ÉCOLE D'ÉTÉ FRANCO-VIETNAMIENNE — FRENCH-VIETNAMESE SUMMER SCHOOL

MATHÉMATIQUES DU TRAITEMENT DE L'IMAGE
MATHEMATICS OF IMAGE PROCESSING

Institut de Mathématique — Institute of Mathematics
18 Hoang Quoc Viet Road, Cau Giay District
10307 Hanoi, Vietnam
10-18 septembre 2004 — 10-18 September 2004

FIXED POINT METHODS IN IMAGE RECOVERY

P. L. Combettes
Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie – Paris 6
4, Place Jussieu, 75005 Paris, France

Contents

1 Image recovery: An overview 3

2 Classical linear image recovery methods 7

2.1 Inverse filtering 7

2.2 Wiener filtering 8

3 Hilbertian convex analysis and fixed point theory: A primer 11

3.1 Basic facts 14

3.2 Convex sets 14

3.3 Convex functions 16

3.4 Nonlinear operators and fixed points 17

3.5 Properties of $\mathfrak{T}$-class operators 18

3.6 Projection onto a convex set 19

3.7 Subgradient projection onto a level set 20

3.8 Proximity operator 21

4 Fixed point formulation of image recovery problems 22

5 Solving feasibility problems 23

5.1 Cyclic projection method (POCS) 23

5.2 Cyclic subgradient projection method 25

5.3 General $\mathfrak{T}$-class algorithm 26

5.4 Numerical illustration of Algorithm 5.4 27

5.5 Parallel methods for inconsistent problems 32

6 Constrained minimization methods 35

7 Remarks on nonconvex problems 36

1 Image recovery: An overview

Image recovery is a broad discipline that encompasses the large body of inverse problems in which an image x is to be inferred from the observation of data y 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. General references on image recovery and its applications are [2], [3], [22], [35], [65], [66], [50], [73], and [69].

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 x of a degraded image y. Hence, in image restoration the data consist of measurements taken directly on the image to be estimated, y being a blurred and noise-corrupted version of x. 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 y are indirectly related to the form of the original image x. 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:

(i) **A data formation model:** The data formation model is essentially a model of the imaging system, i.e., a mathematical description of the relation between the original image x and the recorded data y. The original image and the recorded data are assumed to belong to some real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}$, respectively. In most applications, $\mathcal{H}$ and $\mathcal{G}$ are closed subspaces of $L^2(\Omega, \mathcal{F}, \mu)$ for an appropriate measure space $(\Omega, \mathcal{F}, \mu)$ (this framework covers in particular the Lebesgue space L^2, the Sobolev space H^1, the sequence space ℓ^2, and the usual Euclidean space $\mathbb{R}^N$). One of the most common data formation models in image restoration is

$$y = Lx + u, \tag{1.1}$$

where the operator $L\colon \mathcal{H} \to \mathcal{G}$ represents the blurring process and u an additive noise component. Within this generic model, various sub-models can be distinguished, according as L is linear or nonlinear, deterministic or stochastic, or according as the noise depends on Lx or not, etc. Different models can also be considered to describe situations in which 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, Lx will stand for a low-passed Fourier transform in band-limited extrapolation and a Radon transform in tomography.

(ii) ***A priori* information:** Information may be available to describe the original image x or some constituents of the data formation model. For instance, in (1.1), L may be known as well as some statistical properties of the noise u. Each piece of *a priori* information yields a constraint Ψ_i on the original image, which confines estimates to the set

$$S_i = \{x \in \mathcal{H} \mid x \text{ satisfies } \Psi_i\}. \tag{1.2}$$

Here are a few typical examples [22, 65, 66, 75]:

(a) Lower and upper bounds on the amplitude of the original image are known. Then

$$S_i = \{x \in \mathcal{H} \mid \text{range}(x) \subset [\gamma, \delta]\}. \tag{1.3}$$

(b) The image has limited region of support K. Then[1]

$$S_i = \{x \in \mathcal{H} \mid x = x1_K\}. \tag{1.4}$$

(c) The image is known over some domain K. Then

$$S_i = \{x \in \mathcal{H} \mid x1_K = z1_K\}. \tag{1.5}$$

(d) A bound γ^2 is available on the energy of the original image. Then

$$S_i = \{x \in \mathcal{H} \mid ||x|| \leq \gamma\}. \tag{1.6}$$

(e) More generally, a bound γ is available on the maximum deviation of x from a reference image r. Then

$$S_i = \{x \in \mathcal{H} \mid \|x - r\| \leq \gamma\}. \tag{1.7}$$

(f) A further generalization of the previous example is

$$S_i = \{x \in \mathcal{H} \mid \|Tx - r\| \leq \gamma\}, \tag{1.8}$$

where $T : \mathcal{H} \to \mathcal{G}$ is a bounded linear operator. For instance, if $r = 0$ and T is a differential operator, (1.8) is a set of smooth images; if $r = 0$ and $T = \text{Id} - T'$, (1.8) is the set of images that are nearly invariant under the operator T'.

(g) Moment constraints yield sets of the form

$$S_i = \{x \in \mathcal{H} \mid \gamma \leq \langle x \mid b \rangle \leq \delta\}. \tag{1.9}$$

(h) The original image is band-limited. If we designate by K the corresponding low frequency band and $\widehat{x}$ the Fourier transform of x,

$$S_i = \{x \in \mathcal{H} \mid \widehat{x} = \widehat{x}1_K\}. \tag{1.10}$$

(i) A stronger hypothesis is that the Fourier transform of the original image is a known function z over some frequency band K. Then

$$S_i = \{x \in \mathcal{H} \mid \widehat{x}1_K = z1_K\}. \tag{1.11}$$

(j) The previous constraint can be generalized by considering the set of images that match approximately a reference image r over some frequency band K,

$$S_i = \{x \in \mathcal{H} \mid \|(\widehat{x} - \widehat{r})\, 1_K\| \leq \gamma\}. \tag{1.12}$$

[1] 1_A is the characteristic function of a set A: it takes the value 1 on A and 0 on $\complement A$.

(k) The image has a real and nonnegative Fourier transform. Then

$$S_i = \{x \in \mathcal{H} \mid \text{range}(\widehat{x}) \subset \mathbb{R}_+\} \tag{1.13}$$

(l) The Fourier magnitude of the image is a known function m (phase retrieval problem, e.g., diffraction imaging, electron microscopy, crystallography). Then

$$S_i = \{x \in \mathcal{H} \mid |\widehat{x}| = m\}. \tag{1.14}$$

(m) The phase of the image is a known function φ (magnitude recovery, e.g., holography). Then

$$S_i = \{x \in \mathcal{H} \mid \angle\widehat{x} = \varphi\}. \tag{1.15}$$

(n) The above sets are based on the Fourier transform. Sets can also be constructed from constraints on other transforms, such as the wavelet transform [15, 31, 56].

(o) Some upper bound is available on the total variation of the image [30]

$$S_i = \{x \in \mathcal{H} \mid \int_\Omega |\nabla x(\omega)|_2 d\omega \leq \eta\}. \tag{1.16}$$

(p) Properties of the noise [33]: A known probabilistic property Ψ_i of the noise u in (1.1) can be turned into a confidence region

$$S_i = \{x \in \mathcal{H} \mid y - Lx \;\text{ is consistent with } \Psi_i\}. \tag{1.17}$$

The confidence level of that set is derived from the distribution of the underlying statistic of the residual $y - Lx$ that is associated with Ψ_i. Examples will be given in the applications.

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

$$S = \bigcap_{i \in I} S_i = \{x \in \mathcal{H} \mid (\forall i \in I) \;\; x \text{ satisfies } \Psi_i\}. \tag{1.18}$$

(iii) **A recovery criterion:** The recovery criterion defines those images that are acceptable as solutions to the problem.

Our goal is to formulate the recovery problem so as to extract as much information from the data and the prior knowledge as possible, while avoiding to bias the search for a solution in a particular direction. In doing so, one will obtain a reliable estimate that the practitioner can safely use. If little *a priori* information is available and/or the data are severely degraded, then one will simply not be able to scientifically produce a very good estimate and should not attempt to artificially "invent information." Rather, the challenge is to develop mathematical techniques that can appropriately model most of the existing information and then effectively incorporate it in the recovery scheme.

These considerations lead us to look for a solution in the feasibility set defined in (1.18). In the set theoretic approach [19, 22, 70, 75], all feasible solutions are equally valid. In some problems, it may be judicious to pick a particular feasible solution on the basis of optimality with respect to a certain cost $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ that conveys prior information that cannot be conveniently associated with a constraint. The recovery problem then takes the form

$$\text{Find } \; x \in S = \bigcap_{i \in I} S_i \; \text{ such that } \; f(x) = \min f(S). \tag{1.19}$$

Among the costs that have been considered, let us mention in particular

(a) Probabilistic costs: negative likelihood, Bayesian costs, etc [2, 35].

(b) Various forms of (neg)entropy [2, 9, 12, 47, 48].

(c) Energy and other quadratic functionals (see Section 6).

(d) L^p norm.

(e) Residual energy [2, 66].

(f) Total variation [17, 16, 29, 46, 54, 72].

Most costs in $H^1(\Omega)$ can be expressed as integral functionals of the form

$$f\colon x \mapsto \int_\Omega \varphi\big(\omega, x(\omega), \nabla x(\omega)\big) d\omega. \tag{1.20}$$

For instance:

- $\varphi\colon (a, b, c) \mapsto \phi(a)b$: "generalized moment"
- $\varphi\colon (a, b, c) \mapsto -\ln b$: "Burg's entropy"
- $\varphi\colon (a, b, c) \mapsto b \ln b$: "Boltzmann-Shannon's entropy"
- $\varphi\colon (a, b, c) \mapsto |c|_2^2$: "Heat equation"
- $\varphi\colon (a, b, c) \mapsto \int_0^{|c|_2} \phi(s) ds$: "Perona-Malik"
- $\varphi\colon (a, b, c) \mapsto |c|_2$: "Osher-Rudin"
- $\varphi\colon (a, b, c) \mapsto |c|_2^2/b$: "Fisher information"
- etc...

(iv) **A solution method:** The solution method is a numerical algorithm that will produce a solution to the recovery problem, i.e., an image that solves (1.19). 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 *should* therefore not be adopted. As a result, it may be necessary to simplify (1.19) by altering f or getting rid of some constraints.

The reader should be warned that in the literature this ultimate, yet crucial, component of the recovery problem is often neglected; thus, numerically untractable problems are "solved" via *ad hoc* numerical methods that are not guaranteed to produce meaningful solutions.

A clear trade-off arises from the above discussion. On the one hand, obtaining high quality solutions to a recovery problem demands a sophisticated problem formulation in which as much *a priori* information as possible should be incorporated. On the other hand, a problem so formulated may not be numerically tractable. In pondering over this trade-off, one should be guided by a clear objective, namely ***to develop a recovery scheme that reliably produces solutions which are consistent with the data and prior information***.

The purpose of this course is to give a brief introductory account of the field of convex projection methods in image recovery. In such techniques, one considers only convex constraints and costs. This is admittedly a restriction since some essential constraints are not convex, most notably the magnitude constraint (1.14) in phase retrieval [6, 7, 65]. At the same time, this convexity framework possesses very attractive features:

- It is reliable in the sense that it utilizes algorithms with global convergence properties, irrespective of the starting point.
- The solutions it produces solutions with well-defined properties, namely those that yield convex constraint sets.

In the end, this framework owes its success to the powerful theoretical basis on which they rest, namely ***hilbertian convex analysis and fixed point theory***, and the fact that many of the constraints and costs encountered in applications turn out to be convex.

