L_p-NESTED SYMMETRIC DISTRIBUTIONS

FABIAN SINZ, EERO SIMONCELLI, MATTHIAS BETHGE

1. Introduction

A important part in statistical analysis of data is to find a class of models that is flexible and rich enough to model the regularities in the data, but at the same time exhibits enough symmetry and structure itself to still be computationally and analytically tractable. One special way of introducing such a symmetry is to fix the general form of the isodensity contour lines. This approach was taken by [2] who modelled the contour lines by the level sets of a positively homogeneous function of degree one. Unfortunately, in the general case it is hard to derive the normalization constant for an arbitrary such function. For a special kind of ν-spherical distributions, the L_p-spherically symmetric distributions [5; 3] this problem becomes tractable by restricting the contour lines to L_p-spheres, but at the prize of introducing permutation symmetry. The L_p-spherically symmetric distribution itself generalize the class of L_2-spherically symmetric distributions which exhibit rotational symmetry [4; 1]. In some cases permutation or even rotational symmetry might be an appropriate assumption for the data. However, in other cases such symmetries might actually make the model miss important structure present in the data.

Here, we present a generalization of the class of L_p-spherically symmetric distribution within the class of ν-spherical distributions. Instead of using a single L_p-norm to define the contour of the density, we use nested L_p-norms where the coefficients, the L_p-norm is computed over, can be L_p-norms themselves—with possibly different p. This preserves positive homogeneity and replaces permutational invariance with invariance under reflection at the coordinate axes. Due to the nested structure, we call this new class of distributions *L_p-nested symmetric distributions*. As we demonstrate below, this construction still bears enough structure to define polar-like coordinates similar to those of [6; 3] and thereby to compute the normalization constant of the distribution given an arbitrary univariate distribution on the function values. By that construction, we can leverage most important properties of the L_p-spherically symmetric distributions to the L_p-nested distributions.

The remaining part of the paper is structured as follows: In section 2 we introduce some helpful nomenclature and define L_p-nested functions. In section 3 we define coordinates in the spirit of [3] and derive the Jacobian of the determinant. In section 4 we introduce the uniform distribution on the L_p-nested unit sphere which allows us to leverage some of the results of [3] to L_p-nested symmetric distributions in section 5. In section 6 we derive a sampling scheme for L_p-nested symmetric distributions. We conclude by presenting a potential application for the class of L_p-nested symmetric distributions.

Date: October 30, 2009.

2. Nomenclature and Definitions

Definition 2.1 (L_p-nested functions)**.** We call a function $f : \mathbb{R}^n \to \mathbb{R}_0^+$ *L_p-nested* if f fulfills the following recursive definition:

(i) The function $f : \mathbb{R}^n \to \mathbb{R}$ is the L_p-norm of its ℓ children $(f_1(\mathbf{x}_1), ..., f_\ell(\mathbf{x}_\ell))^\top$:
$$f(\mathbf{x}) = ||(f_1(\mathbf{x}_1), ..., f_\ell(\mathbf{x}_\ell))^\top||_p,$$
where the $\mathbf{x}_j \in \mathbb{R}^{n_j}$ are a partition of the vector $\mathbf{x}$ into ℓ parts.

(ii) The children f_i are either L_p-nested functions themselves or compute the absolute value of a single coefficient x_i, i.e. $f_j(\mathbf{x}_j) = |x_i|$ if and only if $\mathbf{x}_j = x_i \in \mathbb{R}$.

This gives rise to a tree structure of f which is depicted in Figure 1. Note, that every L_p-nested function is positively homogeneous by construction. In order to present results for arbitrary L_p-nested functions, we start by introducing some helpful notation.

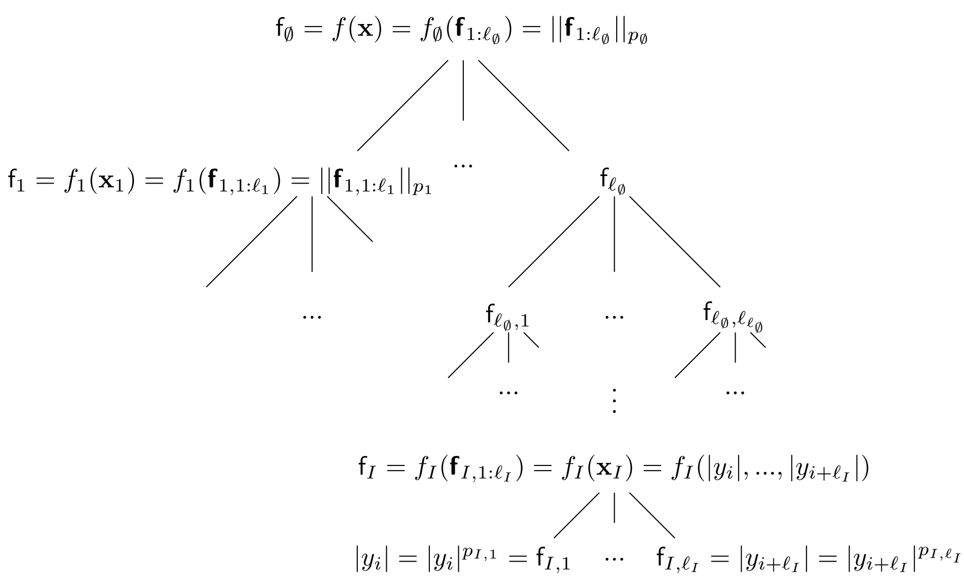


Figure 1. **Tree structure associated with an L_p-nested function f:** Every parent node I gets its value f_I by computing the L_{p_I}-norm of the values of its children $\mathbf{f}_{I,1:\ell_I}$. The leafs of the tree correspond to the (absolute values) of the coefficients in the vector $\mathbf{x}$. The values of the p at the leaf nodes are set to the value $p = 1$ by definition, e.g. $p_{I,1} = ... = p_{I,\ell_I} = 1$ in the diagram.

Definition 2.2 (Notation and Conventions for L_p-nested functions)**.** We use the following notational conventions:

(i) We use multi-indices to denote the different nodes of the tree corresponding to an L_p-nested function f. The function f itself corresponds to the root node and is denoted by $f_\emptyset$. The functions corresponding to its children are denoted by $f_1, ..., f_{\ell_\emptyset}$. The children of the i^{th} child are denoted by $f_{i,1}, ..., f_{i,\ell_i}$. In this manner, an index is added for each layer of the tree.

(ii) We always use the letter "ℓ" to denote the total amount of children of a node.

(iii) For notational convenience, we assign a p to each of the leaf nodes (i.e. the absolute values $|x_i|$) but fix their values to $p = 1$ by definition.

(iv) For the sake of compact notation, we denote a list of indices with a single multi-index $I = i_1, ..., i_\ell$. The range of the single indices and the length of the multi-index should be clear from the context. Multi-indices are always denoted by upper-case letters. A concatenation I, k of a multi-index I with another index k corresponds to adding k to the index list, i.e. $I, k = i_1, ..., i_m, k$. We use the convention that $I, \emptyset = I$.

(v) Those coefficients of the vector $\mathbf{x}$ that correspond to leafs of the subtree under a node with the index I are denoted by $\mathbf{x}_I$. The number of leafs in a subtree under a node I is denoted by n_I. If I denotes a leaf then $n_I = 1$.

(vi) The L_p-nested function associated with the subtree under a node I is denoted by

$$f_I(\mathbf{x}_I) = ||(f_{I,1}(\mathbf{x}_{I,1}), ..., f_{I,\ell_I}(\mathbf{x}_{I,\ell_I}))^\top||_{p_I}.$$

We use sans-serif font to denote the function value $\mathsf{f}_I = f_I(\mathbf{x}_I)$ of a subtree I. In many cases we use f_I and $f_I(\mathbf{x}_I)$ interchangeably. Whether f_I is to be considered as a function of its children or merely the value of the node I should always be clear from the context.

A vector with the function values of the children of I is denoted with bold sans-serif font and the following index-list notation:

$$\begin{aligned} f_I(\mathbf{x}_I) &= ||(f_{I,1}(\mathbf{x}_{I,1}), ..., f_{I,\ell_I}(\mathbf{x}_{I,\ell_I}))^\top||_{p_I} \\ &= ||(\mathsf{f}_{I,1}, ..., \mathsf{f}_{I,\ell_I})^\top||_{p_I} \\ &= ||\boldsymbol{\mathsf{f}}_{I,1:\ell_I}||_{p_I} \end{aligned}$$

(vii) The function computing the value of the ℓ^{th} —and therefore by convention last—child of a node I when fixing the value f_I of that node, is denoted by

$$\begin{aligned} g_{I,\ell_I}(\mathsf{f}_I, \mathsf{f}_{I,1}, ..., \mathsf{f}_{I,\ell_I-1}) &= \left(\mathsf{f}_I^{p_I} - \sum_{k=1}^{\ell_I-1} \mathsf{f}_{I,k}^{p_I}\right)^{\frac{1}{p_I}} \\ &= g_{I,\ell_I}(\boldsymbol{\mathsf{f}}_{I,\emptyset:\ell_I-1}) \\ &= \mathsf{g}_{I,\ell_I}. \end{aligned}$$

Notice the small but important difference that the value f_I depends only on the values of its children $\mathsf{f}_{I,1}, ..., \mathsf{f}_{I,\ell_I}$, while the value g_{I,ℓ_I} depends on the value of its neighbors $\mathsf{f}_{I,1}, ..., \mathsf{f}_{I,\ell_I-1}$ and its parent $\mathsf{f}_I = \mathsf{f}_{I,\emptyset}$.

(viii) Vectors in $\mathbb{R}^n$ that lie on the L_p-nested unit sphere, i.e. that fulfill $f(\mathbf{u}) = 1$ are denoted by the letter $\mathbf{u}$.

Vectors $\tilde{\mathbf{u}} \in \mathbb{R}^{\ell_I}$ that lie on the L_{p_I} unit sphere associated with the inner node I, i.e. that fulfill $\boldsymbol{\mathsf{f}}_{I:1:\ell_I} = \mathsf{f}_I \tilde{\mathbf{u}}$ are denoted by the letter $\tilde{\mathbf{u}}$. Note that the coordinates $\mathbf{u}$ and $\tilde{\boldsymbol{u}}$ are different: $f_I(\tilde{\boldsymbol{u}}) = 1$ while $f_I(\mathbf{u}_I) \leq 1$.

When defining polar-like coordinates in section 3 only all but the last coefficients of $\mathbf{u}$ or $\tilde{\boldsymbol{u}}$ are needed, since the last can be computed from the remaining ones. We often still denote this shorter vectors by $\mathbf{u}$ or $\tilde{\boldsymbol{u}}$. The actual dimensionality should be clear from the context.

Let us demonstrate the above definitions with a simple example.

Example 2.1. Consider the L_p-nested function

$$\begin{aligned}
f(\mathbf{x}) &= \left((|x_1|^{p_1} + |x_2|^{p_1})^{\frac{p_\emptyset}{p_1}} + |x_3|^{p_\emptyset}\right)^{\frac{1}{p_\emptyset}} \\
&= \left(\left(\left(\mathsf{f}_{1,1}^{p_{1,1}}\right)^{\frac{p_1}{p_{1,1}}} + \left(\mathsf{f}_{1,2}^{p_{1,2}}\right)^{\frac{p_1}{p_{1,2}}}\right)^{\frac{p_\emptyset}{p_1}} + \left(\mathsf{f}_2^{p_2}\right)^{\frac{p_\emptyset}{p_2}}\right)^{\frac{1}{p_\emptyset}} \\
&= \left(f_1\left(\mathbf{f}_{1,1:2}\right)^{p_\emptyset} + f_2\left(\mathsf{f}_{2,1}\right)^{p_\emptyset}\right)^{\frac{1}{p_\emptyset}} \\
&= f_\emptyset\left(\mathbf{f}_{1:2}\right)
\end{aligned}$$

with $\ell_\emptyset = 2$, $\ell_1 = 2$ and $p_{1,1} = p_{1,2} = p_2 = 1$ by definition. Resolving $f(x_1, x_2, x_3) = a$ for $|x_3|$ yields the functions g

$$\begin{aligned}
|x_3| &= \mathsf{g}_2 \\
&= g_2\left(\mathsf{f}_\emptyset, \mathsf{f}_1\right) \\
&= \left(\mathsf{f}_\emptyset^{p_\emptyset} - \mathsf{f}_1^{p_\emptyset}\right)^{\frac{1}{p_\emptyset}} \\
&= \left(a^{p_\emptyset} - f_1\left(\mathbf{f}_{1,1:2}\right)^{p_\emptyset}\right)^{\frac{1}{p_\emptyset}} \\
&= \left(a^{p_\emptyset} - (|x_1|^{p_1} + |x_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1}{p_\emptyset}}
\end{aligned}$$

3. L_p-nested Coordinate Transformation and the Determinant of its Jacobian

The most important consequence of the positive homogeneity of f is that it can be used to normalized vectors and, by that property, to generalize the polar-like coordinates using L_p-norms of [3].

Definition 3.1 (Polar-like Coordinates)**.** We define the following polar-like coordinates for a vector $\mathbf{x} \in \mathbb{R}^n$:

$$\begin{aligned}
u_i &= \frac{x_i}{f(\mathbf{x})} \text{ for } i = 1, ..., n-1 \\
r &= f(\mathbf{x}).
\end{aligned}$$

The inverse coordinate transformation is given by

$$\begin{aligned}
x_i &= r u_i \text{ for } i = 1, ..., n-1 \\
x_n &= r \Delta_n u_n
\end{aligned}$$

where we define $\Delta_n = \operatorname{sgn} x_n$ and u_n to be the value of the leaf corresponding to $|x_n|$ when setting $\mathsf{f}_\emptyset = 1$.

The definition of the coordinates is basically equivalent to that of [3] with the difference that the L_p-norm is replaced by an L_p-nested function. Just as in the case of L_p-spherical coordinates, it will turn out that the Jacobian of the coordinate transformation does not depend on the value of Δ_n. This is basically a consequence of the invariance under reflection at the coordinate axes.

The remaining part of this section will be devoted to computing the determinant of the Jacobian. We start by stating the general form of the determinant in terms of the partial derivatives $\frac{\partial u_n}{\partial u_k}$, u_k and r. Afterwards we demonstrate that those partial derivatives have a special form and that most of them cancel in the Laplace expansion of the determinant.

