CONSTRAINT CONSTRUCTION IN CONVEX SET THEORETIC SIGNAL RECOVERY VIA STEIN'S PRINCIPLE

P. L. Combettes[1] *and J.-C. Pesquet*[2]

[1]Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France
[2]Institut Gaspard Monge and URA-CNRS 820, Université de Marne la Vallée, 77454 Marne la Vallée Cedex 2, France

ABSTRACT

Convex set theoretic estimation methods have been shown to be effective in numerous signal recovery problems due to their ability to incorporate a wide range of deterministic and probabilistic information in the form of constraints on the solution. To date, probabilistic information has been used exclusively to constrain statistics of the estimation residual to be consistent with known properties of the noise. In this paper, we propose a new technique to construct constraint sets from probabilistic information based on Stein's identity. In this framework, probabilistic attributes of the signal to be recovered are estimated from the data. The proposed approach is applicable to signal formation models involving additive Gaussian noise and it leads to geometrically simple sets that can easily be handled via projection methods. An application to image denoising is demonstrated.

1. INTRODUCTION

We consider the classical linear recovery (reconstruction or restoration) problem of estimating a signal $\overline{x}$ in a real Hilbert space $\mathcal{H}$ from the observation of a signal

$$y = L\overline{x} + u, \tag{1}$$

in a Euclidean space $\mathcal{G}$, where $L\colon \mathcal{H} \to \mathcal{G}$ is a linear operator modeling the signal formation process and $u \in \mathcal{G}$ is a noise component. Solving this problem as a convex feasibility problem consists in finding a signal $x \in \mathcal{H}$ that satisfies all the convex constraints derived from *a priori* knowledge and from the observed data [2, 12, 15] (see also [7, 9, 13, 14] for recent applications of this framework). If we let S_i denote the closed and convex subset of $\mathcal{H}$ of signals satisfying the ith of m constraints, the problem can be conveniently recast in the set theoretic format

$$\text{Find } \; x \in S = \bigcap_{i=1}^{m} S_i. \tag{2}$$

The main asset of the set theoretic approach is to allow for the incorporation of a broad range of information in the definition of a solution. On the other hand, its numerical viability stems from the availability of efficient numerical methods for solving the convex feasibility problem (2) [2]. More generally, one can select a specific feasible signal by minimizing a suitable convex function J over S, i.e., by solving

$$\text{Find } \; x \in S = \bigcap_{i=1}^{m} S_i \quad \text{such that } \; J(x) = \inf J(S). \tag{3}$$

When J is a quadratic function, this problem can be solved with the parallel decomposition methods proposed in [3].

Assuming the operator L to be known in the signal formation model (1), there are various techniques for constructing constraint sets from deterministic *a priori* information about $\overline{x}$ itself [2, 10, 15] or about the additive noise process [2, 5, 12]. To date, probabilistic information has been used exclusively to constrain some statistics of the estimation residual $y - Lx$ to be consistent with known properties of the additive noise u. This approach was initiated in [12] and further developed in [5] and [2].

The purpose of this paper is to introduce a new technique to exploit probabilistic information in set theoretic methods. Under an additive Gaussian noise assumption, we show that statistical constraints on the signal to be recovered can be constructed from certain statistics of the observed signal y via an identity due to Stein [11]. These additional constraint sets enrich the set theoretic formulation of the problem and lead to improved estimates. Numerically, the sets thus obtained turn out to be quite simple geometrically and they therefore lend themselves to straightforward processing via projection techniques. Although Stein's identity has found various applications in statistics [1] as well as in wavelet-based denoising problems [6, 8], its use in set theoretic estimation appears to be new.

In Section 2, we state our basic assumptions, review Stein's identity, describe the proposed set construction technique, and provide a statistical confidence analysis. Numerical aspects are also discussed. Section 3 is devoted to an application of the proposed framework to image denoising. The numerical results illustrate the benefits of incorporating these new constraints.

