What Is the Limit of Redundancy Reduction with Divisive Normalization?

Fabian H. Sinz
fabian.sinz@epagoge.de
Institute for Neurobiology, Department for Neuroethology, Eberhard
Karls University Tübingen, 72076 Tübingen, Germany

Matthias Bethge
Werner Reichardt Centre for Integrative Neuroscience; Bernstein
Center for Computational Neuroscience,
and Institute for Theoretical Physics, Eberhard Karls University
Tübingen, 72074 Tübingen, Germany

Divisive normalization has been proposed as a nonlinear redundancy reduction mechanism capturing contrast correlations. Its basic function is a radial rescaling of the population response. Because of the saturation of divisive normalization, however, it is impossible to achieve a fully independent representation. In this letter, we derive an analytical upper bound on the inevitable residual redundancy of any saturating radial rescaling mechanism.

1 Introduction

In divisive normalization the activity y_i of a neuron is normalized by the activities of other neurons,

$$y_i \mapsto \frac{\kappa y_i}{\left(\sigma^2 + \sum_j |y_j|^p\right)^{\frac{1}{p}}} \quad \Leftrightarrow \quad \mathbf{y} \mapsto \frac{\kappa \mathbf{y}}{\left(\sigma^2 + \|\mathbf{y}\|_p^p\right)^{\frac{1}{p}}}. \tag{1.1}$$

This simple but elegant mechanism is so apt in capturing the behavior of neurons throughout the brain that it has rececently been termed *canonical computation* (Carandini & Heeger, 2011). One possible computational goal of this widespread mechanism could be the reduction of redundancies among neural responses to natural signals in accordance with Barlow's idea of redundancy exploitation (Barlow, 1961, 2002) as demonstrated in the seminal paper by Schwartz and Simoncelli (2001).

The redundancy reduction achieved with divisive normalization models takes place after the removal of second-order correlations via linear filtering. For natural images, it has previously been demonstrated that linear filters cannot remove all redundancies and have little influence on the reduction of higher-order dependencies (Bethge, 2006; Eichhorn, Sinz, & Bethge,

2009). However, it has also been demonstrated that nonlinear rescalings $\mathbf{y} \mapsto g(\|\mathbf{y}\|_p) \cdot \|\mathbf{y}\|_p^{-1} \cdot \mathbf{y}$ of the norm, as in divisive normalization, can substantially reduce higher-order correlations. For the class of L_p-spherically symmetric distributions, which provide a good fit to the joint filter responses $\mathbf{y}$ on natural images, there is even a unique and optimal radial rescaling mechanism called *radial factorization* (for general p) or *radial gaussianization* (for $p = 2$) (Sinz & Bethge, 2009, 2013; Lyu & Simoncelli, 2009; Sinz, Simoncelli, & Bethge, 2009). The basic underlying mechanism is to map the radial distribution $\varrho(\|\mathbf{y}\|_p)$ of the filter reponses $\mathbf{y}$ into the radial distribution $\varrho_{\mathcal{N}}$ of a p-generalized normal distribution with independent marginals via $g(\|\mathbf{y}\|_p) = \mathcal{F}_{\varrho_{\mathcal{N}}}^{-1}(\mathcal{F}_{\varrho}(\|\mathbf{y}\|_p))$, where $\mathcal{F}$ denotes the cumulative distribution function of the respective probability density (Sinz, Gerwinn, & Bethge, 2009; see also Figure 1a). Since these distributions have infinite support, full redundancy reduction can be achieved through radial rescaling mechanism only if it does not saturate but maps onto the entire positive real axis (Lyu & Simoncelli, 2009). Most divisive normalization mechanisms as well as real neurons, however, do saturate for large $\|\mathbf{y}\|_p$. Thus, an important issue is how critical this principal limitation of saturating radial rescaling mechanisms is.

Here, we show that the inevitable redundancy caused by saturation is relatively small and approaches zero in the limit of many dimensions. More specifically, we investigate the multi-information

$$I[\mathbf{Z}] = \sum_i h[Z_i] - h[\mathbf{Z}] \tag{1.2}$$

of a radially truncated p-generalized normal. Here, $h[Z_i]$ and $h[\mathbf{Z}]$ denote the marginal and the joint differential Shannon entropy, respectively. The truncation threshold will be denoted by κ, as it is this parameter in the divisive normalization model (see equation 1.1) that determines the saturation threshold. We show numerically that the multi-information rate is a decreasing function of κ and then address the two limiting cases $\kappa \to 0$ and $\kappa \to \infty$ analytically. The limit of the latter case is simply the p-generalized normal without truncation and thus has independent marginals (Sinz, Gerwinn, & Bethge, 2009). For $p \geq 1$, the limiting case $\kappa \to 0$ is the multi-information of the uniform distribution within the L_p-unit sphere because $\lim_{\mathbf{x} \to 0} \nabla_{\mathbf{x}} \mathcal{N}_p(\mathbf{x}) = \mathbf{0}$ for the p-generalized normal distribution $N_p(\mathbf{x})$. The main contribution of this letter is an analytic expression for this case, which provides a useful upper bound on the minimal multi-information that can be achieved with a saturating radial rescaling mechanism.

