s. uniformly bounded, say

$$(\forall n \in \mathbb{Z}) \;\; \|\widetilde{T}_n\| \leq \kappa_1 \qquad \text{and} \qquad |V_n| \leq \kappa_2 \qquad \text{a.s.} \tag{4.27}$$

Then, it follows from (4.17) and the Cauchy-Schwarz inequality that

$$\begin{aligned}
(\forall n \in \mathbb{Z}) \;\; |U_n| &\leq |\langle h \mid \widetilde{T}_n\rangle| + |V_n| && (4.28)\\
&\leq \|h\| \cdot \|\widetilde{T}_n\| + |V_n| && (4.29)\\
&\leq \|h\|\kappa_1 + \kappa_2 \qquad \text{a.s.} && (4.30)
\end{aligned}$$

This property places estimates in the set (4.26) where $\delta_n \triangleq \|h\|\kappa_1 + \kappa_2$. Let us note that h lies in each S_n almost surely. These sets can therefore be employed with a 100 percent confidence coefficient.

4.3.4.2 I.I.D. Model

Suppose that the $\langle h \mid \widetilde{T}_n\rangle$s and the V_ns are i.i.d. and that the distribution functions of $\|\widetilde{T}_0\|$ and V_0 are known. Then, by virtue of (4.29), we can find $\delta \in \mathbb{R}$ such that

$$\mathsf{P}\{\omega \in \Omega \mid |U_0(\omega)| \leq \delta\} = 1 - \epsilon_n, \tag{4.31}$$

where $1 - \epsilon_n$ is our preset confidence coefficient defined in (4.25). Since the U_ns are also i.i.d., all the points in the residual path should lie in the confidence interval $[-\delta, \delta]$ with probability $1 - \epsilon_n$. Therefore, the L sets of images that satisfy this constraint are given in (4.26), where $(\forall n \in \{1, \cdots, L\}) \;\; \delta_n = \delta$.

4.3.4.3 General Model

Let us assume that the distribution functions of the r.v.s $(\|\widetilde{T}_n\|)_{1\leq n\leq L}$ are known as well as those of the r.v.s $(V_n)_{1\leq n\leq L}$. Then, thanks to (4.29), we can find parameters $(\delta_n)_{1\leq n\leq L}$ for the sets (4.26) such that

$$\mathsf{P}\{\omega\in\Omega \mid |U_n(\omega)|\leq\delta_n\}=1-\epsilon_n. \tag{4.32}$$

4.3.5 Sets Based on Moment Information

It is assumed that the uncertainty process $(U_n)_{n\in\mathbb{Z}}$ consists of i.i.d.r.v.s. Extensions of the following results to dependent variables are possible thanks to the various Central Limit Theorems that exist for mixing processes (see, e.g., [13]).

4.3.5.1 Mean

The sample mean of the uncertainty process is the statistic

$$M=\frac{1}{L}\sum_{n=1}^{L}U_n. \tag{4.33}$$

A straightforward application of the standard Central Limit Theorem shows that under our assumptions M is asymptotically normal with mean zero and variance $\sigma^2=\mathsf{E}|U_0|^2/L$ [67]. The same property should therefore be satisfied by the residual process. Whence, for a given confidence coefficient, a confidence interval $[-\alpha,\alpha]$ for M/σ is determined from the tables of the standard normal distribution by making the normal approximation. The set of images that yield a residual sample mean within this confidence interval is

$$S_m=\{a\in\Xi \mid \left|\sum_{n=1}^{L}X_n-\overline{\mathcal{T}}_n(a)\right|\leq\alpha\sqrt{L\mathsf{E}|U_0|^2}\}. \tag{4.34}$$

The second moment $\mathsf{E}|U_0|^2$ can be obtained from (4.18) if the required conditions are met. In general, note that the above assumptions on $(U_n)_{n\in\mathbb{Z}}$ are satisfied when the $\langle h \mid \widetilde{T}_n\rangle$s and the V_ns are i.i.d. and that $\mathsf{v}_1\triangleq\mathsf{E}\|\widetilde{T}_0\|^2$ and $\mathsf{v}_2\triangleq\mathsf{E}|V_0|^2$ are known. Thanks to our hypotheses, the variance of U_0 can then be majorized by

$$\mathsf{E}|U_0|^2=\mathsf{E}|\langle h \mid \widetilde{T}_0\rangle|^2+\mathsf{E}|V_0|^2\leq\|h\|^2\mathsf{v}_1+\mathsf{v}_2 \tag{4.35}$$

to produce a useful bound in (4.34).

Proposition 4.2 [50] *S_m is closed and convex.*

4.3.5.2 Absolute Moments

Suppose that, for a fixed $p \in [1, +\infty[$, $U_0 \in \mathcal{L}^{2p}(\mathsf{P})$ and that the pth and $2p$th absolute moments of U_0 are known. The pth sample absolute moment of the uncertainty process is the statistic

$$M_p = \frac{1}{L} \sum_{n=1}^{L} |U_n|^p. \tag{4.36}$$

Under the above hypotheses, as the sample size L tends to infinity, M_p is asymptotically normal with mean $\mathsf{E}|U_0|^p$ and variance $\sigma_p^2 = (\mathsf{E}|U_0|^{2p} - \mathsf{E}^2|U_0|^p)/L$ [67]. Therefore, by invoking the limiting distribution, one can compute a confidence interval $[-\alpha, \alpha]$ for $(M_p - \mathsf{E}|U_0|^p)/\sigma_p$ based on some confidence coefficient. Hence, the subset of Ξ of images which yield a residual sample absolute moment within the desired confidence interval is

$$S_p = \{a \in \Xi \mid \eta_p \leq \sum_{n=1}^{L} |X_n - \overline{\mathcal{T}}_n(a)|^p \leq \zeta_p\}, \tag{4.37}$$

where

$$\eta_p = \begin{cases} L(\mathsf{E}|U_0|^p - \alpha\sigma_p) & \text{if } \mathsf{E}|U_0|^p > \alpha\sigma_p \\ 0 & \text{otherwise} \end{cases} \tag{4.38}$$

and

$$\zeta_p = L(\mathsf{E}|U_0|^p + \alpha\sigma_p). \tag{4.39}$$

Let

$$S_p^- = \{a \in \Xi \mid \sum_{n=1}^{L} |X_n - \overline{\mathcal{T}}_n(a)|^p < \eta_p\} \tag{4.40}$$

denote the convex deficiency of S_p and

$$S_p^+ = \{a \in \Xi \mid \sum_{n=1}^{L} |X_n - \overline{\mathcal{T}}_n(a)|^p \leq \zeta_p\} \tag{4.41}$$

its convex hull. Then $S_p = S_p^+ \smallsetminus S_p^-$.

Proposition 4.3 [50] *S_p^+ is closed and convex.*

In the particular case when $p = 2$ and U_0 is zero mean Gaussian, the exact distribution of $LM_p/\mathsf{E}|U_0|^2$ is a χ_L^2 [67]. Thus, from the tables of the χ_L^2, one can obtain a value of ζ_2 which is more accurate than that resulting from the normal approximation.

As an example of computation of the parameter ζ_p, consider the case when the $\langle h \mid \widetilde{T}_n \rangle$s and the V_ns are i.i.d. and $\mathsf{E}\|\widetilde{T}_0\|^{2p}$, $\mathsf{E}\|\widetilde{T}_0\|^{p}$, $\mathsf{E}|V_0|^{2p}$, and $\mathsf{E}|V_0|^{p}$ are known. Then the U_ns are also i.i.d. and their pth absolute moment can be majorized as

$$\begin{aligned} \mathsf{E}|U_0|^p &\leq \mathsf{E}\left(|\langle h \mid \widetilde{T}_0\rangle| + |V_0|\right)^p && (4.42)\\ &\leq 2^{p-1}\left(\mathsf{E}|\langle h \mid \widetilde{T}_0\rangle|^p + \mathsf{E}|V_0|^p\right) && (4.43)\\ &\leq 2^{p-1}\left(\|h\|^p\mathsf{E}\|\widetilde{T}_0\|^p + \mathsf{E}|V_0|^p\right). && (4.44) \end{aligned}$$

On the other hand, we can majorize σ_p^2 by $\mathsf{E}|U_0|^{2p}$, which can itself be approximated as above.

4.3.6 Sets Based on Second Order Information

Since the processes are real-valued, the spectral distributions are defined on $[0, 1/2]$ (see [58] for details). It is assumed that L is even (if not, $L/2$ should be replaced by $(L-1)/2$ thereafter).

4.3.6.1 Gaussian White Uncertainty Process

Theorem 4.1 [143] *Let $(U_n)_{n\in\mathbb{Z}}$ be a zero mean Gaussian discrete white noise process with power σ^2. Define*

$$(\forall k \in \{0, \cdots, L/2\}) \;\; I_k = \frac{2}{L}\left|\sum_{n=1}^{L} U_n \exp(-\imath\frac{2\pi}{L}kn)\right|^2. \tag{4.45}$$

Then

(i) *The statistics $(I_k)_{0\leq k\leq L/2}$ are independent.*

(ii) *The statistics $I_0/2\sigma^2$ and $I_{L/2}/2\sigma^2$ have a χ_1^2 distribution.*

(iii) *The statistics $(I_k/\sigma^2)_{1\leq k\leq L/2-1}$ have a χ_2^2 distribution.*

Now, suppose that $(U_n)_{n\in\mathbb{Z}}$ satisfies the assumptions of Theorem 4.1. Then, from Theorem 4.1 and the tables of the χ_1^2 and χ_2^2 distributions, one can determine confidence intervals $[0, \beta_1]$ and $[0, \beta_2]$ for the r.v.s in (ii) and (iii) respectively. Consequently, the sets

of images that produce a residual path consistent, to within a desired confidence coefficient $1-\epsilon_k$, with the whiteness and normality of the uncertainty process are

$$S_k = \{a \in \Xi \mid \left| \sum_{n=1}^{L} (X_n - \overline{\mathcal{T}}_n(a)) \exp(-\imath \frac{2\pi}{L} kn) \right|^2 \leq \xi_k\} \qquad \text{for} \quad 0 \leq k \leq L/2, \tag{4.46}$$

where

$$(\forall k \in \{0, \cdots, L/2\}) \;\; \xi_k = \begin{cases} L\sigma^2\beta_1 & \text{if } \; k = 0 \;\; \text{or } L/2 \\ L\sigma^2\beta_2/2 & \text{if } \; 0 < k < L/2. \end{cases} \tag{4.47}$$

Proposition 4.4 [50] *The sets $(S_k)_{0\leq k\leq L/2}$ are closed and convex.*

Note that since the χ^2_2 distribution is simply an exponential distribution with parameter $1/2$, (4.47) reduces to

$$(\forall k \in \{0, \cdots, L/2\}) \;\; \xi_k = \begin{cases} L\sigma^2\beta_1 & \text{if } \; k = 0 \;\; \text{or } L/2 \\ -L\sigma^2 \ln(\epsilon_k) & \text{if } \; 0 < k < L/2. \end{cases} \tag{4.48}$$

In addition, it should be observed that, for $k = 0$, S_k is essentially the same set as the mean set (4.34).

4.3.6.2 Non-Gaussian White Uncertainty Process

Suppose that $(U_n)_{n\in\mathbb{Z}}$ is a discrete white noise process consisting of i.i.d.r.v.s all distributed as a zero mean r.v. $U_0 \in \mathcal{L}^4(\mathsf{P})$, with variance σ^2. Then the r.v.s in (ii) and (iii) of Theorem 4.1 are asymptotically distributed as a χ^2_1 and a χ^2_2 respectively [93]. Thus, under relatively mild conditions, the conclusions of Theorem 4.1 hold in an asymptotic sense. Consequently, since in image processing applications L is typically large, the sets $(S_k)_{0\leq k\leq L/2}$ of (4.46) can be used.

4.3.6.3 Correlated Uncertainty Process

In this section, we further generalize the analysis by dropping the whiteness assumption. We shall base the construction of a spectral set in this case on the following theorem.

Theorem 4.2 [143] *Let $(U_n)_{n\in\mathbb{Z}}$ be a zero mean strictly stationary strongly mixing process with summable second and fourth order cumulant functions and spectral density g. Let $0 = \nu_0 < \nu_1 < \cdots < \nu_m = 1/2$ and*

$$(\forall k \in \{0, \cdots, m\}) \;\; I_k = \frac{2}{L} \left| \sum_{n=1}^{L} U_n \exp(-\imath 2\pi\nu_k n) \right|^2. \tag{4.49}$$

Then

(i) *The statistics* $(I_k)_{0\leq k\leq m}$ *are asymptotically independent.*

(ii) *The statistics* $I_0/g(0)$ *and* $I_m/g(1/2)$ *are asymptotically distributed as a* χ^2_1.

(iii) *The statistics* $(2I_k/g(\nu_k))_{1\leq k\leq m-1}$ *are asymptotically distributed as a* χ^2_2.

Loosely speaking, Theorem 4.2 states that if the span of dependence of the process is small enough, the results of Theorem 4.1 can be generalized for large L. Now suppose that $(U_n)_{n\in\mathbb{Z}}$ satisfies the hypotheses of Theorem 4.2 and that its spectral density g is known at points $0 \leq \nu_0 < \nu_1 < \cdots < \nu_m \leq 1/2$. Then, given a confidence coefficient $1-\epsilon_k$, one can compute the confidence intervals $[0,\beta_1]$ and $[0,\beta_2]$ for the r.v.s in (ii) and (iii) respectively by invoking their asymptotic properties (as before, note that $\beta_2 = -2\ln(\epsilon_k)$). $(U_n)_{n\in\mathbb{Z}}$ and $(Y_n(h))_{n\in\mathbb{Z}}$ being equivalent, this leads to the sets

$$S_k = \{a \in \Xi \mid \left|\sum_{n=1}^{L}(X_n - \overline{\mathcal{T}}_n(a))\exp(-\imath 2\pi\nu_k n)\right|^2 \leq \xi_k\} \qquad \text{for} \quad 0\leq k\leq m, \tag{4.50}$$

where

$$\xi_0 = \begin{cases} Lg(0)\beta_1/2 & \text{if } \nu_0 = 0 \\ Lg(\nu_0)\beta_2/4 & \text{if } \nu_0 > 0, \end{cases} \tag{4.51}$$

$$(\forall k \in \{1,\cdots,m-1\}) \;\; \xi_k = Lg(\nu_k)\beta_2/4, \tag{4.52}$$

$$\xi_m = \begin{cases} Lg(\nu_m)\beta_2/4 & \text{if } \nu_m < 1/2 \\ Lg(1/2)\beta_1/2 & \text{if } \nu_m = 1/2. \end{cases} \tag{4.53}$$

Naturally, Proposition 4.4 still holds.

The processes $(\langle h \mid \widetilde{T}_n\rangle)_{n\in\mathbb{Z}}$ and $(V_n)_{n\in\mathbb{Z}}$ have mean zero and are independent from each other. Therefore, if they possess respectively spectral densities g_1 and g_2, the spectral density of $(U_n)_{n\in\mathbb{Z}}$ will be $g = g_1 + g_2$. In particular, if the $\widetilde{T}_n$s are i.i.d. and if $(V_n)_{n\in\mathbb{Z}}$ is white with power v_2, g will be defined as

$$(\forall \nu \in [0,1/2]) \;\; g(\nu) = 2\left(\mathsf{E}|\langle h \mid \widetilde{T}_0\rangle|^2 + \mathsf{v}_2\right). \tag{4.54}$$

This expression will be evaluated as in (4.18) under suitable hypotheses or majorized as in (4.35) in general.

4.4 Information Management

In order to produce the most accurate set theoretic estimates, one should exploit all the information available in a given problem. Indeed, the larger the number of sets intersected in (1.6), the smaller the resulting feasibility set S. This statement, however, should be tempered by the requirement that the information be utilized efficiently and reliably.

To process the available information efficiently, all the constraints that do not contribute to a significantly smaller feasibility set should be discarded, especially if their processing cost is high (meaning, for instance, that a projection method is employed to find a feasible solution and that the projections onto the associated sets are computationally involved).

The issue of reliability comes into play when statistical constraints are present, as in Section 4.3. In that case, the feasibility set depends on a realization of the stochastic data process (4.16) and one will obtain a reliable set theoretic formulation only if the confidence level

$$\mathsf{c} = \mathsf{P}\{\omega \in \Omega \mid h \in S(\omega)\} \tag{4.55}$$

on the solution set is sufficiently large, say $\mathsf{c} \geq 0.90$. In the jargon of Section 3.1.3, c is the probability of obtaining a fair set theoretic formulation. Of course, one has control only over the confidence coefficient $1 - \epsilon_i$ placed on each property set in (4.25). It should be borne in mind that the value of these coefficients should be determined in terms of the sets used and not preset to some *ad hoc* value. To illustrate this point, consider the scenario of Section 4.3.4.2 and suppose that the L sets (4.26) are to be used. If, as suggested in certain digital image recovery studies, one took $1 - \epsilon = 0.99$ as a confidence coefficient on each set, one would arrive at an overall confidence of $\mathsf{c} = 0.99^L = 0.99^{N^2} \approx 0$. Consequently, such a set theoretic formulation would be unlikely to be fair or even consistent, and would fail to represent reliably the original image. A 99 percent confidence on each set might be acceptable when just a few sets are used, e.g., mean and second moment, but not in large scale problems. In general, the statistics $(Q_i)_{i \in I}$ defining the property sets (4.26), (4.34), (4.37), (4.46), and (4.50) may be dependent and the relation between c and $1 - \epsilon$ may be difficult to establish when joint distribution functions are not available. Such simultaneous inference problems are discussed in [118].

Coming back to the problem of using information efficiently, let us stress that in the presence of statistical constraints, a trade-off arises in the selection of property sets. Indeed, the confidence coefficient on each set must increase with the number of sets selected in order to maintain a fixed overall confidence in (4.55). Consequently, one ends up intersecting a larger number of larger sets. This certainly increases the complexity of the resulting feasibility problem while possibly having little effect on reducing the feasibility set. For instance, the information that the uncertainty process is white and gaussian with mean zero and known power leads to an infinite number of sets of type (4.37) since all the

absolute moments are then known. Of course, not all of them should be used. Thus, efficiency and reliability appear as two intertwined factors that should be carefully considered in selecting property sets.

5 Solving the Convex Feasibility Problem

5.1 Introduction

The goal of this section is to describe methods to solve the convex feasibility problem (1.6). Recall that, unless otherwise stated, $(S_i)_{i\in I}$ is a countable family of closed and convex subsets of Ξ with nonempty intersection S.

The convex feasibility problem is a central problem in applied mathematics [11], [25], [38], [56], [111], [141], which can be formulated in various ways, such as:

1. Finding a common point of closed and convex sets.
2. Finding a common fixed-point of nonexpansive operators.
3. Finding a common minimum of convex functionals.
4. Finding a common zero of maximal monotone operators.
5. Solving a system of variational inequalities.
6. Solving a system of convex inequalities.

Surveys of methods for solving such problems can be found in [25] and [38]. Since these surveys were written, the convex feasibility problem has been the focus of a significant research effort. As a result, a good part of the material presented here will be new. It should also be mentioned that two very important papers in this area were published in 1967 by Browder [19] and Gubin *et al.* [79]. The importance of fundamental concepts such as Fejér-monotonicity, admissibility, and bounded regularity was stressed in these papers and basic proof techniques were established. More recent work has mainly been geared towards various generalizations, especially in the direction of parallel algorithms.

In Section 5.2, we shall first discuss the limitations of the popular POCS algorithm, which will motivate the subsequent developments on alternative algorithms to solve the convex feasibility problem. In Section 5.3, we discuss a parallel projection method for solving in a least-squares sense inconsistent image feasibility problems. We then go back to consistent problems and discuss successively projection methods in Section 5.4, approximate projection methods in Section 5.5, subgradient projection methods in Section 5.6, and finally fixed-point methods in Section 5.7. These various approaches are considered from a higher perspective in Section 5.8.

For the sake of completeness, we shall maintain the discussion at a fairly general theoretical level. We are nontheless aware of the more practical concerns of practicing engineers and scientists who are interested mainly in digital image processing applications, in

which recovery is performed on a digital computer with a finite number of constraints. Section 5.9 will be devoted to this framework and a number of practical issues will be discussed there. A few proofs have been included to illustrate the relevance of certain assumptions and give more theoretical insight into convergence issues.

5.2 The Limitations of the POCS Method

Let us recall that the POCS algorithm is defined by the iteration process

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda_n (P_{i(n)}(a_n) - a_n), \tag{5.1}$$

where the control is periodic, i.e.,

$$(\forall n \in \mathbb{N}) \;\; i(n) = n \,(\text{modulo}\, m) + 1 \qquad \text{with} \qquad m = \text{card}\, I < +\infty, \tag{5.2}$$

and where the relaxation parameters satisfy

$$(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq 2 - \varepsilon \quad \text{with} \quad 0 < \varepsilon < 1. \tag{5.3}$$

As mentioned in Section 3.3, POCS has been the prevalent solution method in convex set theoretic image recovery. It is nonetheless limited in several respects.

5.2.1 Serial Structure

A salient feature of POCS is its serial algorithmic structure: at each iteration only one of the property sets can be activated. Clearly, such a structure does not lend itself naturally to implementations on architectures with parallel processors.

5.2.2 Slow Convergence

A problem with POCS which has long been recognized is its slow convergence. Conceptually, the algorithm can be accelerated by properly relaxing the projections at each iteration. Unfortunately, even for simple set theoretic formulations, there is no systematic method for determining $(\lambda_n)_{n \geq 0}$ so as to speed up the iterations. For instance, when all the S_is are affine half-spaces, there is no systematic answer as to whether underrelaxations are faster than overrelaxations or vice-versa [85], [115]. Likewise, in the studies reported in [159], only heuristic rules for specific problems are given.

5.2.3 Inconsistent Problems

The convergence properties of the unrelaxed version of POCS, that is

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = P_{n \,(\text{modulo}\, m)+1}(a_n), \tag{5.4}$$

in inconsistent problems were studied in [79] (additional convergence results were recently established in [12]).

Theorem 5.1 [79] *Let $(a_n)_{n\geq 0}$ be any sequence generated by (5.4) and suppose that one of the sets in $(S_i)_{1\leq i\leq m}$ is bounded. Then 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, for every $i \in \{1, \cdots, m\}$, the periodic 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 theorem simply states that the sequence $(a_{2n+1})_{n\geq 0}$ converges weakly to a point $\overline{a}_1 \in S_1$ such that $P_1(P_2(\overline{a}_1)) = \overline{a}_1$, i.e., to an image that satisfies property Ψ_1 and which is closest to satisfying Ψ_2 (this result is also discussed in [33], [76], and [183]). Beyond two sets, however, the above result has no useful interpretation and little practical value. It merely indicates that the limit image $\overline{a}_i$ lies in S_i and, thereby, satisfies Ψ_i. Aside from Ψ_i, however, the properties of $\overline{a}_i$ 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 image. Thus, the convergence behavior of POCS in the inconsistent case is generally unsatisfactory.

5.2.4 Countable Set Theoretic Formulations

Countable set theoretic formulations are of great theoretical interest and they are also encountered in certain analog problems. POCS is limited to finite set theoretic formulations and it cannot be used in such problems.

5.3 Inconsistent Problems

In this section, $(S_i)_{i\in I}$ is a finite family of m sets whose intersection may be empty and the strictly convex weights $(w_i)_{i\in I}$ are those of (3.7)/(3.8). Following [42], we present parallel projection methods to find least-squares solutions to inconsistent convex image feasibility problems. The problem of finding an image that minimizes a weighted average of the squares of the distances to the property sets is reformulated in the product space $\boldsymbol{\Xi}$ of Section 2.8, where it is equivalent to that of finding a point that lies in the diagonal

subspace $\mathbf{D}$ and at minimum distance from the cartesian product $\mathbf{S}$ of the original sets. A solution is obtained in $\boldsymbol{\Xi}$ via methods of alternating projections which lead naturally to methods of parallel projections in the original space Ξ.

5.3.1 Least-Squares Solutions

In inconsistent problems, there exists no image possessing exactly all the properties $(\Psi_i)_{i\in I}$ but one can look for an image 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 solution. In set theoretic terms, if $(S_i)_{i\in I}$ represents the family of hyperplanes of $\mathbb{R}^k$ associated with the equations, this is equivalent to looking for a point a^* which minimizes $\sum_{i\in I} d(a, S_i)^2$, the sum of the squares of the distances to the S_is. Along the same lines, the exact feasibility problem (1.6) can be replaced by the weighted least-squares feasibility problem of minimizing the proximity function (3.7), that is

$$\text{Find } a^\star \in G = \{a \in \Xi \mid (\forall b \in \Xi)\ \ \Phi(a) \leq \Phi(b)\}. \tag{5.5}$$

Of course, if $\bigcap_{i\in I} S_i \neq \emptyset$, the minimum value of the proximity function is 0 which is attained only on $G = \bigcap_{i\in I} S_i$, so that (1.6) and (5.5) coincide. In general, (5.5) can be viewed as an extension of (1.6) and G is the set of least-squares solutions of the (possibly inconsistent) image feasibility problem. From an image processing point of view, such solutions are clearly more acceptable and useful than those generated by POCS, whose properties were seen to be elusive.

It should be noted that in finite dimensional spaces, and under certain conditions on the problem, (5.1) can solve (5.5) if the sequence of relaxation parameters $(\lambda_n)_{n\geq 0}$ approaches zero [27], [133]. Experimental evidence first suggested this property in the inconsistent tomographic reconstruction problems of [85], where POCS was reported to provide better results with strong underrelaxations than without relaxations, as in (5.4). 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.

5.3.2 Alternating Projections in a Product Space

As shown in Section 2.8, the original convex feasibility problem (1.6) can be recast in the m-fold product space $\boldsymbol{\Xi}$ as the new feasibility problem (2.47) of finding a point $\mathbf{a}^*$ common to the product $\mathbf{S}$ of the property sets and the diagonal subspace $\mathbf{D}$ of $\boldsymbol{\Xi}$. When $\bigcap_{i\in I} S_i = \emptyset$, then $\mathbf{S} \bigcap \mathbf{D} = \emptyset$ and the best approximate solution will be to find a point

$\mathbf{a}^*$ in $\mathbf{D}$ which is at minimum distance from $\mathbf{S}$. This statement can be formalized by introducing the functional

$$\begin{aligned} \mathbf{\Phi} : \mathbf{D} &\to \mathbb{R}_+ \\ \mathbf{a} &\mapsto \frac{1}{2}\mathbf{d}(\mathbf{a}, \mathbf{S})^2, \end{aligned} \tag{5.6}$$

and calling $\mathbf{G}$ its set of minimizers.

Proposition 5.1 [42] *In the product space $\mathbf{\Xi}$, the weighted least-squares problem (5.5) is equivalent to minimizing $\mathbf{\Phi}$, i.e., to solving*

$$\text{Find } \mathbf{a}^\star \in \mathbf{G}. \tag{5.7}$$

Now let $P_{\mathbf{D}}$ and $P_{\mathbf{S}}$ be the operators of projection onto the sets $\mathbf{D}$ and $\mathbf{S}$. Then $\mathbf{G} = \operatorname{Fix} P_{\mathbf{D}} \circ P_{\mathbf{S}}$ [33] and the following Theorem provides an alternating projection method to solve (5.7).

Theorem 5.2 *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*

$$(\forall n \in \mathbb{N}) \;\; \mathbf{a}_{n+1} = \mathbf{a}_n + \lambda_n (P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n), \tag{5.8}$$

where the relaxation parameters $(\lambda_n)_{n\geq 0}$ satisfy (5.3) converges weakly to a point in $\mathbf{G}$.

Proof. Let $(\mathbf{a}_n)_{n\geq 0}$ be any sequence generated by the algorithm. Let $T = P_{\mathbf{D}} \circ P_{\mathbf{S}}$ and fix $\mathbf{c} \in \mathbf{G} = \operatorname{Fix} T$, $n \in \mathbb{N}$. Then $T : \mathbf{D} \to \mathbf{D}$ is firmly nonexpansive, as shown by the relations

$$\begin{aligned} (\forall (\mathbf{a}, \mathbf{b}) \in \mathbf{D}^2) \;\; |||T(\mathbf{a}) - T(\mathbf{b})|||^2 &\leq |||P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b})|||^2 \\ &\leq \langle\langle \mathbf{a} - \mathbf{b} \mid P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b}) \rangle\rangle \\ &= \langle\langle \mathbf{a} - \mathbf{b} \mid P_{\mathbf{D}}(P_{\mathbf{S}}(\mathbf{a}) - P_{\mathbf{S}}(\mathbf{b})) \rangle\rangle \\ &= \langle\langle \mathbf{a} - \mathbf{b} \mid T(\mathbf{a}) - T(\mathbf{b}) \rangle\rangle, \end{aligned} \tag{5.9}$$

where we have used successively the nonexpansivity of $P_{\mathbf{D}}$, then the firm nonexpansivity of $P_{\mathbf{S}}$ (see Proposition 2.8(i)), and then (2.21) since $P_{\mathbf{D}}$ is linear and $\mathbf{a} - \mathbf{b} \in \mathbf{D}$. As $\mathbf{c} \in \operatorname{Fix} T$, (5.9) yields $\langle\langle \mathbf{a}_n - \mathbf{c} \mid T(\mathbf{a}_n) - \mathbf{c} \rangle\rangle \geq |||T(\mathbf{a}_n) - \mathbf{c}|||^2$ and therefore

$$\langle\langle T(\mathbf{a}_n) - \mathbf{c} \mid T(\mathbf{a}_n) - \mathbf{a}_n \rangle\rangle \leq 0. \tag{5.10}$$

Whence,

$$\begin{aligned} \langle\langle \mathbf{a}_n - \mathbf{c} \mid T(\mathbf{a}_n) - \mathbf{a}_n \rangle\rangle &= -|||T(\mathbf{a}_n) - \mathbf{a}_n|||^2 + \langle\langle T(\mathbf{a}_n) - \mathbf{c} \mid T(\mathbf{a}_n) - \mathbf{a}_n \rangle\rangle \\ &\leq -|||T(\mathbf{a}_n) - \mathbf{a}_n|||^2. \end{aligned} \tag{5.11}$$

Then (5.8), (5.11), and (5.3) imply

$$\begin{aligned}
|||\mathbf{a}_{n+1}-\mathbf{c}|||^2 &= |||\mathbf{a}_n-\mathbf{c}|||^2+2\langle\langle \mathbf{a}_n-\mathbf{c} \mid \mathbf{a}_{n+1}-\mathbf{a}_n\rangle\rangle + |||\mathbf{a}_{n+1}-\mathbf{a}_n|||^2 \\
&= |||\mathbf{a}_n-\mathbf{c}|||^2+2\lambda_n\langle\langle \mathbf{a}_n-\mathbf{c} \mid T(\mathbf{a}_n)-\mathbf{a}_n\rangle\rangle + \lambda_n^2|||T(\mathbf{a}_n)-\mathbf{a}_n|||^2 \\
&\le |||\mathbf{a}_n-\mathbf{c}|||^2-\lambda_n(2-\lambda_n)|||T(\mathbf{a}_n)-\mathbf{a}_n|||^2 \qquad (5.12)\\
&\le ||||\mathbf{a}_n-\mathbf{c}|||^2-\varepsilon^2|||T(\mathbf{a}_n)-\mathbf{a}_n|||^2 \qquad (5.13)\\
&\le |||\mathbf{a}_n-\mathbf{c}|||^2. \qquad (5.14)
\end{aligned}$$

Hence, $(\mathbf{a}_n)_{n\ge 0}$ is Fejér-monotone with respect to $\mathbf{G}$. According to Proposition 2.11, it possesses a weak cluster point $\mathbf{a}$, say $\mathbf{a}_{n_k} \overset{k}{\rightharpoonup} \mathbf{a}$, and it remains to show $\mathbf{a}\in\mathbf{G}$. In view of (5.13), we have

$$|||\mathbf{a}_n-T(\mathbf{a}_n)|||^2 \le \varepsilon^{-2}\left(|||\mathbf{a}_n-\mathbf{c}|||^2-|||\mathbf{a}_{n+1}-\mathbf{c}|||^2\right). \qquad (5.15)$$

But since the nonnegative sequence $(|||\mathbf{a}_n-\mathbf{c}|||)_{n\ge 0}$ is nonincreasing, it converges and therefore $\mathbf{a}_n-T(\mathbf{a}_n) \overset{n}{\to} \mathbf{0}$. According to Proposition 2.8(i), $\mathrm{Id}-T$ is demiclosed and therefore $(\mathrm{Id}-T)(\mathbf{a})=\mathbf{0}$ since $\mathbf{a}_{n_k}\overset{k}{\rightharpoonup} a$ and $(\mathrm{Id}-T)(\mathbf{a}_{n_k})\overset{k}{\to}\mathbf{0}$. Whence $\mathbf{a}\in\operatorname{Fix}T=\mathbf{G}$. □

A pictorial description of (5.8) is given in Fig. 9: $\mathbf{s}_n = P_{\mathbf{S}}(\mathbf{a}_n)$ 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 $\varepsilon\le\lambda\le 1$ or $1\le\lambda_n\le 2-\varepsilon$. As discussed in Section 5.2.3, in the unrelaxed case Theorem 5.2 follows from Theorem 5.1. A noteworthy property of (5.8) is that it can be viewed as a gradient method, as stated in the following proposition.