Lemma 3.1 (Determinant of the Jacobian)**.** *Let r and $\mathbf{u}$ be defined as in Definition (3.1). The general form of the determinant of the Jacobian $\mathcal{J}$ of the inverse coordinate transformation is given by*

$$|\det \mathcal{J}| = r^{n-1}\left(-\sum_{k=1}^{n-1}\frac{\partial u_n}{\partial u_k}\cdot u_k + u_n\right). \tag{1}$$

Proof. The partial derivatives of the inverse coordinate transformation are given by:

$$\begin{aligned}
\frac{\partial}{\partial u_k} y_i &= \delta_{ik} r \text{ for } 1 \leq i, k \leq n-1 \\
\frac{\partial}{\partial u_k} y_n &= \Delta_n r \frac{\partial u_n}{\partial u_k} \text{ for } 1 \leq k \leq n-1 \\
\frac{\partial}{\partial r} y_i &= u_i \text{ for } 1 \leq i \leq n-1 \\
\frac{\partial}{\partial r} y_n &= \Delta_n u_n.
\end{aligned}$$

Therefore, the structure of the Jacobian is given by

$$\mathcal{J} = \begin{pmatrix} r & \dots & 0 & u_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & r & u_{n-1} \\ \Delta_n r \frac{\partial u_n}{\partial u_1} & \dots & \Delta_n r \frac{\partial u_n}{\partial u_{n-1}} & \Delta_n u_n \end{pmatrix}.$$

Since we are only interested in the absolute value of the determinant and since $\Delta_n \in \{-1, 1\}$, we can factor out Δ_n and drop it. Furthermore, we can factor out r from the first $n-1$ columns which yields

$$|\det \mathcal{J}| = r^{n-1}\left|\det \begin{pmatrix} 1 & \dots & 0 & u_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & u_{n-1} \\ \frac{\partial u_n}{\partial u_1} & \dots & \frac{\partial u_n}{\partial u_{n-1}} & u_n \end{pmatrix}\right|.$$

Now we can use Laplace formula to expand the determinant with respect to the last column. For that purpose, let $\mathcal{J}_i$ denote the matrix which is obtained by deleting the last column and the ith row from $\mathcal{J}$. This matrix has the following structure

$$\mathcal{J}_i = \begin{pmatrix} 1 & & & 0 & & & \\ & \ddots & & 0 & & & \\ & & 1 & 0 & & & \\ & & & \vdots & 1 & & \\ & & & 0 & & \ddots & \\ & & & 0 & & & 1 \\ \frac{\partial u_n}{\partial u_1} & & & \frac{\partial u_n}{\partial u_i} & & & \frac{\partial u_n}{\partial u_{n-1}} \end{pmatrix}.$$

We can transform $\mathcal{J}_i$ into a lower triangular matrix by moving the column with all zeros and $\frac{\partial u_n}{\partial u_i}$ bottom entry to the rightmost column of $\mathcal{J}_i$. Each swapping of two columns introduces a factor of -1. In the end, we can compute the value of $\det \mathcal{J}_i$ by simply taking the product of the diagonal entries and obtain $\det \mathcal{J}_i = (-1)^{n-1-i}\frac{\partial u_n}{\partial u_i}$. This yields

$$\begin{aligned}
|\det \mathcal{J}| &= r^{n-1}\left(\sum_{k=1}^{n}(-1)^{n+k}u_k \det \mathcal{J}_k\right)\\
&= r^{n-1}\left(\sum_{k=1}^{n-1}(-1)^{n+k}u_k \det \mathcal{J}_k + (-1)^{2n}\frac{\partial y_n}{\partial r}\right)\\
&= r^{n-1}\left(\sum_{k=1}^{n-1}(-1)^{n+k}u_k(-1)^{n-1-k}\frac{\partial u_n}{\partial u_k} + u_n\right)\\
&= r^{n-1}\left(-\sum_{k=1}^{n-1}u_k\frac{\partial u_n}{\partial u_k} + u_n\right).
\end{aligned}$$

□

For a given L_p-nested function f, the terms r, u_k and $\frac{\partial u_n}{\partial u_k}$ needed to compute the determinant with equation (1) can be computed easily. However, as already mentioned, most constituents of those terms cancel each other as the following example demonstrates. We urge the reader to follow the next example as it contains the important ideas for the general case below.

Example 3.1. Consider the function from the previous example

$$f(\mathbf{y}) = \left(\left(|x_1|^{p_1} + |x_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}} + |x_3|^{p_\emptyset}\right)^{\frac{1}{p_\emptyset}}.$$

Setting $\mathbf{u} = \frac{\mathbf{x}}{f(\mathbf{x})}$ and solving for u_3 yields

$$\begin{aligned}
& f(\mathbf{u}) = 1\\
\Leftrightarrow\ \ & u_3 = \left(1 - \left(|u_1|^{p_1} + |u_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1}{p_\emptyset}}
\end{aligned}$$

Now, let G_2 and F_1 denote

$$\begin{aligned}
\mathsf{G}_2 &= \left(1 - \left(|u_1|^{p_1} + |u_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1-p_\emptyset}{p_\emptyset}}\\
\mathsf{F}_1 &= \left(|u_1|^{p_1} + |u_2|^{p_1}\right)^{\frac{p_\emptyset - p_1}{p_1}}.
\end{aligned}$$

Essentially, G_2 and F_1 are terms that evolve from the application from the chain rule when computing the partial derivative. G_2 originates from using the chain rule upwards in the tree and F_1 from using it downwards. The indices correspond the multi-indices of the respective nodes. Computing the derivative yields

$$\begin{aligned}
\frac{\partial u_3}{\partial u_k} &= \frac{\partial}{\partial u_k}\left(1-(|u_1|^{p_1}+|u_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1}{p_\emptyset}} \\
&= \frac{1}{p_\emptyset}\mathsf{G}_2\cdot-\frac{\partial}{\partial u_k}\left(|u_1|^{p_1}+|u_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}} \\
&= \frac{1}{p_\emptyset}\frac{p_\emptyset}{p_1}\mathsf{G}_2\cdot-\mathsf{F}_1\frac{\partial}{\partial u_k}|u_k|^{p_1} \\
&= -\mathsf{G}_2\mathsf{F}_1\Delta_k {u_k}^{p_1-1}.
\end{aligned}$$

By inserting the results in equation (1) we obtain

$$\begin{aligned}
\frac{1}{r^2}|\mathcal{J}| &= -\sum_{k=1}^{2}\frac{\partial u_n}{\partial u_k}\cdot u_k + u_3 \\
&= \sum_{k=1}^{2}\mathsf{G}_2\mathsf{F}_1|u_k|^{p_1}+u_3 \\
&= \mathsf{G}_2\left(\sum_{k=1}^{2}\mathsf{F}_1|u_k|^{p_1}+\mathsf{G}_2^{-1}\left(1-(|u_1|^{p_1}+|u_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1}{p_\emptyset}}\right)
\end{aligned}$$

$$\begin{aligned}
&= \mathsf{G}_2\left(\sum_{k=1}^{2}\mathsf{F}_1|u_k|^{p_1}+\left(1-(|u_1|^{p_1}+|u_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right)^{-\frac{1-p_\emptyset}{p_\emptyset}}\left(1-(|u_1|^{p_1}+|u_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right)^{\frac{1}{p_\emptyset}}\right) \\
&= \mathsf{G}_2\left(\sum_{k=1}^{2}\mathsf{F}_1|u_k|^{p_1}+1-(|u_1|^{p_1}+|u_2|^{p_1})^{\frac{p_\emptyset}{p_1}}\right) \\
&= \mathsf{G}_2\left(\mathsf{F}_1\sum_{k=1}^{2}|u_k|^{p_1}+1-\mathsf{F}_1\mathsf{F}_1^{-1}\left(|u_1|^{p_1}+|u_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}}\right) \\
&= \mathsf{G}_2\left(\mathsf{F}_1\sum_{k=1}^{2}|u_k|^{p_1}+1-\mathsf{F}_1\left(|u_1|^{p_1}+|u_2|^{p_1}\right)^{-\frac{p_\emptyset-p_1}{p_1}}\left(|u_1|^{p_1}+|u_2|^{p_1}\right)^{\frac{p_\emptyset}{p_1}}\right) \\
&= \mathsf{G}_2\left(\mathsf{F}_1\sum_{k=1}^{2}|u_k|^{p_1}+1-\mathsf{F}_1\sum_{k=1}^{2}|u_k|^{p_1}\right) \\
&= \mathsf{G}_2.
\end{aligned}$$

In the example above, the terms from using the chain rule downwards in the tree canceled while the terms from using the chain rule upwards remained. It will turn out that this is true in general. Before we state the general equation we introduce a short notation for the terms that cancel and for those that remain.

Definition 3.2. Let $I = i_1, ..., i_{r-1}$. In the following, we denote

$$\mathsf{G}_{I,\ell_I} = \mathsf{g}_{I,\ell_I}^{p_{I,\ell_I}-p_I}$$

$$= \left(\mathsf{g}_I^{p_I} - \sum_{j=1}^{\ell-1} \mathsf{f}_{I,j}^{p_I}\right)^{\frac{p_{I,\ell_I}-p_I}{p_I}} \tag{2}$$

and

$$\mathsf{F}_{I,i_r} = \mathsf{f}_{I,i_r}^{p_I - p_{I,i_r}}$$

$$= \left(\sum_{k=1}^{\ell} \mathsf{f}_{I,i_r,k}^{p_{I,i_r}}\right)^{\frac{p_I - p_{I,i_r}}{p_{I,i_r}}}.$$

Note that the term F_{I,i_r} is a function of its children while G_{I,i_r} is a function of the parent node and all but the last children.

Before going on, let us quickly outline the essential mechanism when taking the chain rule $\frac{u_n}{u_q}$. Imagine the tree corresponding to f. By definition u_n is the rightmost leaf of the tree. Let L, ℓ_L be the multi-index of u_n. The calculation of $\frac{\partial u_n}{\partial u_q}$ will obviously involve heavy usage of the chain rule. As in the example, the chain rule starts at the leaf u_n ascends in the the tree until it reaches the lowest node whose subtree contains both, u_n and u_q. At this point, it starts descending the tree until it reaches the leaf u_q. Depending on whether the chain rule ascends or descends, two different forms of derivatives occur: At $u_n = \mathsf{g}_{L,\ell_L}$ the chain rule will start ascending by taking the derivative of the term

$$\mathsf{g}_{L,\ell_L} = \left(\mathsf{g}_L^{p_L} - \sum_{k=1}^{\ell_L-1} \mathsf{f}_{L,k}^{p_L}\right)^{\frac{1}{p_L}}$$

which will produce a G-term and move the chain rule one step up in the tree.

If the parent of u_n is already the lowest node whose subtree contains u_q a u_n, then u_q is hidden somewhere in the f-terms and the g-term is independent of u_q. However, if this node is still higher in the tree, then the situation is reversed, i.e. the f-terms are independent of u_q which is hidden in the g-term. When going on, the chain rule will produce a G-term when ascending the tree and an F-term when descending. The situation is depicted in Figure 2. The next lemma states a few helpful properties of the F- and G-terms.

Lemma 3.2. *Let $I = i_1, ..., i_{r-1}$ and f_{I,i_r} be any node of the tree associated with an L_p-nested function f. Then the following recursions hold for the derivatives of $\mathsf{g}_{I,i_r}^{p_{I,i_r}}$ and $\mathsf{f}_{I,i_r}^{p_I}$ w.r.t u_q: If u_q is not in the subtree under the node I, i_r, i.e. $u_k \not\in \mathsf{f}_{I,i_r}$, then (remember that $p_{I,i_r} = 1$ for leaf nodes by notational convention):*

$$\frac{\partial}{\partial u_q} \mathsf{f}_{I,i_r}^{p_I} = 0$$

and

$$\frac{\partial}{\partial u_q} \mathsf{g}_{I,i_r}^{p_{I,i_r}} = \frac{p_{I,i_r}}{p_I} \mathsf{G}_{I,i_r} \cdot \begin{cases} \frac{\partial}{\partial u_q} \mathsf{g}_I^{p_I} & \text{if } u_q \in \mathsf{g}_I \\ -\frac{\partial}{\partial u_q} \mathsf{f}_{I,j}^{p_I} & \text{if } u_q \in \mathsf{f}_{I,j} \end{cases}$$

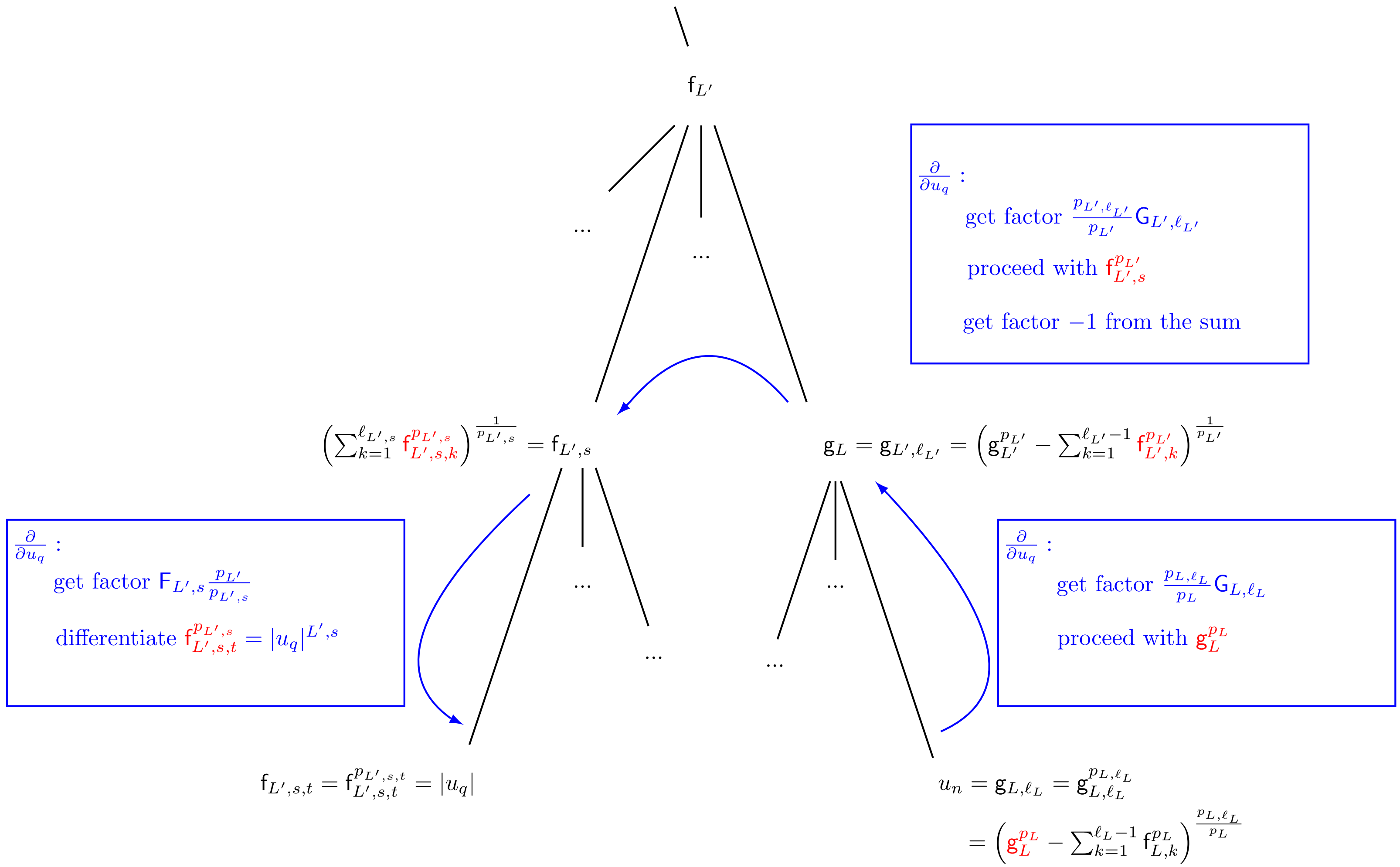


FIGURE 2. **Example scheme for the steps of the chain rule when calculating $\frac{\partial u_n}{\partial u_q}$:** Note that, $L = L', \ell_{L'}$, that the $p_{L,\ell_L}, p_{L,s,t}$ at the leafs are equal to one by definition and that succeeding ratios of the p cancel each other.

for $u_q \in \mathsf{f}_{I,j}$ *and* $u_q \notin \mathsf{f}_{I,k}$ *for* $k \neq j$. *Otherwise*

$$\frac{\partial}{\partial u_q}\mathsf{g}_{I,i_r}^{p_{I,i_r}} = 0$$

and

$$\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r}^{p_I} = \frac{p_I}{p_{I,i_r}}\mathsf{F}_{I,i_r}\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r,s}^{p_{I,i_r}}$$

for $u_q \in \mathsf{f}_{I,i_r,s}$ *and* $u_q \notin \mathsf{f}_{I,i_r,k}$ *for* $k \neq s$.