2 Analytical Results

In order to compute the multi-information of the uniform distribution within the L_p-unit ball $\mathcal{B}_p$, we first compute the entropy of its univariate
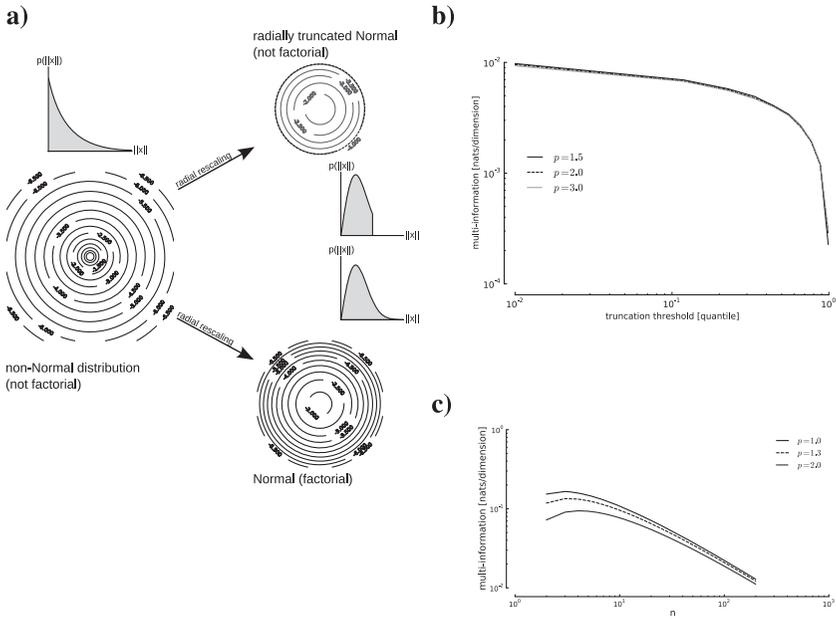
Figure 1: (a) Redundancy reduction via radial rescaling (radial factorization/ gaussianization). The input distribution is a nonfactorial L_p-spherically symmetric distribution (here, $p = 2$). Contour lines show iso-likelihood contours; insets show the corresponding radial distribution on $\|\mathbf{x}\|_p$. If the radii are rescaled such that they follow a χ distribution, the resulting joint distribution becomes a factorial p-generalized normal. If the radial rescaling function saturates, the resulting joint distribution is radially truncated and cannot be fully factorial since the p-generalized gaussian is the only factorial L_p distribution (Sinz, Gerwinn, & Bethge, 2009). (b) Multi-information in nats/dimension for $n = 100$ dimensions, different values of p, and varying truncation thresholds (saturation levels). The truncation thresholds are determined by quantiles of the respective radial distributions. For small truncation thresholds, the radially truncated p-generalized normal distribution becomes a uniform distribution within the L_p-unit ball. The multi-information is monotonic in the truncation threshold. Therefore, the multi-information bounds the multi-information of a p-generalized normal from above. (c) Multi-information of the uniform distribution within the L_p-unit ball as a function of n in nats/dimension.

marginal. The marginal densities of the uniform distribution belong to the family of β-distributions (Sinz & Bethge, 2010). The expression for the marginal results from solving the integral in the general formula for the marginal distribution of L_p-spherically symmetric distributions (Gupta &

Song, 1997),

$$\rho_i(x_i) = \frac{p\Gamma\left(\frac{n}{p}\right)}{2\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p}\right)} \int_{|x_i|}^1 \frac{\left(r^p - |x_i|^p\right)^{\frac{n-1}{p}-1}}{r^{n-p}} nr^{n-1} dr$$

$$= \frac{p\Gamma\left(\frac{n}{p}\right) n}{2\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p}\right)(n-1)} \left(1 - x_i^p\right)^{\frac{n-1}{p}},$$

where Γ denotes the gamma function.

From that, we can compute the marginal entropy

$$h[X_i] = \frac{p\Gamma\left(\frac{n}{p}\right) n}{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p}\right)(n+p-1)^2}$$

$$\times {}_3F_2\left(1 - \frac{1}{p}, \frac{n}{p} - \frac{1}{p} + 1, \frac{n}{p} - \frac{1}{p} + 1; \frac{n}{p} - \frac{1}{p} + 2, \frac{n}{p} - \frac{1}{p} + 2; 1\right)$$

$$+ \log\left(\frac{2\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p}\right)(n-1)}{p\Gamma\left(\frac{n}{p}\right) n}\right),$$ (2.1)

where ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z)$ denotes the generalized hypergeometric function.[1]