Proposition 5.2 [42] *Let $(\mathbf{a}_n)_{n\ge 0}$ be any sequence of iterates in Theorem 5.2. Then $(\mathbf{\Phi}(\mathbf{a}_n))_{n\ge 0}$ decreases until convergence and*

$$(\forall n\in\mathbb{N})\quad \mathbf{a}_{n+1}=\mathbf{a}_n-\lambda_n\nabla_{\mathbf{D}}\mathbf{\Phi}(\mathbf{a}_n), \qquad (5.16)$$

where $\nabla_{\mathbf{D}}$ is the gradient operator in the Hilbert space $\mathbf{D}$. Moreover, at iteration n, the relaxation parameter which is optimal in terms of bringing $\mathbf{a}_{n+1}$ closest to an arbitrary point $\mathbf{a}^*$ in $\mathbf{G}$ is

$$\lambda_n^* = \frac{\langle\langle P_{\mathbf{D}}\circ P_{\mathbf{S}}(\mathbf{a}_n)-\mathbf{a}_n \mid \mathbf{a}^*-\mathbf{a}_n\rangle\rangle}{|||P_{\mathbf{D}}\circ P_{\mathbf{S}}(\mathbf{a}_n)-\mathbf{a}_n|||^2} \ge 1. \qquad (5.17)$$

We observe that the optimal relaxation parameter λ_n^* depends on a solution point $\mathbf{a}^*$, which of course is not known. Hence, optimal relaxations cannot be achieved. However, the above proposition indicates that they are always overrelaxations.

Strong convergence of the unrelaxed version of (5.8) can be proved if one makes additional assumptions on $\mathbf{S}$, such as compactness [33], finite dimensionality [33], or uniform convexity [79]. The next theorem presents a strong convergence result for a variant of (5.8) which does not require special conditions.

Theorem 5.3 *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*

$$(\forall n \in \mathbb{N}) \;\; \mathbf{a}_{n+1} = (1-\alpha_n)\mathbf{a}_0 + \alpha_n(\lambda P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) + (1-\lambda)\mathbf{a}_n), \tag{5.18}$$

where $0 < \lambda \leq 2$ *and where* $(\alpha_n)_{n\geq 0} \subset [0,1[$ *satisfies*

$$\begin{cases} \lim_{n\to+\infty} \alpha_n = 1 \\ \sum_{n\geq 0}(1-\alpha_n) = +\infty \\ \sum_{n\geq 0} |\alpha_{n+1} - \alpha_n| < +\infty, \end{cases} \tag{5.19}$$

converges strongly to $P_{\mathbf{G}}(\mathbf{a}_0)$.

Proof. Similar to that found in [42], except that we now use the more general conditions (5.19) allowed by a fixed-point theorem of [179]. □

Note that, as n increases, (5.18) tends to behave like a constant-relaxation version of (5.8). Moreover, a simple example of sequence $(\alpha_n)_{n\geq 0}$ that satisfies (5.19) is

$$(\forall n \in \mathbb{N}) \;\; \alpha_n = \frac{n}{n+1}. \tag{5.20}$$

5.3.3 Simultaneous Projection Methods

In the previous section we have solved the least-squares feasibility problem (5.5) in the product space $\mathbf{\Xi}$. It remains to reformulate the solution methods in the original signal space Ξ, where they will actually be employed. First, we must secure conditions under which (5.5) admits solutions.

Proposition 5.3 [42], [55] *Suppose that either of the following conditions holds.*

(i) *One of the* S_i*s is bounded.*

(ii) *All of the* S_i*s are closed affine half-spaces.*

Then $G \neq \emptyset$*.*

Next, we need a point of passage from $\boldsymbol{\Xi}$ to Ξ.

Proposition 5.4 [132] *We have*

$$\begin{cases} (\forall\, \mathbf{a} \in \mathbf{D}) \;\; P_{\mathbf{S}}(\mathbf{a}) = (P_i(a))_{i\in I} \\ (\forall\, \mathbf{a} \in \mathbf{S}) \;\; P_{\mathbf{D}}(\mathbf{a}) = (\sum_{i\in I} w_i a^{(i)}, \cdots, \sum_{i\in I} w_i a^{(i)}). \end{cases} \tag{5.21}$$

It follows from this proposition that

$$(\forall \mathbf{a} \in \mathbf{D}) \;\; P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}) = (\sum_{i\in I} w_i P_i(a), \cdots, \sum_{i\in I} w_i P_i(a)). \tag{5.22}$$

Therefore the alternating projection method (5.8) in $\boldsymbol{\Xi}$ yields the simultaneous projection method

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda_n(\sum_{i\in I} w_i P_i(a_n) - a_n), \tag{5.23}$$

in Ξ. We shall call the algorithm (5.23) with relaxation scheme (5.3) the parallel projection method (PPM). 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 of 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 $d_n = \sum_{i\in I} w_i P_i(a_n)$. The last phase consists of positioning the new iterate a_{n+1} on the segment between a_n and $2d_n - a_n$. This procedure is illustrated in Fig. 10. The weak convergence of PPM is a direct consequence of Theorem 5.2 and Proposition 2.12.

Theorem 5.4 *Suppose that $G \neq \emptyset$ (see Proposition 5.3). Then every orbit of PPM converges weakly to a point in G.*

Special cases of PPM have already been studied in the literature via direct approaches in the original space. Thus, Theorem 5.4 generalizes a result of [54], 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 [55], which assumed constant relaxations in (5.23). The following proposition is a consequence of (2.40) and Propositions 5.2 and 5.4.

Proposition 5.5 *Let $(a_n)_{n\geq 0}$ be any orbit of PPM. Then $(\Phi(a_n))_{n\geq 0}$ decreases until convergence and*

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n - \lambda_n \nabla\Phi(a_n), \tag{5.24}$$

where ∇ is the gradient operator in Ξ. Moreover, at iteration n, the relaxation parameter that will bring a_{n+1} closest to a solution point a^* in G is

$$\lambda_n^* = \frac{\langle \sum_{i\in I} w_i P_i(a_n) - a_n \mid a^* - a_n \rangle}{\| \sum_{i\in I} w_i P_i(a_n) - a_n \|^2} \geq 1. \tag{5.25}$$

Although the product space formalism is well suited to analyze and develop projection methods, it is sometimes limited when it comes to strong convergence properties, as it imposes conditions on the whole set $\mathbf{S}$. For instance compactness of $\mathbf{S}$ guarantees strong convergence of (5.8) in $\boldsymbol{\Xi}$ but it translates into compactness of all the sets $(S_i)_{i\in I}$ in Ξ. As we shall now see, much less restrictive conditions can be obtained via a direct approach in Ξ.

Theorem 5.5 [46] *Every orbit of PPM converges strongly to a point in G if any of the following conditions is satisfied.*

(i) *$(S_i)_{i\in I}$ contains only closed affine half-spaces.*

(ii) *$(S_i)_{i\in I}$ contains only uniformly convex sets.*

(iii) *$(S_i)_{i\in I}$ contains a boundedly compact set and a bounded set.*

An alternative strong convergence result which does not place any restriction on the sets is the following.

Theorem 5.6 *Suppose that $G \neq \emptyset$ (see Proposition 5.3). Then, for any a_0 in Ξ, every sequence of iterates $(a_n)_{n\geq 0}$ defined by*

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = (1-\alpha_n)a_0 + \alpha_n(\lambda \sum_{i\in I} w_i P_i(a_n) + (1-\lambda)a_n), \tag{5.26}$$

where $(\alpha_n)_{n\geq 0}$ is as in (5.19) and $0 < \lambda \leq 2$, converges strongly to the projection of a_0 onto G.

Proof. Thanks to (5.22), (5.18) in $\boldsymbol{\Xi}$ yields (5.26) in Ξ. It then follows from Proposition 2.12 that Theorem 5.6 is a corollary of Theorem 5.3. □

It is worth noting that (5.26) 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. Even in consistent problems, this property is very valuable in certain image recovery applications, when one seeks the best feasible approximation of a reference image a_0 [39] (in

comparison, the method developed in [105] is limited to the case $m = 2$ and is relatively involved).[5] As an example, one who adopts the aim of finding a least-squares-feasible image with minimum energy can take a_0 to be the zero image. It then follows from Theorem 5.6 that the iterations

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = \frac{n}{n+1}(\lambda \sum_{i \in I} w_i P_i(a_n) + (1-\lambda)a_n), \tag{5.27}$$

will converge strongly to the desired solution.

5.4 Projection Methods

5.4.1 Panorama

Although POCS has been the focus of most of the attention in image recovery, other projection methods have been available, some for almost three decades, that overcome some of its shortcomings. We discuss here three frameworks that, in our opinion, contain interesting features.

5.4.1.1 Framework 1: Browder's Admissible Control

In POCS the control sequence $(i(n))_{n \geq 0}$ imposes that the sets be activated in periodic order. As mentioned in Section 5.2.4, this periodic control mode can be implemented only when card $I < +\infty$. An alternative way of defining the control sequence $(i(n))_{n \geq 0}$ is to require that each set S_i be activated at least once within any cycle of M_i consecutive iterations, that is

$$(\forall i \in I)(\exists M_i \in \mathbb{N}^*)(\forall n \in \mathbb{N}) \;\; i \in \{i(n), \cdots, i(n + M_i - 1)\}. \tag{5.28}$$

[5] In consistent problems (i.e, $G = S$), a slight extension of a result of [112] shows that (5.26) can be replaced by

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = (1-\alpha_n)a_0 + \alpha_n P_1 \circ \cdots \circ P_m(a_n),$$

in Theorem 5.6. In addition, for both methods, strong convergence to the projection of a_0 onto S remains true if each P_i is replaced by any firmly nonexpansive operator T_i such that $\operatorname{Fix} T_i = S_i$ [43], [112].

For $I = \mathbb{N}^*$ and $M_i = 2^i$, an example of admissible control sequence is

$$\begin{aligned}(i(n))_{n\geq 0} = (\ & 1,2,1,3,1,2,1,4,1,2,1,3,1,2,1,5,1,2,1,3,1,2,1,4,\\ & 1,2,1,3,1,2,1,6,1,2,1,3,1,2,1,4,1,2,1,3,1,2,1,5,\\ & 1,2,1,3,1,2,1,4,1,2,1,3,1,2,1,7,1,2,1,3,1,2,1,4,\\ & 1,2,1,3,1,2,1,5,1,2,1,3,1,2,1,4,1,2,1,3,1,2,1,6,\\ & 1,2,1,3,1,2,1,4,1,2,1,3,1,2,\cdots).\end{aligned} \tag{5.29}$$

It is noted that periodic control is a particular case of admissible control. Hence, the following theorem due to Browder generalizes the weak convergence result of POCS found in Theorem 3.2.

Theorem 5.7 [19] *Suppose that I is any nonempty subset of $\mathbb{N}$. Then every sequence generated by the serial algorithm (5.1) with relaxation strategy (5.3) and admissible control scheme (5.28) converges weakly to a point in S.*

An even more general control scheme is the so-called chaotic control scheme, which imposes only that every set be used infinitely often, i.e.,

$$(\forall i \in I)(\forall n \in \mathbb{N})\ \ i \in \{i(n), i(n+1), \cdots\}. \tag{5.30}$$

This condition goes back to the work of Poincaré on boundary problems [134], who gave the following example for $I = \mathbb{N}^*$

$$(i(n))_{n\geq 0} = (1,2,1,2,3,1,2,3,4,1,2,3,4,5,1,2,3,4,5,6,\cdots). \tag{5.31}$$

However, the result of Theorem 5.7 no longer holds in this case (even in finite dimensional spaces [40]) and some restrictions are needed.

Theorem 5.8 *Every sequence generated by the unrelaxed version of the serial algorithm (5.1) under chaotic control converges weakly to a point in S if any of the following conditions holds.*

(i) *$(S_i)_{i\in I}$ is a finite family of closed vector subspaces* [4].

(ii) *$(S_i)_{i\in I}$ is a finite family containing a weak interior point* [61]*, i.e.,*

$$(\exists w \in S)(\forall c \in S)(\exists \rho \in \mathbb{R}_+^*)\ \ w + \rho(w - c) \in S. \tag{5.32}$$

(iii) card $I = 3$ [61].

A result similar to (ii) can also be found in [182]. If instead of merely a weak interior point, we require the existence of an interior point for S, then strong convergence takes place for countable families.

Theorem 5.9 *Suppose that I is any nonempty subset of $\mathbb{N}$ and that $\overset{\circ}{S} \neq \emptyset$. Then every sequence $(a_n)_{n\geq 0}$ generated by the serial algorithm (5.1) with relaxation strategy (5.3) and chaotic control scheme (5.30) converges strongly to a point in S.*

Proof. First of all, $(a_n)_{n\geq 0}$ is Fejér-monotone with respect to S. Indeed by fixing $c \in S$ and following a procedure similar to that of the proof of Theorem 5.2, we arrive at

$$\begin{aligned} (\forall n \in \mathbb{N}) \;\; ||a_{n+1} - c||^2 &\leq ||a_n - c||^2 - \varepsilon^2 ||P_{i(n)}(a_n) - a_n||^2 && (5.33) \\ &\leq ||a_n - c||^2. && (5.34) \end{aligned}$$

According to Proposition 2.11(iv), there exists a point $a \in \Xi$ such that $a_n \overset{n}{\to} a$, and it remains to show $a \in S$. Take an arbitrary $i \in I$. Since the control is chaotic, there exists an increasing sequence $(n_k)_{k\geq 0} \subset \mathbb{N}$ such that $(\forall k \in \mathbb{N}) \;\; i = i(n_k)$. Therefore (5.33) yields

$$\begin{aligned} (\forall k \in \mathbb{N}) \;\; ||P_i(a_{n_k}) - a_{n_k}||^2 &= ||P_{i(n_k)}(a_{n_k}) - a_{n_k}||^2 && (5.35) \\ &\leq \varepsilon^{-2} \left(||a_{n_k} - c||^2 - ||a_{n_k+1} - c||^2 \right). && (5.36) \end{aligned}$$

As in the proof of Theorem 5.2, we obtain $P_i(a_{n_k}) - a_{n_k} \overset{k}{\to} 0$. But since $a_{n_k} \overset{k}{\to} a$, we get $P_i(a_{n_k}) \overset{k}{\to} a$. However $(P_i(a_{n_k}))_{k\geq 0} \subset S_i$ and S_i is closed. Therefore $a \in S_i$. Since i was arbitrary, we conclude $a \in \bigcap_{i\in I} S_i = S$. □

5.4.1.2 Framework 2: Pierra's Extrapolated Iterations

We have seen in Sections 5.2.1 and 5.2.2 that POCS suffered from slow convergence and that it was not well suited to take advantage of parallel computing. It would be erroneous, however, to conclude that a parallel projection method is always faster than a serial one just because it can process projections simultaneously as opposed to sequentially. Thus, the parallel algorithm SIRT (3.11) was found to be actually slower than the serial algorithm ART (3.10) in tomographic image reconstruction [85]. In our numerical simulations, we have also found that (3.11) is usually slower than unrelaxed POCS (5.4) in a number of problems involving general convex constraints. This fact can be illustrated by comparing Fig. 6 and Fig. 11.

An advantage of a parallel projection method such as PPM (5.23) is that it can be accelerated by overrelaxations, which is not true for serial algorithms. In fact, overrelaxations

have been reported to accelerate parallel projection methods in a number of studies, e.g., [21], [46], [60], and [92]. To explain this, note that the efficient progression of a general relaxed algorithm of the type $(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n + \lambda_n(d_n - a_n)$ towards a solution depends on two factors at every iteration n:

1) Centering: in order to avoid "zigzagging", the iterations should remain centered with respect to the sets so that the directions taken by the algorithm keep pointing to the solution set S.

2) Relaxation: at every iteration, λ_n should place a_{n+1} close to S on the ray emanating from a_n and going through d_n.

In the case of a serial algorithm such as (5.1), d_n is the projection onto a single set $S_{i(n)}$ and therefore the algorithm will keep moving in different directions and will tend to zigzag. By contrast, since PPM averages the projections its centering is much better, which takes care of condition 1) above. On the other hand, Proposition 5.2 takes care of condition 2), as it indicates that overrelaxations will bring the update closer to S. In PPM, however, overrelaxations were limited to 2 in order to guarantee convergence in inconsistent problems. We shall now follow the work of Pierra [132], who showed that in consistent problems this condition can be bypassed and much larger relaxations can be obtained.

In order to define an alternative relaxation strategy, let us return to the product space formalism of Section 2.8, in which the convex feasibility problem was seen to reduce to (2.47). Now consider Fig. 12, where $\mathbf{a}_n \in \mathbf{D} \bigcap \complement\mathbf{S}$, $\mathbf{s}_n = P_{\mathbf{S}}(\mathbf{a}_n)$ and $\mathbf{d}_n = P_{\mathbf{D}}(\mathbf{s}_n) = P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n)$. Let $\mathbf{H}_n$ be the affine hyperplane supporting $\mathbf{S}$ at $\mathbf{s}_n$. Then $\mathbf{H}_n$ separates $\mathbf{a}_n$ from $\mathbf{S}$ and intersects $\mathbf{D}$ at a point $\mathbf{e}_n$. Note that

$$\frac{|||\mathbf{e}_n - \mathbf{a}_n|||}{|||\mathbf{s}_n - \mathbf{a}_n|||} = \frac{|||\mathbf{s}_n - \mathbf{a}_n|||}{|||\mathbf{d}_n - \mathbf{a}_n|||}. \tag{5.37}$$

Hence, returning to the alternating projection method (5.8), an update $\mathbf{a}_{n+1}$ on the segment between $\mathbf{a}_n$ and $\mathbf{e}_n$ will be obtained by taking relaxations up to

$$L_n = \frac{|||\mathbf{e}_n - \mathbf{a}_n|||}{|||\mathbf{d}_n - \mathbf{a}_n|||} = \frac{|||\mathbf{s}_n - \mathbf{a}_n|||^2}{|||\mathbf{d}_n - \mathbf{a}_n|||^2} = \frac{|||P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2}. \tag{5.38}$$

Note that we always have $L_n \geq 1$. Indeed, since $\mathbf{a}_n \in \mathbf{D}$, the nonexpansivity of $P_{\mathbf{D}}$ yields

$$\begin{aligned} |||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n||| &= |||P_{\mathbf{D}}(P_{\mathbf{S}}(\mathbf{a}_n)) - P_{\mathbf{D}}(\mathbf{a}_n)||| && (5.39) \\ &\leq |||P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||. && (5.40) \end{aligned}$$

Proposition 5.6 [132] *Every sequence* $(\mathbf{a}_n)_{n\geq 0} \subset \mathbf{D}$ *constructed as in (5.8) with relaxation strategy*

$$(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq L_n, \quad \textit{where} \;\; 0 < \varepsilon < 1, \tag{5.41}$$

converges weakly to a point in $\mathbf{S} \bigcap \mathbf{D}$.

We can recast this result in the original image space Ξ via Proposition 5.4 to obtain Pierra's extrapolated parallel projection method (EPPM), which is described by the algorithm (5.23) with relaxation range

$$(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq L_n \quad \text{where} \quad L_n = \begin{cases} \dfrac{\sum_{i\in I} w_i \|P_i(a_n) - a_n\|^2}{\|\sum_{i\in I} w_i P_i(a_n) - a_n\|^2} & \text{if } a_n \notin S \\ 1 & \text{otherwise.} \end{cases} \tag{5.42}$$

The weak convergence of PPM follows immediately from Propositions 5.6 and 2.12.

Theorem 5.10 [132] *Every orbit of EPPM converges weakly to a point in* S.

It was observed in [132] that the fast convergence of EPPM was due to the large overrelaxations allowed by (5.42). In fact, L_n can attain values much larger than 2 and eliminate the "angle problem" of conventional methods: one can see in Figs. 6 and 11 that the iterations will slow down as the angle between the two sets diminishes. On the other hand, EPPM is not sensitive to this problem, as seen in Fig. 12. Fig. 13 shows a realization of EPPM with equal weights on the projections and $(\forall n \in \mathbb{N}) \;\; \lambda_n = L_n$. It compares favorably with POCS (Fig. 6) and SIRT (Fig. 11).

In order to mitigate the possible zigzagging that could take place with large relaxations and could reduce the effectiveness of the algorithm, it was suggested in [132] to re-center the orbit every 3 iterations by halving the extrapolations, namely

$$(\forall n \in \mathbb{N}) \;\; \lambda_n = \begin{cases} L_n/2 & \text{if } n = 2 \text{ modulo } 3 \\ L_n & \text{otherwise.} \end{cases} \tag{5.43}$$

We saw in Proposition 5.5, that PPM was a steepest-descent method for the proximity function of (3.7). By invoking (2.16), EPPM can be written in the form

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n - \frac{\alpha_n \Phi(a_n)}{\|\nabla \Phi(a_n)\|^2} \nabla \Phi(a_n) \quad \text{with} \quad \varepsilon \leq \alpha_n \leq 2 - \varepsilon, \tag{5.44}$$

which was shown in [40] to be a particular case of the Gauss-Newton method studied in [136]. In that paper the Gauss-Newton approach was reported to converge more efficiently than the standard steepest-descent approach (5.24). This furnishes another justification for the superiority of EPPM over PPM.[6]

[6] This statement applies only to consistent problems. PPM was developed for inconsistent problems, where relaxations cannot be extended beyond 2 (see Fig. 9).

It should be noted that EPPM does not generalize PPM for the relaxation ranges (5.42) are not necessarily wider than (5.3). Indeed, we have seen that the extrapolation parameter L_n was at least equal to 1 but it may not necessarily be greater than 2.[7] Therefore to unify and extend both PPM and EPPM, we shall now consider relaxations up to $2L_n$, i.e.

$$(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq (2-\varepsilon)L_n. \tag{5.45}$$

This extension will also allow faster convergence in certain problems through the use of larger overrelaxations than those allowed by PPM and EPPM. To justify this extension more rigorously, let us go back to Proposition 5.2. In the consistent case, it states that at iteration n the relaxation parameter which brings $\mathbf{a}_{n+1}$ closest to an arbitrary point $\mathbf{a}^*$ in $\mathbf{S} \bigcap \mathbf{D}$ is

$$\lambda_n^* = \frac{\langle\langle P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n \mid \mathbf{a}^* - \mathbf{a}_n \rangle\rangle}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2}. \tag{5.46}$$

Using the linearity of $P_{\mathbf{D}}$, (2.21), and (2.19), we obtain

$$\begin{aligned}
\lambda_n^* &= \frac{\langle\langle P_{\mathbf{D}}(P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n) \mid \mathbf{a}^* - \mathbf{a}_n \rangle\rangle}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2} && (5.47)\\
&= \frac{\langle\langle P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n \mid \mathbf{a}^* - \mathbf{a}_n \rangle\rangle}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2} && (5.48)\\
&= \frac{|||P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2} + \frac{\langle\langle P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n \mid \mathbf{a}^* - P_{\mathbf{S}}(\mathbf{a}_n) \rangle\rangle}{|||P_{\mathbf{D}} \circ P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n|||^2} && (5.49)\\
&\geq L_n. && (5.50)
\end{aligned}$$

Thus, extending the relaxation range to $]0, 2L_n[$ opens the possibility of getting closer to the optimal relaxation parameter, which lies in $[L_n, +\infty[$. Let us also note that, according to (2.21), $\langle\langle P_{\mathbf{S}}(\mathbf{a}_n) - \mathbf{a}_n \mid \mathbf{a}^* - P_{\mathbf{S}}(\mathbf{a}_n) \rangle\rangle = 0$ if $\mathbf{S}$ is a closed affine subspace. Whence, (5.49) shows that $\lambda_n^* = L_n$ in this case. These results can be routinely transferred to the original space Ξ as follows.

Proposition 5.7 *At iteration n, the relaxation parameter that will bring a_{n+1} closest to a solution point a^* in S is*

$$\lambda_n^* = \frac{\langle \sum_{i \in I} w_i P_i(a_n) - a_n \mid a^* - a_n \rangle}{\| \sum_{i \in I} w_i P_i(a_n) - a_n \|^2} \geq L_n. \tag{5.51}$$

In addition, if $(S_i)_{i \in I}$ is a finite family of closed affine subspaces, then

$$\lambda_n^* = L_n. \tag{5.52}$$

We shall call EPPM2 the algorithm obtained by combining (5.23) and (5.45).

Theorem 5.11 [41] *Every orbit of EPPM2 converges weakly to a point in S.*

[7] Cases when $L_n \leq 2$ can easily be constructed, e.g., [22].

5.4.1.3 Framework 3: Block-Parallel Methods

A limitation of parallel methods such as EPPM2 is that all the sets must be acted upon at each iteration. If the number of sets is larger than the number of concurrent processors available, the implementation of the algorithm will not be fully parallel. At iteration n, a flexible adaptation of the computational load to the parallel computing architecture at hand can be obtained by activating only a subfamily $(S_i)_{i\in I_n\subset I}$ of property sets. If, in addition, we allow the weights on the projections to vary at each iteration, we obtain an iterative method of the form

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n + \lambda_n(\sum_{i\in I_n} w_{i,n} P_i(a_n) - a_n), \tag{5.53}$$

where $(I_n)_{n\geq 0}$ is a sequence of subsets of I and $((w_{i,n})_{i\in I_n})_{n\geq 0}$ a sequence of convex weights. Let us observe that if only one set is processed at each iteration, say $(\forall n \in \mathbb{N}) \quad I_n = \{i(n)\}$, then (5.53) reverts to the serial method (5.1). On the other hand, if all the sets are processed at each iteration, i.e., $(\forall n \in \mathbb{N}) \quad I_n = I$, we obtain the simultaneous projection method (5.23). Algorithms of the general form (5.53) have been proposed with various assumptions on the dimension of Ξ, the sets $(S_i)_{i\in I}$, the relaxation parameters $(\lambda_n)_{n\geq 0}$, the weights $((w_{i,n})_{i\in I_n})_{n\geq 0}$, and the control sequence $(I_n)_{n\geq 0}$ [3], [21], [22], [26], [46], and [126]. In the next section we present a projection method based on (5.53) which encompasses and generalizes these approaches.

5.4.2 Extrapolated Method of Parallel Projections (EMOPP)

Each of the three frameworks discussed in the previous section has an attractive feature. Framework 1 provides flexible control schemes that can handle an infinite number of sets; Framework 2 provides extrapolated iterations that converge efficiently; Framework 3 provides a flexible management of the property sets that can easily be adapted to the configuration of a parallel computer. These attractive features can be combined into a single algorithm, the extrapolated method of parallel projections (EMOPP), which we now describe.

Given an initial point $a_0 \in \Xi$ and numbers $C \in \mathbb{N}^*$, $\delta \in]0, 1/C[$, and $\varepsilon \in]0, 1[$, EMOPP is defined by the iterative process

$$(\forall n \in \mathbb{N}) \quad a_{n+1} = a_n + \lambda_n(\sum_{i\in I_n} w_{i,n} P_i(a_n) - a_n), \tag{5.54}$$

where at each iteration n:

a) The family I_n of indices of selected sets satisfies

$$\emptyset \neq I_n \subset I \quad \text{and} \quad \text{card}\,\{i \in I_n \mid a_n \notin S_i\} \leq C. \tag{5.55}$$

b) The weights on the projections satisfy

$$\sum_{i\in I_n} w_{i,n} = 1 \quad \text{and} \quad (\forall i \in I_n)\ \ w_{i,n} \geq \delta 1_{\complement S_i}(a_n). \tag{5.56}$$

c) The relaxation parameter λ_n lies in $[\varepsilon, (2-\varepsilon)L_n]$, where

$$L_n = \begin{cases} \dfrac{\sum_{i\in I_n} w_{i,n}\|P_i(a_n) - a_n\|^2}{\|\sum_{i\in I_n} w_{i,n}P_i(a_n) - a_n\|^2} & \text{if } \ a_n \notin \bigcap_{i\in I_n} S_i \\ 1 & \text{otherwise.} \end{cases} \tag{5.57}$$

Practically, iteration n of EMOPP is performed as follows. First, one selects the sets to be activated; I_n contains the indices of these sets. One then computes the projections $(P_i(a_n))_{i\in I_n}$ of the current iterate a_n onto the selected sets and determines a convex combination $d_n = \sum_{i\in I_n} w_{i,n}P_i(a_n)$ of these projections as well as the extrapolation parameter L_n. The position of the new iterate a_{n+1} on the segment between a_n and $a_n + 2L_n(d_n - a_n)$ is determined by the relaxation parameter λ_n.