Proof. Both first equations are obvious, since only those nodes have a non-zero derivative for which the subtree actually depends on u_q. The second equations can be seen by computation

$$\begin{aligned}
\frac{\partial}{\partial u_q}\mathsf{g}_{I,i_r}^{p_{I,i_r}} &= p_{I,i_r}\mathsf{g}_{I,i_r}^{p_{I,i_r}-1}\frac{\partial}{\partial u_q}\mathsf{G}_{I,i_r}\\
&= p_{I,i_r}\mathsf{g}_{I,i_r}^{p_{I,i_r}-1}\frac{\partial}{\partial u_q}\left(\mathsf{g}_I^{p_I} - \sum_{j=1}^{\ell_I-1}\mathsf{f}_{I,j}^{p_I}\right)^{\frac{1}{p_I}}\\
&= \frac{p_{I,i_r}}{p_I}\mathsf{g}_{I,i_r}^{p_{I,i_r}-1}\mathsf{g}_{I,i_r}^{1-p_I}\frac{\partial}{\partial u_q}\left(\mathsf{g}_I^{p_I} - \sum_{j=1}^{\ell_I-1}\mathsf{f}_{I,j}^{p_I}\right)\\
&= \frac{p_{I,i_r}}{p_I}\mathsf{G}_{I,i_r}\cdot\begin{cases}\frac{\partial}{\partial u_q}\mathsf{g}_I^{p_I} & \text{if } u_q \in \mathsf{g}_I\\[2ex] -\frac{\partial}{\partial u_q}\mathsf{f}_{I,j}^{p_I} & \text{if } u_q \in \mathsf{f}_{I,j}\end{cases}
\end{aligned}$$

Similarly

$$\begin{aligned}
\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r}^{p_I} &= p_I\mathsf{f}_{I,i_r}^{p_I-1}\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r}\\
&= p_I\mathsf{f}_{I,i_r}^{p_I-1}\frac{\partial}{\partial u_q}\left(\sum_{k=1}^{\ell_{I,i_r}}\mathsf{f}_{I,i_r,k}^{p_{I,i_r}}\right)^{\frac{1}{p_{I,i_r}}}\\
&= \frac{p_I}{p_{I,i_r}}\mathsf{f}_{I,i_r}^{p_I-1}\mathsf{f}_{I,i_r}^{1-p_{I,i_r}}\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r,s}^{p_{I,i_r}}\\
&= \frac{p_I}{p_{I,i_r}}\mathsf{F}_{I,i_r}\frac{\partial}{\partial u_q}\mathsf{f}_{I,i_r,s}^{p_{I,i_r}}
\end{aligned}$$

for $u_k \in \mathsf{f}_{I,i_r,s}$. □

The next lemma states the form of the derivative $\frac{\partial u_n}{\partial u_q}$ in terms of the G- and F-terms.

Lemma 3.3. *Let* $|u_q| = \mathsf{f}_{\ell_1,...,\ell_r,i_1,...,i_t}$, $|u_n| = \mathsf{f}_{\ell_1,...,\ell_d}$ *with* $r < d$ *and, therefore, the shortest path from* u_n *to* u_q *be* $(\ell_1,...,\ell_d)$, $(\ell_1,...,\ell_{d-1})$, ..., $(\ell_1,...,\ell_r)$, $(\ell_1,...,\ell_r,i_1)$, ..., $(\ell_1,...,\ell_r,i_1,...,i_t)$. *The derivative of* u_n *w.r.t.* u_q *is given by*

$$\frac{\partial}{\partial u_q}u_n = -\mathsf{G}_{\ell_1,...,\ell_d}\cdot ...\cdot\mathsf{G}_{\ell_1,...,\ell_{r+1}}\cdot\mathsf{F}_{\ell_1,...,\ell_r,i_1}\cdot\mathsf{F}_{\ell_1,...,\ell_r,i_1,...,i_{t-1}}\cdot\Delta_q u_q^{p_{\ell_1,...,\ell_r,i_1,...,i_{t-1}}-1}$$

with $\Delta_q = sgn\, u_q$ *and* $|u_q|^p = (\Delta_q u_q)^p$. *In particular*

$$u_q \frac{\partial}{\partial u_q} u_n = -\mathsf{G}_{\ell_1,...,\ell_d} \cdot ... \cdot \mathsf{G}_{\ell_1,...,\ell_{r+1}} \cdot \mathsf{F}_{\ell_1,...,\ell_r,i_1} \cdot \mathsf{F}_{\ell_1,...,\ell_r,i_1,...,i_{t-1}} \cdot |u_q|^{p_{\ell_1,...,\ell_r,i_1,...,i_{t-1}}}.$$

Proof. Successive application of Lemma (3.2). □

Before finally deriving the expression for the determinant, we state two more helpful equations.

Lemma 3.4. *Let* $I = i_1, ..., i_{r-1}$, *then*

$$\mathsf{G}^{-1}_{I,i_r}\mathsf{g}^{p_{I,i_r}}_{I,i_r} = \mathsf{g}^{p_I}_{I,i_r} \tag{3}$$

$$= \mathsf{g}^{p_I}_I - \sum_{k=1}^{\ell_I - 1} \mathsf{F}_{I,k}\mathsf{f}^{p_{I,k}}_{I,k} \tag{4}$$

and

$$\mathsf{f}^{p_{I,i_r}}_{I,i_r} = \sum_{k=1}^{\ell_{I,i_r}} \mathsf{F}_{I,i_r,k}\mathsf{f}^{p_{I,i_r,k}}_{I,i_r,k} \tag{5}$$

Proof. First, we prove the equalities (3) and (4):

$$\begin{aligned}
\mathsf{G}^{-1}_{I,i_r}\mathsf{g}^{p_{I,i_r}}_{I,i_r} &= \mathsf{g}^{-(p_{I,i_r}-p_I)}_{I,i_r}\mathsf{g}^{p_{I,i_r}}_{I,i_r} \\
&= \mathsf{g}^{p_I}_{I,i_r} \text{ q.e.d. } (3) \\
&= \left(\mathsf{g}^{p_I}_I - \sum_{k=1}^{\ell_I-1} \mathsf{f}^{p_I}_{I,k}\right)^{\frac{p_I}{p_I}} \\
&= \mathsf{g}^{p_I}_I - \sum_{k=1}^{\ell_I-1} \mathsf{f}^{p_I - p_{I,k}}_{I,k}\mathsf{f}^{p_{I,k}}_{I,k} \\
&= \mathsf{g}^{p_I}_I - \sum_{k=1}^{\ell_I-1} \mathsf{F}_{I,k}\mathsf{f}^{p_{I,k}}_{I,k} \text{ q.e.d. } (4).
\end{aligned}$$

In a similar fashion, equality (5) can be proven by substituting definitions and introducing one in the exponent.

□

Proposition 3.1 (Determinant of the Jacobian)**.** *Let* $\mathcal{L}$ *be the set of multi-indices of the path from the leaf* u_n *to the root node (excluding the root node). The determinant of the Jacobian for an* L_p*-nested function is given by*

$$\det|\mathcal{J}| = r^{n-1}\prod_{L\in\mathcal{L}} \mathsf{G}_L.$$

Proof. Let $L = \ell_1, ...\ell_{d-1}$ be the multi-index of the parent of u_n. We compute $\frac{1}{r^{n-1}}|\det \mathcal{J}|$ and obtain the result by solving for $|\det \mathcal{J}|$. As shown in Lemma (3.1) $\frac{1}{r^{n-1}}|\det \mathcal{J}|$ has the form

$$\frac{1}{r^{n-1}}|\det \mathcal{J}| = -\sum_{k=1}^{n-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + u_n.$$

By definition $u_n = \mathsf{g}_{L,\ell_d} = \mathsf{g}_{L,\ell_d}^{p_{L,\ell_d}}$. Now, assume that $u_r, ..., u_{n-1}$ are children of f_L, i.e. $u_k = \mathsf{f}_{L,I,i_t}$ for some $I, i_t = i_1, ..., i_t$ and $r \leq k < n$. Remember, that by Lemma (3.3) the terms $u_q \frac{\partial}{\partial u_q} u_n$ for $r \leq q < n$ have the form

$$u_q \frac{\partial}{\partial u_q} u_n = -\mathsf{G}_{L,\ell_d} \cdot \mathsf{F}_{L,i_1} \cdot ... \cdot \mathsf{F}_{L,I} \cdot |u_q|^{p_{\ell_1,...,\ell_{d-1},i_1,...,i_{t-1}}}.$$

Now, we can expand the determinant as follows

$$\begin{aligned}
&-\sum_{k=1}^{n-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{g}_{L,\ell_d}^{p_{L,\ell_d}} \\
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k - \sum_{k=r}^{n-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{g}_{L,\ell_d}^{p_{L,\ell_d}} \\
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d} \left(-\sum_{k=r}^{n-1} \mathsf{G}_{L,\ell_d}^{-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d}^{-1} \mathsf{g}_{L,\ell_d}^{p_{L,\ell_d}} \right) \\
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d} \left(-\sum_{k=r}^{n-1} \mathsf{G}_{L,\ell_d}^{-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{g}_L^{p_L} - \sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}} \right)
\end{aligned}$$

by equality (4) of Lemma (3.4). Note that all terms $\mathsf{G}_{L,\ell_d}^{-1} \frac{\partial u_n}{\partial u_k} \cdot u_k$ for $r \leq k < n$ now have the form

$$\mathsf{G}_{L,\ell_d}^{-1} u_k \frac{\partial}{\partial u_k} u_n = -\mathsf{F}_{L,i_1} \cdot ... \cdot \mathsf{F}_{L,I} \cdot |u_q|^{p_{\ell_1,...,\ell_{d-1},i_1,...,i_{t-1}}}$$

since we constructed them to be neighbors of u_n. However, with equation (5) of Lemma (3.4), we can further expand the sum $\sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}}$ down to the leafs $u_r, ..., u_{n-1}$. When doing so we end up with the same factors $\mathsf{F}_{L,i_1} \cdot ... \cdot \mathsf{F}_{L,I} \cdot |u_q|^{p_{\ell_1,...,\ell_{d-1},i_1,...,i_{t-1}}}$ as in the derivatives $\mathsf{G}_{L,\ell_d}^{-1} u_q \frac{\partial}{\partial u_q} u_n$. This means exactly that

$$-\sum_{k=r}^{n-1} \mathsf{G}_{L,\ell_d}^{-1} \frac{\partial u_n}{\partial u_k} \cdot u_k = \sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}}$$

and, therefore,

$$\begin{aligned}
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d} \left(-\sum_{k=r}^{n-1} \mathsf{G}_{L,\ell_d}^{-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{g}_L^{p_L} - \sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}} \right) \\
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d} \left(\sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}} + \mathsf{g}_L^{p_L} - \sum_{k=1}^{\ell_d - 1} \mathsf{F}_{L,k} \mathsf{f}_{L,k}^{p_{L,k}} \right) \\
&= -\sum_{k=1}^{r-1} \frac{\partial u_n}{\partial u_k} \cdot u_k + \mathsf{G}_{L,\ell_d} \mathsf{g}_L^{p_L}.
\end{aligned}$$

By factoring out G_{L,ℓ_d} from the equation, the terms $\frac{\partial u_n}{\partial u_k} \cdot u_k$ loose the G_{L,ℓ_d} in front and we get basically the same equation as before, only that the new leaf (the new "u_n") is $\mathsf{g}_L^{p_L}$ and we got rid of all the children of f_L. By repeating that procedure up to the root node, we successively factor out all $\mathsf{G}_{L'}$ for $L' \in \mathcal{L}$ until all terms of the sum vanish and we are only left with $\mathsf{f}_\emptyset = 1$. Therefore, the determinant is

$$\frac{1}{r^{n-1}}|\det \mathcal{J}| = \prod_{L\in\mathcal{L}} \mathsf{G}_L$$

which completes the proof. □

4. L_p-NESTED UNIFORM DISTRIBUTION

In analogy to [6] we define a uniform distribution on the L_P-nested sphere. Naturally, the density of this distribution is the inverse of the surface area of the L_p-nested unit sphere. In this section we first compute the surface of the L_p-nested sphere and then define the L_p-nested uniform distribution in terms of the polar-like coordinates from the section before. Before we start, we start by computing the surface and the volume of an arbitrary L_p-nested sphere.

