Max–Planck–Institut für biologische Kybernetik
Max Planck Institute for Biological Cybernetics

Technical Report No. 169

How much can Orientation Selectivity and Contrast Gain Control Reduce Redundancies in Natural Images.

Fabian Sinz,[1] Matthias Bethge,[1]

March 2008

This Technical Report has been approved by:

Director at MPIK

Postdoc at MPIK

Max–Planck–Institut für biologische Kybernetik
Max Planck Institute for Biological Cybernetics

Technical Report No. 169

How much can Orientation Selectivity and Contrast Gain Control Reduce Redundancies in Natural Images.

Fabian Sinz,[1] Matthias Bethge,[1]

March 2008

[1] Department Schölkopf, email: fabee,mbethge@tuebingen.mpg.de

This report is available in PDF–format via anonymous ftp at ftp://ftp.kyb.tuebingen.mpg.de/pub/mpi-memos/pdf/Sinz_Bethge_2008.pdf. The complete series of Technical Reports is documented at: http://www.kyb.tuebingen.mpg.de/techreports.html

How much can Orientation Selectivity and Contrast Gain Control Reduce Redundancies in Natural Images.

Fabian Sinz and Matthias Bethge

Abstract. The two most prominent features of early visual processing are orientation selective filtering and contrast gain control. While the effect of orientation selectivity can be assessed within in a linear model, contrast gain control is an inherently nonlinear computation. Here we employ the class of L_p elliptically contoured distributions to investigate the extent to which the two features, orientation selectivity and contrast gain control, are suited to model the statistics of natural images. Within this model we find that contrast gain control can play a significant role for the removal of redundancies in natural images. Orientation selectivity, in contrast, has only a very limited potential for linear redundancy reduction.

---

1 Introduction

Both orientation selectivity and contrast gain control have been suggested to serve the reduction of redundancy in neural responses to natural images. In order to *quantitatively* verify *how much* such mechanisms can actually reduce the redundancy in natural images it is necessary to investigate their effects in terms of a statistical model.

The feature of orientation selectivity can be related to a linear filter bank model for which the different filter shapes are chosen to optimally reduce the dependencies among different filter outputs. In fact, orientation selectivity is a highly robust feature of such optimal filter shapes, reproducibly found if linear independent component analysis (ICA) is applied to ensembles of natural image patches [13; 2].

The effect underlying the redundancy reduction achieved with orientation selective filtering is based on an increase in peakedness of the histograms over the filter responses. In other words, orientation selective signal decompositions can increase the "sparsity" with which the filter respond without discounting any information.

The linear filtering framework used in ICA, however, is not powerful enough to achieve a representation with truly independent components. In particular, it has been shown that co-variations in the variance of different filters is an ubiquitous finding [14]. In addition, it has been shown *qualitatively* that a contrast gain mechanism can be used to remove these variance dependencies to a large extent [14].

In this study we would like to investigate the two mechanisms of orientation selective filtering and contrast gain control in a common framework. In particular, we would like to determine *quantitatively*, how much each mechanisms can be useful for redundancy reduction when they interact conjointly.

In previous studies [3; 6] we have shown that the effect of orientation selectivity alone has only a very limited potential for redundancy reduction as the advantage of the optimal ICA solution turned out to be less than 4% relative to a entirely random decorrelation transformation in the case of color as well as gray level images. Moreover, we have demonstrated that after whitening, a simple spherical symmetric distribution with a radial Gamma distribution with solely two parameters fits the ensemble of natural image patches significantly better than the factorial ICA model with $\frac{1}{2}n(n-1)$ parameters (n being the number of dimensions). Since the marginals of this spherically symmetric distribution are not independent, this finding is a striking evidence against the independence assumption on the linear filter responses imposed by ICA. Another interesting conclusion can be drawn from the fact that spherical distributions are invariant against orthogonal transformations. In the framework of any spherically symmetric distribution, in particular the one used by [6], all bases are equivalent. Particulary, the basis found with ICA is as good as any other random basis within the spherical symmetric model.

In order to use the spherical symmetric model for the purpose of redundancy reduction, a nonlinear transformation is required. In fact, for this model independent marginals can be achieved by using a contrast gain control mechanism: If we denote the cumulative distribution function (cdf) of the radial distribution by $F(r)$ then inde-

pendent marginals can be achieved by the transformation

$$\mathbf{z} = \frac{G^{-1}(F(r))}{r}\,\mathbf{x}$$

where $r = ||\mathbf{x}||_2$ and G is the cdf of the χ-distribution (i.e. the radial distribution of a spherical symmetric Gaussian). The function $G^{-1}(F(r))/r$ can be interpreted as a gain control mechanism. Since $r = ||\mathbf{x}||_2 = ||Q\mathbf{x}||_2$ is invariant against arbitrary rotations Q the choice of the basis is irrelevant. Consequently, there is no role for orientation selectivity in this simple gain control model. By comparing its performance to the ICA model which exploits orientation selectivity but no contrast gain control we can conclude that contrast gain control is much more effective for redundancy reduction than orientation selectivity.

If we wanted to incorporate both, spherical symmetry and marginal independence in a distribution of whitened natural images, the only distribution serving those two goals would be the Gaussian (see e.g. [11; 1]). Yet, it is a well known observation that the marginal distributions of natural images in the whitened space look strikingly non Gaussian as the coefficient histograms are much more peaked at zero (see Figure 2A). Therefore, when dropping the requirement of rotational invariance, the ICA model may seem to be a good candidate for fitting the statistics of natural images as it allows one to model a factorial distribution with arbitrarily shaped marginal distributions. Indeed, the ICA model favors non-Gaussian marginals along its coordinate axes. However, it would not predict the marginal distributions to be non Gaussian in any arbitrary direction. In particular, if we look at the marginal distribution which is obtained by summing over all different ICA coefficients, one would expect to find an approximately Gaussian distribution whenever the conditions of the central limit theorem are fulfilled. This is not the case for natural images (see Figure 1A). In addition, we can verify that the central limit theorem holds if we destroy the statistical dependencies between the coefficients of the different components by random shuffling (see Figure 1B). This clearly shows that the marginals of natural images in the ICA basis are not independent.

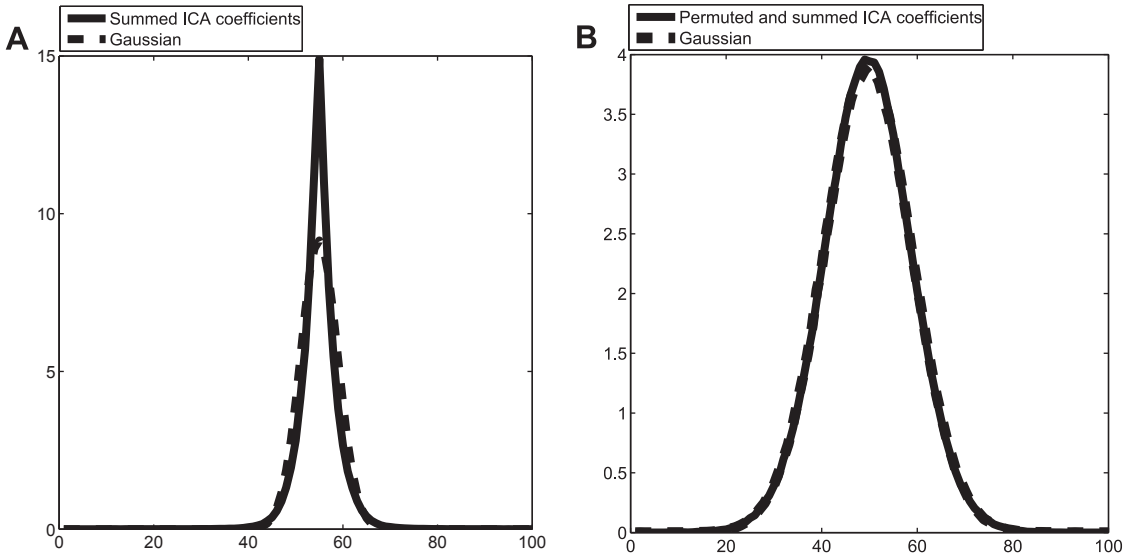

Figure 1: **Lack of Independence between the ICA Marginals** **A.** Histogram over the sum of coefficients in the ICA basis and over a corresponing Gaussian. If the coefficients of the single ICA basis vectors were really independent, the histogram should look similar to the one of a Gaussian random variable due to the central limit theorem. **B.** Central Limit theorem for natural images. Histogram of the summed coefficients for natural images in the ICA basis, after the sample index of the coefficients has been shuffled separately for each dimension. By this random permutation all statistical dependencies between the coefficients are destroyed. Since the resulting histogram is almost indistinguishable from a Gaussian density function, we can conclude that the missing Gaussianity for the sum over the ICA coefficients of natural images shown in the left image is due to the lack of statistical independence.

In summary, the empirical findings are:

(i) For natural images, the independence assumption of the ICA model is significantly violated as the distribution over the sum over the ICA coefficients does not approximate a Normal distribution while it does if statistical dependencies are destroyed via shuffling (see Figure 1). In other words, linear filtering is not powerful enough to remove all redundancies in natural images.

(ii) A spherical symmetric distribution yields a better fit in the whitened space than the factorial ICA model [6] (see Figure 2). Nevertheless, the convergence of ICA to orientation selective filters is a robust phenomenon and the resulting marginal distribution exhibit an increased kurtosis in comparison to the marginals of any arbitrary direction. Ideally, a model distribution should be able to take this kind of symmetry breaking into account.

In this paper we study a model which allows one to determine the effect of both orientation selectivity and contrast gain control: The class of L_p-spherically symmetric distributions (see section 2.3) includes as special cases both the L_2-spherical symmetric distribution considered above as well as the ICA model for which the marginals are all fitted by the same distribution within the exponential power family. In addition, it also generalizes to the case where both mechanisms are used conjointly for the removal of statistical dependencies.

To specify the model parameters, we first fit the whitened coefficients of natural images to the class of L_p-spherical distributions. In this way we determine the optimal p-norm for a selection of different bases. Subsequently, it is easy to determine the transformation which yields independent marginals for the specified L_p-spherical distribution in the ideal case. In a similar vein to the special case discussed above when $p = 2$, we can factorize any L_p-spherical distribution $\rho(\mathbf{y}) = \rho_u(\mathbf{y}/r_p)\rho_r(r_p)$ into a uniform distribution over the L_p-sphere and an arbitrary distribution over the p-norm $r_p = ||\mathbf{y}||_p = \left(\sum |y_k|^p\right)^{1/p}$ (see Appendix 5.1). In contrast to the case of $p = 2$, the L_p-sphere is not rotationally invariant and we need to distinguish between different data distributions that are obtained by $\mathbf{y} = V\,\mathbf{x}$ with $VV^\top = I$ being an orthogonal transform. We will see below that for any p-spherical distribution there is a scaling transformation

$$\mathbf{z} = h(r_p)\,\mathbf{y}$$

such that the resulting representation has independent components. This non-linear step in our cascade can be interpreted as a generalization of the contrast gain control mechanism for the L_2-spherical symmetric case. The case $p = 2$ is distinct in being invariant under arbitrary orthogonal transformations. In contrast, for arbitrary $p \neq 2$ we only have the weaker symmetry of permutational invariance among the coordinate axis. This symmetry breaking for $p \neq 2$ opens the possibility for orientation selectivity to play a role for redundancy reduction in combination with contrast gain control. We will use this combined model to evaluate the contribution of both mechanisms quantitatively.

2 Model

In this section, we describe the model and introduce the necessary mathematical tools. Rigorous versions of the formal statements in this section will be treated in the Appendix. For a rough overview, we start with natural image data obtained by randomly sampling 7×7 patches from photographic color images. Thus, each data point is a $3 \times 49 = 147$ dimensional vector whose components specify the pixel intensities of the corresponding patch.

