2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Efficient JPEG decompression by the alternating direction method of multipliers

Michal Šorel and Michal Bartoš

Institute of Information Theory and Automation
Czech Academy of Sciences, Prague, Czechia
e-mail: {sorel,bartos}@utia.cas.cz

***Abstract*—Standard decompression of JPEG images produces artifacts along edges and a disturbing checkerboard pattern. To reduce these artifacts, decompression can be formulated as an image reconstruction problem within Bayesian maximum a posteriori probability framework. In this type of problem, the prior information about an image is typically given by the l_1 norm of its sparse domain representation. In this paper, we show how the solution of this problem can be achieved very efficiently using the alternating direction method of multipliers if the sparsity domain forms a tight frame. The proposed algorithm restores images without disturbing JPEG artifacts in several iterations, typically considerably less than competing algorithms. The quality of reconstruction both visually and in terms of SNR primarily depends on the tight frame used.**

***Index Terms*—image processing, image restoration, sparsity, JPEG, ADMM**

I. Introduction

Lossy compression using the JPEG standard [1] is probably the most common method to store image data. The loss of information caused by the compression process typically results in artifacts along strong edges and a visually disturbing checkerboard pattern. For this reason it is natural to look at JPEG decompression as an image restoration problem.

There are many JPEG decompression methods, from simple filters, locally reducing what is disturbing from a human perspective, to more elaborate methods based on more rigorous statistical formulations. The latter usually uses the maximum a posteriori probability (MAP) principle, looking for a solution with the highest posterior probability. Following the Bayes formula, this posterior probability is proportional to the product of the likelihood, describing the error introduced during the compression process, and of an approximation of the prior probability. As a rule, instead of maximizing the probability, we equivalently minimize negative log-probability, which transforms the product to the sum of negative log-likelihood and a regularization function.

In JPEG, the log-likelihood corresponds to the quantization constraint set (QCS), which is the set of images compression of which would result in what has actually been stored in the JPEG file. Since the compression is based on rounding the weighted coefficients of the discrete cosine transform (DCT), this set is given by intervals around the stored integer coefficients. This formulation was used in [2], [3], [4], [5]. An alternative approach is to approximate the QCS by a Gaussian distribution, which simplifies optimization and speeds up convergence [6], [7].

The quality of reconstruction depends mainly on the choice of image prior probability distribution represented by the corresponding regularization function. Early publications used smooth priors [6], [2], later methods incorporated non-differentiable sparse priors providing state-of-the-art results for other image restoration problems [8]. These include the total variation (TV) [3], [9], fields of experts (FoE) [7], total generalized variation (TGV) [4], [5], wavelets [10], sparse dictionaries [11], [12] and non-local means based algorithms [13]. State-of-the-art algorithms usually build on the ideas of sparsity and non-local means denoising [14], [15], [16], [17] or neural networks [18], achieving good results at the cost of longer run times.

The choice of the algorithm to optimize the MAP criterion depends on image priors. Early methods that used smooth priors applied gradient descent or for the QCS the projected gradient descent [6], [9]. Convexity of the QCS motivated the use of the projection on convex sets (POCS) method [19], with smoothness priors replaced by inequalities.

Bayesian estimation with sparse priors based on the l_1 norm found its use in many image processing, compressed sensing and machine learning applications. Resulting non-smooth functions are difficult to optimize by standard methods, which helped spread first-order techniques for non-smooth optimization, which are relatively simple to implement and fast enough to be practical [20], [21]. Probably the most popular are the alternating direction method of multipliers (ADMM) [22], also known as the split-Bregman method [20], and the Arrow-Hurwicz type algorithm [23].

For JPEG decompression, the algorithm [23] with TV regularization was used in [3]. The main disadvantage of TV in this context is that it favors unnatural piecewise constant functions, which is exactly the character of JPEG artifacts on block boundaries. To alleviate this problem, in [24], the same authors proposed a modified regularization term, the TGV.

In this paper, we investigate the use of ADMM to solve the MAP formulation of the JPEG decompression problem with sparse priors. This corresponds to the minimization of a regularization function subject to the convex QCS constraint. The regularization function is considered in the form of the l_1 norm of a linear operator applied on the image, where the

978-1-5090-4847-2/16/$31.00 ©2016 IEEE

linear operator is constrained to be a tight frame. In Sec. IV, we derive a closed-form solution to the most demanding algorithmic step of ADMM, which makes the proposed approach very efficient. This is our main contribution.