Proposition 4.1 (Volumen and Surface of the L_p-nested Sphere)**.** *Let f be an L_p-nested function and let $\mathcal{I}$ be the set of all multi-indices denoting the inner nodes of the tree structure associated with f. Let n_I denote the number of leafs contained in the subtree under the node I (if I is a leaf already, $n_I = 1$). The volumen $\mathcal{V}_f(R)$ and the surface $\mathcal{S}_f(R)$ of the L_p-nested sphere with radius R is given by*

$$\mathcal{V}_f(R) = \frac{R^n 2^n}{n}\prod_{I\in\mathcal{I}}\frac{1}{p_I^{\ell_I-1}}\prod_{k=1}^{\ell_I-1} B\left[\frac{\sum_{i=1}^k n_{I,k}}{p_I}, \frac{n_{I,k+1}}{p_I}\right] \tag{6}$$

$$= \frac{R^n 2^n}{n}\prod_{I\in\mathcal{I}}\frac{\prod_{k=1}^{\ell_I}\Gamma\left[\frac{n_{I,k}}{p_I}\right]}{p_I^{\ell_I-1}\Gamma\left[\frac{n_I}{p_I}\right]} \tag{7}$$

$$\mathcal{S}_f(R) = R^{n-1}2^n\prod_{I\in\mathcal{I}}\frac{1}{p_I^{\ell_I-1}}\prod_{k=1}^{\ell_I-1} B\left[\frac{\sum_{i=1}^k n_{I,k}}{p_I}, \frac{n_{I,k+1}}{p_I}\right] \tag{8}$$

$$= R^{n-1}2^n\prod_{I\in\mathcal{I}}\frac{\prod_{k=1}^{\ell_I}\Gamma\left[\frac{n_{I,k}}{p_I}\right]}{p_I^{\ell_I-1}\Gamma\left[\frac{n_I}{p_I}\right]} \tag{9}$$

Proof. We obtain the volumen by computing the integral $\int_{f(\mathbf{x})\leq R} d\mathbf{x}$. Differentiation with respect to R yields the surface area. For symmetry reasons we can compute the volume only on the positive quadrant $\mathbb{R}^n_+$ and multiply the result with 2^n later to obtain the full volumen and surface area. The strategy for computing the volumen is as follows. We start off with inner nodes I that are parents of leafs only. The value f_I of such a node is simply the L_{p_I} norm of its children. Therefore, we can convert the integral over the children of I with the transformation of [3]. This maps the leafs $\mathbf{f}_{I,1:\ell_I}$ into f_I and "angular" variables $\tilde{\boldsymbol{u}}_{\ell_I-1}$. Since integral borders of the original integral depend only on the value of f_I and not on $\tilde{\boldsymbol{u}}$, we can separate the variables $\tilde{\boldsymbol{u}}$ from the radial variables f_I and integrate the variables $\tilde{\boldsymbol{u}}_{\ell_I-1}$ separately. The integration over $\tilde{\boldsymbol{u}}_{\ell_I-1}$ yields a certain factor, while the variable f_I effectively becomes a new leaf.

Now suppose I is the parent of leafs only. W.l.o.g. let the ℓ_I leafs correspond to the last ℓ_I coefficients of $\mathbf{x}$. Let $\mathbf{x}\in\mathbb{R}^n_+$. Carrying out the first transformation and integration yields

$$\begin{aligned}\int_{f(\mathbf{x})\leq R} d\mathbf{x} &= \int_{f(\mathbf{x}_{1:n-\ell_I},\mathsf{f}_I)\leq R}\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}} \mathsf{f}_I^{\ell_I-1}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{1-p_I}{p_I}} d\mathsf{f}_I d\tilde{\boldsymbol{u}}_{\ell_I-1} d\mathbf{x}_{1:n-\ell_I}\\ &= \int_{f(\mathbf{x}_{1:n-\ell_I},\mathsf{f}_I)\leq R} \mathsf{f}_I^{n_I-1} d\mathsf{f}_I d\mathbf{x}_{1:n-\ell_I} \times \int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{n_{I,\ell_I}-p_I}{p_I}} d\tilde{\boldsymbol{u}}_{\ell_I-1}.\end{aligned}$$

For solving the second integral we make the pointwise transformation $s_i = \tilde{u}_i^{p_I}$ and obtain

$$\begin{aligned}\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{n_{I,\ell_I}-p_I}{p_I}} d\tilde{\boldsymbol{u}}_{\ell_I-1} &= \frac{1}{p_I^{\ell_I-1}}\int_{\sum s_i\leq 1}\left(1-\sum_{i=1}^{\ell_I-1}s_i\right)^{\frac{n_{I,\ell_I}}{p_I}-1}\prod_{i=1}^{\ell_I-1} s_i^{\frac{1}{p_I}-1} d\boldsymbol{s}_{\ell_I-1}\\ &= \frac{1}{p_I^{\ell_I-1}}\prod_{k=1}^{\ell_I-1} B\left[\frac{\sum_{i=1}^k n_{I,k}}{p_I},\frac{n_{I,k+1}}{p_I}\right]\\ &= \frac{1}{p_I^{\ell_I-1}}\prod_{k=1}^{\ell_I-1} B\left[\frac{k}{p_I},\frac{1}{p_I}\right]\end{aligned}$$

by using the fact that the transformed integral has the form of an unnormalized Dirichlet distribution and, therefore, the value of the integral must equal its normalization constant.

Now, we go on with solving the integral

$$\int_{f(\mathbf{x}_{1:n-\ell_I},\mathsf{f}_I)\leq R} \mathsf{f}_I^{n_I-1} d\mathsf{f}_I d\mathbf{x}_{1:n-\ell_I}. \tag{10}$$

We carry this out in exactly the same manner as we solved the previous integral. We only need to make sure that we only contract nodes that have only leafs as children (remember that radii of contracted nodes become leafs) and we need to find a formula how the factors $\mathsf{f}_I^{n_I-1}$ propagate through the tree.

For the latter, we first state the formula and then prove it via induction. For notational convenience let $\hat{\mathbf{x}}$ denote the remaining coefficients of $\mathbf{x}$, $\hat{\mathbf{f}}$ the vector of leafs resulting from contraction and $\mathcal{J}$ the set of multi-indices corresponding to the contracted leafs. The integral which is left to solve after integrating over all $\tilde{\boldsymbol{u}}$ is given by (remember that n_J denotes real leafs, i.e. the ones corresponding to coefficients of $\mathbf{x}$):

$$\int_{f(\hat{\mathbf{x}},\hat{\mathbf{f}})\leq R}\prod_{J\in\mathcal{J}}\mathsf{f}_J^{n_J-1} d\hat{\mathbf{f}} d\hat{\mathbf{x}}.$$

We already proved the first induction step by computing equation (10). For computing the general induction step suppose I is an inner node whose children are leafs or contracted leafs. Let $\mathcal{J}'$ be the set of contracted leafs under I and $\hat{\mathcal{J}} = \mathcal{J}\backslash\mathcal{J}'$. Furthermore, let $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{x}}$ be the leafs belonging to the set $\hat{\mathcal{J}}$. For notational convenience, we will denote all children of I with $\mathsf{f}_{I,k}$ no matter whether they are real leafs y_i or result from a previous contraction. Transforming the children of I into radial coordinates by [3] yields

$$\begin{aligned}
\int_{f(\hat{\mathbf{x}},\hat{\mathbf{f}})\leq R}\prod_{J\in\mathcal{J}}\mathsf{f}_J^{n_J-1}d\hat{\mathbf{f}}d\hat{\mathbf{x}} &= \int_{f(\hat{\mathbf{x}},\hat{\mathbf{f}})\leq R}\left(\prod_{\hat{J}\in\hat{\mathcal{J}}}\mathsf{f}_{\hat{J}}^{n_{\hat{J}}-1}\right)\cdot\left(\prod_{J'\in\mathcal{J}'}\mathsf{f}_{J'}^{n_{J'}-1}\right)d\hat{\mathbf{f}}d\hat{\mathbf{x}}\\
&= \int_{f(\tilde{\mathbf{x}},\tilde{\mathbf{f}},\mathsf{f}_I)\leq R}\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{1-p_I}{p_I}}\mathsf{f}_I^{\ell_I-1}\right)\cdot\left(\prod_{\hat{J}\in\hat{\mathcal{J}}}\mathsf{f}_{\hat{J}}^{n_{\hat{J}}-1}\right)\\
&\quad\times\left(\left(\mathsf{f}_I\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)\right)^{\frac{n_{\ell_I}-1}{p_I}}\prod_{k=1}^{\ell_I-1}(\mathsf{f}_I\tilde{u}_k)^{n_k-1}\right)d\tilde{\mathbf{x}}d\tilde{\mathbf{f}}d\mathsf{f}_Id\tilde{\boldsymbol{u}}_{\ell_I-1}\\
&= \int_{f(\tilde{\mathbf{x}},\tilde{\mathbf{f}},\mathsf{f}_I)\leq R}\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(\prod_{\hat{J}\in\hat{\mathcal{J}}}\mathsf{f}_{\hat{J}}^{n_{\hat{J}}-1}\right)\\
&\quad\times\left(\mathsf{f}_I^{\ell_I-1+\sum_{i=1}^{\ell_I}(n_i-1)}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{n_{\ell_I}-p_I}{p_I}}\prod_{k=1}^{\ell_I-1}\tilde{u}_k^{n_k-1}\right)d\tilde{\mathbf{x}}d\tilde{\mathbf{f}}d\mathsf{f}_Id\tilde{\boldsymbol{u}}_{\ell_I-1}\\
&= \int_{f(\tilde{\mathbf{x}},\tilde{\mathbf{f}},\mathsf{f}_I)\leq R}\left(\prod_{\hat{J}\in\hat{\mathcal{J}}}\mathsf{f}_{\hat{J}}^{n_{\hat{J}}-1}\right)\mathsf{f}_I^{n_I-1}d\tilde{\mathbf{x}}d\tilde{\mathbf{f}}d\mathsf{f}_I\\
&\quad\times\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{n_{\ell_I}-p_I}{p_I}}\prod_{k=1}^{\ell_I-1}\tilde{u}_k^{n_k-1}d\tilde{\boldsymbol{u}}_{\ell_I-1}.
\end{aligned}$$

Again, by transforming it into a Dirichlet distribution, the latter integral has the solution

$$\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}}\left(1-\sum_{i=1}^{\ell_I-1}\tilde{u}_i^{p_I}\right)^{\frac{n_{\ell_I}-p_I}{p_I}}\prod_{k=1}^{\ell_I-1}\tilde{u}_k^{n_k-1}d\tilde{\boldsymbol{u}}_{\ell_I-1}=\prod_{k=1}^{\ell_I-1}B\left[\frac{\sum_{i=1}^{k}n_{I,k}}{p_I},\frac{n_{I,k+1}}{p_I}\right]$$

while the remaining former integral has the form

$$\int_{f(\tilde{\mathbf{x}},\tilde{\mathbf{f}},\mathsf{f}_I)\leq R}\left(\prod_{\hat{J}\in\hat{\mathcal{J}}}\mathsf{f}_{\hat{J}}^{n_{\hat{J}}-1}\right)\mathsf{f}_I^{n_I-1}d\tilde{\mathbf{x}}d\tilde{\mathbf{f}}d\mathsf{f}_I=\int_{f(\hat{\mathbf{x}},\hat{\mathbf{f}})\leq R}\prod_{J\in\mathcal{J}}\mathsf{f}_J^{n_J-1}d\hat{\mathbf{f}}d\hat{\mathbf{x}}$$

as claimed.

By carrying out the integration up to the root node the remaining integral becomes

$$\int_{\mathsf{f}_\emptyset\leq R}\mathsf{f}_\emptyset^{n-1}df_\emptyset=\int_0^R\mathsf{f}_\emptyset^{n-1}df_\emptyset=\frac{R^n}{n}.$$

Collecting the factors from integration over the $\tilde{\boldsymbol{u}}$ proves the equations (6) and (8). Using $B[a,b]=\frac{\Gamma[a]\Gamma[b]}{\Gamma[a+b]}$ yields equations (7) and (9). □

In order to clarify the proof we explicitly carry out the integration for our first example.

Example 4.1. Again, let the L_p-nested function be given by

$$f(\mathbf{x}) = \left((|x_1|^{p_1} + |x_2|^{p_1})^{\frac{p_\emptyset}{p_1}} + |x_3|^{p_\emptyset} \right)^{\frac{1}{p_\emptyset}}.$$

Let $\mathbf{x} \in \mathbb{R}^3_+$. Carrying out the steps from the proof above yields

$$\begin{aligned}\int_{f(\mathbf{x})\leq R} d\mathbf{x} &= \int_{f(\mathsf{f}_1,x_3)\leq R} \int_0^1 (1-\tilde{u}^{p_1})^{\frac{1-p_1}{p_1}} \mathsf{f}_1^{\ell_1-1} d\tilde{u}d\mathsf{f}_1 dx_3 \\ &= \int_{f(\mathsf{f}_1,x_3)\leq R} \mathsf{f}_1^{\ell_1-1} d\mathsf{f}_1 dx_3 \times \int_0^1 (1-\tilde{u}^{p_1})^{\frac{1-p_1}{p_1}} d\tilde{u} \\ &= \int_{f(\mathsf{f}_1,x_3)\leq R} \mathsf{f}_1^{\ell_1-1} d\mathsf{f}_1 dx_3 \times \frac{1}{p_1} B\left[\frac{1}{p_1}, \frac{1}{p_1}\right].\end{aligned}$$

Solving the first integral yields

$$\begin{aligned}\int_{f(\mathsf{f}_1,x_3)\leq R} \mathsf{f}_1^{\ell_1-1} d\mathsf{f}_1 &= \int_{\mathsf{f}_\emptyset\leq R} \int_0^1 \mathsf{f}_\emptyset^{\ell_\emptyset-1} (\mathsf{f}_\emptyset \tilde{u}^{p_0})^{\ell_1-1} (1-\tilde{u}^{p_0})^{\frac{1-p_\emptyset}{p_\emptyset}} d\tilde{u}d\mathsf{f}_\emptyset \\ &= \int_{\mathsf{f}_\emptyset\leq R} \int_0^1 \mathsf{f}_\emptyset^{\ell_\emptyset+\ell_1-2} \tilde{u}^{\ell_1-1} (1-\tilde{u}^{p_0})^{\frac{1-p_\emptyset}{p_\emptyset}} d\tilde{u}d\mathsf{f}_\emptyset \\ &= \int_{\mathsf{f}_\emptyset\leq R} \mathsf{f}_\emptyset^2 d\mathsf{f}_\emptyset \times \int_0^1 \tilde{u}\,(1-\tilde{u}^{p_0})^{\frac{1-p_\emptyset}{p_\emptyset}} d\tilde{u} \\ &= \frac{R^3}{3} \cdot \frac{1}{p_\emptyset} B\left[\frac{2}{p_\emptyset}, \frac{1}{p_\emptyset}\right].\end{aligned}$$

Collecting all factors yields

$$\int_{f(\mathbf{x})\leq R} d\mathbf{x} = \frac{R^3}{3} \cdot \frac{1}{p_\emptyset}\frac{1}{p_1} B\left[\frac{2}{p_\emptyset}, \frac{1}{p_\emptyset}\right] B\left[\frac{1}{p_1}, \frac{1}{p_1}\right].$$

