RECURSIVE RISK ESTIMATION FOR NON-LINEAR IMAGE DECONVOLUTION WITH A WAVELET-DOMAIN SPARSITY CONSTRAINT

Cédric Vonesch, Sathish Ramani and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland

ABSTRACT

We propose a recursive data-driven risk-estimation method for non-linear iterative deconvolution. Our two main contributions are 1) a solution-domain risk-estimation approach that is applicable to non-linear restoration algorithms for ill-conditioned inverse problems; and 2) a risk estimate for a state-of-the-art iterative procedure, the thresholded Landweber iteration, which enforces a wavelet-domain sparsity constraint. Our method can be used to estimate the SNR improvement at every step of the algorithm; e.g., for stopping the iteration after the highest value is reached. It can also be applied to estimate the optimal threshold level for a given number of iterations.

Index Terms— Risk estimation, parameter adjustment, deconvolution, nonlinear, iterative, wavelets, sparsity.

1. INTRODUCTION

Deconvolution software plays an increasingly important role in various imaging applications, such as optical microscopy, medical imaging, astronomy, or satellite imaging. While there are many different deconvolution algorithms available [1, 2], virtually all of them depend on a set of parameters that must be properly adjusted to obtain satisfying results. Depending on the type of algorithm, one may have to tune various parameters such as regularization factor, step size, number of iterations, etc. Many end-users do not feel comfortable with this task, unless they have a good practical experience of the algorithm. As a result, they often resort to a subjective choice based on purely qualitative visual criteria. This reduces the reproducibility of the results and hinders comparison between data sets.

Consequently, the problem of parameter adjustment is of considerable importance for deconvolution-software manufacturers. There are various methods for automating this task based on quantitative criteria, such as Generalized Cross Validation, the L-curve method, the discrepancy principle (see for example Chapter 7 of [2]) or the Predicted MSE [3]. To our knowledge, however, these approaches have only been applied to linear deconvolution methods. Furthermore, they often rely on measurement-domain performance measures (except for [3] and our recent work [4]), which give only a partial account of the actual restoration quality.

Blu and Luisier [5] recently revitalized the principle of data-driven image restoration in the context of denoising, based on a risk estimate introduced by Stein [6]. In this paper, we use a similar estimate for ill-conditioned deconvolution problems. Specifically, we construct a recursive risk estimate—which depends only on the measurement—for a non-linear iterative deconvolution algorithm: the so-called "thresholded-Landweber" (TL) algorithm [7, 8, 9, 10]. This wavelet-based method belongs to the state-of-the-art in image deconvolution (we refer the reader to the experimental results presented in [7]). Nevertheless, the optimal result is typically obtained before the algorithm has fully converged, that is, for a particular number of iterations (see the example in Fig. 1). In such a situation, our method can be applied to monitor the (estimated) SNR improvement during the execution, so as to automatically stop the algorithm after the highest value has been reached. Our experiments show that it can also be used to adjust the threshold level.

2. THE PROBLEM OF DECONVOLUTION

In this paper, we consider the algebraic image-formation equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b},$$

where $\mathbf{H}$ is a block-circulant matrix that represents the effect of a convolutive image-formation device (e.g. a microscope, see [11]). $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is a vector that models the errors introduced by the measurement device (e.g. a CCD camera); we assume that its entries are realizations of IID Gaussian random variables of mean 0 and variance σ^2.

A deconvolution method can be seen as a function $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$ that returns an estimate of the original image $\mathbf{x}$ from the measurement $\mathbf{y}$ (N stands for the dimension of these vectors, that is, the number of pixels of the corresponding images). The design of a good deconvolution method is a difficult task, because the inversion of $\mathbf{H}$ is typically an ill-posed problem [1]. Most approaches are based on variational principles or on statistical assumptions whose discussion would go beyond the scope of this paper. The performance of a deconvolution method $\mathbf{f}$ is usually measured in terms of the

quadratic error (or "risk") $r = \|\mathbf{x} - \mathbf{f}(\mathbf{y})\|^2$, where $\|\cdot\|$ is the ℓ_2 norm. Of course, evaluating the risk is only possible in phantom experiments where the original image $\mathbf{x}$ is known. Such a quantity (which requires the knowledge of the ground-truth image) is often refered to as an "oracle" measure.