2 Classical linear image recovery methods

$\mathcal{H}$ and $\mathcal{G}$ are two real Hilbert spaces, $L\colon \mathcal{H} \to \mathcal{G}$ a bounded linear operator, $x \in \mathcal{H}$ the original signal to be recovered, y the observed signal, and n additive noise. The model is

$$y = Lx + u. \tag{2.1}$$

The goal is to find an operator $Q \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ such that the linear estimate $\widetilde{x} = Qy$ is optimal in some sense. We describe two standard methods of this type.

2.1 Inverse filtering

Assume a finite dimensional setting. The goal is to minimize the *a posteriori* mean-square fitting error

$$f(Q) = \frac{1}{2}\mathsf{E}\|y - L\widetilde{x}\|^2 = \frac{1}{2}\mathsf{E}\|y - LQy\|^2 \tag{2.2}$$

To find Q, we assume that the noise u is uncorrelated with the original signal x and that it has zero mean. Hence

$$\mathsf{E}u^\top x = \mathsf{E}u^\top \mathsf{E}x = \underline{0}. \tag{2.3}$$

Moreover, x and u are second order stationary processes. Hence the covariance matrices $R_x = \mathsf{E}xx^\top$ and $R_u = \mathsf{E}uu^\top$ are Toeplitz with rapid decay and will be assumed to be circulant. The above assumptions on the noise also impose

$$R_y = \mathsf{E}yy^\top = \mathsf{E}(Lx+u)(Lx+u)^\top = LR_xL^\top + R_u. \tag{2.4}$$

This matrix will be assumed to be invertible.

Let us find the optimal Q. We have

$$f(Q) = \frac{1}{2}\mathsf{E}(y-LQy)^\top(y-LQy) = \frac{1}{2}\mathsf{E}\big(y^\top y - 2y^\top LQy + y^\top Q^\top L^\top LQy\big) \tag{2.5}$$

Using the identities (for A symmetric)

$$\frac{\partial a^\top Qb}{\partial Q} = ab^\top \quad \text{and} \quad \frac{\partial a^\top Q^\top AQa}{\partial Q} = 2AQaa^\top, \tag{2.6}$$

we obtain

$$\frac{\partial f(Q)}{\partial Q} = \mathsf{E}\big(-L^\top yy^\top + L^\top LQyy^\top\big) = -L^\top R_y + L^\top LQR_y. \tag{2.7}$$

Since R_y is invertible, setting this to zero yields

$$L^\top LQ = L^\top. \tag{2.8}$$

The solution to this equation is the pseudo-inverse of L, i.e., $Q = L^\dagger$. The solution is

$$\widetilde{x} = L^\dagger y \tag{2.9}$$

Assuming that L stands for a convolution process with kernel l (i.e., $y = l \star x + u$), we have in the frequency domain

$$\widehat{\widetilde{x}} = \frac{\widehat{y}}{\widehat{l}} = \frac{\widehat{y}}{\widehat{l}} = \widehat{x} + \frac{\widehat{u}}{\widehat{l}}. \tag{2.10}$$

The second term produces unacceptable noise amplification for frequencies at which $\widehat{l} \approx 0$, see Fig. 1.

2.2 Wiener filtering

Assume a finite dimensional setting. The goal is to minimize the mean-square error

$$f(Q) = \frac{1}{2}\mathsf{E}\|x-\widetilde{x}\|^2 = \frac{1}{2}\mathsf{E}\|x-Qy\|^2, \tag{2.11}$$

under the same assumptions as above. Let us find the optimal Q. We have

$$f(Q) = \frac{1}{2}\mathsf{E}(x-Qy)^\top(x-Qy) = \frac{1}{2}\mathsf{E}\big(x^\top x - 2x^\top Qy + y^\top Q^\top Qy\big) \tag{2.12}$$

Figure 1: Inverse filtering: Convolution by a 9×9 flat kernel L and addition of uniform noise u (SNR 86, 66, and 46 dB). Left: original x; center: degraded $y = Lx + u$; right: restored. Even though the noise level is low, the performance is very poor.

```
#! /usr/bin/octave -qf
#load data
load -force "/home/plc/recherche/images/lena128.mat";
a=lena128; a=255*a/max(max(a)); A=fft2(a); N=128;

#DEGRADE IMAGE
b=zeros(size(a)); M=9;
b(1:ceil(M/2),1:ceil(M/2))=ones(ceil(M/2),ceil(M/2));
b(1:ceil(M/2),N-floor(M/2)+1:N)=ones(ceil(M/2),floor(M/2));
b(N-floor(M/2)+1:N,1:ceil(M/2))=ones(floor(M/2),ceil(M/2));
b(N-floor(M/2)+1:N,N-floor(M/2)+1:N)=ones(floor(M/2),floor(M/2));
b=b/(M*M);
B=fft2(b);
d=abs(ifft2(fft2(a).*B));

disp('Inverse filter - Uniform noise, SNR:');
n=0.01*rand(size(a));
20*log10(norm(d,'fro')/norm(n,'fro'))
x1=d+n;x1=x1*255/max(max(x1));
X1=fft2(x1);

n=0.1*rand(size(a));
20*log10(norm(d,'fro')/norm(n,'fro'))
x2=d+n;x2=x2*255/max(max(x2));
X2=fft2(x2);

n=1.0*rand(size(a));
20*log10(norm(d,'fro')/norm(n,'fro'))
x3=d+n;x3=x3*255/max(max(x3));
X3=fft2(x3);

# INVERSE FILTER
ni1=abs(ifft2(X1./B));
ni2=abs(ifft2(X2./B));
ni3=abs(ifft2(X3./B));
ni1=ni1*255/max(max(ni1));
ni2=ni2*255/max(max(ni2));
ni3=ni3*255/max(max(ni3));
imagesc([a,x1,ni1;a,d,ni2;a,d,ni3],2);
```

Figure 2: Octave [49] code for the experiment of Fig. 1.

Using (2.6), we obtain

$$\frac{\partial f(Q)}{\partial Q} = \mathsf{E}\big(-xy^\top + Qyy^\top\big) = -\mathsf{E}x(Lx+u)^\top + QR_y = -R_xL^\top + QR_y. \tag{2.13}$$

Since R_y is invertible, setting this to zero yields

$$Q = R_xL^\top R_y^{-1} = R_xL^\top\big(LR_xL^\top + R_u\big)^{-1}. \tag{2.14}$$

The solution is therefore

$$\widetilde{x} = Qy = R_xL^\top\big(LR_xL^\top + R_u\big)^{-1}y. \tag{2.15}$$

Assuming that L stands for a convolution process with kernel l (i.e., $y = l \star x + u$), we have in the frequency domain (since all matrices are circulant they are diagonalized by the DFT (Discrete Fourier Transform) matrix [2])

$$\widehat{\widetilde{x}} = \frac{P_x\overline{\widehat{l}}\,\widehat{y}}{|\widehat{l}\,|^2P_x + P_u}, \tag{2.16}$$

where we denote by P_x and P_u the power spectral densities of x and u, respectively. This operation can be implemented very efficiently with the FFT (Fast Fourier Transform). Note that

$$\widehat{\widetilde{x}} = \frac{\overline{\widehat{l}}\,\widehat{y}}{|\widehat{l}\,|^2 + P_u/P_x}. \tag{2.17}$$

If no noise is present, $\widehat{\widetilde{x}} = \overline{\widehat{l}}\,\widehat{y}/|\widehat{l}\,|^2 = \widehat{y}/\widehat{l}$ and we recover (2.10). The term P_u/P_x prevents the denominator from being too small and therefore stabilizes the inversion process. However, the solutions thus obtained are usually very smooth (see Fig. 3).

3 Hilbertian convex analysis and fixed point theory: A primer

The classical methods described in Section 2 are simple to implement but fall short on several scores. In particular, they are extremely limited in terms of incorporating *a priori* information. These shortcomings can potentially be circumvented by bringing into play nonlinear techniques. The main tools we shall be using are convex analysis[2] and fixed point theory. We provide here a very basic account that will cover our immediate needs.

Throughout, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle\cdot\mid\cdot\rangle$, associated norm $\|\cdot\|$, and associated distance d. For more details on convex analysis see [8, 37, 53, 76]; for nonexpansive operators see [39, 40]; for $\mathfrak{T}$-class operators see [5, 26]; for convex projection operators, see [36].

[2]Modern convex analysis was developed by Jean-Jacques Moreau in France and R. Terry Rockafellar in the US in the 1960s.

Figure 3: Wiener filtering: Convolution by a 9×9 flat kernel L and addition of uniform noise u (SNR 65, 45, and 23 dB). Left: original x; center: degraded $y = Lx + u$; right: restored. Even though the noise level is higher than in Fig. 1, the results are better, although the edges are very smooth.

```
#! /usr/bin/octave -qf
#load data
load -force "/home/plc/recherche/images/lena128.mat";
a=lena128; a=255*a/max(max(a)); N=128;

#DEGRADE IMAGE
b=zeros(size(a));
M=9;
b(1:ceil(M/2),1:ceil(M/2))=ones(ceil(M/2),ceil(M/2));
b(1:ceil(M/2),N-floor(M/2)+1:N)=ones(ceil(M/2),floor(M/2));
b(N-floor(M/2)+1:N,1:ceil(M/2))=ones(floor(M/2),ceil(M/2));
b(N-floor(M/2)+1:N,N-floor(M/2)+1:N)=ones(floor(M/2),floor(M/2));
b=b/(M*M);
B=fft2(b);
d=abs(ifft2(fft2(a).*B));

disp('Wiener filter - Uniform noise, SNR:');
s1=0.01;
n1=sqrt(s1)*rand(size(a));
20*log10(norm(d,'fro')/norm(n1,'fro'))
x1=d+n1;x1=x1*255/max(max(x1)); X1=fft2(x1);

s2=1.2;
n2=sqrt(s2)*rand(size(a));
20*log10(norm(d,'fro')/norm(n2,'fro'))
x2=d+n2;x2=x2*255/max(max(x2));
X2=fft2(x2);

s3=200;
n3=sqrt(s3)*rand(size(a));
20*log10(norm(d,'fro')/norm(n3,'fro'))
x3=d+n3;x3=x3*255/max(max(x3));
X3=fft2(x3);

#WIENER FILTER
w1=abs(ifft2((conj(B).*X1)./(abs(B).^2+abs(fft2(n1)).^2./(abs(X1).^2))));
w1=255*w1/max(max(w1));
w2=abs(ifft2((conj(B).*X2)./(abs(B).^2+abs(fft2(n2)).^2./(abs(X2).^2))));
w2=255*w2/max(max(w2));
w3=abs(ifft2((conj(B).*X3)./(abs(B).^2+abs(fft2(n3)).^2./(abs(X3).^2))));
w3=255*w3/max(max(w3));
imagesc([a,x1,w1;a,x2,w2;a,x3,w3],2);
```

Figure 4: Octave [49] code for the experiment of Fig. 3.

3.1 Basic facts

Fact 3.1 *For every x and y in $\mathcal{H}$, the following properties hold.*

(i) $\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x \mid y\rangle + \|y\|^2$.

(ii) $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$ *(Parallelogram identity).*

(iii) $\|x+y\|^2 - \|x-y\|^2 = 4\langle x \mid y\rangle$ *(Polarization identity).*

(iv) $|\langle x \mid y\rangle| \leq \|x\| \cdot \|y\|$ *(Cauchy-Schwarz).*

(v) $(\forall \alpha \in \mathbb{R})\ \ \|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2$.

(vi) $\langle x \mid y\rangle \leq 0 \quad\Leftrightarrow\quad (\forall \alpha \in \,]0,1])\ \ \|x\| \leq \|x - \alpha y\| \quad\Leftrightarrow\quad (\forall \alpha \in \,]0,+\infty[)\ \ \|x\| \leq \|x-\alpha y\|$.

(vii) $x \perp y \quad\Leftrightarrow\quad (\forall \alpha \in \mathbb{R}^*)\ \ \|x\| \leq \|x - \alpha y\|$.

