

Restoration by Compression

Yehuda Dar , Michael Elad , *Fellow, IEEE*, and Alfred M. Bruckstein

Abstract—In this paper, we study the topic of signal restoration using complexity regularization, quantifying the compression bit-cost of the signal estimate. While complexity-regularized restoration is an established concept, solid practical methods were suggested only for the Gaussian denoising task, leaving more complicated restoration problems without a generally constructive approach. Here, we present practical methods for complexity-regularized restoration of signals, accommodating deteriorations caused by a known linear degradation operator of an arbitrary form. Our iterative procedure, obtained using the alternating direction method of multipliers (ADMM) approach, addresses the restoration task as a sequence of simpler problems involving ℓ_2 -regularized estimations and rate-distortion optimizations (considering the squared-error criterion). We replace the rate-distortion optimizations with an arbitrary standardized compression technique and thereby restore the signal by leveraging underlying models designed for compression. Additionally, we propose a shift-invariant complexity regularizer, measuring the bit-cost of all the shifted forms of the estimate, extending our method to use averaging of decompressed outputs gathered from compression of shifted signals. On the theoretical side, we present an analysis of complexity-regularized restoration of a cyclo-stationary Gaussian signal from deterioration by a linear shift-invariant operator and an additive white Gaussian noise. The theory shows that optimal complexity-regularized restoration relies on an elementary restoration filter and compression spreading reconstruction quality unevenly based on the energy distribution of the degradation filter. These ideas are realized also in the proposed practical methods. Finally, we present experiments showing good results for image deblurring and inpainting using the JPEG2000 and HEVC compression standards.

Index Terms—Complexity regularization, rate-distortion optimization, signal restoration, image deblurring, alternating direction method of multipliers (ADMM).

I. INTRODUCTION

SIGNAL restoration methods are often posed as inverse problems using regularization terms. While many solutions can explain a given degraded signal, using regularization will provide signal estimates based on prior assumptions on signals.

Manuscript received November 15, 2017; revised July 21, 2018; accepted September 4, 2018. Date of publication September 20, 2018; date of current version October 8, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yuantao Gu. This work was supported in part by the European Research Council under European Union's Seventh Framework Program, ERC Grant Agreement 320649, and in part by the Israel Science Foundation Grant 2597/16. (*Corresponding author: Yehuda Dar.*)

The authors are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 3200003, Israel (e-mail: ydar@cs.technion.ac.il; elad@cs.technion.ac.il; freddy@cs.technion.ac.il).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>. The material consists of a pdf, which includes the mathematical proof details for Section IV. The file is 184 KB in size.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2018.2871388

One interesting regularization type measures the complexity of the candidate solution in terms of its compression bit-cost. Indeed, encoders (that yield the bit cost) rely on signal models and allocate shorter representations to more likely signal instances. This approach of complexity-regularized restoration is an attractive meeting point of signal restoration and compression, two fundamental signal-processing problems.

Numerous works [1]–[7] considered the task of denoising a signal corrupted by an additive white Gaussian noise using complexity regularization. In [2], [7], this idea is translated to practically estimating the clean signal by employing a standard lossy compression of its noisy version. However, more complex restoration problems (e.g., deblurring, super resolution, inpainting), involving non-trivial degradation operators, do not lend themselves to a straightforward treatment by compression techniques designed for the squared-error distortion measure. Moulin and Liu [8] studied the complexity regularization idea for general restoration problems, presenting a thorough theoretical treatment together with a limited practical demonstration of Poisson denoising based on a suitably designed compression method. Indeed, a general method for complexity-regularized restoration remained as an open question for a long while until our recent preliminary publication [9], where we presented a generic and practical approach flexible in both the degradation model addressed and the compression technique utilized.

Our strategy for complexity-regularized signal restoration relies on the alternating direction method of multipliers (ADMM) approach [10], decomposing the difficult optimization problem into a sequence of easier tasks including ℓ_2 -regularized inverse problems and standard rate-distortion optimizations (with respect to a squared-error distortion metric). A main part of our methodology is to replace the rate-distortion optimization with standardized compression techniques enabling an indirect utilization of signal models used for efficient compression designs. Moreover, our method relates to various contemporary concepts in signal and image processing. The recent frameworks of Plug-and-Play Priors [11], [12] and Regularization-by-Denoising [13] suggest leveraging a Gaussian denoiser for more complicated restoration tasks, achieving impressive results (see, e.g., [11]–[16]). Essentially, our approach is the compression-based counterpart for denoising-based restoration concepts from [11]–[13].

Commonly, compression methods process the given signal based on its decomposition into non-overlapping blocks, yielding block-level rate-distortion optimizations based on block bit-costs. The corresponding complexity measure sums the bit-costs of all the non-overlapping blocks, however, note that this evaluation is shift sensitive. This fact motivates us to propose a shift-invariant complexity regularizer by quantifying the bit-costs of all the overlapping blocks of the signal estimate. This

improved regularizer calls for our restoration procedure to use averaging of decompressed signals obtained from compressions of shifted signals. Our shift-invariant approach conforms with the Expected Patch Log-Likelihood (EPLL) idea [17], where a full-signal regularizer is formed based on a block-level prior in a way leading to averaging MAP estimates of shifted signal versions. Our extended method also recalls the cycle spinning concept, presented in [18] for wavelet-based denoising. Additional resemblance is to the compression postprocessing techniques in [19], [20] enhancing a given decompressed image by averaging supplementary compression-decompression results of shifted versions of the given image, thus, our method generalizes this approach to any restoration problem with an appropriate consideration of the degradation operator. Very recent works [21], [22] suggested the use of compression techniques for compressive sensing of signals and images, but our approach examines other perspectives and settings referring to restoration problems as will be explained below.

In this paper we extend our previous conference publication [9] with improved algorithms and new theoretical and experimental results. In [9] we implemented our concepts in procedures relying on the half quadratic splitting optimization technique, in contrast, here we present improved algorithms designed based on the ADMM approach. The new ADMM-based methods introduce the following benefits (with respect to using half quadratic splitting as in [9]): significant gains in the restoration quality, reduction in the required amount of iterations, and an easier parameter setting. In addition, in this paper we provide an extensive experimental section. While in [9] we experimentally examined only the inpainting problem, in this paper we present new results demonstrating the practical complexity-regularized restoration approach for image deblurring. While deblurring is a challenging restoration task, we present compelling results obtained using the JPEG2000 method and the image compression profile of the high efficiency video coding (HEVC) standard [23]. An objective comparison to other deblurring techniques showed that the proposed HEVC-based implementation provides good deblurring results. Moreover, we also extend our evaluation given in [9] for image inpainting, where here we use the JPEG2000 and HEVC compression standards in our ADMM-based approach to restore images from a severe degradation of 80% missing pixels. Interestingly, our compression-based image inpainting approach can be perceived as the dual concept of inpainting-based compression of images and videos suggested in, e.g., [24]–[26] and discussed also in [27].

Another prominent contribution of this paper is the new theoretical study of the problem of complexity-regularized restoration, considering the estimation of a cyclo-stationary Gaussian signal from a degradation procedure consisting of a linear shift-invariant operator and additive white Gaussian noise. We gradually establish few equivalent optimization forms, emphasizing two main concepts for complexity-regularized restoration: the degraded signal should go through a simple inverse filtering procedure, and then should be compressed so that the decomposition components will have a varying quality distribution determined by the degradation-filter energy-distribution. We ex-

plain how these ideas materialize in the practical approach we propose, thus, establishing a theoretical reasoning for the feasible complexity-regularized restoration.

This paper is organized as follows. In Section II-C we overview the settings of the complexity-regularized restoration problem. In Section III we present the proposed practical methods for complexity-regularized restoration. In Section IV we theoretically analyze particular problem settings where the signal is a cyclo-stationary Gaussian process. In Section V we provide experimental results for image deblurring and inpainting. Section VI concludes this paper.

II. COMPLEXITY-REGULARIZED RESTORATION: PROBLEM SETTINGS

A. Regularized Restoration of Signals

In this paper we address the task of restoring a signal $\mathbf{x}_0 \in \mathbb{R}^N$ from a degraded version, $\mathbf{y} \in \mathbb{R}^M$, obeying the prevalent deterioration model:

$$\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{n} \quad (1)$$

where \mathbf{H} is a $M \times N$ matrix being a linear degradation operator (e.g., blur, pixel omission, decimation) and $\mathbf{n} \in \mathbb{R}^M$ is a white Gaussian noise vector having zero mean and variance σ_n^2 .

Maximum A-Posteriori (MAP) estimation is a widely-known statistical approach forming the restored signal, $\hat{\mathbf{x}}$, via

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y}) \quad (2)$$

where $p(\mathbf{x}|\mathbf{y})$ is the posterior probability. For the above defined degradation model (1), incorporating additive white Gaussian noise, the MAP estimate reduces to the form of

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2\sigma_n^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 - \log p(\mathbf{x}) \quad (3)$$

where $p(\mathbf{x})$ is the prior probability that, here, evaluates the probability of the candidate solution.

Another prevalent restoration approach, embodied in many contemporary techniques, forms the estimate via the optimization

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu s(\mathbf{x}) \quad (4)$$

where $s(\mathbf{x})$ is a general regularization function returning a lower value for a more likely candidate solution, and $\mu \geq 0$ is a parameter weighting the regularization effect. This strategy for restoration based on arbitrary regularizers can be interpreted as a generalization of the MAP approach in (3). Specifically, comparing the formulations (4) and (3) exhibits the regularization function $s(\mathbf{x})$ and the parameter μ as extensions of $(-\log p(\mathbf{x}))$ and the factor $2\sigma_n^2$, respectively.

Among the various regularization functions that can be associated with the general restoration approach in (4), we explore here the class of complexity regularizers measuring the required number of bits for the compressed representation of the candidate solution. The practical methods presented in this section focus on utilizing existing (independent) compression techniques, implicitly employing their underlying signal models for the restoration task.

B. Operational Rate-Distortion Optimization

The practical complexity-regularized restoration methods in this section are developed with respect to a compression technique obeying the following conceptual design. The signal is segmented to equally-sized non-overlapping blocks (each is consisted of N_b samples) that are independently compressed. The block compression procedure is modeled as a general variable-rate vector quantizer relying on the following mappings. The compression is done by the mapping $Q: \mathbb{R}^{N_b} \rightarrow \mathcal{W}$ from the N_b -dimensional signal-block domain to a discrete set \mathcal{W} of binary compressed representations (that may have different lengths). The decompression procedure is associated with the mapping $F: \mathcal{W} \rightarrow \mathcal{C}$, where $\mathcal{C} \subset \mathbb{R}^{N_b}$ is a finite discrete set (a codebook) of block reconstruction candidates. For example, consider the block $\mathbf{x}_{block} \in \mathbb{R}^{N_b}$ that its binary compressed representation in \mathcal{W} is given via $b = Q(\mathbf{x}_{block})$ and the corresponding reconstructed block in \mathcal{C} is $\hat{\mathbf{x}}_{block} = F(b)$. Importantly, it is assumed that shorter codewords are coupled with block reconstructions that are, in general, more likely.

The signal \mathbf{x} is compressed based on its segmentation into a set of blocks $\{\mathbf{x}_i\}_{i \in \mathcal{B}}$ (where \mathcal{B} denotes the index set of blocks in the non-overlapping partitioning of the signal). In addition we introduce the function $r(\mathbf{z})$ that evaluates the bit-cost (i.e., the length of the binary codeword) for the block reconstruction $\mathbf{z} \in \mathcal{C}$. Then, the operational rate-distortion optimization corresponding to the described architecture and a squared-error distortion metric is

$$\{\tilde{\mathbf{x}}_i\}_{i \in \mathcal{B}} = \underset{\{\mathbf{v}_i\}_{i \in \mathcal{B}} \in \mathcal{C}}{\operatorname{argmin}} \sum_{i \in \mathcal{B}} \|\mathbf{x}_i - \mathbf{v}_i\|_2^2 + \lambda \sum_{i \in \mathcal{B}} r(\mathbf{v}_i), \quad (5)$$

where $\lambda \geq 0$ is a Lagrange multiplier corresponding to some total compression bit-cost. Importantly, the independent representation of non-overlapping blocks allows solving (5) separately for each block [28], [29].