Extending the domain such that $\mathbf{x} \in \mathbb{R}^3$, simply introduces a factor 2^3. The surface is obtained by differentiating with respect to R. This yields the final equations

$$\begin{aligned}\mathcal{V}_f(R) &= \frac{R^3 2^3}{3} \cdot \frac{1}{p_\emptyset}\frac{1}{p_1} B\left[\frac{2}{p_\emptyset}, \frac{1}{p_\emptyset}\right] B\left[\frac{1}{p_1}, \frac{1}{p_1}\right] \\ \mathcal{S}_f(R) &= R^2 2^3 \cdot \frac{1}{p_\emptyset}\frac{1}{p_1} B\left[\frac{2}{p_\emptyset}, \frac{1}{p_\emptyset}\right] B\left[\frac{1}{p_1}, \frac{1}{p_1}\right]\end{aligned}$$

Proposition 4.2 (L_p-nested Uniform Distribution)**.** *Let f be an L_p-nested function. Let $\mathcal{L}$ be set set of multi-indices on the path from the root node to the leaf corresponding to y_n and let $\tilde{L}$ be the multi-index of x_n. The uniform distribution on the L_p-nested unit sphere, i.e. the set $\{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = 1\}$ is given by*

$$\rho(\mathbf{u}) = \left(\frac{1}{2^{n-1}} \prod_{I\in\mathcal{I}} p_I^{\ell_I-1} \prod_{k=1}^{\ell_I-1} B\left[\frac{\sum_{i=1}^k n_{I,k}}{p_I}, \frac{n_{I,k+1}}{p_I}\right]^{-1} \right) \cdot \prod_{L\in\mathcal{L}} \mathsf{G}_L$$

where the support of $p(\mathbf{u})$ is given by

$$\operatorname{supp} \rho = \left\{\mathbf{u} \in \mathbb{R}^{n-1} | f(\mathbf{u}, g_{\tilde{L}}(\mathbf{u})) = 1\right\}$$

Proof. Since the L_p-nested sphere is a compact set, the density of the uniform distribution is simply one over the surface area of the unit L_p-nested sphere. The surface $\mathcal{S}_f(1)$ is given by Proposition 4.1. Transforming $\frac{1}{\mathcal{S}_f(1)}$ into the coordinates of Definition 3.1 introduces the determinant of the Jacobian from Proposition 3.1 and an additional factor of 2 since the $\mathbf{u} \in \mathbb{R}^{n-1}$ have to account for both half-shells of the L_p-nested unit sphere. This yields the expression above. □

Example 4.2 (L_p-spherically symmetric uniform distribution). We consider L_p-norm as a special case of an L_p-nested function

$$f(\mathbf{x}) = ||\mathbf{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

The corresponding tree has only one single inner node, which is the root node. Using Proposition 4.1, the surface area is given by

$$\begin{aligned}
\mathcal{S}_{||\cdot||_p} &= 2^n \frac{1}{p_\emptyset^{\ell_\emptyset - 1}} \prod_{k=1}^{\ell_\emptyset - 1} B\left[\frac{\sum_{i=1}^{k} n_k}{p_\emptyset}, \frac{n_{k+1}}{p_\emptyset}\right] \\
&= 2^n \frac{1}{p^{n-1}} \prod_{k=1}^{n-1} B\left[\frac{k}{p}, \frac{1}{p}\right] \\
&= 2^n \frac{1}{p^{n-1}} \prod_{k=1}^{n-1} \frac{\Gamma\left[\frac{k}{p}\right]\Gamma\left[\frac{1}{p}\right]}{\Gamma\left[\frac{k+1}{p}\right]} \\
&= \frac{2^n \Gamma^n\left[\frac{1}{p}\right]}{p^{n-1}\Gamma\left[\frac{n}{p}\right]}.
\end{aligned}$$

The factor G_n is given by $\left(1 - \sum_{i=1}^{n-1} |u_i|^p\right)^{\frac{1-p}{p}}$, which together with the factor 2 yields the uniform distribution on the L_p-sphere as defined in [6]

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

5. L_p-Nested Symmetric Distributions

Definition 5.1 (L_p-Nested Symmetric Distribution)**.** A n-dimensional random vector $\boldsymbol{X}$ is called *L_p-nested symmetrically distributed with respect to f* if f is an L_p-nested function, $\boldsymbol{X} = R\boldsymbol{U}$ for two independent random variables R and $\boldsymbol{U}$, where R is a non-negative univariate random variable and $\boldsymbol{U}$ is a n-dimensional random variable uniformly distributed on the L_p-nested unit sphere corresponding to f, i.e. $f(\boldsymbol{U}) = 1$ and $U_1, .., U_{n-1}$ follow the distribution of Proposition 4.2.

This definition of L_p-nested symmetric distribution is a straightforward generalization of Gupta and Song's definition of L_p-spherically symmetric distributions. By exactly the same reasoning as their's [3] the definition implies that $f(X) \doteq R$ and $\frac{\boldsymbol{X}}{f(\boldsymbol{X})} \doteq \boldsymbol{U}$ and, therefore, that $f(\boldsymbol{X})$ and $\frac{\boldsymbol{X}}{f(\boldsymbol{X})}$ are independent. This also means that being able to sample from any L_p-nested symmetric distribution makes it possible to sample from any other L_p-nested symmetric distribution as long as the radial distribution of it is known. One simply has to normalize the samples $\boldsymbol{X}$ from the first distribution to obtain an instance of a uniformly distributed random variable on the L_p-unit sphere, sample a new radius and scale the normalized sample with it. Based on that idea, we derive a sampling scheme for L_p-nested distributions in section 6.

Another consequence resulting from the definition of L_p-nested symmetric distributions is the following proposition, which is almost equivalent to Lemma 2.1 and Theorem 2.1 in [3] which themselves are a special case of the results in [2].

Proposition 5.1. *Each L_p-nested symmetric density on $\mathbb{R}^n$ (with zero probability mass at zero) has the form $\tilde{\rho}(\boldsymbol{X}) = \rho(f(\boldsymbol{X}))$ and gives rise to a univariate (radial) density ϱ on $\mathbb{R}_+$. On the other hand, each univariate density ρ on $\mathbb{R}_+$ gives rise to a L_p-nested symmetric distribution on $\mathbb{R}^n$. The relation between the two densities is given by*

$$\begin{aligned}\varrho(r) &= \mathcal{S}_f(1) r^{n-1} \rho(r)\\ &= \mathcal{S}_f(r)\rho(r)\end{aligned}$$

and

$$\begin{aligned}\rho(\mathbf{x}) &= \frac{1}{\mathcal{S}_f(1)\cdot f^{n-1}(\mathbf{x})}\varrho(f(\mathbf{x}))\\ &= \frac{1}{\mathcal{S}_f(f(\mathbf{x}))}\varrho(f(\mathbf{x})).\end{aligned}$$

This shows again, that L_p-nested symmetric distributions are parameterized over univariate radial distributions. The maximum likelihood estimation of the parameters of L_p-nested symmetric distributions therefore becomes very easy since $\operatorname{argmax}_\vartheta \log \rho(\boldsymbol{X}|\vartheta) = \operatorname{argmax}_\vartheta \log \varrho(f(\boldsymbol{X})|\vartheta)$ which means that parameter estimation can be carried out over a univariate instead of an n-dimensional multivariate distribution, which is more robust and computationally efficient.

By the form of a general L_p-nested function and the corresponding symmetric distribution, one might suspect, that the children of the root node, i.e. the $\mathsf{f}_{1:\ell_\emptyset}$ are $L_{p_\emptyset}$-spherically symmetric distributed. This is actually not the case as the next proposition shows.

Proposition 5.2. *Let f be an L_p-nested function. Suppose we remove complete subtrees (not single branches) from the tree associated with f. Let $\hat{\mathbf{x}} \in \mathbb{R}^m$ denote a subset of the coefficients of $\mathbf{x} \in \mathbb{R}^n$ that are still part of that smaller tree and let $\hat{\mathsf{f}}$ denote the vector of inner nodes that became new leafs. The joint distribution of $\hat{\mathbf{x}}$ and $\hat{\mathsf{f}}$ is given by.*

$$\rho(\hat{\mathbf{x}}, \hat{\mathsf{f}}) = \frac{\varrho(f(\hat{\mathbf{x}}, \hat{\mathsf{f}}))}{S_f(f(\hat{\mathbf{x}}, \hat{\mathsf{f}}))} \prod_{J\in\mathcal{J}} \mathsf{f}_J^{n_J - 1}$$

where J is the set of multi-indices for the elements of $\hat{\mathsf{f}}$ and n_J is the number of leafs (in the original tree) in the subtree under the node J.

Proof.

$$\begin{aligned}\rho(\mathbf{x}) &= \frac{\varrho(f(\mathbf{x}))}{\mathcal{S}_f(f(\mathbf{x}))}\\ &= \frac{\varrho(f(\mathbf{x}_{1:n-\ell_I}, \mathsf{f}_I, \tilde{\boldsymbol{u}}_{\ell_I-1}, \Delta_n))}{\mathcal{S}_f(f(\mathbf{x}))} \cdot \mathsf{f}_I^{\ell_I - 1}\left(1 - \sum_{i=1}^{\ell_I-1} |\tilde{u}_i|^{p_I}\right)^{\frac{1-p_I}{p_I}}\end{aligned}$$

where $\Delta_n = \text{sign}(x_n)$. Note that f is invariant to the actual value of Δ_n. However, when integrating it out, it yields a factor of 2. Integrating out $\tilde{\boldsymbol{u}}_{\ell_I - 1}$ and Δ_n now yields

$$\begin{aligned}\rho(\mathbf{x}_{1:n-\ell_I}, \mathsf{f}_I) &= \frac{\varrho(f(\mathbf{x}_{1:n-\ell_I}, \mathsf{f}_I))}{\mathcal{S}_f(f(\mathbf{x}))} \cdot \mathsf{f}_I^{\ell_I - 1} \frac{2^{\ell_I}\Gamma^{\ell_I}\left[\frac{1}{p_I}\right]}{p_I^{\ell_I - 1}\Gamma\left[\frac{\ell_I}{p_I}\right]} \\ &= \frac{\varrho(f(\mathbf{x}_{1:n-\ell_I}, \mathsf{f}_I))}{\mathcal{S}_f(f(\mathbf{x}_{1:n-\ell_I}, \mathsf{f}_I))} \cdot \mathsf{f}_I^{\ell_I - 1}\end{aligned}$$

Now, we can go on an integrate out more subtrees. For that purpose, let $\hat{\mathbf{x}}$ denote the remaining coefficients of $\mathbf{x}$, $\hat{\mathbf{f}}$ the vector of leafs resulting from the kind of contraction just shown for f_I and $\mathcal{J}$ the set of multi-indices corresponding to the "new leafs", i.e the node f_I after contraction. We obtain the following equation

$$\rho(\hat{\mathbf{x}}, \hat{\mathbf{f}}) = \frac{\varrho(f(\hat{\mathbf{x}}, \hat{\mathbf{f}}))}{S_f(f(\hat{\mathbf{x}}, \hat{\mathbf{f}}))} \prod_{J \in \mathcal{J}} \mathsf{f}_J^{n_J - 1}.$$

where n_J denotes the number of leafs in the subtree under the node J. The proof is basically the same as the one for proposition (4.1).

□

Corollary 5.1. *The children of the root node $\mathbf{f}_{1:\ell_\emptyset} = (\mathsf{f}_1, ..., \mathsf{f}_{\ell_\emptyset})^\top$ follow the distribution*

$$\rho(\mathbf{f}_{1:\ell_\emptyset}) = \frac{p_\emptyset^{\ell_\emptyset - 1}\Gamma\left[\frac{n}{p_\emptyset}\right]}{f^{n-1}(\mathsf{f}_1, ..., \mathsf{f}_{\ell_\emptyset}) 2^m \prod_{k=1}^{\ell_\emptyset} \Gamma\left[\frac{n_k}{p_\emptyset}\right]} \varrho\left(f(\mathsf{f}_1, ..., \mathsf{f}_{\ell_\emptyset})\right) \prod_{i=1}^{\ell_\emptyset} \mathsf{f}_i^{n_i - 1}$$

where $m \leq \ell_\emptyset$ is the number of leafs directly attached to the root node. In particular, $\mathbf{f}_{1:\ell_\emptyset}$ can be written as the product RU, where R is the L_p-nested radius and the single $|U_i|^{p_\emptyset}$ are Dirichlet distributed, i.e. $(|U_1|^{p_\emptyset}, ..., |U_{\ell_\emptyset}|^{p_\emptyset}) \sim Dir\left[\frac{n_1}{p_\emptyset}, ... \frac{n_{\ell_\emptyset}}{p_\emptyset}\right]$.

Proof. The joint distribution is simply the application of Proposition (5.2). Note that $f(\mathsf{f}_1, ..., \mathsf{f}_{\ell_\emptyset}) = ||\mathbf{f}_{1:\ell_\emptyset}||_{p_\emptyset}$. Applying the pointwise transformation $s_i = |u_i|^{p_\emptyset}$ yields $(|U_1|^{p_\emptyset}, ..., |U_{\ell_\emptyset - 1}|^{p_\emptyset}) \sim \text{Dir}\left[\frac{n_1}{p_\emptyset}, ... \frac{n_{\ell_\emptyset}}{p_\emptyset}\right]$ (see also [6]).

□

6. Sampling from L_p-Nested Symmetric Distributions

In this section, we derive a sampling scheme for L_p-nested symmetric distributions. Since the radial and the uniform component are independent, normalizing a the sample from any L_p-nested distribution to f-length one yields samples from the uniform distribution on the L_p-unit sphere. By multiplying those uniform samples with new samples from another radial distribution, one obtains samples from another L_p-nested distribution. Therefore, for each L_p-nested function f one needs to find only a single L_p-nested distribution one is able to sample from. Sampling from all other L_p-nested distributions with respect to f then comes for free due to the trick just described. Gupta and Song [3] sample from the L_p-generalized Normal distribution since it has independent marginals which makes it easy to sample from it. Due to the tree structure of L_p-nested distributions, this is not possible in general. Instead we choose to sample from the uniform distribution inside the L_P-nested unit ball.

From Proposition (4.1) we already know the normalization constant. Therefore, the distribution has the form $\rho(\mathbf{x}) = \frac{1}{\mathcal{V}_f(1)}$. In order to sample from that distribution, we will first only consider the uniform distribution in the positive quadrant of the unit L_p-nested ball which has the form $\rho(\mathbf{x}) = \frac{2^n}{\mathcal{V}_f(1)}$. Samples from the uniform distributions in the whole ball can be obtained by multiplying each coordinate of a sample with independent samples from the uniform distribution in $\{-1, 1\}$.