In the experimental section (Sec. V), we compare the convergence of ADMM with [23] used in the recent papers [3], [4], [5] and show that ADMM can be set to converge to acceptable visual reconstruction in several iterations. We also compare the behavior of the algorithm for several tight frames, including the dual-tree complex wavelets [25], a tight frame composed of several orders of standard Daubechies wavelets and the recent tight frames learned from data [26]. The authors of [26] claim that they can achieve performance similar to standard sparse dictionaries.

II. MAP formulation of JPEG decompression problem

JPEG compression [27] is a lossy type of compression based on the quantization of DCT coefficients, where DCT is applied to small blocks of usually 8×8 pixels. We first describe this process in detail and introduce mathematical notation we use. For grayscale images, the lossy part of JPEG compression can be expressed as

$$y = [QCx], \tag{1}$$

where x is a vectorized original image, y integer coefficients stored in the JPEG file, Q and C are matrices, and the square brackets denote the operator of rounding. C is the matrix of the block DCT made up of the square matrices of the DCT. C is orthogonal, because all the DCT submatrices are orthogonal. Q is a diagonal matrix corresponding to element-wise division by quantization coefficients from the quantization table stored in each JPEG file (64 values for 8×8 blocks) replicated for each block along diagonal. We will denote the vector of these coefficients as q, i.e. $Q = \mathrm{diag}(1/q)$. In the next section, we will see that matrices Q and C do not need to be formed explicitly. For color images, the image is first transformed into $Y'C_BC_R$ space and individual channels are stored separately. The brightness Y' is treated as described above but chrominance channels are often stored at smaller resolution, which complicates the degradation model. Formally,

$$y = [QCDx], \tag{2}$$

where D is a matrix describing the process of decreasing resolution (see Sec. IV). The grayscale case can be considered a special case of (2) with $D = I$, i.e. identity.

Given an observation y, a model describing the probability distribution of possible observations $p(y|x)$ and a prior probability $p(x)$, the Bayesian maximum a priori probability (MAP) approach maximizes the posterior probability $p(x|y) \sim p(y|x)p(x)$. Since in our case of rounding to the nearest integer the likelihood $p(\tilde{y}|x)$ is uniform within the quantization constraint set $-0.5 < y - QCx \leq 0.5$ (QCS), we get

$$\max_x p(x) \text{ s.t. } QCDx \in \left\langle y - \frac{1}{2}, y + \frac{1}{2} \right). \tag{3}$$

It is quite common in the MAP based image restoration that the prior probability $p(x)$ is considered in the form $p(x) \sim \exp(-\tau \left\|\Phi^T x\right\|_1)$, where Φ^T is a linear transform to a sparse domain such as the gradient, wavelets, or an overcomplete dictionary. Importantly, the inner part $\left\|\Phi^T x\right\|_1$ is convex. We in addition constrain the transform to satisfy $\Phi\Phi^T = I$, which we call a tight frame [28]. Note that sometimes tight frames are defined more generally as $\Phi\Phi^T = tI$ for a positive constant t. Nevertheless, since tight frames can always be normalized, we derive algorithms for $t = 1$ but use the general term tight frame.

In our experiments we use three type of tight frames. First, the dual-tree complex wavelet transform (DT-CWT) [25] represents well natural images and has a linear computational complexity. For comparison, we show results for a union of orthogonal Daubechies wavelets normalized to satisfy $\Phi\Phi^T = I$. Finally, frames can be learned from data [26]. For the TV and FoE priors in general $\Phi\Phi^T \neq I$ and therefore the procedure we derive cannot be used directly. Instead, we could use the algorithm [23].

Since the exponential function is increasing, the MAP formulation (3) is in our case equivalent to minimization

$$\min_x \|\Phi^T x\|_1 \text{ s.t. } QCDx \in \left\langle y - \frac{1}{2}, y + \frac{1}{2} \right). \tag{4}$$

In this paper, we restore the $Y'C_BC_R$ channels independently of each other.