EMOPP features doubly extrapolated relaxation ranges as EPPM2, it can process variable blocks of sets as (5.53), and it can be driven by flexible control schemes that extend in particular the admissible and chaotic control schemes of the serial algorithm (5.1). The condition (5.56) imposes that the weights be bounded away from 0 on violated sets and add up to 1. This, in turn, implies that the number of violated sets processed at each iteration must be bounded, whence condition (5.55). It can always be assumed that nonviolated sets are selected since they can be assigned a 0 weight. Finally, let us observe that $L_n \geq 1$ (since the function $\|\cdot\|^2$ is convex) and that $L_n = 1$ when less than two violated sets are used ($\text{card}\,\{i \in I_n \mid a_n \notin S_i\} < 2$). In this case, the relaxation range reduces to the usual interval $[\varepsilon, 2-\varepsilon]$ and no extrapolation takes place.

5.4.3 Control

We shall consider the following control strategies for EMOPP. They constitute extensions to parallel projection methods of schemes which have been proposed for serial ones.

We shall say that the control is:

(i) Static if all the sets are activated at each iteration, i.e.,

$$(\forall n \in \mathbb{N})\ \ I_n = I. \tag{5.58}$$

This control condition goes back to Cimmino's algorithm [35]. It was used in SIRT, PPM, EPPM, and EPPM2.

(ii) Cyclic if there exists a positive integer M such that

$$(\forall n \in \mathbb{N}) \;\; I = \bigcup_{k=n}^{n+M-1} I_k. \tag{5.59}$$

Thus, if the control is M-cyclic, all the sets must be activated at least once within any M consecutive iterations. This condition was utilized in [46] and [126].

(iii) Quasi-cyclic if there exists an increasing sequence $(M_m)_{m\geq 0} \subset \mathbb{N}$ such that

$$\begin{cases} M_0 = 0 \\ \sum_{m\geq 0} (M_{m+1} - M_m)^{-1} = +\infty \\ (\forall m \in \mathbb{N}) \;\; I = \bigcup_{k=M_m}^{M_{m+1}-1} I_k. \end{cases} \tag{5.60}$$

In words, if the control is $(M_m)_{m\geq 0}$-quasi-cyclic, all the sets are activated at least once within each quasi-cycle of iterations $\{M_m, \cdots, M_{m+1} - 1\}$. The nonsummability condition imposes that the lengths $(M_{m+1} - M_m)_{m\geq 0}$ of the quasi-cycles do not increase too fast eventually. For instance the linear growth condition $(\forall m \in \mathbb{N}) \;\; M_{m+1} - M_m = \alpha m + 1$ is acceptable. Quasi-cyclic control was introduced in [175] for a serial method.

(iv) Admissible if there exist positive integers $(M_i)_{i\in I}$ such that

$$(\forall (i,n) \in I \times \mathbb{N}) \;\; i \in \bigcup_{k=n}^{n+M_i-1} I_k \tag{5.61}$$

Hence, the set S_i is activated at least once within any M_i consecutive iterations, which extends Browder's definition (5.28) to the parallel case. Of course, if card $I < +\infty$, this control mode coincides with the cyclic mode (5.59) for $M = \max_{i\in I} M_i$.

(v) Chaotic if each set is activated infinitely often in the iteration process, i.e.

$$(\forall n \in \mathbb{N}) \;\; I = \bigcup_{k\geq n} I_k. \tag{5.62}$$

This is a direct generalization of (5.30), which was used in the parallel method of [126]. Clearly, static $\Rightarrow$ cyclic $\Rightarrow$ quasi-cyclic $\Rightarrow$ chaotic, and cyclic $\Rightarrow$ admissible $\Rightarrow$ chaotic.

(vi) Coercive if

$$(\exists (i(n))_{n\geq 0} \in \bigtimes_{n\geq 0} I_n) \;\; d(a_n, S_{i(n)}) \overset{n}{\to} 0 \Rightarrow \sup_{i\in I} d(a_n, S_i) \overset{n}{\to} 0. \tag{5.63}$$

In the serial case, this control mode was proposed in [79] as a generalization of the most-remote set control scheme

$$(\forall n \in \mathbb{N})(\exists i(n) \in I_n) \;\; d(a_n, S_{i(n)}) = \sup_{i \in I} d(a_n, S_i), \tag{5.64}$$

which is not always applicable when card $I = +\infty$.

(vii) Chaotically coercive if $(I_n)_{n \geq 0}$ contains a subsequence $(I_{n_k})_{k \geq 0}$ such that

$$(\exists (i(k))_{k \geq 0} \in \bigtimes_{k \geq 0} I_{n_k}) \;\; d(a_{n_k}, S_{i(k)}) \stackrel{k}{\to} 0 \;\Rightarrow\; \sup_{i \in I} d(a_{n_k}, S_i) \stackrel{k}{\to} 0. \tag{5.65}$$

This condition generalizes (5.63) as well as the control strategy consisting in activating one of the most remote sets infinitely often in the course of the iterations.

5.4.4 Convergence Results

In this section we present results on the convergence of EMOPP. As usual, a key step in proving the convergence to a feasible image is (2.37).

Proposition 5.8 [40] *Every orbit of EMOPP is Fejér-monotone with respect to S.*

The next step is then to determine suitable conditions on the control and the sets so that weak or strong convergence to a point in S is actually achieved by every orbit. Let us note that, since the sequence $(\text{card}\,\{i \in I_n \mid a_n \notin S_i\})_{n \geq 0}$ is bounded, quasi-cyclic control cannot be applied to countable set theoretic formulations in general as it requires that all the sets be activated over a finite number of iterations.

5.4.4.1 Weak Convergence

The following theorem, which generalizes results of [46] as well as Theorems 5.7 and 5.11, appears to be the most general result available on the weak convergence of projection methods.

Theorem 5.12 [40] *Under coercive or admissible control, every orbit of EMOPP converges weakly to a point in S.*

The next theorem does not guarantee weak convergence as weak cluster points could exist outside of S but it is nonetheless of interest.

Theorem 5.13 [40] *Each orbit of EMOPP possesses one and only one weak cluster point in S if either of the following conditions holds.*

(i) *The control is chaotically coercive.*

(ii) card $I < +\infty$ *and the control is quasi-cyclic.*

In the special case of algorithm (5.1), Theorem 5.13(ii) was obtained in [174].

5.4.4.2 Strong Convergence

Following the terminology of [10], $(S_i)_{i\in I}$ is boundedly regular if for any bounded sequence $(a_n)_{n\geq 0}$ we have

$$\sup_{i\in I} d(a_n, S_i) \stackrel{n}{\to} 0 \Rightarrow d(a_n, S) \stackrel{n}{\to} 0. \tag{5.66}$$

The concept of bounded regularity was first used extensively in [79] to prove the strong convergence of several serial projections algorithms. Conditions for bounded regularity were previously discussed in [110] in the case of two sets. The importance of this notion stems from the following fact.

Proposition 5.9 [79] *Let $(a_n)_{n\geq 0}$ be a Fejér-monotone sequence with respect to S and suppose that $(S_i)_{i\in I}$ is boundedly regular. Then*

$$\sup_{i\in I} d(a_n, S_i) \stackrel{n}{\to} 0 \Rightarrow (\exists\, a \in S)\ a_n \stackrel{n}{\to} a. \tag{5.67}$$

Since under chaotically coercive or quasi-cyclic control each orbit $(a_n)_{n\geq 0}$ contains a suborbit $(a_{n_k})_{k\geq 0}$ such that $\sup_{i\in I} d(a_{n_k}, S_i) \stackrel{k}{\to} 0$ [40], Propositions 5.8, 5.9, and 2.11(iii) lead to the following result, which generalizes results of [46] and [132], as well as Theorem 3.2.

Theorem 5.14 [40] *Suppose that $(S_i)_{i\in I}$ is boundedly regular. Then every orbit of EMOPP converges strongly to a point in S if either of the following conditions holds.*

(i) *The control is chaotically coercive.*

(ii) card $I < +\infty$ *and the control is quasi-cyclic.*

We now give specific conditions when this theorem can be applied. The following definition is motivated by [110] (see also Proposition 2.4(ix)). We shall say that a set S_i is a Levitin-Polyak set if, for every sequence $(a_n)_{n\geq 0} \subset \Xi$ such that $d(a_n, S_i) \overset{n}{\to} 0$, the following property holds: if $a_n \overset{n}{\underset{S_i}{\to}} a$ and $a \in \partial S_i$, then $a_n \overset{n}{\to} a$. Locally uniformly convex sets, and *a fortiori* uniformly convex sets, are Levitin-Polyak sets. However, unlike uniformly convex sets, locally uniformly convex sets need not be bounded [110].[8]

Proposition 5.10 [11], [40], [79] *$(S_i)_{i\in I}$ is boundedly regular if any of the following conditions is satisfied.*

(i) $(\exists j \in I)\ \ S_j \bigcap (\bigcap_{i\in I\setminus\{j\}} S_i)^\circ \neq \emptyset$.

(ii) *All, except possibly one, of the sets in $(S_i)_{i\in I}$ are f-uniformly convex.*

(iii) *One of the sets in $(S_i)_{i\in I}$ is boundedly compact. In particular:*

- *One of the sets in $(S_i)_{i\in I}$ is compact.*
- *One of the sets in $(S_i)_{i\in I}$ is contained in a finite dimensional affine subspace.*
- $\dim \Xi < +\infty$.

(iv) *$(S_i)_{i\in I}$ is a finite family and all, except possibly one, of its sets are Levitin-Polyak sets. In particular:*

- *$(S_i)_{i\in I}$ is a finite family and all, except possibly one, of its sets are locally uniformly convex.*
- *$(S_i)_{i\in I}$ is a finite family and all, except possibly one, of its sets are uniformly convex.*

(v) *$(S_i)_{i\in I}$ is a finite family of closed affine subspaces such that $\sum_{i\in I} S_i^\perp$ is closed. In particular:*

- *$(S_i)_{i\in I}$ is a finite family of closed affine subspaces, all of which, except possibly one, have finite codimension.*
- *$(S_i)_{i\in I}$ is a finite family of closed affine subspaces, all of which, except possibly one, are affine hyperplanes.*

[8] The fact that a locally uniformly convex set S_i is a Levitin-Polyak set can be proved as follows. Take a sequence $(a_n)_{n\geq 0}$ such that $d(a_n, S_i) \overset{n}{\to} 0$ and $a_n \overset{n}{\underset{S_i}{\to}} a \in \partial S_i$. Then $P_i(a_n) \overset{n}{\underset{S_i}{\to}} a$. Now take a point $b \notin S_i$ such that $P_i(b) = a$ and consider the half-space $\{h \in \Xi \mid \langle h - a \mid b - a\rangle \leq 0\}$ containing S_i and whose boundary supports S_i at a. Since S_i is locally uniformly convex, there exists a nondecreasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ that vanishes only at 0 such that $(\forall h \in S_i)\ \ \langle h - a \mid b - a\rangle \leq -f(\|h - a\|)$. Whence, $(\forall n \in \mathbb{N})\ \ \langle P_i(a_n) - a \mid a - b\rangle \geq f(\|P_i(a_n) - a\|)$. But since $P_i(a_n) \overset{n}{\underset{S_i}{\to}} a$, we get $f(\|P_i(a_n) - a\|) \overset{n}{\to} 0$ and therefore $P_i(a_n) \overset{n}{\to} a$. As $d(a_n, S_i) \overset{n}{\to} 0$, we conclude $a_n \overset{n}{\to} a$.

We also note that (ii) in Theorem 5.5 can be generalized to: $(S_i)_{i\in I}$ contains only Levitin-Polyak sets, one of which is bounded.

(vi) $(S_i)_{i\in I}$ *is a finite family of closed polyhedrons (finite intersections of closed affine half-spaces).*

We now move to the most general type of control, namely chaotic control. To obtain strong convergence in that case, the hypotheses on the sets will have to be strengthened.

A point $c \in S$ is a strongly regular point of $(S_i)_{i\in I}$ if [126]

$$(\forall(\rho_1,\rho_2)\in\mathbb{R}_+^{*2})(\exists\rho\in\mathbb{R}_+)(\forall(i,a,b)\in I\times\Xi\times\Xi)$$
$$\begin{cases} \|P_i(a)-c\|\geq\rho_1 \\ \|b-c\|\leq\rho_2 \end{cases} \Rightarrow\ d(b,H_i(a))\leq\rho d(c,H_i(a)), \tag{5.68}$$

where $H_i(a)=\{h\in\Xi \mid \langle h-P_i(a)\mid a-P_i(a)\rangle=0\}$.

Theorem 5.15 [40] *Under chaotic control, every orbit of EMOPP converges strongly to a point in S if any of the following conditions is satisfied.*

(i) $(S_i)_{i\in I}$ *has a strongly regular point. In particular:*

- $\overset{\circ}{S}\neq\emptyset$.
- $(S_i)_{i\in I}$ *is a family of f-uniformly convex sets.*

(ii) $(S_i)_{i\in I}$ *is a finite family and one of its sets is boundedly compact. In particular:*

- $(S_i)_{i\in I}$ *is a finite family and one of its sets is compact.*
- $(S_i)_{i\in I}$ *is a finite family and one of its sets is contained in a finite dimensional affine subspace.*
- $(S_i)_{i\in I}$ *is a finite family and* $\dim\Xi<+\infty$.

(iii) $(S_i)_{i\in I}$ *is a finite family of closed affine subspaces with finite codimensions. In particular,* $(S_i)_{i\in I}$ *is a finite family of affine hyperplanes.*

(iv) $(S_i)_{i\in I}$ *is a finite family of closed polyhedrons. In particular,* $(S_i)_{i\in I}$ *is a finite family of closed affine half-spaces.*

For relaxations only up to L_n, Theorem 5.15(i) was established in [126]; in the special case of the unrelaxed version of algorithm (5.1)+(5.30), Theorem 5.15(ii) was proved in [20] with the compactness condition. A related result is the following.

Theorem 5.16 [10] *Suppose that $(S_i)_{i\in I}$ is a finite family and that any of its nonvoid subfamilies $(S_i)_{i\in J\subset I}$ is boundedly regular, in particular:*

(i) $(\exists j \in I)\ \ S_j \bigcap (\bigcap_{i \in I \smallsetminus \{j\}} \overset{\circ}{S}_i) \neq \emptyset$.

(ii) *All, except possibly one, of the sets in* $(S_i)_{i \in I}$ *are boundedly compact.*

(iii) *Each set in* $(S_i)_{i \in I}$ *is a closed affine subspace and* $\sum_{i \in J} S_i^{\perp}$ *is closed for every* $\emptyset \neq J \subset I$.

(iv) $(S_i)_{i \in I}$ *is a family of closed polyhedrons.*

Then every sequence generated by the unrelaxed version of the serial algorithm (5.1) under chaotic control (5.30) converges strongly to a point in S.

5.5 Extrapolated Method of Parallel Approximate Projections (EMOPAP)

5.5.1 Problem Statement

In Section 2.5.2 we have given examples of sets whose projection operators admit closed-form expressions. There are many cases, however, when projection operators are not so easy to determine, which constitutes a serious obstacle in the implementation of a projection algorithm. To illustrate this point, consider the problem of projecting a digital image a onto the set

$$S_i = \{b \in \mathbb{E}^{N^2} \mid g_i(b) \leq 0\}, \tag{5.69}$$

where g_i is a convex functional (in digital image processing, this is typically how sets are specified). The projection $P_i(a)$ is obtained by solving the constrained quadratic minimization problem

$$\min \frac{1}{2}\|b - a\|^2 \qquad \text{subject to} \qquad g_i(b) = 0, \tag{5.70}$$

which can be recast as the problem of minimizing

$$\Theta(b) = \frac{1}{2}\|b - a\|^2 + \mu g_i(b), \tag{5.71}$$

where μ is a Lagrange multiplier to be adjusted so that $g_i(P_i(a)) = 0$. Assuming that g_i is differentiable, $P_i(a)$ should therefore satisfy

$$\begin{cases} P_i(a) = a + \mu \nabla g_i(P_i(a)) \\ g_i(P_i(a)) = 0. \end{cases} \tag{5.72}$$

If g_i is an affine or quadratic functional, as in the examples (2.3)-(2.6), this system is easily solved. Otherwise, it may require a costly solution method to adjust μ iteratively. For instance, consider the second moment set

$$S_i = \{a \in \mathbb{E}^{N^2} \mid \|x - Ta\|^2 \leq \zeta_2\} \tag{5.73}$$

of (4.41), where T is an $N^2 \times N^2$ matrix.[9] In this case, (5.72) was solved in [171] via a Newton method initialized at $\mu_0 = 0$. Another example is the minimum entropy set proposed in [47] for images that exhibit a low level of structure. If we denote by $\ln(a)$ the vector ${}^t[\ln(a^{(i)})]_{0 \leq i \leq N^2-1}$, this set takes the form[10]

$$S_i = \{a \in \Delta \mid -\langle a \mid \ln(a) \rangle \geq \eta\}, \tag{5.74}$$

where $\Delta = \{a \in \mathbb{E}^{N^2} \mid \sum_{i=0}^{N^2-1} a^{(i)} = 1$ and $(\forall i \in \{0, \cdots, N^2 - 1\})\ a^{(i)} \geq \tau > 0\}$, τ being a lower bound on the pixel values. The closedness and convexity of this set follow from the convexity of the functional $a \mapsto \langle a \mid \ln(a) \rangle$ on Δ (see Proposition 2.6). There too, (5.72) must be solved via iterative methods similar to those proposed in [70] and [167] for the maximum entropy method.

A way to circumvent the sometimes tedious computation of projections is to replace them by approximate ones. By an approximate projection of $a_n \in \complement S_i$ onto S_i, we shall mean the projection of a_n onto any closed and convex superset $S_{i,n}$ of S_i which does not contain a_n.[11] A natural candidate for $S_{i,n}$ is a closed affine half-space whose boundary hyperplane $H_{i,n}$ separates a_n from S_i (see Fig. 14). The approximate projection is then simply given by (2.22), meaning that the nonlinear constraint defining S_i has been "affinized". More formally, we shall say that $((S_{i,n})_{i \in I_n})_{n \geq 0}$ are approximating sets if they are closed and convex and satisfy

$$(\exists \eta \in]0,1[)(\forall n \in \mathbb{N})(\forall i \in I_n)\ \ S_i \subset S_{i,n} \quad \text{and} \quad d(a_n, S_{i,n}) \geq \eta d(a_n, S_i). \tag{5.75}$$

This condition has been used in [2], [11], [45], and [68].

5.5.2 Algorithm

Given an initial point $a_0 \in \Xi$ and numbers $C \in \mathbb{N}^*$, $\delta \in]0, 1/C[$, $\eta \in]0,1[$, and $\varepsilon \in]0,1[$, EMOPAP is defined by the iterative process [40]

$$(\forall n \in \mathbb{N})\ \ a_{n+1} = a_n + \lambda_n(\sum_{i \in I_n} w_{i,n} P_{i,n}(a_n) - a_n), \tag{5.76}$$

[9] Note that the set (5.73) has the same analytic expression as (4.6).

[10] In this context, each gray level $a^{(i)}$ is viewed as a probability and $\sum_{i=0}^{N^2-1} a^{(i)} = 1$.

[11] The idea of replacing exact projections by approximate ones was actually suggested in [132]. Naturally, this approach will be numerically advantageous if the determination of this superset is less costly than the computation of the exact projection.

where at each iteration n:

a) The family I_n of indices of selected sets satisfies

$$\emptyset \neq I_n \subset I \quad \text{and} \quad \text{card}\,\{i \in I_n \mid a_n \notin S_i\} \leq C. \tag{5.77}$$

b) $(P_{i,n}(a_n))_{i\in I_n}$ are the projections of a_n onto the approximating sets $(S_{i,n})_{i\in I_n}$ defined by (5.75).

c) The weights $(w_{i,n})_{i\in I_n}$ conform to (5.56).

d) The relaxation parameter λ_n lies in $[\varepsilon, (2-\varepsilon)L_n]$, where

$$L_n = \begin{cases} \dfrac{\sum_{i\in I_n} w_{i,n}\|P_{i,n}(a_n) - a_n\|^2}{\|\sum_{i\in I_n} w_{i,n}P_{i,n}(a_n) - a_n\|^2} & \text{if } a_n \notin \bigcap_{i\in I_n} S_i \\ 1 & \text{otherwise.} \end{cases} \tag{5.78}$$

Methods involving projections onto separating hyperplanes have been proposed previously for less general projection algorithms in [2] and [68].

5.5.3 Convergence Results

Theorem 5.17 [40] *Theorems 5.12, 5.13, and 5.14 remain true for EMOPAP. In addition, under chaotic control, every orbit of EMOPAP converges strongly to a point in S if any of the following conditions holds.*

(i) $\overset{\circ}{S} \neq \emptyset$.

(ii) $(S_i)_{i\in I}$ *is a finite family and one of its sets is boundedly compact. In particular:*

- $(S_i)_{i\in I}$ *is a finite family and one of its sets is compact.*
- $(S_i)_{i\in I}$ *is a finite family and one of its sets is contained in a finite dimensional affine subspace.*
- $(S_i)_{i\in I}$ *is a finite family and* $\dim \Xi < +\infty$.

(iii) $(S_i)_{i\in I}$ *is a finite family of closed affine subspaces with finite codimensions. In particular,* $(S_i)_{i\in I}$ *is a finite family of affine hyperplanes.*

(iv) $(S_i)_{i\in I}$ *is a finite family of closed affine half-spaces.*

In the finite dimensional case, part (ii) of this theorem was established in [45] and generalizes results of [2] and [68] which considered $(\lambda_n)_{n\geq 0} \subset [\varepsilon, 2-\varepsilon]$.

5.6 Extrapolated Method of Parallel Subgradient Projections (EMOPSP)

5.6.1 Problem Statement

The previous framework gives of lot of latitude in the choice of the approximating supersets. In practice, however, it is often convenient to have at hand a systematic method for determining the separating hyperplanes $((H_{i,n})_{i\in I_n})_{n\geq 0}$ in Fig 14.[12] In this section, we follow [41] and define one such method.

First of all, let us observe that a closed and convex property set S_i can always be expressed as the 0-section

$$S_i = \{a \in \Xi \mid g_i(a) \leq 0\} \tag{5.79}$$

of a convex, (lower semi-)continuous functional $g_i : \Xi \to \mathbb{R}$. This representation is quite general since (3.18) indicates that one can always take $g_i = d(\cdot, S_i)$. More practically, let us note that a convex constraint Ψ_i is usually formulated through a convex inequality, which leads directly to (5.79). Now suppose $a_n \in \complement S_i$. Then the closed affine half-space

$$S_{i,n} = \{a \in \Xi \mid \langle a_n - a \mid t_{i,n}\rangle \geq g_i(a_n)\} \qquad \text{where} \quad t_{i,n} \in \partial g_i(a_n), \tag{5.80}$$

is a valid outer approximation of S_i at iteration n. Indeed, $a_n \in S_{i,n}$ would imply $0 \geq g_i(a_n)$, which is impossible since $a_n \notin S_i$. Moreover, take any $a \in S_i$. Then $g_i(a) \leq 0$. But according to (2.14)

$$\begin{aligned} g_i(a_n) &\leq g_i(a) + \langle a_n - a \mid t_{i,n}\rangle & (5.81)\\ &\leq \langle a_n - a \mid t_{i,n}\rangle. & (5.82) \end{aligned}$$

Therefore $a \in S_{i,n}$ and $S_i \subset S_{i,n}$. Note that we have

$$S_{i,n} = \{a \in \Xi \mid \langle a \mid t_{i,n}\rangle \leq \langle a_n \mid t_{i,n}\rangle - g_i(a_n)\} \tag{5.83}$$

Therefore, the projection of a_n onto $S_{i,n}$ is given by (2.23) and reads

$$P_{i,n}(a_n) = a_n - \frac{g_i(a_n)}{\|t_{i,n}\|^2} t_{i,n}. \tag{5.84}$$

This projection is called a subgradient projection. With such projections, only the computation of a subgradient $t_{i,n}$ (of the gradient $\nabla g_i(a_n)$ if g_i is differentiable at a_n) is needed to process the set S_i at iteration n as opposed to the potentially involved exact projection

[12] This is essentially the same problem that arises in cutting plane methods in nonlinear programming [107].

$P_i(a_n)$. It is important to note that subgradient projections generalize the notion of projections. Indeed, if we let $g_i = d(\cdot, S_i)$, then (5.84) yields the exact projections thanks to (2.17). In general, for an arbitrary $a_n \in \Xi$, the subgradient projection of a_n onto S_i in (5.79) will be defined by

$$P_{i,n}(a_n) = \begin{cases} a_n - \dfrac{g_i(a_n)}{\|t_{i,n}\|^2}\, t_{i,n} & \text{if } a_n \in \complement S_i \\ a_n & \text{otherwise} \end{cases} \qquad \text{where} \quad t_{i,n} \in \partial g_i(a_n), \tag{5.85}$$

and it can be taken as the conventional projection $P_i(a_n)$ whenever this exact projection is easy to compute.

5.6.2 Examples of Subgradient Projections

We have seen that the projections onto the sets (5.73) and (5.74) needed to be computed iteratively. By contrast the subgradient projection of an image a_n onto (5.73) is simply obtained via (5.85) as

$$P_{i,n}(a_n) = \begin{cases} a_n + \dfrac{\|y(a_n)\|^2 - \zeta_2}{2\,\|\,^t T y(a_n)\|^2}\; {}^t T y(a_n) & \text{if } \|y(a_n)\|^2 > \zeta_2 \\ a_n & \text{otherwise,} \end{cases} \tag{5.86}$$

where $y(a_n) = x - T a_n$ and where we have used the identity

$$\nabla g_i(a_n) = \nabla(\|x - T a_n\|^2 - \zeta_2) = -2\; {}^t T(x - T a_n). \tag{5.87}$$

Likewise, subgradient projection of an image a_n onto (5.74) is obtained via

$$P_{i,n}(a_n) = \begin{cases} a_n - \dfrac{\langle a_n \mid \ln(a_n)\rangle + \eta}{\|\ln(a_n) + \vec{1}\|^2}\,(\ln(a_n) + \vec{1}) & \text{if } \; -\langle a_n \mid \ln(a_n)\rangle < \eta \\ a_n & \text{otherwise,} \end{cases} \tag{5.88}$$

where $\vec{1}$ denotes the vector of ones in $\mathbb{E}^{N^2}$ and where we have used the identity

$$\nabla g_i(a_n) = \nabla(\langle a_n \mid \ln(a_n)\rangle + \eta) = \ln(a_n) + \vec{1}. \tag{5.89}$$

5.6.3 Algorithm

Given an initial point $a_0 \in \Xi$ and numbers $C \in \mathbb{N}^*$, $\delta \in]0, 1/C[$, and $\varepsilon \in]0, 1[$, EMOPSP is defined by the iterative process [41]

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda_n\Big(\sum_{i \in I_n} w_{i,n} P_{i,n}(a_n) - a_n\Big), \tag{5.90}$$

where at each iteration n:

a) The family I_n of indices of selected sets satisfies

$$\emptyset \neq I_n \subset I \quad \text{and} \quad \operatorname{card}\{i \in I_n \mid a_n \notin S_i\} \leq C. \tag{5.91}$$

b) The subgradient projections $(P_{i,n}(a_n))_{i\in I_n}$ are defined by (5.85).

c) The weights $(w_{i,n})_{i\in I_n}$ conform to (5.56).

d) The relaxation parameter λ_n lies in $[\varepsilon, (2-\varepsilon)L_n]$, where L_n is as in (5.78).

5.6.4 Convergence Results

Recall that, for every $i \in I$, S_i is defined in (5.79) via a (lower semi-)continuous convex functional $g_i : \Xi \to \mathbb{R}$. We shall say that the subdifferentials of $(g_i)_{i\in I}$ are locally uniformly bounded if

$$(\forall \gamma \in \mathbb{R}_+^*)(\exists \zeta \in \mathbb{R}_+^*)(\forall i \in I)(\forall a \in B(0,\gamma)) \;\; \partial g_i(a) \subset B(0,\zeta). \tag{5.92}$$

Theorem 5.18 [41] *Suppose that the subdifferentials of $(g_i)_{i\in I}$ are locally uniformly bounded. Then, under admissible control, every orbit of EMOPSP converges weakly to a point in S.*

The next theorem pertains to strong convergence under chaotic control. Naturally, additional hypotheses are required.

Theorem 5.19 [41] *Suppose that the subdifferentials of $(g_i)_{i\in I}$ are locally uniformly bounded. Then, under chaotic control, every orbit of EMOPSP converges strongly to a point in S if either of the following conditions is satisfied.*

(i) $\overset{\circ}{S} \neq \emptyset$*;*

(ii) *The family $(g_i)_{i\in I}$ is finite and contains a lower semi-boundedly-compact functional.*

To our knowledge, these results are the most general ones available for the subgradient methods governed by (5.90). In particular, the following corollary of Theorem 5.19(ii) generalizes results of [29], which considered serial, cyclic control, as well as results of [60], which considered static control.

Proposition 5.11 *Suppose that $\dim \Xi < +\infty$ and $\operatorname{card} I < +\infty$. Then, under chaotic control, every orbit of EMOPSP converges (strongly) to a point in S.*

Proof. If $\dim \Xi < +\infty$ and $\operatorname{card} I < +\infty$, then $(g_i)_{i\in I}$ satisfies (5.92) [137]. In addition, each g_i is l.s.b.co. by virtue of Proposition 2.3(i). □

5.7 Extrapolated Method of Parallel Nonexpansive Operators (EMOPNO)

5.7.1 Problem Statement

Another generalization of the projection framework of Section 5.4 can be obtained by replacing the projection operators $(P_i)_{i\in I}$ by arbitrary firmly nonexpansive operators $(T_i)_{i\in I}$ such that

$$(\forall i \in I)\ \ S_i = \operatorname{Fix} T_i. \tag{5.93}$$

This framework is of interest when constraints are specified as invariance properties, say $h = T_i(h)$, where T_i is nonexpansive.[13] For instance, T_i may be a local rotation or reflection operator to model local symmetries in the image, or a translation operator to model certain periodicities, etc. In such cases, activating the property set $S_i = \operatorname{Fix} T_i$ through the projection operator P_i may be difficult, whereas activating it through the readily available operator T_i is straightforward. In this regard, it should be noted that, by virtue of (2.32), the elementary update $a_{n+1} = T_i(a_n)$ is still a step in the direction of S_i.