Fact 3.2 *Let C be a nonempty subset of $\mathcal{H}$. The distance to C is the function*

$$d_C\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \inf_{y \in C} d(x,y). \tag{3.1}$$

This function is continuous (actually 1-Lipschitz).

Fact 3.3 (weak and strong convergence) *A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ converges weakly to $x \in \mathcal{H}$ if, for every $y \in \mathcal{H}$, $\langle x_n - x \mid y\rangle \to 0$, in symbols, $x_n \rightharpoonup x$. By Cauchy-Schwarz $x_n \to x$ (i.e, $\|x_n - x\| \to 0$) $\Rightarrow x_n \rightharpoonup x$; if $\dim \mathcal{H} < +\infty$, $x_n \rightharpoonup x \Rightarrow x_n \to x$.*

3.2 Convex sets

Definition 3.4 *For every $(x,y) \in \mathcal{H}^2$, the* open segment *between x and y is*

$$]x,y[\; = \big\{\alpha x + (1-\alpha)y \mid 0 < \alpha < 1\big\}. \tag{3.2}$$

A subset C of $\mathcal{H}$ is convex *if $(\forall (x,y) \in C^2)\ ]x,y[\; \subset C$. By convention, $\varnothing$ is convex.*

Example 3.5 The following subsets of $\mathcal{H}$ are convex.

(i) A ball (open or closed).

(ii) An *affine subspace*, i.e., a nonempty subset C such that

$$(\forall \alpha \in \mathbb{R})\ \ \alpha C + (1-\alpha)C = C. \tag{3.3}$$

(iii) A *closed affine hyperplane*, i.e.,

$$C = \big\{x \in \mathcal{H} \mid \langle x \mid t\rangle = \eta\big\}, \tag{3.4}$$

where $t \in \mathcal{H} \smallsetminus \{0\}$ and $\eta \in \mathbb{R}$.

(iv) An *affine halfspace*, i.e.,

$$C = \big\{x \in \mathcal{H} \mid \langle x \mid t\rangle \leq \eta\big\} \;\text{ or }\; C = \big\{x \in \mathcal{H} \mid \langle x \mid t\rangle < \eta\big\}, \tag{3.5}$$

where $t \in \mathcal{H} \smallsetminus \{0\}$ and $\eta \in \mathbb{R}$.

(v) Any intersection of convex sets.

(vi) Let $\mathcal{G}$ be a real Hilbert space, let $T\colon \mathcal{H} \to \mathcal{G}$ an affine[3] operator, and let C and D be convex subsets of $\mathcal{H}$ and $\mathcal{G}$, respectively. Then $T(C)$ and $T^{-1}(D)$ are convex subsets of $\mathcal{G}$ and $\mathcal{H}$, respectively.

(vii) Let $(C_i)_{1\leq i\leq m}$ be convex subsets of $\mathcal{H}$. Then

(a) $\displaystyle\mathop{\times}_{i=1}^{m} C_i$ is convex.

(b) $(\forall(\alpha_i)_{1\leq i\leq m} \in \mathbb{R}^m)\;\; \sum_{i=1}^m \alpha_i C_i$ is convex.

Fact 3.6 (topological properties) *Let C be a convex subset of $\mathcal{H}$. Then*

(i) $(\forall x \in \operatorname{int} C)(\forall y \in \overline{C})\;\; [x, y[\; \subset \operatorname{int} C$.

(ii) $\overline{C}$ *is convex.*

(iii) $\operatorname{int} C$ *is convex.*

(iv) *If* $\operatorname{int} C \neq \varnothing$, $\operatorname{int} C = \operatorname{int} \overline{C}$ *and* $\overline{C} = \overline{\operatorname{int} C}$.

(v) *If C is closed and convex, it is weakly closed.*

Fact 3.7 (separation) *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x \in \mathcal{H} \smallsetminus C$. Then x is separated from C, i.e.,*

$$(\exists\, t \in \mathcal{H} \smallsetminus \{0\})(\exists\, \varepsilon \in\,]0, +\infty[)(\forall z \in C) \quad \langle z - x \mid t\rangle \leq -\varepsilon. \tag{3.6}$$

[3] $T\colon x \mapsto g + Lx$, L linear, $g \in \mathcal{G}$. Alternatively, $(\forall \alpha \in \mathbb{R})(\forall (x, y) \in \mathcal{H}^2)\; T(\alpha x + (1-\alpha)y) = \alpha T(x) + (1-\alpha)Ty$.

3.3 Convex functions

Definition 3.8 *A function* $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ *is* proper *if* $\operatorname{dom} f = \big\{x \in \mathcal{H} \mid f(x) \neq +\infty\big\} \neq \varnothing$ *and* convex *if its epigraph*

$$\operatorname{epi} f = \big\{(x,\eta) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \eta\big\}, \tag{3.7}$$

is convex in $\mathcal{H} \times \mathbb{R}$*. In other words,*

$$(\forall \alpha \in \,]0,1[)(\forall (x,y) \in (\operatorname{dom} f)^2) \;\; f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha) f(y). \tag{3.8}$$

The class of all proper lower semicontinous convex functions from $\mathcal{H}$ *to* $]-\infty,+\infty]$ *is denoted by* $\Gamma_0(\mathcal{H})$*.*

Example 3.9

(i) Let C be a convex subset of $\mathcal{H}$. Then the *indicator function*

$$\iota_C\colon x \mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C, \end{cases} \tag{3.9}$$

is convex.

(ii) Let C be a nonempty convex subset of $\mathcal{H}$. Then d_C is convex. In particular, $\|\cdot\|$ is convex.

(iii) Let $(f_i)_{i\in I}$ be a family of convex functions from $\mathcal{H}$ into $\mathbb{R}$. Then

(a) If I is finite, for every $(\alpha_i)_{i\in I}$ in $[0,+\infty[$, $\sum_{i\in I} \alpha_i f_i$ is convex.

(b) $\sup_{i\in I} f_i$ is convex (indeed, $\operatorname{epi} \sup_{i\in I} f_i = \bigcap_{i\in I} \operatorname{epi} f_i$ is convex as an intersection of convex sets).

(iv) Let $f\colon \mathcal{H} \to \mathbb{R}$ be a convex function and $\varphi\colon \mathbb{R} \to \mathbb{R}$ a nondecreasing convex function. Then $\varphi \circ f$ is convex (with the convention $\varphi(+\infty) = +\infty$).

Definition 3.10 *Let* $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ *be a convex function and let* $(x,y) \in \mathcal{H}^2$*. It follows from the convexity of* f *that the function* $\lambda \mapsto (f(x+\lambda y) - f(x))/\lambda$ *is increasing on* $]0,+\infty[$*. Therefore the* directional derivative *of* f *at* x *in the direction* y*,*

$$f'(x;y) = \lim_{\lambda \downarrow 0+} \frac{f(x+\lambda y) - f(x)}{\lambda} = \inf_{\lambda > 0} \frac{f(x+\lambda y) - f(x)}{\lambda} \tag{3.10}$$

exists in $[-\infty,+\infty]$*. If* $f'(x;\cdot)$ *is linear and continuous on* $\mathcal{H}$*, then there exists a (unique) vector* $\nabla f(x) \in \mathcal{H}$ *such that*

$$(\forall y \in \mathcal{H}) \;\; f'(x;y) = \lim_{\lambda \to 0} \frac{f(x+\lambda y) - f(x)}{\lambda} = \langle y \mid \nabla f(x)\rangle, \tag{3.11}$$

called the Gâteaux-derivative *of* f *at* x*.*

Definition 3.11 *The* subdifferential *of a proper function* $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ *is*

$$\partial f\colon x \mapsto \big\{t \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \ \langle y - x \mid t\rangle + f(x) \leq f(y)\big\}. \tag{3.12}$$

f *is* subdifferentiable *at* x *if* $\partial f(x) \neq \varnothing$. *The elements of* $\partial f(x)$ *are the* subgradients *of* f *at* x.

Example 3.12 (normal cone) Let C be a convex subset of $\mathcal{H}$. Then the normal cone operator of C is

$$N_C = \partial\iota_C\colon x \mapsto \begin{cases} \big\{t \in \mathcal{H} \mid (\forall y \in C)\ \ \langle y - x \mid t\rangle \leq 0\big\} & \text{if } x \in C\\ \varnothing & \text{otherwise.}\end{cases} \tag{3.13}$$

Fact 3.13 *Let* $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ *be a proper convex function and* $x \in \mathcal{H}$. *Then:*

(i) $f(x) = \min f(\mathcal{H}) \Leftrightarrow 0 \in \partial f(x)$ *(Fermat's rule).*

(ii) *If* f *is differentiable at* x, *then* $\partial f(x) = \{\nabla f(x)\}$.

(iii) *Suppose* f *is continuous at* $x \in \operatorname{dom} f$. *Then*

  (a) $\partial f(x) \neq \varnothing$.

  (b) *If* $\partial f(x)$ *contains a single element* t, *then* $t = \nabla f(x)$.

3.4 Nonlinear operators and fixed points

Let $T\colon \mathcal{H} \to \mathcal{H}$ be an operator and let $k \in \,]0,+\infty[$. The set of *fixed points* of T is

$$\operatorname{Fix} T = \big\{x \in \mathcal{H} \mid Tx = x\big\}. \tag{3.14}$$

T is said to be *k-Lipschitz continuous* if

$$(\forall (x,y) \in \mathcal{H}^2)\ \ \|Tx - Ty\| \leq k\|x - y\|. \tag{3.15}$$

If $k < 1$, then T is called a *strict contraction*; if $k = 1$, then T is said to be *nonexpansive*. Alternatively, T is nonexpansive if and only if $F = (\mathrm{Id} + T)/2$ is *firmly nonexpansive*, i.e.,

$$(\forall (x,y) \in \mathcal{H}^2)\ \ \|Fx - Fy\|^2 \leq \|x - y\|^2 - \|(\mathrm{Id} - F)x - (\mathrm{Id} - F)y\|^2. \tag{3.16}$$

T is said to be a *Fejér operator* if

$$(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \|Tx - y\| \leq \|x - y\| \tag{3.17}$$

and a $\mathfrak{T}$-class operator if

$$(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \langle y - Tx \mid x - Tx\rangle \leq 0. \tag{3.18}$$

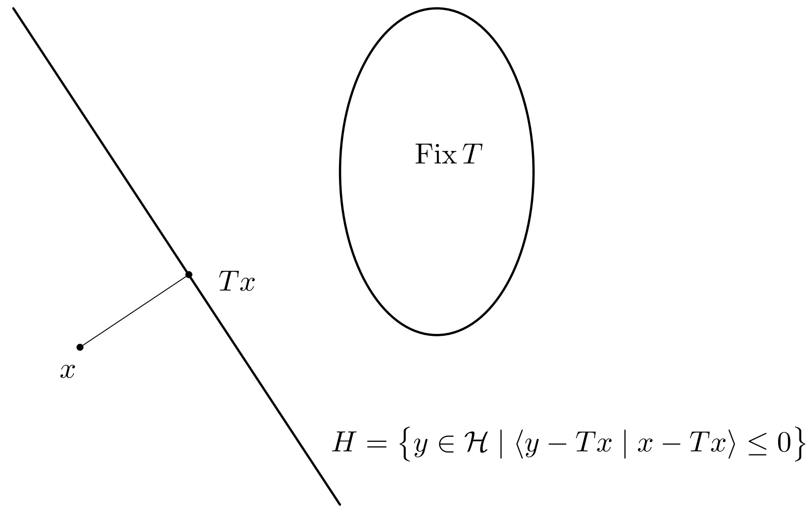

Figure 5: A $\mathfrak{T}$-class operator.