Using a suitable linear transformation, it is possible to represent the image patches with "whitened" coefficients which all have the same variance and zero pairwise correlations. In fact, many different whitening transformations are possible. If C_{pixel} denotes the covariance matrix between the pixel intensities, then $C_{\text{pixel}}^{-1/2}$ defines the *symmetric* whitening transformation which stays as close as possible to the Pixel basis (in Frobenius norm) as possible. Here, our analysis starts after this whitening has already taken place. That is, we are considering an ensemble of N image patches described by vectors $\mathbf{x}_k \in \mathbb{R}^{147}, k = 1, \ldots, N$ each of which describing the coefficients of the image patch with respect to the symmetric whitening basis.

The details of the whitening procedure are described in the Appendix 5.1. What should be mentioned again though is that the three DC components defining the spatially constant components within the three different color channels are modeled separately because of their extraordinary statistical properties. That is, we can decompose each vector into a direct sum of two parts $\mathbf{x} = \mathbf{x}^{\text{DC}} \oplus \mathbf{x}^{\text{AC}}$ with $\mathbf{x}^{\text{DC}} \in \mathbb{R}^3$ describing the DC components and $\mathbf{x}^{\text{AC}} \in \mathbb{R}^{144}$ describing the spatially structured AC components.

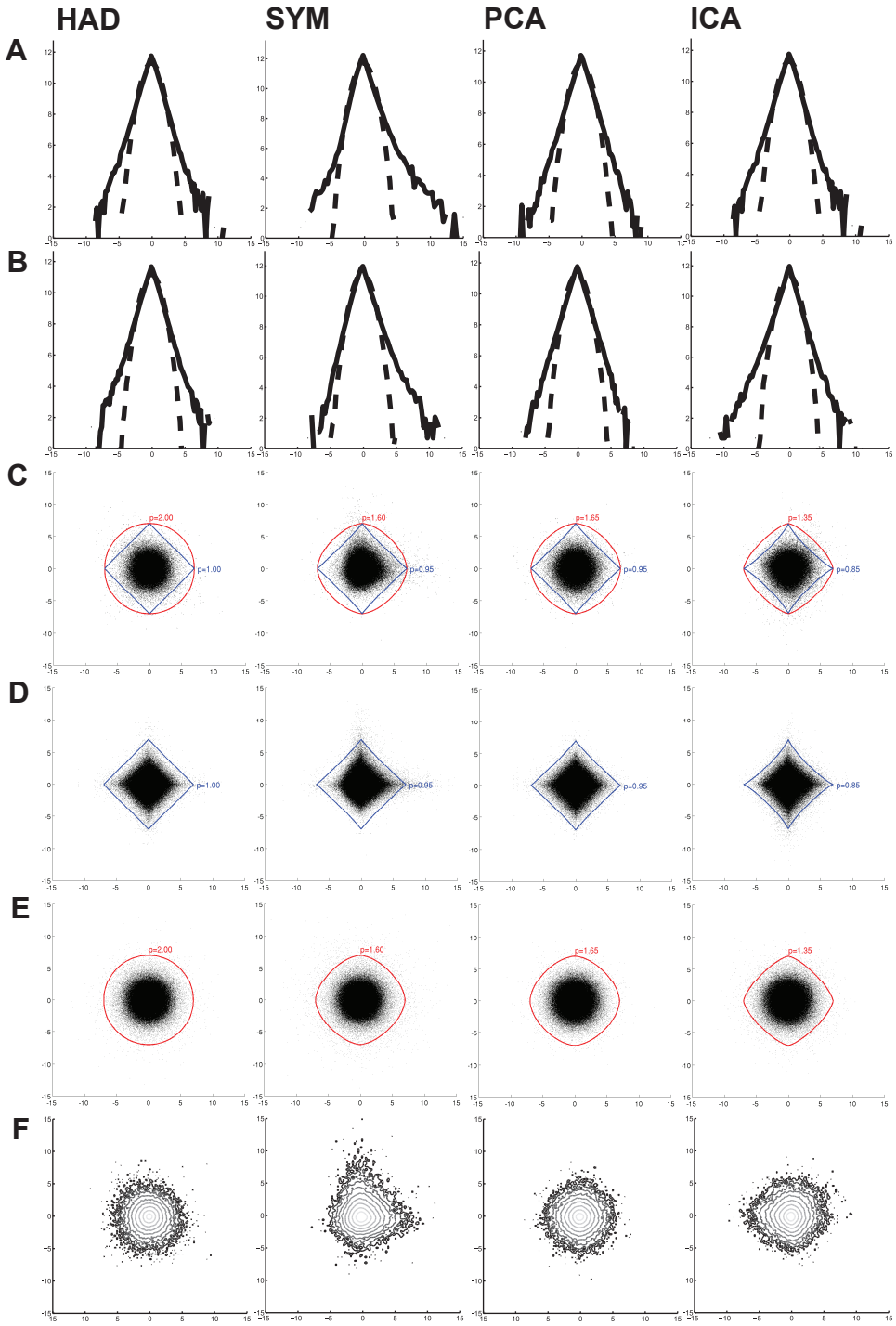

Figure 2: **Marginal Histograms for Different Transforms** **A-B** show show the log marginal histograms on whitened 7×7 image patches for randomly chosen basis elementes of transformations indicated at the top of each column. The dashed line corresponds to a normal distribution. One can nicely see that the marginal distribution of the coefficients is considerably more peaked and has heavier tails than a Normal distribution. **C** shows a scatter plot of the corresponding coefficients. **D-E** show scatter plots where the single coefficients have been permutated among the examples or the data points have been randomly rotated, respectively. If the coefficients where independent, **C** and **D** should look exactly the same. If the data was rotationally symmetric, **C** and **E** should look the same. However, the difference between the scatter plots is more dramatic between the former, suggesting to drop the independence assumption and to use a spherically symmetric distribution to model the data. The red and the blue line depict one contour line of the best fitting L_p-spherically symmetric distribution with radial gamma component or independent exponential power distributed marginals. Since the blue contour does not resemble the shape of the distribution at all, it seems apprropriate to drop the linear independence constraint. By looking at the red contours one can see that rotational symmetry is also not exatly fulfilled. However the deviation between the shapes of the distribution seems less severe in that case. **F** shows a contour plots of the log joint histograms for the two dimensions used above.

The fitting of the L_p-spherical distribution that we will describe in the following only concerns the 144-dimensional subspace of AC components. In order to avoid too much clutter in the notation, we will write $\mathbf{x}$ instead of $\mathbf{x}^{\text{AC}}$ in the following while we will distinguish between $\mathbf{x}$ and $\mathbf{x}^{\text{AC}}$ in the notation only if their is an increased risk of confusion. As another convention, we use uppercase letters to denote several images patches collected as column vectors in a matrix $X = (\mathbf{x}_1, ..., \mathbf{x}_m)$.

Fitting the L_p-spherical distribution to the ensemble X can then be separated into two steps:

1. First, we can apply different orthogonal transforms $\mathbf{y} = V\mathbf{x}$. For instance, V can be chosen in order to maximize the non-Gaussanity of the marginal distributions as it is the case for ICA. This is the part where orientation selectivity enters the model by a specific choice of the transform. Since all whitening transforms only differ by an orthogonal transform in whitened space each choice of basis determines one transform. We describe all transforms used in this report in more detail in section 2.1.

2. Direction-sensitive rescaling $\mathbf{z} = h(||\mathbf{y}||_p)\,\mathbf{y}$ where $f : (0, \infty) \to (0, \infty)$ is a positive, invertible and differentiable function which is chosen such that the resulting variable $\mathbf{z}$ has independent marginals. This part of the model corresponds to the contrast gain control mechanism. The details are described in section 2.3.

Summarizing the two steps in an single equation, the model can be written as

$$p(\mathbf{z}) \quad = \quad \prod_{k=1}^{n} p_k \left(h(||V\mathbf{y}||_p) \cdot (V\mathbf{y})_k\right) \tag{1}$$

with p_k being the density of the univariate marginal distributions.

2.1 Orientation Selectivity and Whitening

Since we have assumed that $E[\mathbf{x}] = 0$ and $E[\mathbf{x}\mathbf{x}^\top] = I$ the covariance matrix of the transformed variable $\mathbf{y} = V\,\mathbf{x}$ also equals the identity matrix as $E[\mathbf{y}\mathbf{y}^\top] = VV^\top = I$ if V is orthogonal. In the following we will introduce a selection of four different choices for V.

SYM $\mathbf{x}$ by default contains the coefficients for the symmetric whitening basis. Therefore, we refer to the choice where $V = I$ is chosen to be the identity matrix as "symmetric whitening". From a biological perspective, this case is of particular interest as it resembles the filter properties of retinal Ganglion cells.

wPCA Let $C_{\text{pixel}} = UDU^\top$ be the eigenvalue decomposition of C_{pixel}. By choosing $V_{\text{wPCA}} = U^\top$ we obtain the coefficients for the basis which is aligned with the principal axis of the pixel intensities. As this basis is usually determined via principal component analysis (PCA) we refer to this transform as *whitening PCA*. From a technical perspective this case is particularly interesting as the PCA basis is optimal in terms of coding efficiency [6; 16].

ICA In Independent Component Analysis (ICA), a transformation V_{ICA} is determined which maximizes the non-Gaussanity of the marginal distributions. For natural image patches, ICA is known to yield orientation selective filters in resemblance to V1 simple cells. While other orientation selective basis are possible, the filters defined by V_{ICA} correspond to the optimal choice for the purpose of redundancy reduction.

HAD Finally we consider the case $V_{\text{HAD}} = \frac{1}{\sqrt{m}} H V_{\text{ICA}}$ with H denoting an arbitrary Hadamard matrix. Hadamard matrices are defined by the two properties $H_{ij} = \pm 1$ and $HH^\top = mI$. Therefore, the resulting marginal distributions will all correspond to a sum over the different ICA coefficients modulo sign flips. In a certain sense, this case can be seen as the opposite extreme to the case of ICA. Instead of running an independent search for the most Gaussian marginals we rather make use of the central limit theorem here: If the independence assumption underlying ICA applies, we expect to find the most Gaussian components by using the Hadamard transformation which mixes all ICA coefficients with equal weight.

By comparing the effect of the different choices for V on the average log-likelihood, we can evaluate the potential role of orientation selectivity for redundancy reduction in a quantitative way.

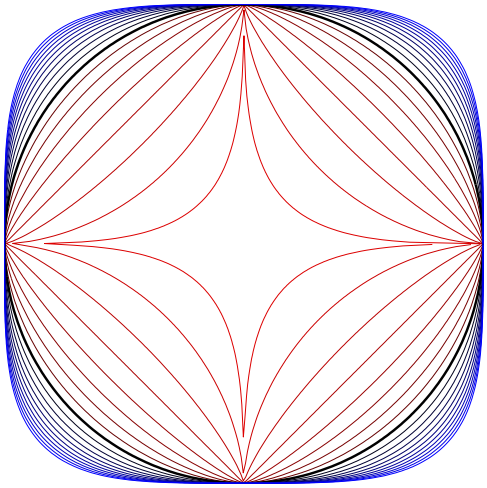

Figure 3: **Contour Lines of p-Spheres** The figure shows several unit p-spheres. The red contour lines depict p-spheres with $0 < p < 2$, the blue lines correspond to p-spheres with $p > 2$. The black line shows the standard Euclidean sphere.

2.2 Constrast-Gain Control

According to the classical contrast normalization model [8] the normalization of the filter responses is described by

$$h(r) = \frac{1}{c + r}.$$

where r denotes the sum over the squared responses of a subset of filters. So far, it has been shown only qualitatively that contrast gain control can remove variance correlations among pairs of filter responses [14]. In the following we will show how contrast gain control can systematically be exploited for the purpose of redundancy reduction.

Following [10] we use the L_p-spherically symmetric distribution [7] to model the distribution of coefficients in the whitened space. In difference to [10], we show here how this distribution can be used for the design of a contrast gain control mechanism and compare *how much* it can contribute to the reduction of redundancies relative to orientation selectivity.

By definition (see [7] and Definition 5.6), an L_p-spherically symmetric distributed variable Y is a product $\mathsf{Y} = \mathsf{U} \cdot \mathsf{R}$ of a random variable U, which is uniformly distributed on the unit p-sphere and a non-negative random variable R with an arbitrary distribution. Intuitively speaking, U specifies the direction, while R accounts for the radius. This implies that one can change from one L_p-spherically symmetric distribution to another one with identical p by appropriate transformation of the radial component R.