5.7.2 Algorithm

Given an initial point $a_0 \in \Xi$ and numbers $C \in \mathbb{N}^*$, $\delta \in]0, 1/C[$, and $\varepsilon \in]0, 1/2[$, EMOPNO is defined by the recursion [43]

$$(\forall n \in \mathbb{N})\ \ a_{n+1} = a_n + \lambda_n(\sum_{i\in I_n} w_{i,n} T_i(a_n) - a_n), \tag{5.94}$$

where

a) The family I_n of indices of selected operators satisfies

$$\emptyset \neq I_n \subset I \quad \text{and} \quad \operatorname{card}\{i \in I_n \mid a_n \notin S_i\} \leq C. \tag{5.95}$$

b) The weights $(w_{i,n})_{i\in I_n}$ conform to (5.56).

c) $\varepsilon \leq \lambda_n \leq (2-\varepsilon)L_n$ with $L_n = \begin{cases} \dfrac{\sum_{i\in I_n} w_{i,n}\|T_i(a_n) - a_n\|^2}{\|\sum_{i\in I_n} w_{i,n} T_i(a_n) - a_n\|^2} & \text{if } a_n \notin \bigcap_{i\in I_n} S_i \\ 1 & \text{otherwise.} \end{cases}$

[13] We actually address the problem of finding a common fixed-point of firmly nonexpansive operators but it is closely related to that of finding a common fixed-point of nonexpansive operators. Indeed, Proposition 2.8(iv) indicates that a nonexpansive operator T_i' can be associated with a firmly nonexpansive operator $T_i = (T_i' + \mathrm{Id})/2$ where, by construction, $\operatorname{Fix} T_i = \operatorname{Fix} T_i'$.

5.7.3 Convergence Results

Theorem 5.20 [43] *Under admissible control, every orbit of EMOPNO converges weakly to a point in S. The convergence is strong if $(T_i)_{i\in I}$ contains a demicompact mapping.*

Theorem 5.21 [43] *Under chaotic control, every orbit of EMOPNO converges strongly to a point in S if either of the following conditions is satisfied.*

(i) $\overset{\circ}{S} \neq \emptyset$.

(ii) *The family $(T_i)_{i\in I}$ is finite and contains a demicompact mapping.*

Theorem 5.20 improves upon results of [19], which considered a serial scheme. Theorem 5.20 and Theorem 5.21(i) improve, respectively, upon results of [131] and [121], which both considered the successive approximation scheme $a_{n+1} = T(a_n)$. The above theorems also generalize certain results of Section 5.4.4, which were restricted to projection operators. In particular, thanks to Proposition 2.10, condition (ii) above generalizes condition (ii) in Theorem 5.15. Finally, since in finite dimensional spaces any operator is demicompact, we obtain the following corollary of Theorem 5.21(ii). It generalizes a result of [174], which considered card $I_n = 1$ in (5.95).

Proposition 5.12 *Suppose* card $I < +\infty$ *and* $\dim \Xi < +\infty$. *Then, under chaotic control, every orbit of EMOPNO converges (strongly) to a point in S.*

5.8 Towards Unification

EMOPAP, EMOPSP, and EMOPNO are three separate generalizations of EMOPP which are not related in general. However, given their similar structure, it is natural to contemplate the possibility of unifying them in a single framework.

An important step towards unification was made in [11] where, under the more restrictive assumptions card $I < +\infty$ and $(\lambda_n)_{n\geq 0} \subset [\varepsilon, 2-\varepsilon]$, some of the results of Sections 5.4-5.7 were obtained by investigating a general iterative method for solving convex feasibility

problems. The algorithm proposed there was of the form[14]

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = a_n + \lambda_n(\sum_{i\in I_n} w_{i,n} T_{i,n}(a_n) - a_n), \tag{5.96}$$

where at each iteration n:

a) The family I_n of indices of selected sets satisfies

$$\emptyset \neq I_n \subset I. \tag{5.97}$$

b) $(T_{i,n})_{i\in I_n}$ is a family of firmly nonexpansive operators such that

$$(\forall i \in I_n) \;\; S_i \subset \operatorname{Fix} T_{i,n}. \tag{5.98}$$

c) The weights $(w_{i,n})_{i\in I_n}$ satisfy a condition similar to (5.56).

d) The relaxation parameter λ_n lies in $[\varepsilon, 2-\varepsilon]$.

In addition, a so-called focusing condition was introduced to study convergence. It requires that for every $i \in I$ and every subsequence $(a_{n_k})_{k\geq 0}$ of an orbit of the algorithm, we have

$$\left\{ \begin{array}{l} a_{n_k} \overset{k}{\rightharpoonup} a \\ a_{n_k} - T_{i,n_k}(a_{n_k}) \overset{k}{\to} 0 \\ (w_{i,n_k})_{k\geq 0} \subset]0,1[\end{array} \right. \quad \Rightarrow \quad a \in S_i. \tag{5.99}$$

This study also contains a number of results on geometrical convergence rates.

It appears reasonable to investigate the algorithms presented above in a single framework described by the recursion (5.96) with a)-c) but that would allow, as in Sections 5.4-5.7, countable families of property sets under suitable control modes, as well as extrapolated relaxations, i.e.,

$$\text{d)} \;\; \varepsilon \leq \lambda_n \leq (2-\varepsilon) L_n \;\; \text{with} \;\; L_n = \left\{ \begin{array}{ll} \dfrac{\sum_{i\in I_n} w_{i,n} \|T_{i,n}(a_n) - a_n\|^2}{\| \sum_{i\in I_n} w_{i,n} T_{i,n}(a_n) - a_n\|^2} & \text{if } \; a_n \notin \bigcap_{i\in I_n} S_i \\ 1 & \text{otherwise.} \end{array} \right.$$

[14] Actually, the algorithm of [11] proceeds by averaging relaxed operators, i.e.,

$$(\forall n \in \mathbb{N}) \;\; a_{n+1} = \sum_{i\in I_n} w_{i,n}((1-\lambda_{i,n})a_n + \lambda_{i,n} T_{i,n}(a_n)).$$

But this is equivalent to relaxing averaged operators, as in (5.96).

5.9 Practical Considerations for Digital Image Processing

In this section, we discuss the practical issues pertaining to the numerical realization of the proposed methods on a digital computer. This places us in the context of a finite number of sets $(S_i)_{i\in I}$ in the euclidean space $\mathbb{E}^{N^2}$. In other words, all the above results should now be viewed from the perspective $\operatorname{card} I < +\infty$ and $\dim \Xi < +\infty$. Fortunately, this is the context in which the most powerful convergence results were obtained.

Inconsistent problems will be considered first.

5.9.1 Inconsistent Problems

When the property sets do not intersect POCS has been seen to be inadequate[15] and two parallel methods producing weighted least-squares solutions were developed in Section 5.3. The method (5.26) is intersting theoretically for it converges strongly and it provides the closest least-squares solution from a starting point a_0. The first aspect is irrelevant in digital processing since weak and strong convergence modes coincide. As to the second, it may be of interest in certain best approximation problems, but since our chief interest here is just feasibility, we shall discuss only the second method, namely PPM (5.23)+(5.3).

First of all, it follows from Theorem 5.4 that any sequence generated by PPM converges to a solution of the weighted least-squares problem (5.5). In practice, PPM will provide an approximate minimum of the proximity function Φ in a finite number of steps. According to Proposition 5.5, 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

$$\Phi(a_n) - \Phi(a_{n+1}) \leq \epsilon, \tag{5.100}$$

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

$$\|\nabla\Phi(a_n)\| = \|a_n - \sum_{i\in I} w_i P_i(a_n)\| \leq \epsilon. \tag{5.101}$$

In implementing PPM, one should also be aware of the influence of the weights $(w_i)_{i\in I}$ on solutions. 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

[15] It goes without saying that the same is true of any serial method of type (5.1).

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, $w_i = 1/\text{card } I$.

We have seen that overrelaxations had an accelerating effect on the algorithm. One could therefore blindly choose relaxations in $[1, 2 - \varepsilon]$. However, an explicit relaxation rule can be determined by going back to Proposition 5.5. Since PPM behaves as a steepest-descent method, we can use the relaxation scheme devised by Armijo [6] 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. In our applications, this adaptation scheme yielded overrelaxations that converged efficiently.

Based on numerical experience and the recommendations of [135] 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 $(w_i)_{i\in I}$, and $\epsilon \in]0, +\infty[$. Set $n = 0$.
2. Set $\nabla\Phi(a_n) = a_n - \sum_{i\in I} w_i P_i(a_n)$ 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.

5.9.2 Consistent Problems

In consistent problems, we highly recommend that the extrapolated parallel methods of Section 5.4-5.7 be used. We shall discuss EMOPSP here since in practice sets are most frequently specified in the format (5.79). The convergence of EMOPSP is guaranteed by Proposition 5.11, which indicates that the sets can be activated in any order so long as every set is used repeatedly in the course of the iterations.

EMOPSP is superior to the widely used POCS algorithm on three counts.

1. It is straightforward to implement on any parallel machine, as the number of activated sets is variable.
2. It converges very efficiently thanks to its extrapolated relaxations.

3. It does not rely on the often cumbersome computation of exact projections and involves only the evaluation of subgradients.

EMOPSP is faster than POCS in that each iteration has a lower computational cost (item 3) and the whole iterative process converges in a smaller number of steps (item 2). EMOPSP is also very versatile as all of its parameters can be changed at each iteration (sets selected, approximating supersets, weights on the projections, relaxations). However, a standard implementation can be obtained with the following guidelines [41].

5.9.2.1 Control

If the number P of parallel processors is at least equal to the number m of sets, one can implement the algorithm with static control. It may not be worth activating only the violated sets since checking for membership in a set is usually done before projecting and there will be no savings in terms of computation. When $m > P$, then only violated sets should be activated. The chaotic control mode does not impose any specific scheduling for the processing of the sets but, for the sake of simplicity, one may want to sweep through the constraints circularly and activate blocks of P consecutive violated sets.

5.9.2.2 Weights

Although the weights can be defined in a number of ways and may have some influence on the centering of the algorithm, it is usually best to keep them uniform, that is

$$(\forall n \in \mathbb{N})(\forall i \in I_n) \;\; w_{i,n} = 1/\text{card}\, I_n. \tag{5.102}$$

5.9.2.3 Relaxations

Although no general conclusion is intended, our intensive simulations with EMOPSP in various problems has revealed the following behavior. When a small number of sets is used, very large extrapolations (say $1.5L_n \leq \lambda_n \leq 1.99L_n$) often create a lot of zigzagging and are not as effective as the centered extrapolations (5.43). On the other hand, large extrapolations accelerate the iterations significantly in more sizable problems.

5.9.2.4 Stopping Rule

If static control is used with exact projections (in which case EMOPSP reduces to EPPM2), a stopping rule involving Φ, such as (5.100), can be used. In other cases, the exact projections $(P_i(a_n))_{i\in I}$ will not be available at iteration n and alternative stopping criteria must

be considered, e.g., $\|a_{n+1} - a_n\| \leq \epsilon$, $\sum_{k=0}^{M} \sum_{i \in I_{n-k}} \|P_{i,n-k}(a_{n-k}) - a_{n-k}\|^2 \leq \epsilon$ for some $M \in \mathbb{N}$, etc.

6 Numerical Examples

This section is devoted to concrete numerical examples of convex set theoretic digital signal and image recoveries. The results of previous sections will therefore be applied in the context described in Sections 3.1.2.4 and 5.9.

6.1 Recovery with Inconsistent Constraints

This example is taken from [42] and illustrates an application of PPM to the set theoretic restoration of a one-dimensional signal in the presence of an inconsistent family of constraints.

6.1.1 Experiment

The problem is to deconvolve a noisy discrete-time N-point signal, i.e., to estimate 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 $N = 64$ and the solution space is the N-dimensional euclidean space $\mathbb{E}^N$. The original signal h is shown in Fig. 15. The recorded signal x of Fig. 16 was obtained via the standard convolutional model

$$x = Th + u, \tag{6.1}$$

in which the $N \times N$ Tœplitz matrix T models a shift-invariant linear blur and u is a vector of noise samples uniformly distributed in $[-\delta, \delta]$, with $\delta = 0.15$. The blurring kernel is a Gaussian function with a variance of 2 samples^2. If T_n designates the nth row of T and x_n the nth component of the data vector x, (6.1) can be written as

$$(\forall n \in \{0, \cdots, N-1\}) \;\; x_n = \langle T_n \mid h \rangle + u_n, \tag{6.2}$$

which is a special case of (4.16).

6.1.2 Set Theoretic Formulation

The set theoretic formulation for the problem consists of $m = 66$ closed and convex sets. The sets $(S_n)_{0 \le n \le N-1}$ are based on the knowledge of the blurring operator T and the information that the noise samples are distributed in $[-\delta, \delta]$. According to the analysis of Section 4.3.4, they take the form of hyperslabs defined by (4.26), namely

$$S_n = \{a \in \mathbb{E}^N \mid x_n - \delta \le \langle T_n \mid a \rangle \le x_n + \delta\}, \quad \text{for} \quad 0 \le n \le N-1. \tag{6.3}$$

Therefore, the projection $P_n(a)$ of a signal a onto S_n is given by (2.24) and reads

$$\begin{cases} a + [(x_n + \delta - \langle T_n \mid a \rangle)/\|T_n\|^2]T_n & \text{if } \langle T_n \mid a \rangle > x_n + \delta \\ a + [(x_n - \delta - \langle T_n \mid a \rangle)/\|T_n\|^2]T_n & \text{if } \langle T_n \mid a \rangle < x_n - \delta \\ a & \text{otherwise.} \end{cases} \tag{6.4}$$

The next set is constructed by assuming knowledge of the phase of h. From (4.13), we obtain

$$S_{m-1} = \{a \in \mathbb{E}^N \mid (\forall k \in \{0, \cdots, N-1\})\ \angle \widehat{a}(k) = \angle \widehat{h}(k)\}, \tag{6.5}$$

where $\widehat{h}$ is the N-point DFT of h. Since the DFT operator is an isometry (up to a factor N), the projection onto S_{m-1} can be performed in the DFT domain for each frequency individually. It is then easy to show that the projection of a signal a onto S_{m-1} is the signal $P_{m-1}(a) = b$, where for every k in $\{0, \cdots, N-1\}$

$$\widehat{b}(k) = \begin{cases} 0 & \text{if } \cos(\angle \widehat{a}(k) - \angle \widehat{h}(k)) \leq 0 \\ |\widehat{a}(k)| \cos(\angle \widehat{a}(k) - \angle \widehat{h}(k)) \exp(\imath \angle \widehat{h}(k)) & \text{otherwise.} \end{cases} \tag{6.6}$$

The last set arises from the prior knowledge that the components of h are nonnegative and bounded by $\|h\|_\infty = 12$. This leads to the bounded set (4.1) defined by

$$S_m = \{a \in \mathbb{E}^N \mid (\forall i \in \{0, \cdots, N-1\})\ 0 \leq a^{(i)} \leq \|h\|_\infty\}. \tag{6.7}$$

The projection of a signal a onto S_m is given by $P_m(a) = b$, where

$$(\forall i \in \{0, \cdots, N-1\})\ \ b^{(i)} = \begin{cases} 0 & \text{if } a^{(i)} < 0 \\ \|h\|_\infty & \text{if } a^{(i)} > \|h\|_\infty \\ a^{(i)} & \text{otherwise.} \end{cases} \tag{6.8}$$

6.1.3 Results

All the algorithms are initialized with the degraded signal, i.e., $a_0 = x$. The feasible signal of Fig. 17 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 10dB of background noise. The limiting signal of the subsequence $(a_{nm})_{n \geq 0}$ generated by POCS in this case is depicted in Fig. 18. As discussed in Section 5.2.3, the only definite property of this signal is to lie in S_m and, thereby, to satisfy the amplitude constraints. The convergence behavior of POCS in the inconsistent case is shown in Fig. 19, where the values taken by

the proximity function $(\Phi(a_{nm}))_{n\geq 0}$ are plotted. 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. 20. This least-squares solution to the inconsistent feasibility problem has fewer artifacts than the solution generated by POCS. The sequence $(\Phi(a_n))_{n\geq 0}$ produced by PPM is shown in Fig. 21. PPM achieves a much lower asymptotic degree of unfeasibility than POCS of $\Phi(a_\infty) = 0.035$.

6.1.4 Numerical Performance

Fig. 22 depicts the convergence behavior of PPM subjected to various relaxations schemes. In the underrelaxed case A, $(\lambda_n)_{n\geq 0}$ is drawn randomly 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 drawn randomly from the interval $[1, 2]$; in the adapted case D, $(\lambda_n)_{n\geq 0}$ is obtained as in Section 5.9.1. These plots support the claims of Section 5.9.1 to the effect 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 5.5.

6.2 Deconvolution with Bounded Uncertainty

We demonstrate an example of deconvolution with bounded kernel disturbances and bounded measurement noise along the lines of [51].

6.2.1 Experiment

In this experiment, the length of the signals is set to $N = 512$. We consider the problem of recovering the original signal h of Fig. 23 from the data signal

$$x = \overline{T}h + \widetilde{T}h + v, \tag{6.9}$$

where $\overline{T}$ is an $N \times N$ Tœplitz matrix representing a known shift-invariant linear blur with kernel $\overline{\mathfrak{b}}$, $\widetilde{T}$ an $N \times N$ matrix representing an unknown shift-variant linear blur with kernel $\widetilde{\mathfrak{b}}$, and v a vector of bounded noise samples. We note that (6.9) is a special case of (4.16)-(4.17) which can be written component-wise as

$$(\forall n \in \{0, \cdots, N-1\}) \;\; x_n = \langle \overline{T}_n \mid h \rangle + \langle \widetilde{T}_n \mid h \rangle + v_n \triangleq \langle \overline{T}_n \mid h \rangle + u_n, \tag{6.10}$$

where $\overline{T}_n$ and $\widetilde{T}_n$ denote, respectively, the nth rows of $\overline{T}$ and $\widetilde{T}$. The known convolutional kernel $\overline{\mathfrak{b}}$ which makes up the rows of $\overline{T}$ is uniform and has length $l = 16$ points, i.e.,

$$(\forall i \in \{0, \cdots, N-1\}) \;\; \overline{\mathfrak{b}}^{(i)} = \begin{cases} 1/l & \text{if } 0 \leq i \leq l-1, \\ 0 & \text{otherwise.} \end{cases} \tag{6.11}$$

The unknown blurring kernel $\widetilde{\mathfrak{b}}$ which makes up the rows of $\widetilde{T}$ is shift-variant and has the same l-point region of support as $\overline{\mathfrak{b}}$. Furthermore, for every n, a bound on the ℓ^1-norm of each $\widetilde{T}_n$ is available, say $\|\widetilde{T}_n\|_1 \leq \alpha_n$ a.s. It is also assumed that h is nonnegative with maximum value $\|h\|_\infty = 7.4$. Finally, the absolute bound on the noise samples $(v_n)_{0\leq n\leq N-1}$ is $\beta = 0.1$.

6.2.2 Set Theoretic Formulation

From the above information, a bound on the uncertainty signal samples can be derived as[16]

$$(\forall n \in \{0, \cdots, N-1\}) \;\; |u_n| \leq \alpha_n \cdot \|h\|_\infty + \beta \triangleq \delta_n. \tag{6.12}$$

As seen in Section 6.1.2, we obtain from (4.26) the sets

$$S_n = \{a \in \mathbb{E}^N \mid x_n - \delta_n \leq \langle \overline{T}_n \mid a \rangle \leq x_n + \delta_n\}, \quad \text{for} \quad 0 \leq n \leq N-1. \tag{6.13}$$

The projection of a signal a onto S_n is

$$P_n(a) = \begin{cases} a + [(x_n + \delta_n - \langle \overline{T}_n \mid a \rangle)/\|\overline{T}_n\|^2]\overline{T}_n & \text{if } \langle \overline{T}_n \mid a \rangle > x_n + \delta_n \\ a + [(x_n - \delta_n - \langle \overline{T}_n \mid a \rangle)/\|\overline{T}_n\|^2]\overline{T}_n & \text{if } \langle \overline{T}_n \mid a \rangle < x_n - \delta_n \\ a & \text{otherwise.} \end{cases} \tag{6.14}$$

The last set S_N is based on the information on the amplitude of h, which yields the same set as in (6.7). The feasibility set is $S = \bigcap_{i=0}^{N} S_i$.

6.2.3 Results

As all the projections are easy to evaluate, EMOPP is used with $a_0 = x$ to obtain feasible solutions. First, we simulate an instance when the blur is known exactly, i.e., $\alpha_n = 0$. The degraded signal is shown in Fig. 24 and the set theoretic deconvolution in Fig. 25. Then, we introduce shift-varying perturbations in the blurring kernel, with $\alpha_n = 0.04$, to obtain the degraded signal of Fig. 26, whose restoration is shown in Fig. 27. Clearly, the added uncertainty has increased the bounds $(\delta_n)_{0\leq n\leq N-1}$ and therefore the feasibility set, which results in a poorer restoration.

Besides the knowledge of the component $\overline{T}$ of the system, the information used in this experiment is limited to upper and lower bounds on the input signal and the noise, and upper

[16] In general, such a bound can be obtained via Hölder's inequality as long as the ℓ^p-norm of h is known as well as bounds $(\alpha_n)_{0\leq n\leq N-1}$ for the ℓ^q-norms of the random vectors $(\widetilde{T}_n)_{0\leq n\leq N-1}$ where $p \in [1, +\infty]$ and $1/p + 1/q = 1$. For instance, $p = 2$ was chosen to derive (4.29), which assumed prior knowledge of the energy of h.

bounds on the ℓ^1-norm of the shift-variant disturbances affecting the blurring kernel. Let us emphasize that no statistical assumption has been made and that the only conventional deconvolution method that could be implemented with such little information would be inverse filtering, which is known to give unacceptable results [5].

6.3 Image Restoration with Bounded Noise

In this section, a two-dimensional version of the previous experiment is investigated. It leads to a set theoretic formulation with $m = 16385$ sets, which will allow us to demonstrate the flexibility of EMOPP in large scale problems. Such large set theoretic formulations have also been encountered in other studies, e.g., [127], [157], [171], where they were solved using POCS.

6.3.1 Preliminaries

All images have $N \times N$ pixels ($N = 128$) and will be represented using stacked-vector notations as in Section 3.1.2.4. Ξ is the usual N^2-dimensional euclidean space $\mathbb{E}^{N^2}$. Every algorithm will be initialized with the degraded image, i.e., $a_0 = x$, and the progression of its orbit $(a_n)_{n\geq 0}$ will be tracked by plotting the normalized decibel values $(10\log_{10}(\Phi(a_n)/\Phi(a_0)))_{n\geq 0}$ of the proximity function (3.7), where

$$(\forall i \in I) \;\; w_i = 1/(\text{card } I). \tag{6.15}$$

As a practical stopping rule to compare performance, we shall use the criterion

$$\Phi(a_n) \leq \frac{\|h\|_\infty^2}{1300\,\text{card } I}. \tag{6.16}$$

As seen in Section 5.9.2.4, the sequence $(\Phi(a_n))_{n\geq 0}$ will usually not be computed in actual applications, but we use it here as we need a pertinent and uniform quantification of the notion of unfeasibility to compare accurately the performance of the algorithms.

6.3.2 Experiment

The original image h of Fig. 28 is degraded by convolutional blur with a uniform 7×7 kernel $\mathfrak{b}$ and addition of noise. The noise samples are distributed in the interval $[0, R]$ and the resulting blurred image-to-noise ratio is 32dB. The degraded image x is shown in Fig. 29. It can be written as

$$x = Th + u, \tag{6.17}$$

where T is the $N^2 \times N^2$ block-Tœplitz matrix associated with the point spread function $\mathfrak{b}$ [5] and u is a noise vector.

6.3.3 Set Theoretic Formulation

First, we assume that the point spread function $\mathfrak{b}$ (or equivalently T) is known. No probabilistic information is available about the noise vector u, except that its components lie in $[0, R]$. As before, this information leads to the N^2 hyperslabs

$$S_n = \{a \in \mathbb{E}^{N^2} \mid 0 \leq x_n - \langle T_n \mid a \rangle \leq R\} \quad \text{for} \quad 0 \leq n \leq N^2 - 1, \tag{6.18}$$

where T_n is the nth row of T. Then, by using the fact that the pixel values are nonnegative, the last property set we obtain is the nonnegative orthant

$$S_{N^2} = (\mathbb{R}_+)^{N^2}. \tag{6.19}$$

The projection of an image a onto S_{N^2} is simply

$$P_{N^2}(a) = a^+ = {}^t[\max\{0, a^{(i)}\}]_{0 \leq i \leq N^2 - 1}. \tag{6.20}$$

The set theoretic formulation is $(\mathbb{E}^{N^2}, (S_i)_{0 \leq i \leq N^2})$ and it comprises $m = N^2 + 1 = 16385$ sets. Since all the projections are easily computed, EMOPP will be used to solve the feasibility problem.

6.3.4 Numerical Performance

POCS (5.4) is implemented by skipping the nonviolated sets so that each iteration actually produces an update. The convergence pattern of POCS is shown in Fig. 30. To implement EMOPP, computer architectures with $P = 8$ and 64 parallel processors are considered.[17] At each iteration, the control selects P sets as follows: S_{N^2}, if it is violated, and a block of consecutive violated sets in (6.18). In addition, over the iterations, the sets $(S_n)_{0 \leq n \leq N^2 - 1}$ are swept through in a circular fashion. Three values of λ_n are considered: 1, L_n, and $1.9L_n$. In Figs. 31 and 32, the corresponding algorithms are labeled as EMOPP(1), EMOPP(L), and EMOPP(1.9L), respectively. These plots clearly show the numerical superiority of EMOPP and the remarkable acceleration provided by extrapolated overrelaxations. Thus, the -55dB mark corresponding to the stopping rule (6.16) was reached by POCS in 44700 iterations. By contrast, it took EMOPP(1.9L) only 5346 iterations to reach this point with 8 processors and 1168 iterations with 64 processors.

[17] Our AT&T Pixel Machines have 64 parallel processors.

6.3.5 Results

The restored image obtained by EMOPP is shown in Fig. 33. Again, it is important to stress that the only information available about the noise consists of amplitude bounds and that no probabilistic assumption whatsoever has been made. None of the conventional methods could operate with such little information (except inverse filtering, but it is unacceptable in the presence of noise).

6.3.6 Bounded versus Unbounded Noise

As was mentioned in Section 4.3.4.1, in the presence of bounded noise, the confidence coefficient c on the solution set defined in (4.55) is 100%. A question that naturally arises is what happens when the noise is unbounded. To answer this question, let us assume that the components of the noise vector u in (6.17) are i.i.d. and distributed as a zero mean normal r.v. U_0 with known second moment $\sigma^2 = \mathsf{E}|U_0|^2$, adjusted so that the blurred image-to-noise ratio is again 32dB. The degraded image x thus obtained is shown in Fig. 34. According to the results of Section 4.3.4.2, the sets $(S_n)_{0\le n\le N^2-1}$ take the form

$$S_n = \{a \in \mathbb{E}^{N^2} \mid x_n - \alpha\sigma \le \langle T_n \mid a\rangle \le x_n + \alpha\sigma\} \quad \text{for} \quad 0 \le n \le N^2 - 1, \tag{6.21}$$

where α is to be determined from the tables of the standard normal distribution in terms of the confidence coefficient $1-\epsilon$ placed on each set. Now suppose that we fix the global confidence coefficient at $\mathsf{c} = 95\%$ in (4.55). Then, since the noise samples are independent, we must have

$$1 - \epsilon = \mathsf{c}^{1/N^2} = 99.999687\%, \tag{6.22}$$

which gives $\alpha = 4.662$. Of course, these sets are "wider" than those obtained in the case of bounded noise. For instance, assume that in the experiment of Section 6.3.2 the noise samples were i.i.d. and distributed uniformly in $[0, R]$ as a r.v. V_0 of same power as U_0, i.e., $\mathsf{E}|V_0|^2 = \sigma^2$. Then the residual samples were constrained in (6.18) to fall in the interval $[0, R]$ which has length $R = 1.732\sigma$. In comparison, for gaussian noise, they were constrained in (6.21) to fall in a wider interval of length $2\alpha\sigma = 9.324\sigma$. As a result, the restoration obtained in this case with the sets of (6.19) and (6.21) is seen in Fig. 35 to be very poor.

We conclude that in the presence of bounded noise the sets (4.26) are quite effective and require minimal information to be constructed. On the other hand, when the noise is unbounded, they must be made large in order to secure a reasonable confidence coefficient c on their intersection. Consequently, they usually fail to describe the original image accurately and must be accompanied by other sets. It should also be noted that a substantial

amount of information is required to construct these sets in the case of unbounded noise. For instance, in the above example, the i.i.d. assumption was used and knowledge of the distribution of U_0 was assumed. Fortunately, when such information is available, all the sets described in Section 4.3 can be constructed to refine the set theoretic formulation, as we shall see in the next section.

6.4 Image Restoration via Subgradient Projections

We consider here an application of EMOPSP to set theoretic image restoration. A similar example was presented in [41]. $\mathfrak{F}$ is the two-dimensional DFT operator defined in (3.6) and the basic setup is as in Section 6.3.1.

6.4.1 Experiment

The experiment is the same as in Section 6.3.6: the degraded image x of Fig. 34 is obtained by convolving the original image h of Fig. 28 with a known uniform 7×7 kernel $\mathfrak{b}$ and addition of zero mean white gaussian noise with power σ^2. The blurred image-to-noise ratio is 32 dB.

The number of parallel processors is $P = 4$.