We have

$$\begin{array}{ccccc}
 & & \mathrm{Id} - T \text{ firmly nonexpansive} & & \\
 & & \Updownarrow & & \\
2T - \mathrm{Id} \text{ nonexpansive} & \Leftrightarrow & T \text{ firmly nonexpansive} & \Rightarrow & T \text{ nonexpansive} \\
\Downarrow & & \Downarrow & & \Downarrow \\
2T - \mathrm{Id} \text{ Fejér} & \Leftrightarrow & T \in \mathfrak{T} & \Rightarrow & T \text{ Fejér} \\
 & & & & \Downarrow \\
 & & & & \mathrm{Fix}\, T \text{ closed and convex.}
\end{array} \tag{3.19}$$

The last implication follows from the identity

$$\mathrm{Fix}\, T = \bigcap_{x \in \mathcal{H}} \big\{ y \in \mathcal{H} \mid \langle y \mid x - Tx \rangle \leq \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}\|Tx\|^2 \big\}. \tag{3.20}$$

3.5 Properties of $\mathfrak{T}$-class operators

Proposition 3.14 *Let $T \in \mathfrak{T}$. Then*

(i) *Set $R = \mathrm{Id} + \kappa(T - \mathrm{Id})$, where $\kappa \in \,]0, +\infty[$. Then*

$$(\forall (x,y) \in \mathcal{H} \times \mathrm{Fix}\, T) \quad \|Rx - y\|^2 \leq \|x - y\|^2 - \kappa(2 - \kappa)\|Tx - x\|^2. \tag{3.21}$$

In particular,

$$(\forall (x,y) \in \mathcal{H} \times \mathrm{Fix}\, T) \quad \|Tx - y\|^2 \leq \|x - y\|^2 - \|Tx - x\|^2. \tag{3.22}$$

(ii) $(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \|Tx - x\|^2 \leq \langle y - x \mid Tx - x \rangle$

(iii) $(\forall (x,y) \in (\mathcal{H} \smallsetminus \operatorname{Fix} T) \times \operatorname{Fix} T)\ \ \|Tx - y\| < \|x - y\|$.

(iv) $(\forall x \in \mathcal{H})\ \ \|Tx - x\| \leq d_{\operatorname{Fix} T}(x)$.

(v) $\operatorname{Fix} T$ *is closed and convex.*

Proposition 3.15 *Let $(T_i)_{1 \leq i \leq m}$ be a finite family of operators in $\mathfrak{T}$ such that $\bigcap_{i=1}^m \operatorname{Fix} T_i \neq \varnothing$. Then $\operatorname{Fix} T_m \cdots T_1 = \bigcap_{i=1}^m \operatorname{Fix} T_i$.*

Proposition 3.16 *Let I be a finite ordered index set, let $(T_i)_{i\in I}$ be a family of operators in $\mathfrak{T}$ such that $F = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, and let $(\omega_i)_{i\in I}$ be numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$. Define*

$$(\forall x \in \mathcal{H})\ \ L\big(x, (T_i)_{i\in I}, (\omega_i)_{i\in I}\big) = \begin{cases} \dfrac{\sum_{i\in I} \omega_i \|T_i x - x\|^2}{\|\sum_{i\in I} \omega_i T_i x - x\|^2}, & \text{if } x \notin F; \\ 1, & \text{otherwise,} \end{cases} \tag{3.23}$$

and

$$T\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto x + \lambda(x) \left(\sum_{i\in I} \omega_i T_i x - x \right) \quad \text{where } 0 < \lambda(x) \leq L\big(x, (T_i)_{i\in I}, (\omega_i)_{i\in I}\big). \tag{3.24}$$

Then

(i) *For every $x \in \mathcal{H}$, $L\big(x, (T_i)_{i\in I}, (\omega_i)_{i\in I}\big)$ is a well defined number in $[1, +\infty[$.*

(ii) $\operatorname{Fix} T = F$.

(iii) $T \in \mathfrak{T}$.

Corollary 3.17 *Let I be a finite ordered index set, let $(T_i)_{i\in I}$ be a family of operators in $\mathfrak{T}$ such that $F = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, and let $(\omega_i)_{i\in I}$ be numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$. Set*

$$T = \sum_{i\in I} \omega_i T_i. \tag{3.25}$$

Then $\operatorname{Fix} T = F$ and $T \in \mathfrak{T}$.

3.6 Projection onto a convex set

Theorem 3.18 *Let C be a nonempty closed and convex subset of $\mathcal{H}$ and x a point in $\mathcal{H}$. Then there exists a unique point $y \in C$ such that $d_C(x) = \|y - x\|$. This point, called the* projection of x onto C*, is characterized by the variational inequality*

$$y \in C \ \text{ and } \ (\forall z \in C)\ \ \langle z - y \mid x - y \rangle \leq 0. \tag{3.26}$$

The projector *onto C that maps every point in $\mathcal{H}$ into its projection onto C is denoted by P_C.*

Some useful corollaries of Theorem 3.18 follow.

Fact 3.19 *The projector onto a nonempty closed convex subset C of $\mathcal{H}$ is firmly nonexpansive (hence nonexpansive and therefore continuous):*

$$(\forall (x,y) \in \mathcal{H}^2) \ \ \|P_C x - P_C y\|^2 \leq \|x-y\|^2 - \|(\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y\|^2. \tag{3.27}$$

In particular, $P_C \in \mathfrak{T}$ and

$$(\forall x \in \mathcal{H})(\forall y \in C) \ \ \|P_C x - y\|^2 \leq \|x-y\|^2 - \|P_C x - x\|^2. \tag{3.28}$$

Fact 3.20 *Let V be a closed vector subspace of $\mathcal{H}$. Then the projector P_V is linear and characterized by the variational equality*

$$(\forall x \in \mathcal{H}) \ \ P_V x \in V \ \ \text{and} \ \ (\forall v \in V) \ \ \langle v \mid x - P_V x\rangle = 0. \tag{3.29}$$

Moreover, if $V \neq \{0\}$, then $\|P_V\| = \sup_{\|x\|=1} \|P_V x\| = 1$.

Fact 3.21 *Let C be a nonempty closed convex set and N_C its normal cone operator (See Example 3.12). Then*

$$(\forall x \in \mathcal{H}) \ \ \partial d_C(x) = \begin{cases} \left\{\dfrac{x - P_C x}{d_C(x)}\right\} & \text{if } x \notin C \\ N_C x \cap B(0;1) & \text{if } x \in C, \end{cases} \tag{3.30}$$

where $B(0;1)$ is the closed unit ball in $\mathcal{H}$. In particular,

$$(\forall x \in \mathcal{H}) \ \ \partial\|x\| = \begin{cases} \left\{\dfrac{x}{\|x\|}\right\} & \text{if } x \neq 0 \\ B(0;1) & \text{if } x = 0. \end{cases} \tag{3.31}$$

3.7 Subgradient projection onto a level set

Let C be a nonempty closed and convex subset of $\mathcal{H}$ and let x be a point in $\mathcal{H}$. Then we have seen that there exists a unique point $P_C x \in C$ such that $\|x - P_C x\| = d_C(x)$. By definition, computing the projection onto C amounts to solving a constrained quadratic minimization problem, which usually involves some iterative procedure. A notable exception is when one deals with an affine halfspace, say $H = \big\{y \in \mathcal{H} \mid \langle y \mid t\rangle \leq \eta\big\}$, where $t \neq 0$. The projection of x onto H is computable explicitly as

$$P_H x = \begin{cases} x + \dfrac{\eta - \langle x \mid t\rangle}{\|t\|^2} t & \text{if } \langle x \mid t\rangle > \eta, \\ x & \text{if } \langle x \mid t\rangle \leq \eta. \end{cases} \tag{3.32}$$

Therefore, if C can be approximated at x by a halfspace H_x, an economical approximation to $P_C x$ will be $P_{H_x} x$. We shall now describe a general context in which it is possible to do so.

Suppose $C = \operatorname{lev}_{\leq 0} f \triangleq \big\{y \in \mathcal{H} \mid f(y) \leq \eta\big\} \neq \varnothing$, where $f\colon \mathcal{H} \to \mathbb{R}$ is continuous and convex (this level set representation of a closed convex set is quite general since one can always put $C = \operatorname{lev}_{\leq 0} d_C$). Fix $x \in \mathcal{H}$ and a subgradient $t \in \partial f(x)$. Then (3.12) provides a linearization of f at x as

$$(\forall y \in \mathcal{H}) \;\; \langle y - x \mid t\rangle + f(x) \leq f(y). \tag{3.33}$$

Moreover, if $f(y) \leq 0$, then $\langle y - x \mid t\rangle + f(x) \leq 0$ and, therefore, the closed affine halfspace

$$H_x = \big\{y \in \mathcal{H} \mid \langle x - y \mid t\rangle \geq f(x)\big\} \tag{3.34}$$

contains C. The projection of x onto H_x is a *subgradient projection* of x onto C. In view of (3.32), it is given by

$$G_f x = P_{H_x} x = \begin{cases} x - \dfrac{f(x)t}{\|t\|^2} & \text{if } f(x) > 0, \\ x & \text{if } f(x) \leq 0. \end{cases} \tag{3.35}$$

To sum up,

$$C \subset H_x = \big\{y \in \mathcal{H} \mid \langle y - P_{H_x} x \mid x - P_{H_x} x\rangle \leq 0\big\}. \tag{3.36}$$

We also note that, in view of (3.30), if $f = d_C$ the notion of subgradient projection reverts to the usual notion of projection, i.e., $P_{H_x} x = P_C x$.

It should be stressed that, although a subgradient projector is not continuous in general, it is a $\mathfrak{T}$-class operator [5].

Fact 3.22 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function. Then a subgradient projector G_f onto a level set $\operatorname{lev}_{\leq 0} f \neq 0$ is a $\mathfrak{T}$-class operator. In particular*

$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{lev}_{\leq 0} f) \;\; \|G_f x - y\|^2 \leq \|x - y\|^2 - \|G_f x - x\|^2. \tag{3.37}$$

Moreover, $\operatorname{Fix} G_f = \operatorname{lev}_{\leq 0} f$.

3.8 Proximity operator

Fact 3.23 *Let $f \in \Gamma_0(\mathcal{H})$ and $x \in \mathcal{H}$. Then $\operatorname{prox}_f x$ is the unique minimizer of the function $y \mapsto f(y) + \frac{1}{2}\|y - x\|^2$. We have $\operatorname{prox}_f \in \mathfrak{T}$ (actually prox_f is firmly nonexpansive) and $\operatorname{Fix} \operatorname{prox}_f = \operatorname{Argmin} f$.*

Note that if $f = \iota_C$, where C is a nonempty closed convex set, then $\operatorname{prox}_f = P_C$. Moreover we have the following extension of (3.26): $y = \operatorname{prox}_f x$ if and only if

$$(\forall z \in \mathcal{H}) \;\; \langle z - y \mid x - y\rangle + f(y) \leq f(z). \tag{3.38}$$

4 Fixed point formulation of image recovery problems

As suggested by (3.19), a closed and convex property set may be represented as the fixed point set of some Fejér operator. Since $\operatorname{Fix} T = \operatorname{Fix}(\operatorname{Id}+T)/2$, it is equivalent (and actually more convenient) to work with $\mathfrak{T}$ class operators. We observe that this $\mathfrak{T}$-class representation is quite general since any nonempty closed convex set S_i can always be written as

$$S_i = \operatorname{Fix} P_i, \tag{4.1}$$

where $P_i \in \mathfrak{T}$ (see Fact 3.19) is the projector onto S_i. In some problems, however, the projector may not be easy to implement and it is more advantageous to use other operators. For instance, suppose that a certain property of the image is described by a convex inequality, say

$$S_i = \operatorname{lev}_{\leq 0} f_i, \tag{4.2}$$

where $f_i \colon \mathcal{H} \to \mathbb{R}$ is a continuous convex function. This formulation occurs explicitly in (1.6), (1.7), (1.8), (1.9), (1.12), and (1.16) for example. In such instances it is more economical to write (see Fact 3.22)

$$S_i = \operatorname{Fix} G_i, \tag{4.3}$$

where $G_i = G_{f_i}$ is the subgradient projector defined in (3.35). Indeed, implementing $G_i x$ merely amounts to calculating a subgradient (the gradient in the differentiable case) of f_i at x. Now suppose that original image is known to be invariant under a unitary operator U_i (e.g., symmetry property if U_i is a rotation, etc.). Then $T_i = (\operatorname{Id} + U_i)/2 \in \mathfrak{T}$ and

$$S_i = \operatorname{Fix} T_i. \tag{4.4}$$

Next, suppose that the original image is known to be a minimizer of a proper lower semicontinuous convex function $f_i \colon \mathcal{H} \to \left]-\infty, +\infty\right]$, i.e.,

$$S_i = \operatorname{Argmin} f_i. \tag{4.5}$$

Then we can write (see Fact 3.23)

$$S_i = \operatorname{Fix} \operatorname{prox}_{f_i}. \tag{4.6}$$

To sum up, in its general form, the recovery problem (1.19) can be written as

$$\text{Find } x \in \underbrace{S = \bigcap_{i \in I} \operatorname{Fix} T_i}_{\text{a priori information}} \quad \text{such that} \quad \underbrace{f(x) = \min f(S)}_{\text{selection criterion}}, \quad \text{where} \quad (\forall i \in I)\ T_i \in \mathfrak{T}. \tag{4.7}$$

Henceforth, we assume to avoid technical details that I is a finite, say $I = \{1, \dots, m\}$.