Again, from the proof of Proposition (4.1), we are now able to derive the sampling scheme. The idea of the proof is to successively transform the inner nodes of the tree associated with f into L_p-radial coordinates as defined by [6]. This yields a series of independent integrals over expressions like

$$\int_{\tilde{\boldsymbol{u}}_{\ell_I-1}\in\mathcal{V}_+^{\ell_I-1}} \left(1 - \sum_{i=1}^{\ell_I-1} \tilde{u}_i^{p_I}\right)^{\frac{n_{\ell_I}-p_I}{p_I}} \prod_{k=1}^{\ell_I-1} \tilde{u}_k^{n_k-1} d\tilde{\boldsymbol{u}}_{\ell_I-1}$$

and a final integral over the radius $\mathsf{f}_\emptyset$ which always is

$$\int_0^1 \mathsf{f}_\emptyset^{n-1} d\mathsf{f}_\emptyset.$$

Since all variables together integrate to one, $\rho(\mathbf{x})$ is still a density on those variables. Because we can integrate the independently, the final radial variable $\mathsf{f}_\emptyset$ and the uniform variables are independent. Now, it is easy to see that $\mathsf{f}_\emptyset$ can be drawn from a β-distribution and the single u^{p_I} can be drawn from a Dirichlet distribution. By reversing the transformations we obtain samples from the uniform distribution inside the unit L_p-nested ball. Normalizing those samples yields uniformly distributed points on the L_p-nested unit sphere which can be transformed into samples from any L_p-nested distribution by multiplying with the appropriate radial samples.

This provides us with the following sampling scheme:

(1) Sample $\mathsf{f}_\emptyset$ from a beta distribution $\beta\left[n, 1\right]$.
(2) For each inner node I of the tree associated with f sample $\boldsymbol{s}_I$ from a Dirichlet distribution Dir $\left[\frac{n_{I,1}}{p_I}, ..., \frac{n_{I,\ell_I}}{p_I}\right]$ where $n_{I,k}$ are the number of leafs in the subtree under node I, k. Obtain uniform coordinates on the L_p-sphere by $s_k \mapsto s_k^{\frac{1}{p_I}} = \tilde{u}_k$.
(3) Apply the reverse transformation to map the $\tilde{\boldsymbol{u}}$ and $\mathsf{f}_\emptyset$ into Cartesian coordinates $\mathbf{x}$.
(4) Normalize $\mathbf{x}$ to get a uniform sample from the sphere $\boldsymbol{z} = \frac{\mathbf{x}}{f(\mathbf{x})}$.
(5) Sample a new radius $\tilde{\mathsf{f}}_\emptyset$ from the radial distribution of the target L_p-nested distribution $\rho_\emptyset$ and obtain the sample via $\tilde{\mathbf{x}} = \tilde{\mathsf{f}}_\emptyset \cdot \boldsymbol{z}$.

References

[1] K. T. Fang, S. Kotz, and K. W. Ng. *Symmetric multivariate and related distributions*. Chapman and Hall New York, 1990. 1

[2] Carmen Fernandez, Jacek Osiewalski, and Mark F. J. Steel. Modeling and inference with ν-spherical distributions. *Journal of the American Statistical Association*, 90(432):1331–1340, Dec 1995. 1, 18

[3] A.K. Gupta and D. Song. l_p-norm spherical distribution. *Journal of Statistical Planning and Inference*, 60:241–260, 1997. 1, 4, 13, 15, 17, 18, 19

[4] Douglas Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya: The Indian Journal of Statistics, Series A*, 32(4):419–430, Dec 1970. 1

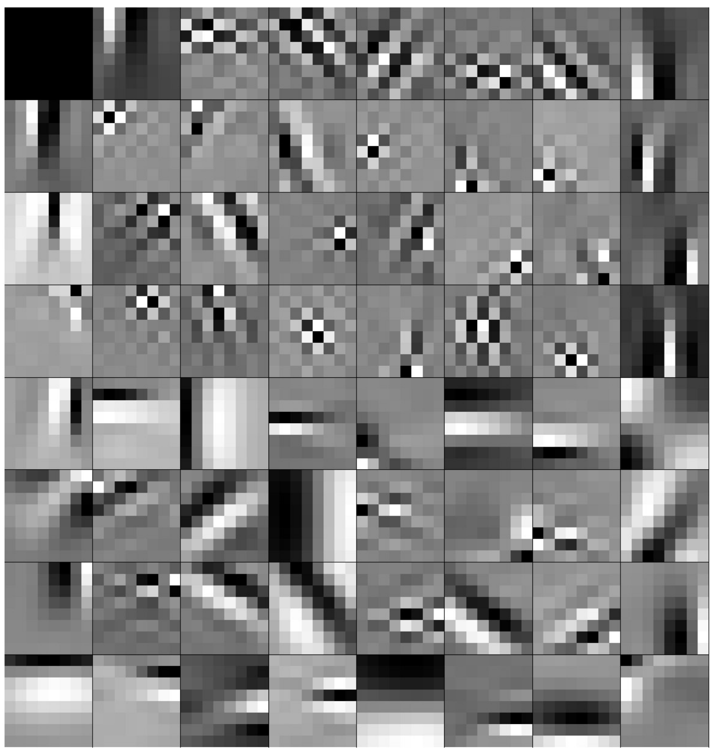
[5] Jacek Osiewalski and Mark F. J. Steel. Robust bayesian inference in l_q-spherical models. *Biometrika*, 80(2):456–460, Jun 1993. 1

[6] D. Song and A.K. Gupta. l_p-norm uniform distribution. *Proceedings of the American Mathematical Society*, 125:595–601, 1997. 1, 13, 17, 19, 20

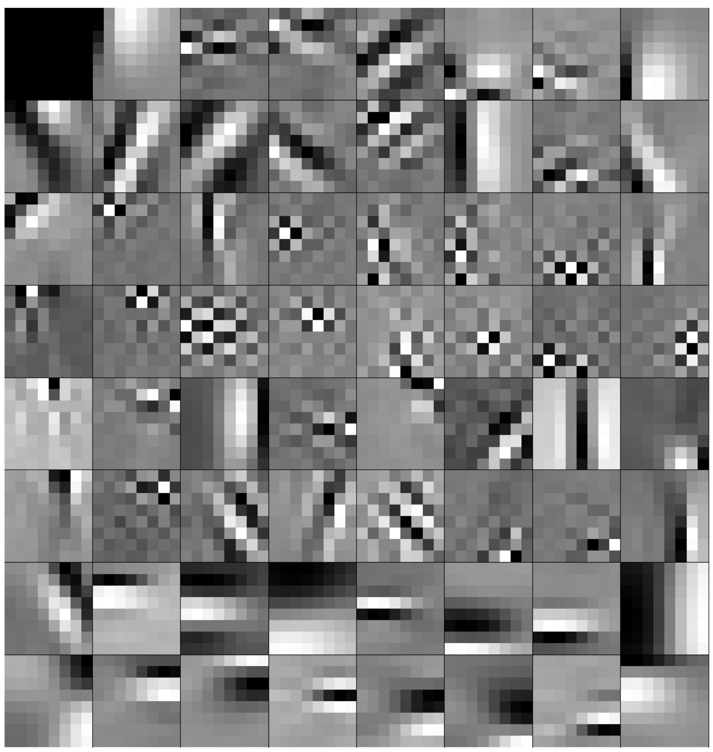
1. Optimal Filters For All Different Models

Independent Subspace Models

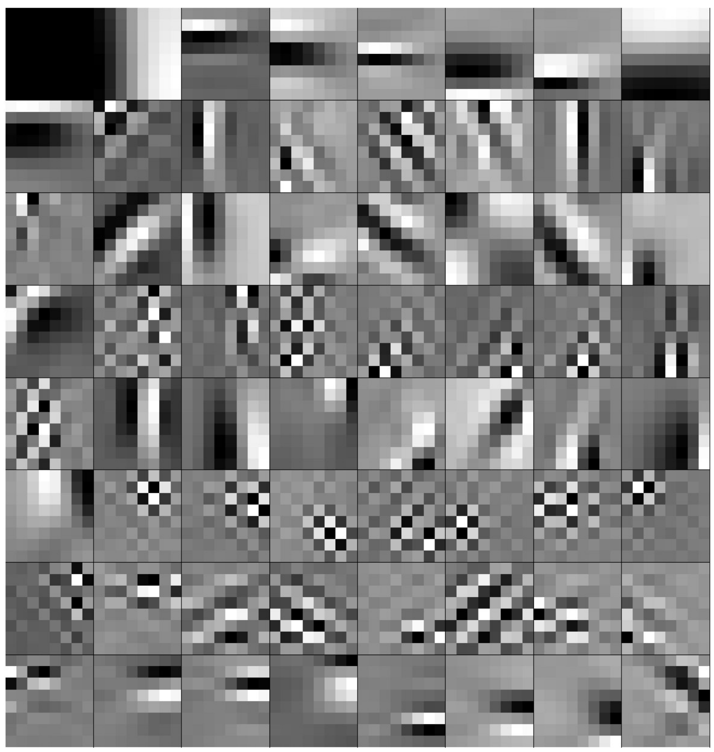
Independent Subspace ISA for 2 Subspaces without CGC.



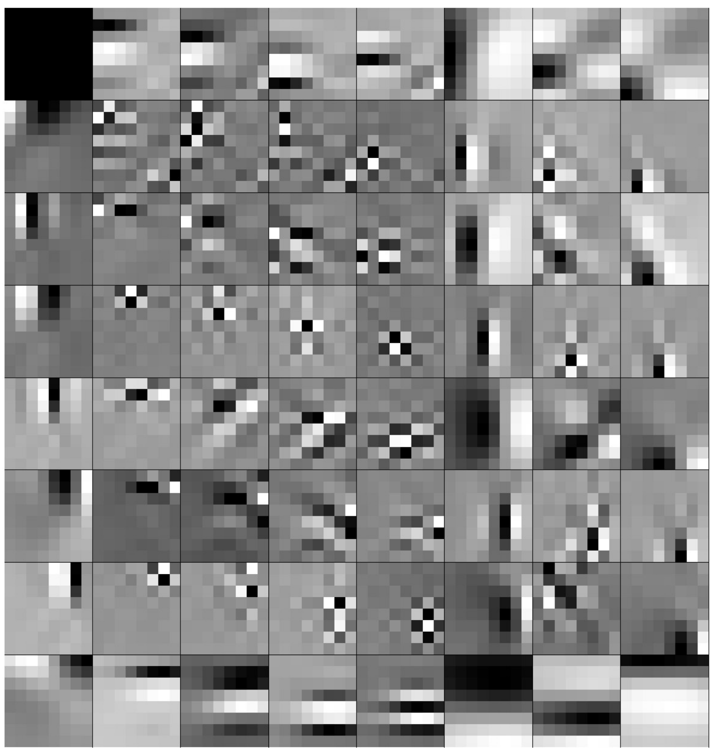
Independent Subspace ISA for 4 Subspaces without CGC.



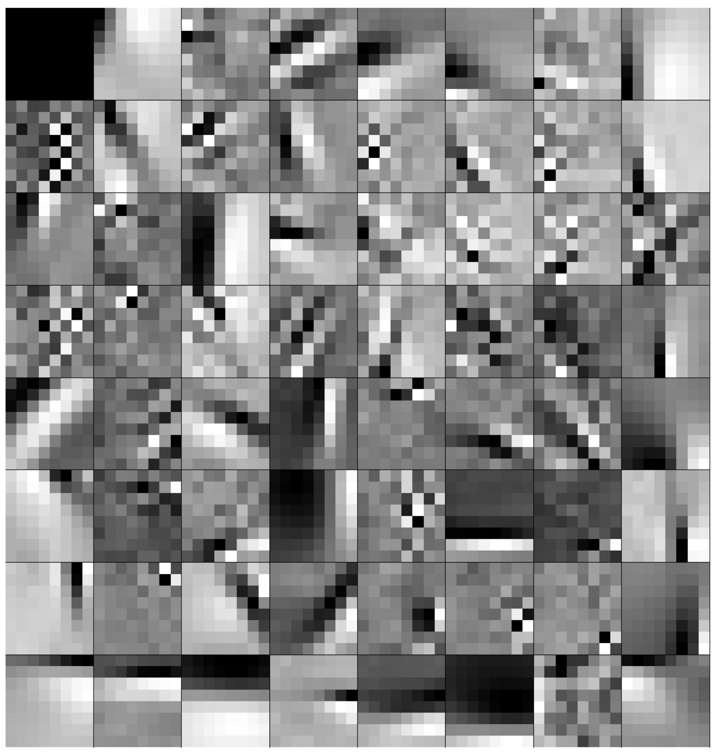
Independent Subspace ISA for 8 Subspaces without CGC.



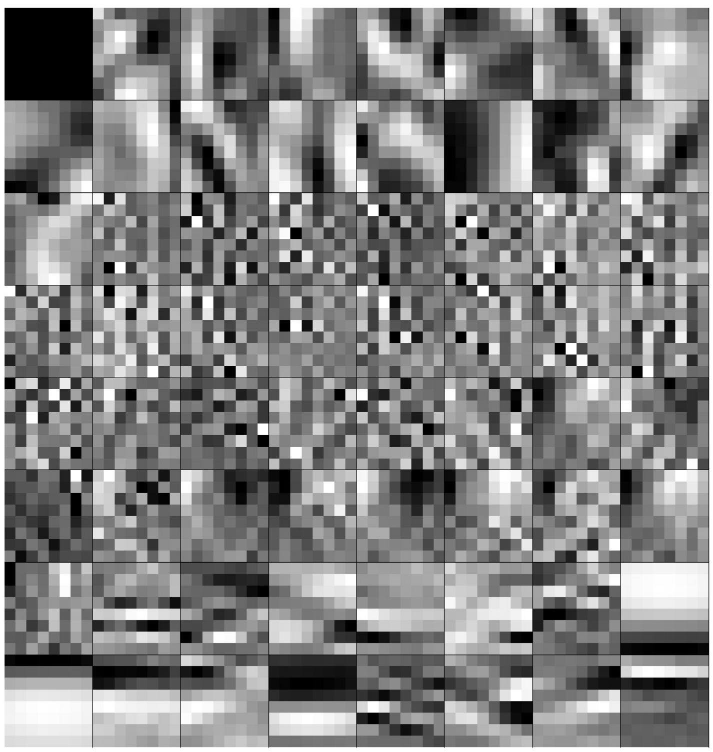
Independent Subspace ISA for 16 Subspaces without CGC.