Our mathematical developments require the following algebraic tools for block handling. The matrix \mathbf{P}_i is defined to provide the i th block from the complete signal via the standard multiplication $\mathbf{P}_i \mathbf{x} = \mathbf{x}_i$. Note that \mathbf{P}_i can extract any block of the signal, even one that is not in the non-overlapping grid \mathcal{B} . Accordingly, the matrix \mathbf{P}_i^T locates a block in the i th block-position in a construction of a full-sized signal and, therefore, lets to express the a complete signal as $\mathbf{x} = \sum_{i \in \mathcal{B}} \mathbf{P}_i^T \mathbf{x}_i$.

Now we can use the block handling operator \mathbf{P}_i for expressing the block-based rate-distortion optimization in its corresponding full-signal formulation:

$$\tilde{\mathbf{x}} = \underset{\mathbf{v} \in \mathcal{C}_B}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda r_{tot}(\mathbf{v}). \quad (6)$$

where \mathcal{C}_B is the full-signal codebook, being the discrete set of candidate reconstructions for the full signal, defined using the block-level codebook \mathcal{C} as

$$\mathcal{C}_B = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{i \in \mathcal{B}} \mathbf{P}_i^T \mathbf{v}_i, \{\mathbf{v}_i\}_{i \in \mathcal{B}} \in \mathcal{C} \right\}. \quad (7)$$

Moreover, the regularization function in (6) is the total bit cost of the reconstructed signal defined for $\mathbf{v} \in \mathcal{C}_B$ as $r_{tot}(\mathbf{v}) \triangleq \sum_{i \in \mathcal{B}} r(\mathbf{P}_i \mathbf{v})$.

C. Complexity-Regularized Restoration: Basic Optimization Formulation

While the regularized-restoration optimization in (4) is over a continuous domain, the operational rate-distortion optimization in (6) is a discrete problem with solutions limited to the set \mathcal{C}_B . Therefore, we extend the definition of the block bit-cost evaluation function such that it is defined for any $\mathbf{z} \in \mathbb{R}^{N_b}$ via

$$\bar{r}(\mathbf{z}) = \begin{cases} r(\mathbf{z}), & \mathbf{z} \in \mathcal{C} \\ \infty, & \mathbf{z} \notin \mathcal{C} \end{cases}, \quad (8)$$

and the corresponding extension of the total bit-cost $\bar{r}_{tot}(\mathbf{x}) \triangleq \sum_{i \in \mathcal{B}} \bar{r}(\mathbf{P}_i \mathbf{x})$ is defined for any $\mathbf{x} \in \mathbb{R}^N$.

Now we define the complexity regularization function as

$$s(\mathbf{x}) = \bar{r}_{tot}(\mathbf{x}) \quad (9)$$

and the corresponding restoration optimization is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \bar{r}_{tot}(\mathbf{x}). \quad (10)$$

Due to the definition of the extended bit-cost evaluation function, $\bar{r}_{tot}(\mathbf{x})$, the solution candidates of (10) are limited to the discrete set \mathcal{C}_B as defined in (7).

Examining the complexity-regularized restoration in (10) for the Gaussian denoising task, where $\mathbf{H} = \mathbf{I}$, shows that the optimization reduces to the regular rate-distortion optimization in (6), namely, the compression of the noisy signal \mathbf{y} . However, for more complicated restoration problems, where \mathbf{H} has an arbitrary structure, the optimization in (10) is not easy to solve and, in particular, it does not correspond to standard compression designs that are optimized for the regular squared-error distortion metric.

III. PROPOSED METHODS

In this section we present three restoration methods leveraging a given compression technique. The proposed algorithms result from two different definitions for the complexity regularization function. While the first approach regularizes the total bit-cost of the non-overlapping blocks of the restored signal, the other two refer to the total bit-cost of all the overlapping blocks of the estimate.

A. Regularize Total Complexity of Non-Overlapping Blocks

Here we establish a practical method addressing the optimization problem in (10) based on the alternating direction method of multipliers (ADMM) approach [10] (for additional uses see, e.g., [11], [12], [14], [16], [30]). The optimization (10) can be expressed also as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \sum_{i \in \mathcal{B}} \bar{r}(\mathbf{P}_i \mathbf{x}), \quad (11)$$

where the degradation matrix \mathbf{H} , having a general structure, renders a block-based treatment infeasible.

Addressing this structural difficulty using the ADMM strategy [10] begins with introducing the auxiliary variables $\{\mathbf{z}_i\}_{i \in \mathcal{B}}$, where \mathbf{z}_i is coupled with the i th non-overlapping block.

Specifically, we reformulate the problem (11) into

$$\begin{aligned} (\hat{\mathbf{x}}, \{\hat{\mathbf{z}}_i\}_{i \in \mathcal{B}}) &= \underset{\mathbf{x}, \{\mathbf{z}_i\}_{i \in \mathcal{B}}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \sum_{i \in \mathcal{B}} \bar{r}(\mathbf{z}_i) \\ \text{s.t. } \mathbf{z}_i &= \mathbf{P}_i \mathbf{x}, \text{ for } i \in \mathcal{B}. \end{aligned} \quad (12)$$

Then, reformulating the constrained optimization (12) using the augmented Lagrangian (in its scaled form) and the method of multipliers (see [10, Ch. 2]) leads to the following iterative procedure

$$\begin{aligned} (\hat{\mathbf{x}}^{(t)}, \{\hat{\mathbf{z}}_i^{(t)}\}_{i \in \mathcal{B}}) &= \underset{\mathbf{x}, \{\mathbf{z}_i\}_{i \in \mathcal{B}}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \sum_{i \in \mathcal{B}} \bar{r}(\mathbf{z}_i) \\ &\quad + \frac{\beta}{2} \sum_{i \in \mathcal{B}} \left\| \mathbf{P}_i \mathbf{x} - \mathbf{z}_i + \mathbf{u}_i^{(t)} \right\|_2^2 \end{aligned} \quad (13)$$

$$\mathbf{u}_i^{(t+1)} = \mathbf{u}_i^{(t)} + \left(\mathbf{P}_i \hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}_i^{(t)} \right) \quad i \in \mathcal{B}, \quad (14)$$

where t is the iteration number, β is a parameter originating in the augmented Lagrangian, and $\mathbf{u}_i^{(t)} \in \mathbb{R}^{N_b}$ is the scaled dual variable corresponding to the i th block (where $i \in \mathcal{B}$).

Each of the optimization variables in (13) participates only in part of the terms of the cost function and, therefore, employing one iteration of alternating minimization (see [10, Ch. 2]) leads to the ADMM form of the problem, where the included optimizations are relatively simple. Accordingly, the t th iteration of the proposed iterative solution is

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\beta}{2} \sum_{i \in \mathcal{B}} \left\| \mathbf{P}_i \mathbf{x} - \tilde{\mathbf{z}}_i^{(t)} \right\|_2^2 \quad (15)$$

$$\hat{\mathbf{z}}_i^{(t)} = \underset{\mathbf{z}_i}{\operatorname{argmin}} \frac{\beta}{2} \left\| \tilde{\mathbf{x}}_i^{(t)} - \mathbf{z}_i \right\|_2^2 + \mu \bar{r}(\mathbf{z}_i), \quad i \in \mathcal{B} \quad (16)$$

$$\mathbf{u}_i^{(t+1)} = \mathbf{u}_i^{(t)} + \left(\mathbf{P}_i \hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}_i^{(t)} \right), \quad i \in \mathcal{B}, \quad (17)$$

where $\tilde{\mathbf{z}}_i^{(t)} \triangleq \hat{\mathbf{z}}_i^{(t-1)} - \mathbf{u}_i^{(t)}$ and $\tilde{\mathbf{x}}_i^{(t)} \triangleq \mathbf{P}_i \hat{\mathbf{x}}^{(t)} + \mathbf{u}_i^{(t)}$ for $i \in \mathcal{B}$.

The analytic solution of the first stage optimization in (15) is

$$\hat{\mathbf{x}}^{(t)} = \left(\mathbf{H}^T \mathbf{H} + \frac{\beta}{2} \mathbf{I} \right)^{-1} \left(\mathbf{H}^T \mathbf{y} + \frac{\beta}{2} \sum_{i \in \mathcal{B}} \mathbf{P}_i^T \tilde{\mathbf{z}}_i^{(t)} \right) \quad (18)$$

rendering this stage as a weighted averaging of the deteriorated signal with the block estimates obtained in the second stage of the previous iteration. While the analytic solution (18) explains the underlying meaning of the ℓ_2 -constrained deconvolution stage (15), it includes matrix inversion that, in general, may lead to numerical instabilities. Accordingly, in the implementation of the proposed method we suggest to address (15) via numerical optimization techniques (for example, we used the biconjugate gradients method).

The optimizations in the second stage of each iteration (16) are rate-distortion optimizations corresponding to each of the non-overlapping blocks of the signal estimate $\hat{\mathbf{x}}^{(t)}$ obtained in the first stage. Accordingly, the set of block-level optimizations in (16) can be interpreted as a single full-signal rate-distortion optimization with respect to a Lagrange multiplier value of $\lambda = \frac{2\mu}{\beta}$. We denote the compression-decompression procedure

Algorithm 1: Proposed Method Based on Total Complexity of Non-Overlapping Blocks.

- 1: Inputs: $\mathbf{y}, \beta, \theta$.
 - 2: Initialize $\hat{\mathbf{z}}^{(0)}$ (depending on the deterioration type).
 - 3: $t = 1$ and $\mathbf{u}^{(1)} = \mathbf{0}$
 - 4: **repeat**
 - 5: $\tilde{\mathbf{z}}^{(t)} = \hat{\mathbf{z}}^{(t-1)} - \mathbf{u}^{(t)}$
 - 6: Solve the ℓ_2 -constrained deconvolution:

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\beta}{2} \|\mathbf{x} - \tilde{\mathbf{z}}^{(t)}\|_2^2$$
 - 7: $\tilde{\mathbf{x}}^{(t)} = \hat{\mathbf{x}}^{(t)} + \mathbf{u}^{(t)}$
 - 8: $\hat{\mathbf{z}}^{(t)} = \operatorname{CompressDecompress}_\theta(\tilde{\mathbf{x}}^{(t)})$
 - 9: $\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + (\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}^{(t)})$
 - 10: $t \leftarrow t + 1$
 - 11: **until** stopping criterion is satisfied
-

that replaces (16) as

$$\hat{\mathbf{z}}^{(t)} = \operatorname{CompressDecompress}_\lambda(\tilde{\mathbf{x}}^{(t)}), \quad (19)$$

where $\tilde{\mathbf{x}}^{(t)} \triangleq \sum_{i \in \mathcal{B}} \mathbf{P}_i^T \tilde{\mathbf{x}}_i^{(t)}$ is the signal to compress, assembled from all the non-overlapping blocks, and $\hat{\mathbf{z}}^{(t)}$ is the corresponding decompressed full signal. Moreover, by defining a full-sized scaled dual variable $\mathbf{u}^{(t)} \triangleq \sum_{i \in \mathcal{B}} \mathbf{P}_i^T \mathbf{u}_i^{(t)}$ we get that $\tilde{\mathbf{x}}^{(t)} = \hat{\mathbf{x}}^{(t)} + \mathbf{u}^{(t)}$. Then, using the definitions established here we can translate the block-level computations (15)–(17) into the full-signal formulations described in Algorithm 1.

We further suggest using a standardized compression method as the compression-decompression operator (19). While many compression methods do not follow the exact rate-distortion optimizations we have in our mathematical development, we still encourage utilizing such techniques as an approximation for (16). Additionally, since many compression methods do not rely on Lagrangian optimization, their operating parameters may have different definitions such as quality parameters, compression ratios, or output bit-rates. Accordingly, we present the suggested algorithm with respect to a general compression-decompression procedure with output bit-cost directly or indirectly affected by a parameter denoted as θ . These generalizations are also implemented in the proposed Algorithm 1. In Section V we elaborate on particular settings of θ that were empirically found appropriate for utilization of the HEVC and the JPEG2000 standard. In cases where the compression method significantly deviates from a Lagrangian optimization form, it can be useful to appropriately update the compression parameter in each iteration (this is the case for JPEG2000 as explained in Section V).

Importantly, Algorithm 1 does not only restore the deteriorated input image, but also provides the signal estimate in a compressed form by employing the output of the compression stage of the last iteration.

B. Regularize Total Complexity of All Overlapping Blocks

Algorithm 1 emerged from complexity regularization measuring the total bit-cost of the estimate based on its decomposition into non-overlapping blocks (see Eq. (11)), resulting

in a restored signal available in a compressed form compatible with the compression technique in use. Obviously, the above approach provides estimates limited to the discrete set of signals supported by the compression architecture, thus, having a somewhat reduced restoration ability with respect to methods providing estimates from an unrestricted domain of solutions. This observation motivates us to develop a complexity-regularized restoration procedure that provides good estimates from the continuous unrestricted domain of signals while still utilizing a standardized compression technique as its main component.

