Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 607–610
COMPTES RENDUS
MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Theory of Signals

Estimating the probability law of the codelength as a function of the approximation error in image compression

François Malgouyres

LAGA/L2TI, université Paris 13, 99, avenue Jean-Baptiste-Clément, 93430 Villetaneuse, France

Received 23 October 2006; accepted after revision 6 March 2007

Available online 25 April 2007

Presented by Yves Meyer

Abstract

After recalling the subject of the compression of images using a projection onto a polyhedral set (which generalizes the compression by coordinate quantization), we express, in this framework, the probability that an image is coded with K coefficients as an explicit function of the approximation error. ***To cite this article: F. Malgouyres, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Résumé

Estimation de la loi de probabilité suivie par la longueur du code comme une fonction de l'erreur d'approximation, en compression d'images. Après des rappels sur la compression d'images par une projection sur un polyèdre, nous explicitons, dans ce cadre, la probabilité qu'une image soit codée par K coefficients, comme une fonction de l'erreur d'approximation. ***Pour citer cet article : F. Malgouyres, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

1. Introduction

In the past twenty years, many image processing tasks have been approached using two distinct mathematical tools: image decomposition in a basis and optimization.

The first mathematical approach has proved very useful and is supported by solid theoretical foundations: these guarantee its efficiency, as long as the basis captures the information contained in images. Modelling the image content by appropriate function spaces (of infinite dimension), the mathematical theory tells us how the coordinates of an image, in a given basis, behave. For example, it is possible to characterize Besov spaces (see [12]) and the space of bounded variation (which is 'almost characterized' in [3]) with wavelet coefficients. As a consequence of these characterizations, one can obtain performance estimates for practical algorithms (see Th. 9.6, p. 386, in [11] and [5,4] for some analyses in more complex situations). Image compression and restoration are the typical applications where such analyses are meaningful.

E-mail address: malgouy@math.univ-paris13.fr.
URL: http://www.math.univ-paris13.fr/~malgouy/.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.03.007

The optimization methods which have been applied to solve those practical problems have also proved very efficient (see [13], for a very famous example). However, the theory is not able to assess how well they perform, given an image model.

Interestingly, many in the community who were primarily involved in the image decomposition approach are now focusing on optimization models (see, for instance, the work on Basis Pursuit [2] or compressed sensing [6]). The main reason for this is probably that optimization provides a more general framework [1,7,8].

The framework which seems to allow both a good flexibility for practical applications (see [2] and other papers on Basis Pursuit) and good properties for theoretical analysis is the method of projection onto polyhedra or polytopes. For theoretical studies, it shares simple geometrical properties with the usual image decomposition models (see [10]); this should allow the derivation of approximation results.

The aim of this Note is to state a theorem[1] which relates, asymptotically as the precision grows, the approximation error and the number of coefficients which are coded (which we abusively call codelength, for simplicity). More precisely, when the initial datum is assumed random in a convex set, we give the probability for the datum to be coded by K coefficients, as a function of the approximation error (see Theorem 3.1 for details).

This result is given in a framework which generalizes the usual coding of the quantized coefficients ('non-linear approximation'), as usually performed by compression standards (for instance, JPEG and JPEG2000).

2. Recollection on variational compression

Here, and throughout the Note, N is a positive integer, $I = \{1, \ldots, N\}$ and $\mathcal{B} = (\psi_i)_{i \in I}$ is a basis of $\mathbb{R}^N$. We will also denote, for $\tau > 0$ (throughout the paper τ stands for a positive real number) and for all $k \in \mathbb{Z}$, $\tau_k = \tau(k - \frac{1}{2})$.