6.4.2 Set Theoretic Formulation

We first assume that the maximum intensity value $\|h\|_\infty$ of h is known to obtain the set

$$S_1 = \{a \in \mathbb{E}^{N^2} \mid (\forall i \in \{0, \cdots, N^2 - 1\})\ 0 \leq a^{(i)} \leq \|h\|_\infty\}. \tag{6.23}$$

The projection operator P_1 is described in (6.8). Next, we assume that the discrete Fourier transform of h is known over the low frequency region

$$K' = \{(k, l) \in \{0, \cdots, N-1\}^2 \mid 0 \leq k, l \leq M\}, \tag{6.24}$$

where $M = 21$. Recall that the two-dimensional DFT of real images possesses the conjugate-symmetry properties

$$(\forall (k,l) \in \{0, \cdots, N-1\}^2) \quad \begin{cases} \widehat{h}(k,0) = \overline{\widehat{h}}(N-k, 0) & \text{if } k \neq 0 \\ \widehat{h}(0,l) = \overline{\widehat{h}}(0, N-l) & \text{if } l \neq 0 \\ \widehat{h}(k,l) = \overline{\widehat{h}}(N-k, N-l) & \text{if } kl \neq 0. \end{cases} \tag{6.25}$$

The set K' must therefore be extended accordingly to a set K including all the symmetric pairs. The associated property set is then given by (4.9) as

$$S_2 = \{a \in \mathbb{E}^{N^2} \mid \widehat{a}1_K = \widehat{h}1_K\}. \tag{6.26}$$

Note that we have $\widehat{a} = \widehat{a}1_K + \widehat{a}1_{\complement K}$. Hence, the projection of a onto S_2 is given by

$$P_2(a) = \mathfrak{F}^{-1}\left(\widehat{h}1_K + \widehat{a}1_{\complement K}\right). \tag{6.27}$$

The information that the noise is zero mean white and gaussian with power σ^2 provides a complete description of its probabilistic structure. Hence, all the sets described in Section 4.3 can be constructed. For instance, since the noise samples are i.i.d. with second and fourth moments given, respectively, by σ^2 and $3\sigma^4$, the second moment set (4.41) becomes

$$S_3 = \{a \in \mathbb{E}^{N^2} \mid \|x - Ta\|^2 \leq \zeta_2\} \quad \text{where} \quad \zeta_2 = N(N + \alpha\sqrt{2})\sigma^2. \tag{6.28}$$

This set has proven quite useful in several applications, e.g., [48], [171]. Unfortunately, we have seen in Section 5.5.1 that its projection operator must be determined iteratively via a costly procedure, which precludes its use in certain applications [127]. However, with EMOPSP, S_3 can simply be activated via (5.86), where the subgradient projection of an image a onto S_3 was seen to be

$$\widetilde{P_3}(a) = \begin{cases} a + \dfrac{\|y(a)\|^2 - \zeta_2}{2\,\|\,^t Ty(a)\|^2}\;{}^t Ty(a) & \text{if } \|y(a)\|^2 > \zeta_2 \\ a & \text{otherwise,} \end{cases} \tag{6.29}$$

where $y(a) = x - Ta$ is the residual image. Upon making the standard block-circulant approximation on the matrix T [5], we obtain $\widehat{y(a)} = \widehat{a} - \widehat{\mathfrak{b}}\widehat{a}$. Whence, the upper line in (6.29) can be computed efficiently via the Fast Fourier Transform (FFT) as

$$\widetilde{P_3}(a) = \mathfrak{F}^{-1}\left(\widehat{a} + \frac{\|\widehat{y(a)}\|^2 - N^2\zeta_2}{2\|\overline{\widehat{\mathfrak{b}}}\,\widehat{y(a)}\|^2}\;\overline{\widehat{\mathfrak{b}}}\;\widehat{y(a)}\right), \tag{6.30}$$

where we have kept the notation $\|\cdot\|$ to designate the norm in the Fourier space, i.e.,

$$(\forall a \in \mathbb{E}^{N^2}) \;\; \|\widehat{a}\|^2 = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} |\widehat{a}(k,l)|^2. \tag{6.31}$$

The exact computation of $P_3(a)$ proposed in [171] typically requires 10 to 20 iterations of much higher complexity than (6.30). Consequently the subgradient projection reduces the cost of processing S_3 by at least an order of magnitude. To define the last set based on the spectral properties of the noise, let

$$D = \{1, \cdots, N/2 - 1\} \times \{1, \cdots, N - 1\}. \tag{6.32}$$

Then we can define

$$S_4 = \bigcap_{(k,l)\in D} \{a \in \mathbb{E}^{N^2} \mid \left|\widehat{y(a)}\,(k,l)\right|^2 \leq \xi\} \quad \text{where} \quad \xi = -N^2\sigma^2\ln(\epsilon). \tag{6.33}$$

We observe that this is not exactly the form in which the sets were given in Section 4.3.6.1. Indeed, it is more convenient here to replace (4.45) by the two-dimensional periodogram

$$(\forall (k,l) \in D)\;\; I_{k,l} = \frac{2}{N^2}\left|\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} U_{mN+n}\exp(-\imath\frac{2\pi}{N}(mk+nl))\right|^2, \tag{6.34}$$

where $(U_n)_{n\in\mathbb{Z}}$ is the noise process. It can be shown that Theorem 4.1(i)+(iii) remains true for the statistics of (6.34), which leads to the above definition of S_4. As was done for (6.5), the projection $P_4(a)$ of an image a onto S_4 can be performed in the Fourier domain for every frequency pair (k,l) individually. Note that, for any frequencies $(k,l) \in D$ such that $\widehat{\mathfrak{b}}(k,l) \neq 0$, the constraint on the residual can be written as

$$\widehat{a}(k,l) \in B\left(\frac{\widehat{x}(k,l)}{\widehat{\mathfrak{b}}(k,l)}, \frac{\sqrt{\xi}}{|\widehat{\mathfrak{b}}(k,l)|}\right). \tag{6.35}$$

The projection onto this ball is given by (2.25). Consequently, by taking (6.25) into account, we obtain

$$P_4(a) = \mathfrak{F}^{-1}\left(\left(\frac{\widehat{x} - \sqrt{\xi}\,\widehat{y(a)}/|\widehat{y(a)}|}{\widehat{\mathfrak{b}}}\right)1_E + \widehat{a}1_{\complement E}\right), \tag{6.36}$$

where

$$E = \left\{(k,l) \in \{0,\cdots,N-1\}^2 \mid \left\{\begin{array}{l} |\widehat{y(a)}\,(k,l)|^2 > \xi \\ \widehat{\mathfrak{b}}(k,l) \neq 0 \\ (k,l) \in D \quad \text{or} \quad (N-k,N-l) \in D \end{array}\right.\right\}. \tag{6.37}$$

To fully specify the sets S_3 and S_4 it remains to choose the confidence parameters α and ϵ. To this end, let us impose a global confidence coefficient of $\mathsf{c} = 95\%$ on the feasibility set in (4.55) and let us call p the confidence coefficient to be placed on S_3 and S_4. Consider the events

$$\left\{\begin{array}{l} A_3 = \{\omega \in \Omega \mid h \in S_3(\omega)\} \\ A_4 = \{\omega \in \Omega \mid h \in S_4(\omega)\}. \end{array}\right. \tag{6.38}$$

Note that since the statistics (4.36) and (6.34) are not mutually independent we cannot take $\mathsf{p} = \sqrt{\mathsf{c}}$. We can nonetheless derive the value of p from the relations

$$\begin{aligned} \mathsf{c} &= \mathsf{P}A_3 \cap A_4 && (6.39)\\ &= 1 - \mathsf{P}\complement A_3 \cup \complement A_4 && (6.40)\\ &\geq 1 - \mathsf{P}\complement A_3 - \mathsf{P}\complement A_4 && (6.41)\\ &= 2\mathsf{p} - 1. && (6.42) \end{aligned}$$

Hence, we should take $\mathsf{p} = 97.5\%$, which yields $\alpha = 2.241$ in (6.28). Moreover, since the statistics $(I_{k,l})_{(k,l)\in D}$ are independent, the confidence coefficient $1-\epsilon$ on the $(N-1)(N/2-1)$ sets defining S_4 should satisfy $(1-\epsilon)^{(N-1)(N/2-1)} = \mathsf{p}$, which yields $\epsilon = 3.164 \times 10^{-6}$ in (6.33).

We have now completely defined the set theoretic formulation $(\mathbb{E}^{N^2}, (S_i)_{1\leq i\leq 4})$ for this problem.

6.4.3 Numerical Performance

Various subgradient projection methods are compared here. The exact projection operators will be used for the sets S_1, S_2, and S_4 for they admit closed-form expressions. On the other hand, S_3 will be activated through its subgradient projection. We shall call subgradient POCS (SPOCS) the subgradient version of (5.4) thus obtained. Since we have $P = 4$ processors and $m = 4$ sets, EMOPSP is implemented with static control. Several relaxation schemes are considered. We shall call EMOPSP(1), EMOPSP(1.9), and EMOPSP(L) the algorithms obtained by taking at each iteration n relaxations $\lambda_n = 1$, $\lambda_n = 1.9$, and $\lambda_n = L_n$, respectively. Finally, EMOPSP(C) designates the algorithm obtained with the centering technique (5.43). Since the control is static, EMOPSP(1) can be regarded as the subgradient version of SIRT (3.11), EMOPSP(1.9) as an overrelaxed subgradient version of PPM (5.23), and EMOPSP(L) as the fully extrapolated subgradient version of EPPM.

The convergence patterns are shown in Fig. 36. We notice, that the unrelaxed EMOPSP(1) algorithm is slower than SPOCS and that overrelaxations in EMOPSP(1.9) have an accelerating effect. However, the extrapolated algorithm EMOPSP(L) is much faster and centering in EMOPSP(C) further accelerates the progression of the iterates towards a solution. Thus, the -51 dB mark corresponding to the stopping rule (6.16) was reached by SPOCS in 64 iterations and by EMOPSP(C) in only 14 iterations.

6.4.4 Results

Fig. 37 shows the image restored by EMOPSP. To give a more complete demonstration of the effectiveness of this particular set theoretic formulation, the same experiment was repeated on the image of Fig. 38. The degraded image appears in Fig. 39 and its restoration in Fig. 40.

7 Summary

Every image recovery problem is accompanied by some *a priori* knowledge. Together with the observed data, this *a priori* knowledge defines constraints on the solutions to the problem. In the conventional approach, an optimality criterion is introduced to define a unique solution and computational tractability imposes that many constraints be left out of the recovery process. As a result, the end product may violate known properties of the image being estimated.

In the set theoretic approach, the notion of feasibility prevails: any image which satisfies all the constraints arising from the data and *a priori* knowledge is an acceptable solution. A set of solutions is thus defined, whose elements are equally likely to have generated the observed data in the light of the available information. The main asset of this framework is to provide great flexibility in the incorporation of statistical as well as nonstatistical constraints. In addition, the recovered images thus obtained have - by construction - well defined, tangible and meaningful properties, which is often more valuable than satisfying some conceptual optimality criterion.

The focus of this survey has been placed on problems in which the property sets associated with the constraints are closed and convex in some Hilbert image space. In this context, the set theoretic image recovery problem can be abstracted into the problem of finding a common point of convex sets, i.e., into a convex feasibility problem. This framework is certainly limited by the restriction to convex constraints. However, this limitation is advantageously counterbalanced by the existence of efficient algorithms that are guaranteed to find feasible solutions. In addition, a wide range of useful constraints was seen to yield convex property sets.

The field originated in the early 1970's with the formulation of tomographic reconstruction and band-limited extrapolation problems as affine feasibility problems. Because these approaches lacked a general abstract formalism and powerful analytical tools, their scope remained limited both in the nature of the problems and in the amount of information that could be used. As convex feasibility algorithms entered the image recovery toolbox in the early 1980's, the restriction to subspace and half-space property sets disappeared, and a much wider range of information became exploitable. As a result, the set theoretic approach soon gained widespread recognition and found applications in numerous image recovery problems. Very recently, the field has benefited from a regained interest in the convex feasibility problem on the part of several groups of researchers, and efficient parallel alternatives to the rudimentary POCS algorithm have been proposed. We expect such developments to further broaden the scope of set theoretic image recovery by making it less involved computationally and therefore more widely applicable.

Naturally, the next logical extension would be to relax the convexity requirement on the

property sets. In this regard, the lack of a general-purpose, globally convergent method for solving nonconvex feasibility problems seems to be an unsurmountable obstacle. On the other hand, it is quite conceivable that suitable methods could be developed for specific problems.

Before closing this survey, the unavoidable question should be posed: when should an image recovery problem be formulated as a feasibility problem rather than an optimization problem? A complete and systematic answer is of course not possible and it would set the stage for endless philosophical discussions. In addition, some methods are simply known to work better in certain problems, which makes such a debate a rather academic one. Nonetheless, our view is that any optimization approach is acceptable as long as it yields a feasible image. If not, it simply produces a solution which is inconsistent with known facts about the image being estimated.

Appendix A - Acronyms

ART: algebraic reconstruction technique (Section 3.2.1)

EMOPP: extrapolated method of parallel projections (Section 5.4.2)

EMOPAP: extrapolated method of parallel approximate projections (Section 5.5)

EMOPNO: extrapolated method of parallel nonexpansive operators (Section 5.7)

EMOPSP: extrapolated method of parallel subgradient projections (Section 5.6)

EPPM: extrapolated parallel projection method (Section 5.4.1.2)

EPPM2: (generalized) extrapolated parallel projection method (Section 5.4.1.2)

POCS: projection onto convex sets (Section 3.2.4)

PPM: parallel projection method (Section 5.3.3)

SIRT: simultaneous iterative reconstruction technique (Section 3.2.1)

References

[1] S. Agmon, "The relaxation method for linear inequalities," *Canadian Journal of Mathematics,* vol. 6, no. 3, pp. 382-392, 1954.

[2] R. Aharoni, A. Berman, and Y. Censor, "An interior points algorithm for the convex feasibility problem," *Advances in Applied Mathematics,* vol. 4, no. 4, pp. 479-489, December 1983.

[3] 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, August 1989.

[4] I. Amemiya and T. Ando, "Convergence of random products of contractions in Hilbert space," *Acta Scientiarum Mathematicarum (Szeged),* vol. 26, no. 3, pp. 239-244, 1965.

[5] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[6] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics,* vol. 16, no. 1, pp. 1-3, 1966.

[7] J. Arsac, *Transformation de Fourier et Théorie des Distributions*. Paris: Dunod, 1961.

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

[9] R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, band-limited Fourier transform pairs from noisy data," *Journal of the Optical Society of America A,* vol. 2, no. 11, pp. 2027-2039, November 1985.

[10] H. H. Bauschke, "A norm convergence result on random products of relaxed projections in Hilbert space," *Transactions of the American Mathematical Society,* vol. 347, no. 4, pp. 1365-1373, April 1995.

[11] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," accepted for publication in *SIAM Review*.

[12] H. H. Bauschke, J. M. Borwein, and A. S. Lewis, "On the method of cyclic projections for convex sets in Hilbert space." Research report, Simon Fraser University, 1994.

[13] P. Billingsley, *Convergence of Probability Measures.* New York: Wiley, 1968.

[14] J. M. Boone, B. A. Arnold, and J. A. Seibert, "Characterization of the point spread function and modulation transfer function of scattered radiation using a digital imaging system," *Medical Physics,* vol. 13, pp. 254-256, 1986.

[15] G. Bouligand, *Introduction à la Géométrie Infinitésimale Directe.* Paris: Vuibert, 1932.

[16] N. Bourbaki, *Eléments de Mathématique - Espaces Vectoriels Topologiques, Chapitres 1 à 5.* Paris: Masson, 1981.

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

[18] 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.

[19] F. E. Browder, "Convergence theorems for sequences of nonlinear operators in Banach spaces," *Mathematische Zeitschrift,* vol. 100, no. 3, pp. 201-225, July 1967.

[20] R. E. Bruck, "Random products of contractions in metric and Banach spaces," *Journal of Mathematical Analysis and Applications,* vol. 88, no. 2, pp. 319-332, August 1982.

[21] D. Butnariu and Y. Censor, "On the behavior of a block-iterative projection method for solving convex feasibility problems," *International Journal of Computer Mathematics,* vol. 34, nos. 1-2, pp. 79-94, 1990.

[22] D. Butnariu and Y. Censor, "Strong convergence of almost simultaneous block-iterative projection methods in Hilbert spaces," *Journal of Computational and Applied Mathematics,* vol. 53, no. 1, pp. 33-42, July 1994.

[23] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Transactions on Image Processing,* vol. 2, no. 1, pp. 96-103, January 1993. ("Erratum and addendum," vol. 4, no. 2, pp. 226-227, February 1995.)

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

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

[26] Y. Censor, "Parallel application of block-iterative methods in medical imaging and radiation therapy," *Mathematical Programming,* vol. 42, no. 2, pp. 307-325, 1988.

[27] 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, April 1983.

[28] Y. Censor and G. T. Herman, "On some optimization techniques in image reconstruction from projections," *Applied Numerical Mathematics,* vol. 3, pp. 365-391, 1987.

[29] Y. Censor and A. Lent, "Cyclic subgradient projections," *Mathematical Programming,* vol. 24, no. 2, pp. 233-235, 1982.

[30] A. E. Çetin, "An iterative algorithm for signal reconstruction from bispectrum," *IEEE Transactions on Signal Processing,* vol. 39, no. 12, pp. 2621-2628, December 1991.

[31] A. E. Çetin and R. Ansari, "Convolution-based framework for signal recovery and applications," *Journal of the Optical Society of America A,* vol. 5, no. 8, pp. 1193-1200, August 1988.

[32] A. E. Çetin and R. Ansari, "Signal recovery from wavelet transform maxima," *IEEE Transactions on Signal Processing,* vol. 42, no. 1, pp. 194-196, January 1994.

[33] W. Cheney and A. A. Goldstein, "Proximity maps for convex sets," *Proceedings of the American Mathematical Society,* vol. 10, no. 3, pp. 448-450, June 1959.

[34] R. T. Chin, C. L. Yeh, and W. S. Olson, "Restoration of multichannel microwave radiometric images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 7, no. 4, pp. 475-484, July 1985.

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

[36] M. R. Civanlar and H. J. Trussell, "Digital signal restoration using fuzzy sets," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 34, no. 4, pp. 919-936, August 1986.

[37] D. Cochran, "Phase and magnitude in normalized images," *IEEE Transactions on Image Processing,* vol. 3, no. 6, pp. 858-862, November 1994.

[38] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE,* vol. 81, no. 2, pp. 182-208, February 1993.

[39] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Transactions on Image Processing,* vol. 2, no. 2, pp. 269-271, April 1993.

[40] P. L. Combettes, "Hilbertian convex feasibility problem: Convergence of projection methods," accepted for publication in *Applied Mathematics and Optimization.*

[41] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections." *IEEE Transactions on Image Processing,* submitted.

[42] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Transactions on Signal Processing,* vol. 42, no. 11, pp. 2955-2966, November 1994.

[43] P. L. Combettes, "Construction d'un point fixe commun à une famille de contractions fermes," *Comptes Rendus de l'Académie des Sciences de Paris,* Série I, vol. 320, no. 11, pp. 1385-1390, June 1995.

[44] P. L. Combettes, M. Benidir, and B. Picinbono, "A general framework for the incorporation of uncertainty in set theoretic estimation," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol. 3, pp. 349-352. San Francisco, CA, March 23-26, 1992.

[45] P. L. Combettes and H. Puh, "Extrapolated projection method for the euclidean convex feasibility problem." Technical report, City University of New York, 1993.

[46] P. L. Combettes and H. Puh, "Iterations of parallel convex projections in Hilbert spaces," *Numerical Functional Analysis and Optimization,* vol. 15, nos. 3-4, pp. 225-243, 1994.

[47] P. L. Combettes and H. J. Trussell, "Modèles et algorithmes en vue de la restauration numérique d'images rayons-X," *Proceedings of MARI-Cognitiva Electronic Image,* pp. 146-151. Paris, France, May 18-22, 1987.

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

[49] P. L. Combettes and H. J. Trussell, "Method of successive projections for finding a common point of sets in metric spaces," *Journal of Optimization Theory and Applications,* vol. 67, no. 3, pp. 487-507, December 1990.

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

[51] P. L. Combettes and H. J. Trussell, "Deconvolution with bounded uncertainty," *International Journal of Adaptive Control and Signal Processing,* vol. 9, no. 1, pp. 3-17, January 1995.

[52] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Statistics,* vol. 19, no. 4, pp. 2032-2066, December 1991.

[53] G. Demoment "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 37, no. 12, pp. 2024-2036, December 1989.

[54] 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, January 1985.

[55] 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.

[56] F. Deutsch, "The method of alternating orthogonal projections," in *Approximation Theory, Spline Functions and Applications,* (S. P. Singh, Editor), pp. 105-121. The Netherlands: Kluwer, 1992.

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

[58] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1953.

[59] J. L. Doob, *Measure Theory.* New York: Springer-Verlag, 1994.

[60] L. T. Dos Santos, "A parallel subgradient projections method for the convex feasibility problem," *Journal of Computational and Applied Mathematics,* vol. 18, no. 3, pp. 307-320, June 1987.

[61] J. M. Dye and S. Reich, "Unrestricted iterations of nonexpansive mappings in Hilbert space," *Nonlinear Analysis - Theory, Methods, and Applications,* vol. 18, no. 2, pp. 199-207, January 1992.

[62] S. Ebstein, "Stellar speckle interferometry energy spectrum recovery by convex projections," *Applied Optics,* vol. 26, no. 8, pp. 1530-1536, April 1987.

[63] B. Efron, "Controversies in the foundations of statistics," *The American Mathematical Monthly,* vol. 85, no. 4, pp. 231-246, April 1978.

[64] B. Efron, "Why isn't everyone a Bayesian?" *The American Statistician,* vol. 40, no. 1, pp. 1-5, February 1986.

[65] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels.* Paris: Dunod, 1974.

[66] P. Erdös, "Some remarks on the measurability of certain sets," *Bulletin of the American Mathematical Society,* vol. 51, no. 10, pp. 728-731, October 1945.

[67] M. Fisz, *Probability Theory and Mathematical Statistics,* third edition. New York: Wiley, 1963.

[68] S. D. Flåm and J. Zowe, "Relaxed outer projections, weighted averages, and convex feasibility," *BIT,* vol. 30, no. 2, pp. 289-300, 1990.

[69] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America,* vol. 56, no. 10, pp 1372-1379, October 1966.

[70] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *Journal of the Optical Society of America,* vol. 62, no. 4, pp. 511-518, April 1972.

[71] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, November 1984.

[72] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta,* vol. 21, no. 9, pp. 709-720, September 1974.

[73] 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, April 1972.

[74] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology,* vol. 36, no. 1, pp. 105-117, July 1972.

[75] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory.* Cambridge: Cambridge University Press, 1990.

[76] M. Goldburg and R. J. Marks II, "Signal synthesis in the presence of an inconsistent set of constraints," *IEEE Transactions on Circuits and Systems,* vol. 32, no. 7, pp. 647-663, July 1985.

[77] J. W. Goodman, *Statistical Optics.* New York: Wiley Interscience, 1985.

[78] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *Journal of Theoretical Biology,* vol. 29, no. 3, pp. 471-481, December 1970.

[79] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *USSR Computational Mathematics and Mathematical Physics,* vol. 7, no. 6, pp. 1-24, 1967.

[80] S. S. Gupta and J. O. Berger (Editors), *Statistical Decision Theory and Related Topics IV,* vol. 1. New York: Springer-Verlag, 1988.

[81] I. Halperin, "The product of projection operators," *Acta Scientiarum Mathematicarum (Szeged),* vol. 23, no. 1, pp. 96-99, 1962.

[82] M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 30, no. 2, pp. 140-154, April 1982.

[83] S. Hein and A. Zakhor, "Halftone to continuous-tone conversion of error-diffusion coded images," *IEEE Transactions on Image Processing,* vol. 4, no. 2, pp. 208-216, February 1995.

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

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

[86] G. T. Herman, "Mathematical optimization versus practical performance: A case study based on the maximum entropy criterion in image reconstruction," *Mathematical Programming Study,* vol. 20, pp. 96-112, October 1982.

[87] G. T. Herman, H. Hurwitz, A. Lent, and H. P. Lung, "On the Bayesian approach to image reconstruction," *Information and Control,* vol. 42, no. 1, pp. 60-71, July 1979.

[88] G. T. Herman, A. Lent, and P. H. Lutz, "Relaxation methods for image reconstruction," *Communications of the ACM,* vol. 21, no. 2, pp. 152-158, February 1978.

[89] B. R. Hunt, "The application of constrained least-squares estimation to image restoration by digital computer," *IEEE Transactions on Computers,* vol. 22, no. 9, pp. 805-812, September 1973.

[90] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Transactions on Computers,* vol. 26, no. 3, pp. 219-229, March 1977.

[91] N. E. Hurt, "Signal enhancement and the method of successive projections," *Acta Applicandae Mathematicae,* vol. 23, no. 2, pp. 145-162, May 1991.

[92] 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, August 1986.

[93] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. Oakland, CA: Holden Day, 1968.

[94] B. Jessen, "To saetninger om konvekse punktmaengder," *Matematisk Tiddskrift B,* vol. 1940, pp. 66-70, 1940.

[95] M. Jiang, "On Johnson's example of a nonconvex Chebyshev set," *Journal of Approximation Theory,* vol. 74, no. 2, pp. 152-158, August 1993.

[96] G. G. Johnson, "A nonconvex set which has the unique nearest point property," *Journal of Approximation Theory,* vol. 51, no. 4, pp. 289-332, December 1987.

[97] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen," *Bulletin de l'Académie des Sciences de Pologne,* vol. A35, pp. 355-357, 1937.

[98] A. K. Katsaggelos (Ed.), *Digital Image Restoration.* New York: Springer-Verlag, 1991.

[99] A. K. Katsaggelos, J. Biemond, R. W. Schafer, and R. M. Mersereau, "A regularized iterative image restoration algorithm," *IEEE Transactions on Signal Processing,* vol. 39, no. 4, pp. 914-929, April 1991.

[100] H. Kudo and T. Saito, "Sinogram recovery with the method of convex projections for limited-data reconstruction in computed tomography," *Journal of the Optical Society of America A,* vol. 8, no. 7, pp. 1148-1160, July 1991.

[101] S. S. Kuo and R. J. Mammone, "Image restoration by convex projections using adaptive constraints and the L_1 norm," *IEEE Transactions on Signal Processing,* vol. 40, no. 1, pp. 159-168, January 1992.

[102] H. J. Landau and W. L. Miranker, "The recovery of distorted band-limited signals," *Journal of Mathematical Analysis and Applications,* vol. 2, no. 1, pp. 97-104, February 1961.

[103] 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, January 1991.

[104] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Transactions on Medical Imaging,* vol. 6, no. 2, pp. 106-114, June 1987.

[105] R. M. Leahy and C. E. Goutis, "An optimal technique for constraint-based image restoration and reconstruction," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 34, no. 6, pp. 1629-1642, December 1986.

[106] A. Lent and H. Tuy, "An iterative method for the extrapolation of band-limited functions," *Journal of Mathematical Analysis and Applications,* vol. 83, no. 2, pp. 554-565, October 1981.

[107] D. G. Luenberger, *Linear and Nonlinear Programming,* 2nd ed. Redwood City, CA: Addison-Wesley, 1984.

[108] A. Levi and H. Stark, "Signal reconstruction from phase by projection onto convex sets," *Journal of the Optical Society of America,* vol. 73, no. 6, pp. 810-822, June 1983.

[109] A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," *Journal of the Optical Society of America A,* vol. 1, no. 9, pp. 932-943, September 1984.

[110] E. S. Levitin and B. T. Polyak, "Convergence of minimizing sequences in conditional extremum problems," *Soviet Mathematics - Doklady,* vol. 7, no. 3, pp. 764-767, May 1966.

[111] J. L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires.* Paris: Dunod, 1969.

[112] P. L. Lions, "Approximation de points fixes de contractions," *Comptes Rendus de l'Académie des Sciences de Paris,* vol. A284, no. 21, pp. 1357-1359, June 1977.

[113] G. E. Mailloux, F. Langlois, P. Y. Simard, and M. Bertrand, "Restoration of the velocity field of the heart from two-dimensional echograms," *IEEE Transactions on Medical Imaging,* vol. 8, no. 2, pp. 143-153, June 1989.

[114] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, no. 7, pp. 710-732, July 1992.

[115] J. Mandel, "Convergence of the cyclical relaxation method for linear inequalities," *Mathematical Programming,* vol. 30, no. 2, pp. 218-228, 1984.

[116] C. P. Mariadassou and B. Yegnanarayana, "Image reconstruction from noisy digital holograms," *IEE Proceedings-F,* vol. 137, no. 5, pp. 351-356, October 1990.

[117] C. L. Matson, "Fourier spectrum extrapolation and enhancement using support constraints," *IEEE Transactions on Signal Processing,* vol. 42, no. 1, pp. 156-163, January 1994.

[118] R. G. Miller, Jr., *Simultaneous Statistical Inference,* 2nd ed. New York: Springer-Verlag, 1981.

[119] A. Mohammad-Djafari and G. Demoment, "Maximum entropy image reconstruction in x-ray and diffraction tomography," *IEEE Transactions on Medical Imaging,* vol. 7, no. 4, pp. 345-354, December 1988.

[120] W. D. Montgomery, "Optical applications of Von Neumann's alternating-projection theorem," *Optics Letters,* vol. 7, no. 1, pp. 1-3, January 1982.

[121] J. J. Moreau, "Un cas de convergence des itérées d'une contraction d'un espace hilbertien," *Comptes Rendus de l'Académie des Sciences de Paris,* vol. A286, no. 3, pp. 143-144, January 1978.

[122] T. S. Motzkin and I. J. Schoenberg, "The relaxation method for linear inequalities," *Canadian Journal of Mathematics,* vol. 6, no. 3, pp. 393-404, 1954.

[123] 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 Transactions on Biomedical Engineering,* vol. 40, no. 4, pp. 323-328, April 1993.

[124] P. Oskoui-Fard and H. Stark, "Tomographic image reconstruction using the theory of convex projections," *IEEE Transactions on Medical Imaging,* vol. 7, no. 1, pp. 45-58, March 1988.

[125] P. Oskoui and H. Stark, "A comparative study of three reconstruction methods for a limited-view computer tomography problem,"' *IEEE Transactions on Medical Imaging,* vol. 8, no. 1, pp. 43-49, March 1989.

[126] N. Ottavy, "Strong convergence of projection-like methods in Hilbert spaces," *Journal of Optimization Theory and Applications,* vol. 56, no. 3, pp. 433-461, March 1988.

[127] M. K. Özkan, A. M. Tekalp, and M. I. Sezan, "POCS-based restoration of space-varying blurred images," *IEEE Transactions on Image Processing,* vol. 3, no. 4, pp. 450-454, July 1994.

[128] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Transactions on Circuits and Systems,* vol. 22, no. 9, pp. 735-742, September 1975.

[129] H. Peng and H. Stark, "Signal recovery with similarity constraints," *Journal of the Optical Society of America A,* vol. 6, no. 6, pp. 844-851, June 1989.

[130] H. Peng and H. Stark, "Image recovery in computer tomography from partial fan-beam data by convex projections," *IEEE Transactions on Medical Imaging,* vol. 11, no. 4, pp. 470-478, December 1992.

[131] W. V. Petryshyn, "Construction of fixed points of demicompact mappings in Hilbert space," *Journal of Mathematical Analysis and Applications,* vol. 14, no. 2, pp. 276-284, May 1966.

[132] G. Pierra, "Méthodes de projections parallèles extrapolées relatives à une intersection de convexes." Rapport de Recherche, INPG, Grenoble, France, September 1975. See also "Decomposition through formalization in a product space," *Mathematical Programming,* vol. 28, no. 1, pp. 96-115, January 1984.

[133] 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, Editors), pp. 60-66. Sverdlovsk: Akademiia Nauk SSSR, 1982.

[134] H. Poincaré, "Sur les équations aux dérivées partielles de la physique mathématique," *American Journal of Mathematics,* vol. 12, pp. 211-294, 1890.