3. A GENERAL RISK-ESTIMATION APPROACH

In the sequel, we consider the case where the deconvolution algorithm depends on a parameter p; we thus use the notation $\mathbf{f}_p(\mathbf{y})$ instead of $\mathbf{f}(\mathbf{y})$. Ideally, the parameter p should be selected so as to minimize the corresponding risk

$$r_p = \|\mathbf{x} - \mathbf{f}_p(\mathbf{y})\|^2.$$

Since the ground-truth is obviously not available in real situations, our first goal is to derive an estimate of this quantity that depends only on the available information about $\mathbf{x}$, that is, the measurement $\mathbf{y}$. We have already constructed and successfully employed such a risk estimate in the case where $\mathbf{f}_p$ corresponds to a (linear) Wiener-type inverse filter with regularization parameter p [4]. Here, we consider a general (possibly non-linear) deconvolution algorithm. To obtain an estimate in this case, we rewrite the risk as

$$\begin{aligned} r_p &= \|\mathbf{x}\|^2 - 2\mathbf{x}^T\mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2 \\ &= \|\mathbf{x}\|^2 - 2(\mathbf{H}_{\text{inv}}\mathbf{H}\mathbf{x})^T\mathbf{f}_p(\mathbf{y}) \\ &\quad - 2([\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T\mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2 \\ &= \|\mathbf{x}\|^2 - 2\mathbf{y}^T\mathbf{g}_p(\mathbf{y}) + 2\mathbf{b}^T\mathbf{g}_p(\mathbf{y}) \\ &\quad - 2([\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T\mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2, \end{aligned}$$

where $\mathbf{H}_{\text{inv}}$ is a stabilized approximation of the inverse of $\mathbf{H}$ (see e.g. [1]) and $\mathbf{g}_p(\mathbf{y}) = \mathbf{H}_{\text{inv}}^T\mathbf{f}_p(\mathbf{y})$. There are three terms that are related to unknown quantities ($\mathbf{x}$ and $\mathbf{b}$) in the last expression. The first one, $\|\mathbf{x}\|^2$, is a constant with respect to p and is thus irrelevant for our purpose (we are only interested in *minimizing* the risk). The second one, $\mathbf{b}^T\mathbf{g}_p(\mathbf{y})$, can be estimated using a technique that is quite common in the field of statistics, which follows from the equality

$$\mathsf{E}[\mathbf{b}^T\mathbf{g}_p(\mathbf{y})] = \sigma^2\mathsf{E}[\operatorname{div}\mathbf{g}_p(\mathbf{y})],$$

where div is the divergence operator (see [6] or [5], which also state adequate conditions on $\mathbf{g}$). If $\mathbf{H}_{\text{inv}}\mathbf{H}$ is reasonably close to $\mathbf{I}$, then the third term $([\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T\mathbf{f}_p(\mathbf{y})$ is negligible.

Based on these considerations, we propose to estimate the risk using the quantity

$$e_p = c - 2\mathbf{y}^T\mathbf{g}_p(\mathbf{y}) + 2\sigma^2\operatorname{div}\mathbf{g}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2.$$

This quantity only depends on the measurement $\mathbf{y}$ (here, c represents an arbitrary constant, which does not influence the minimum of e_p with respect to p).

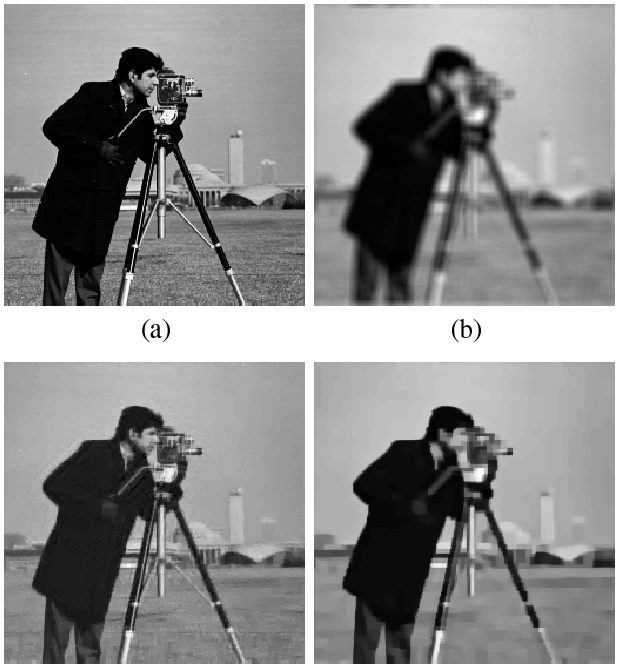

Fig. 1. (a) Original image. (b) Blurred and noisy image (9×9 uniform blur, Gaussian white noise with BSNR = 40 dB). (c) Restored image for the optimal number of iterations (16) and (d) after 300 iterations.