For any $(k_i)_{i \in I} \in \mathbb{Z}^N$, we set

$$\mathcal{C}\big((k_i)_{i \in I}\big) = \left\{ \sum_{i \in I} u_i \psi_i, \forall i \in I, \tau_{k_i} \leqslant u_i \leqslant \tau_{k_i+1} \right\}. \tag{1}$$

We then consider the optimization problem

$$(\tilde{P})\big((k_i)_{i \in I}\big): \quad \begin{cases} \text{minimize } f(v) \\ \text{under the constraint } v \in \mathcal{C}\big((k_i)_{i \in I}\big), \end{cases}$$

where f is a norm which is continuously differentiable away from 0 and has strictly convex level sets. In order to state Theorem 3.1, we also need f to be *curved*. This means that the inverse of the homeomorphism h below[2] is Lipschitz.

$$h : \big\{u \in \mathbb{R}^N,\ f(u) = 1\big\} \to \big\{g \in \mathbb{R}^N, \|g\|_2 = 1\big\},$$
$$u \mapsto \frac{\nabla f(u)}{\|\nabla f(u)\|_2}.$$

(The notation $\|\cdot\|_2$ refers to the euclidean norm in $\mathbb{R}^N$.)

We denote, for any $(k_i)_{i \in I} \in \mathbb{Z}^N$,

$$\tilde{J}\big((k_i)_{i \in I}\big) = \{i \in I, u_i^* = \tau_{k_i} \text{ or } u_i^* = \tau_{k_i+1}\},$$

where $u^* = \sum_{i \in I} u_i^* \psi_i$ is the solution to $(\tilde{P})((k_i)_{i \in I})$.

The interest in these optimization problems comes from the fact that, as explained in [8], we can recover $(k_i)_{i \in I}$ from the knowledge of $(\tilde{J}, (u_i^*)_{j \in \tilde{J}})$ (where $\tilde{J} = \tilde{J}((k_i)_{i \in I})$).

The problem $(\tilde{P})$ can therefore be used for compression. Given a datum $u = \sum_{i \in I} u_i \psi_i \in \mathbb{R}^N$, we consider the unique $(k_i(u))_{i \in I} \in \mathbb{Z}^N$ such that (for instance)

$$\forall i \in I, \quad \tau_{k_i(u)} \leqslant u_i < \tau_{k_i(u)+1}. \tag{2}$$

[1] The theorem concerning compression in [10] is incorrect. The situation turns out to be more complex than we thought at the time that [10] was written.

[2] We prove in [10] that, under the above hypotheses, h actually is an homeomorphism.

The information $(\tilde{J}, (u_i^*)_{j\in\tilde{J}})$, where $\tilde{J} = \tilde{J}((k_i(u))_{i\in I})$, is then used to encode u. In the following, we denote the set of indexes that need to be coded to describe u by $\tilde{J}(u) = \tilde{J}((k_i(u))_{i\in I})$.

Notice that we can also show (see [8]) that the coding performed by the standard image processing compression algorithms (JPEG and JPEG2000) corresponds to the above model when, for instance,

$$f\left(\sum_{i\in I} u_i\psi_i\right) = \left(\sum_{i\in I} |u_i|^2\right)^{1/2}.$$

Observe that the above compression scheme works for any quantization table (see [8]); we restrict to the uniform quantization because Theorem 3.1 only applies in this context. However, several levels of generalization are possible, if one wants to generalize it to more general quantization tables. Notice that, in the theorem, we assume that the data belong to a given level set, denoted $\mathcal{L}_{f_d}(\tau')$, of a norm f_d. Therefore, the code attributed to each coefficient need not to be infinite.

3. The estimate

Theorem 3.1. *Let $\tau' > 0$ and U be a random variable whose low is uniform in $\mathcal{L}_{f_d}(\tau')$, for a norm f_d. Assume f satisfies the hypotheses given in Section 2. For any norm $\|\cdot\|$ and any $K \in \{1, \dots, N\}$ there exists D_K such that for all $\varepsilon > 0$, there exists $T > 0$ such that for all $\tau < T$*

$$\mathbb{P}\big(\#\tilde{J}(U) = K\big) \leqslant D_K E^{(N-K)/(N+1)} + \varepsilon,$$

where E is the approximation error:[3]

$$E = \mathbb{E}\left(\left\|U - \tau\sum_{i\in I} k_i(U)\psi_i\right\|\right).$$

