

Sparse Coding by Bayesian Variational Marginalization

Koray Kavukcuoglu¹, Marc’Aurelio Ranzato², Rob Fergus¹, Yann LeCun¹

¹The Courant Institute of Mathematical Sciences - New York University

²Department of Computer Science - University of Toronto

koray@cs.nyu.edu, ranzato@cs.toronto.edu, fergus@cs.nyu.edu, yann@cs.nyu.edu

Sparse coding algorithms are shown to produce good models for natural images [5, 1, 3]. An input signal $x \in \mathcal{R}^m$ is represented using a linear combination of features which are columns of the dictionary matrix $\mathcal{D} \in \mathcal{R}^{m \times n}$, using coefficients $z \in \mathcal{R}^n$, with $n > m$. Many sparse coding algorithms have been proposed in the literature and in this work we focus on the following convex formulation:

$$\min_z \frac{1}{2} \|x - \mathcal{D}z\|_2^2 + \lambda \sum_i |z_i| \quad (1)$$

This particular formulation has been extensively studied and it has also been extended to the case when the dictionary \mathcal{D} is learnt, thus adapting to the statistics of the input. Dictionaries learned using sparse coding models contain edge filters at different locations, orientations and scales resembling gabor functions. The model that is learned by sparse coding models is local patch based model of images, therefore a significant portion of the dictionary is used for modeling a particular filter at many locations. However, for many practical applications, these models are applied on large images, thus producing highly redundant representations. In order to avoid this problem a convolutional sparse coding model can be used to learn a dictionary that does not encode translations. This is effectively equivalent to enforcing a repeating toeplitz structure for the dictionary in equation 1.

Moreover, in many computer vision applications, one is interested in extracting a *feature set* for a given image and sparse representations are shown to be useful features in object recognition tasks [3, 2, 6]. However, an important problem is stability. It is desirable for a representation to be stable under small perturbations of the input. Sparse representations as produced by equation 1 are not stable, because the component of z that best explains a given input, suppresses all other similar components. This behaviour may cause abrupt changes in the representation for very similar inputs. Sample inputs and corresponding representations are given in figure 1, which are obtained by using a convolutional extension to the coordinate descent method given in [4]. It can be seen that, generally the input is explained using only a handful of components. In order to avoid such a behaviour, rather than minimizing the sparse coding energy function itself, given in eq. (1), we propose to minimize the variational free energy corresponding to the sparse coding energy:

$$F_\beta(x; \mathcal{D}) = -\frac{1}{\beta} \log \int_z e^{-\beta E(x,z;\mathcal{D})} \quad (2)$$

One can find an upper bound to this formulation using an approximating distribution $Q(z)$ which can be assumed to be a gaussian with mean m and diagonal covariance S (s.t $\sigma_i^2 = s_i = S_{ii}, i = 1..d$). After necessary simplifications, the final form reduces to:

$$F_\beta(x; \mathcal{D}) \leq \int_z Q(z) \|X - Dz\|_2^2 + \lambda \int_z Q(z) \|z\|_1 - \frac{1}{\beta} H(Q(z)) \quad (3)$$

$$F_\beta(x; \mathcal{D}) \leq \|X - Dm\|_2^2 + \mathbf{Tr} [D^T DS] + \lambda \sum_i \left[\sqrt{\frac{2}{\pi}} \sigma_i e^{-m_i^2/2\sigma_i^2} + m_i \left[1 - 2\Phi\left(\frac{-m_i}{\sigma_i}\right) \right] \right] - \frac{1}{2\beta} \ln \left[(2\pi e)^d \prod_i s_i \right] \quad (4)$$

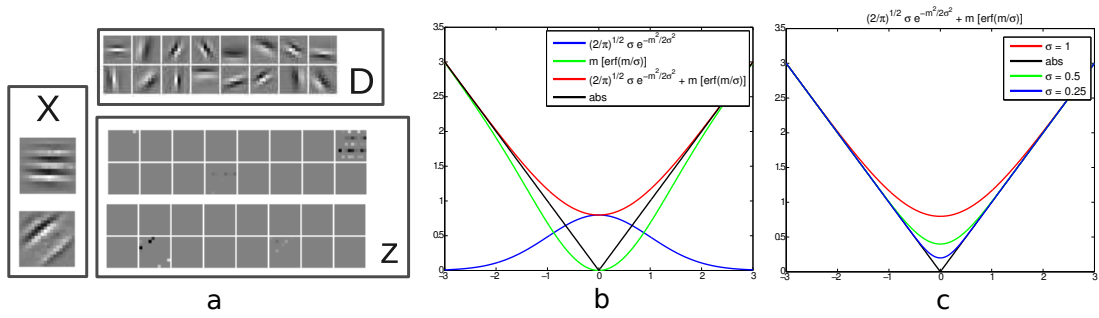


Figure 1: **(a)** Two sample image parts (20×20 pixels) are represented using the dictionary \mathcal{D} shown on the top. Each element of the dictionary is 9×9 . Corresponding representations z (12×12) are shown below. (Mid gray level corresponds to zero) **(b)** The regularization term in eq. (4) is shown for $\sigma = 1$ together with absolute value function. **(c)** The regularization term is shown for different values of σ .

where Φ is the cdf of standard normal distribution. After solving for the mean and standard deviation of the approximating distribution $Q(z)$, the mean values can be taken as the solution. The complicated sum in the third term of equation 4 is derived from the second term in equation 3 which corresponds to the sparsity regularization penalty. It can be seen from figure 1 that this term corresponds to a smoothed version of the harsh L1 penalty function. We show that the representations obtained by minimizing the variational free energy are more stable and smooth compared the representations obtained by minimizing the sparse coding energy.

References

- [1] M. Aharon, M. Elad, and A. M. Bruckstein. K-SVD and its non-negative variant for dictionary design. In M. Papadakis, A. F. Laine, and M. A. Unser, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5914 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 327–339, August 2005.
- [2] Kevin Jarrett, Koray Kavukcuoglu, Marc’Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? In *Proc. International Conference on Computer Vision (ICCV’09)*. IEEE, 2009.
- [3] Koray Kavukcuoglu, Marc’Aurelio Ranzato, and Yann LeCun. Fast inference in sparse coding algorithms with applications to object recognition. Technical report, Computational and Biological Learning Lab, Courant Institute, NYU, 2008. Tech Report CBLL-TR-2008-12-01.
- [4] Y. Li and S. Osher. Coordinate Descent Optimization for ℓ^1 Minimization with Application to Compressed Sensing: a Greedy Algorithm. *Inverse Problems and Imaging*, 3(3):487–503, 2009.
- [5] B A Olshausen and D J Field. Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.
- [6] Jianchao Yang, Kai Yu, Yihong Gong, and Thomas Huang. Linear spatial pyramid matching using sparse coding for image classification. In *CVPR*, 2009.