III. ADMM

The main optimization tool we use is the alternating direction method of multipliers (ADMM) [22], [21] with many applications in image restoration and machine learning [29]. ADMM is a method to minimize the sum of two functions

$$\min_x f(x) + g(Gx), \tag{5}$$

where f and g are convex not necessarily differentiable functions and G is a linear operator. ADMM consists of iteratively executing three update steps

$$x \leftarrow \arg\min_x f(x) + \frac{\mu}{2} \|Gx - a - d\|^2, \tag{6}$$

$$a \leftarrow \arg\min_a g(a) + \frac{\mu}{2} \|Gx - a - d\|^2, \tag{7}$$

$$d \leftarrow d - (Gx - a), \tag{8}$$

where scalar $\mu > 0$ is a parameter, a is an auxiliary variable representing a sparse domain counterpart of x and d a dual variable. As a rule, the variable a is initialized by an initial estimate x_0 and d is initially zero. Iterations are stopped, when $||x-a||$ is smaller than a threshold or the number of iterations exceeds a limit. The convergence of the algorithm is proved in [22].

IV. Algorithm

At the end of Sec. II we explained how JPEG decompression can be formulated as the solution of problem (4) for each color

Algorithm 1 JPEG decompression by ADMM for one chrominance channel. The grayscale version of the algorithm is a special case with $D = I$ and $k = 1$. The operators used denote: D - down-sampling, C - block DCT, Q - division by quantization coefficients q, Φ - transform from sparse domain; D^T, C^T, Φ^T are respective transposes and y quantized block DCT coefficients stored in JPEG file. $P_{\langle l,u\rangle}(x)$ is projection of x on interval $\langle l, u\rangle$ taken element-wise.

1) Initialize $a_0 = \Phi^T D^T Q^{-1} C^T y$, $d_0 = 0$
2) repeat
3)
$$x_{k+1} = \left(I - \frac{1}{k} D^T D\right) \Phi\left(a_k + d_k\right) + D^T C^T \operatorname{diag}(\frac{1}{kq}) P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}\left(QCD\Phi\left(a_k + d_k\right)\right)$$
4)
$$a_{k+1} = \operatorname{sgn}\left(\Phi^T x_{k+1} - d_k\right) \cdot \max\left(0; \left|\Phi^T x_{k+1} - d_k\right| - \frac{1}{2\mu}\right)$$
5) $d_{k+1} = d_k - \Phi^T x_{k+1} + a_{k+1}$
6) until stopping criterion is satisfied

channel. For the purpose of optimization, we first reformulate (4) using an indicator function as

$$\min_x \|\Phi^T x\|_1 + I_{QCDx \in \left\langle y-\frac{1}{2}, y+\frac{1}{2}\right)}(x). \tag{9}$$

The indicator function of a set gives zero for vectors in the set and infinity otherwise. The indicator function of a convex set is convex. For technical reasons we replace the interval $\left\langle y-\frac{1}{2}, y+\frac{1}{2}\right)$ by its closed approximation $\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle$ for a sufficiently small constant $\varepsilon > 0$, getting a special case of (5) with $f(x) = I_{QCDx \in \left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}(x)$ and $g(x) = \|\Phi^T x\|_1$. Its global minimum can be found by ADMM. ADMM alternates solution of two convex problems

$$\arg\min_x \|\Phi^T x - a - d\| \text{ s.t. } QCDx \in \left\langle y - \frac{1}{2}, y + \frac{1}{2} - \varepsilon \right\rangle \tag{10}$$

and

$$\arg\min_a \|a\|_1 + \frac{\mu}{2}\|\Phi^T x - a - d\|^2, \tag{11}$$

supplemented by a simple update of the dual variable d. Equation (11) is the fast element-wise soft thresholding

$$a \leftarrow \operatorname{sgn}\left(\Phi^T x - d\right) \max\left(0; \left|\Phi^T x - d\right| - \frac{1}{2\mu}\right). \tag{12}$$