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

[136] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics,* vol. 9, no. 3, pp. 14-29, 1969.

[137] B. T. Polyak, *Introduction to Optimization.* New York: Optimization Software Inc., 1987.

[138] W. K. Pratt, *Digital Image Processing,* 2nd ed. New York: Wiley, 1991.

[139] R. Ramaseshan and B. Yegnanarayana, "Image reconstruction from multiple frames of sparse data," *Multidimensional Systems and Signal Processing,* vol. 4, no. 2, pp. 167-179, April 1993.

[140] R. Rangayyan, A. P. Dhawan, and R. Gordon, "Algorithms for limited-view computed tomography: An annoted bibliography and a challenge," *Applied Optics,* vol. 24, no. 23, pp. 4000-4012, December 1985.

[141] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization,* vol. 14, no. 5, pp. 877-898, August 1976.

[142] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projections," *IEEE Transactions on Nuclear Science,* vol. 23, no. 4, pp. 1428-1432, August 1976.

[143] M. Rosenblatt, *Stationary Sequences and Random Fields.* Boston, MA: Birkhäuser, 1985.

[144] C. Sánchez-Avila, "An adaptive regularized method for deconvolution of signals with edges by convex projections," *IEEE Transactions on Signal Processing,* vol. 42, no. 7, pp. 1849-1851, July 1994.

[145] J. L. C. Sanz and T. S. Huang, "Unified Hilbert space approach to iterative least-squares linear signal restoration," *Journal of the Optical Society of America,* vol. 73, no. 11, pp. 1455-1465, November 1983.

[146] O. Sasaki and T. Yamagami, "Image restoration in singular vector space by the method of convex projections," *Applied Optics,* vol. 26, no. 7, pp. 1216-1221, April 1987.

[147] K. D. Sauer and J. P. Allebach, "Iterative reconstruction of band-limited images from nonuniformly spaced samples," *IEEE Transactions on Circuits and Systems,* vol. 34, no. 12, pp. 1497-1506, December 1987.

[148] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative restoration algorithms," *Proceedings of the IEEE,* vol. 69, no. 4, pp. 432-450, April 1981.

[149] L. Schwartz, *Analyse III - Calcul Intégral.* Paris: Hermann, 1993.

[150] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Transactions on Automatic Control,* vol. 13, no. 1, pp. 22-28, February 1968.

[151] M. I. Sezan and H. Stark, "Image restoration by the method of convex projections: Part 2 - applications and numerical results," *IEEE Transactions on Medical Imaging,* vol. 1, no. 2, pp. 95-101, October 1982. ("Correction," vol. 1, no. 3, p. 204, December 1982.)

[152] M. I. Sezan and H. Stark, "Image restoration by convex projections in the presence of noise," *Applied Optics,* vol. 22, no. 18, pp. 2781-2789, September 1983.

[153] M. I. Sezan and H. Stark, "Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion," *IEEE Transactions on Medical Imaging,* vol. 3, no. 2, pp. 91-98, June 1984.

[154] M. I. Sezan and H. Stark, "Incorporation of a priori moment information into signal recovery and synthesis problems," *Journal of Mathematical Analysis and Applications,* vol. 122, no. 1, pp. 172-186, February 1987.

[155] M. I. Sezan and A. M. Tekalp, "Adaptive image restoration with artifact suppression using the theory of convex projections," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 38, no. 1, pp. 181-185, January 1990.

[156] M. I. Sezan and A. M. Tekalp, "Survey of recent developments in digital image restoration," *Optical Engineering,* vol. 29, no. 5, pp. 393-404, May 1990.

[157] M. I. Sezan and H. J. Trussell, "Prototype image constraints for set-theoretic image restoration," *IEEE Transactions on Signal Processing,* vol. 39, no. 10, pp. 2275-2285, October 1991.

[158] P. Y. Simard and G. E. Mailloux, "Vector field restoration by the method of convex projections," *Computer Vision, Graphics, and Image Processing,* vol. 52, no. 3, pp. 360-385, December 1990.

[159] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, CA: Academic Press, 1987.

[160] H. Stark, D. Cahana, and H. Webb, "Restoration of arbitrary finite energy optical objects from limited spatial and spectral information," *Journal of the Optical Society of America,* vol. 71, no. 6, pp. 635-642, June 1981.

[161] H. Stark and E. T. Olsen, "Projection-based image restoration," *Journal of the Optical Society of America A,* vol. 9, no. 11, pp. 1914-1919, November 1992.

[162] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A,* vol. 6, no. 11, pp. 1715-1726, November 1989.

[163] W. J. Stiles, "Closest point maps and their product II," *Nieuw Archief voor Wiskunde,* vol. 13, no. 3, pp. 212-225, November 1965.

[164] A. M. Tekalp and H. J. Trussell, "Comparative study of some statistical and set-theoretic methods for image restoration," *CVGIP: Graphical Models and Image Processing,* vol. 53, no. 2, pp. 108-120, March 1991.

[165] V. T. Tom, T. F. Quatieri, M. H. Hayes, and J. H. McClellan, "Convergence of iterative nonexpansive signal reconstruction algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 29, no. 5, pp. 1052-1058, October 1981.

[166] H. J. Trussell, "Notes on linear image restoration by maximizing the *a posteriori* probability," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 26, no. 2, pp. 174-176, April 1978.

[167] H. J. Trussell, "The relationship between image restoration by maximum *a posteriori* and maximum entropy methods," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 28, no. 1, pp. 114-117, February 1980.

[168] H. J. Trussell, "Maximum power signal restoration," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 29, no. 5, pp. 1059-1061, October 1981.

[169] H. J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 31, no. 1, pp. 129-136, February 1983.

[170] H. J. Trussell, "*A priori* knowledge in algebraic reconstruction methods," in *Advances in Computer Vision and Image Processing* (T. S. Huang, Ed.), vol. 1, pp. 265-316. Greenwich, CT: JAI Press, 1984.

[171] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 32, no. 2, pp. 201-212, April 1984.

[172] H. J. Trussell, H. Orun-Ozturk, and M. R. Civanlar, "Errors in reprojection methods in computerized tomography," *IEEE Transactions on Medical Imaging,* vol. 6, no. 3, pp. 220-227, September 1987.

[173] H. J. Trussell and P. L. Vora, "Bounds on restoration quality using a priori information," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp. 1758-1761. New York, NY, April 11-14, 1988.

[174] P. Tseng, "On the convergence of products of firmly nonexpansive mappings," *SIAM Journal on Optimization,* vol. 2, no. 3, pp. 425-434, August 1992.

[175] P. Tseng and D. P. Bertsekas, "Relaxation methods for problems with strictly convex separable costs and linear constraints," *Mathematical Programming,* vol. 38, no. 3, pp. 303-321, 1987.

[176] J. Von Neumann, "On rings of operators. Reduction theory," *Annals of Mathematics,* vol. 50, no. 2, pp. 401-485, April 1949 (the result of interest first appeared in 1933 in lecture notes).

[177] S. J. Wernecke and L. R. D'Addario, "Maximum entropy image reconstruction," *IEEE Transactions on Computers,* vol. 26, no. 4, pp. 351-364, April 1977.

[178] M. N. Wernick and C. T. Chen, "Superresolved tomography by convex projections and detector motion," *Journal of the Optical Society of America A,* vol. 9, no. 9, pp. 1547-1553, September 1992.

[179] R. Wittmann, "Approximation of fixed points of nonexpansive mappings," *Archiv der Mathematik,* vol. 58, no. 5, pp. 486-491, May 1992.

[180] S. J. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *Journal of the Optical Society of America A,* vol. 7, no. 3, pp. 491-499, March 1990.

[181] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Transactions on Circuits and Systems,* vol. 25, no. 9, pp. 694-702, September 1978.

[182] D. C. Youla, "On deterministic convergence of iterations of relaxed projection operators," *Journal of Visual Communication and Image Representation,* vol. 1, no. 1, pp. 12-20, September 1990.

[183] D. C. Youla and V. Velasco, "Extensions of a result on the synthesis of signals in the presence of inconsistent constraints," *IEEE Transactions on Circuits and Systems,* vol. 33, no. 4, pp. 465-468, April 1986.

[184] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 - theory," *IEEE Transactions on Medical Imaging,* vol. 1, no. 2, pp. 81-94, October 1982.

[185] H. T. Yura and S. G. Hanson, "Second-order statistics for wave propagation through complex optical systems," *Journal of the Optical Society of America A,* vol. 6, no. 4, pp. 564-575, April 1989.

[186] L. A. Zadeh, "What is optimal?," *IRE Transactions on Information Theory,* vol. 4, no. 1, p. 3, March 1958.

[187] E. H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory," in *Contributions to Nonlinear Functional Analysis,* (E. H. Zarantonello, Editor) pp. 237-424. Academic Press: New York, 1971.

[188] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems,* 2nd ed. New York: Springer-Verlag, 1993.

[189] E. Zeidler, *Nonlinear Functional Analysis and Its Applications II/A: Linear Monotone Operators.* New York: Springer-Verlag, 1990.

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

List of Figures

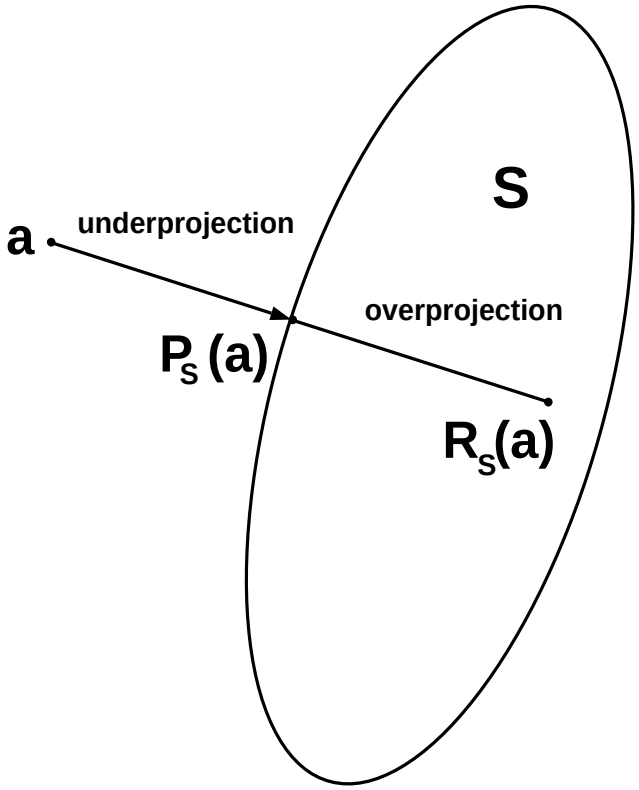

Figure 1: Relaxed projection.

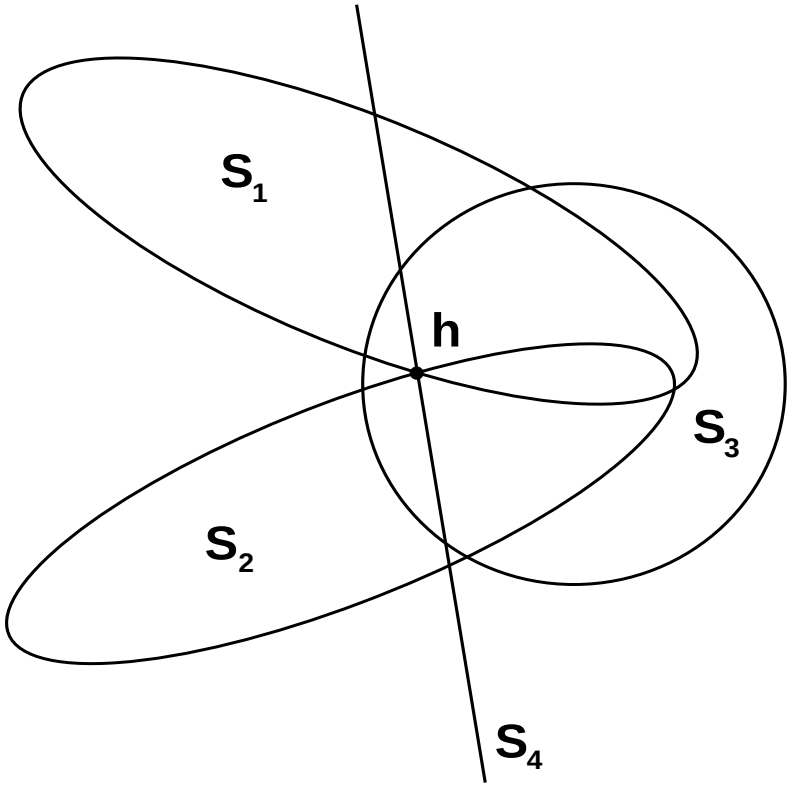

Figure 2: Ideal set theoretic formulation.

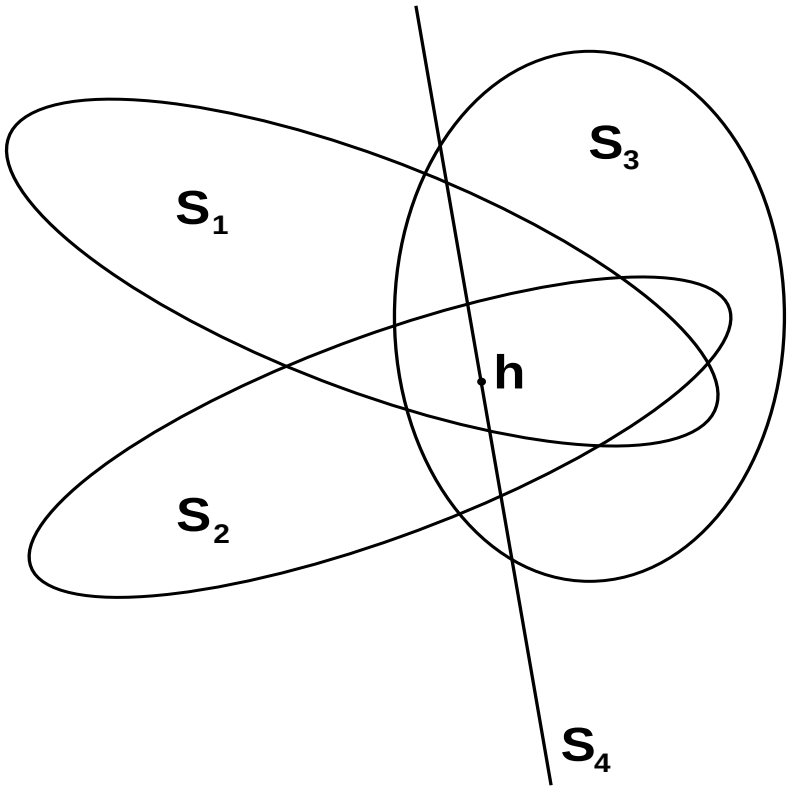

Figure 3: Fair set theoretic formulation.

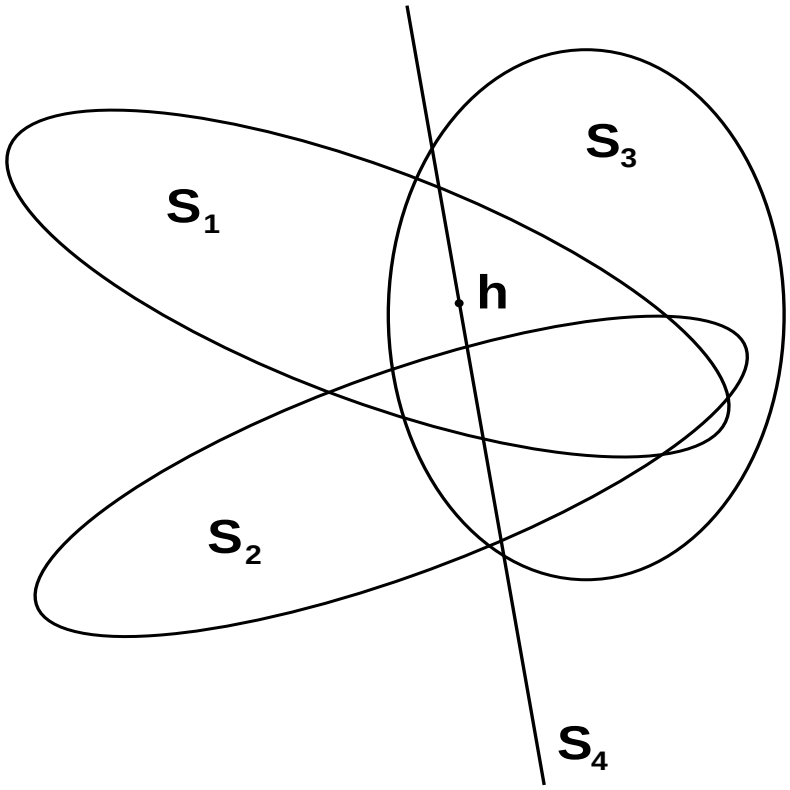

Figure 4: Unfair set theoretic formulation.

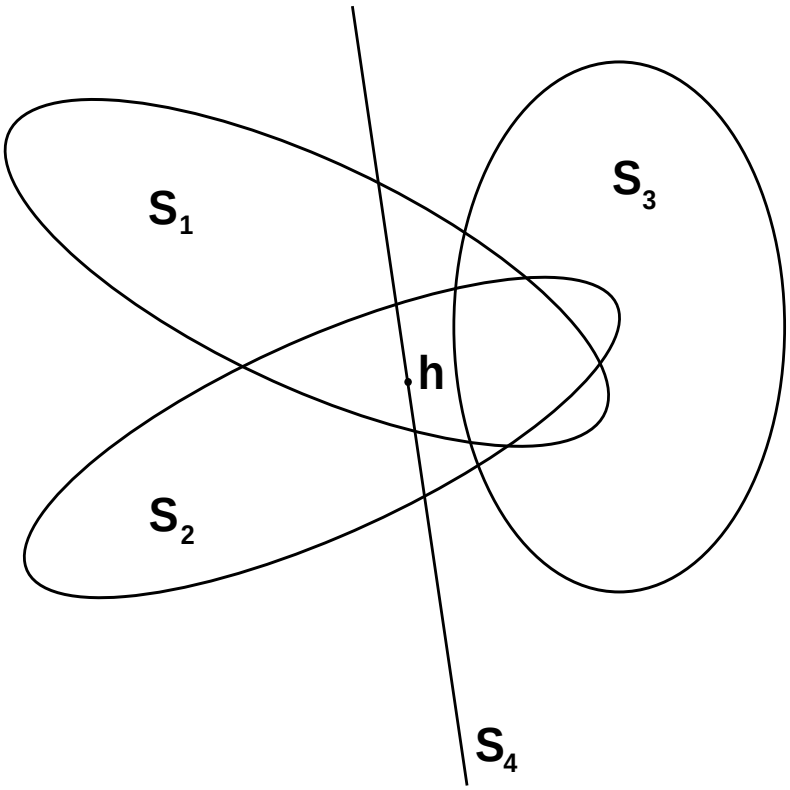

Figure 5: Inconsistent set theoretic formulation.

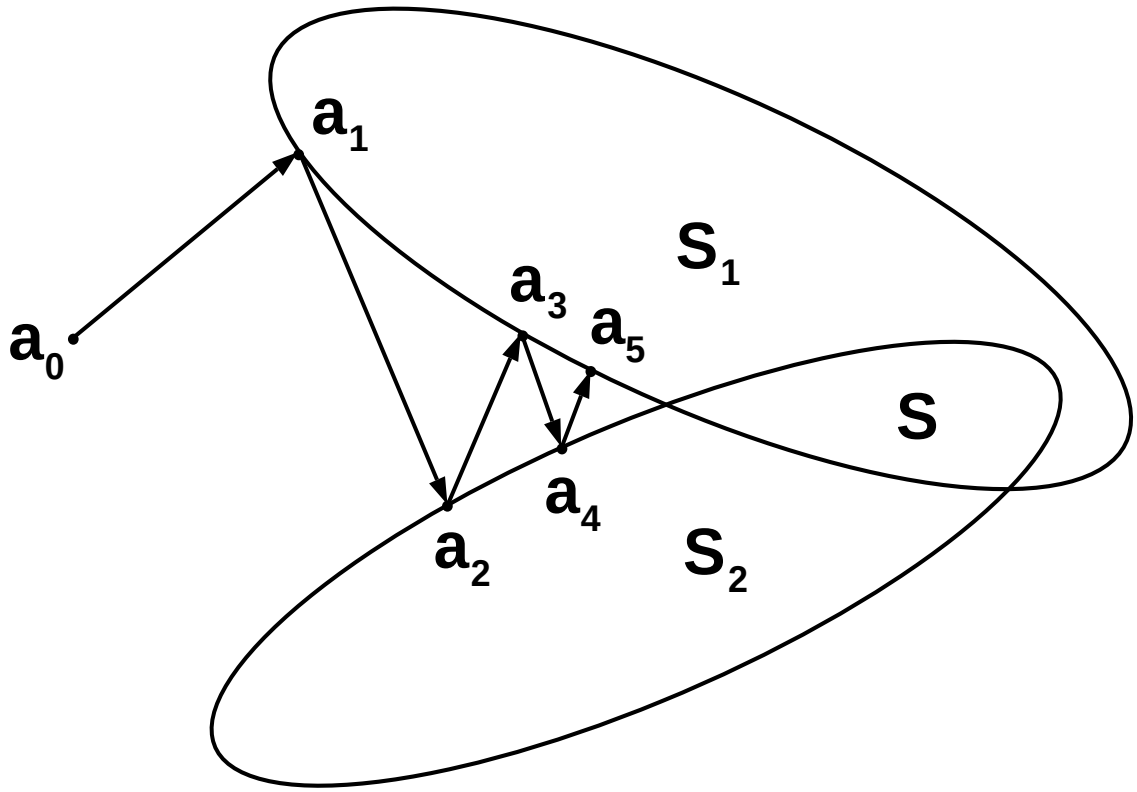

Figure 6: Unrelaxed POCS algorithm ($\lambda_n = 1.0$).

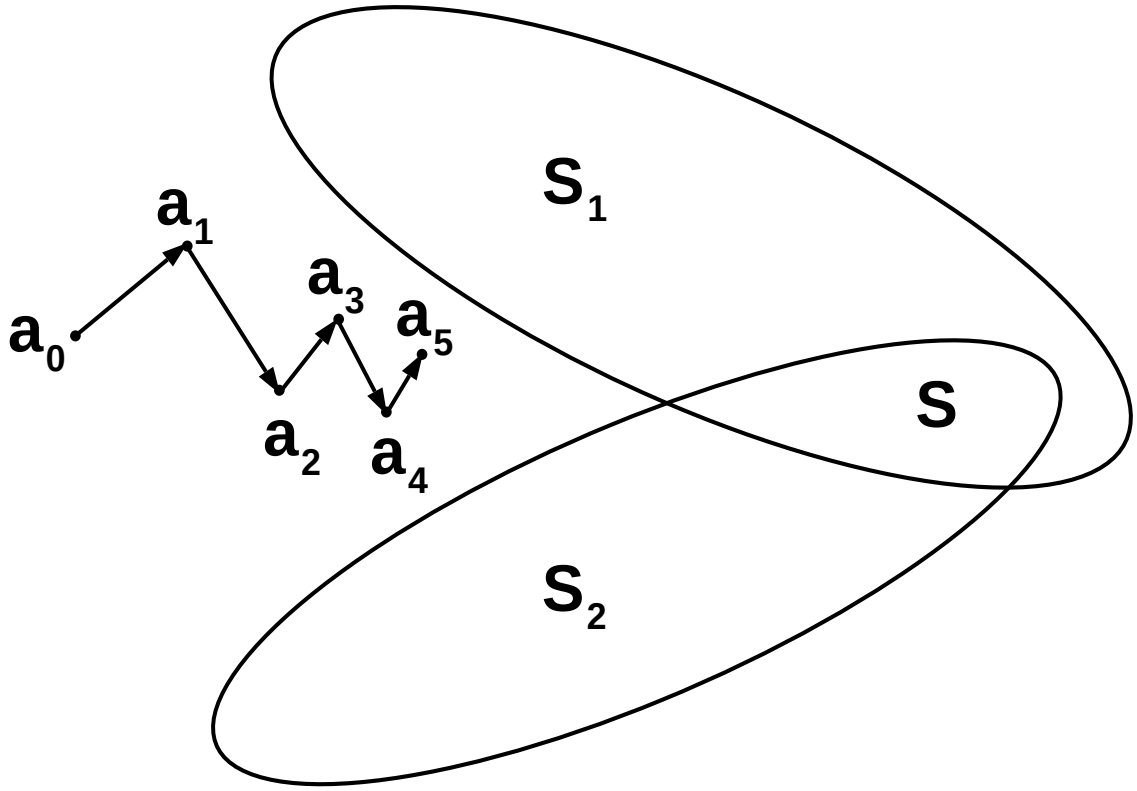

Figure 7: Underrelaxed POCS algorithm ($\lambda_n = 0.5$).

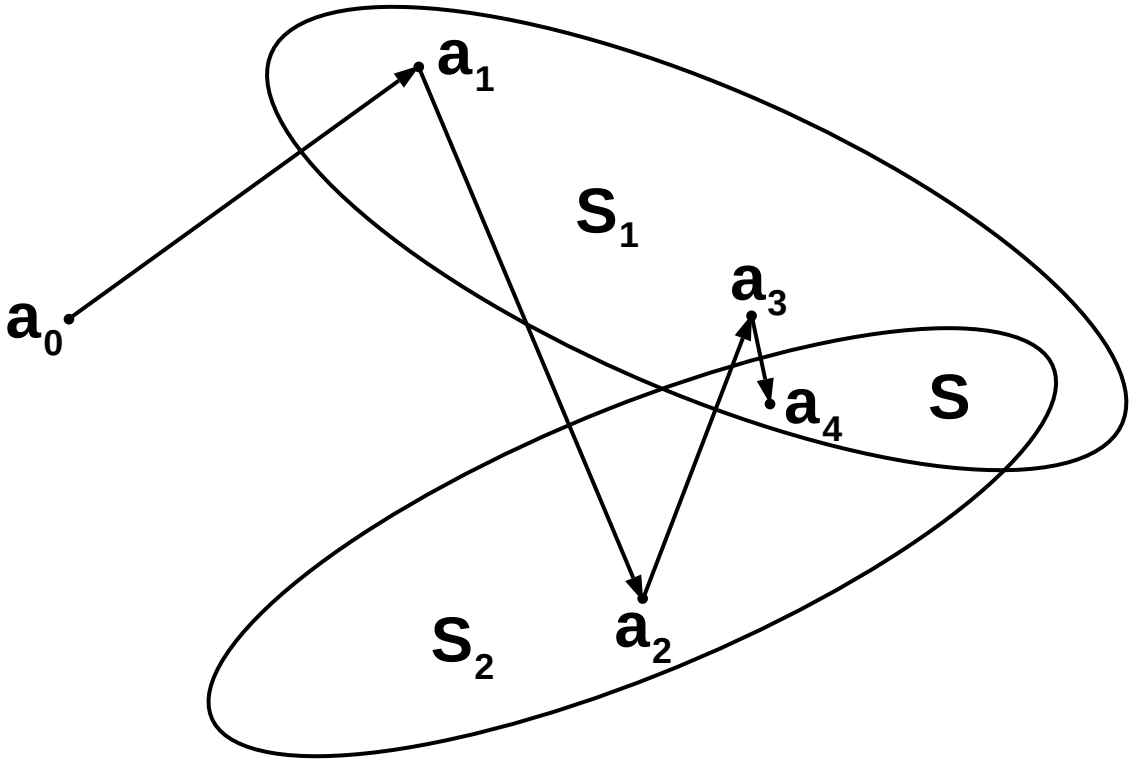

Figure 8: Overrelaxed POCS algorithm ($\lambda_n = 1.5$).

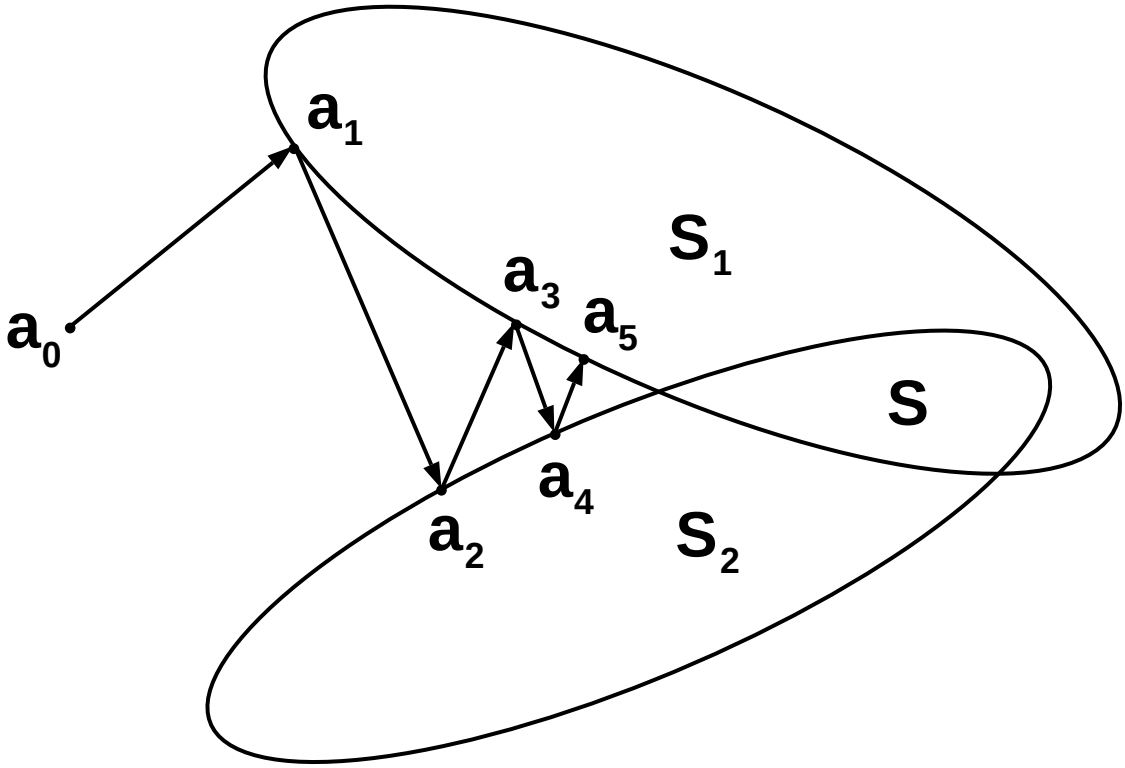

Figure 9: PPM algorithm in the product space.