4. RECURSIVE RISK ESTIMATION

The general approach described in the previous section must be further specified in the case of an iterative method such as the TL algorithm. For simplicity, we consider the case of soft-thresholding in an orthonormal wavelet basis; however, the following ideas can readily be extended to a more general setting (biorthogonal or redundant decompositions; other thresholding functions).

When it is run until convergence, the classical Landweber iteration [1] induces considerable noise amplification for ill-conditioned inverse problems such as deconvolution. The TL algorithm can be understood as a way to compensate for this effect using wavelet-coefficient thresholding, a technique known for its effectiveness in image-denoising applications. Starting from some initialization $\mathbf{x}^{(0)}$, it produces a sequence of images $\mathbf{x}^{(k)}$ according to the update rule[1]

$$\mathbf{x}^{(k+1)} = \mathbf{W}\mathcal{T}_\theta\Big\{\mathbf{W}^T\big[\mathbf{x}^{(k)} + \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(k)})\big]\Big\}.$$

Here, $\mathbf{W}$ is an orthonormal wavelet reconstruction matrix (thus $\mathbf{W}^T$ is the corresponding decomposition matrix) and $\mathcal{T}_\theta\{\cdot\}$ denotes a component-wise application of the so-called "soft-thresholding" function (with threshold level θ):

[1] Assuming that $\mathbf{H}$ is adequately normalized—see [8].

$\mathcal{T}_\theta(w) = \text{sign}(w)\max(|w|-\theta, 0)$. This operation guarantees that the deconvolved images have a certain degree of sparsity in the wavelet domain (depending on the value of θ). This property makes the TL algorithm very attractive, because many natural images can be well approximated using only a few non-zero coefficients in a given wavelet basis.

Our goal is now to obtain a risk estimate for every iteration k of the TL algorithm. In this case, we can interpret the number of iterations as our parameter ($p = k$). The only difficulty for applying the estimate e_p lies in the computation of the divergence term. We propose to use the identity $\text{div}\ \mathbf{g}_p(\mathbf{y}) = \text{Tr}\ \mathbf{J}_{\mathbf{g}_p}(\mathbf{y})$, where Tr is the trace operator and $\mathbf{J}_{\mathbf{g}_p}(\mathbf{y})$ denotes the Jacobian matrix of $\mathbf{g}_p$, evaluated at $\mathbf{y}$. In our case, $\mathbf{J}_{\mathbf{g}_p}(\mathbf{y}) = \mathbf{H}_{\text{inv}}^T \mathbf{J}_{\mathbf{f}_p}(\mathbf{y})$; using a functional description of the TL algorithm, we obtain an explicit expression of this matrix. Let us define

$$\mathbf{t}(\mathbf{u}, \mathbf{v}) = \mathbf{W}\mathcal{T}_\theta\Big\{\mathbf{W}^T(\mathbf{I} - \mathbf{H}^T\mathbf{H})\mathbf{u} + \mathbf{W}^T\mathbf{H}^T\mathbf{v}\Big\}.$$

Then the result at iteration $p = k$ is given by $\mathbf{x}^{(k)} = \mathbf{f}_k(\mathbf{y})$, where the function $\mathbf{f}_k$ is defined recursively by

$$\begin{cases} \mathbf{f}_0 & \text{(initialization method);} \\ \mathbf{f}_{k+1} = \mathbf{t}(\mathbf{f}_k, \cdot) & \text{(update rule).} \end{cases}$$

This leads to a recursive relation between the Jacobian matrices of the functions $\mathbf{f}_k$:

$$\begin{aligned} \mathbf{J}_{\mathbf{f}_{k+1}}(\mathbf{y}) &= \mathbf{J}_{\mathbf{t}(\mathbf{f}_k, \cdot)}(\mathbf{y}) \\ &= \mathbf{J}_{\mathbf{t}(\mathbf{f}_k, \mathbf{y})}(\mathbf{y}) + \mathbf{J}_{\mathbf{t}(\mathbf{f}_k(\mathbf{y}), \cdot)}(\mathbf{y}) \\ &= \mathbf{J}_{\mathbf{t}(\cdot, \mathbf{y})}(\mathbf{f}_k(\mathbf{y}))\, \mathbf{J}_{\mathbf{f}_k}(\mathbf{y}) + \mathbf{J}_{\mathbf{t}(\mathbf{f}_k(\mathbf{y}), \cdot)}(\mathbf{y}) \\ &= \mathbf{W}\mathbf{D}_k\mathbf{W}^T([\mathbf{I} - \tau\mathbf{H}^T\mathbf{H}]\mathbf{J}_{\mathbf{f}_k}(\mathbf{y}) + \mathbf{H}^T). \end{aligned}$$