As before, our developments refer to a general block-based compression method relying on a codebook \mathcal{C} as a discrete set of block reconstruction candidates. We consider here the segmentation of the signal-block space, \mathbb{R}^{N_b} , given by the voronoi cells corresponding to the compression reconstruction candidates, namely, for each $\mathbf{c} \in \mathcal{C}$ there is a region

$$V_{\mathbf{c}} \triangleq \left\{ \mathbf{w} \in \mathbb{R}^{N_b} \mid \mathbf{c} = \arg \min_{\mathbf{c} \in \mathcal{C}} \|\mathbf{w} - \mathbf{c}\|_2^2 \right\} \quad (20)$$

defining all the vectors in \mathbb{R}^{N_b} that \mathbf{c} is their nearest member of \mathcal{C} . We use the voronoi cells in (20) for defining an alternative extension to the bit-cost evaluation of a signal block (i.e., the new definition, $\bar{r}_v(\mathbf{z})$, will replace $\bar{r}(\mathbf{z})$ given in (8) that was used for the development of Algorithm 1). Specifically, we associate a finite bit-cost to any $\mathbf{z} \in \mathbb{R}^{N_b}$ based on the voronoi cell it resides in, i.e.,

$$\bar{r}_v(\mathbf{z}) = r(\mathbf{c}) \text{ for } \mathbf{z} \in V_{\mathbf{c}} \quad (21)$$

where $r(\mathbf{c})$ is the regular bit-cost evaluation defined in Section II-B only for blocks in \mathcal{C} .

The method proposed here emerges from a new complexity regularization function that quantifies the total complexity of all the overlapping blocks of the estimate. Using the extended bit-cost measure $\bar{r}_v(\cdot)$, defined in (21), we introduce the full-signal regularizer as

$$s^*(\mathbf{x}) = \sum_{i \in \mathcal{B}^*} \bar{r}_v(\mathbf{P}_i \mathbf{x}) \quad (22)$$

where $\mathbf{x} \in \mathbb{R}^N$, and \mathcal{B}^* is a set containing the indices of all the overlapping blocks of the signal. The associated restoration optimization is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \sum_{i \in \mathcal{B}^*} \bar{r}_v(\mathbf{P}_i \mathbf{x}). \quad (23)$$

Importantly, in contrast to the previous subsection, the function $s^*(\mathbf{x})$ evaluates the complexity of any $\mathbf{x} \in \mathbb{R}^N$ with a finite value and, thus, does not restrict the restoration to the discrete set of codebook-based constructions, $\mathcal{C}_{\mathcal{B}}$, defined in (7).

While the new regularizer in (23) is not separable into complexity evaluation of non-overlapping blocks, the ADMM approach can accommodate it as well. This is explained next. We define the auxiliary variables $\{\mathbf{z}_i\}_{i \in \mathcal{B}^*}$, where each \mathbf{z}_i is coupled with the i th overlapping block. Then, the optimization (23) is expressed as

$$\begin{aligned} (\hat{\mathbf{x}}, \{\hat{\mathbf{z}}_i\}_{i \in \mathcal{B}^*}) &= \underset{\mathbf{x}, \{\mathbf{z}_i\}_{i \in \mathcal{B}^*}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \sum_{i \in \mathcal{B}^*} \bar{r}_v(\mathbf{z}_i) \\ \text{s.t. } \mathbf{z}_i &= \mathbf{P}_i \mathbf{x} \text{ for } i \in \mathcal{B}^*. \end{aligned} \quad (24)$$

As in Section III-A, employing the augmented Lagrangian (in its scaled form) and the method of multipliers results in an iterative solution provided by the following three steps in each iteration (as before t denotes the iteration number):

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\beta}{2} \sum_{i \in \mathcal{B}^*} \left\| \mathbf{P}_i \mathbf{x} - \hat{\mathbf{z}}_i^{(t)} \right\|_2^2 \quad (25)$$

$$\hat{\mathbf{z}}_i^{(t)} = \underset{\mathbf{z}_i}{\operatorname{argmin}} \frac{\beta}{2} \left\| \hat{\mathbf{x}}_i^{(t)} - \mathbf{z}_i \right\|_2^2 + \mu \bar{r}_v(\mathbf{z}_i), \quad i \in \mathcal{B}^* \quad (26)$$

$$\mathbf{u}_i^{(t+1)} = \mathbf{u}_i^{(t)} + \left(\mathbf{P}_i \hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}_i^{(t)} \right), \quad i \in \mathcal{B}^*, \quad (27)$$

where $\mathbf{u}_i^{(t)}$ is the scaled dual variable for the i th block, $\hat{\mathbf{z}}_i^{(t)} \triangleq \hat{\mathbf{z}}_i^{(t-1)} - \mathbf{u}_i^{(t)}$ and $\hat{\mathbf{x}}_i^{(t)} \triangleq \mathbf{P}_i \hat{\mathbf{x}}^{(t)} + \mathbf{u}_i^{(t)}$ for $i \in \mathcal{B}^*$.

While the procedure above resembles the one from the former subsection, the treatment of overlapping blocks has different interpretations to the optimizations in (25) and (26). Indeed, note that the block-level rate-distortion optimizations in (26) are not discrete due to the extended bit-cost evaluation $\bar{r}_v(\cdot)$ defined in (21). Due to the definition of $\bar{r}_v(\cdot)$, the rate-distortion optimizations (26) can be considered as continuous relaxations of the discrete optimizations done by the practical compression technique. Since we intend using a given compression method without explicit knowledge of its underlying codebook, we cannot construct the voronoi cells defining $\bar{r}_v(\cdot)$ and, thus, it is impractical to accurately solve (26). Consequently, we suggest to approximate the optimizations (26) by the discrete forms of

$$\hat{\mathbf{z}}_i^{(t)} = \underset{\mathbf{z}_i}{\operatorname{argmin}} \frac{\beta}{2} \left\| \mathbf{P}_i \hat{\mathbf{x}}^{(t)} - \mathbf{z}_i \right\|_2^2 + \mu \bar{r}(\mathbf{z}_i), \quad i \in \mathcal{B}^* \quad (28)$$

where $\bar{r}(\cdot)$ is the discrete evaluation of the block bit-cost, defined in (8), letting to identify the problems as operational rate-distortion optimizations of the regular discrete form.

Each block-level rate-distortion optimization in (28) is associated with one of the overlapping blocks of the signal. Accordingly, we interpret this group of optimizations as multiple applications of a full-signal compression-decompression procedure, each associates to a shifted version of the signal (corresponding to different sets of non-overlapping blocks). Specifically, for a signal \mathbf{x} and a compression block-size of N_b samples, there are N_b shifted grids of non-overlapping blocks. For mathematical convenience, we consider here cyclic shifts such that the j th shift ($j = 1, \dots, N_b$) corresponds to a signal of N samples taken cyclically starting at the j th sample of \mathbf{x} (in practice other definitions of shifts may be used, e.g., see Section V for a suggested treatment of two-dimensional signals). We denote the j th shifted signal as $\mathit{shift}_j\{\mathbf{x}\}$. Moreover, we denote the index set of blocks included in the j th shifted signal as \mathcal{B}^j (noting that $\mathcal{B}^1 = \mathcal{B}$), hence, $\mathcal{B}^* = \cup_{j=1}^{N_b} \mathcal{B}^j$. Therefore, the decompressed blocks $\{\hat{\mathbf{z}}_i^{(t)}\}_{i \in \mathcal{B}^*}$ can be decomposed into N_b subsets, $\{\hat{\mathbf{z}}_i^{(t)}\}_{i \in \mathcal{B}^j}$ for $j = 1, \dots, N_b$, each contains non-overlapping blocks corresponding to a different shifted grid. Moreover, the j th set of blocks, $\{\hat{\mathbf{z}}_i^{(t)}\}_{i \in \mathcal{B}^j}$, is associated with the full signal $\hat{\mathbf{z}}^{j,(t)} \triangleq \sum_{i \in \mathcal{B}^j} \mathbf{P}_i^T \hat{\mathbf{z}}_i^{(t)}$. Then, the set of full signals $\{\hat{\mathbf{z}}^{j,(t)}\}_{j=1}^{N_b}$ can be obtained by multiple full-signal compression-decompression applications, namely, for

$j = 1, \dots, N_b$: $\hat{\mathbf{z}}_{shifted}^{j,(t)} = CompressDecompress_{\lambda}(\tilde{\mathbf{x}}_{shifted}^{j,(t)})$, where the Lagrangian multiplier value is $\lambda = \frac{2\mu}{\beta}$ and

$$\tilde{\mathbf{x}}_{shifted}^{j,(t)} \triangleq shift_j \left\{ \hat{\mathbf{x}}^{(t)} + \mathbf{u}^{j,(t)} \right\} \quad (29)$$

is the compression input formed as the j th shift of $\hat{\mathbf{x}}^{(t)}$ combined with the full-sized dual variable defined via

$$\mathbf{u}^{j,(t)} \triangleq \sum_{i \in \mathcal{B}^j} \mathbf{P}_i^T \mathbf{u}_i^{(t)} \quad (30)$$

assembled from the block-level dual variables corresponding to the j th grid of non-overlapping blocks. Notice that inverse shifts are required for obtaining the desired signals, i.e.,

$$\hat{\mathbf{z}}^{j,(t)} = shift_j^{-1} \left\{ \hat{\mathbf{z}}_{shifted}^{j,(t)} \right\} \quad (31)$$

where $shift_j^{-1} \{ \cdot \}$ is the inverse shift operator that (cyclically) shifts back the given full-size signal by j samples.

The deconvolution stage (25) of the iterative process can be rewritten as

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\| \mathbf{H}\mathbf{x} - \mathbf{y} \right\|_2^2 + \frac{\beta}{2} \sum_{j=1}^{N_b} \left\| \mathbf{x} - \tilde{\mathbf{z}}^{j,(t)} \right\|_2^2 \quad (32)$$

where the regularization part (the second term) considers the distance of the estimate from the N_b full signals defined via

$$\tilde{\mathbf{z}}^{j,(t)} \triangleq \hat{\mathbf{z}}^{j,(t)} - \mathbf{u}^{j,(t)} \quad (33)$$

for $j = 1, \dots, N_b$, where $\hat{\mathbf{z}}^{j,(t)}$ and $\mathbf{u}^{j,(t)}$ were defined above. The analytic solution of the optimization (32) is

$$\hat{\mathbf{x}}^{(t)} = \left(\mathbf{H}^T \mathbf{H} + \frac{\beta}{2} N_b \right)^{-1} \left(\mathbf{H}^T \mathbf{y} + \frac{\beta}{2} \sum_{j=1}^{N_b} \tilde{\mathbf{z}}^{j,(t)} \right), \quad (34)$$

showing that the first stage of each iteration is a weighted averaging of the given deteriorated signal with all the decompressed signals (and the dual variables) obtained in the former iteration. It should be noted that the analytic solution (34) is developed here for showing the essence of the ℓ_2 -constrained deconvolution part of the method. Nevertheless, the possible numerical instabilities due to the matrix inversion appearing in (34) motivate the practical direct treatment of (25) via numerical optimization techniques.

Algorithm 2 summarizes the practical restoration method for a compression technique operated by the general parameter θ for determining the bit-cost (see details in Section III-A).

The computational cost of Algorithm 2 stems from its reliance on repeated applications of compressions, decompressions, and ℓ_2 -constrained deconvolution procedures. While the actual run-time depends on the computational complexity of the utilized compression technique, we can generally state that the total run-time will be of at least the run-time of compression and decompression processes for a total number of applications equal to the product of the number of iterations and the number of shifts considered.

The ADMM is known for promoting distributed optimization structures [10]. In Algorithm 2 the distributed nature of the ADMM is expressed in the separate optimization of each of the shifted block-grids (see stages 8–11). In particular, the

Algorithm 2: Proposed Method Based on Total Complexity of All the Overlapping Blocks.

