Derivatives and Inverse of a Linear-Nonlinear Multi-Layer Spatial Vision Model

B. Galán*[1], M. Martinez-Garcia[1,2], P. Cyriac[3], T. Batard[3], M. Bertalmio[3] and J. Malo[1]

[1]Image Processing Lab. Parc Científic, Universitat de València, Spain
[2]Instituto de Neurociencias, CSIC, Spain
[3]Dept. Tecnol. Inf. Comunic., Universitat Pompeu Fabra, Spain

Motivation: The association of a linear transform and a saturating nonlinear transform is ubiquitous in vision models, either psychophysical [1–3] or physiological [4,5]. Therefore, an appropriate mathematical formulation of these linear-nonlinear units is extremely convenient.

The consideration of the linear-nonlinear units as multidimensional transforms in vector spaces allows to understand the visual system in the same (abstract/geometrical) terms used in current machine learning. This abstract formulation also has relevant consequences in vision science experimentation. For example, the *derivatives of the model* (or Jacobian matrices) determine the discrimination regions (JNDs) and the subjective metric in the stimulus space [6,7]. Point-dependence of such matrices would be related to masking/adaptation. Applications involving the optimization of subjective distortion will critically depend on the Jacobian (e.g. to implement gradient descent). This includes engineering applications, but also psychophysical techniques such as Maximum Differentiation (MAD), in which the stimuli are designed to maximize/minimize the perceived distortion [7–9]. Note also that the Jacobian describes the deformation of the response space. This deformation determines the change of the redundancy in the signal [10, 11] and this will have an impact in interpretations of the visual function from the efficient coding perspective. On the other hand, the *inverse of the model* is key to define the features the system is tuned to. Applications of the inverse range from designing stimuli that isolate the response of a specific sensor, to methods to recover the input from a set of responses [12].

Given the above, the aim of this work is providing a full report of the the d-dimensional formulation of an illustrative model consisting on a cascade of standard linear-nonlinear units. The interest of these analytic results and the required numerical tricks transcend the particular model because of the ubiquity of this architecture.

The linear-nonlinear multi-layer model: The model considered for this illustration was originally intended to provide a psychophysically meaningful alternative to the *modular concept* in Structural Similarity measures (SSIM) [13]. Its authors suggest a separate consideration of luminance, contrast and structure (which is a sensible approach), but the definition of such factors has no obvious perceptual meaning in SSIM. The idea for a *more perceptual* alternative in [7] was addressing one psychophysical factor at a time (i.e. modular), by using a cascade of linear-nonlinear transforms. Here we extend [7] by considering 4 layers: (1) linear spectral integration and nonlinear brightness response, (2) definition of local contrast by using linear filters and divisive normalization, (3) linear CSF filter and nonlinear local contrast masking, and (4) linear wavelet-like decomposition and nonlinear divisive normalization to account for orientation and scale-dependent masking. On top of its interpretability, the modular structure simplifies the use of MAD to set the free parameters by determining only one layer at a time.

Reproducible results [1] We list all the analytic expressions for (1) every linear-nonlinear layer, (2) the Jacobian matrices, (3) the inverses, and (4) the gradient of the perceptual distance. When the analytic result is not available or a naive implementation is not feasible we discuss the tricks to face these problems. We provide a `Matlab` toolbox to apply the presented theory to actual image patches. The analytic Jacobian matrices are numerically checked using finite differences. Conditions for the existence of the inverses are given, and it is straightforward to check the convergence of the inversion methods using the toolbox. In summary, the formulation and toolbox are ready to explore the issues addressed in the introductory section (giving all the information that was missing in [7]).

Conclusion We studied the mathematical issues found in perceptually meaningful linear-nonlinear units (e.g. huge kernels and divisive normalization). The proposed analytic and numeric solutions may be applicable (or inspiring) in similar models, having impact in psychophysics and in statistical/geometrical interpretation of models.

[1] M.D. Fairchild. *Color appearance models*. Wiley, 2013.

[2] D. Burr and P. Thompson. Motion psychophysics: 1985–2010. *Vision research*, 51(13):1431–1456, 2011.

[3] A. B. Watson and J. A. Solomon. Model of visual contrast gain control and pattern masking. *JOSA A*, 14(9):2379–2391, 1997.

[4] M. Carandini and D. J Heeger. Normalization as a canonical neural computation. *Nature Rev. Neurosci.*, 13(1):51–62, 2012.

[5] K. N. Kendrick, J. Winawer, A. Rokem, A. Mezer, and . A Wandell. A two-stage cascade model of bold responses in human visual cortex. *PLoS Comput Biol*, 9(5):e1003079, 2013.

[6] V. Laparra, J. Muñoz-Marí, and J. Malo. Divisive normalization image quality metric revisited. *JOSA A*, 27(4):852–864, 2010.

[7] J. Malo and E. P. Somincelli. Geometrical and statistical properties of vision models obtained via MAD. SPIE, 2015.

[8] Z. Wang and E. P Simoncelli. Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):8–8, 2008.

[9] J. Malo, D. Kane, and M. Bertalmio. The MAD competition depends on the viewing conditions. Accept. VSS meeting, 2016.

[10] S. Lyu and E. P Simoncelli. Nonlinear extraction of independent components of natural images using radial gaussianization. *Neural Comp.*, 21(6):1485–1519, 2009.

[11] V. Laparra, G. Camps-Valls, and J. Malo. Iterative gaussianization: from ica to random rotations. *IEEE Trans. Neu. Net.*, 22(4):537–549, 2011.

[12] J. Malo, Y. Kamitani, N. Kriegeskorte, A. Lazar, K.N. Kay, S. Nishimoto, and O. Marr. What is in your mind?: inversion problems in vision science. Simposium ECVP-16. http://www.ub.edu/ecvp/symposia#sim2, 2016.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Im. Proc.*, 13(4):600–612, 2004.

*Correspondence: borja.galan@uv.es, jesus.malo@uv.es

[1]Supplementary material (Formulae, code and illustrations) available at: http://isp.uv.es/docs/MODVIS16linearnonlinear.zip