Here $\mathbf{D}_k = \mathbf{J}_{\mathcal{T}_\theta}(\mathbf{W}^T[\mathbf{I} - \mathbf{H}^T\mathbf{H}]\mathbf{f}_k(\mathbf{y}) + \mathbf{W}^T\mathbf{H}^T\mathbf{y})$ is a diagonal matrix, since thresholding is a pointwise operation.

This result can not be applied directly, however. The Jacobian matrices $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})$ are indeed extremely large for typical image sizes, so that it is unrealistic to compute and store them explicitly. Instead, we propose to exploit the following property: if $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{J}$, then $\mathsf{E}[\mathbf{n}^T\mathbf{J}\mathbf{n}] = \text{Tr}\ \mathbf{J}$ (where the expectation is taken over $\mathbf{n}$). In other words, given a normally distributed noise realization $\mathbf{n}$, we propose to compute the *vector* $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})\mathbf{n}$ rather than the *matrix* $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})$. This vector is updated recursively using the above recursion relation, which is computationally equivalent to performing one iteration of the TL algorithm. To estimate the divergence term at a given iteration k, we take the inner product with the filtered noise realization $\mathbf{H}_{\text{inv}}\mathbf{n}$. Thus, the risk estimate that we use in practice is

$$e'_p = c - 2(\mathbf{H}_{\text{inv}}\mathbf{y})^T\mathbf{f}_p(\mathbf{y}) + 2\sigma^2(\mathbf{H}_{\text{inv}}\mathbf{n})^T\mathbf{J}_{\mathbf{f}_p}(\mathbf{y})\mathbf{n} + \|\mathbf{f}_p(\mathbf{y})\|^2.$$

Typically, one realization of $\mathbf{n}$ is enough to yield an accurate estimate, since the number of samples N is relatively large for images.

5. PRACTICAL APPLICATIONS OF THE METHOD

We performed two phantom experiments to illustrate the potential applications of the method. These experiments were conducted under the following general conditions. The value of σ^2 was assumed to be known; in practical situations it can be estimated using a method similar to the one described in [4]. The TL algorithm was applied with four decomposition levels of the orthonormal Haar wavelet transformation. In addition to the description given in Section 4, we applied the random-shift strategy proposed in [7] which was found to yield the best deconvolution results (for non-redundant decompositions). Our risk-estimation method can be easily adapted to account for this additional processing step, without notably changing its computational complexity (which is of the same order as the TL algorithm itself). The initialization $\mathbf{x}^{(0)}$ was generated with the same method as in [7].

The estimate was computed using the simplest form of regularized inverse filter, $\mathbf{H}_{\text{inv}} = (\mathbf{H}^T\mathbf{H} + \epsilon\mathbf{I})^{-1}\mathbf{H}^T$; the constant ϵ was set to $\sigma^2/m_{\mathbf{y}}^2$, where $m_{\mathbf{y}}$ stands for the average (spatial) intensity of the measurement. For graphical-representation purposes, the constant c was chosen such that the risk and our estimate would coincide for the initialization $\mathbf{x}^{(0)}$.

5.1. Optimal number of iterations

In our first experiment, we convolved the standard Cameraman image with a 9×9 uniform-blur kernel and added noise to the result; this is illustrated in Fig. 1 for a BSNR of 40 dB. We considered the case of a threshold level that is suboptimal for a large number of iterations, which is representative of practical situations. We looked at the evolution of the true risk r_p and of our estimate e'_p as a function of the parameter $p = k$ (number of iterations). The results (signal-to-noise improvement in dB) are presented in Fig. 2; SNRI = $10\log_{10}(\|\mathbf{y} - \mathbf{x}\|^2/q_p)$, where q_p stands for one of the two previous quantities. The optimal value of the parameter according to the oracle and the estimate are indicated with a star and a circle, respectively.

It is seen that the curves of the true risk and of the estimate are in good agreement. In particular, the maximum of the esimate is sufficiently close to the maximum of the risk to allow for a prediction of the optimal number of iterations; such a prediction would yield the optimal number with an accuracy of about 0.1 dB in terms of SNRI. For BSNR levels of 40 and 30 dB, one would even obtain the optimal result (shown in Fig. 1 for a BSNR of 40 dB). In the 20 dB case, using the estimate would result in stopping the algorithm a few iterations before the optimum, which is favorable from a computational standpoint.