- 1: Inputs: $\mathbf{y}, \beta, \theta$.
 - 2: Initialize $\{ \hat{\mathbf{z}}^{j,(0)} \}_{j=1}^{N_b}$ (depending on the deterioration type).
 - 3: $t = 1$ and $\mathbf{u}^{j,(1)} = \mathbf{0}$ for $j = 1, \dots, N_b$.
 - 4: **repeat**
 - 5: $\tilde{\mathbf{z}}^{j,(t)} = \hat{\mathbf{z}}^{j,(t-1)} - \mathbf{u}^{j,(t)}$ for $j = 1, \dots, N_b$
 - 6: Solve the ℓ_2 -constrained deconvolution:

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\| \mathbf{H}\mathbf{x} - \mathbf{y} \right\|_2^2 + \frac{\beta}{2} \sum_{j=1}^{N_b} \left\| \mathbf{x} - \tilde{\mathbf{z}}^{j,(t)} \right\|_2^2$$
 - 7: **for** $j = 1, \dots, N_b$ **do**
 - 8: $\tilde{\mathbf{x}}_{shifted}^{j,(t)} = shift_j \left\{ \hat{\mathbf{x}}^{(t)} + \mathbf{u}^{j,(t)} \right\}$
 - 9: $\hat{\mathbf{z}}_{shifted}^{j,(t)} = CompressDecompress_{\theta} \left(\tilde{\mathbf{x}}_{shifted}^{j,(t)} \right)$
 - 10: $\hat{\mathbf{z}}^{j,(t)} = shift_j^{-1} \left\{ \hat{\mathbf{z}}_{shifted}^{j,(t)} \right\}$
 - 11: $\mathbf{u}^{j,(t+1)} = \mathbf{u}^{j,(t)} + \left(\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}^{j,(t)} \right)$
 - 12: **end for**
 - 13: $t \leftarrow t + 1$
 - 14: **until** stopping criterion is satisfied
-

dual variables $\{ \mathbf{u}^{j,(t)} \}_{j=1}^{N_b}$, associated with the various grids (see stages 5, 8, and 11 in Algorithm 2), are updated independently in stage 11 such that each considers only its respective $\hat{\mathbf{z}}^{j,(t)}$. However, the dual variables $\{ \mathbf{u}^{j,(t)} \}_{j=1}^{N_b}$ essentially refer to the same data based on different block-grids. Accordingly, we suggest to merge the independent dual variables to form a single, more robust, dual variable defined as

$$\mathbf{u}_{total}^{(t)} = \frac{1}{N_b} \sum_{j=1}^{N_b} \mathbf{u}^{j,(t)} \quad (35)$$

where the averaging tends to reduce particular artifacts that may appear due to specific block-grids. We utilize the averaged dual variable (35) to extend Algorithm 2 into Algorithm 3. Notice stages 5, 8, and 13 of Algorithm 3, where the averaged dual variable is used instead of the independent ones.

In Section V we further discuss practical aspects of the proposed Algorithms 1–3 and evaluate their performance for deblurring and inpainting of images.

IV. RATE-DISTORTION THEORETIC ANALYSIS FOR THE GAUSSIAN CASE

In this section we theoretically study the complexity-regularized restoration problem from the perspective of rate-distortion theory. While our analysis is focused on the particular settings of a cyclo-stationary Gaussian signal and deterioration caused by a linear shift-invariant operator and additive white Gaussian noise, the results clearly explain the main principles of complexity-regularized restoration.

In general, theoretical studies of rate-distortion problems for the Gaussian case provide to the signal processing practice optimistic beliefs about which design concepts perform well for the real-world non-Gaussian instances of the problems (see the excellent discussion in [31, Sec. 3]). Moreover, theoretical and

Algorithm 3: Proposed Method Based on Total Complexity of All the Overlapping Blocks with Robust Dual Variables.

- 1: Inputs: \mathbf{y} , β , θ .
 - 2: Initialize $\{\hat{\mathbf{z}}^{j,(0)}\}_{j=1}^{N_b}$ (depending on the deterioration type).
 - 3: $t = 1$ and $\mathbf{u}_{total}^{(1)} = \mathbf{0}$.
 - 4: **repeat**
 - 5: $\tilde{\mathbf{z}}^{j,(t)} = \hat{\mathbf{z}}^{j,(t-1)} - \mathbf{u}_{total}^{(t)}$ for $j = 1, \dots, N_b$
 - 6: Solve the ℓ_2 -constrained deconvolution:

$$\hat{\mathbf{x}}^{(t)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\beta}{2} \sum_{j=1}^{N_b} \|\mathbf{x} - \tilde{\mathbf{z}}^{j,(t)}\|_2^2$$
 - 7: **for** $j = 1, \dots, N_b$ **do**
 - 8: $\tilde{\mathbf{x}}_{shifted}^{j,(t)} = \operatorname{shift}_j \left\{ \hat{\mathbf{x}}^{(t)} + \mathbf{u}_{total}^{(t)} \right\}$
 - 9: $\hat{\mathbf{z}}_{shifted}^{j,(t)} = \operatorname{CompressDecompress}_{\theta} \left(\tilde{\mathbf{x}}_{shifted}^{j,(t)} \right)$
 - 10: $\hat{\mathbf{z}}^{j,(t)} = \operatorname{shift}_j^{-1} \left\{ \hat{\mathbf{z}}_{shifted}^{j,(t)} \right\}$
 - 11: $\mathbf{u}^{j,(t+1)} = \mathbf{u}^{j,(t)} + \left(\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}^{j,(t)} \right)$
 - 12: **end for**
 - 13: $\mathbf{u}_{total}^{(t+1)} = \frac{1}{N_b} \sum_{j=1}^{N_b} \mathbf{u}^{j,(t+1)}$
 - 14: $t \leftarrow t + 1$
 - 15: **until** stopping criterion is satisfied
-

practical solutions may embody in a different way the same general concepts. Therefore, one should look for connections between theory and practice in the form of high-level analogies.

The optimal solution presented in this section considers the classical framework of rate-distortion theory and a particular, however, important case of a Gaussian signal and a linear shift-invariant degradation operator. Our rate-distortion analysis below will show that the optimal complexity-regularized restoration consists of the following two main ideas: pseudoinverse filtering of the degraded input, and compression with respect to a squared-error metric that is weighted based on the degradation-filter squared-magnitude (considering a processing in the Discrete Fourier Transform (DFT) domain). In SubSection IV-D we explain how these two concepts connect to more general themes having different realizations in the practical approach proposed in Section III.

In this section, consider the signal $\mathbf{x} \in \mathbb{R}^N$ modeled as a zero-mean cyclo-stationary Gaussian random vector with a circulant autocorrelation matrix \mathbf{R}_x , i.e., $\mathbf{x} \sim \mathcal{N}(0, \mathbf{R}_x)$. The degradation model studied remains

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (36)$$

where here \mathbf{H} is a real-valued $N \times N$ circulant matrix representing a linear shift-invariant deteriorating operation and $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I})$ is a length N vector of white Gaussian noise. Clearly, the degraded observation \mathbf{y} is also a zero-mean cyclo-stationary Gaussian random vector with a circulant autocorrelation matrix $\mathbf{R}_y = \mathbf{H}\mathbf{R}_x\mathbf{H}^* + \sigma_n^2 \mathbf{I}$.

A. Prevalent Restoration Strategies

We precede the analysis of the complexity-regularized restoration with mentioning three well-known estimation meth-

ods. The restoration procedure is a function

$$\hat{\mathbf{x}} = f(\mathbf{y}), \quad (37)$$

where f maps the degraded signal \mathbf{y} to an estimate of \mathbf{x} denoted as $\hat{\mathbf{x}}$. In practice, one gets a realization of \mathbf{y} denoted here as \mathbf{y}_r and forms the corresponding estimate as $\hat{\mathbf{x}}_r = f(\mathbf{y}_r)$.

1) *Minimum Mean Squared Error (MMSE) Estimate:* This restoration minimizes the expected MSE of the estimate, i.e.,

$$f_{MMSE} = \underset{f}{\operatorname{argmin}} E \left\{ \|\mathbf{x} - f(\mathbf{y})\|_2^2 \right\}, \quad (38)$$

yielding that the corresponding estimate is the conditional expectation of \mathbf{x} given \mathbf{y}

$$\hat{\mathbf{x}}_{MMSE} = f_{MMSE}(\mathbf{y}) = E \{ \mathbf{x} | \mathbf{y} \}. \quad (39)$$

Nicely, for the Gaussian case considered in this section, the MMSE estimate (39) reduces to a linear operator, presented below as the Wiener filter.

2) *Wiener Filtering:* The Wiener filter is also known as the Linear Minimum Mean Squared Error (LMMSE) estimate, corresponding to a restoration function of the form

$$\hat{\mathbf{x}} = f_{Wiener}(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}, \quad (40)$$

optimized via

$$\left\{ \hat{\mathbf{A}}, \hat{\mathbf{b}} \right\} = \underset{\mathbf{A}, \mathbf{b}}{\operatorname{argmin}} E \left\{ \|\mathbf{x} - (\mathbf{A}\mathbf{y} + \mathbf{b})\|_2^2 \right\}. \quad (41)$$

In our case, where \mathbf{x} and \mathbf{y} are zero mean, $\hat{\mathbf{b}} = \mathbf{0}$ and

$$\hat{\mathbf{A}} = \mathbf{R}_x \mathbf{H}^* (\mathbf{H} \mathbf{R}_x \mathbf{H}^* + \sigma_n^2 \mathbf{I})^{-1}. \quad (42)$$

If the distributions are Gaussian, this linear operator coincides with the optimal MMSE estimator.

3) *Constrained Deconvolution Filtering:* This approach considers a given degraded signal $\mathbf{y}_r = \mathbf{H}\mathbf{x}_0 + \mathbf{n}_r$, with the noise vector a realization of a random process while the signal \mathbf{x}_0 is considered as a deterministic vector, with perhaps some known properties. Then, the restoration is carried out by minimizing a carefully-designed penalty function, g , that assumes lower values for \mathbf{x} vectors that fit the prior knowledge on \mathbf{x}_0 . Note that for a sufficiently large signal dimension we get that $\|\mathbf{y}_r - \mathbf{H}\mathbf{x}_0\|_2^2 = \|\mathbf{n}_r\|_2^2 \approx N\sigma_n^2$. The last result motivates to constrain the estimate, $\hat{\mathbf{x}}$, to conform with the known degradation model (36), by demanding the similarity of $\mathbf{y}_r - \mathbf{H}\hat{\mathbf{x}}$ to the additive noise term via the equality relation $\|\mathbf{y}_r - \mathbf{H}\hat{\mathbf{x}}\|_2^2 = N\sigma_n^2$. The above idea is implemented in an optimization of the form

$$\begin{aligned} \min_{\hat{\mathbf{x}}} \quad & g(\hat{\mathbf{x}}) \\ \text{subject to} \quad & \|\mathbf{y}_r - \mathbf{H}\hat{\mathbf{x}}\|_2^2 = N\sigma_n^2. \end{aligned} \quad (43)$$

Our practical methods presented in Section III emerge from an instance of the constrained deconvolution optimization (43), in its Lagrangian version, where the penalty function g is the cost in bits measuring the complexity in describing the estimate $\hat{\mathbf{x}}$. In the remainder of this section, we study the complexity-regularized restoration problem from a statistical perspective.

B. The Complexity-Regularized Restoration Problem and Its Equivalent Forms

Based on rate-distortion theory (e.g., see [32]), we consider the estimate of \mathbf{x} as a random vector $\hat{\mathbf{x}} \in \mathbb{R}^N$ with the probability density function (PDF) $p_{\hat{\mathbf{x}}}(x)$. The estimate characterization, $p_{\hat{\mathbf{x}}}(x)$, is determined by optimizing the conditional PDF $p_{\hat{\mathbf{x}}|y}(x|y)$, statistically representing the mapping between the given data \mathbf{y} and the decompression result $\hat{\mathbf{x}}$. Moreover, the rate is measured as the mutual information between $\hat{\mathbf{x}}$ and \mathbf{y} , defined via

$$I(\mathbf{y}; \hat{\mathbf{x}}) = \int p_{\mathbf{y}, \hat{\mathbf{x}}}(y, \hat{x}) \log \frac{p_{\mathbf{y}, \hat{\mathbf{x}}}(y, \hat{x})}{p_{\mathbf{y}}(y) p_{\hat{\mathbf{x}}}(x)} dy d\hat{x}. \quad (44)$$

Then, the basic form of the complexity-regularized restoration optimization is expressed as

Problem 1 (Basic Form):

$$\begin{aligned} \min_{p_{\hat{\mathbf{x}}|y}} \quad & I(\mathbf{y}; \hat{\mathbf{x}}) \\ \text{subject to} \quad & E \left\{ \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_2^2 \right\} = N\sigma_n^2. \end{aligned} \quad (45)$$

Here the estimate rate is minimized while maintaining suitability to the degradation model (36) using a distortion constraint set to achieve an a-priori known total noise energy. In general, Problem 1 is complicated to solve since the distortion constraint considers $\hat{\mathbf{x}}$ through the degradation operator \mathbf{H} , while the rate is directly evaluated for $\hat{\mathbf{x}}$.