We conclude this preamble by noting that the first explicit formulation of a signal recovery problem as a fixed point problem seems to be [67] (see also [57]), where a single nonexpansive operator in $\mathbb{R}^N$ was used to model certain spatial and spectral properties in a signal reconstruction problem.

5 Solving feasibility problems

If f is a constant function in (4.7), the recovery problem becomes a *convex feasibility problem*, namely

$$\text{Find} \quad x \in S = \bigcap_{i \in I} S_i, \quad \text{where} \quad (\forall i \in I) \ \ S_i = \operatorname{Fix} T_i \ \text{ and } \ T_i \in \mathfrak{T}. \tag{5.1}$$

Unless stated otherwise, it will be assumed that $S \neq \varnothing$ (see Section 5.5 for the inconsistent case). This feasibility approach is known as the *set theoretic approach.* In set theoretic estimation, all the members of the feasibility set S are acceptable solutions. They can be regarded as the images which, in light of all available information, may have given rise to the observed data. The only way to restrict objectively the feasibility set is to incorporate more information in the formulation. If some of the feasible solutions are not acceptable, then it must be the case that the formulation fails to include some constraint which has not been identified. Once the set based on this constraint is incorporated, any point in the feasibility set should be acceptable; if not the cycle is repeated. Usually, there is more than one solution, which reflects the inherent uncertainty surrounding the problem formulation.

Historically, the set theoretic approach finds its root in [38, 51, 74]. Major developments came in the 1980's with the papers [14, 18, 33, 58, 65, 70, 75]. These concepts were later formalized in [19]; see also [22] for an image recovery perspective. Further applications and a selection of more recent developments can be found in [1, 10, 13, 15, 23, 30, 31, 32, 41, 44, 52, 56, 62, 63, 64, 66, 71].

5.1 Cyclic projection method (POCS)

Because of its historical [19, 75] and conceptual importance, we start by presenting the cyclic method of projections onto convex sets known as POCS (see Fig. 6). Underlying this approach is the assumption that the projections onto the sets are relatively easy to compute.

Denote by P_i the projector onto S_i. It follows from (3.28) that

$$(\forall (x,y) \in S_i \times \mathcal{H}) \ \ \|P_i y - x\|^2 \leq \|y - x\|^2 - \|P_i y - y\|^2. \tag{5.2}$$

Lemma 5.1 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, S a nonempty subset of $\mathcal{H}$, and $(\nu_n)_{n\in\mathbb{N}}$ a sequence in $[0, +\infty[$ such that*

$$(\forall x \in S)(\forall n \in \mathbb{N}) \ \ \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - \nu_n. \tag{5.3}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in S if for every strictly increasing sequence $(l_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$

$$\begin{cases} \nu_{l_n} \to 0 \\ x_{l_n} \rightharpoonup y \end{cases} \quad \Rightarrow \quad y \in S. \tag{5.4}$$

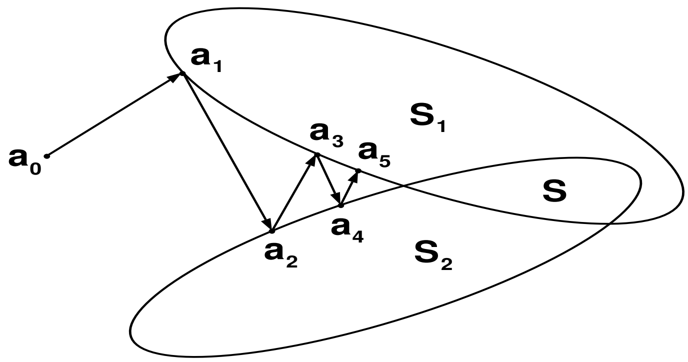

Figure 6: POCS.

Algorithm 5.2 (POCS) Fix $x_0 \in \mathcal{H}$ and do $(\forall n \in \mathbb{N})$ $x_{n+1} = P_m \cdots P_1 x_n$.

This scheme converges weakly to a point in S.

Proof. Fix $x \in S$ and $n \in \mathbb{N}$. Repeated applications of (5.2) yield

$$\begin{aligned} \|x_{n+1} - x\|^2 &= \|P_m(P_{m-1} \cdots P_1 x_n) - x\|^2 \\ &\leq \|P_{m-1}(P_{m-2} \cdots P_1 x_n) - x\|^2 - \|P_m \cdots P_1 x_n - P_{m-1} \cdots P_1 x_n\|^2 \\ &= \|x_n - x\|^2 - \nu_n, \end{aligned} \tag{5.5}$$

where

$$\nu_n = \|P_1 x_n - x_n\|^2 + \|P_2 P_1 x_n - P_1 x_n\|^2 + \cdots + \|P_m \cdots P_1 x_n - P_{m-1} \cdots P_1 x_n\|^2. \tag{5.6}$$

In view of Lemma 5.1, it remains to check (5.4). Suppose $x_{l_n} \rightharpoonup x$ and $\nu_{l_n} \to 0$. Then $P_1 x_{l_n} - x_{l_n} \to 0$ and therefore $S_1 \ni P_1 x_{l_n} \rightharpoonup x$, whence $x \in S_1$ since S_1 is closed and convex, hence weakly closed (Fact 3.6). Next, since $P_1 x_{l_n} \rightharpoonup x$ and $P_2 P_1 x_{l_n} - P_1 x_{l_n} \to 0$, we get $S_2 \ni P_2 P_1 x_{l_n} \rightharpoonup x$ and therefore $x \in S_2$ since S_2 is closed. Continuing this process, we obtain $x \in S_i$ for every $i \in \{1, \ldots, m\}$. ☐

5.2 Cyclic subgradient projection method

Each set takes the form $S_i = \operatorname{lev}_{\leq 0} f_i$, where $f_i \colon \mathcal{H} \to \mathbb{R}$ is convex and continuous. Then the subgradient projector onto S_i is given by (3.35) to be

$$G_i \colon x \mapsto \begin{cases} x - \dfrac{f_i(x)}{\|t\|^2} t \ \text{ where } \ t \in \partial f_i(x) & \text{if } \ f_i(x) > 0 \\ x & \text{if } \ f_i(x) \leq 0. \end{cases} \tag{5.7}$$

The subgradient inequality (3.33) yields

$$\begin{aligned} (\forall (x,y) \in S_i \times \mathcal{H}) \ \ \|G_i y - x\|^2 &= \|y - x\|^2 - 2\langle y - G_i y \mid G_i y - x\rangle - \|G_i y - y\|^2 \\ &\leq \|y - x\|^2 - \|G_i y - y\|^2. \end{aligned} \tag{5.8}$$

Algorithm 5.3 Fix $x_0 \in \mathcal{H}$ and do $(\forall n \in \mathbb{N})$ $x_{n+1} = G_m \cdots G_1 x_n$.

This scheme converges weakly to a point in S if, for every i, ∂f_i maps bounded sets to bounded sets.

Proof. Fix $x \in S$ and $n \in \mathbb{N}$. Repeated applications of (5.8) yield

$$\begin{aligned} \|x_{n+1} - x\|^2 &= \|G_m(G_{m-1} \cdots G_1 x_n) - x\|^2 \\ &\leq \|G_{m-1}(G_{m-2} \cdots G_1 x_n) - x\|^2 - \|G_m \cdots G_1 x_n - G_{m-1} \cdots G_1 x_n\|^2 \\ &= \|x_n - x\|^2 - \nu_n, \end{aligned} \tag{5.9}$$

where

$$\nu_n = \|G_1 x_n - x_n\|^2 + \|G_2 G_1 x_n - G_1 x_n\|^2 + \cdots + \|G_m \cdots G_1 x_n - G_{m-1} \cdots G_1 x_n\|^2. \tag{5.10}$$

In view of Lemma 5.1, it remains to check (5.4). Suppose $x_{l_n} \rightharpoonup x$ and $\nu_{l_n} \to 0$. Since f_i is convex and continuous, it is weak lower semicontinuous and therefore $f_1(x) \leq \underline{\lim} f_1(x_{l_n})$. However, since $(x_{l_n})_{n\in\mathbb{N}}$ is bounded and ∂f_1 maps bounded sets to bounded sets (without loss of generality assume $(x_{l_n})_{n\in\mathbb{N}}$ lies outside of S_1),

$$\|G_1 x_{l_n} - x_{l_n}\| \to 0 \ \Rightarrow \ \frac{f_1(x_{l_n})}{\|t_{1,l_n}\|} \to 0 \ \Rightarrow \ f_1(x_{l_n}) \to 0. \tag{5.11}$$

Hence $f_1(x) \leq 0$, i.e., $x \in S_1$. Next, since $G_1 x_{l_n} \rightharpoonup x$, $f_2(x) \leq \underline{\lim} f_2(G_1 x_{l_n})$. However,

$$\|G_2 G_1 x_{l_n} - G_1 x_{l_n}\| \to 0 \ \Rightarrow \ \frac{f_2(G_1 x_{l_n})}{\|t_{2,l_n}\|} \to 0 \ \Rightarrow \ f_2(G_1 x_{l_n}) \to 0. \tag{5.12}$$

Hence $f_2(x) \leq 0$, i.e., $x \in S_2$. Continuing this process, we obtain $x \in S_i$ for every $i \in \{1, \ldots, m\}$. ☐

5.3 General $\mathfrak{T}$-class algorithm

We now present a general algorithm to solve (4.7) in the general case [26]. It extends the above methods in several directions:

- Arbitrary $\mathfrak{T}$-class operators are used.
- A set S_i is not assigned a fixed operator T_i but, rather, iteration-dependent operators $(T_{i,n})_{n\in\mathbb{N}}$.
- Several operators can be used simultaneously at each iteration (parallel, block-iterative processing).
- Relaxation parameters are introduced.
- The operators are assigned varying weights.

In the following algorithm, the function L is that defined in (3.23).

Algorithm 5.4 Fix $x_0 \in \mathcal{H}$. At iteration $n \in \mathbb{N}$, x_n designates the current iterate and the update x_{n+1} is constructed according to the following steps.

① $\varnothing \neq I_n \subset I$.

② $(\forall i \in I_n)\ \ T_{i,n} \in \mathfrak{T}$ and $\operatorname{Fix} T_{i,n} = S_i$.

③ $(\forall i \in I_n)\ \omega_{i,n} \in \left]0,1\right]$ and $\sum_{i\in I_n} \omega_{i,n} = 1$.