The joint entropy of the uniform distribution within the L_p-unit ball $\mathcal{B}_p$ is computed via its volume V,

$$h[X] = -\int_{\mathcal{B}_p} \frac{1}{V} \log\left(\frac{1}{V}\right) dx$$

$$= \int_{\mathcal{B}_p} \frac{1}{V} \log(V) dx$$

$$= \log(V),$$ (2.2)

---

[1]The integral can be computed by using the substitution $u = 1 - x^p$, the fact that the marginal is symmetric around zero, and the identities

$$\left(\frac{1}{\alpha_1} z \frac{d}{dz} + 1\right) f(z) = \frac{1}{\alpha_1} z^{1-\alpha_1} \frac{d}{dz}\left(z^{\alpha_1} f(z)\right)$$

$$\left(\frac{1}{\alpha_1} z \frac{d}{dz} + 1\right) {}_2F_1\left(\alpha_1, \alpha_2; \alpha_1 + 1 \middle| z\right) = (1-z)^{-\alpha_2}$$

$$\frac{d}{dz}\left(z^{\alpha_1} \cdot {}_3F_2\left(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2 \middle| z\right)\right) = \alpha_1 z^{\alpha_1 - 1} \cdot \left(z^{\alpha_1} \cdot {}_3F_2\left(\alpha_1 + 1, \alpha_2, \alpha_3; \beta_1, \beta_2 \middle| z\right)\right).$$

with $V = \frac{2^n \Gamma^n\left(\frac{1}{p}\right)}{n\Gamma\left(\frac{n}{p}\right)p^{n-1}}$ (Gupta & Song, 1997).

The marginal entropy $h[X_i]$ given by equation 2.1 and the joint entropy $h[\mathbf{X}]$ given by equation 2.2 together determine the multi-information $I[\mathbf{X}] = \sum_i h[X_i] - h[\mathbf{X}] = n\,h[X_i] - h[\mathbf{X}]$ of the uniform distribution within the L_p-unit ball.

3 Discussion

After whitening, natural image patches can be well modeled by L_p-spherically symmetric distributions (Sinz & Bethge, 2009). These can be transformed into factorial distribution by a nonlinear radial rescaling similar to divisive normalization (see Figure 1a). Since the radial rescaling of divisive normalization saturates at κ, it cannot achieve full redundancy reduction. Numerical computations show that the multi-information rate of a radially truncated p-generalized normal distribution is monotonicly decreasing with the truncation threshold (see Figure 1b). Therefore, the limiting case $\kappa \to 0$ provides an upper bound on the information rate for arbitrary radially truncated p-generalized normal distributions. This upper bound is given by the multi-information rate of the uniform distribution within the L_p-unit ball that we derived here. It turns out that the upper bound is quite low compared to a lower bound on the multi-information rate of natural images of ≈ 2.26 nats/pixel (Hosseini, Sinz, & Bethge, 2010; see also Figure 1c). This means that the dependencies due to radial truncation are negligible compared to the dependencies present in unnormalized natural images. Therefore, the multi-information of the uniform distribution on the L_p-unit ball can serve as a meaningful lower bound on the redundancy reduction that radial rescaling mechanisms should be able to achieve at least.

Acknowledgments

References

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Barlow, H. B. (2002). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*(4), 602–607.

Bethge, M. (2006). Factorial coding of natural images: How effective are linear models in removing higher-order dependencies? *Journal of the Optical Society of America A*, *23*(6), 1253–1268.

Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.

Eichhorn, J., Sinz, F., & Bethge, M. (2009). Natural image coding in V1: How much use is orientation selectivity? *PLoS Comput Biol*, *5*(4).

Gupta, A. K., & Song, D. (1997). Lp-norm spherical distribution. *Journal of Statistical Planning and Inference*, *60*(2), 241–260.

Hosseini, R., Sinz, F., & Bethge, M. (2010). Lower bounds on the redundancy of natural images. *Vision Research*, *50*(22), 2213–2222.

Lyu, S., & Simoncelli, E. P. (2009). Nonlinear extraction of independent components of natural images using radial gaussianization. *Neural Computation*, *21*(6), 1485–1519.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.*, *4*(8), 819–825.

Sinz, F., & Bethge, M. (2009). The conjoint effect of divisive normalization and orientation selectivity on redundancy reduction. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems, 21* (pp. 1521–1528). Red Hook, NY: Curran.

Sinz, F., & Bethge, M. (2010). Lp-nested symmetric distributions. *Journal of Machine Learning Research*, *11*, 3409–3451.

Sinz, F., & Bethge, M. (2013). Temporal adaptation enhances efficient contrast gain control on natural images. *PLoS Computational Biology*, *9*(1), e1002889.

Sinz, F., Gerwinn, S., & Bethge, M. (2009). Characterization of the p-generalized normal distribution. *Journal of Multivariate Analysis*, *100*, 817–820.

Sinz, F., Simoncelli, E. P., & Bethge, M. (2009). Hierarchical modeling of local image features through Lp-nested symmetric distributions. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems, 22* (pp. 1696–1704). Red Hook, NY: Curran.