2. PROPOSED SET CONSTRUCTION METHOD

2.1. Standing assumptions

The signal space $\mathcal{H}$ is a real Hilbert space and the observation space $\mathcal{G}$ is a Euclidean (finite dimensional real Hilbert) space with scalar product $\langle \cdot \mid \cdot \rangle$. Regarding (1), it is assumed that $L\colon \mathcal{H} \to \mathcal{G}$ is a known bounded linear operator and that $\overline{x}$, u, and y are realizations of some random vectors. Finally, $(e_k)_{1 \le k \le K}$ is an orthonormal basis of $\mathcal{G}$ and ψ_i is a real-valued function defined on $\mathbb{R}$.

2.2. Stein's identity

The proposed approach relies on the following fact.

Lemma 1 [11] *Suppose that A and B are real-valued random variables such that*

i) $\mathsf{E}|A|^2 < +\infty$;

ii) *$B - A$ is a zero-mean Gaussian random variable with variance σ^2;*

iii) *A is independent of $B - A$;*

iv) *ψ_i is continuous, piecewise differentiable, and*

$$(\forall \theta \in \mathbb{R}) \quad \lim_{|v| \to +\infty} \psi_i(v) \exp\Big(-\frac{(v-\theta)^2}{2\sigma^2}\Big) = 0;$$

v) $0 < \mathsf{E}|\psi_i(B)|^2 < +\infty$ *and* $\mathsf{E}|\psi_i'(B)| < +\infty$.

Then $\mathsf{E}(A\psi_i(B)) = \mathsf{E}(B\psi_i(B)) - \sigma^2 \mathsf{E}\psi_i'(B)$.

2.3. Set construction

We can write

$$L\overline{x} = \sum_{k=1}^{K} \overline{\alpha}_k e_k, \; y = \sum_{k=1}^{K} \beta_k e_k, \; \text{and } u = \sum_{k=1}^{K} \gamma_k e_k, \quad (4)$$

where, for every $k \in \{1, \ldots, K\}$,

$$\overline{\alpha}_k = \langle L\overline{x} \mid e_k \rangle, \; \beta_k = \langle y \mid e_k \rangle, \; \text{and } \gamma_k = \langle u \mid e_k \rangle. \quad (5)$$

In view of the assumptions made in Section 2.1, $\overline{\alpha}_k$, β_k, and γ_k are realizations of random variables $\overline{A}_k$, B_k, and C_k, respectively. Under the provision that ψ_i, $\overline{A}_k$, and B_k satisfy the assumptions required in Lemma 1 (in particular C_k is Gaussian with variance σ^2), Stein's identity yields

$$\mathsf{E}\big(\overline{A}_k \psi_i(B_k)\big) = \mathsf{E}\big(B_k \psi_i(B_k)\big) - \sigma^2 \mathsf{E}\psi_i'(B_k). \quad (6)$$

Naturally, this probabilistic identity is not directly enforceable and it must be approximated by a statistical one. Provided that K is large enough and under the technical conditions to be made precise in Assumption 3, the above expectations may be empirically estimated by the consistent statistics

$$\mathsf{E}\big(\overline{A}_k \psi_i(B_k)\big) \approx \frac{1}{K} \sum_{k=1}^{K} \overline{\alpha}_k \psi_i(\beta_k) \quad (7)$$

and

$$\mathsf{E}\big(B_k \psi_i(B_k)\big) - \sigma^2 \mathsf{E}\psi_i'(B_k) \approx \frac{1}{K}\Big(\sum_{k=1}^{K} \beta_k \psi_i(\beta_k) - \sigma^2 \sum_{k=1}^{K} \psi_i'(\beta_k)\Big). \quad (8)$$

Example 2 Let $\psi_i\colon v \mapsto |v|^{p-1}\,\mathrm{sign}(v)$, where $p > 1$ and $\mathrm{sign}(v) = 1$, if $v \ge 0$ and -1 otherwise. Then $\mathsf{E}(B_k \psi_i(B_k)) = \mathsf{E}|B_k|^p$, which shows that ψ_i enforces a constraint close to a p-th order absolute moment constraint. In particular, if $p = 2$, then $\psi_i = \mathrm{Id}$ and we obtain precisely a constraint on the correlation between $\overline{A}_k$ and B_k. Since in this simple case Stein's formula reduces to

$$\mathsf{E}(\overline{A}_k \psi_i(B_k)) = \mathsf{E}|B_k|^2 - \sigma^2 = \mathsf{E}|\overline{A}_k|^2, \quad (9)$$

the resulting constraint is very similar to one on the energy of $L\overline{x}$. One potential advantage of the proposed approach with respect to the classical convex energy constraint is that the latter guarantees only that an upper bound is satisfied, whereas the former provides an equality.