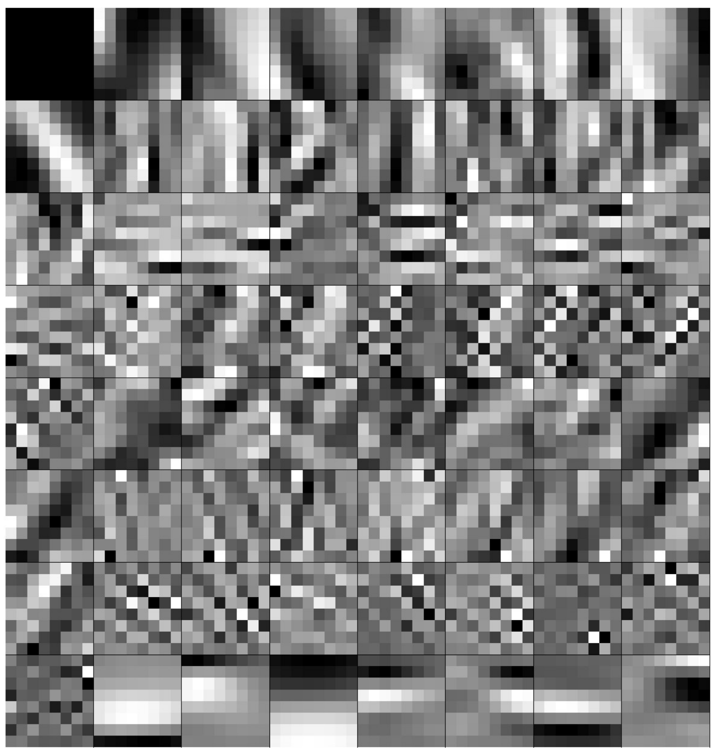
Independent Subspace ISA for 2 Subspaces with CGC.



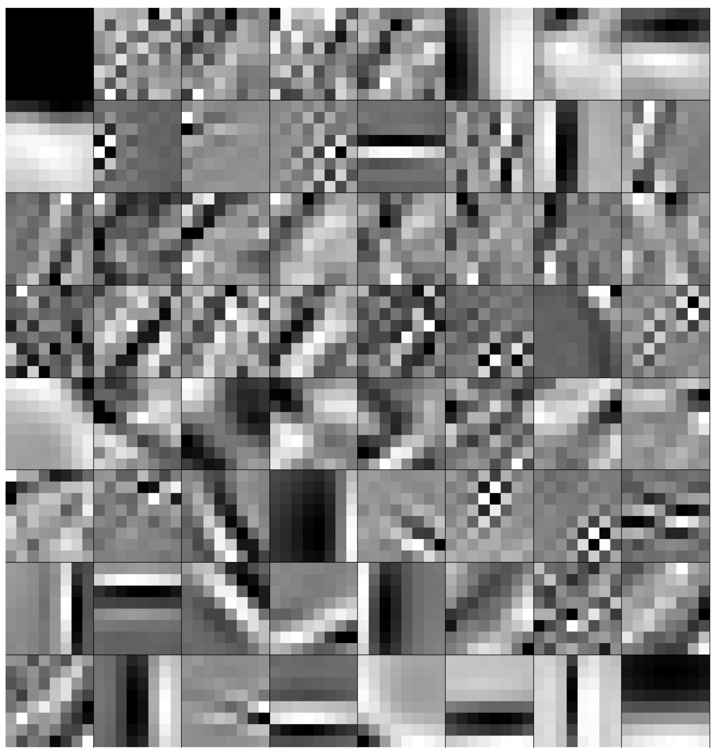
Independent Subspace ISA for 4 Subspaces with CGC.



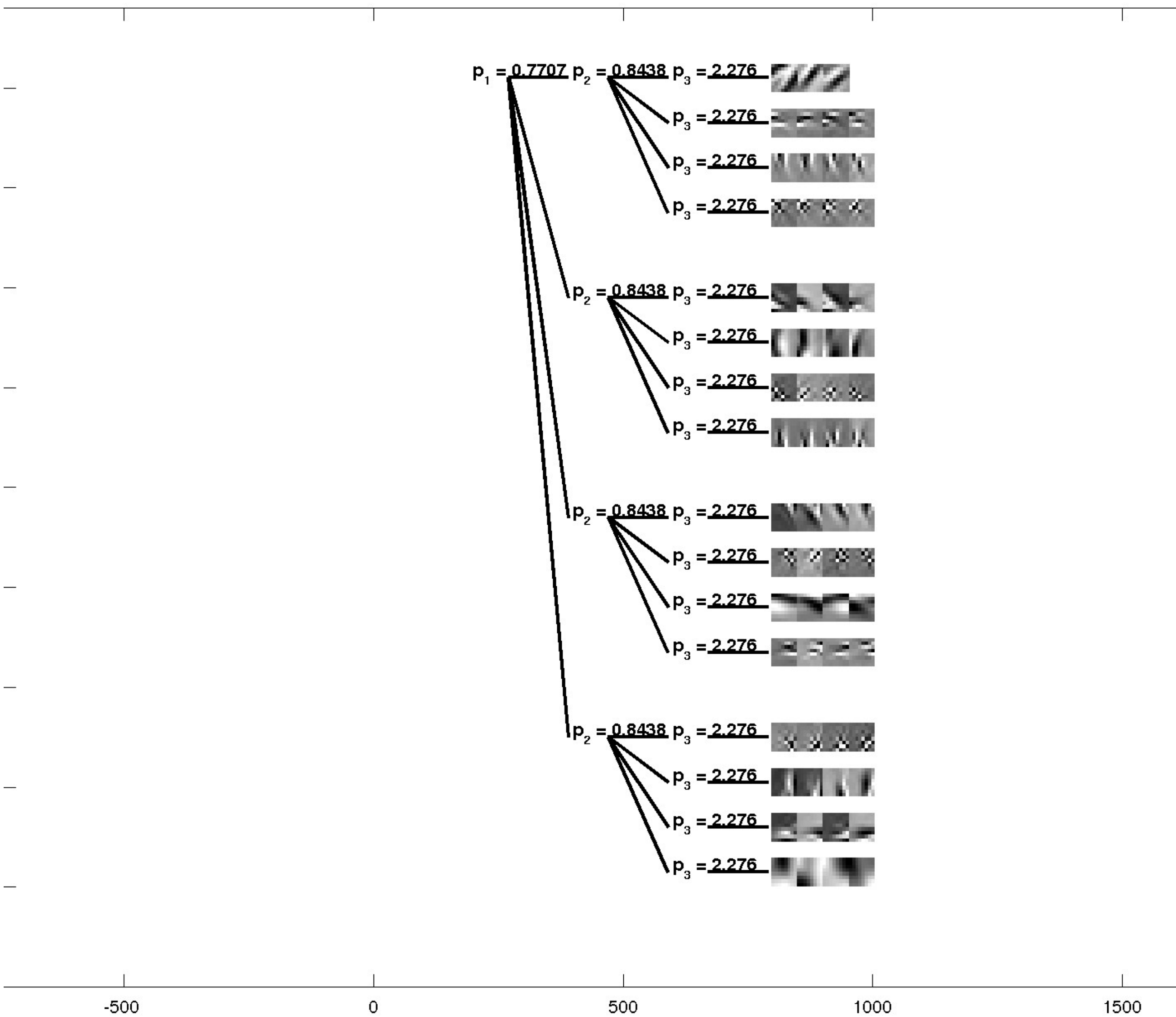
Independent Subspace ISA for 8 Subspaces with CGC.



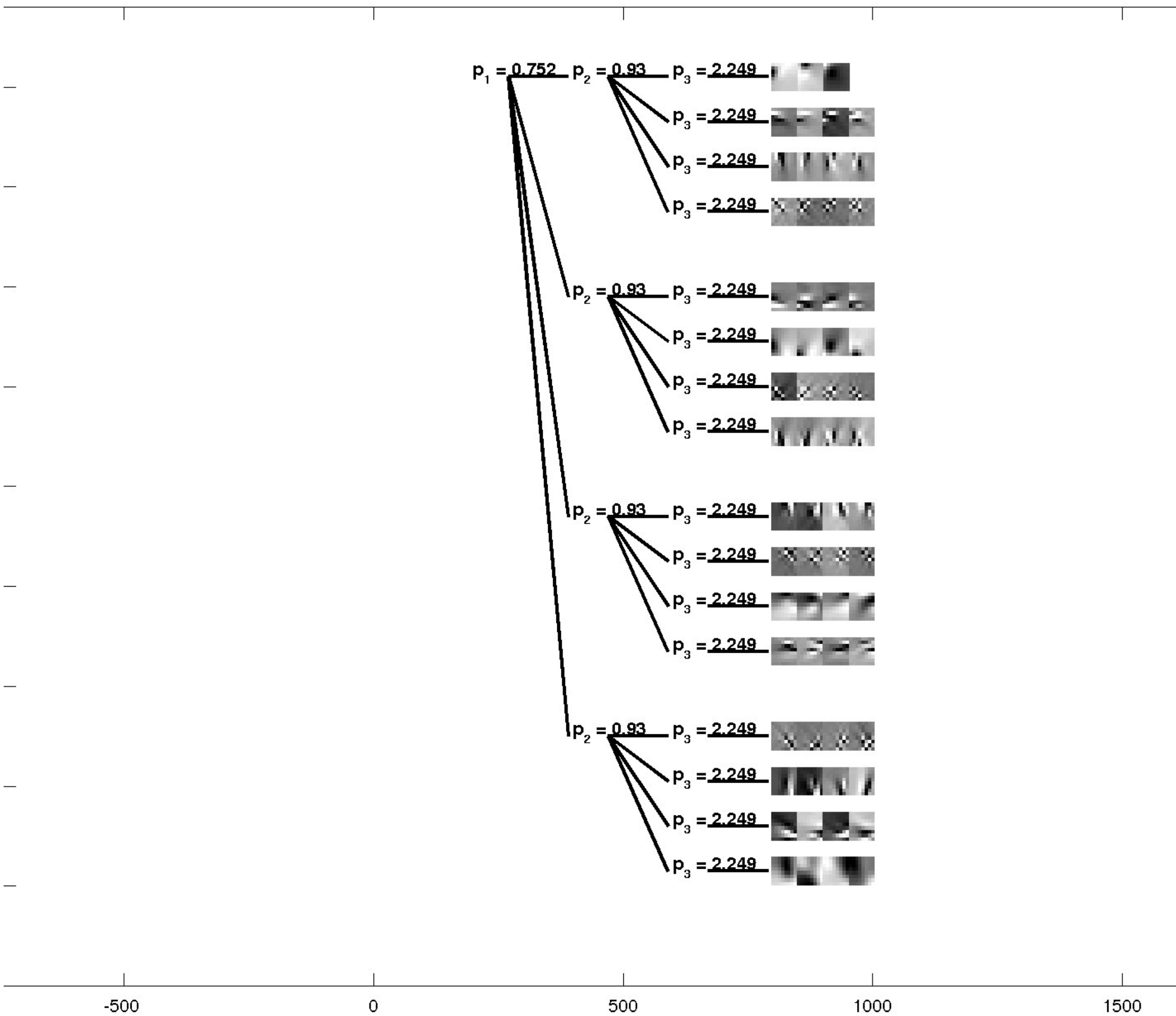
Independent Subspace ISA for 16 Subspaces with CGC.



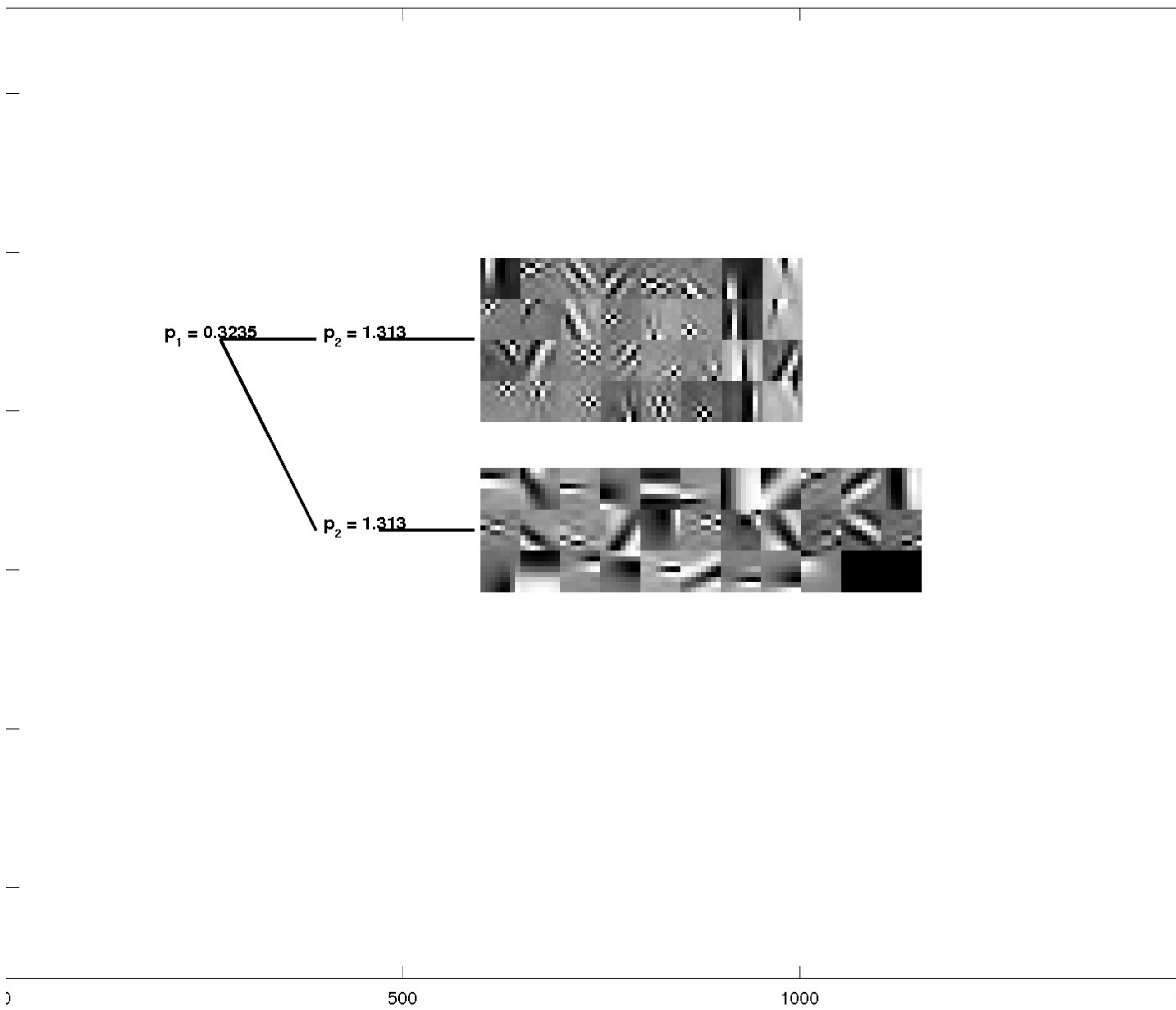
L_p-nested model with DT tree structure without CGC.