The unit p-sphere is defined as the set of points $\{\mathbf{y} \in \mathbb{R}^m : ||\mathbf{y}||_p = 1\}$ for which the *p-norm*

$$||\mathbf{y}||_p \equiv \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}$$

equals one[1]. By construction, the contour lines of the L_p-spherically symmetric distribution are all rescaled versions of the unit p-sphere. For illustration, several examples of contour lines for different choices of p are depicted in Figure 3.

The case $p = 2$ is the only case for which the distribution is completely isotropic (i.e. invariant under arbitrary orthogonal transformations). For all other cases with $p \neq 2$ the rotational symmetry is broken and reduced to a permutational symmetry. While classical contrast normalization models are based on the 2-norm, the symmetry breaking with respect to rotations is a necessary requirement for orientation selectivity to have an effect on redundancy reduction.

[1] Note, that for $p < 1$ the p-norm is not a norm in the strict sense. However, this does not matter for our purposes and, as a convention, we will further use the term for all $p > 0$.

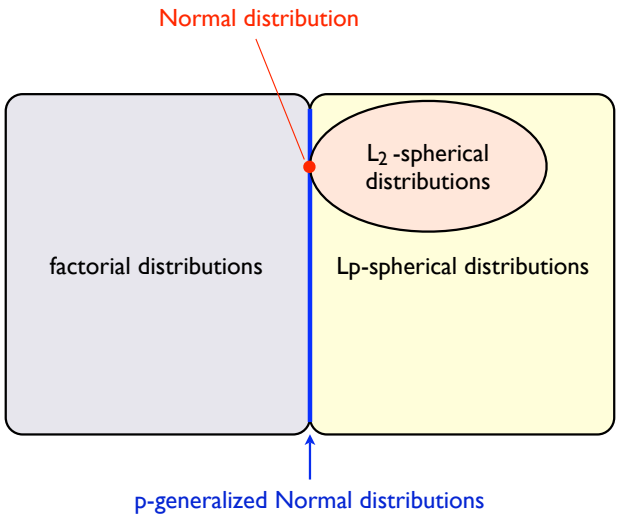

Figure 4: **Properties of the p-generalized Normal distribution**
The Gaussian is the only L_2-spherically symmetric distribution with independent marginals. Like the Gaussian all p-generalized Normal distributions have independent marginals and the property of spherical symmetry is a special case of the L_p-spherical symmetry in this class. We suspect that the p-generalized Normal distributions are the only distributions which combine these two properties simultaneously but this still needs to be proven.

In order to make use of the L_p-spherical distribution for the purpose of contrast gain control, we will use a L_p-generalized version of the Gaussian, called the p-generalized Normal distribution $\mathcal{N}_p$. We suspect, that this distribution plays the same special role for L_p-spherically symmetric distributions as the Normal distribution does for L_2-spherically symmetric distributions, i.e. to be the only L_2-spherically symmetric distribution with independent marginals [1; 11].

At the moment, we are working on proving or refuting this hypothesis. For using contrast gain control as a nonlinear redundancy reduction mechanism, it suffices to show that for each given p the p-generalized Normal distribution is in fact a factorial L_p-spherically symmetric distribution.

Properties of the p-generalized Normal distribution

(P1) The p-generalized Normal distribution has independent marginals which all belong to the exponential power family with identical shape parameter p.

(P2) The radial cumulative distribution function of the p-generalized Normal distribution over $r_p = ||\mathbf{y}||_p$ is given by

$$\mathcal{F}_{\mathcal{N}_p}(r_p) = \frac{\Gamma\left(\frac{n}{p}, \frac{r_p^p}{2\sigma^2}\right)}{\Gamma\left(\frac{n}{p}\right)}$$

where $\Gamma\left(\frac{n}{p}, a\right) = \int_0^a y^{\frac{n}{p}-1} e^{-y} dy$ denotes the incomplete Γ-function (see section 5.3.3)

(P3) Uniqueness conjecture (remains to be shown): For every p and given variance the p-generalized Normal distribution is the only L_p-spherical distribution with independent marginals.

A schematic illustration of the properties of the p-generalized Normal distribution is depicted in Figure 4. The properties (P1) and (P2), in particular, provide the fundamental basis for the L_p-spherical distribution to be usable as a contrast gain control model for redundancy reduction: Given the cdf $\mathcal{F}_p(r_p)$ over the 'radial' component $r_p = ||\mathbf{y}||_p$ of an arbitrary L_p-spherically distribution we can use (P2) for the construction of a contrast gain control function

$$h(r_p) = \frac{(\mathcal{F}_{\mathcal{N}_p}^{-1} \circ \mathcal{F}_p)(r_p)}{r_p} \tag{2}$$

which thus transforms the distribution into a p-generalized Gaussian. Thus, by property (P1), the new random variable $\mathbf{z} = h(r_p)\mathbf{y}$ has a *factorial* distribution with marginals belonging to the exponential power family.

2.3 Fitting L_p-spherically Symmetric Distributions to Natural Images

In principle, the radial component of an L_p-spherically symmetric distribution can be of arbitrary shape, and because $r_p = ||\mathbf{y}||_p$ is a one-dimensional random variable, precise non-parametric density estimation is feasible. For the sake of simplicity, however, we here consider two cases of parametric density models only, the Gamma and the Log-normal distribution.

First, we shortly recap the foundations of L_p-spherically symmetric distributions. In [7] a multivariate random variabe Y is defined to have an L_p-spherically symmetric distribution if Y can be written as a product $\mathsf{Y} = \mathsf{R} \cdot \mathsf{U}$ of two independent random variables U and R, where U is a multivariate random variable which is uniformly distributed on the unit p-sphere and R is a non-negative univariate random variable R with arbitrary distribution (see Definition 5.6). Subsequently, [7] show that this definition is equivalent to Y having a density of the form $g(||\mathbf{y}||_p^p)$ (see Theorem 5.8). Furthermore, they show that if Y is p-spherically symmetric distributed, then the random variables $||\mathsf{Y}||_p$ and $\frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ are independent and $\frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ is uniformly distributed on the unit p-sphere (see Lemma 5.7).

This suggests two ways of constructing p-spherically symmetric distributions. The first is to simply choose a radial distribution q_r and define

$$p(\mathbf{y})d\mathbf{y} = \frac{q_r(r)}{S_p^{n-1}} d\mathbf{s} dr,$$

where S_p^{n-1} is the surface area, and $\mathbf{s}$ is an isometric parametrization of the p-sphere in n dimensions. With this construction, the distribution is defined in a polar parametrization that separates the radial from the directional component. The density over $\mathbf{y}$ is obtained by reversing the coordinate transformation. Throughout this paper, we use the parametrization of [7; 15] that transforms $\mathbf{y} \in \mathbb{R}^n$ into

$$\begin{aligned} \mathbf{y} \quad &\mapsto \quad (r, \mathbf{u}) \\ \text{with} \quad & \quad r = ||\mathbf{y}||_p \\ \text{and} \quad & \quad u_i = \frac{y_i}{||\mathbf{y}||_p} \text{ for } i = 1, ..., n-1. \end{aligned}$$

Note, that this parametrization is not isometric and only one-to-one on the half-p-sphere. Because of the former, we get an additional correction factor due to the transformation of the differential. The latter does not harm us, since the determinant of the Jacobian is the same on both half-p-spheres and since we ignore the specific direction of $\mathbf{y}$ anyway. In this parametrization, the uniform distribution on the unit p-sphere is given by (see Theorem 5.4)

$$q_u(u_1, ..., u_{n-1})d\mathbf{u} \quad = \quad \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}$$

and the radial distribution, corresponding to the density $g(||\mathbf{y}||_p^p)$, by

$$\begin{aligned} q_r(r)dr \quad &= \quad \frac{r^{n-1}2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})} g(r^p)dr \\ &= \quad S_p^{n-1}(r)g(r^p)dr, \; r > 0, \end{aligned}$$

where $S_p^{n-1}(r)$ is the surface area of the p-sphere with radius r (see Lemma 5.5).

By choosing an arbitrary q_r, multiplying it with q_u and reversing the transformation, will yield a p-spherically symmetric distribution on $\mathbf{y}$. The distributions[2] $SC_p[\mu, \Gamma[s, u]]$ and $SC_p[\mu, \log\mathcal{N}[\mu, \sigma]]$ below are obtained in that way.

[2] Analogous to elliptically contoured distributions, which are denoted by EC, we denote p-spherically symmetric or p-spherically contoured distributions by SC_p. Note, that SC_2 corresponds to an elliptically contoured distribution in the original pixel space.

The other way of obtaining a p-spherically symmetric distribution is simply to specify a density of the form $g(||\mathbf{y}||_p^p)$. In order to obtain the radial distribution, one has to transform $\mathbf{y}$ into $(r, \mathbf{u})$ and integrate out $\mathbf{u}$. The p-generalized Normal, denoted by $\mathcal{N}_p[\mu, \sigma]$, below is an example of that approach.

We finish the part about p-spherically symmetric distributions by introducing the three p-spherically symmetric distributions used in our experiments. More details can be found in section 5.3. In the following we assume that Y has mean zero and, therefore, skip the mean μ.

- $SC_p[\mu, \Gamma[s, u]]$: We obtain the *p-spherically symmetric distribution with Γ-distributed radial component*, short $SC_p[\mu, \Gamma[s, u]]$, by multiplying the uniform distribution on the unit p-sphere (see Theorem 5.4) with a Γ distribution

$$q_r(r) = \frac{r^{u-1}e^{-\frac{r}{s}}}{s^u\Gamma(u)}$$

with shape parameter u and scale parameter s.

In Euclidean coordinates the density is given by

$$p(\mathbf{y}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^n s^u \Gamma(u)\Gamma^n\left(\frac{1}{p}\right)}||\mathbf{y}||_p^{u-n}e^{-\frac{||\mathbf{y}||_p}{s}},$$

The maximum likelihood estimations of its parameters u and s are

$$\hat{s} = \frac{\hat{\mathbb{V}}[\mathsf{R}]}{\hat{\mathbb{E}}[\mathsf{R}]} \quad \text{and} \quad \hat{u} = \frac{\hat{\mathbb{E}}^2[\mathsf{R}]}{\hat{\mathbb{V}}[\mathsf{R}]},$$

where $\hat{\mathbb{E}}[\mathsf{R}]$ and $\hat{\mathbb{V}}[\mathsf{R}]$ denote the empirical mean and the empirical variance.