It follows from (6) that the difference of the statistics in (7) and (8) should lie within some confidence interval, say

$$\Big|\sum_{k=1}^{K} \overline{\alpha}_k \psi_i(\beta_k) - \sum_{k=1}^{K} \beta_k \psi_i(\beta_k) + \sigma^2 \sum_{k=1}^{K} \psi_i'(\beta_k)\Big| \le \zeta_i, \quad (10)$$

where ζ_i is derived from the asymptotic distribution of the statistics and some confidence level.

Now let $x \in \mathcal{H}$ be a candidate solution to (1) and let $(\alpha_k)_{1 \le k \le K}$ denote the coefficients of Lx in the basis $(e_k)_{1 \le k \le K}$. Then (10) leads at once to the constraint set

$$S_i = \Big\{x \in \mathcal{H} \;\Big|\; \Big|\sum_{k=1}^{K} \alpha_k \psi_i(\beta_k) - \eta_i\Big| \le \zeta_i\Big\}, \quad (11)$$

where

$$\eta_i = \sum_{k=1}^{K} \beta_k \psi_i(\beta_k) - \sigma^2 \sum_{k=1}^{K} \psi_i'(\beta_k). \quad (12)$$

The problem of determining the bound ζ_i is addressed next.

2.4. Confidence analysis

The families $(\overline{A}_k)_{1 \le k \le K}$, $(B_k)_{1 \le k \le K}$, and $(C_k)_{1 \le k \le K}$ are extracted from random sequences $(\overline{A}_k)_{k \in \mathbb{Z}}$, $(B_k)_{k \in \mathbb{Z}}$, and $(C_k)_{k \in \mathbb{Z}}$ such that

$$(\forall k \in \mathbb{Z}) \quad B_k = \overline{A}_k + C_k. \quad (13)$$

Assumption 3

i) $(\overline{A}_k)_{k\in\mathbb{Z}}$ *is an i.i.d. sequence with finite variance;*

ii) $(C_k)_{k\in\mathbb{Z}}$ *is a zero-mean Gaussian i.i.d. sequence with variance* σ^2*;*

iii) $(\overline{A}_k)_{k\in\mathbb{Z}}$ *and* $(C_k)_{k\in\mathbb{Z}}$ *are independent;*

iv) ψ_i *is continuous, piecewise differentiable, and*

$$(\forall\theta\in\mathbb{R})\quad \lim_{|v|\to+\infty} v\,\psi_i(v)^2\exp\Big(-\frac{(v-\theta)^2}{2\sigma^2}\Big)=0;$$

v) $0<\mathsf{E}|\psi_i(B_0)|^2<+\infty$ *and* $\mathsf{E}|\psi_i'(B_0)|^2<+\infty$.

Theorem 4 [4] *Let* N *be a standard normal random variable, suppose that Assumption 3 is in force, and define*

$$D_K=\sigma^2\sum_{k=1}^{K}\psi_i(B_k)^2+\sigma^4\sum_{k=1}^{K}\psi_i'(B_k)^2$$

and

$$E_K=\sum_{k=1}^{K}\overline{A}_k\psi_i(B_k)-\sum_{k=1}^{K}\big(B_k\psi_i(B_k)-\sigma^2\psi_i'(B_k)\big).$$

Then $D_K/\mathsf{var}E_K\overset{\text{p.s.}}{\to}1$ *and* $E_K/\sqrt{D_K}\overset{\text{d}}{\to}N$*, as* $K\to+\infty$

Thus, for K large enough, the bound ζ_i in (11) can be determined from the tables of the standard normal distribution for a preset confidence level.

We conclude this section by noting that, for the sake of brevity, assumptions i)–iii) above are stated with independence conditions. It is important to emphasize that, at the expense of added technicalities, these conditions can be replaced by much weaker mixing assumptions [4].