The critical point of the algorithm is a fast solution of (10). Let us assume that Φ is a tight frame ($\Phi\Phi^T = I$). Denoting $a + d$ as w, the definition of norm implies

$$\|\Phi^T x - w\|^2 = \|x - \Phi w\|^2 + w^T \Phi^T \Phi w - w^T w. \tag{13}$$

Because the right terms of (13) are independent of x, this transforms (10) to

$$\arg\min_x \|x - \Phi w\| \text{ s.t. } QCDx \in \left\langle y - \frac{1}{2}, y + \frac{1}{2} - \varepsilon \right\rangle, \tag{14}$$

which can be interpreted as a projection of Φw on the convex set given by the QCS. Let us assume that $D = I$ (grayscale image or the brightness channel of a color image). Recall that the DCT is orthogonal, i.e. $CC^T = C^T C = I$. Since Q is diagonal, $Q^{-1}\left\langle y - \frac{1}{2}, y + \frac{1}{2}\right\rangle$ is a box (with edges of different lengths) aligned with axes and $C^{-1}Q^{-1}\left\langle y - \frac{1}{2}, y + \frac{1}{2}\right\rangle$ a rotated version of the box. Therefore projection can be computed as

$$x \leftarrow C^{-1} Q^{-1} P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}(QC\Phi w), \tag{15}$$

where $P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}$ denotes the element-wise projection on the interval $\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle$, defined as $P_{\langle b_1,b_2\rangle}(y) = \min(\max(b_1, y), b_2)$. A formal proof will be given later as a special case of what holds for the chrominance channels.

For the decompression of chrominance channels, the problem is complicated by the down-sampling operation D in (2). Unfortunately, D is not fully specified in the JPEG standard and there are two options how to implement it. It is either the direct sampling at half resolution, i.e. taking every second pixel in each direction, or D computes the mean value of each square of size 2×2 pixels (in general, averaging over $m \times n$ pixels). On the other hand, it turns out that in both cases $DD^T = kI$[1] and therefore, denoting $A = QCD$, $AA^T = QCDD^T C^T Q^T = kQQ^T = \operatorname{diag}(kq^2)$ is a diagonal matrix. For simple down-sampling, $DD^T = I$. In the variant with averaging over $m \times n$ pixels $k = 1/(mn)$, i.e. for the most common $2:1$ subsampling [1] $k = 1/4$. Under these assumptions we can use Lemma 1 and compute the x-update projection as

$$x \leftarrow \Phi w - A^T\left(AA^T\right)^{-1}\left(A\Phi w - P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}(A\Phi w)\right). \tag{16}$$

We are obtaining

$$x \leftarrow \Phi w - A^T \operatorname{diag}(\frac{1}{kq^2})\left(A\Phi w - P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}(A\Phi w)\right), \tag{17}$$

$$x \leftarrow \Phi w - D^T C^T \operatorname{diag}(\frac{1}{kq}) \cdot \cdot \left(QCD\Phi w - P_{\left\langle y-\frac{1}{2}, y+\frac{1}{2}-\varepsilon\right\rangle}(QCD\Phi w)\right), \tag{18}$$

[1]Proof is trivial for direct subsampling. Because D^T replicates one pixel to a rectangle of $m \times n$ pixels and D picks one of them, we get $DD^T = I$. In the latter case, D^T replicates each pixel to $m \times n$ pixels multiplied by a scalar $k = 1/(mn)$. D computes the average of the same values, together giving $DD^T = kI$.

and finally

$$x \leftarrow \left(I - \frac{1}{k} D^T D\right) \Phi\left(a + d\right) + D^T C^T \operatorname{diag}\left(\frac{1}{kq}\right) P_{\left\langle y - \frac{1}{2}, y + \frac{1}{2} - \varepsilon \right\rangle} \left(QCD\Phi\left(a + d\right)\right). \quad (19)$$

Note that for D being identity and $k = 1$, the update (15) is a special case of (19). Eq. (15) only consists of element-wise operations, two DCTs and the transform to the sparse domain and back, Eq. (17) requires one additional up-sampling D^T and one down-sampling D. The resulting algorithm is summarized as Alg. 1.

V. Experiments

The first experiment was focused on the quality of reconstruction in comparison with ground truth. It was performed on a database of 60 images of historical city architecture. The images were taken from full-frame SLR camera under good lighting conditions. We further reduced noise by decreasing resolution eight times in both directions. Each image was compressed into the JPEG format with quality factors 10 to 90 with step 10 and decompressed/restored by the proposed algorithm with 3 different tight frames (details at the end of section II). The algorithm did 50 iterations for every preset ADMM parameter μ and the common μ giving best reconstructions in terms of improvement of signal-to-noise ratio (ISNR) was used to statistically compare the frames (Fig. 1). The best ISNR performance around 0.6 dB across all the quality factors gave the tight frame learned from data.