The shift invariant operator \mathbf{H} is a circulant $N \times N$ matrix, thus, diagonalized by the $N \times N$ Discrete Fourier Transform (DFT) matrix \mathbf{F} . The (k, l) component of the DFT matrix $(k, l = 0, \dots, N-1)$ is $\mathbf{F}_{k,l} = W_N^{kl}$ where $W_N \triangleq \frac{1}{\sqrt{N}} e^{-i2\pi/N}$. Then, the diagonalization of \mathbf{H} is expressed as

$$\mathbf{F}\mathbf{H}\mathbf{F}^* = \mathbf{\Lambda}_H, \quad (46)$$

where $\mathbf{\Lambda}_H$ is a diagonal matrix formed by the components h_k^F for $k = 0, \dots, N-1$. Using $\mathbf{\Lambda}_H$ we define the pseudoinverse of \mathbf{H} as

$$\mathbf{H}^+ = \mathbf{F}^* \mathbf{\Lambda}_H^+ \mathbf{F}, \quad (47)$$

where $\mathbf{\Lambda}_H^+$ is the pseudoinverse of $\mathbf{\Lambda}_H$, an $N \times N$ diagonal matrix with the k th diagonal element:

$$h_k^{F,+} = \begin{cases} \frac{1}{h_k^F}, & \text{for } h_k^F \neq 0 \\ 0, & \text{for } h_k^F = 0. \end{cases} \quad (48)$$

We denote by N_H the number of nonzero diagonal elements in $\mathbf{\Lambda}_H$, the rank of \mathbf{H} .

The first main result of our analysis states that Problem 1, being the straightforward formulation for complexity-regularized restoration, is equivalent to the next problem.

Problem 2 (Pseudoinverse-filtered input):

$$\begin{aligned} \min_{p_{\hat{\mathbf{x}}|\tilde{\mathbf{y}}}} \quad & I(\tilde{\mathbf{y}}; \hat{\mathbf{x}}) \\ \text{subject to} \quad & E \left\{ \|\mathbf{H}(\tilde{\mathbf{y}} - \hat{\mathbf{x}})\|_2^2 \right\} = N_H \sigma_n^2, \end{aligned} \quad (49)$$

where

$$\tilde{\mathbf{y}} = \mathbf{H}^+ \mathbf{y} \quad (50)$$

is the pseudoinverse filtered version of the given degraded signal \mathbf{y} . One should note that Problem 2 has a more convenient form than Problem 1 since the distortion is an expected weighted squared error between the two random variables determining the rate. The equivalence of Problems 1 and 2 is proved in Appendix A (the appendices are provided in the supplementary material).

In this section, \mathbf{x} is a cyclo-stationary Gaussian signal, hence, having a circulant autocorrelation matrix \mathbf{R}_x . Consequently, and also because \mathbf{H} is circulant, the deteriorated signal \mathbf{y} is also a cyclo-stationary Gaussian signal. Moreover, \mathbf{H}^+ is also a circulant matrix, thus, by (50) the pseudoinverse filtering result, $\tilde{\mathbf{y}}$, is also cyclo-stationary and zero-mean Gaussian. Specifically, the autocorrelation matrix of $\tilde{\mathbf{y}}$ is

$$\mathbf{R}_{\tilde{\mathbf{y}}} = \mathbf{H}^+ \mathbf{R}_y \mathbf{H}^{+*} \quad (51)$$

$$= \mathbf{H}^+ \mathbf{H} \mathbf{R}_x \mathbf{H}^* \mathbf{H}^{+*} + \sigma_n^2 \mathbf{H}^+ \mathbf{H}^{+*}, \quad (52)$$

and, as a circulant matrix, it is diagonalized by the DFT matrix yielding the eigenvalues

$$\lambda_k^{(\tilde{\mathbf{y}})} = \begin{cases} \lambda_k^{(\mathbf{x})} + \frac{\sigma_n^2}{|h_k^F|^2}, & \text{for } h_k^F \neq 0 \\ 0, & \text{for } h_k^F = 0. \end{cases} \quad (53)$$

The DFT-domain representation of $\tilde{\mathbf{y}}$ is

$$\tilde{\mathbf{y}}^F = \mathbf{F}\tilde{\mathbf{y}}, \quad (54)$$

consisted of the coefficients $\{\tilde{y}_k^F\}_{k=0}^{N-1}$, being independent zero-mean Gaussian variables with variances corresponding to the eigenvalues in (53).

Transforming Problem 2 to the DFT domain, where $\tilde{\mathbf{y}}$ becomes a set of independent Gaussian variables to be coded under a joint distortion constraint, simplifies the optimization structure to the following separable form (see proof sketch in Appendix B).

Problem 3 (Separable form in DFT domain):

$$\begin{aligned} \min \quad & \sum_{k=0}^{N-1} I(\tilde{y}_k^F; \hat{x}_k^F) \\ \text{subject to} \quad & \sum_{k=0}^{N-1} |h_k^F|^2 E \left\{ |\tilde{y}_k^F - \hat{x}_k^F|^2 \right\} = N_H \sigma_n^2, \end{aligned} \quad (55)$$

where $\{\hat{x}_k^F\}_{k=0}^{N-1}$ are the elements of $\hat{\mathbf{x}}^F = \mathbf{F}\hat{\mathbf{x}}$. Nicely, the separable distortion in Problem 3 considers each variable using a squared error that is weighted by the squared magnitude of the corresponding degradation-filter coefficient.

The rate-distortion function of a single Gaussian variable with variance σ^2 has the known formulation [32]:

$$R(D) = \left[\frac{1}{2} \log \left(\frac{\sigma^2}{D} \right) \right]_+ \quad (56)$$

evaluating the minimal rate for a squared-error allowed reaching up to $D \geq 0$. In addition, the operator $[\cdot]_+$ is defined for real scalars as $[\alpha]_+ \triangleq \max\{\alpha, 0\}$, hence, $R(D) = 0$ for $D \geq \sigma^2$. Accordingly, the rate-distortion function of the Gaussian

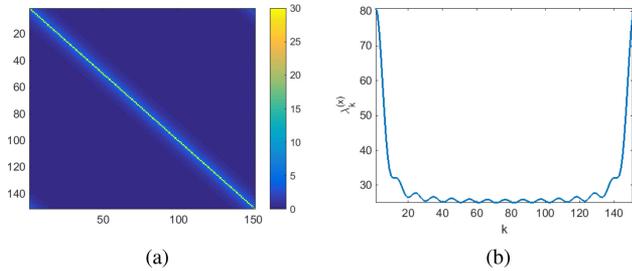


Fig. 1. The autocorrelation of the cyclo-stationary Gaussian signal used in the demonstration. (a) The circulant autocorrelation matrix in the signal domain, and (b) the corresponding eigenvalues obtained using the DFT decomposition.

variable \tilde{y}_k^F is

$$R_k(D_k) = \left[\frac{1}{2} \log \left(\frac{\lambda_k^{(\tilde{y})}}{D_k} \right) \right]_+ \quad (57)$$

where D_k denotes the maximal squared-error allowed for this component. Now, similar to the famous case of jointly coding independent Gaussian variables with respect to a regular (non-weighted) squared-error distortion [32], we explicitly express Problem 3 as the following distortion-allocation optimization.

Problem 4 (Explicit distortion allocation):

$$\begin{aligned} \min_{D_0, \dots, D_{N-1}} \quad & \sum_{k=0}^{N-1} \left[\frac{1}{2} \log \left(\frac{\lambda_k^{(\tilde{y})}}{D_k} \right) \right]_+ \\ \text{subject to} \quad & \sum_{k=0}^{N-1} |h_k^F|^2 D_k = N_H \sigma_n^2 \\ & D_k \geq 0, k = 0, \dots, N-1. \end{aligned} \quad (58)$$

The optimal distortion-allocation satisfying the last optimization is

$$D_k^{opt} = \begin{cases} \frac{\sigma_n^2}{|h_k^F|^2}, & \text{for } h_k^F \neq 0 \\ 0, & \text{for } h_k^F = 0 \end{cases} \quad (59)$$

and the associated optimal rates are

$$R_k^{opt} = \begin{cases} \frac{1}{2} \log \left(|h_k^F|^2 \frac{\lambda_k^{(x)}}{\sigma_n^2} + 1 \right), & \text{for } h_k^F \neq 0 \\ 0, & \text{for } h_k^F = 0. \end{cases} \quad (60)$$

Results (59) and (60) are proved in Appendix C.

C. Demonstration of the Explicit Results

Let us exemplify the optimal rate-distortion results (59)–(60) for a cyclo-stationary Gaussian signal, \mathbf{x} , having the circulant autocorrelation matrix presented in Fig. 1(a), corresponding to the eigenvalues $\{\lambda_k^{(x)}\}_{k=0}^{N-1}$ (Fig. 1(b)) obtained by a DFT-based decomposition. We first examine the denoising problem, where the signal-domain degradation matrix is $\mathbf{H} = \mathbf{I}$ (Fig. 2(a)) and its respective DFT-domain spectral representation consists of $h_k^F = 1$ for any k (see Fig. 2(b)). The additive white Gaussian noise has a sample variance of $\sigma_n^2 = 5$. Fig. 2(c) exhibits the optimal distortion allocation using a reverse-waterfilling diagram, where the signal-energy distribution $\{\lambda_k^{(x)}\}_{k=0}^{N-1}$ (black solid line) and the additive noise energy (the light-red region) defining together

the noisy-signal energy level (purple solid line) corresponding to $\lambda_k^{(\tilde{y})} = \lambda_k^{(x)} + \sigma_n^2$. The blue dashed line in Fig. 2(c) shows the water level associated with the uniform distortion allocation. The optimal rate-allocation, corresponding to Fig. 2(c) and Eq. (60), is presented in Fig. 2(d) showing that more bits are spent on components with higher signal-to-noise ratios.

Another example considers the same Gaussian signal described in Fig. 1 and the noise level of $\sigma_n^2 = 5$, but here the degradation operator is the circulant matrix shown in Fig. 3(a) having a DFT-domain representation given in magnitude-levels in Fig. 3(b) exhibiting its frequency attenuation and amplification effects. The waterfilling diagram in Fig. 3(c) includes the same level of signal energy (black solid line) as in the denoising experiment, but the effective additive noise levels and the allocated distortions are clearly modulated in an inversely proportional manner by the squared magnitude of the degradation operator. For instance, frequencies corresponding to degradation-filter magnitudes lower than 1 lead to increase in the effective noise-energy addition and in the allocated distortion. The optimal rate allocation (Fig. 3(d)) is affected by the signal-to-noise ratio and by the squared-magnitude of the degradation filter (see also Eq. (60)), e.g., components that are attenuated by the degradation operator get less bits in the rate allocation.

D. Conceptual Relation to the Proposed Approach

As explained at the beginning of this section, theoretical and practical solutions may include different implementations of the same general ideas. Accordingly, connections between theory and practice should be established by pointing on high-level analogies. Our rate-distortion analysis (for a Gaussian signal and a LSI degradation operator) showed that the optimal complexity-regularized restoration relies on two prominent ideas: pseudoinverse filtering of the degraded input, and compression with respect to a squared-error metric that is weighted based on the degradation-filter squared-magnitude (considering the DFT-domain procedure). We will now turn to explain how these two concepts connect to more general themes having different realizations in the practical approach proposed in Section III.¹

• *Design Concept #1. Apply simple restoration filtering:* The general idea of using an elementary restoration filter is implemented in the Gaussian case as pseudoinverse filtering. Correspondingly, our practical approach relies on a simple filtering mechanism, extending the pseudoinverse filter as explained next. Stage 6 of Algorithm 1 is an ℓ_2 -constrained deconvolution filtering that its analytic solution can be rewritten, using the relation $\mathbf{H}^* (\mathbf{I} - \mathbf{H}\mathbf{H}^*) = \mathbf{0}$, as (see proof in Appendix D)

$$\hat{\mathbf{x}}^{(t)} = \left(\mathbf{H}^* \mathbf{H} + \frac{\beta}{2} \mathbf{I} \right)^{-1} \left(\mathbf{H}^* \mathbf{H} \tilde{\mathbf{y}} + \frac{\beta}{2} \tilde{\mathbf{z}}^{(t)} \right). \quad (61)$$

As before, $\tilde{\mathbf{y}} = \mathbf{H}^+ \mathbf{y}$, i.e., the pseudoinverse-filtered version of \mathbf{y} . The expression (61) can be interpreted as an initial pseudoinverse filtering of the degraded input, followed by a simple

¹Since the differences between Algorithm 1 and Algorithms 2 and 3 are for a shift-invariance purpose, an issue that we do not concern in this section, we compare our theoretic results only to Algorithm 1.