2.5. Numerical issues

Let L^* be the adjoint of L and let $a_i=\sum_{k=1}^{K}\psi_i(\beta_k)L^*e_k$. The set S_i of (11) can be rewritten explicitly as an affine hyperslab, namely

$$S_i=\big\{x\in\mathcal{H}\mid\ \eta_i-\zeta_i\leq\langle x\mid a_i\rangle_{\mathcal{H}}\leq\eta_i+\zeta_i\big\}.\qquad(14)$$

Since the projection onto such a set can be computed in closed form [2], the constraint it represents can easily be handled by the convex projection algorithms which are available to solve (2) in the case of feasibility formulations [2], or (3) in the case of quadratic formulations [3].

It is noteworthy that in a given problem several sets of type (11) can be constructed via the proposed technique. Indeed, several functions ψ_i can be considered as well as several bases $(e_k)_{1\leq k\leq K}$. From a statistical standpoint, however, care should be taken to compute the confidence level on each S_i so as to achieve a reasonable confidence level on the feasibility set S in (2) (see the analysis of [2]).

3. DENOISING APPLICATION

The image of Fig. 1 is obtained by adding zero-mean Gaussian white noise to the standard 256×256 8-bit Lena image. The image-to-noise ratio is 11.90 dB. We assume knowledge of the noise variance σ^2 and of the pixel range values. This produces the constraint sets $S_1=[0,255]^K$ (with $K=256^2$) and $S_2=\big\{x\in\mathbb{R}^K\mid\|x-y\|^2\leq\delta_2\big\}$, where δ_2 is determined as in [12]. Additional constraints related to the proposed approach are defined by choosing as a basis $(e_k)_{1\leq k\leq K}$ a set of discrete 2D separable orthonormal wavelets (symlets of length 8). Three functions are involved in this example:

- $\psi_3\colon v\mapsto\tanh(v/a)$,
- $\psi_4\colon v\mapsto v\big(\tanh((v+\chi)/a)-\tanh((v-\chi)/a)\big)$,
- $\psi_5\colon v\mapsto\tanh(v/a)\big(\tanh((v+\chi)/a)-\tanh((v-\chi)/a)\big)$,

where $a\in\mathbb{R}_+^*$ and $\chi\in\mathbb{R}_+^*$. Guidelines for the choice of these functions are discussed in [4]. The wavelet decomposition is realized over 4 resolutions and at each resolution level constraints of the form (11) based on the above functions have been introduced. The denoised image shown in Fig. 3 is obtained by setting

$$J\colon x\mapsto\|x-r\|^2\qquad(15)$$

in (3), where the reference image r is the image produced by the SUREshink thresholding method [6] (see Fig. 2). An inspection of Fig. 3 and of the values of the mean square error (MSE) show that the proposed signal-dependent constraints lead to an improvement of the quality of the recovery. If only S_1 and S_2 are considered, the improvement is much less significant, as illustrated by Fig. 4. Using constraints constructed in different bases leads to further improvements of these results [4].

4. REFERENCES

[1] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis,* 2nd ed. New York: Springer-Verlag, 1985.

[2] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics*, vol. 95, pp. 155-270. New York: Academic, 1996.

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

[4] P. L. Combettes and J.-C. Pesquet, "Stein's principle in set theoretic estimation," technical report, 2003.

[5] 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, July 1991.

[6] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.,* vol. 90, pp. 1200-1224, 1995.

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

[8] H. Krim, D. Tucker, S. Mallat and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Information Theory*, vol. 5, pp. 2225-2238, Nov. 1999.

[9] 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, Jan. 2001.

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

[11] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, pp. 1135-1151, 1981.

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

[13] 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, Dec. 2001.

[14] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. Signal Processing,* vol. 50, pp. 1091–1101, May 2002.

[15] 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, Oct. 1982.

Fig. 1. Noisy image (MSE=904.02).

Fig. 2. SUREshrink denoising (MSE=201.13).

Fig. 3. Proposed approach (MSE=166.68).

Fig. 4. As in Fig. 3 but w/o proposed sets (MSE=192.53).