The second experiment focused on convergence properties of ADMM in comparison with [23], using for both cases the same priors. Note that [23] was used in recent papers [3], [4], [5] with different priors. This was done for the same image database with fixed quality factor 50 and minimization parameters (μ and two others in [23]) manually tuned to achieve the fastest possible convergence using the tight frame learned from data. The mean value of the functional (4) over all images in each iteration is plotted in Fig. 2. Our algorithm converged in several iterations to acceptable visual reconstructions, a plateau of the ADMM curve. The competing algorithm could not be set up to converge so fast at the beginning but had better convergence close to the minimum. Behavior of ADMM in later iterations can be improved by increasing μ in each iteration, though [30].

Fig. 3 shows an example of ADMM-based reconstruction using the DT-CWT tight frame. We show the result with parameter μ and number of iterations optimized to achieve the best ISNR (Fig. 3(b)), in contrast to the case of parameters optimized to converge fast to the minimum of the functional (Fig. 3(c)). Both approaches reduced the typical JPEG checkerboard pattern (sky) but the artifacts along edges (towers) are removed only for the converged image. Although the ISNR got worse in this case, visual perception of this reconstruction is better then for the image with the best ISNR.

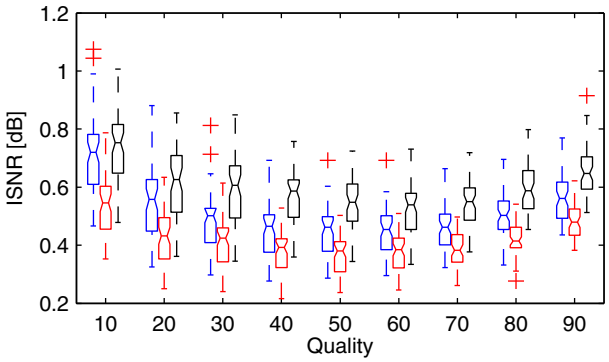

Figure 1. Statistical comparison of three different tight frames on a set of 60 images. In each group of boxplots (from left to right): DT-CWT (in blue), concatenated Daubechies wavelets (in red) and the tight frame learned from data [26] with kernels of size 8x8 pixels (in black). The box shows the median value and first/third quartiles, whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. The notches indicate the confidence intervals of the median.

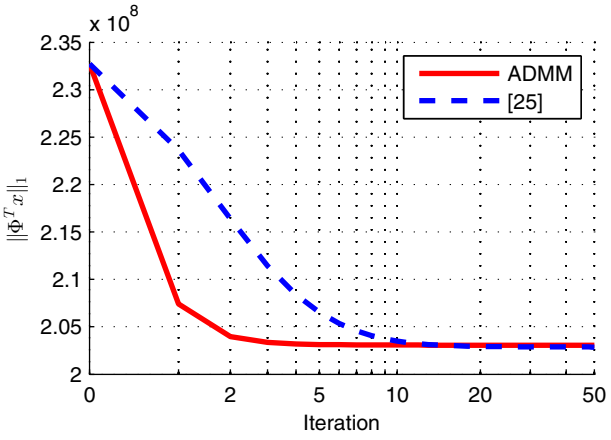

Figure 2. Mean convergence curves of ADMM and algorithm [23] obtained from the reconstructions of 60 images with minimization parameters tuned to achieve the maximum speed performance.

VI. Conclusion

We have presented an algorithm for image reconstruction from quantized DCT coefficients stored in the standardized JPEG format. The algorithm is formulated as a Bayesian maximum a posteriori probability problem with the tight frame sparse priors based on the l_1-norm. Convex minimization problems with convex constraints can be solved using ADMM but its application to the JPEG compression model consisting of the DCT, quantization and down-sampling operations is not straightforward. We derive a closed-form solution for the projection necessary to efficiently compute the most demanding algorithmic step, under the assumption that the linear operator used in regularization forms a tight frame. This result reduces the projection only to element-wise operations, two DCTs and the transform to a sparse domain and back. In the end, the algorithm alternately computes the derived projection and a simple operation of soft thresholding in the sparse domain. In

(a) Original image

(b) Best ISNR image

(c) Converged image

(d) JPEG image, quality $= 30$

Figure 3. Example of the restoration from a JPEG image stored at quality factor 30 using the DT-CWT tight frame. Whereas image b) is restored to achieve the best possible ISNR, image c) is close to the minimum of (4) after the algorithm converged.

our case a satisfactory solution is found within few iterations, typically less than for alternatives. In our experience, the number of iterations does not depend on image size, which means that the algorithm is approximately linear in the number of pixels.