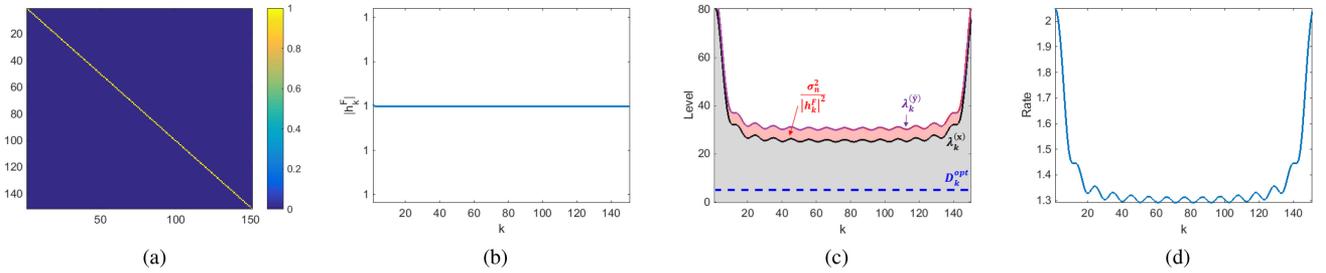


Fig. 2. Demonstrating the theoretic results for a denoising problem with a noise level of $\sigma_n^2 = 5$. (a) The degenerated degradation operator in the signal domain $\mathbf{H} = \mathbf{I}$. (b) DFT-domain magnitude of the degradation filter. (c) Optimal waterfilling solution in DFT domain. (d) Optimal rate allocation in DFT domain.

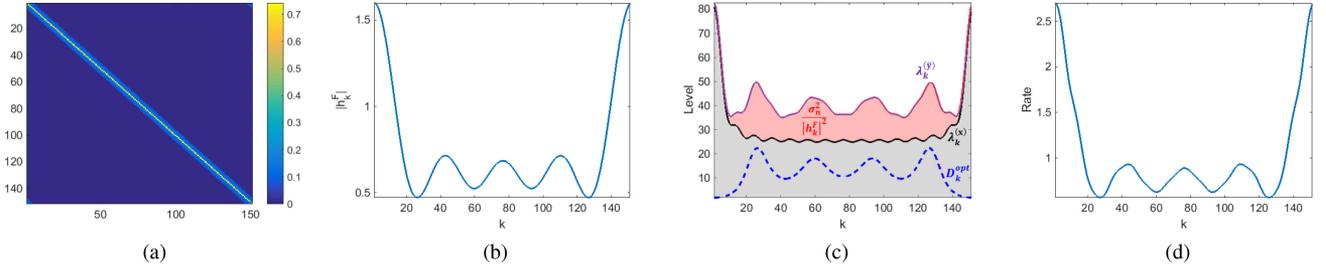


Fig. 3. Demonstrating the theoretic results for a restoration problem with a noise level of $\sigma_n^2 = 5$. (a) The degradation operator in the signal domain \mathbf{H} . (b) DFT-domain magnitude of the degradation filter. (c) Optimal waterfilling solution in DFT domain. (d) Optimal rate allocation in DFT domain.

weighted averaging with $\tilde{\mathbf{z}}^{(t)}$ (that includes the decompressed signal obtained in the last iteration). Evidently, the filtering in (61) is determined by the β value, specifically, for $\beta = 0$ the estimate coincides with the pseudoinverse filtering solution and for a larger β it is closer to $\tilde{\mathbf{z}}^{(t)}$.

• *Design Concept #2: Compress by promoting higher quality for signal-components matching to higher h-operator magnitudes:* This principle is realized in the theoretic Gaussian case as weights attached to the squared-errors of DFT-domain components (see Problems 3 and 4). Since the weights, $(|h_0^F|^2, \dots, |h_{N-1}^F|^2)$, are the squared magnitudes of the corresponding degradation-filter coefficients, in the compression of the pseudoinverse-filtered input the distortion is spread unevenly being larger where the degradation filter-magnitude is lower. Remarkably, this concept is implemented differently in the proposed procedure (Algorithm 1) where regular compression techniques, optimized for the squared-error distortion measure, are applied on the filtering result of the preceding stage. We will consider the essence of the effective compression corresponding to these two stages together. Let us revisit (61), expressing stage 6 of Algorithm 1. Assuming \mathbf{H} is a circulant matrix, we can transform (61) into its Fourier domain representation

$$\hat{\mathbf{x}}^{F,(t)} = \left(\Lambda_H^* \Lambda_H + \frac{\beta}{2} \mathbf{I} \right)^{-1} \left(\Lambda_H^* \Lambda_H \tilde{\mathbf{y}}^F + \frac{\beta}{2} \tilde{\mathbf{z}}^{F,(t)} \right) \quad (62)$$

where $\hat{\mathbf{x}}^{F,(t)}$ and $\tilde{\mathbf{z}}^{F,(t)}$ are the Fourier representations of $\hat{\mathbf{x}}^{(t)}$ and $\tilde{\mathbf{z}}^{(t)}$, respectively. Furthermore, (62) reduces to the componentwise formulation

$$\hat{x}_k^{F,(t)} = \frac{|h_k^F|^2 \tilde{y}_k^F + \frac{\beta}{2} \tilde{z}_k^{F,(t)}}{|h_k^F|^2 + \frac{\beta}{2}} \quad (63)$$

where $\hat{x}_k^{F,(t)}$ and $\tilde{z}_k^{F,(t)}$ are the k th Fourier coefficients of $\hat{\mathbf{x}}^{F,(t)}$ and $\tilde{\mathbf{z}}^{F,(t)}$, respectively. Equation (63) shows that signal

elements (of the pseudoinverse-filtered input) corresponding to degradation-filter components of weaker energies will be retracted more closely to the respective components of $\tilde{z}_k^{F,(t)}$ —thus, will be farther from \tilde{y}_k^F , yielding that the corresponding components in the standard compression applied in the next stage of this iteration will be of a relatively lower quality with respect their matching components of \tilde{y}_k^F (as they were already retracted relatively far from them in the preceding deconvolution stage).

To conclude this section, we showed that the main architectural ideas expressed in theory (for the Gaussian case) appear also in our practical procedure. The iterative nature of our methods (Algorithms 1–3) as well as the desired shift-invariance property provided by Algorithms 2–3 are outcomes of treating real-world scenarios such as non-Gaussian signals, general linear degradation operators, and computational limitations leading to block-based treatments—these all relate to practical aspects, hence, do not affect the fundamental treatment given in this section.

V. EXPERIMENTAL RESULTS

In this section we present experimental results for image restoration. Our main study cases include deblurring and inpainting using the image-compression profile of the HEVC standard (in its BPG implementation [33]). We also provide evaluation of our method in conjunction with the JPEG2000 technique for the task of image deblurring.

We empirically found it sufficient to consider only a part of all the shifts, i.e., a portion of \mathcal{B}^* . When using HEVC, the limited amount of shifts is compensated by the compression architecture that employs inter-block spatial predictions, thus, improves upon methods relying on independent block treatment. The shifts are defined by the rectangular images having their upper-left corner pixel relatively close to the upper-left

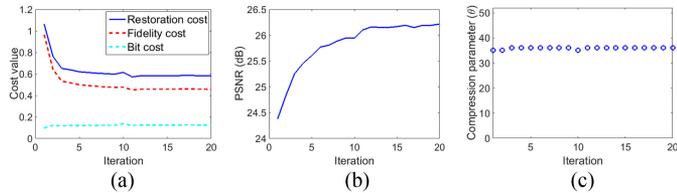


Fig. 4. Empirical analysis for deblurring the Cameraman image using Algorithm 1 (i.e., without overlapping blocks and shifted grids). The parameter settings here are: $\mu = (6.67 \times 10^{-6})$ and $\beta = 0.01$.

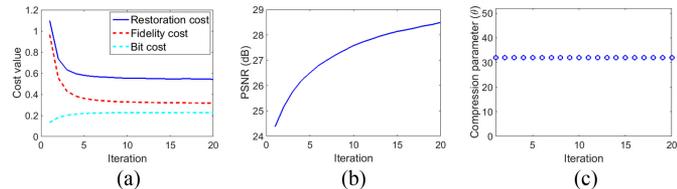


Fig. 5. Empirical analysis for deblurring the Cameraman image using Algorithm 2 considering 9 shifted block grids. The parameter settings here are: $\mu = \frac{(6.67 \times 10^{-6})}{\text{number of shifts}}$ and $\beta = 0.01$.

corner of the full image, and their bottom-right corner pixel coincides with that of the full image. This extends the mathematical developments in Section III as practical compression handles arbitrarily sized rectangular images.

Many image regularizers have a visual interpretation, for example, the classical image-smoothness evaluation. In our framework, the regularization part in (10) measures the complexity in terms of the compression bit-cost with respect to a specific compression architecture, designed based on some image model. Our complexity regularization also has a general visual meaning since, commonly, low bit-cost compressed images tend to be overly smooth or piecewise-smooth.

A. Image Deblurring

Here we consider two deterioration settings taken from [34]. The first setting, denoted here also as ‘Set. 1’, considers a noise variance $\sigma_n^2 = 2$ coupled with a blur operator defined by the two-dimensional point-spread-function (PSF) $h(x_1, x_2) = 1/(1 + x_1^2 + x_2^2)$ for $x_1, x_2 = -7, \dots, 7$, and zero-valued otherwise. The second setting, denoted here also as ‘Set. 2’ (named in [34] as ‘Scenario 3’), considers a noise variance $\sigma_n^2 \approx 0.3$ joint with a blur operator defined by the two-dimensional uniform blur PSF of size 9×9 .

We precede the deblurring experiments with empirical evaluations of four important aspects of the proposed method.

1) *Iterative Reduction of the Fundamental Restoration Cost:* In Section III we established the basic optimization problems for restoration by regularizing the bit-costs of the non-overlapping and the overlapping blocks of the estimate (see (11) and (23), respectively). As explained above, these two fundamental optimization tasks cannot be directly addressed and, therefore, we developed the ADMM-based Algorithms 1–3, that iteratively employ simpler optimization problems. Figs. 4(a) and 5(a) demonstrate that, for appropriate parameter settings, the fundamental optimization cost reduces in each iteration. The provided figures also show the fidelity term, $\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$, and the regularizing bit-cost (multiplied by μ) of each iteration.

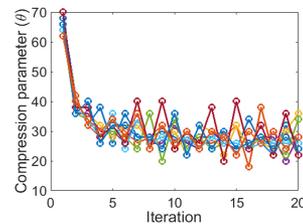


Fig. 6. Empirical analysis for deblurring the Cameraman image using Algorithm 2 with JPEG2000 considering 9 shifted block grids. The compression parameter is the compression ratio given to the JPEG2000 compression. The parameter settings here are: 9 shifts, $\mu = \frac{(1.33 \times 10^{-4})}{\text{number of shifts}}$, $\beta = 25 \times 10^{-4}$.

