Nonlinear Transform Coding: Polar Coordinates Revisited

Demba E. Ba and Vivek K Goyal

Massachusetts Institute of Technology

Transform coding is central to standards for audio, image and video coding. In conventional transform coding, the encoder first performs a linear transformation, followed by *uniform scalar quantization* and scalar entropy coding. With scalar quantization, to minimize the distortion at high rate, the best one can hope for is a quantizer that partitions into hypercubes; and with scalar entropy coding, one desires independent transform coefficients to minimize the rate. Transform coding has a clean analytical theory for Gaussian sources because these two virtues can be achieved simultaneously. However, for other sources, ideal cell shape and independence of transform coefficients may be competing, and in fact the optimal transform will generally give neither. The integer-to-integer (i2i) transform coding framework introduced in [1] suggests a way to deal with this conflict. In linear i2i transform coding, the source is first quantized, which allows for more degrees of freedom in designing an approximately-linear transform that—for certain sources—will give approximately-independent components. (The transform step is merely a reordering of a discrete probability mass function.) In this work, we go much further than in [1] by removing approximate linearity as a design constraint.

If cell shape were not an issue, one would simply try to obtain independent transform coefficients. For many sources, polar coordinates (or spherical coordinates for higher dimensions) will give independent coefficients or coefficients that are closer to independent than any linear coordinate change. Examples include certain Gaussian scale mixtures (popular for modeling image transform coefficients [2]) and distributions arising from the conditional uncertainty in tracking problems with range-only sensing modalities [3].

We designed a family of i2i approximations to the Cartesian-to-polar transformation and analyzed its behavior for high-rate transform coding. Denoting (ordinary, continuous) polar coordinates by (r, θ), our precise high-rate analysis relates the performance to the differential entropies of r^2 and θ, which are often easy to evaluate. One may thus predict when there is an improvement over linear transform coding. The analysis matches our simulations for coding of Gaussian scale mixtures and other polar-separable sources. The advantage over the best linear transform coder can be large.

Our hope is to extend the polar-coordinate results to a general theory for nonlinear transform coding based on i2i implementations of arbitrary nonlinear transformations.

[1] V. K Goyal, "Transform coding with integer-to-integer transforms," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 465–473, Mar. 2000.

[2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[3] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *IEEE J. Sel. Areas Comm.*, vol. 23, no. 4, pp. 809–819, Apr. 2005.