④ $T_n = \mathrm{Id} + L\big(\,\cdot\,,(T_{i,n})_{i\in I_n},(\omega_{i,n})_{i\in I_n}\big)\big(\sum_{i\in I_n}\omega_{i,n}T_{i,n} - \mathrm{Id}\big)$.

⑤ $x_{n+1} = x_n + \rho_n(T_n x_n - x_n)$, where $0 < \rho_n < 2$.

In view of Proposition 3.16, $T_n \in \mathfrak{T}$ and the convergence of Algorithm 5.4 can be analyzed using general results on the convergence of relaxed $\mathfrak{T}$-class iterations [26]. Thus, under suitable conditions, every orbit generated by Algorithm 5.4 converges weakly to a point in S.

More practically, the updating rule is (for $x_n \notin \bigcap_{i\in I_n} S_i$)

$$x_{n+1} = x_n + \lambda_n\left(\sum_{i\in I_n}\omega_{i,n}T_{i,n}x_n - x_n\right), \quad \text{where } \ 0 < \lambda_n < 2\frac{\sum_{i\in I_n}\omega_{i,n}\|T_{i,n}x_n - x_n\|^2}{\|\sum_{i\in I_n}\omega_{i,n}T_{i,n}x_n - x_n\|^2}. \tag{5.13}$$

In particular if, for every $n \in \mathbb{N}$, I_n is the singleton $\{n \text{ modulo } m+1\}$ (then $L \equiv 1$), $T_{i,n} = G_i$, and $\lambda_n \equiv 1$, then we recover Algorithm 5.3. Various other algorithms, including those of [4, 21, 24, 45] are thus unified and extended by Algorithm 5.4, in particular SIRT (Simultaneous Iterative Reconstruction Technique, see Fig. 7)

$$(\forall n \in \mathbb{N})\ \ x_{n+1} = \frac{1}{m}\sum_{i=1}^{m} P_i x_n. \tag{5.14}$$

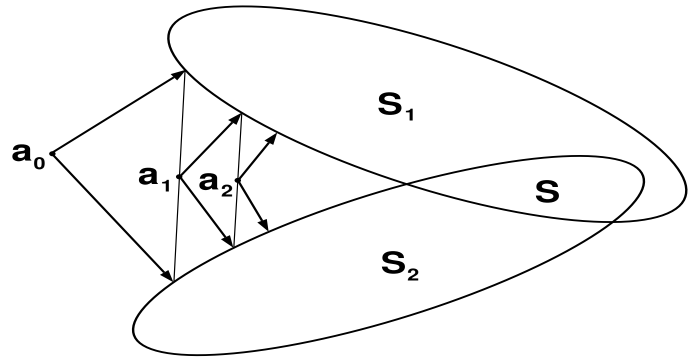

Figure 7: SIRT.

When only subgradient projectors are used in (5.13), we shall call the resulting method EMOPSP (Extrapolated Method of Parallel Subgradient Projections [23]), that is,

$$x_{n+1} = x_n + \lambda_n \left(\sum_{i \in I_n} \omega_{i,n} G_i x_n - x_n \right), \quad \text{where} \;\; 0 < \lambda_n < 2L_n, \;\; L_n = \frac{\sum_{i \in I_n} \omega_{i,n} \|G_i x_n - x_n\|^2}{\|\sum_{i \in I_n} \omega_{i,n} G_i x_n - x_n\|^2}. \tag{5.15}$$

The fact that the extrapolation parameter L_n and, thereby, the relaxation parameter λ_n can attain very large values results in an algorithm which is much faster than conventional methods. This is illustrated next.

5.4 Numerical illustration of Algorithm 5.4

All images have $N \times N$ pixels ($N = 128$). $\mathcal{H}$ is the usual N^2-dimensional Euclidean space and $\mathfrak{F}$ is the two-dimensional DFT operator, i.e., $(\forall x \in \mathcal{H})\;\; \mathfrak{F}(x) = \widehat{x}$, where for every (k,l) in $\{0, \ldots, N-1\}^2$

$$\widehat{x}(k,l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x^{(Ni+j)} \exp(-\imath 2\pi(ik + jl)/N). \tag{5.16}$$

The original image h of Fig. 8 is degraded by convolutional blur with a uniform 9×9 kernel ℓ and addition of uniform purely white noise u with range $[0, R]$ resulting in a blurred image-to-noise

Figure 8: Original image.

Figure 9: Degraded image.

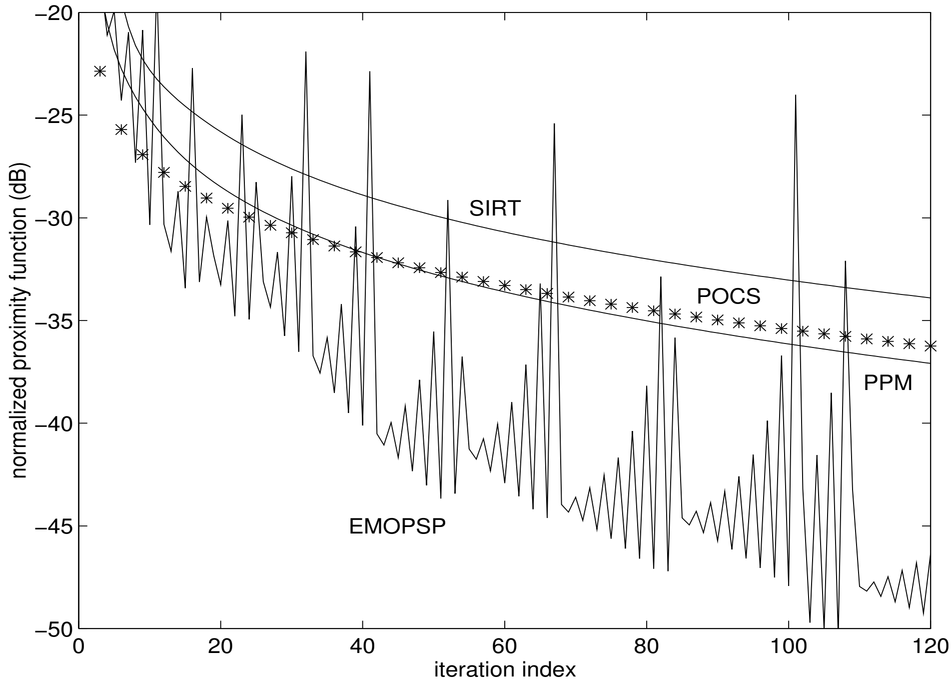

Figure 10: EMOPSP versus POCS and SIRT.

ratio of 35 dB. The degraded image y is shown in Fig. 9. Using the usual column stacking [2], it can be written as $y = Lh + u$, where L is the block-Tœplitz matrix associated with the point spread function ℓ [2]. The problem is to estimate h given x and some *a priori* information about h, ℓ, and u. The first property set $S_1 = (\mathbb{R}_+)^{N^2}$ arises from the nonnegativity of pixel values. Next, it is assumed that the discrete Fourier transform of h is known on one fourth of its support for low frequencies in both directions. The associated property set is $S_2 = \{x \in \mathcal{H} \mid \widehat{x}1_K = \widehat{h}1_K\}$, where K contains the set of frequency pairs $\{0, \ldots, N/8-1\}^2$ as well as all those resulting from the symmetry properties of the two-dimensional DFT of real images. The projections of an image x_n onto S_1 and S_2 are given by the closed-form expressions

$$\begin{cases} P_1 x_n = x_n^+ = [\max\{0, x_n^{(i)}\}]^\top_{0 \leq i \leq N^2-1} \triangleq G_1 x_n, \\ P_2 x_n = \mathfrak{F}^{-1}(\widehat{h}1_K + \widehat{x_n}1_{\complement K}) \triangleq G_2 x_n. \end{cases} \tag{5.17}$$

The algorithm will be initialized with $x_0 = y$ and the progression of its orbit $(x_n)_{n\in\mathbb{N}}$ will be tracked by plotting the normalized decibel values $(10\log_{10}(\Phi(x_n)/\Phi(x_0)))_{n\in\mathbb{N}}$ of the proximity function

$$\begin{aligned} \Phi\colon \mathcal{H} &\to [0, +\infty[\\ x &\mapsto \frac{1}{2}\sum_{i\in I} w_i d_{S_i}(x,)^2, \end{aligned} \tag{5.18}$$

where

$$(\forall i \in I)\ \ w_i = 1/(\operatorname{card} I). \tag{5.19}$$

Figure 11: Restored image (EMOPSP).

It is assumed here that the information available about the noise vector u is that its components are independent and all distributed as a random variable U with known second and fourth moments (the actual distribution of the noise is not supposed to be known here; if it were exploited, results could be further improved [23]). As shown in [33], with a 95 percent confidence level, this information leads to the property set

$$S_3 = \{x \in \mathcal{H} \mid f_3(x) = \|y - Lx\|^2 \leq \rho\}, \tag{5.20}$$

where

$$\rho = N^2 \mathsf{E}|U|^2 + 1.96 N \sqrt{\mathsf{E}|U|^4 - \mathsf{E}^2|U|^2}. \tag{5.21}$$

This set has proven quite useful in a number of applications [68] but, unfortunately, its projection operator must be approximated iteratively [70]. By contrast, using (5.7) and the fact that $\nabla f_3(x_n) = \nabla(\|y - Lx_n\|^2 - \rho) = -2\, L^\top (y - Lx_n)$, we simply process the set S_3 at iteration n with the subgradient projection

$$G_3 x_n = \begin{cases} x_n + \dfrac{\|r_n\|^2 - \rho}{2\, \|L^\top r_n\|^2}\, L^\top r_n & \text{if } \|r_n\|^2 > \rho, \\ x_n & \text{otherwise,} \end{cases} \tag{5.22}$$

Figure 12: Restored image (Wiener filtering).

where $r_n = y - Lx_n$. The upper expression in (5.22) can be evaluated in the frequency domain efficiently via the two-dimensional FFT as

$$G_3 x_n = \mathfrak{F}^{-1}\left(\widehat{x_n} + \frac{\|\widehat{r_n}\|^2 - N^2\rho}{2\|\overline{\widehat{\ell}}\,\widehat{r_n}\|^2}\,\overline{\widehat{\ell}}\,\widehat{r_n}\right), \tag{5.23}$$

where $\widehat{r_n} = \widehat{y} - \widehat{\ell}\,\widehat{x_n}$. The subgradient projection reduces the cost of processing S_3 by at least an order of magnitude compared to a standard projection.