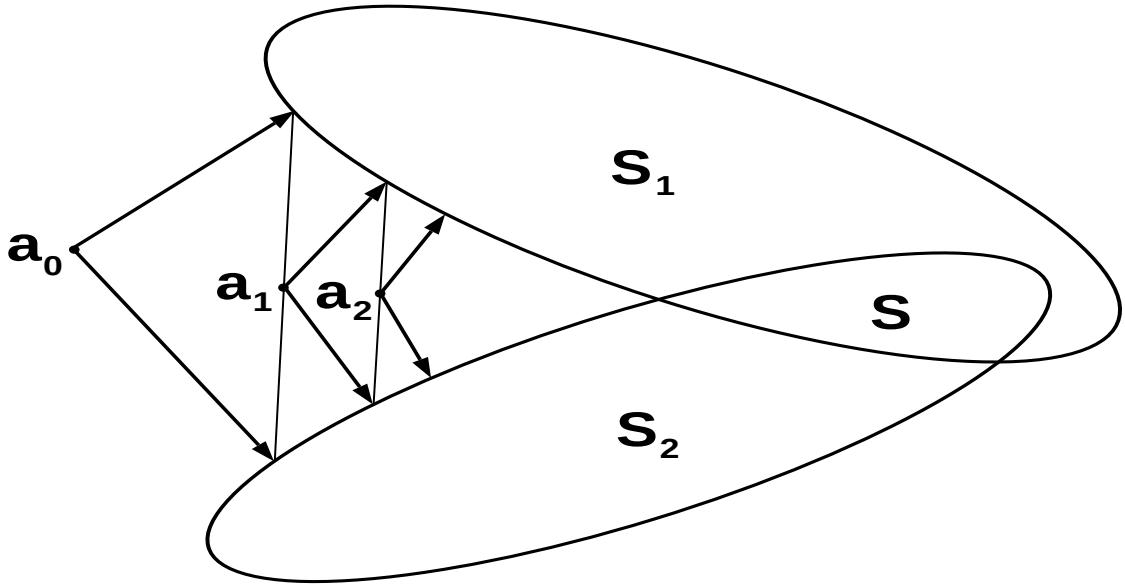

Figure 10: PPM algorithm in the orignal space. ©1994 IEEE.

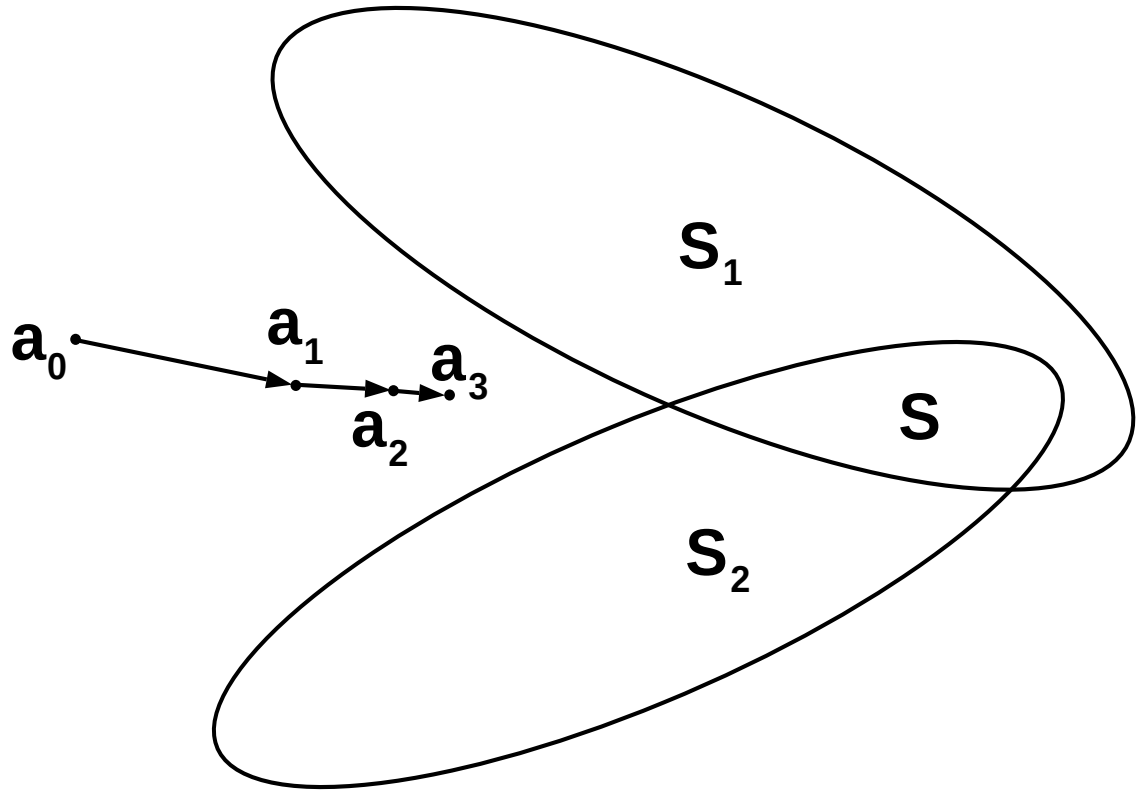

Figure 11: SIRT algorithm.

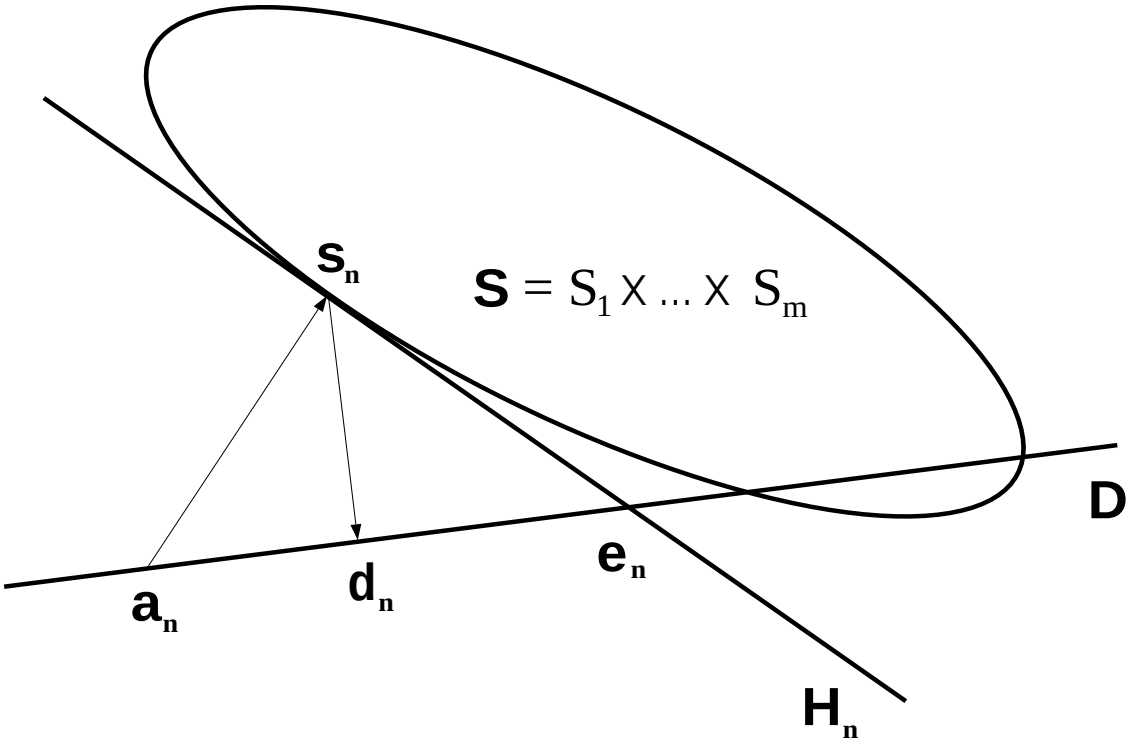

Figure 12: EPPM algorithm in the product space.

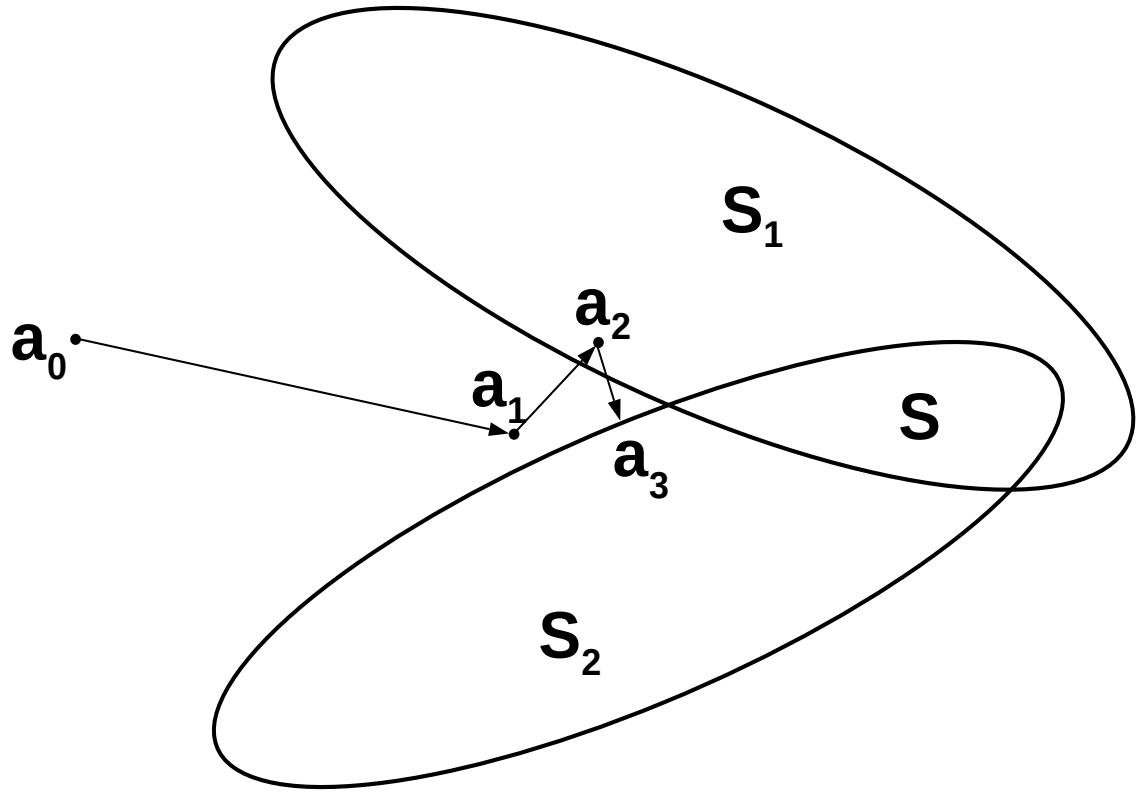

Figure 13: EPPM algorithm in the original space.

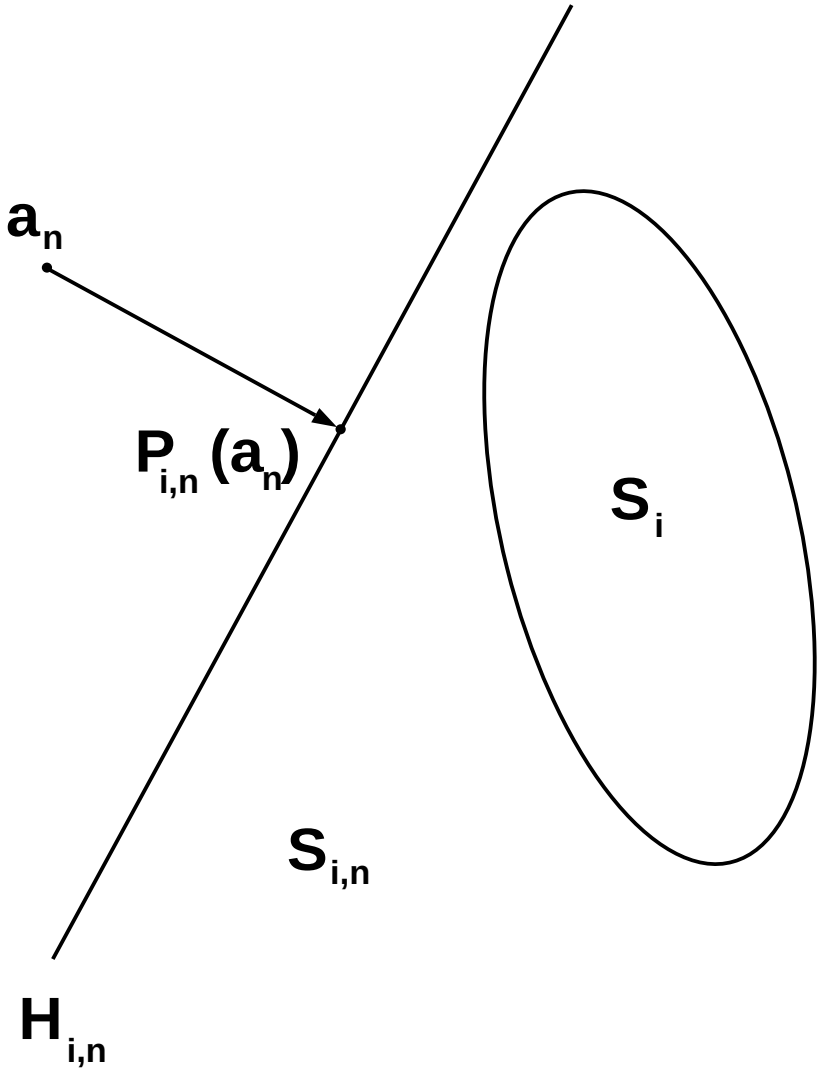

Figure 14: Projection onto a separating hyperplane.

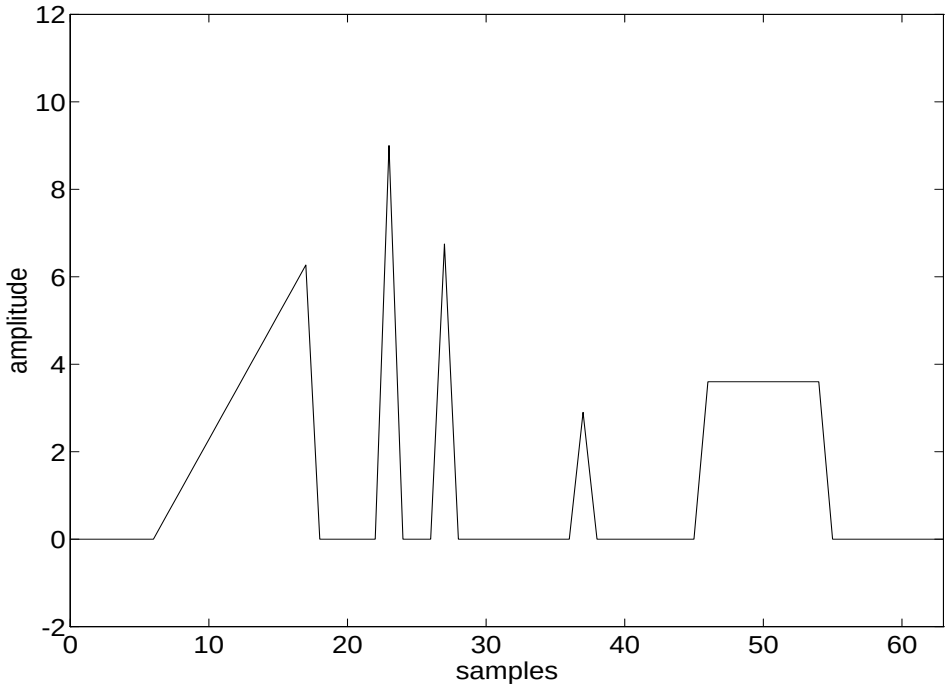

Figure 15: Original signal. ©1994 IEEE.

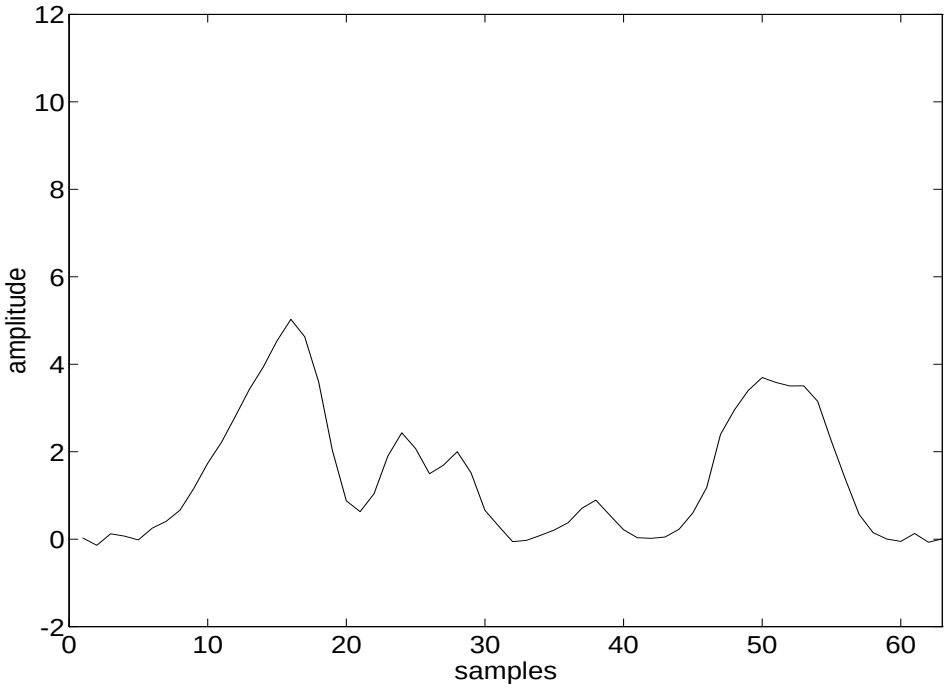

Figure 16: Degraded signal. ©1994 IEEE.

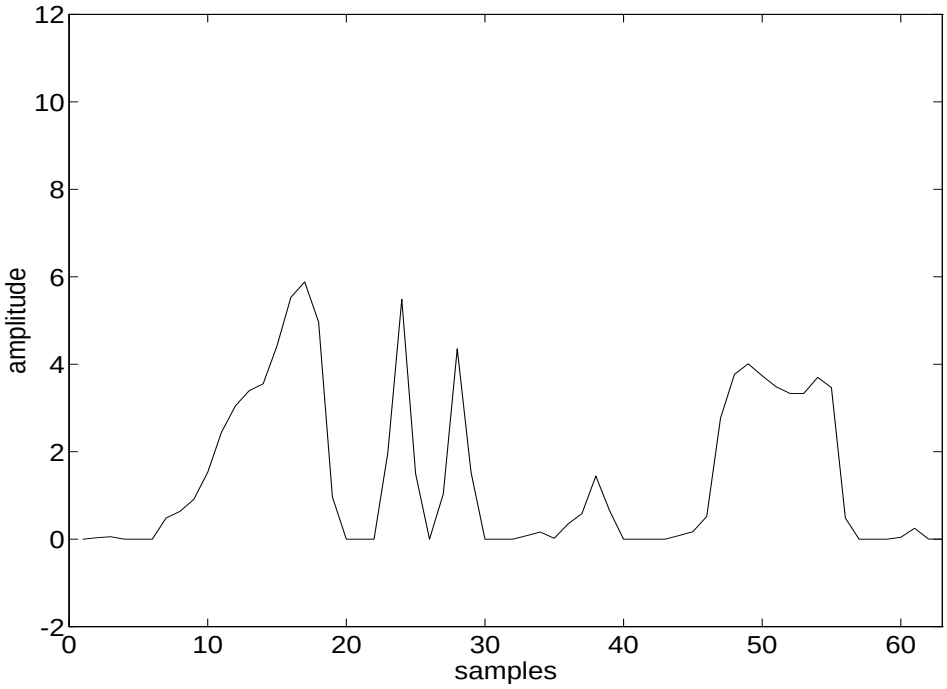

Figure 17: Consistent case - Deconvolution by POCS. ©1994 IEEE.

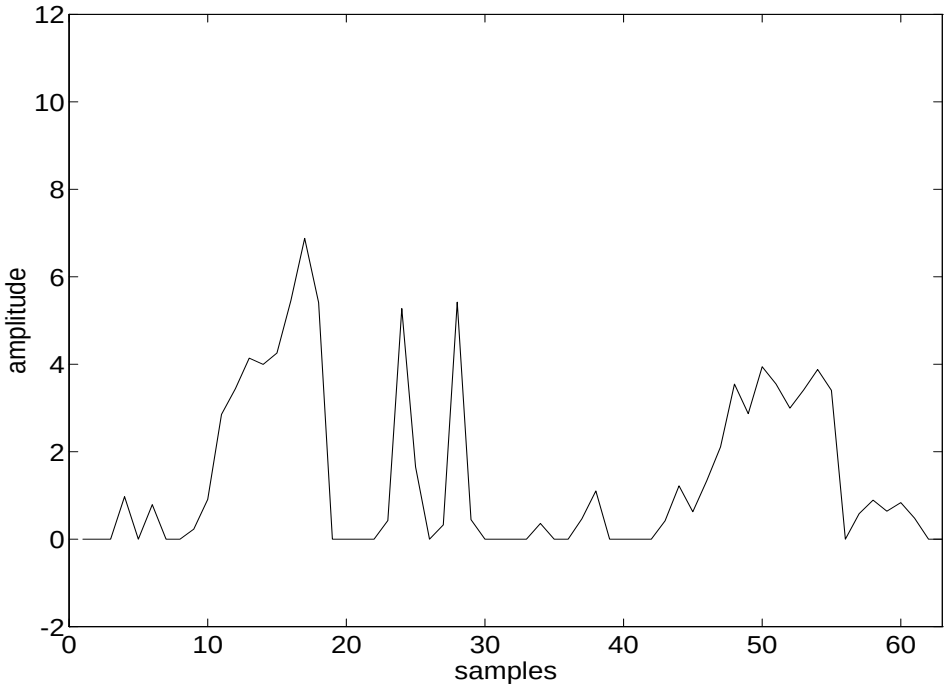

Figure 18: Inconsistent case - Deconvolution by POCS. ©1994 IEEE.

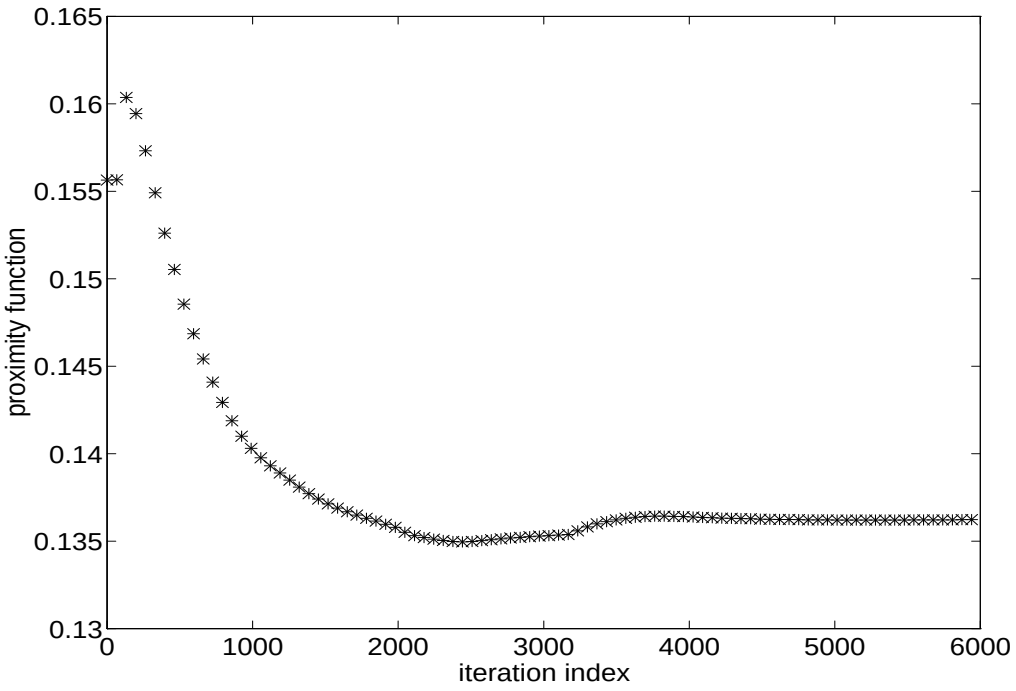

Figure 19: Inconsistent case - Convergence of POCS. ©1994 IEEE.

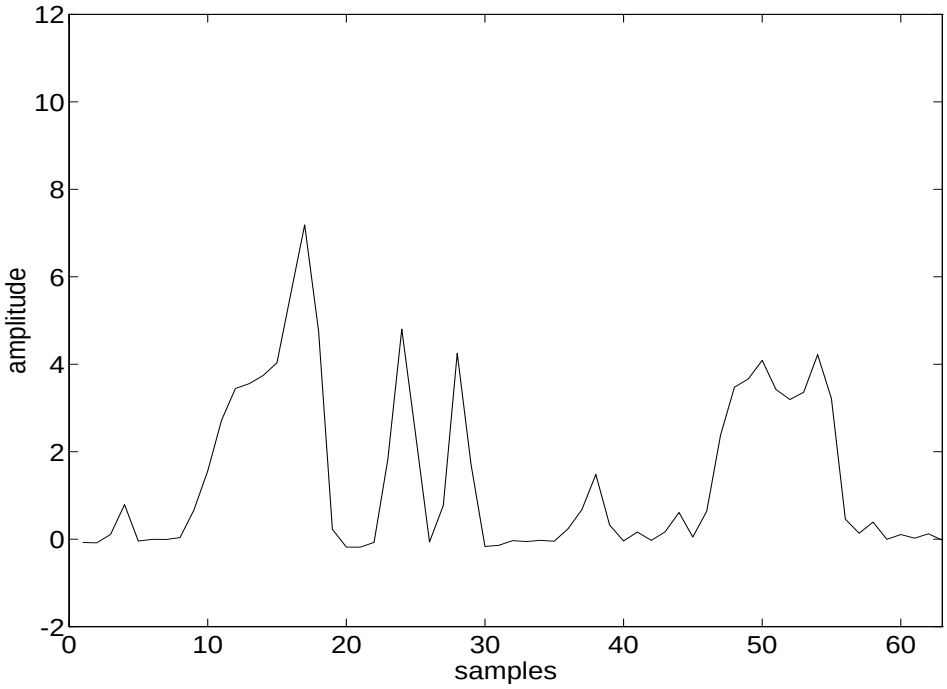

Figure 20: Consistent case - Deconvolution by PPM. ©1994 IEEE.

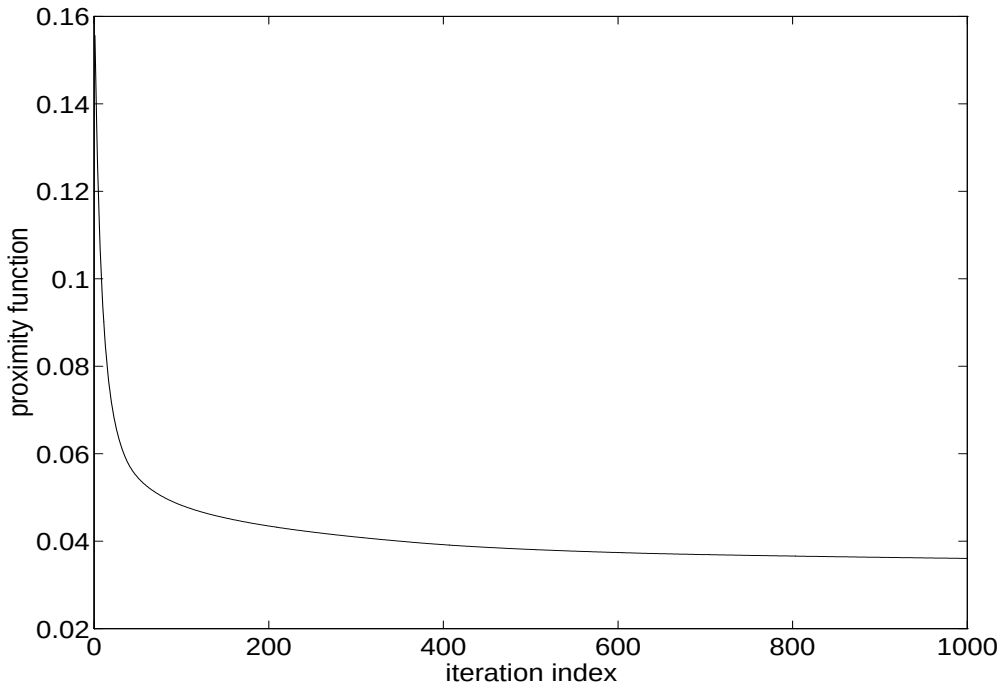

Figure 21: Inconsistent case - Convergence of PPM. ©1994 IEEE.

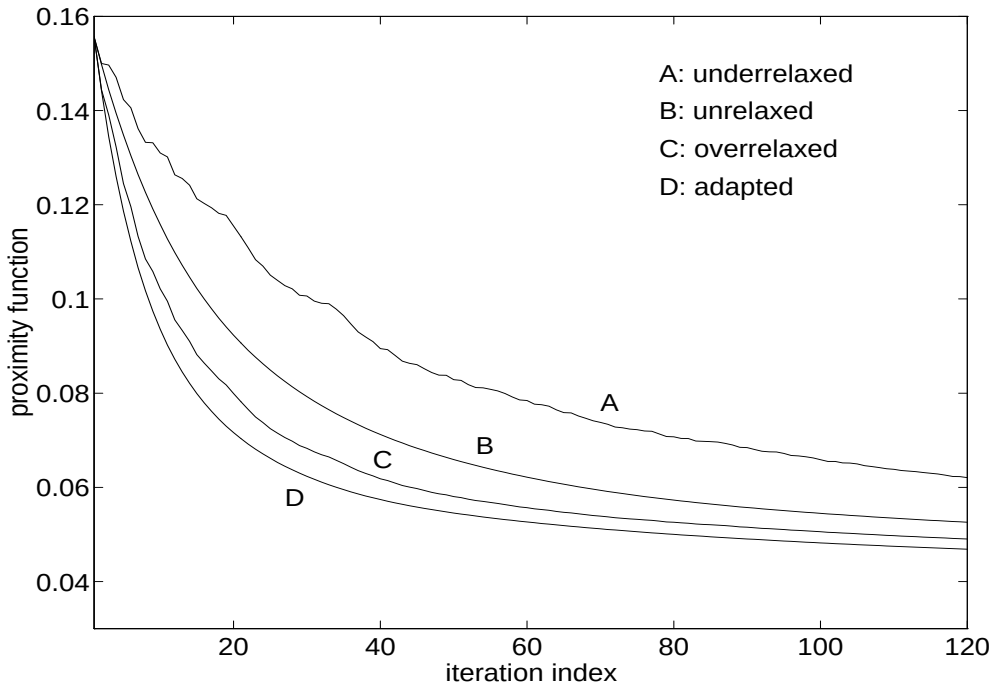

Figure 22: Convergence of PPM for various relaxations schemes. ©1994 IEEE.