- $SC_p[\mu, \log\mathcal{N}[\mu, \sigma]]$: If we choose the log normal distribution

$$q_r(r) = \frac{1}{r\sigma\sqrt{2\pi}}e^{-\frac{(\log r-\mu)^2}{2\sigma^2}}$$

as radial distribution, we obtain

$$p(\mathbf{y}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{||\mathbf{y}||_p^n\sigma\sqrt{2\pi}2^n\Gamma^n\left(\frac{1}{p}\right)}e^{-\frac{(\log||\mathbf{y}||_p-\mu)^2}{2\sigma^2}}.$$

The maximum likelihood estimates of μ and σ are given by

$$\begin{aligned}\hat{\mu} &= \log(\hat{\mathbb{E}}[R]) - \frac{1}{2}\log\left(1 + \frac{\hat{\mathbb{V}}[R]}{\hat{\mathbb{E}}^2[R]}\right)\\ \hat{\sigma} &= \sqrt{\log\left(1 + \frac{\hat{\mathbb{V}}[R]}{\hat{\mathbb{E}}^2[R]}\right)}.\end{aligned}$$

- $\mathcal{N}_p[\sigma^2]$: The p-generalized Normal distribution is obtained by choosing Y to be a collection of n i.i.d. random variables y_i, each distributed according to the exponential power distribution

$$\begin{aligned}\mathsf{y}_i \sim p(y) &= \frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2}e^{-\frac{|y|^p}{2\sigma^2}}\\ \mathsf{Y} \sim p(\mathbf{y}) = \prod_{i=1}^n p(y_i) &= \left(\frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2}\right)^n e^{-\frac{\sum_{i=1}^n |y_i|^p}{2\sigma^2}}\end{aligned}$$

Since $p(\mathbf{y})$ has the form $g(||\mathbf{x}||_p^p)$, it is a proper p-spherically symmetric distribution. Note, that for the case of $p = 2$, the p-generalized Normal distribution reduces to a multivariate isotropic Gaussian. In order to compute the contrast gain control function, we need to compute the radial distribution q_r of $p(\mathbf{y})$. Transforming p according to Lemma 5.3 and integrating out $\mathbf{u}$ yields the radial distribution

$$q_r(r) \quad = \quad \frac{p\, r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} e^{-\frac{r^p}{2\sigma^2}}.$$

The maximum likelihood estimate for σ, given the data $\{r_1, ..., r_m\} = \{||\mathbf{y}_1||_p, ..., \mathbf{y}_m||_p\}$, is

$$\hat{\sigma} \quad = \quad \sqrt{\frac{p}{2mn}\sum_{i=1}^{m} r_i^p}.$$

For the transformation of the radial component, we also need the cumulative distribution function of q_r. As shown in section 5.3.3, it is given by

$$\begin{aligned}\mathcal{F}_{\mathcal{N}_p}(a) \quad &= \quad \int_0^a \frac{p r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} e^{-\frac{r^p}{2\sigma^2}} dr \\ &= \quad \frac{\Gamma\left(\frac{n}{p}, \frac{a^p}{2\sigma^2}\right)}{\Gamma\left(\frac{n}{p}\right)},\end{aligned}$$

where $\Gamma\left(\frac{n}{p}, a\right) = \int_0^a y^{\frac{n}{p}-1} e^{-y} dy$ is the incomplete Γ-function.

2.4 Evaluation Measures

In this work, we aim at quantifying the effect of orientation selectivity and constrast-gain control on the redundancy reduction for natural images. Redundancies can be quantified by a comparison of coding costs. According to Shannon's channel coding theorem the entropy of a discrete random variable is an attainable lower bound on the coding cost for error-free encoding [5]. For the construction of such a code, it is necessary to know the true distribution of the random variable. If the assumed distribution $\hat{P}(k)$ used for the construction of an optimal code is different from the true distribution $P(k)$, the coding cost is given by the log-loss

$$E[-\log(\hat{P}(k))] = -\sum_k P(k) \log \hat{P}(k) = H[k] + D_{KL}[P(k)||\hat{P}(k)]\,.$$

$H[k] = -\sum_k P(k) \log P(k)$ denotes the discrete entropy and $D_{KL}[P(k)||\hat{P}(k)] = \sum_k P(k) \log(P(k)/\hat{P}(k))$ the Kullback-Leibler divergence. The Kullback-Leibler divergence is always positive and quantifies the additional coding cost caused by using a model distribution different from the true one.

For continuous random variables, the total amount of bits required for loss-less encoding is infinite. In order to get a meaningful measure of redundancy, one can compare the coding costs after discretization in the limit of small bin size. Accordingly, we have

$$D_{KL}[p(\mathbf{y})||\hat{p}(\mathbf{y})] = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})}\, d\mathbf{y} = \lim_{|\Delta\mathbf{y}|\to 0} \sum p(\mathbf{y}) \Delta\mathbf{y}\, \log \frac{p(\mathbf{y})\Delta\mathbf{y}}{\hat{p}(\mathbf{y})\Delta\mathbf{y}} \tag{3}$$

where $p(\mathbf{y})$ denotes the density over $\mathbf{y} \in \mathbb{R}^m$ and the right hand side constitutes a Riemann sum with $\Delta\mathbf{y}$ being the m-dimensional volume element.

Now we are in the position to quantitatively measure redundancies of continuous random variables based on the Kullback-Leibler distance of a factorial model distribution $\hat{p}(\mathbf{y}) = \prod_{j=1}^m p_j(y_j)$ from the true joint distribution. The resulting measure is the multi-information which can be defined for any multivariate random variable

$$I[p(\mathbf{y})] \;=\; D_{\mathrm{KL}}[p(\mathbf{y})||\textstyle\prod_{j=1}^m p_j(y_j)] \;=\; \displaystyle\int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{j=1}^m p_j(y_j)} d\mathbf{y} \quad . \tag{4}$$

The goal of redundancy reduction is to map a random variable $\mathbf{y}$ to a new random variable $\mathbf{z} = f(\mathbf{y})$ such that the multi-information is reduced as much as possible. In this study, in particular, we want to compare the redundancy reduction achieved with different linear and also nonlinear mappings. To this purpose it is not necessary to know the true joint distributions as long as we consider differences between the multi-informations of different representations only. The following calculation shows that evaluating the redundancy reduction achieved with a mapping $\mathbf{z} = f(\mathbf{y})$ is equivalent to evaluating the difference between the log-loss of two particular model distributions. Let $p(\mathbf{y})$ denote the density of the input and $q(\mathbf{z}) = \left(\lim_{\Delta\mathbf{y}\to 0} \frac{\Delta\mathbf{y}}{\Delta\mathbf{z}}\right) p\left(f^{-1}(\mathbf{z})\right)$ the resulting density of the output of f. Note, that $\Delta\mathbf{z}$, or $d\mathbf{z}$ in the limit, is a function of $\Delta\mathbf{y}$ or $d\mathbf{y}$, respectively. Consequently, it holds $q(\mathbf{z})\Delta\mathbf{z} = p(\mathbf{y})\Delta\mathbf{y}$ for $\Delta\mathbf{y} \to 0$ and we can write

$$
\begin{aligned}
\Delta I &= D_{\mathrm{KL}}[q(\mathbf{z})||\textstyle\prod_{j=1}^m q_j(z_j)] - D_{\mathrm{KL}}[p(\mathbf{y})||\prod_{j=1}^m p_j(y_j)] \\
&= \lim_{|\Delta\mathbf{y}|\to 0} \sum p(\mathbf{y})\Delta\mathbf{y} \left(\log \frac{q(\mathbf{z})\Delta\mathbf{z}}{\prod_{j=1}^m q_j(z_j)\Delta\mathbf{z}} - \log \frac{p(\mathbf{y})\Delta\mathbf{y}}{\prod_{j=1}^m p_j(y_j)\Delta\mathbf{y}} \right) \\
&= \lim_{|\Delta\mathbf{y}|\to 0} \sum p(\mathbf{y})\Delta\mathbf{y} \, \log \left(\frac{\prod_{j=1}^m p_j(y_j)\Delta\mathbf{y}}{\prod_{j=1}^m q_j(z_j)\Delta\mathbf{z}} \cdot \frac{q(\mathbf{z})\Delta\mathbf{z}}{p(\mathbf{y})\Delta\mathbf{y}} \right) \\
&= \lim_{|\Delta\mathbf{y}|\to 0} \sum p(\mathbf{y})\Delta\mathbf{y} \, \log \left(\frac{\prod_{j=1}^m p_j(y_j)\Delta\mathbf{y}}{\prod_{j=1}^m q_j(z_j)\Delta\mathbf{z}} \right) \\
&= \lim_{|\Delta\mathbf{y}|\to 0} \sum p(\mathbf{y})\Delta\mathbf{y} \, \log \left(\frac{\prod_{j=1}^m p_j(y_j)\Delta\mathbf{y}}{\hat{p}_f(\mathbf{y})\Delta\mathbf{y}} \right) \\
&= \int p(\mathbf{y}) \log \frac{\prod_{j=1}^m p_j(y_j)}{\hat{p}_f(\mathbf{y})} \, d\mathbf{y} \\
&= E[-\hat{p}_f(\mathbf{y})] - E[-\textstyle\prod_{j=1}^m p_j(y_j)]
\end{aligned}
$$

where $\hat{p}_f(\mathbf{y}) = \left(\lim_{\Delta\mathbf{y}\to 0} \frac{\Delta\mathbf{z}}{\Delta\mathbf{y}}\right) \hat{q}(f(\mathbf{y})) = f'(\mathbf{y})\, \hat{q}(f(\mathbf{y}))$ denotes the model density in the input space which corresponds to the factorial model distribution $\hat{q}(\mathbf{z}) = \prod_{j=1}^m q_j(z_j)$ in the output space. Thus, if we have a model density which does not factorize with respect to $\mathbf{y}$ and we have a (possibly nonlinear) mapping $\mathbf{z} = f(\mathbf{y})$ such that the transformed model density with respect to $\mathbf{z}$ becomes factorial, we can evaluate the redundancy reduction achieved with the mapping f simply by estimating the difference in the average log-loss obtained for $\hat{p}_f(\mathbf{y})$ and for $\prod_{j=1}^m p_j(y_j)$. In order to get a measure which is less dependent on the number of dimensions m we define the average log-loss (ALL) to be

$$
ALL = \frac{1}{m} E[-\log \hat{p}(\mathbf{y})] \tag{5}
$$

for any given model distribution $\hat{p}(\mathbf{y})$. In summary, we have here outlined a way to deal with the case of nonlinear redundancy reduction mechanisms such as contrast gain control: Instead of estimating the multi-information directly, it suffices to estimate the differences in the average log-loss between the different models.

3 Experiments and Results

3.1 Data-set

Here, we resort to the data-set used in a previous study [17; 12]. This color image data set contains images from the Bristol Hyperspectral Images Database [4] which consists of multi-spectral recordings of natural scenes taken in the surroundings of Bristol, UK and in the greenhouses of Bristol Botanical Gardens. Our analysis was performed on a subset of eight images similar to the ones used in [12; 6]. The authors of this study kindly provided to us a pre-processed version of the image data where spectral radiance vectors were converted into LMS-values. During subsequent processing the reflectance standard was cut out and images were converted to log-intensities (see . [12]).

All images come at a resolution of 256×256 pixels. From each image circa 5000 patches of size 7×7 pixels were drawn at random locations for training (circa 40000 patches in total) as well as circa 6250 patches per image for testing (circa 50000 patches in total). In total, we sampled ten pairs of training and test sets in that way. All results

below are averaged over ten trials of the experiments on those pairs. All experiments were carried out on gray level and color images for exactly the same sets of images patches. In both cases, we used the log pixel intensities. For chromatic images with three color channels (LMS) each patch is reshaped as a $7 \times 7 \times 3 = 147$-dimensional vector. The conversion to gray level images was done by simply averaging the channels $I = \log(\frac{1}{3}(L + M + S))$. In this case the dimensionality of a data vectors reduced to 49 dimensions. Further details of preprocessing have been mentioned above in section 2.1 and can also be found in Appendix 5.1.

3.2 Experiments

We fitted the distributions $SC_p[\mu, \Gamma[s, u]]$, $SC_p[\mu, \log \mathcal{N}[\mu, \sigma]]$ and $\mathcal{N}_p[\mu, \sigma]$ for various values to the coefficients of chromatic and achromatic natural image patches in the bases HAD, SYM, wPCA and ICA for several values of p. For each model, we computed the maximum likelihood estimate of the model parameters and the value for p with the best average log-loss on a training set. Afterwards, we computed the average log-loss on an independent test set. This procedure was repeated for ten independently sampled splits of training and test set. All results are averaged over those ten splits. The transforms for HAD, SYM and wPCA were obtained as described in section 2.1. For ICA, we performed a gradient descent over the orthogonal group where we used the result of the FastICA algorithm by [9] as initial starting point. All transforms were computed separately for each of the training sets.

In order to compare the redundancy reduction of the different transforms with respect to the pixel basis (PIX), we computed a non-parametric estimate of the marginal entropies of the patches in that representation with DC component still contained [3]. Since the estimation is not bound to a particular parametric model, we used the mean of the marginal entropies as an estimate of the average log-loss in the pixel representation.

3.3 Results

Figure 6 shows the average log-loss for the transforms HAD, SYM, wPCA and ICA, and the factorial $\mathcal{N}_p[\mu, \sigma]$ model (dashed lines) as well as the non-factorial $SC_p[\mu, \Gamma[s, u]]$ model. The curves for the $SC_p[\mu, \log \mathcal{N}[\mu, \sigma]]$ model are almost exactly the same as the ones for the $SC_p[\mu, \Gamma[s, u]]$ model and we therefore omit them.