The constraint sets for this problem are $(S_i)_{1 \leq i \leq 3}$. In the results shown in Fig. 10, POCS is defined in Algorithm 5.2 and SIRT in (5.14). Furthermore, $P = 3$ parallel processors are available. Since only three sets are present, EMOPSP is implemented with static control ($I_n \equiv \{1, 2, 3\}$), fixed weights as in (5.19), and relaxation strategy $(\forall n \in \mathbb{N})\ \ \lambda_n = L_n$. POCS is faster than SIRT, but clearly outperformed by EMOPSP which uses extrapolated relaxations. The restored image appears in Fig. 11 (see also Fig. 12 for a qualitative comparison with Wiener filtering). Let us observe that these results show performance only in terms of the number of iterations required to reach a given degree of unfeasibility. However, to fully appreciate the numerical superiority of EMOPSP, it must be borne in mind that to process the set S_3 the algorithms POCS and SIRT must use the costly projection onto S_3 whereas EMOPSP needs only the approximate projection

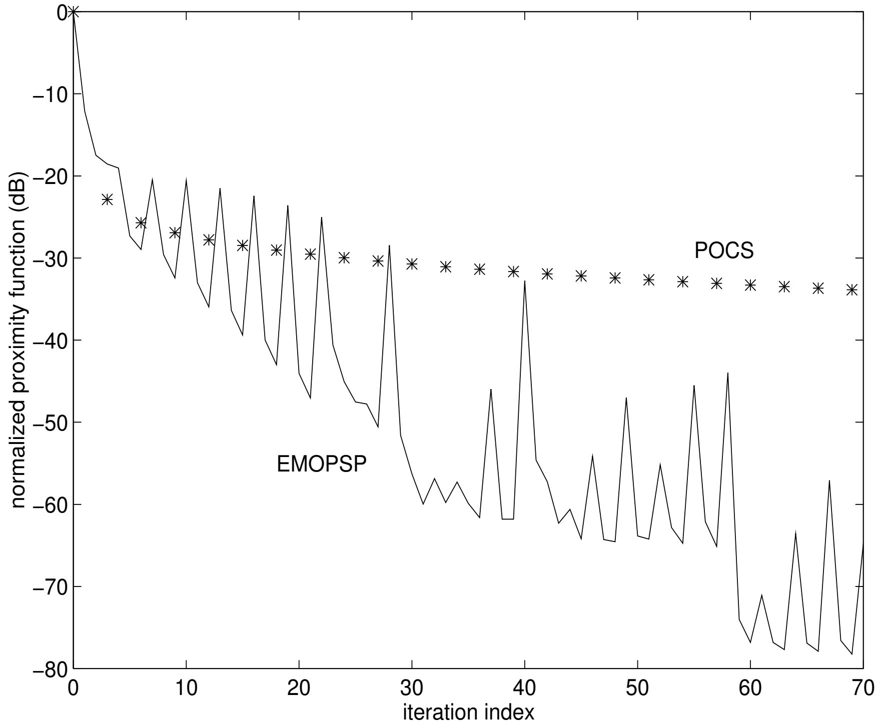

Figure 13: EMOPSP with centering.

(5.22). Finally, the performance can be further improved by using the "centering" technique

$$(\forall n \in \mathbb{N})\ \ \lambda_n = \begin{cases} L_n/2 & \text{if } n = 2 \text{ modulo } 3, \\ L_n & \text{otherwise.} \end{cases} \tag{5.24}$$

The resulting convergence patterns are displayed in Fig. 13.

5.5 Parallel methods for inconsistent problems

In writing (5.1), it is tacitly assumed that the problem is consistent in the sense that the constraints are compatible so that $S \neq \varnothing$. However, signal feasibility problems may turn out to be inconsistent for a variety of reasons. In design problems this situation typically results from the incorporation of too demanding and therefore conflicting specifications. In estimation problems it may be due to inaccurate deterministic constraints, to overly aggressive confidence levels on stochastic constraints, or to inadequate data modeling. Naturally, when the feasibility problem is inconsistent, $S = \varnothing$ and (5.1) must be reformulated in a physically meaningful way. We now follow an approach proposed in [28].

Let us assume, as in (5.1), that the constraints take the form of minimizing properties, say

$$(\forall i \in I)\ \ S_i = \operatorname{Argmin} \varphi_i, \quad \varphi_i \in \Gamma_0(\mathcal{H}). \tag{5.25}$$

Now let $(\gamma_i)_{i\in I}$ be real numbers in $]0,+\infty[$ such that $\sum_{i\in I} 1/\gamma_i = 1$. The *Moreau envelope* of φ_i of index γ_i is the function

$$^{\gamma_i}\varphi_i\colon x \mapsto \min_{y\in\mathcal{H}} \varphi_i(y) + \frac{1}{2\gamma_i}\|y-x\|^2. \tag{5.26}$$

When the original common minimizer problem defined by (5.1)&(5.25) has no solution it can be approximated by the problem of minimizing the function

$$\begin{aligned}\Phi\colon \mathcal{H} &\to \mathbb{R}\\ x &\mapsto \sum_{i\in I} {}^{\gamma_i}\varphi_i(x),\end{aligned} \tag{5.27}$$

where

$$(\forall i\in I)\ \ \gamma_i > 0 \quad \text{and} \quad \sum_{i\in I} 1/\gamma_i = 1. \tag{5.28}$$

This approximation is justified by the fact that if the original problem had solutions, then they can be shown to coincide with the set of minimizers to (5.27). In general, however, (5.27) admits minimizers under mild conditions even though the original problem may not. Indeed, Φ is a very well behaved function: it is finite, convex, continuous, and differentiable with a Lipschitz gradient. Let

$$(\forall i\in I)\ \ \omega_i = 1/\gamma_i. \tag{5.29}$$

Then it can be shown that

$$\nabla\, {}^{\gamma_i}\varphi_i = \omega_i(\mathrm{Id} - \mathrm{prox}_{\gamma_i\varphi_i}). \tag{5.30}$$

Now, since $\mathrm{prox}_{\gamma_i\varphi_i}$ is firmly nonexpansive (Fact 3.23), so is $\mathrm{Id} - \mathrm{prox}_{\gamma_i\varphi_i}$ (see (3.19)), and we conclude that the convex combination

$$\nabla\Phi = \sum_{i\in I}\omega_i(\mathrm{Id} - \mathrm{prox}_{\gamma_i\varphi_i}) = \mathrm{Id} - \sum_{i\in I}\omega_i\,\mathrm{prox}_{\gamma_i\varphi_i} \tag{5.31}$$

is also firmly nonexpansive, hence 1-Lipschitz. Furthermore, it can be shown that the iterates

$$x_{n+1} = x_n + \lambda_n\left(\sum_{i\in I}\omega_i\,\mathrm{prox}_{\gamma_i\varphi_i} x_n - x_n\right),\quad 0<\varepsilon\le\lambda_n\le 2-\varepsilon \tag{5.32}$$

converge weakly to a minimizer of (5.27). Let us provide two special cases:

- Suppose that the original constraints are defined via convex inequalities, as in (4.2). Then the original problem is to

  $$\text{Find } x\in\mathcal{H} \text{ such that } \max_{i\in I} f_i(x) \le 0. \tag{5.33}$$

  This problem can be recast in the format (5.1)&(5.25) by setting, for every $i\in I$, $\varphi_i = \max\{0, f_i\}^2$.

- Suppose that the projectors $(P_i)_{i\in I}$ are easy to implement, i.e., the original problem is formulated via (5.1)&(4.1). Then we certainly recover the formulation (5.1)&(5.25) by setting, for every $i \in I$, $\varphi_i = \iota_{S_i}$. Moreover, ${}^{\gamma_i}\varphi_i = \frac{1}{2}d_{S_i}^2$ and (5.27) becomes

$$\begin{aligned}\Phi\colon \mathcal{H} &\to \mathbb{R}\\ x &\mapsto \frac{1}{2}\sum_{i\in I} w_i d_{S_i}(x)^2.\end{aligned} \tag{5.34}$$

In other words, one seeks an image, which best approximates – in a least squared distance sense – the inconsistent constraints. This framework is discussed in [20], where specific applications are described. In this case, a simple sufficient condition for $\operatorname{Argmin}\Phi \neq \varnothing$ is that one of the sets be bounded. Note also that (5.32) reduces to the parallel projection method

$$x_{n+1} = x_n + \lambda_n\left(\sum_{i\in I}\omega_i P_i x_n - x_n\right), \quad 0 < \varepsilon \leq \lambda_n \leq 2-\varepsilon. \tag{5.35}$$

This process is illustrated in Figure 14: the current iterate is $a_n = x_n$, and the update x_{n+1} is on the segment between a_n and $2d_n - a_n$, where $d_n = \sum_{i\in I}\omega_i P_i x_n$.

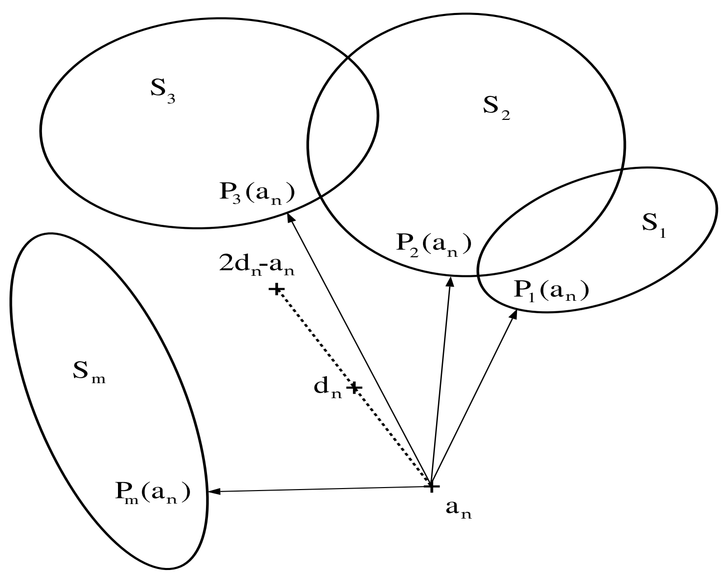

Figure 14: Parallel projection method for inconsistent problems.

6 Constrained minimization methods

In some problems, it may be justified, on the basis of physical or empirical considerations, to select a particular feasible image by minimizing a function $f \in \Gamma_0(\mathcal{H})$ over the feasibility set S, as described in the problem formulation (4.7). Examples of such functions are provided in item (iii) of Section 1.

As before, the feasibility set S is closed and convex, and we assume that it is nonempty. Therefore $\iota_S \in \Gamma_0(\mathcal{H})$. Moreover, (4.7) amounts to minimizing $f + \iota_S$ over $\mathcal{H}$. In view of Fermat's rule (Proposition 3.13(i)), x solves the problem if

$$0 \in \partial(f + \iota_S)(x) = \partial f(x) + \partial \iota_S(x) = \partial f(x) + N_S(x), \tag{6.1}$$

where the first identity assumes some mild qualification condition, e.g., f is finite and continuous at a point in S, and the second follows from Example 3.12. To simplify the analysis, suppose that f is differentiable on S. Then (6.1) becomes $-\nabla f(x) \in N_S(x)$ or, given an arbitrary $\gamma \in \,]0, +\infty[$,

$$-\gamma \nabla f(x) \in N_S(x). \tag{6.2}$$

Equivalently (see Example 3.12)

$$x \in S \ \text{ and } \ (\forall y \in S) \ \ \langle y - x \mid -\gamma \nabla f(x) \rangle \leq 0, \tag{6.3}$$

that is

$$x \in S \ \text{ and } \ (\forall y \in S) \ \ \big\langle y - x \mid \big(x - \gamma \nabla f(x)\big) - x \big\rangle \leq 0. \tag{6.4}$$

In view of (3.26), we conclude that $x = P_S\big(x - \gamma \nabla f(x)\big)$, i.e.,

$$x \in \operatorname{Fix} P_S \circ \big(\mathrm{Id} - \gamma \nabla f\big). \tag{6.5}$$

Now suppose that ∇f is $1/\beta$-Lipschitz-continuous. Then the operator $P_S \circ \big(\mathrm{Id} - \gamma \nabla f\big)$ is nonexpansive for $0 < \gamma \leq 2\beta$. To find a fixed point (that is a minimizer of f over S) consider the extended projected gradient iteration

$$x_{n+1} = x_n + \lambda_n \bigg(P_S\big(x_n - \gamma_n(\nabla f(x_n) + b_n)\big) + a_n - x_n \bigg), \tag{6.6}$$

where $\gamma_n \in \,]0, 2\beta[$, $(a_n, b_n) \in \mathcal{H}^2$ represent numerical errors, and $\lambda_n \in \,]0, 1]$. Then, as shown in [28], every sequence so generated converges weakly to a solution under the conditions

(i) $\underline{\lim}\, \lambda_n > 0$ and $0 < \underline{\lim}\, \gamma_n \leq \overline{\lim}\, \gamma_n < 2\beta$.

(ii) $\sum_{n \in \mathbb{N}} \|a_n\| < +\infty$ and $\sum_{n \in \mathbb{N}} \|b_n\| < +\infty$.

This result is applicable only in instances when P_S can be implemented easily, which is seldom the case in practice.