Moreover, if $f(\sum_{i\in I} u_i\psi_i) = (\sum_{i\in I} |u_i|^2)^{1/2}$, we also have[4]

$$\mathbb{P}\big(\#\tilde{J}(U) = K\big) \geqslant D_K E^{(N-K)/(N+1)} - \varepsilon.$$

The proof of the above theorem is given in [9]. Its two main steps are: the characterization of all the $(k_i)_{i\in I} \in \mathcal{L}_{f_d}(\tau')$ which are coded with K coefficients, for any given $K \in \{1, \dots, N\}$; the census, for each K, of $(k_i)_{i\in I}$ obtained at the first step.

The above theorem differs from the results evoked in Section 1 in several ways.

First, it concerns variational models which are more general than the model for which the results of Section 1 are usually stated. This is probably the main interest of the current result. For instance, by a reasoning similar to the one used in the proof of Theorem 3.1, it is probably possible to obtain approximation results for redundant transforms.

Secondly, it expresses the distribution of the number of coefficients as a function of the approximation error, whereas earlier results do the opposite. Typically, they bound the approximation error (quantified by the L^2 norm) by a function of the number of coefficients that are coded. The advantages and drawbacks of the different kinds of statements are not very clear. In the framework of Theorem 3.1, the larger D_K (for K small), the better the model compresses the data. However, it is clear that, as the approximation error goes to 0, it is more and more likely to obtain a code of size N. In this respect, the constant D_{N-1} seems to play an important role, since it dominates (asymptotically as τ goes to 0) the probability not to obtain a code of length N.

Thirdly, the theorem is stated for data leaving in a finite dimension vector space and, as a consequence, it does not impose a priori links between the data distribution (the function f_d) and the model (the function f and the basis $\mathcal{B}$). The ability of the model to represent the data is always assessed by the C_K. Of course, an analogue of Theorem 3.1 for data in an infinite dimension space would be interesting.

[3] When computing the approximation error, we consider the center of $\mathcal{C}((k_i)_{i\in I})$ has been chosen to represent all the elements u such that $(k_i(u))_{i\in I} = (k_i)_{i\in I}$.

[4] This assumption is very pessimistic. For instance, the lower bound seems to hold for almost every basis $\mathcal{B}$ of $\mathbb{R}^N$, when f is fixed. However, we have not worked out the details of the proof of such a statement.

References

[1] A. Chambolle, R.A. De Vore, N. Lee, B.J. Lucier, Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage, IEEE Transactions on Image Processing 7 (3) (1998) 319–355. Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.
[2] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1999) 33–61.
[3] A. Cohen, W. Dahmen, I. Daubechies, R. De Vore, Harmonic analysis of the space bv, Revista Matematica Iberoamericana 19 (2003) 235–262.
[4] A. Cohen, I. Daubechies, O.G. Guleryus, M.T. Orchard, On the importance of combining wavelet-based nonlinear approximation with coding strategies, IEEE Transactions on Information Theory 48 (7) (July 2002) 1895–1921.
[5] A. Cohen, R. De Vore, P. Petrushev, H. Xu, Nonlinear approximation and the space bv, American Journal of Mathematics 121 (3) (June 1999) 587–628.
[6] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (April 2006) 1289–1306.
[7] F. Malgouyres, Minimizing the total variation under a general convex constraint for image restoration, IEEE Transactions on Image Processing 11 (12) (December 2002) 1450–1456.
[8] F. Malgouyres, Image compression through a projection onto a polyhedral set, Technical Report 2004-22, University Paris 13, August 2004.
[9] F. Malgouyres, Estimating the probability law of the codelength as a function of the approximation error in image compression, Technical Report ccsd-00108319, CCSD, October 2006.
[10] F. Malgouyres, Projecting onto a polytope simplifies data distributions, Technical Report 2006-1, University Paris 13, January 2006.
[11] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, Boston, 1998.
[12] Y. Meyer, Ondelettes et opérateurs, vol. 1, Hermann, 1990.
[13] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.