The standard deviation about the ten trials is indicated via the linewidth of the curves in Figure 6. One striking finding is the superiority of the non-factorial model over the factorial one in terms of the ALL score. This shows again, that the linear independence assumption is not a good choice for a model of natural images. Note, that at $p = 2$ the curves of the non-factorial model intersect for all different basis transforms, since for $p = 2$ the model is invariant under orthogonal transformations.

In order to assess the relative contributions of whitening, orientation selectivity, and contrast gain control to redundancy reduction, we consider the difference in the ALL scores (see section 2.4) of whitening in the Hadamard basis (WHI), plain orientation selectivity (OS), plain contrast gain control (CG), retinal contrast gain control (CG_{ret}) and the joint model (OC) relative to the average marginal entropy of the pixel basis (PIX). The latter is estimated with the nonparametric entropy estimator NPL described in [3].

Absolute Difference [Bits/Comp.]			Relative Difference [%]		
	Color	Gray		Color	Gray
WHI - PIX	-4.2171 ± 0.0041	-3.2470 ± 0.0041	$\frac{\text{WHI-PIX}}{\text{OC-PIX}}$	92.1971 ± 0.0578	91.0850 ± 0.0808
OS - PIX	-4.2171 ± 0.0041	-3.3359 ± 0.0038	$\frac{\text{OS-PIX}}{\text{OC-PIX}}$	95.7591 ± 0.0462	93.5792 ± 0.0773
CG - PIX	-4.4959 ± 0.0057	-3.5373 ± 0.0060	$\frac{\text{CG-PIX}}{\text{OC-PIX}}$	98.2921 ± 0.0109	99.2277 ± 0.0092
CG_{ret} - PIX	-4.5267 ± 0.0057	-3.5466 ± 0.0059	$\frac{\text{CG}_{\text{ret}}\text{-PIX}}{\text{OC-PIX}}$	98.9654 ± 0.0112	99.4875 ± 0.0087

Table 1: **Comparision of the Redundancy Reduction for Orientation Selectivity, Contrast Gain Control and the Combined Model** The table shows the absolute and relative differences in redundancy reduction achieved by whitening in the Hadamard basis (WHI), contrast gain control (CG), retinal contrast gain control (CG_{ret}) which uses the SYM basis, orientation selevtivity (OS) and the full model (OC) which combines both orientation selectivity and contrast gain control. The full model adds a little less than 10% on top of Hadamard whitening. When comparing the contributions of each single model, orientation selectivity always contributes less than contrast gain control. The maximal contribution of orientation selectivity to the full model is less than 1% for gray images and less than 2% for color images. If we compare the full model to the retinal contrast gain control model, the contribution of orientation selectivity is further reduced to 1% for color and 0.5% for gray images. The absolute differences in the average log-loss are specified in bits per component.

For plain whitening, we use the average log-loss for the factorial $\mathcal{N}_p[\mu, \sigma]$ model in the HAD basis with the optimal value for p. This can be seen as an upper bound for the average log-loss of an arbitrary whitening transform, since we expect the HAD basis to produce coefficients which are as Gaussian as possible (as argued in Section 1). Clearly, an even more conservative upper bound would be the average log-loss in the Gaussian case, i.e. the $\mathcal{N}_p[\mu, \sigma]$ model for $p = 2$ which is invariant under orthogonal transformations. However, since the joint distribution of the coefficients look significantly non-Gaussian, even for the HAD case, it is more appropriate to use the basis which is expected to produce normal distributed coefficients, if they were independent (see also Figure 1). The fact that the ALL curve for the HAD transformation under the non-factorial $SC_p[\mu, \Gamma[s, u]]$ model has its minimum at $p = 2$ shows that this is as reasonable choice.

For modelling the contribution of plain orientation selectivity, we use the optimal ALL value of the ICA basis under the factorial $\mathcal{N}_p[\mu, \sigma]$ model, since the ICA filters exhibit the characteristics of orientation selective receptive fields and the $\mathcal{N}_p[\mu, \sigma]$ model assumes the marginals to be independent already without the nonlinear contrast gain control transformation.

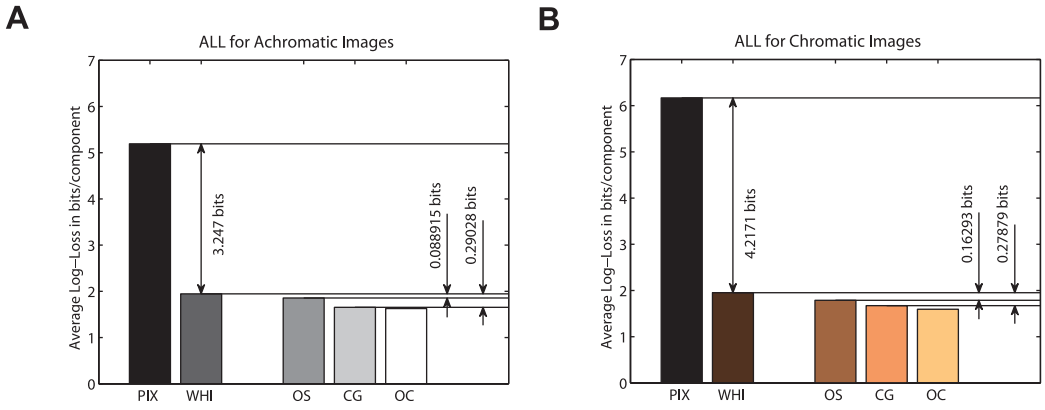

Figure 5: **Comparision of the Redundancy Reduction for Orientation Selectivity, Contrast Gain Control and the Combined Model** The figure shows the absolute and relative differences in redundancy reduction achieved by contrast gain control, orientation selevtivity and the combined model. The combined model only contributes less than 10% compared to the pixel basis PIX. When comparing the contributions of each single model on top of plain whitening, orientation selectivity always contributes less than contrast gain control, although the differences are smaller for chromatic images.

For the situation of plain contrast gain control, we take the average log-loss of the spherical $SC_p[\mu, \Gamma[s, u]]$ model, i.e. $p = 2$, with an arbitrary basis. This corresponds to the traditional constrast-gain control model, where the image is scaled with a function of the Euclidean norm and, therefore, does not dependent on the particular choice of the orthogonal transform V in whitened space. As mentioned before, it also corresponds to the optimal p in case when using the Hadamard basis.

In a similar way as we use the ICA basis to model orientation selectivity as found for V1 simple cells, we can also use the SYM basis to model the center-surround receptive fields observed for retinal ganglion cells. Since retinal ganglion cells are known to exhibit contrast gain control too, it makes sense to compare the $SC_p[\mu, \Gamma[s, u]]$ model with the optimal value of p for the SYM basis to the $SC_p[\mu, \Gamma[s, u]]$ model with the optimal value of p for the ICA basis. In this way we get an idea how much the orientation selectivity adds on top of the contrast gain control in the retina. The interpretation of these models as redundancy reduction mechanisms is justified as the distributions can be transformed into one with independent marginals by applying the deterministic scaling transformation $h(r_p) = (\mathcal{F}_{\mathcal{N}_p}^{-1} \circ \mathcal{F}_p)(r_p)/r_p$.

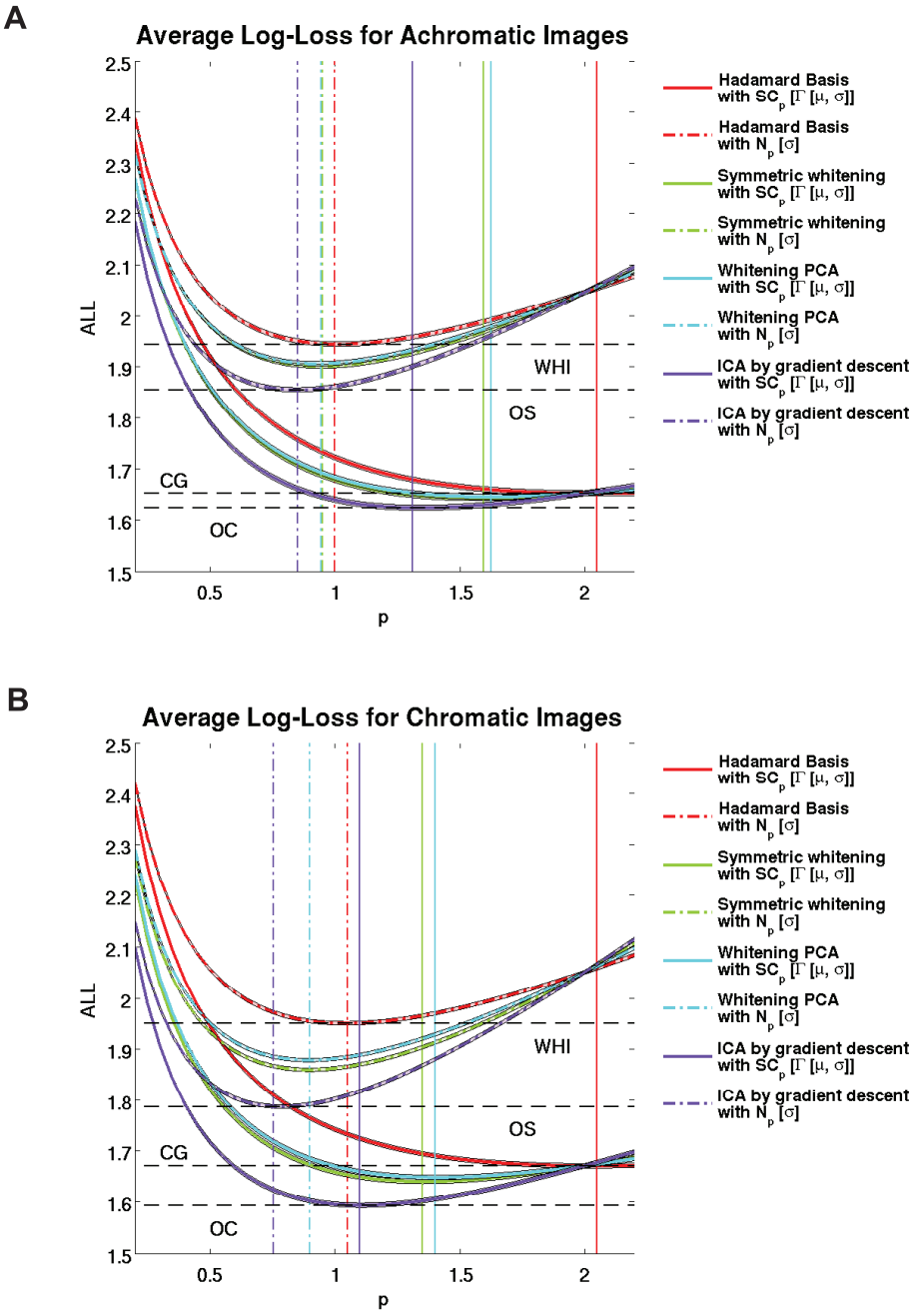

Figure 6: **Average Log-Loss (ALL) as a Function of p** Both figures show the average log-loss in bits per component on an independent test for the factorial (dashed lines) and non-factorial models (solid lines) fit to the coefficients in the bases HAD, SYM, wPCA and ICA. We show the curves of the $SC_p[\mu, \Gamma[s, u]]$ model for the non-factorial case. When using the $SC_p[\mu, \log \mathcal{N}[\mu, \sigma]]$ the curves look almost exactly the same. The shaded hull around the curves depicts the standard deviation of the test error. The vertical lines denote the position of minimal average log-loss on the training set. The horizontal lines indicate the values chosen for arbitrary whitening (WHI), pure orientation selectivity (OS), contrast gain control (CG) and the combined model (OC).