In general, it is necessary, as in Section 5, to decompose the problem and to use operators acting on the constraint sets $(S_i)_{i\in I}$ individually. Such methods are discussed in [25] and [27]. For simplicity, we discuss only the simple case when $f(x) = \|x - r\|^2$, i.e., when one seeks the feasible image which is the closest to a reference image r (a minimum energy solution for $r = 0$).

Given a triple (x, y, z) in $\mathcal{H}^3$, set $\pi = \langle x - y \mid y - z\rangle$, $\mu = \|x-y\|^2$, $\nu = \|y-z\|^2$, and $\rho = \mu\nu - \pi^2$. Now define

$$Q(x,y,z) = \begin{cases} z & \text{if } \rho = 0 \text{ and } \pi \geq 0, \\ x + (1+\pi/\nu)(z-y) & \text{if } \rho > 0 \text{ and } \pi\nu \geq \rho, \\ y + \dfrac{\nu}{\rho}(\pi(x-y) + \mu(z-y)) & \text{if } \rho > 0 \text{ and } \pi\nu < \rho. \end{cases} \tag{6.7}$$

Algorithm 6.1 Set $x_0 = r$ and do $(\forall n \in \mathbb{N})$ $x_{n+1} = Q(x_0, x_n, T_n x_n)$, where T_n is as in Algorithm 5.4.

Under suitable assumptions, the above iterates converge strongly to the projection of r onto S [25]. Image recovery applications of this result can be found in [27] and [31].

7 Remarks on nonconvex problems

Nonconvex problems are much more complex numerically as the powerful tools described in the preceding sections are no longer available. Nonetheless, fixed point concepts are still useful in this context and lead to (locally convergent) numerical methods. Applications to the important problem of phase retrieval – i.e., the reconstruction of an image from the knowledge of the magnitude of its Fourier transform (see (1.14)) and side information – are discussed in [6] and [7].

References

[1] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Trans. Circ. Syst. Video Technol.,* vol. 12, pp. 217–226, 2002.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

[3] R. H. T. Bates and M. J. McDonnell, *Image Restoration and Reconstruction.* New York: Clarendon Press, 1986.

[4] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.,* vol. 38, pp. 367–426, 1996.

[5] H. H. Bauschke and P. L. Combettes, "A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces," *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

[6] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *J. Opt. Soc. Amer. A,* vol. 19, pp. 1334–1345, 2002.

[7] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection-reflection method for phase retrieval," *J. Opt. Soc. Amer. A,* vol. 20, pp. 1025–1034, 2003.

[8] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization—Theory and Examples.* New York: Springer-Verlag, 2000.

[9] J. M. Borwein, A. S. Lewis, and D. Noll, "Maximum entropy reconstruction using derivative information. I: Fisher information and convex duality," *Math. Oper. Res.,* vol. 21, pp. 442–468, 1996.

[10] A. K. Brodzik and J. M. Mooney, "Convex projections algorithm for restoration of limited-angle chromotomographic images," *J. Opt. Soc. Amer. A,* vol. 16, pp. 246–257, 1999.

[11] D. Butnariu, Y. Censor, and S. Reich (Eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications.* North-Holland, Amsterdam, 2001.

[12] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing,* vol. 2, no. 1, pp. 96–103, 1993. ("Erratum and addendum," vol. 4, no. 2, pp. 226–227, 1995.)

[13] A. E. Çetin, "An iterative algorithm for signal reconstruction from bispectrum," *IEEE Trans. Signal Processing,* vol. 39, pp. 2621–2628, 1991.

[14] A. E. Çetin and R. Ansari, "Convolution-based framework for signal recovery and applications," *J. Opt. Soc. Amer. A,* vol. 5, pp. 1193–1200, 1988.

[15] A. E. Çetin and R. Ansari, "Signal recovery from wavelet transform maxima," *IEEE Trans. Signal Processing,* vol. 42, pp. 194–196, 1994.

[16] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision,* vol. 20, pp. 89–97, 2004.

[17] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.,* vol. 76, pp. 167–188, 1997.

[18] M. R. Civanlar and H. J. Trussell, "Digital signal restoration using fuzzy sets," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 34, pp. 919–936, 1986.

[19] P. L. Combettes, "The fondations of set theoretic estimation," *Proc. IEEE,* vol. 81, pp. 182–208, 1993.

[20] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Trans. Signal Process.,* vol. 42, pp. 2955–2966, 1994.

[21] P. L. Combettes, "Construction d'un point fixe commun à une famille de contractions fermes," *C. R. Acad. Sci. Paris Sér. I Math.,* vol. 320, pp. 1385–1390, 1995.

[22] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics,* vol. 95. New York: Academic, 1996.

[23] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Process.,* vol. 6, pp. 493–506, 1997.

[24] P. L. Combettes, "Hilbertian convex feasibility problem: Convergence of projection methods," *Appl. Math. Optim.*, vol. 35, pp. 311–330, 1997.

[25] P. L. Combettes, "Strong convergence of block-iterative outer approximation methods for convex optimization," *SIAM J. Control Optim.,* vol. 38, pp. 538–565, 2000.

[26] P. L. Combettes, "Quasi-Fejérian analysis of some optimization algorithms," in *Inherently Parallel Algorithms for Feasibility and Optimization* (D. Butnariu, Y. Censor, and S. Reich, Eds.), pp. 115–152. New York: Elsevier, 2001.

[27] P. L. Combettes, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Processing,* vol. 51, pp. 1771–1782, 2003.

[28] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization,* to appear.

[29] P. L. Combettes and J. Luo, "An adaptive level set method for nondifferentiable constrained image recovery," *IEEE Trans. Image Processing,* vol. 11, pp. 1295–1304, 2002.

[30] P. L. Combettes and J. C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Processing,* vol. 13, no. 9, 2004.

[31] P. L. Combettes and J. C. Pesquet, "Wavelet-constrained image restoration," *Int. J. Wavelets, Multiresolution, Information Processing,* to appear.

[32] P. L. Combettes and H. J. Trussell, "Methods for digital restoration of signals degraded by a stochastic impulse response," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 393–401, 1989.

[33] P. L. Combettes and H. J. Trussell, "The use of noise properties in set theoretic estimation," *IEEE Trans. Signal Processing,* vol. 39, pp. 1630–1641, 1991.

[34] S. Dharanipragada and K. S. Arun, "A quadratically convergent algorithm for convex set constrained signal recovery," *IEEE Trans. Signal Processing,* vol. 44, pp. 248–266, 1996.

[35] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 2024–2036, 1989.

[36] F. Deutsch, *Best Approximation in Inner Product Spaces.* New York: Springer-Verlag, 2001.

[37] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels.* Paris: Dunod, 1974; *Convex Analysis and Variational Problems.* Philadelphia, PA: SIAM, 1999.

[38] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta,* vol. 21, pp. 709–720, 1974.

[39] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory.* Cambridge: Cambridge University Press, 1990.

[40] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings.* New York: Marcel Dekker, 1984.

[41] F. J. González-Castaño, U. M. García-Palomares, J. L. Alba-Castro, and J. M. Pousada-Carballo, "Fast image recovery using dynamic load balancing in parallel architectures, by means of incomplete projections," *IEEE Trans. Image Process.,* vol. 10, pp. 493-499, 2001.

[42] B. R. Hunt, "The inverse problem of radiography," *Math. Biosciences,* vol. 8, pp. 161–179, 1970.

[43] B. R. Hunt, "The application of constrained least-squares estimation to image restoration by digital computer," *IEEE Trans. Computers,* vol. 22, pp. 805–812, 1973.

[44] P. Ishwar and P. Moulin, "On the equivalence of set-theoretic and maxent MAP estimation," *IEEE Trans. Signal Process.,* vol. 51, pp. 698–713, 2003.

[45] K. C. Kiwiel and B. Łopuch, "Surrogate projection methods for finding fixed points of firmly nonexpansive mappings," *SIAM J. Optim.*, vol. 7, pp. 1084–1102, 1997.

[46] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations.* Providence, RI: AMS, 2001.

[47] A. Mohammad-Djafari and G. Demoment, "Maximum entropy image reconstruction in x-ray and diffraction tomography," *IEEE Trans. Medical Imaging,* vol. 7, pp. 345–354, 1988.

[48] D. Noll, "Restoration of degraded images with maximum entropy," *J. Gobal Optim.,* vol. 10, pp. 91–103, 1997.

[49] Free software package OCTAVE available at `http://www.octave.org`.

[50] S. Osher and N. Paragios (Eds.), *Geometric Level Set Methods in Imaging, Vision, and Graphics.* New York: Springer, 2003.

[51] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.,* vol. 22, pp. 735–742, 1975.

[52] A. J. Patti and Y. Altunbasak, "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Trans. Image Process.,* vol. 10, pp. 179–186, 2001.

[53] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.

[54] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259–268, 1992.

[55] A. Sabharwal and L. C. Potter, "Convexly constrained linear inverse problems: Iterative least-squares and regularization," *IEEE Trans. Signal Processing,* vol. 46, pp. 2345–2352, 1998.

[56] C. Sánchez-Avila, "Wavelet domain signal deconvolution with singularity-preserving regularization," *Math. Comput. Simul.,* vol. 61, pp. 165–176, 2003.

[57] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative restoration algorithms," *Proc. IEEE,* vol. 69, pp. 432–450, 1981.

[58] M. I. Sezan and H. Stark, "Incorporation of a priori moment information into signal recovery and synthesis problems," *J. Math. Anal. Appl.,* vol. 122, pp. 172–186, 1987.

[59] M. I. Sezan, H. Stark, and S. J. Yeh, "Projection method formulations of Hopfield-type associative memory neural networks," *Appl. Optics,* vol. 29, no. 17, pp. 2616–2622, 1990.

[60] M. I. Sezan and A. M. Tekalp, "Adaptive image restoration with artifact suppression using the theory of convex projections," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 38, pp. 181–185, 1990.

[61] M. I. Sezan and A. M. Tekalp, "Survey of recent developments in digital image restoration," *Opt. Eng.,* vol. 29, pp. 393–404, 1990.

[62] M. I. Sezan and H. J. Trussell, "Prototype image constraints for set-theoretic image restoration," *IEEE Trans. Signal Processing,* vol. 39, pp. 2275–2285, 1991.

[63] G. Sharma, "Set theoretic estimation for problems in subtractive color," *Color Res. Appl.,* vol. 25, pp. 333-348, 2000.

[64] G. Sharma and H. J. Trussell, "Set theoretic signal restoration using an error in variables criterion," *IEEE Trans. Image Processing,* vol. 6, pp. 1692–1697, 1997.

[65] H. Stark (Ed.), *Image Recovery: Theory and Application.* San Diego, CA: Academic, 1987.

[66] H. Stark and Y. Yang, *Vector Space Projections : A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* New York: Wiley, 1998.

[67] V. T. Tom, T. F. Quatieri, M. H. Hayes, and J. H. McClellan, "Convergence of iterative nonexpansive signal reconstruction algorithms," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 29, pp. 1052–1058, 1981.

[68] H. J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 31, pp. 129–136, 1983.

[69] 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.

[70] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 32, pp. 201–212, 1984.

[71] G. B. Unal and A. E. Çetin, "Restoration of error-diffused images using projection onto convex sets," *IEEE Trans. Image Processing,* vol. 10, pp. 1836–1841, 2001.

[72] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.,* vol. 19, pp. 553–572, 2003.

[73] J. Weickert, *Anisotropic Diffusion in Image Processing.* Stuttgart, Germany: Teubner-Verlag, 1998.

[74] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.,* vol. 25, pp. 694–702, 1978.

[75] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 – theory," *IEEE Trans. Medical Imaging,* vol. 1, pp. 81–94, 1982.

[76] C. Zălinescu, *Convex Analysis in General Vector Spaces.* River Edge, NJ: World Scientific Publishing, 2002.