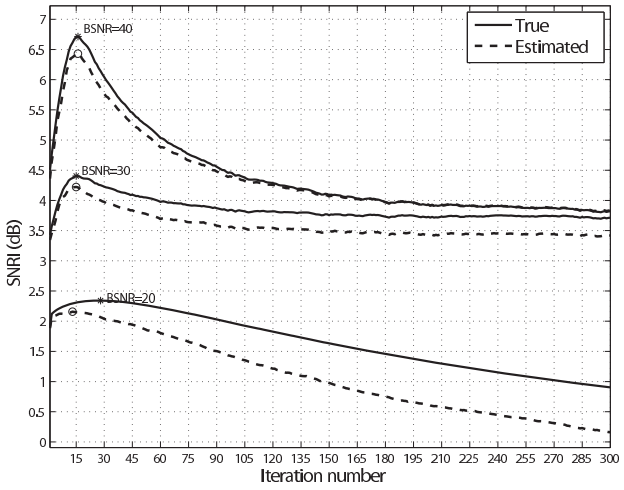

Fig. 2. Estimation of the SNR improvement as a function of the number of iterations, for different noise levels. In each case, the continuous line corresponds to the true value of the risk (oracle measure) and the dashed one to the estimated risk.

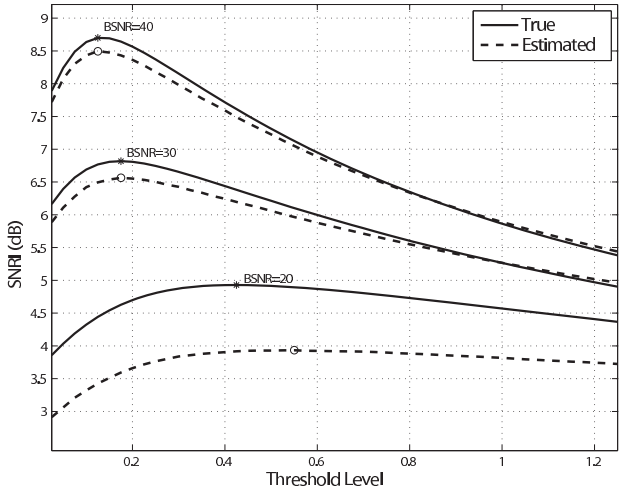

Fig. 3. Estimation of the SNR improvement as a function of the threshold level, for a fixed number of iterations.

5.2. Optimal threshold level

We applied a similar protocol for the parameter $p = \theta$ (threshold level) and a fixed number of iterations ($k = 50$). This time we used the MRI image and a Gaussian convolution kernel of width $\sigma_{\text{kernel}} = 2$ pixels.

The results are shown in Fig. 3. The curves of the true risk and of the estimate have the same shape. In particular, the maxima of the curves are again situated within a range of about 0.1 dB; this would allow for a prediction of the optimal threshold level based on the proposed risk estimate.

6. CONCLUSION

We have presented a promising data-driven method for estimating the risk during the execution of a recent non-linear iterative deconvolution algorithm: the thresholded Landweber iteration. Our results suggest that the method is suitable for estimating the optimal number of iterations, or for estimating the optimal threshold level. Future work will be concerned with a more extensive study of the accuracy of the method, and with the online estimation of the optimal threshold level (during the execution of the algorithm).

7. ACKNOWLEDGEMENTS

The authors would like to thank Thierry Blu for many stimulating discussions that led to this work. This work was funded in part by the Hasler Foundation and in part by the Swiss National Science Foundation under grant 200020-109415.

8. REFERENCES

[1] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*, Institute of Physics Publishing, 1998.

[2] C. Vogel, *Computational methods for inverse problems*, Society for Industrial and Applied Mathematics, 2002.

[3] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, July 1992.

[4] S. Ramani, C. Vonesch, and M. Unser, "Deconvolution of 3D fluorescence micrographs with automatic risk minimization," in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France*, May 14-17, 2008, pp. 732–735.

[5] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, November 2007.

[6] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, November 1981.

[7] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, August 2003.

[8] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, August 2004.

[9] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A ℓ^1-unified variational framework for image restoration," *Lecture Notes in Computer Science - Proceedings of ECCV 2004 Part IV*, vol. 3024, pp. 1–13, 2004.

[10] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539–549, April 2008.

[11] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 20–31, May 2006.