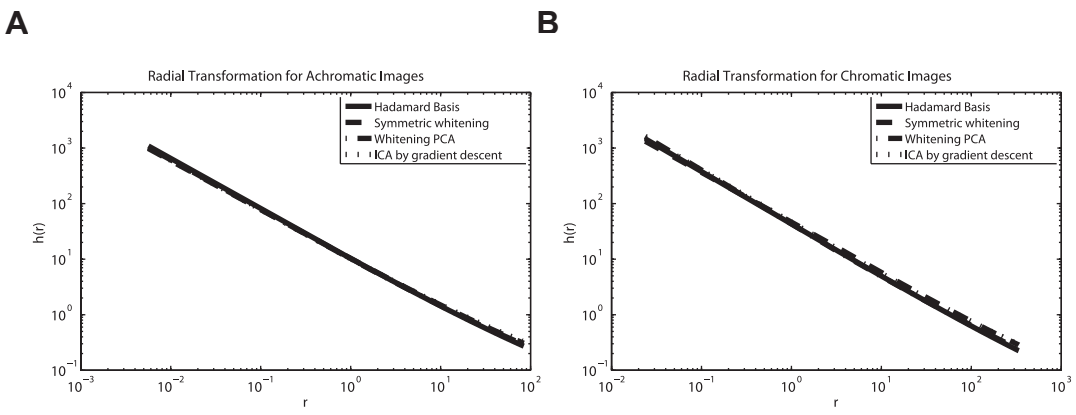

Figure 7: **Radial Scaling Function** The figure shows the radial scaling function $h(r_p) = (\mathcal{F}_{\mathcal{N}_p}^{-1} \circ \mathcal{F}_p)(r_p)/r_p$ for achromatic images (A) and chromatic images (B). As in the traditional contrast gain control model, we find that the function follows a power law of the form $h(r) = \frac{c}{r^\kappa}$. Fixing the output variance of the $\mathcal{N}_p[\mu, \sigma]$ model to one, we find that all κ range between 0.79 and 0.84 for achromatic images, and 0.86 and 0.91 for color images, respectively.

Figure 5 and Table 1 show the redundancy reduction achieved by orientation selectivity (OS), contrast gain control (CG), retinal contrast gain control (CG_{ret}) and the joint model (OC) compared to the reference points PIX and WHI. As already reported in [6; 3], plain orientation selectivity adds only very little to the redundancy reduction achieved by decorrelation and is even less effective than plain contrast gain control. If both orientation selectivity and contrast gain control are combined it is possible to achieve a little less than 10% extra to the redundancy reduction with plain whitening only. When comparing the effect of plain contrast gain control and plain orientation selectivity we find that orientation selectivity always contributes less to the overall redundancy reduction. In fact, when looking at the additional gain for using the joint model instead of plain contrast gain control, we find that the percental difference is only 0.7723% for gray level and 1.7088% for color images, respectively. This means that less than two percent of the redundancy reduction can be attributed to orientation selectivity.

We also examined the contrast gain function h transforming the radial component of the $SC_p[\mu, \Gamma[s, u]]$ model into the radial component of the $\mathcal{N}_p[\mu, \sigma]$ (see Figure 7). When comparing those functions for the different choices of bases in whitened space, we find that they exhibit a power law shape similar to the function assumed in the traditional model , i.e. $h(r) = \frac{c}{r^\kappa}$ with $\kappa = 1$. Here we derive that, for an optimal contrast gain control mechanism, κ should be chosen slightly less than one. The optimal κ parameter ranges between 0.79 and 0.84 for achromatic images, and 0.86 and 0.91 for color images, respectively. When choosing the scale parameter of the $\mathcal{N}_p[\mu, \sigma]$ model such that the output becomes white (under the assumption that the $SC_p[\mu, \Gamma[s, u]]$ model perfectly fits the data), all curves fall on top of each other.

4 Summary

In this report, we studied the conjoint effect of contrast gain control and orientation selectivity on the redundancy reduction for natural images. In particular, we showed how the L_p-spherically distribution can be used to tune the nonlinearity of contrast gain control to remove higher-order redundancies in natural images. Similar to the classical model the gain control function turns out to follow a power law. In difference to the classical model, we find that the absolute value of the exponent is slightly smaller than one and the optimal p-norm is slightly smaller than two. In addition, we find that the relevance of orientation selectivity is further reduced in the conjoint model for orientation selectivity and contrast gain control. In the linear framework (possibly endowed with a point-wise nonlinearity for each neuron) the contribution of orientation selectivity to redundancy reduction has been shown to be smaller than 5% relative to whitening (i.e. bandpass filtering) alone [3; 6]. Here, we found that the contribution of orientation selectivity is even smaller than two percent relative to decorrelation plus gain control. We conclude that orientation selectivity may show a more pronounced advantage for other objectives than redundancy reduction.

5 Appendix

5.1 Preprocessing

In the first step, a projector P_{remDC} is computed such that the first (for each channel) component of $P_{remDC}\mathbf{x}$ corresponds to the DC component(s) of that patch. This intermediate step is performed because the distribution of a DC-component differs strongly from the distribution of the remaining channels. Additionally this separation allows us to distinguish between the contributions of both parts to the total entropy. Any further decorrelation transforms will be performed in the remaining subspace orthogonal to the DC-components. The transpose of the matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & 1 & 0 & \cdots \\ 1 & 0 & 1 & \cdots \\ \vdots & & & \ddots \end{pmatrix}$$

has exactly the required property. However, it is not an orthogonal transformation. Therefore, we decompose P into $P = QR$ where R is upper triangular and Q is an orthogonal transform. Since $P = QR$, the first column of Q must be a multiple of the vector with all coefficients equal to one (due to the upper triangluarity of R). Therefore, the first component of $Q^\top \mathbf{x}$ is a multiple of the DC component. Since Q is an orthonomal transform, using all but the first row of $Q^\top$ for P_{remDC} projects out the DC component. In case of color images the same trick is applied to each channel by making P_{remDC} a block-diagonal matrix with $Q^\top$ as diagonal elements.

Secondly, the data was scaled such that the whitening transform has determinant one, i.e. that the determinant of the globally scaled data is one. This is done by setting $\eta = \prod \lambda_i^{\frac{1}{2n}}$, where λ_i are the eigenvalues of the covariance matrix of the training data and n is their dimension. Therefore, the determinant of the covariance matrix of the data after scaling with $\frac{1}{\eta}$ is

$$\frac{1}{\eta^{2n}} \prod \lambda_i = \frac{\prod \lambda_i}{\left(\prod \lambda_i^{\frac{1}{2n}}\right)^{2n}} = 1.$$

Since the whitening transform consist of $D^{-\frac{1}{2}}U^\top$ with $UDU^\top = C$ (C is the determinant of the scaled data), the whitening must have determinant one due to

$$1 = \det(C) = \det(UDU^\top) = \det(D^{-\frac{1}{2}}U^\top)^2$$

Note, that the same scaling factor is used for the training and test set.

5.2 Definitions, Lemmas and Theorems

In this part, we provide the rigorous definitions, lemmas and theorems used in the text above. Most results and proofs are not new and have been collected from papers and books. Nevertheless, in many cases we adapted the original statements to our need and provided more detailed versions of the proofs. The original sources are mentioned at the respective lemmas and theorems.

Definition 5.1. p**-Norm** Let $\mathbf{y} \in \mathbb{R}^n$ be an arbitrary vector. We define

$$||\mathbf{y}||_p \quad = \quad \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}, \; p > 0$$

as the p-norm of $\mathbf{y}$. Note, that only for $p > 1$, $||\mathbf{y}||_p$ is a strict norm. However, we will also use the term "p-norm" even if only $0 < p$.

Definition 5.2. p**-Sphere**
The unit p-sphere $\mathbb{S}_p^{n-1}$ in n dimensions is the set of points that fulfill

$$\mathbb{S}_p^{n-1} \quad := \quad \{\mathbf{y} \in \mathbb{R}^n | \, ||\mathbf{y}||_p = 1, \, p > 0\}.$$

Lemma 5.3. ***Transformation in Radial and Spherical Coordinates [15]***

Let $\mathbf{y} = (y_1, ...y_n)^\top$ $n \geq 2$ *be a vector in* $\mathbb{R}^n \backslash \{\mathbf{0}\}$. *Consider the transformation*

$$\mathbf{y} \quad \mapsto \quad (r, u_1, ..., u_{n-1}) = \left(||\mathbf{y}||_p, \frac{y_1}{||\mathbf{y}||_p}, ..., \frac{y_n}{||\mathbf{y}||_p} \right).$$

The absolute value of the determinants of the transformation on the upper and lower halfspaces

$$\begin{aligned} \mathbb{R}^n_+ &:= \{\mathbf{y} \in \mathbb{R}^n | \, y_n \geq 0\} \\ \mathbb{R}^n_- &:= \{\mathbf{y} \in \mathbb{R}^n | \, y_n < 0\} \end{aligned}$$

are equal and are given by

$$|\det \mathcal{J}| \quad = \quad r^{n-1} \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}}.$$

Proof. The proof is a more detailed version of the proof found in [15].

Let

$$\Delta_i \quad := \quad \begin{cases} 1, \, u_i \geq 0 \\ -1, \, u_i < 0. \end{cases}$$

Then we can write $|u_i| = \Delta_i u_i$. The above transformation is bijective on each of the regions $\mathbb{R}^n_+$ and $\mathbb{R}^n_-$. Let $\sigma = \text{sign}(y_n)$, then the inverse is given by

$$\begin{aligned} y_i &= u_i r \, , \, 1 \leq i \leq n-1 \\ y_n &= \sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1}{p}} = \sigma r \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p \right)^{\frac{1}{p}}. \end{aligned}$$

Note, that the $\sigma = \text{sign}(y_n)$ determines the halfspace in which the transformation is inverted.

First, we determine the Jacobian $\mathcal{J}$. We start with computing the derivatives

$$\begin{aligned} \frac{\partial y_i}{\partial u_j} &= \delta_{ij} r, \, 1 \leq i, j \leq n-1 \\ \frac{\partial y_n}{\partial u_j} &= -\sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}} \Delta_i^p {u_i}^{p-1}, \, 1 \leq j \leq n-1 \\ \frac{\partial y_i}{\partial r} &= u_i, \, 1 \leq i \leq n-1 \\ \frac{\partial y_n}{\partial r} &= \sigma \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p \right)^{\frac{1}{p}}. \end{aligned}$$

Therefore, the Jacobian, is given by

$$\begin{aligned} \mathcal{J} &= \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & & \frac{\partial y_1}{\partial u_{n-1}} & \frac{\partial y_1}{\partial r} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_n}{\partial u_1} & & \frac{\partial y_n}{\partial u_{n-1}} & \frac{\partial y_n}{\partial r} \end{pmatrix} \\ &= \begin{pmatrix} r & & 0 & \dots & u_1 \\ 0 & & r & & u_2 \\ \vdots & & & \ddots & \vdots \\ -\sigma r \left(1 - \sum_{i=1}^{n-1} |u_i|^p \right)^{\frac{1-p}{p}} \Delta_1^p u_1^{p-1} & & \dots & \dots & \sigma \left(1 - \sum_{i=1}^{n-1} (\Delta_i u_i)^p \right)^{\frac{1}{p}} \end{pmatrix}. \end{aligned}$$

Before actually computing the absolute value of the determinant $|\det \mathcal{J}|$, we can factor out r from the first $n-1$ columns. Furthermore, we can factor out σ from the last row. Since we take the absolute value of $\det \mathcal{J}$ and $\sigma = \{-1, 1\}$, we can remove it completely afterwards. Now we can use Laplace's formula to expand the determinant along the last column. With this, we get

$$\begin{aligned}
\frac{1}{r^{n-1}}|\det \mathcal{J}| &= \sum_{k=1}^{n-1}(-1)^{n+k} \cdot u_k \cdot (-1)^{n-1+k} \cdot -\Delta_k^p u_k^{p-1} \cdot \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} \\
&\quad +(-1)^{2n}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1}{p}} \\
&= \sum_{k=1}^{n-1}|u_k|^p\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} + \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1}{p}} \\
&= \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}\left(\sum_{k=1}^{n-1}|u_k|^p + 1 - \sum_{k=1}^{n-1}|u_k|^p\right) \\
&= \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}.
\end{aligned}$$

Resolving the result for $|\det \mathcal{J}|$ completes the proof. □

Theorem 5.4. ***p-Spherical Uniform Distribution [15]***
Let $\mathsf{Y} = (\mathsf{y}_1, ..., \mathsf{y}_n)^\top$ be a random vector. Let the y_i be i.i.d. distributed with p.d.f.

$$f(\mathbf{y}) = \frac{p^{1-\frac{1}{p}}}{2\Gamma\left(\frac{1}{p}\right)}e^{-\frac{|y|^p}{p}},\ y \in \mathbb{R}.$$

Let $\mathsf{u}_i = \frac{\mathsf{y}_i}{||\mathsf{Y}||_p}$ for $i = 1, ..., n$. Then $\sum_{i=1}^n |\mathsf{u}_i|^p = 1$ and the joint p.d.f of $\mathsf{u}_1, ..., \mathsf{u}_{n-1}$ is

$$q_u(u_1, ..., u_{n-1}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$$

with $-1 < u_i < 1$, $i = 1, ..., n-1$ and $\sum_{i=1}^{n-1}|u_i|^p < 1$.