TABLE I
PARAMETER SETTINGS USED FOR THE DEBLURRING AND INPAINTING RESULTS IN TABLES II AND III

	Maximal Number of Iterations	Number of Shifts	β	Compression Parameter θ
Deblurring Algorithm 2 and 3 with HEVC	10 for ‘Set. 1’ 15 for ‘Set. 2’	400	5×10^{-3} for ‘Set. 1’ 10^{-3} for ‘Set. 2’	Fixed on 40
Deblurring Algorithm 2 and 3 with JPEG2000	10	3600	5×10^{-3} for ‘Set. 1’ 25×10^{-4} for ‘Set. 2’	Start at 90, decrease in 10 in each iteration keep fix when arriving to 10
Inpainting Algorithm 2 and 3 with HEVC	35	400	$\frac{0.1}{\text{number of shifts}}$	Fixed on 35 for Algorithm 2 Fixed on 40 for Algorithm 3
Inpainting Algorithm 2 with JPEG2000	10 for Algorithm 2 20 for Algorithm 3	3600	0 for Algorithm 2 $\frac{0.2}{\text{number of shifts}}$ for Algorithm 3	For Algorithm 2: Start at 78, decrease in 2 in each iteration keep fix when arriving to 25 For Algorithm 3: Start at 98, decrease in 2 in each iteration keep fix when arriving to 50

2) *Iterative Improvement of the Restored Image:* The fundamental optimization costs in (11) and (23) include the fidelity term $\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$ that considers the candidate estimate \mathbf{x} and the given degraded signal \mathbf{y} . However, the ultimate goal of the restoration process is to produce an estimate \mathbf{x} that will be close to the original (unknown!) signal \mathbf{x}_0 . It is common to evaluate proposed methods in experiment settings where \mathbf{x}_0 is known and used only for the evaluation of the squared error $\|\mathbf{x} - \mathbf{x}_0\|_2^2$ or its corresponding PSNR. Accordingly, it is a desired property that our iterative methods will provide increment in the PSNR along the iterations and, indeed, Figs. 4(b) and 5(b) show that this is achievable for appropriate parameter settings (the use of improper parameters may lead to unwanted decrease of the PSNR starting at some unknown iteration that, however, can be detected in many cases by heuristic divergence rules based on the dual variables used in the ADMM process). Interestingly, for some parameter settings, the PSNR may increase with the iterations, whereas the fundamental restoration cost will not necessarily consistently decrease. The last behavior may result from the fact the our optimization problem (with respect to a standard compression technique) is discrete, non-linear, and usually not convex and, therefore, the convergence guarantees of the ADMM [10] do not hold here for the fundamental restoration cost.

3) *The Optimal Compression Parameter:* Another question of practical importance is the value of the parameter θ , determining the compression level of the standard technique utilized in the proposed Algorithms 1–3. Recall that the ADMM-based developments in Section III led to an iterative procedure including a stage of Lagrangian rate-distortion optimization

TABLE II
DEBLURRING: PSNR COMPARISON (THE THREE BEST RESULTS IN EACH COLUMN APPEAR IN BOLD TEXT)

	Cameraman 256x256		House 256x256		Lena 512x512		Barbara 512x512	
	Set. 1	Set. 2	Set. 1	Set. 2	Set. 1	Set. 2	Set. 1	Set. 2
Input PSNR	22.23	20.76	25.61	24.11	27.25	25.84	23.34	22.49
ForWaRD [35]	28.99	28.10	32.96	33.67	33.30	32.81	27.03	26.51
SV-GSM [36]	29.68	28.09	34.25	33.15	-	-	30.19	27.56
BM3DDEB [37]	30.42	29.10	34.93	34.96	35.20	33.81	31.14	28.35
TVMM [38]	29.64	29.30	33.59	34.50	33.61	33.31	26.44	25.98
CGMK [39]	30.03	29.91	33.92	34.86	34.01	33.70	25.79	26.04
IDD-BM3D [34]	31.08	31.21	35.56	37.00	35.22	34.75	30.98	28.54
EPLL [17]	29.40	29.54	33.88	35.87	34.64	34.39	27.62	26.78
Proposed Algorithm 2 with JPEG2000	28.99	27.48	33.55	33.89	34.11	32.56	26.42	25.14
Proposed Algorithm 3 with JPEG2000	29.28	28.10	33.07	34.01	34.38	32.60	26.31	25.18
Proposed Algorithm 2 with HEVC	30.19	30.20	33.17	36.47	33.32	34.40	29.83	27.74
Proposed Algorithm 3 with HEVC	30.35	30.14	34.37	36.57	34.55	34.41	30.20	27.72

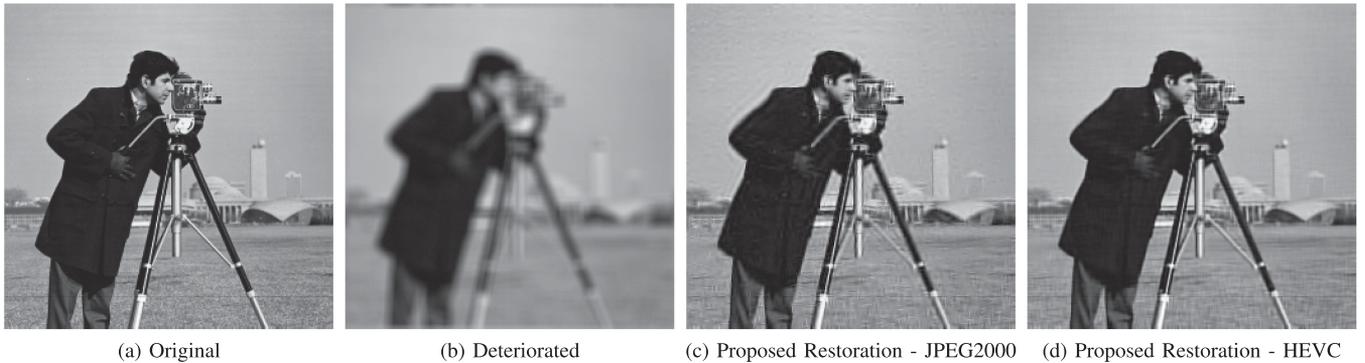


Fig. 7. The deblurring experiment (settings #2) for the Cameraman image (256×256). (a) The underlying image. (b) Degraded image (20.76 dB). (c) Restored image using Algorithm 3 with JPEG2000 compression (28.10 dB). (d) Restored image using Algorithm 3 with HEVC compression (30.14 dB).

operated for a Lagrange multiplier $\lambda \triangleq \frac{2\mu}{\beta}$ and, then, we replaced this optimization with application of a standard compression-decompression technique operated based on a parameter θ . It is clear that θ is a function of λ . In the particular case where the standard compression has the Lagrangian form from our developments, then $\theta = \lambda$, however, this is not the general case. For an arbitrary compression technique, we assume that its parameter θ has K possible values $\theta_1, \dots, \theta_K$ (for example, the HEVC standard supports 52 values for its quantization parameter), then, for a given $\lambda \triangleq \frac{2\mu}{\beta}$ the required θ value in stage 8 of Algorithm 1 can be determined via

$$\theta_{\lambda, opt}^{(t)} = \underset{\theta \in \{\theta_1, \dots, \theta_K\}}{\operatorname{argmin}} \left\| \tilde{\mathbf{x}}^{(t)} - \hat{\mathbf{z}}_{\theta}^{(t)} \right\|_2^2 + \lambda r_{tot, \theta}. \quad (64)$$

where $\hat{\mathbf{z}}_{\theta}^{(t)} = \operatorname{CompressDecompress}_{\theta}(\tilde{\mathbf{x}}^{(t)})$ is the decompressed signal and $r_{tot, \theta}$ is the associated compression bit-cost. We present here experiments (see Figs. 4 and 5) for Algorithm 1

and 2 that in each iteration optimize the θ value corresponding to $\lambda \triangleq \frac{2\mu}{\beta}$ based on procedures similar to (64). Nicely, it is shown in Figs. 4(c) and 5(c) that, for the HEVC compression used here, the best θ values along the iterations are nearly the same (for a specific restoration task). This important property may be a result of the fact that HEVC extensively relies on Lagrangian rate-distortion optimizations (although in much more complex forms than those presented in Section III). Accordingly, in order to reduce the computational load, in the experiments shown below we will use a constant compression parameter given as an input to our methods. Interestingly, when examining the optimal compression parameters (compression ratios in this case) for the JPEG2000 method that applies wavelet-based transform coding, there is a decrease in the optimal compression ratio along the iterations (see Fig. 6(a)). Accordingly, in order to reduce the computational load in the experiments below, we employed a predefined rule for reducing the JPEG2000 compression ratio along the iterations. Importantly, when we use the sub-optimal

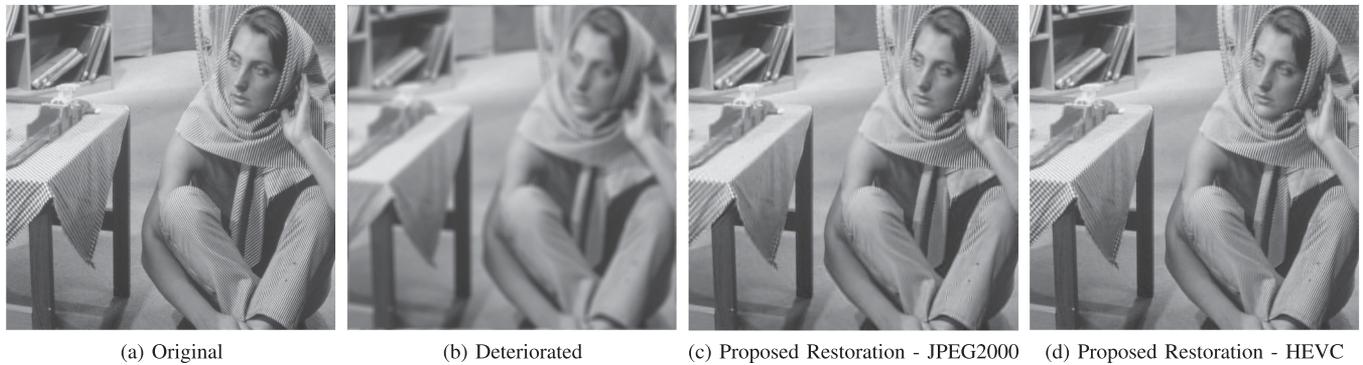


Fig. 8. The deblurring experiment (settings #2) for the Barbara image (512×512). (a) The underlying image. (b) Degraded image (22.49 dB). (c) Restored image using Algorithm 3 with JPEG2000 compression (25.18 dB). (d) Restored image using Algorithm 3 with HEVC compression (27.72 dB).

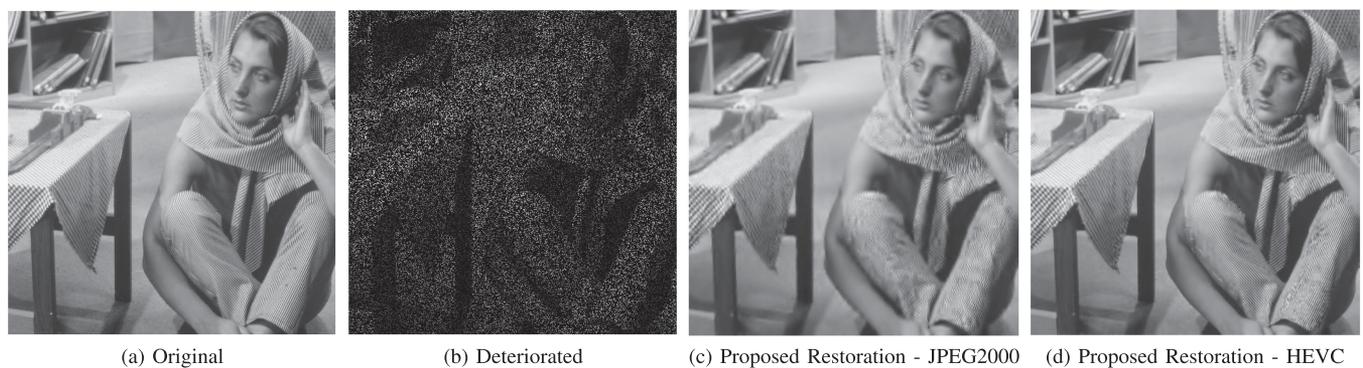


Fig. 9. The inpainting experiment (80% missing pixels) for the Barbara image (512×512). (a) The original image. (b) Deteriorated image. (c) Restored image using Algorithm 3 with JPEG2000 compression (24.83 dB). (d) Restored image using Algorithm 3 with HEVC compression (28.80 dB).

predefined rules for setting the compression parameter θ values (see Table I for the settings used in our evaluation comparisons in Table II) we do not longer need to set a value for μ (since, in these cases, μ is practically unused).

4) *Restoration Improvement for Increased Number of Image Shifts*: In our experiments we noticed that the restoration quality improves as more shifts are used, however, at some point the added gain due to the added shifts becomes marginal. As an example to the benefits due to shifts see Figs. 4(b) and 5(b), where the PSNR obtained for deblurring the Cameraman image using Algorithm 2 and 9 shifts is about 2dB higher than the PSNR obtained using Algorithm 1 (i.e., without additional shifts).