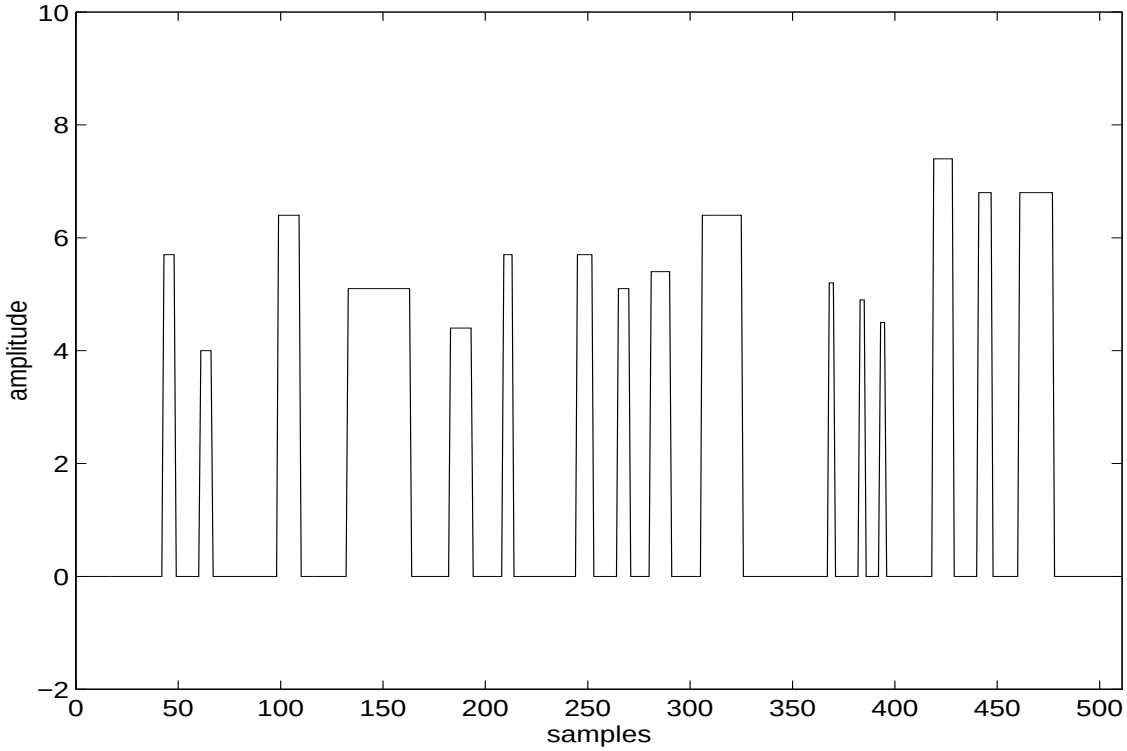

Figure 23: Original signal.

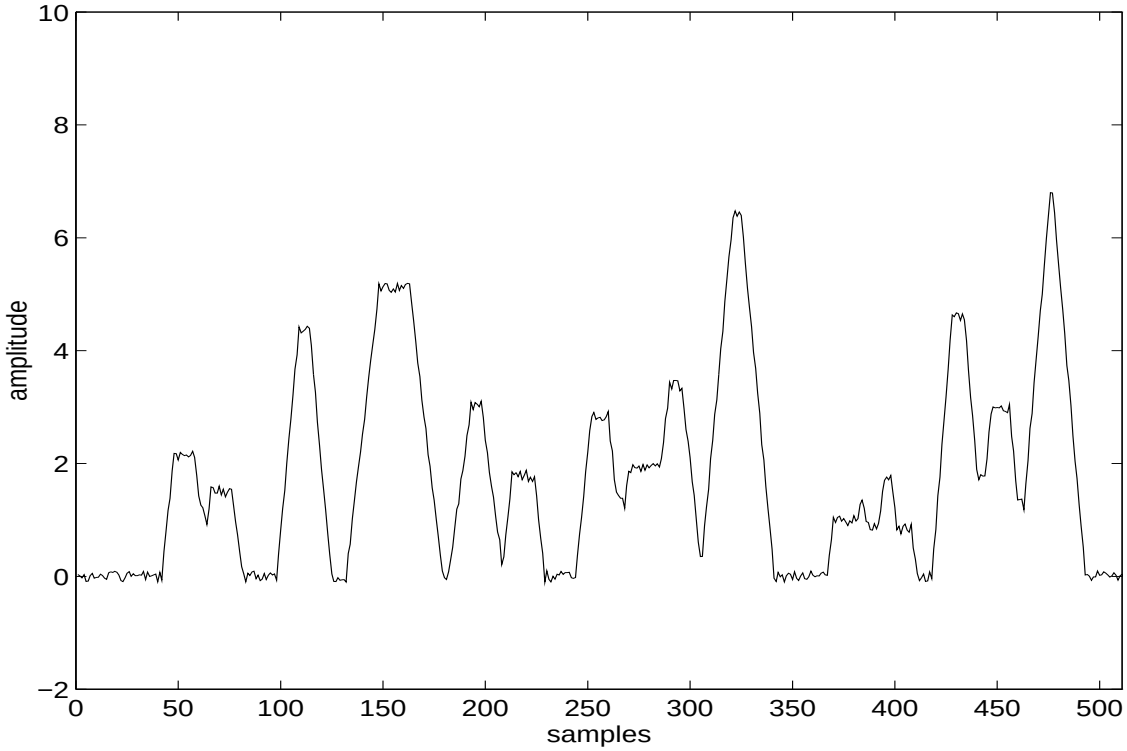

Figure 24: Degraded signal - Known blur.

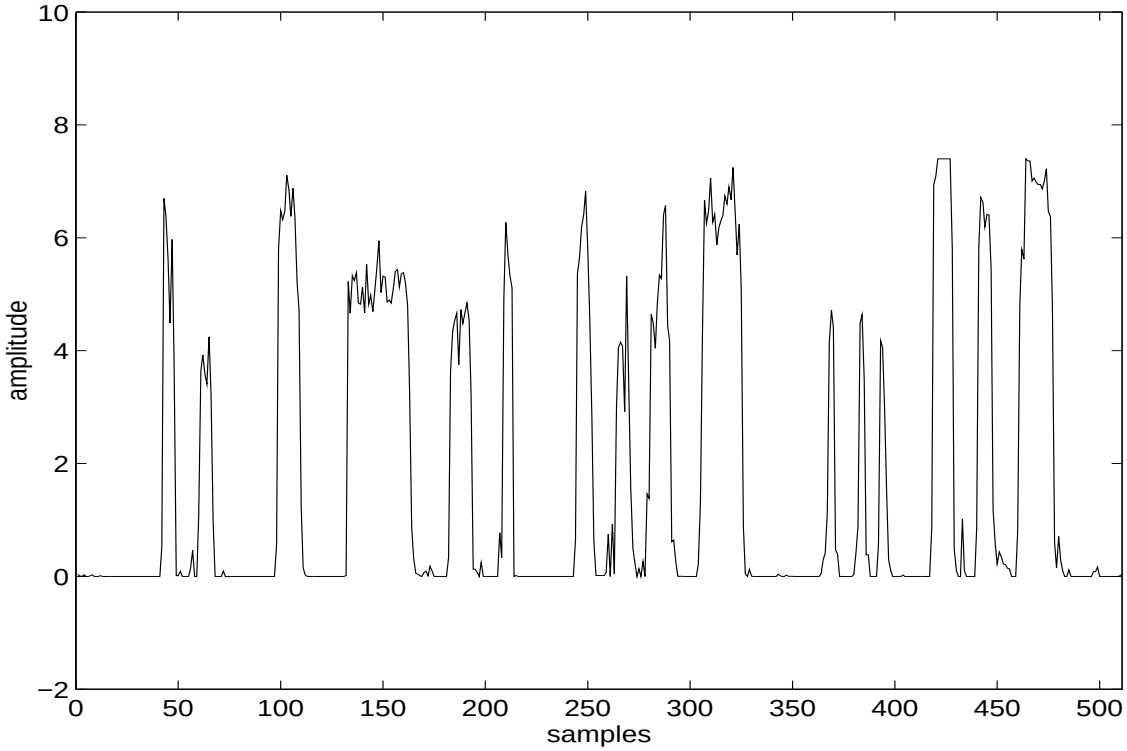

Figure 25: Deconvolved signal - Known blur.

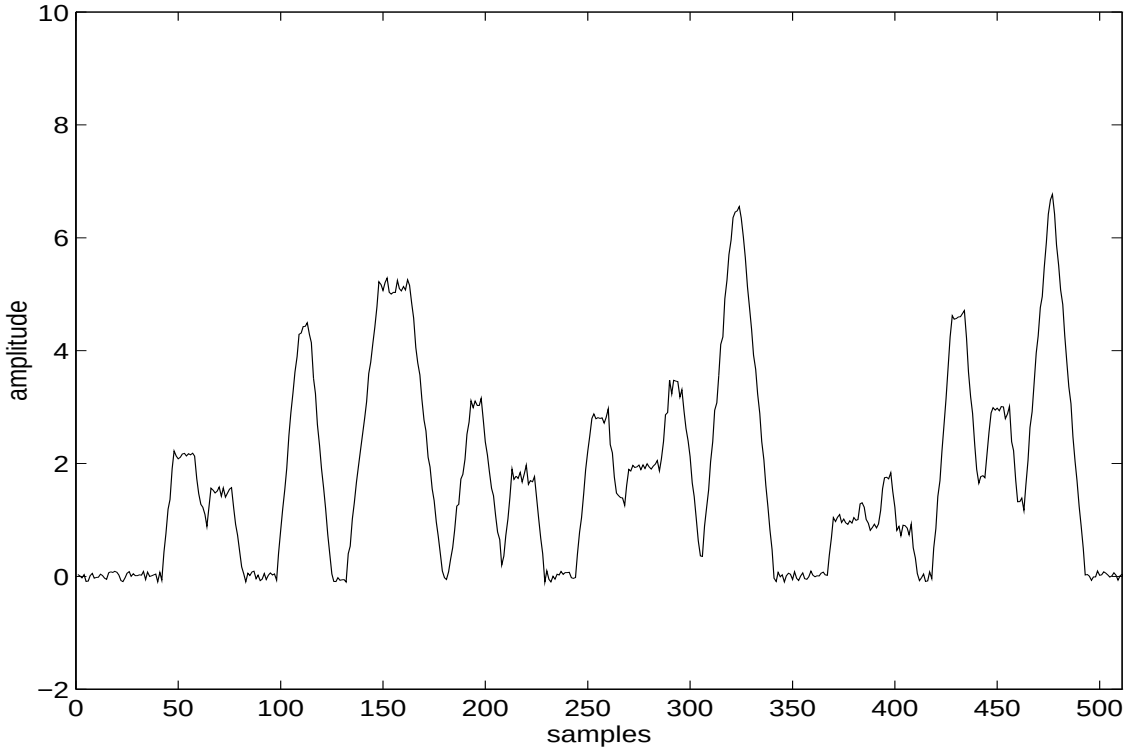

Figure 26: Degraded signal - Perturbed blur.

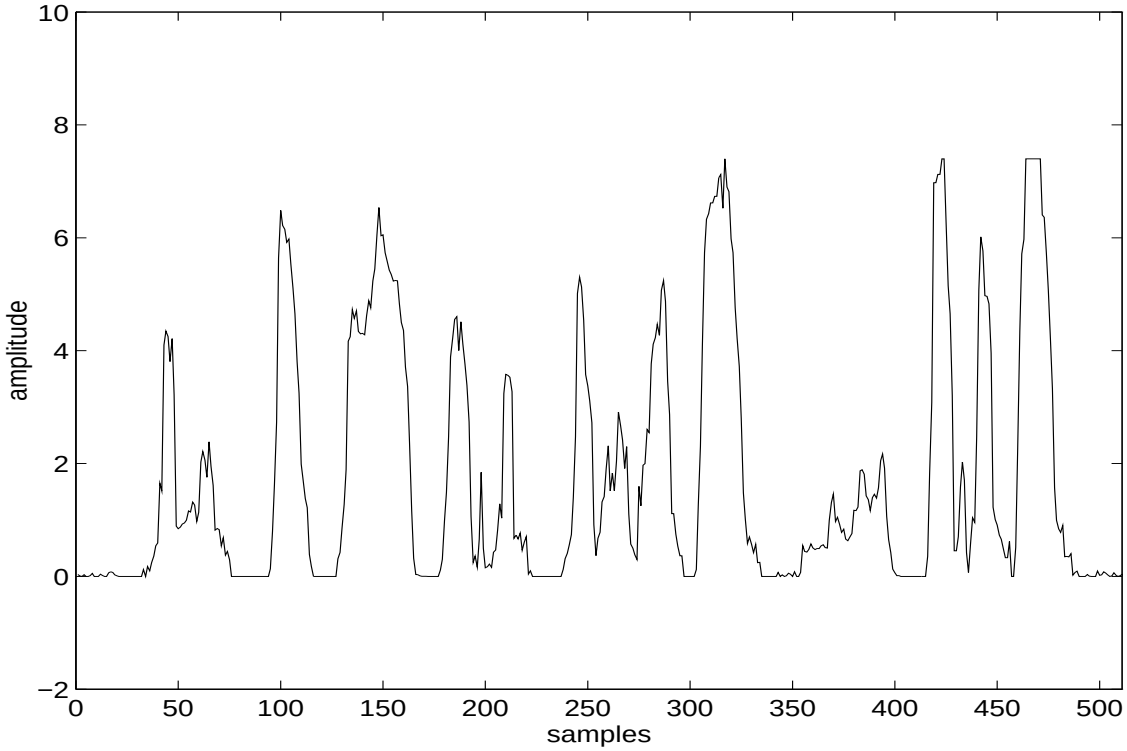

Figure 27: Deconvolved signal - Perturbed blur.

Figure 28: Original image.

Figure 29: Degraded image - Bounded noise.

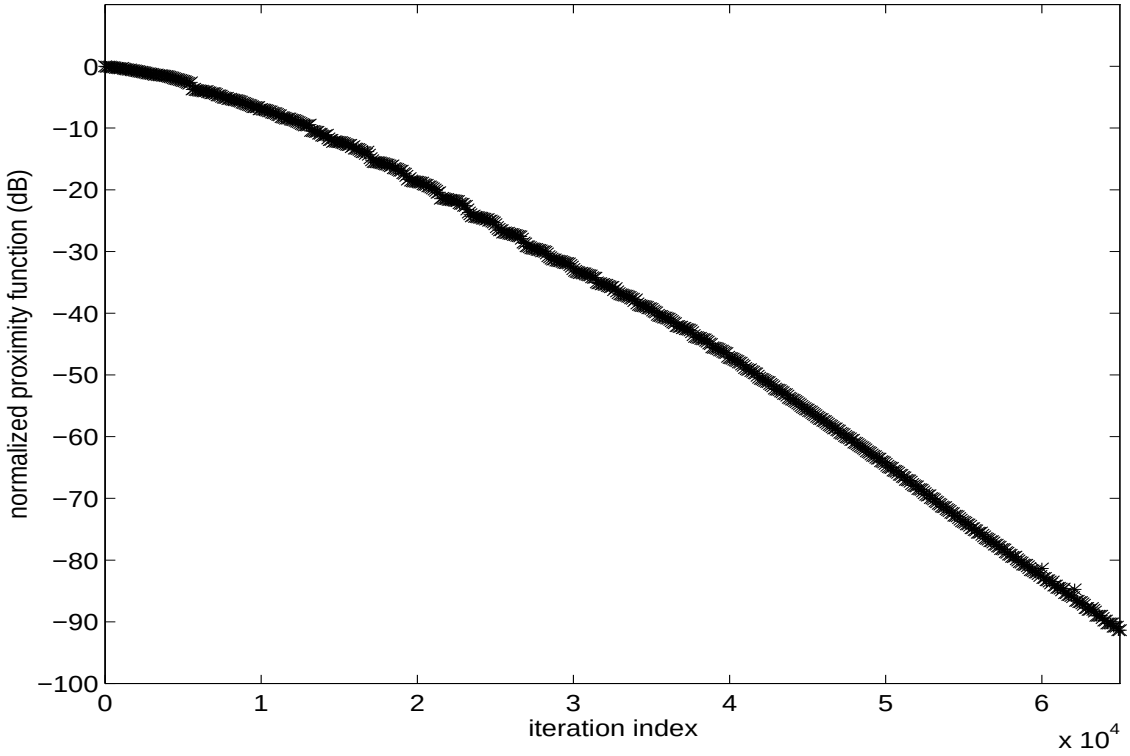

Figure 30: Convergence of POCS.

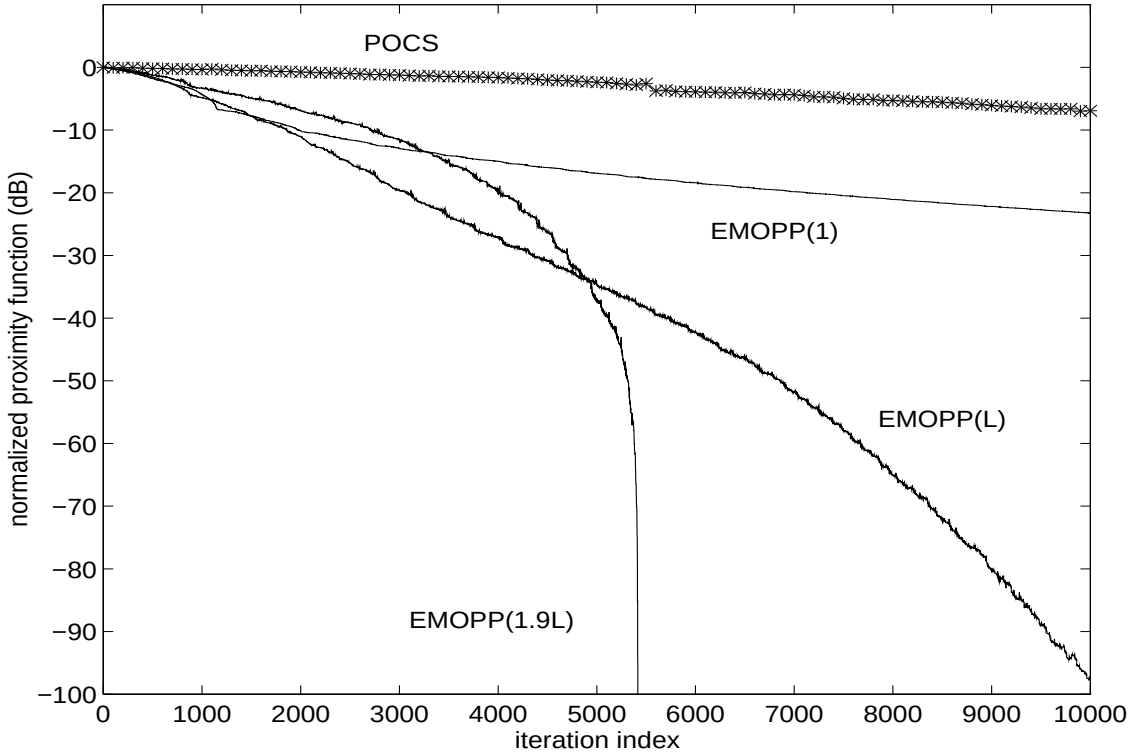

Figure 31: Convergence of EMOPP - 8 parallel processsors.

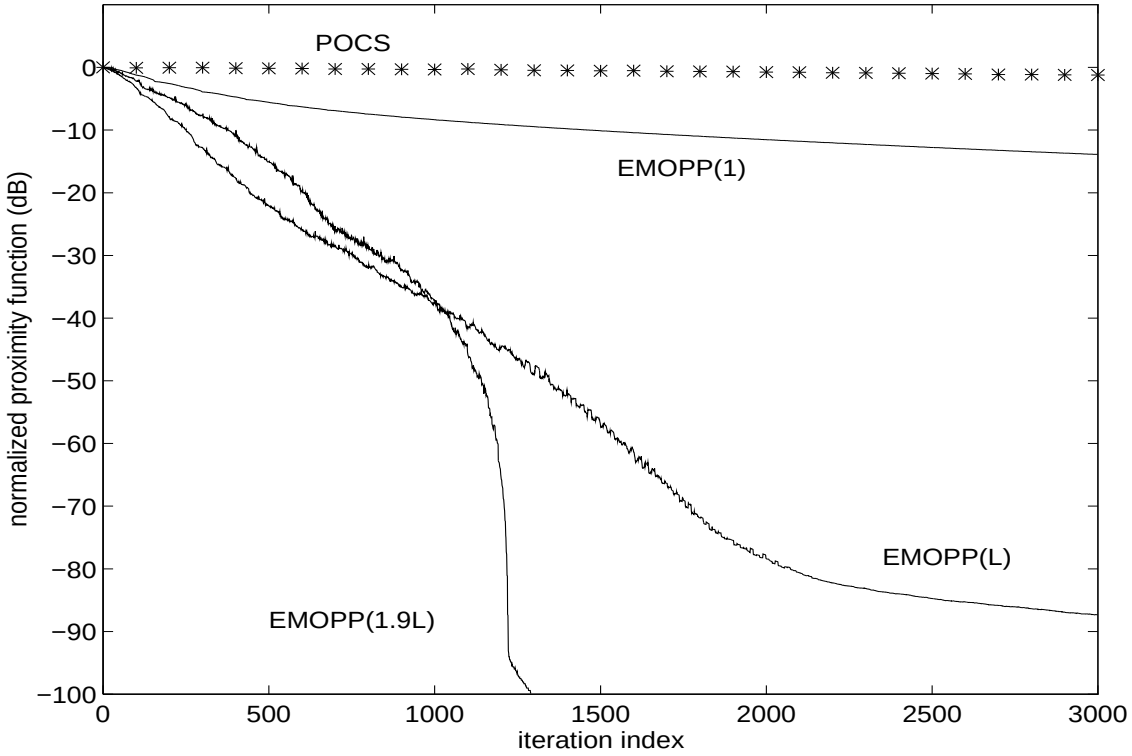

Figure 32: Convergence of EMOPP - 64 parallel processsors.

Figure 33: Restored image - Bounded noise.

Figure 34: Degraded image - Gaussian noise.

Figure 35: Restored image - Gaussian noise.

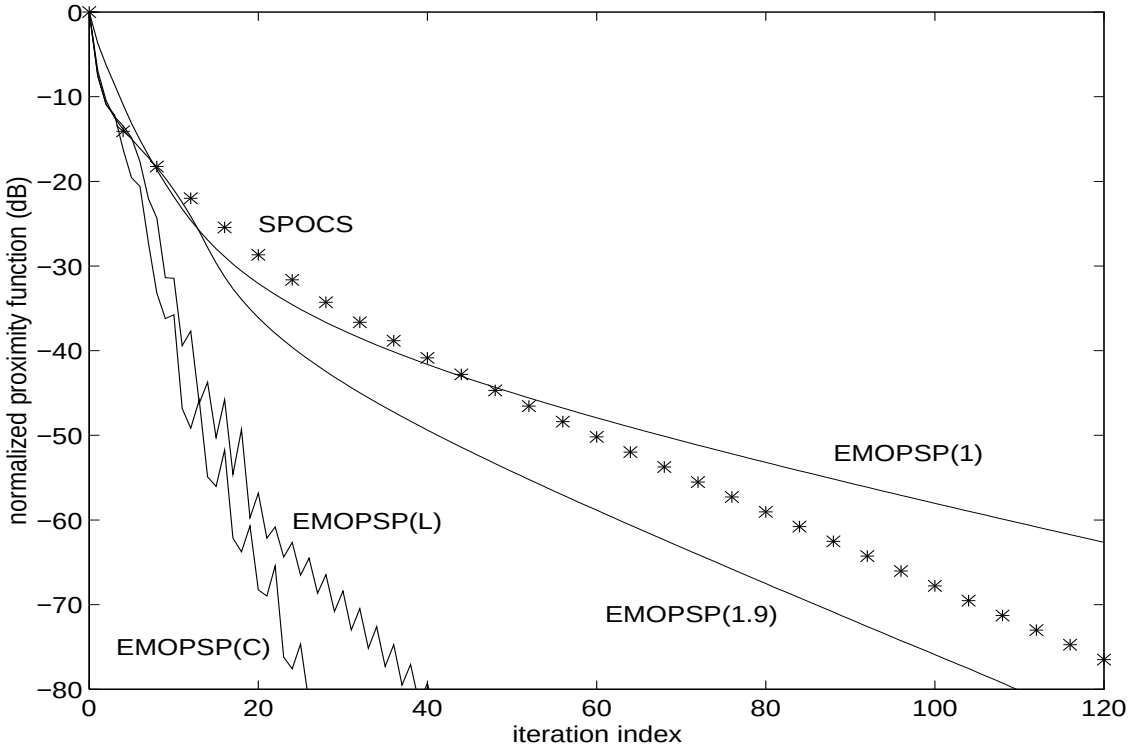

Figure 36: Convergence of subgradient methods.

Figure 37: Restored image.

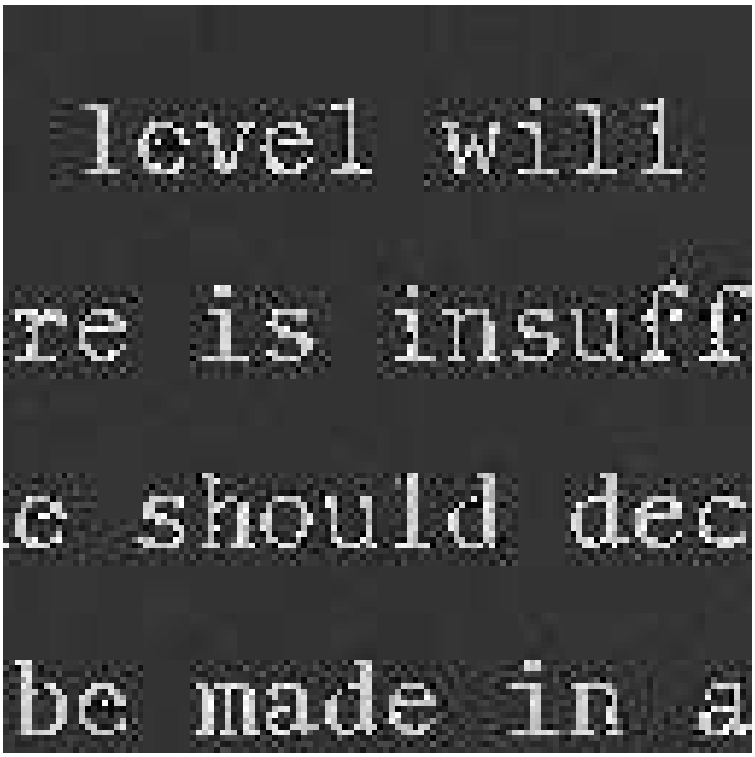

Figure 38: Original image.

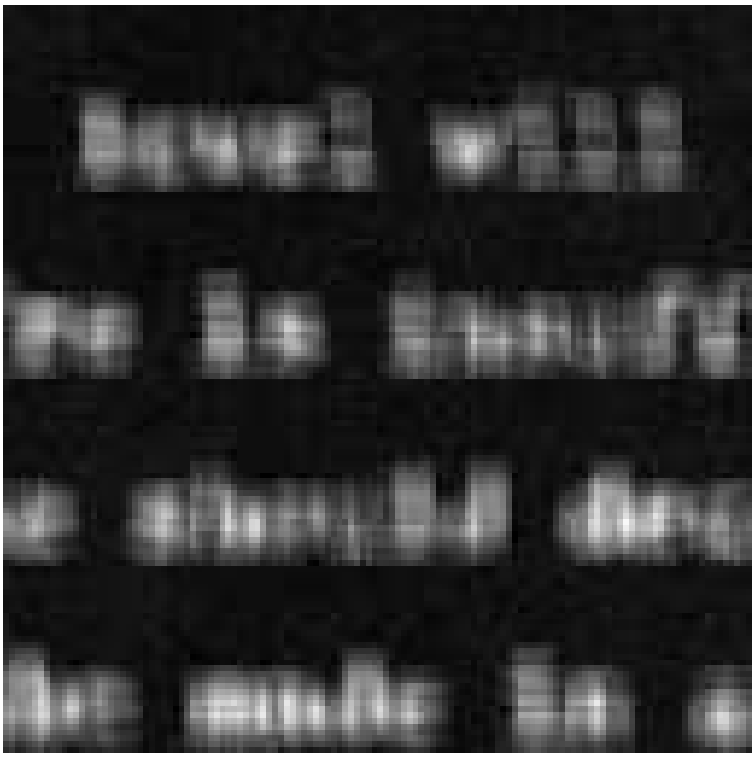

Figure 39: Degraded image.

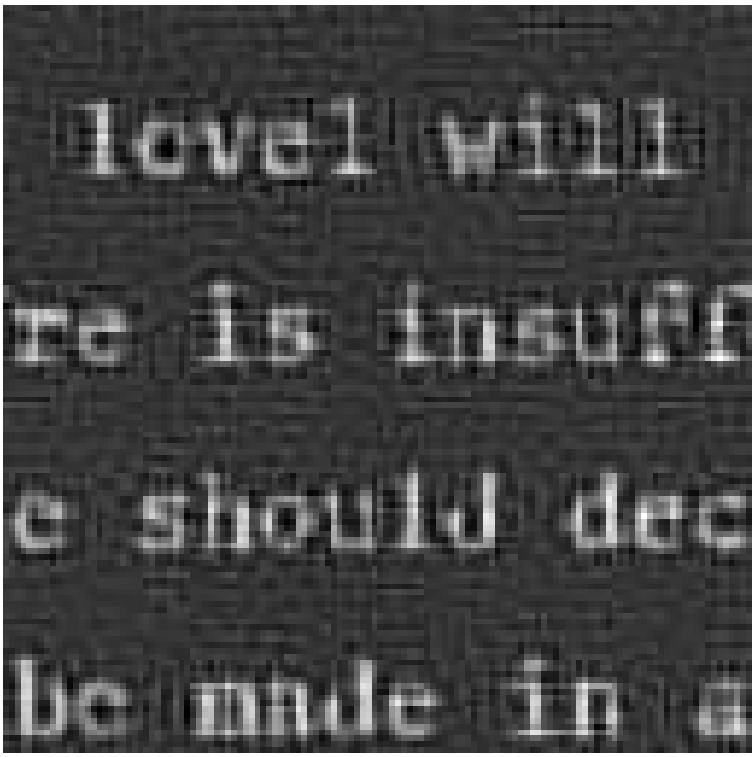

Figure 40: Restored image.