The quality of reconstruction primarily depends on the tight frame used. In our experiments, the tight frame learned from data [26] gives the best performance, followed by the dual-tree complex wavelets [25], irrespective of how the reconstruction quality is measured. Even with the learned tight frame, the proposed algorithm restores slightly less details than state-of-the-art algorithm [17] but is much faster (seconds vs. minutes).

A more detailed treatment of this topic is given in our paper [31], which describes the application of ADMM to the formulation with the Gaussian approximation of QCS and its extension to combined JPEG restoration and denoising.

ACKNOWLEDGMENT

Authors thank Pavel Rajmic from the Brno University of Technology for critical reading of the manuscript. This work was funded by the Czech Science Foundation, grant GA16-13830S.

APPENDIX

Lemma 1. Projection $P_{Ax\in\langle b_1,b_2\rangle}(z) = \arg\min_x \|x-z\|$, s.t. $Ax \in \langle b_1, b_2\rangle$, where , $b_1 \leq b_2$, $A \in R^{m\times n}$, $b_i \in R^m$, $m \leq n$, A full rank, AA^T diagonal, can be written as $P_{Ax\in\langle b_1,b_2\rangle}(z) = z - A^T(AA^T)^{-1}(Az - P_{\langle b_1,b_2\rangle}(Az))$, where $P_{\langle b_1,b_2\rangle}(y) = \min(\max(b_1, y), b_2)$.

Proof: By the Karush-Kuhn-Tucker conditions, $\lambda_+, \lambda_- \geq 0$ exist so that we can instead minimize

$$\frac{1}{2}\|x-z\|^2 + \lambda_+^T(Ax-b_2) + \lambda_-^T(-Ax+b_1). \tag{20}$$

We set its gradient equal to zero

$$x - z + A^T(\lambda_+ - \lambda_-) = 0. \tag{21}$$

From complementary slackness conditions

$$\lambda_+ > 0 \Rightarrow Ax = b_2, \tag{22}$$

$$\lambda_- > 0 \Rightarrow Ax = b_1. \tag{23}$$

For $\lambda_+ > 0$ $Ax \neq b_1$(since $b_1 \neq b_2$) and therefore (23) and non-negativity of λ_- implies $\lambda_- = 0$. Similarly $\lambda_- > 0 \Rightarrow \lambda_+ = 0$. The gradient condition (21) implies $x = z - A^T(\lambda_+ - \lambda_-)$, or in the other direction $\lambda_+ - \lambda_- = \left(AA^T\right)^{-1}(Az - Ax)$, which will be finally inserted into the former. There are three cases according to the position of Az with respect to the interval $\langle b_1, b_2\rangle$.

If $\lambda_+ = \lambda_- = 0$, (21) implies $x = z$, i.e. also $Ax = Az$ and from primal feasibility condition $Ax \in \langle b_1, b_2\rangle$ follows $Az \in \langle b_1, b_2\rangle$. If $\lambda_+ > 0$ (that is by (22) also $Ax = b_2$), $\lambda_+ - \lambda_- \geq 0$. Then (21) and (22) implies $b_2 = Ax = Az - AA^T(\lambda_+ - \lambda_-)$. Since AA^T is diagonal with positive entries (its diagonal contains the squared norms of rows of A), $AA^T(\lambda_+ - \lambda_-) > 0$ and therefore $b_2 < Az$. Similarly $\lambda_- > 0$ implies $b_1 > Az$. Since $b_1 < b_2$, these two cases are exclusive. Taken together with the case $\lambda_+ = \lambda_- = 0$, this defines the projection. Indeed, if $Az < b_1$, necessarily $\lambda_- > 0$ and therefore by (23) $Ax = b_1$. Analogously, if $Az > b_2$, necessarily $\lambda_+ > 0$ and by (22) $Ax = b_2$. Finally, if $Az \in \langle b_1, b_2\rangle$, $\lambda_+ = \lambda_- = 0$ and $Ax = Az$. ■

REFERENCES

[1] "JPEG file interchange format (JFIF)," ECMA TR/98, Ecma International, Tech. Rep., 2009.