Proof. The joint p.d.f. of Y is given by

$$f(\mathbf{y}) = \frac{p^{n-\frac{n}{p}}}{2^n\Gamma^n\left(\frac{1}{p}\right)}e^{-\frac{1}{p}\sum_{i=1}^n |y_i|^p}$$

with $y_i \in \mathbb{R}$ and $i = 1, ..., n$. Applying the transformation

$$(y_1, ..., y_n) = (r, u_1, ..., u_{n-1})$$

from Lemma 5.3 and taking into account that each $(u_1, ..., u_{n-1})$ corresponds to $(y_1, ..., y_n)$ and $(y_1, ..., -y_n)$ we obtain

$$q(u_1, ..., u_{n-1}, r) = 2 \cdot \frac{p^{n-\frac{n}{p}}}{2^n\Gamma^n\left(\frac{1}{p}\right)}r^{n-1}e^{-\frac{r^p}{p}}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}.$$

By integrating out r, we obtain $q_u(u_1, ..., u_n)$:

$$\int_0^\infty q(u_1, ..., u_{n-1}, r)dr = \frac{p^{n-\frac{n}{p}}}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}\int_0^\infty r^{n-1}e^{-\frac{r^p}{p}}dr.$$

In order to compute the integral, we use the substitution $z = \frac{r^p}{p}$ or $r = (zp)^{\frac{1}{p}}$. This yields $dr = (zp)^{\frac{1}{p}-1}dz$ and, therefore,

$$\begin{aligned}\int_0^\infty r^{n-1}e^{-\frac{r^p}{p}}dr &= \int_0^\infty (zp)^{\frac{n-1}{p}}e^{-z}(zp)^{\frac{1-p}{p}}dz \\ &= p^{\frac{n-p}{p}}\int_0^\infty z^{\frac{n}{p}-1}e^{-z}dz \\ &= p^{\frac{n-p}{p}}\Gamma\left(\frac{n}{p}\right).\end{aligned}$$

Hence,

$$\begin{aligned}q_u(u_1,...,u_{n-1}) &= \int_0^\infty q(u_1,...,u_{n-1},r)dr \\ &= \frac{p^{n-\frac{n}{p}}}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}p^{\frac{n-p}{p}}\Gamma\left(\frac{n}{p}\right) \\ &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}.\end{aligned}$$

□

In order to see, why q_u is called *uniform* on $\mathbb{S}_p^{n-1}$, we must recognize, that q_u of $\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ which is due to the coordinate transformation and $\frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}$ which corresponds to twice the surace area of the p-sphere (see Lemma 5.5). Since each $\mathbf{u}$ corresponds to two $\mathbf{y}$ before the coordinate transform (one on the upper and one on the lower halfsphere), the density of $\mathbf{u}$ in $\mathbf{y}$-coordinates corresponds to $\frac{1}{S_p^{n-1}}$ where $S_p^{n-1} = \frac{2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}$ is the surface area of the unit p-sphere.

As we will see in Lemma 5.7, $\frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ is independent of $||\mathsf{Y}||_p$ and, therefore, the specific form of the density f does not matter as long as it is p-spherically symmetric.

Lemma 5.5. ***Volume and Surface of the p-Sphere***

The volume $V_p^{n-1}(r)$ of the p-Sphere with radius r is given by

$$V_p^{n-1}(r) = \frac{r^n 2^n\Gamma(\frac{1}{p})^n}{np^{n-1}\Gamma(\frac{n}{p})}.$$

The surface $S_p^{n-1}(r)$ is given by

$$\begin{aligned}S_p^{n-1}(r) &= \frac{d}{dr}V_p^{n-1}(r) \\ &= \frac{r^{n-1}2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}.\end{aligned}$$

As a convention, we leave out the argument of $V_p^{n-1}(r)$ and $S_p^{n-1}(r)$ when denoting the volume or the surface of the unit p-sphere, i.e.

$$\begin{aligned}V_p^{n-1} &:= V_p^{n-1}(1) \\ S_p^{n-1} &:= S_p^{n-1}(1).\end{aligned}$$

Proof. In order to compute the volume of the p-sphere in n-dimension, we must solve the integral $\int_{\mathbb{S}_p^{n-1}} d\mathbf{u}$. Using the volume element transformation from lemma 5.3, we can transform the integral into

$$\begin{aligned}\int_{\mathbb{S}_p^{n-1}} d\mathbf{u} &= 2\int_0^r \int r^{n-1}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} dr\, d\mathbf{u} \\ &= 2\int_0^r r^{n-1}dr \cdot \int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\ &= \frac{1}{n}r^n \cdot 2\int \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}.\end{aligned}$$

In theorem 5.4 we prove that $q(u_1,...,u_{n-1}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ is a probability density. In particular, this means that

$$\begin{aligned}\int q(u_1,...,u_{n-1})d\mathbf{u} &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\int\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\ &= 1\end{aligned}$$

which is equivalent to

$$\int\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} = \frac{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}.$$

Therefore,

$$\begin{aligned}V_p^{n-1}(r) &= \int_{\mathbb{S}_p^{n-1}} d\mathbf{u} \\ &= \frac{2}{n}r^n \cdot \int\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \\ &= \frac{r^n 2^n\Gamma^n\left(\frac{1}{p}\right)}{np^{n-1}\Gamma\left(\frac{n}{p}\right)}\end{aligned}$$

Differentiation of $V_p^{n-1}(r)$ with respect to r yields the result for the surface area. □

Definition 5.6. **L_p-Spherically Symmetric Distribution [7]** A random vector $\mathsf{Y} = (\mathsf{y}_1,...,\mathsf{y}_n)^\top$ is said to have a L_p-spherically symmetric distribution if Y can be written as a product of two independent random variables $\mathsf{Y} = \mathsf{R}\cdot\mathsf{U}$, where R is a non-negative univariate random variable with density $q_r : \mathbb{R}^+ \to \mathbb{R}^+$ and U is uniformly distributed on the unit p-sphere, i.e.

$$q_u(u_1,...,u_n) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$$

(see Theorem 5.4).

Lemma 5.7. ***Probability Density Functions [7]***

Let $\mathsf{Y} = (\mathsf{y}_1,...,\mathsf{y}_n)^\top$ be an n-dimensional random variable with $P\{\mathsf{Y} = \mathbf{0}\} = 0$ and a density of the form $\mathsf{Y} \sim g(||\mathbf{y}||_p^p)$. Then the following three statements hold:

1. *The random variables* $\mathsf{R} = ||\mathsf{Y}||_p$ *and* $\mathsf{U} = \frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ *are independent.*

2. $\mathsf{U} = \frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ *is uniformly distributed on the unit* p*-sphere* $\mathbb{S}_p^{n-1}$.

3. $\mathsf{R} = ||\mathsf{Y}||_p$ *has a density* q_r*, where* q_r *relates to* g *via*

$$\begin{aligned} q_r(r) &= \frac{r^{n-1}2^n\Gamma(\frac{1}{p})^n}{p^{n-1}\Gamma(\frac{n}{p})}g(r^p) \\ &= S_p^{n-1}(r)g(r^p),\ r > 0. \end{aligned}$$

Proof. The proof is a more detailed version of the proof found in [7].

1. First we transform the density of Y with the transformation of lemma 5.3 and obtain the new density in spherical and radial coordinates

$$\begin{aligned} q(u_1, ..., u_{n-1}, r) &= 2\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} g(r^p)r^{n-1} \\ &\quad -1 < u_i < 1,\ 1 \leq i \leq n-1,\ \sum_{i=1}^{n}|u_i|^p < 1. \end{aligned}$$

   Since q can be written as a product of a function of r and a function of $\mathbf{u} = (u_1, ..., u_{n-1})$, U and R are independent. Thus, $||\mathsf{Y}||_p = \mathsf{R}$ and $\mathsf{U} = \frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ are independent as well.

2. In order to get $q_u(u_1, ..., u_{n-1})$, we must integrate out r. However, we do not know the exact form of g. But since q is a probability density, we know that

$$\int_0^\infty \int q(u_1, ..., u_{n-1}, r)d\mathbf{u}dr = 1.$$

   Since Y and R are independent, we can write this integral as

$$\int_0^\infty \int q(u_1, ..., u_{n-1}, r)d\mathbf{u}dr = 2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} \cdot \int_0^\infty g(r^p)r^{n-1}dr.$$

   From that, we can immediately derive

$$\int_0^\infty g(r^p)r^{n-1}dr = \left(2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}\right)^{-1}.$$

   In order to solve $\left(2\int\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}\right)^{-1}$ we can use theorem 5.4. In this theorem, we showed that $q_u(u_1, ..., u_{n-1}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}$ is the uniform distribution on the p-unit sphere. In particular, we know that $\int q(u_1, ..., u_{n-1})d\mathbf{u} = 1$ and, therefore,

$$\int \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u} = \frac{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}.$$

Thus,

$$\begin{aligned}\int_0^\infty g(r^p)r^{n-1}dr &= \left(2\int\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}d\mathbf{u}\right)^{-1} \\ &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^n\Gamma^n\left(\frac{1}{p}\right)}\end{aligned}$$

and

$$\begin{aligned}q_u(u_1,...,u_{n-1}) &= \int_0^\infty q(u_1,...,u_{n-1},r)dr \\ &= \left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}\frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}.\end{aligned}$$

This shows that Y is uniformly distributed on the unit p-sphere.

3. The density of R can be computed by integrating out $u_1,...,u_{n-1}$

$$\begin{aligned}q_r(r) &= \int q(u_1,...,u_{n-1},r)d\mathbf{u} \\ &= \frac{2^n\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}r^{n-1}g(r^p),\ r>0\end{aligned}$$

by the same argument as in 2. This completes the proof.

□

The next theorem tells us that Y is L_p-spherically symmetric distributed if and only if its density has the form $g(||\mathbf{y}||_p^p)$.