In our main evaluation, we examined the proposed Algorithms 2 and 3 for image deblurring in conjunction with the JPEG2000 and HEVC compression techniques (see the parameter settings in Table I). Table II shows a comparison between various deblurring methods tested in the above two settings for four grayscale images.² In 7 out of the 8 cases, the proposed Algorithms 2 or 3 utilizing the HEVC standard provided one of the best three results. Visual results are presented in Figs. 7 and 8.

²The results in Table II for the methods from [34]–[39] were taken as is from [34].

B. Image Inpainting

We presented in [9] experimental results for the inpainting problem, in its noisy and noiseless settings. Here we focus on the noiseless inpainting problem, where only pixel erasure occurs without an additive noise. The degradation is represented by a diagonal matrix \mathbf{H} of $N \times N$ size with main diagonal values of zeros and ones, indicating positions of missing and available pixels, respectively. Then, the product $\mathbf{H}\mathbf{x}$ equals to an N -length vector where its k th sample is determined by \mathbf{H} : if $\mathbf{H}[k, k] = 0$ then it is zero, and for $\mathbf{H}[k, k] = 1$ it equals to the corresponding sample of \mathbf{x} . The structure of the pixel erasure operator let us to simplify the optimization in step 6 of Algorithm 2. We note that \mathbf{H} is a square diagonal matrix and, therefore, $\mathbf{H}^T = \mathbf{H}$ and $\mathbf{H}^T \mathbf{y}$ is equivalent to a vector \mathbf{y} with zeroed components according to \mathbf{H} 's structure. Additional useful relation is $\mathbf{H}^T \mathbf{H} = \mathbf{H}$. Consequently, step 6 of Algorithm 2 facilitates a componentwise computation that is interpreted to form the k th sample of $\hat{\mathbf{x}}^{(t)}$ as

$$\hat{\mathbf{x}}^{(t)}[k] = \begin{cases} \frac{\mathbf{y}[k] + \frac{\beta}{2} \sum_{j=1}^{N_b} \tilde{\mathbf{z}}^{j,(t)}[k]}{1 + \frac{\beta}{2} N_b} & \text{for } \mathbf{H}[k, k] = 1 \\ \frac{1}{N_b} \sum_{j=1}^{N_b} \tilde{\mathbf{z}}^{j,(t)}[k] & \text{for } \mathbf{H}[k, k] = 0 \end{cases} \quad (65)$$

We initialize the shifted images $\{\tilde{\mathbf{z}}^{j,(0)}\}_{j=1}^{N_b}$ as the given image with the missing pixels set as the corresponding local averages

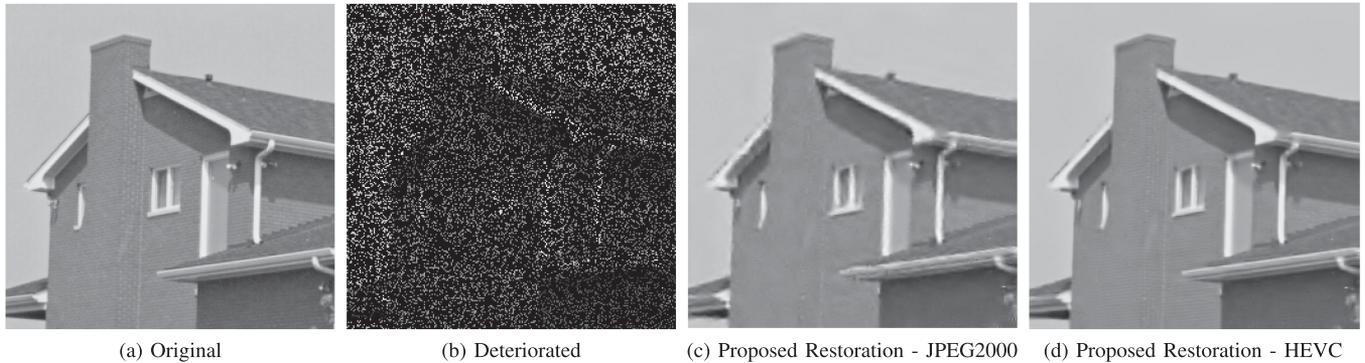


Fig. 10. The inpainting experiment (80% missing pixels) for the House image (256×256). (a) The original image. (b) Deteriorated image. (c) Restored image using Algorithm 3 with JPEG2000 compression (30.50 dB). (d) Restored image using Algorithm 3 with HEVC compression (33.10 dB).

TABLE III
IMAGE INPAINTING FROM 80% MISSING PIXELS: PSNR RESULTS

Image	Delaunay Triang.	ODCT	Li [42]	Yu et al. [43]	Ram et al. [40]	Proposed Algorithm 2 with JPEG2000	Proposed Algorithm 3 with JPEG2000	Proposed Algorithm 2 with HEVC	Proposed Algorithm 3 with HEVC
Lena 512x512	30.25	29.97	31.62	32.22	31.96	30.31	30.96	32.14	32.55
Barbara 512x512	22.88	27.15	25.40	30.94	29.71	24.25	24.83	26.06	28.80
House 256x256	29.21	29.69	32.87	33.05	32.71	29.49	30.50	32.42	33.10

of the available pixels in the respective 7×7 neighborhoods. When the iterative processing ends, we use the fact that the available pixels are noiseless and set them in the reconstructed image. The rest of the procedure remains as before.

We present here implementations of Algorithms 2 and 3 utilizing the JPEG2000 and HEVC image compression (the parameter settings are described in Table I). We consider the experimental settings from [40], where 80% of the pixels are missing (see Fig. 9(b) and 10(b)). Five competing inpainting methods are considered: cubic interpolation of missing pixels via Delaunay triangulation (using Matlab's 'griddata' function); inpainting using sparse representations of patches of 16×16 pixels based on an overcomplete DCT (ODCT) dictionary (see method description in [41, Ch. 15]); using patch-group transformation [42]; based on patch clustering [43]; and via patch reordering [40]. The PSNR values of images restored using the above methods (taken from [40]) are provided in Table III together with our results. For two images our HEVC-based implementation of Algorithm 3 provides the highest PSNR values. Visually, Figs. 9(d) and 10(d) exhibit the effectiveness of our method in repairing the vast amount of absent pixels.

VI. CONCLUSION

In this paper we explored the topic of complexity-regularized restoration, where the likelihood of candidate estimates are determined by their compression bit-costs. Using the alternating direction method of multipliers (ADMM) approach we developed three practical methods for restoration using standard compression techniques. Two of the proposed methods rely on a

new shift-invariant complexity regularizer, evaluating the total bit-cost of the signal shifted versions. We explained the main ideas of our approach using an insightful theoretical analysis of complexity-regularized restoration of a cyclo-stationary Gaussian signal from deterioration of a linear shift-invariant operator and additive white Gaussian noise. Experiments for deblurring and inpainting of images using the JPEG2000 and HEVC technique showed good results.

REFERENCES

- [1] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelet Analysis and Its Applications*, vol. 4. New York, NY, USA: Elsevier, 1994, pp. 299–324.
- [2] B. K. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2595–2605, Nov. 1995.
- [3] S. G. Chang, B. Yu, and M. Vetterli, "Image denoising via lossy compression and wavelet thresholding," in *Proc. Int. Conf. Image Process.*, 1997, vol. 1, pp. 604–607.
- [4] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.
- [5] J. Rissanen, "MDL denoising," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2537–2543, Nov. 2000.
- [6] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
- [7] J. Liu and P. Moulin, "Complexity-regularized image denoising," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 841–851, Jun. 2001.
- [8] P. Moulin and J. Liu, "Statistical imaging and complexity regularization," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1762–1777, Aug. 2000.
- [9] Y. Dar, A. M. Bruckstein, and M. Elad, "Image restoration via successive compression," in *Proc. Picture Coding Symp.*, 2016, pp. 1–5.

- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [11] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 945–948.
- [12] S. Sreehari *et al.*, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.
- [13] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [14] Y. Dar, A. M. Bruckstein, M. Elad, and R. Giryes, "Postprocessing of compressed images via sequential denoising," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3044–3058, Jul. 2016.
- [15] Y. Dar, A. M. Bruckstein, M. Elad, and R. Giryes, "Reducing artifacts of intra-frame video coding via sequential denoising," in *Proc. IEEE Int. Conf. Sci. Elect. Eng.*, 2016, pp. 1–5.
- [16] A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 96–108, 2016.
- [17] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 479–486.
- [18] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*. New York, NY, USA: Springer, 1995, pp. 125–150.
- [19] A. Nosratinia, "Enhancement of JPEG-compressed images by re-application of JPEG," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 27, no. 1/2, pp. 69–79, 2001.
- [20] A. Nosratinia, "Postprocessing of JPEG-2000 images to remove compression artifacts," *IEEE Signal Process. Lett.*, vol. 10, no. 10, pp. 296–299, Oct. 2003.
- [21] S. Beygi, S. Jalali, A. Maleki, and U. Mitra, "Compressed sensing of compressible signals," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 2158–2162.
- [22] S. Beygi, S. Jalali, A. Maleki, and U. Mitra, "An efficient algorithm for compression-based compressed sensing," 2017, arXiv:1704.01992.
- [23] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [24] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, "Image compression with anisotropic diffusion," *J. Math. Imag. Vis.*, vol. 31, no. 2, pp. 255–269, 2008.
- [25] C. Schmaltz, J. Weickert, and A. Bruhn, "Beating the quality of JPEG 2000 with anisotropic diffusion," in *Proc. DAGM Symp.*, 2009, pp. 452–461.
- [26] S. Andris, P. Peter, and J. Weickert, "A proof-of-concept framework for PDE-based video compression," in *Proc. Picture Coding Symp.*, 2016, pp. 1–5.
- [27] R. D. Adam, P. Peter, and J. Weickert, "Denoising by inpainting," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.*, 2017, pp. 121–132.
- [28] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.
- [29] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
- [30] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
- [31] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.
- [32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
- [33] F. Bellard, "BPG 0.9.6." [Online]. Available: <http://bellard.org/bpg/>. Accessed on: Oct. 15, 2017.
- [34] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.
- [35] R. Neelamani, H. Choi, and R. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 418–433, Feb. 2004.
- [36] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 27–41, Jan. 2008.
- [37] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *Proc. SPIE*, vol. 6812, 2008, Art. no. 681207.
- [38] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, 2009.
- [39] G. Chantas, N. P. Galatsanos, R. Molina, and A. K. Katsaggelos, "Variational Bayesian image restoration with a product of spatially weighted total variation image priors," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 351–362, Feb. 2010.
- [40] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2764–2774, Jul. 2013.
- [41] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.
- [42] X. Li, "Patch-based image interpolation: Algorithms and applications," in *Proc. Int. Workshop Local Non-Local Approx. Image Process.*, 2008, pp. 1–6.
- [43] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, May 2012.



Yehuda Dar received the B.Sc. degree in computer engineering and M.Sc. degree in electrical engineering, in 2008 and 2014, respectively, from the Technion–Israel Institute of Technology, Haifa, Israel, where he is currently working toward the Ph.D. degree with the Computer Science Department. He received the 2017 Jacobs Fellowship for excellence. His research interests include signal and image/video processing and their information-theoretic aspects, in particular, in signal compression and restoration.



Michael Elad (F'12) received the B.Sc., M.Sc., and D.Sc. degrees from the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel, in 1986, 1988, and 1997, respectively. Since 2003, he has been a faculty member with the Computer Science Department, Technion–Israel Institute of Technology, and since 2010 he holds a full professorship position. His research interests include signal and image processing, specializing in inverse problems, and sparse representations. He was the recipient of numerous teaching awards, the 2008 and 2015 Henri Taub Prizes for Academic Excellence, and the 2010 Hershel-Rich Prize for innovation. He is a SIAM Fellow (2018). He has been the Editor-in-Chief for *SIAM Journal on Imaging Sciences* since January 2016.



Alfred M. Bruckstein received the B.Sc. and M.Sc. degrees from the Technion–Israel Institute of Technology, Haifa, Israel, in 1976 and 1980, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1984. Since October 1984, he has been with the Technion–Israel Institute of Technology, where he holds the Ollendorff Chair in Science. His research interests include swarm/ant robotics, image and signal processing, analysis and synthesis, pattern recognition, and various aspects of applied geometry. He authored

and coauthored more than 150 journal papers in these fields. He is a SIAM Fellow for contributions to signal processing, image analysis, and ant robotics, and was the recipient of 2014 SIAG/Imaging Science Prize (for the From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images with D. Donoho and M. Elad).