[2] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *Image Processing, IEEE Transactions on*, vol. 4, no. 7, pp. 896–908, 1995.

[3] K. Bredies and M. Holler, "A total variation-based jpeg decompression model." *SIAM J. Imaging Sciences*, vol. 5, no. 1, pp. 366–393, 2012.

[4] ——, "A TGV-based framework for variational image decompression, zooming, and reconstruction. part i: Analytics," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2814–2850, 2015.

[5] ——, "A TGV-based framework for variational image decompression, zooming, and reconstruction. part ii: Numerics," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2851–2886, 2015.

[6] R. L. Stevenson, "Reduction of coding artifacts in transform image coding," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 5. IEEE, 1993, pp. 401–404.

[7] D. Sun and W.-K. Cham, "Postprocessing of low bit-rate block dct coded images based on a fields of experts prior," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2743–2751, Nov 2007.

[8] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.

[9] F. Alter, S. Durand, and J. Froment, "Adapted total variation for artifact free decompression of jpeg images," *J. Math. Imaging Vis.*, vol. 23, pp. 199–211, September 2005.

[10] Z. Xiong, M. Orchard, and Y.-Q. Zhang, "A deblocking algorithm for jpeg compressed images using overcomplete wavelet representations," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, no. 2, pp. 433–437, Apr 1997.

[11] C. Jung, L. Jiao, H. Qi, and T. Sun, "Image deblocking via sparse representation," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 663–677, 2012.

[12] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in JPEG decompression via a learned dictionary," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 718–728, 2014.

[13] C. Wang, J. Zhou, and S. Liu, "Adaptive non-local means filter for image deblocking," *Signal Processing: Image Communication*, vol. 28, no. 5, pp. 522–530, 2013.

[14] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *Image Processing, IEEE Transactions on*, vol. 22, no. 12, pp. 4613–4626, 2013.

[15] J. Zhang, S. Ma, Y. Zhang, and W. Gao, "Image deblocking using group-based sparse representation and quantization constraint prior," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 306–310.

[16] Y. Kwon, K. I. Kim, J. Tompkin, J. H. Kim, and C. Theobalt, "Efficient learning of image super-resolution and compression artifact removal with semi-local gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, pp. 1792–1805, 2015.

[17] X. Liu, X. Wu, J. Zhou, and D. Zhao, "Data-driven soft decoding of compressed images in dual transform-pixel domain," 2016.

[18] P. Svoboda, M. Hradis, D. Barina, and P. Zemcik, "Compression artifacts removal using convolutional neural networks," *arXiv preprint arXiv:1605.00366*, 2016.

[19] A. W.-C. Liew, H. Yan, and N.-F. Law, "POCS-based blocking artifacts suppression using a smoothness constraint set with explicit region modeling," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 6, pp. 795–800, 2005.

[20] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[21] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[22] J. Eckstein and D. P. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, pp. 293–318, June 1992.

[23] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[24] K. Bredies and M. Holler, "Artifact-free jpeg decompression with total generalized variation." in *VISAPP (1)*, G. Csurka and J. Braz, Eds. SciTePress, 2012, pp. 12–21.

[25] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234 – 253, 2001.

[26] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.

[27] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, 1st ed. Norwell, MA, USA: Kluwer Academic Publishers, 1992.

[28] P. G. Casazza, G. Kutyniok, and F. Philipp, "Introduction to finite frame theory," in *Finite Frames*. Springer, 2013, pp. 1–53.

[29] M. Sorel and F. Sroubek, "Fast convolutional sparse coding using matrix inversion lemma," *Digital Signal Processing*, vol. 55, pp. 44–51, 2016.

[30] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, 2013.

[31] M. Sorel and M. Bartos, "Fast bayesian JPEG decompression and denoising with tight frame priors," *IEEE Trans. Image Process.*, 2016.