Theorem 5.8. ***Form of L_p-Spherically Symmetric Distribution [7]*** *Let $\mathsf{Y}=(\mathsf{y}_1,...,\mathsf{y}_n)^\top$ be an n-dimensional random variable with $P\{\mathsf{Y}=\mathbf{0}\}=0$. Then, the density of Y has the form $g(||\mathbf{y}||_p^p)$, where $g:\mathbb{R}_+\to\mathbb{R}_+$ is a measurable function, if and only if $\mathsf{Y}=\mathsf{RU}$ is spherically symmetric distributed, with independent R and U, where R has the density*

$$q_r(r) = \frac{2^n\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)}r^{n-1}g(r^p),\ r>0.$$

Proof. *Sufficiency:* Assume $\mathsf{Y}=\mathsf{RU}$ with independent R and U, where U is uniformly distributed on the p-sphere and R has the density q_r. Then the joint density is given by (see theorem 5.4):

$$\begin{aligned}q(r,u_1,...,u_{n-1}) &= q_r(r)\frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} \\ &\quad -1<u_i<1,\ 1\le i\le n-1,\ \sum_{i=1}^{n-1}|u_i|^p<1,\ r>0.\end{aligned}$$

Now let $y_i=ru_i$ for $1\le i\le n-1$ and $|y_n|=r\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1}{p}}$. We can use 5.3 to see that the absolute value of the determinant of the jacobian is given by

$$\left(r^{n-1}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}}\right)^{-1} = r^{1-n}\left(1-\sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{p-1}{p}}.$$

Therefore,

$$\begin{aligned} p(y_1,...,y_n) &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} q_r(||\mathbf{y}||_p)||\mathbf{y}||_p^{1-n} \\ &= g(||\mathbf{y}||_p^p). \end{aligned}$$

Necessity: Assume $\mathsf{Y} \sim g(||\mathsf{Y}||_p^p)$. According to lemma 5.7 $\frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ and Y are independent and $\frac{\mathsf{Y}}{||\mathsf{Y}||_p}$ is uniformly distributed on the p-sphere. Again in lemma 5.7 we showed that R has the density

$$q_r(r) = \frac{2^n\Gamma^n\left(\frac{1}{p}\right)}{p^{n-1}\Gamma\left(\frac{n}{p}\right)} r^{n-1} g(r^p),\ r > 0.$$

Therefore, Y is L_p-spherically symmetric distributed if and only if $\mathsf{Y} \sim g(||\mathsf{Y}||_p^p)$ for some density g. □

5.3 Distributions

5.3.1 The p-spherically symmetric distribution with Γ-distributed radial component $SC_p[\mu, \Gamma[s,u]]$

We obtain the *p-spherically symmetric distribution with Γ-distributed radial component*, short $SC_p[\mu, \Gamma[s,u]]$, by multiplying the uniform distribution on the unit p-sphere (see Theorem 5.4) with a Γ distribution

$$q_r(r) = \frac{r^{u-1}e^{-\frac{r}{s}}}{s^u\Gamma(u)}$$

with shape parameter u and scale parameter s:

$$q(r,\mathbf{u}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}s^u\Gamma(u)\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} r^{u-1}e^{-\frac{r}{s}}.$$

Transforming $q(r,\mathbf{u})$ back into Euclidean coordinates yields

$$p(\mathbf{y}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^n s^u\Gamma(u)\Gamma^n\left(\frac{1}{p}\right)} ||\mathbf{y}||_p^{u-n} e^{-\frac{||\mathbf{y}||_p}{s}},$$

where we have to remember that $q(r,\mathbf{u}) = 2p(\mathbf{y}(\mathbf{u},r))r^{n-1}\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1-p}{p}} d\mathbf{u}dr$, since a given $\mathbf{u} = (u_1,...,u_{n-1})^\top$ can correspond to $y_n = \pm r\left(1 - \sum_{i=1}^{n-1}|u_i|^p\right)^{\frac{1}{p}}$, i.e. the transformation $(y_1,...,y_n) \mapsto (r,u_1,...,u_{n-1})$ is only one-to-one on the half-p-sphere.

In order to estimate the parameters u and s from data $X = \{\mathbf{y}_1,...,\mathbf{y}_m\}$, we first compute the radial components of all datapoints $\{r_1,...,r_m\} = \{||\mathbf{y}_1||_p,...,||\mathbf{y}_m||_p\}$. Then, we use that

$$\begin{aligned} \mathbb{E}[\mathsf{R}] &= su \\ \mathbb{V}[\mathsf{R}] &= us^2 \end{aligned}$$

and, therefore,

$$\begin{aligned} s &= \frac{\mathbb{V}[\mathsf{R}]}{\mathbb{E}[\mathsf{R}]} \\ u &= \frac{\mathbb{E}^2[\mathsf{R}]}{\mathbb{V}[\mathsf{R}]}. \end{aligned}$$

Using the maximum likelihood estimatior for $\mathbb{E}[\mathsf{R}]$ and $\mathbb{V}[\mathsf{R}]$ yields the estimations for u and s

$$\begin{aligned} \hat{\mathbb{E}}[\mathsf{R}] &= \frac{1}{m}\sum_{i=1}^m r_i \\ \hat{\mathbb{V}}[\mathsf{R}] &= \hat{\mathbb{E}}[\mathsf{R}^2] - \hat{\mathbb{E}}^2[\mathsf{R}]. \end{aligned}$$

5.3.2 The p-spherically symmetric distribution with log-normal distributed radial component $SC_p[\mu, \log \mathcal{N}[\mu, \sigma]]$

If we choose the log normal distribution $q_r(r) = \frac{1}{r\sigma\sqrt{2\pi}} e^{-\frac{(\log r - \mu)^2}{2\sigma^2}}$ as radial distribution, we obtain

$$\begin{aligned} q(u_1, ..., u_{n-1}, r) &= q_r(r) \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} \\ &= \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{r\sigma\sqrt{2\pi}2^{n-1}\Gamma^n\left(\frac{1}{p}\right)} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}} e^{-\frac{(\log r - \mu)^2}{2\sigma^2}}. \end{aligned}$$

Transforming the density back to Cartesian coordinates yields

$$p(\mathbf{y}) = \frac{p^{n-1}\Gamma\left(\frac{n}{p}\right)}{||\mathbf{y}||_p^n \sigma\sqrt{2\pi}2^n\Gamma^n\left(\frac{1}{p}\right)} e^{-\frac{(\log ||\mathbf{y}||_p - \mu)^2}{2\sigma^2}}.$$

The maximum likelihood estimates of μ and σ are the usual estimates for the log-normal distribution and are given by

$$\begin{aligned} \mu &= \log(\hat{\mathbb{E}}[R]) - \frac{1}{2}\log\left(1 + \frac{\hat{\mathbb{V}}[R]}{\hat{\mathbb{E}}^2[R]}\right) \\ \sigma^2 &= \log\left(1 + \frac{\hat{\mathbb{V}}[R]}{\hat{\mathbb{E}}^2[R]}\right). \end{aligned}$$

5.3.3 The p-generalized Normal distribution $\mathcal{N}_p[\sigma^2]$

The p-generalized Normal distribution is obtained by choosing Y to be a collection of n i.i.d. random variables y_i, each distributed according to the exponential power distribution

$$\begin{aligned} \mathsf{y}_i \sim p(x) &= \frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2} e^{-\frac{|x|^p}{2\sigma^2}} \\ \mathsf{Y} \sim p(\mathbf{y}) = \prod_{i=1}^{n} p(y_i) &= \left(\frac{p}{\Gamma\left(\frac{1}{p}\right)(2\sigma^2)^{\frac{1}{p}}2}\right)^n e^{-\frac{\sum_{i=1}^{n}|y_i|^p}{2\sigma^2}} \end{aligned}$$

Since $p(\mathbf{x})$ has the form $g(||\mathbf{x}||_p^p)$, it is a proper p-spherically symmetric distribution due to Theorem 5.8. Note, that for the case of $p = 2$, the p-generalized Normal distribution reduces to a multivariate isotropic Gaussian. In order to compute the contrast gain control function, we need to compute the radial distribution q_r of $p(\mathbf{x})$. Transforming p according to Lemma 5.3 yields

$$q(r, \mathbf{u}) = \frac{p^n r^{n-1}}{\Gamma^n\left(\frac{1}{p}\right)(2\sigma)^{\frac{n}{p}}2^{n-1}} e^{-\frac{r^p}{2\sigma}} \left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}.$$

By integrating over $\mathbf{u}$ (see lemma 5.5 how to carry out the integral) we get

$$q_r(r) = \frac{p\, r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} e^{-\frac{r^p}{2\sigma^2}}$$

In order to estimate the scale parameter σ from data $X = \{r_1, ..., r_m\} = \{||\mathbf{x}_1||_p, ..., ||\mathbf{x}_m||_p\}$, we carry out the usual procedure for maximum likelihood estimation and obtain

$$
\begin{aligned}
\frac{d}{d\sigma}\log q_r(r) &= \frac{d}{d\sigma}\left(-\frac{2n}{p}\log(\sigma) - \frac{r^p}{2\sigma^2}\right) \\
&= \frac{r^p p - 2n\sigma^2}{p\sigma^3} \\
\frac{d}{d\sigma}\sum_{i=1}^{m}\log q_r(r_i) &= \sum_{i=1}^{m}\frac{r_i^p p - 2n\sigma^2}{p\sigma^3} \\
&\stackrel{!}{=} 0.
\end{aligned}
$$

This yields

$$
\hat{\sigma} = \sqrt{\frac{p}{2mn}\sum_{i=1}^{m} r_i^p}.
$$

For the transformation of the radial component, we will also need the cumulative distribution function of

$$
q_r(r) = \frac{p\, r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} e^{-\frac{r^p}{2\sigma^2}}.
$$

It can be computed via simple integration with the substitution $y = \frac{r^p}{2\sigma^2}$

$$
\begin{aligned}
\mathcal{F}_{\mathcal{N}_p}(a) &= \int_0^a \frac{p r^{n-1}}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} e^{-\frac{r^p}{2\sigma^2}} dr \\
&= \frac{p}{\Gamma\left(\frac{n}{p}\right)(2\sigma^2)^{\frac{n}{p}}} \int_0^a r^{n-1} e^{-\frac{r^p}{2\sigma^2}} dr \\
&= \frac{1}{\Gamma\left(\frac{n}{p}\right)} \int_0^{\frac{a^p}{2\sigma^2}} y^{\frac{n}{p}-1} e^{-y} dy \\
&= \frac{\Gamma\left(\frac{n}{p}, \frac{a^p}{2\sigma^2}\right)}{\Gamma\left(\frac{n}{p}\right)},
\end{aligned}
$$

where $\Gamma(z, b) = \int_0^b y^{z-1} e^{-y} dy$ is the incomplete Γ-function.

References

[1] Steven F. Arnold and James Lynch. On ali's characterization of the spherical normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):49–51, 1982. 2, 7

[2] A.J. Bell and T.J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res.*, 37(23):3327–38, 1997. 1

[3] M. Bethge. Factorial coding of natural images: How effective are linear model in removing higher-order dependencies? *J. Opt. Soc. Am. A*, 23(6):1253–1268, June 2006. 1, 12, 15

[4] G. J. Brelstaff, A. Parraga, T. Troscianko, and D. Carr. Hyperspectral camera system: acquisition and analysis. In B. J. Lurie, J. J. Pearson, and E. Zilioli, editors, *Proceedings of SPIE*, volume 2587, pages 150–159, 1995. The database can be downloaded from: `http://psy223.psy.bris.ac.uk/hyper/`. 11

[5] T.M. Cover and J.A. Thomas. *Elements of information theory*. J. Wiley & Sons, New York, 1991. 10

[6] J. Eichhorn, F. Sinz, and M. Bethge. Simple cell coding of natural images in V1: How much use is orientation selectivity? *(in preparation)*, -:–, 2008. 1, 3, 5, 11, 15

[7] A.K. Gupta and D. Song. l_p-norm spherical distribution. *Journal of Statistical Planning and Inference*, 60:241–260, 1997. 6, 8, 20, 21, 22

[8] D.J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992. 6

[9] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001. 12

[10] Aapo Hyvärinen and Urs Köster. Complex cell pooling and the statistics of natural images. *Network*, 18:81–100, 2007. 6

[11] M. Kac. On a characterization of the normal distribution. *American Journal of Mathematics*, 61(3):726–728, July 1939. 2, 7

[12] T.-W. Lee, T. Wachtler, and T. J. Sejnowski. Color opponency is an efficient representation of spectral properties in natural scenes. *Vision Res*, 42(17):2095–2103, Aug 2002. 11

[13] B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:560–561, 1996. 1

[14] O. Schwartz and E. P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, August 2001. 1, 6

[15] D. Song and A.K. Gupta. l_p-norm uniform distribution. *Proceedings of the American Mathematical Society*, 125:595–601, 1997. 8, 16, 17, 18

[16] B. E. Usevitch. A tutorial on modern lossy wavelet image compression: foundations of jpeg 2000. *IEEE Signal Processing Magazine*, 18(5):22–35, 2001. 5

[17] T Wachtler, T W Lee, and T J Sejnowski. Chromatic structure of natural scenes. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 18:65–77, 2001. PMID: 11152005. 11