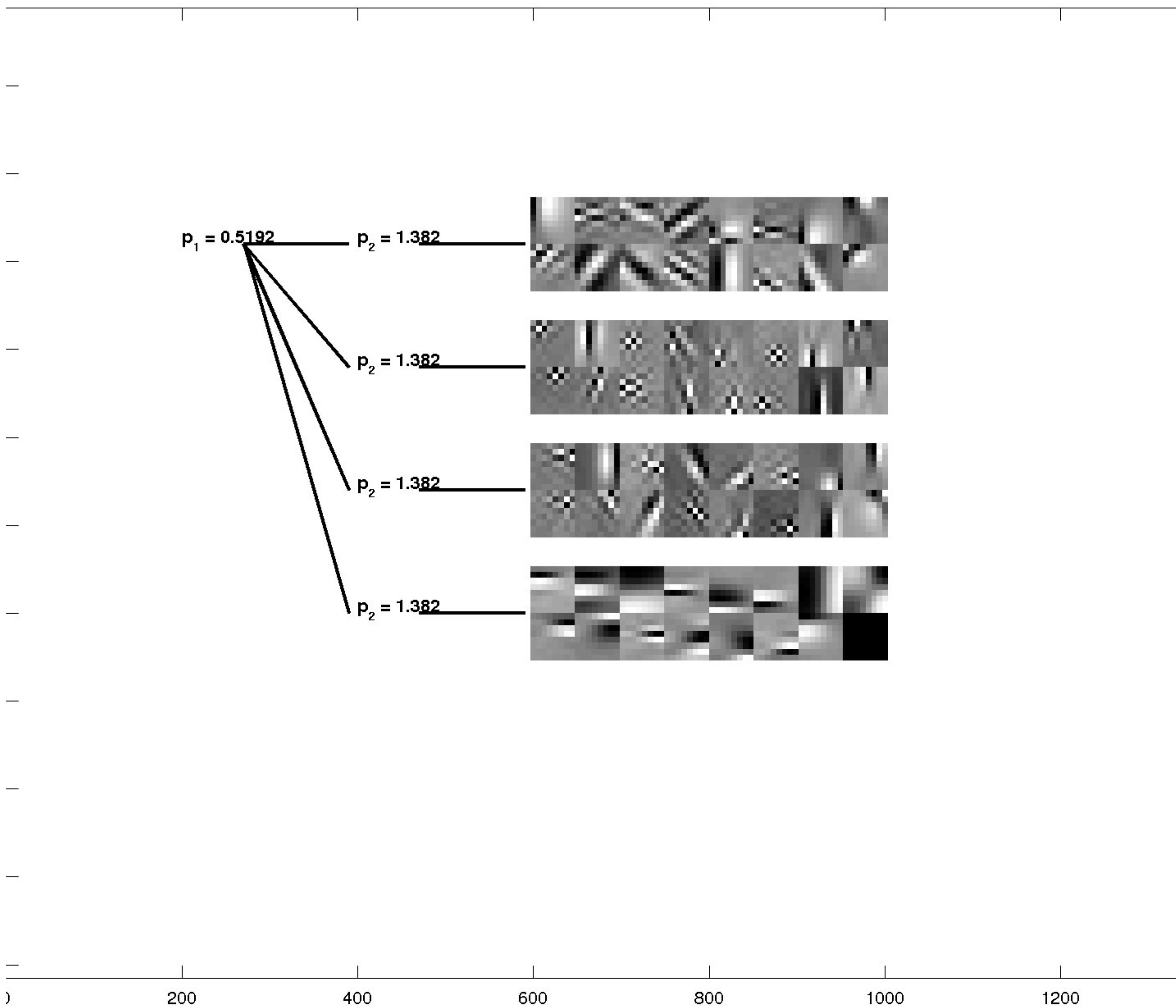
L_p-nested model with DT tree structure with CGC.



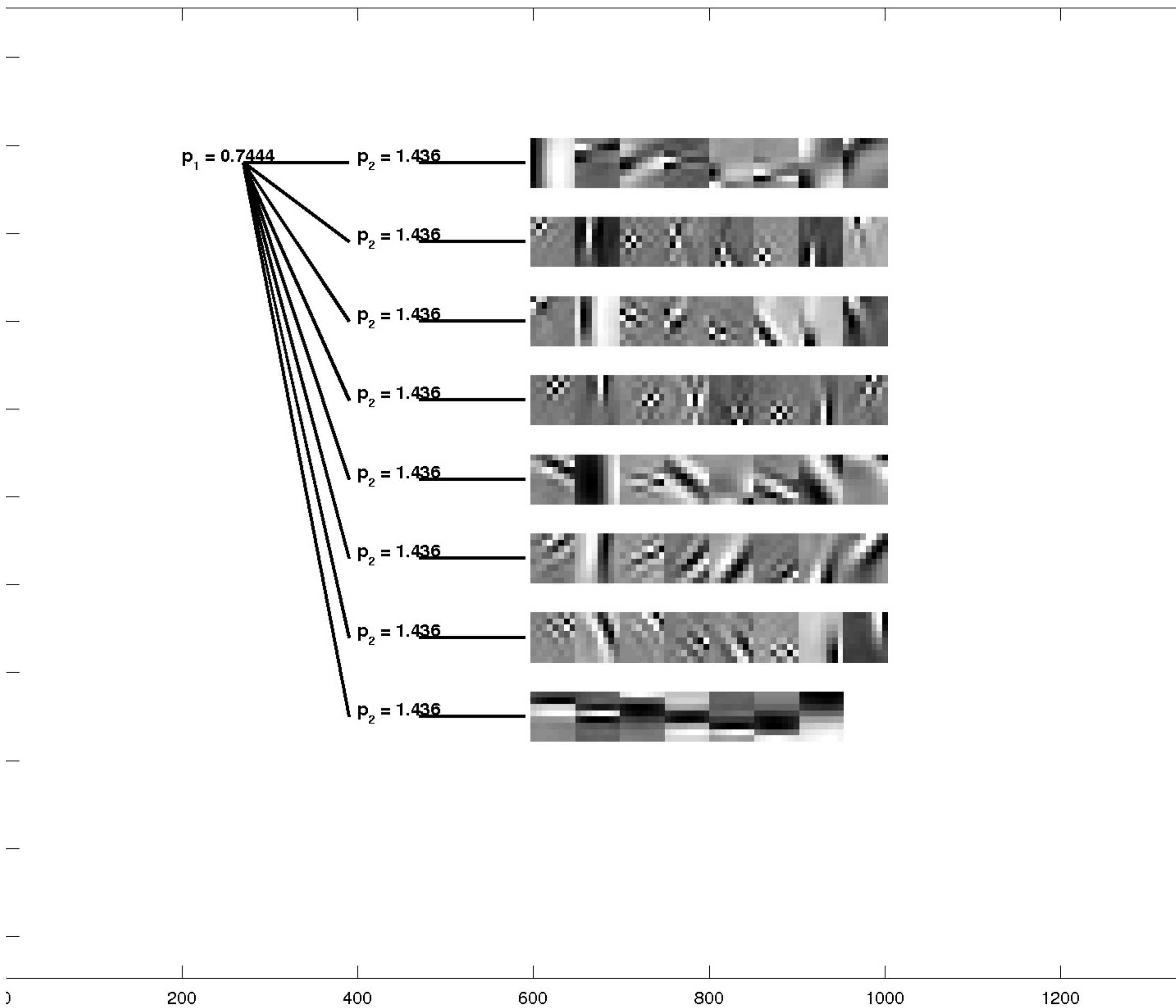
L_p-nested model with PND_2 tree structure without CGC.



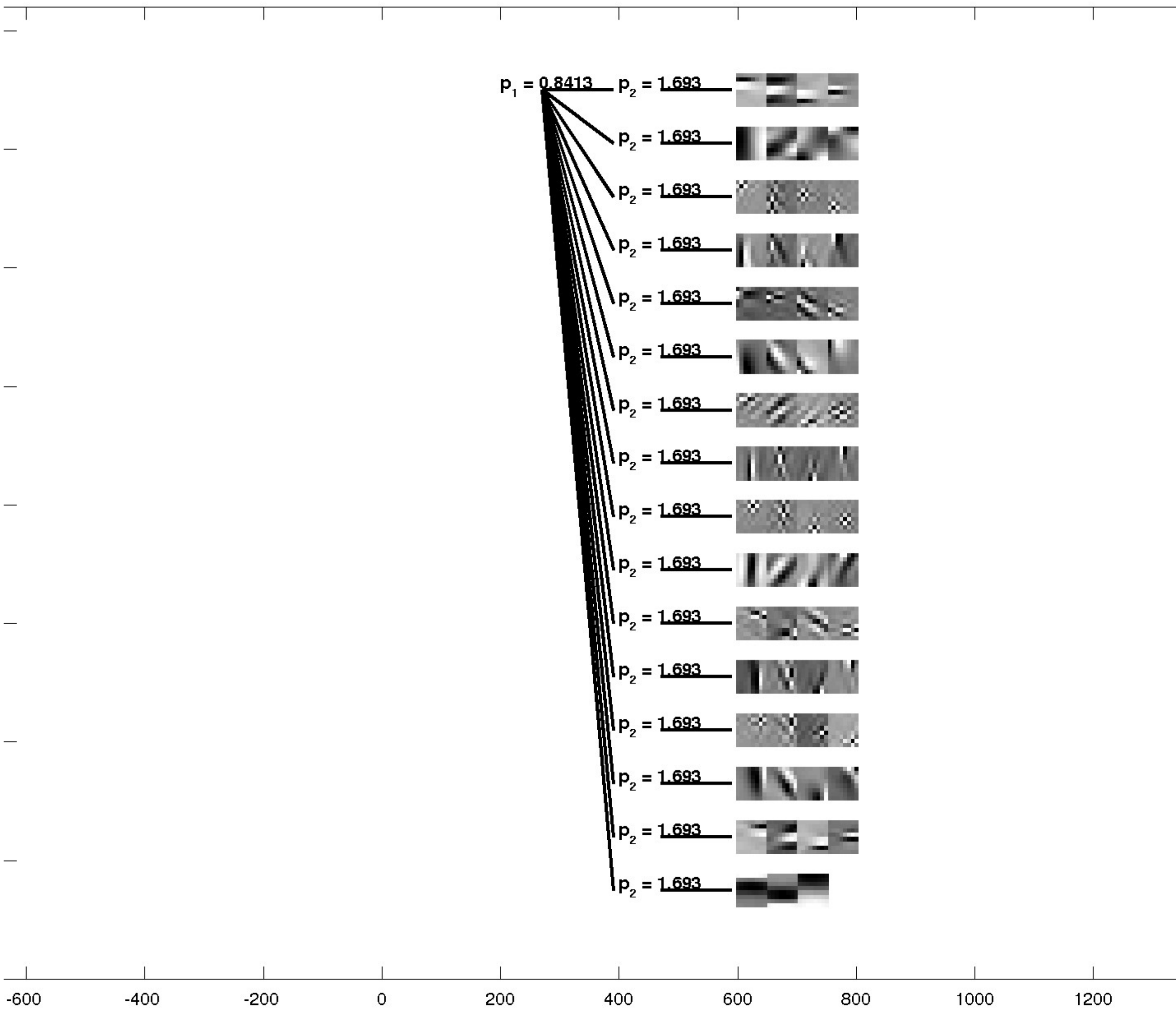
L_p-nested model with PND_4 tree structure without CGC.



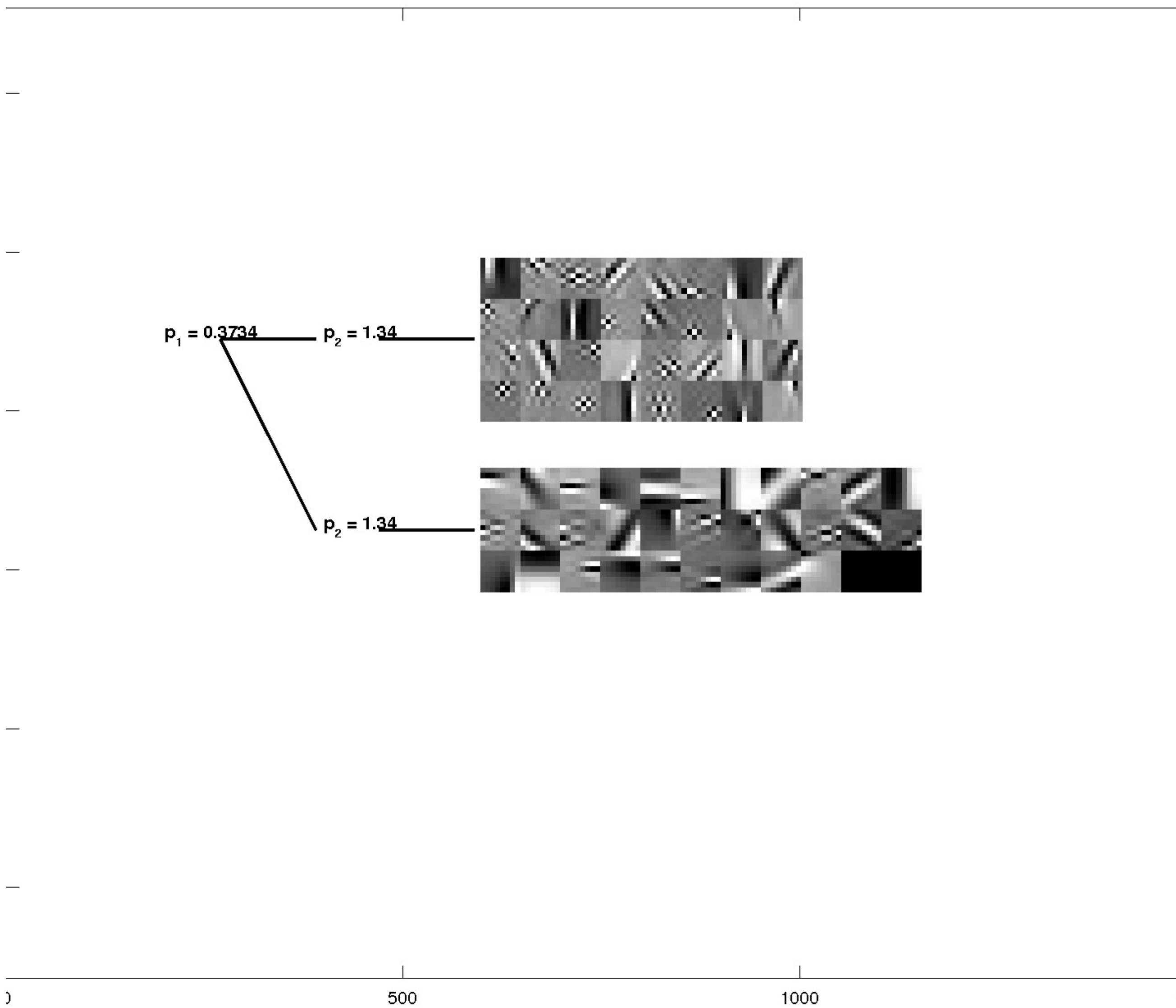
L_p-nested model with PND_8 tree structure without CGC.



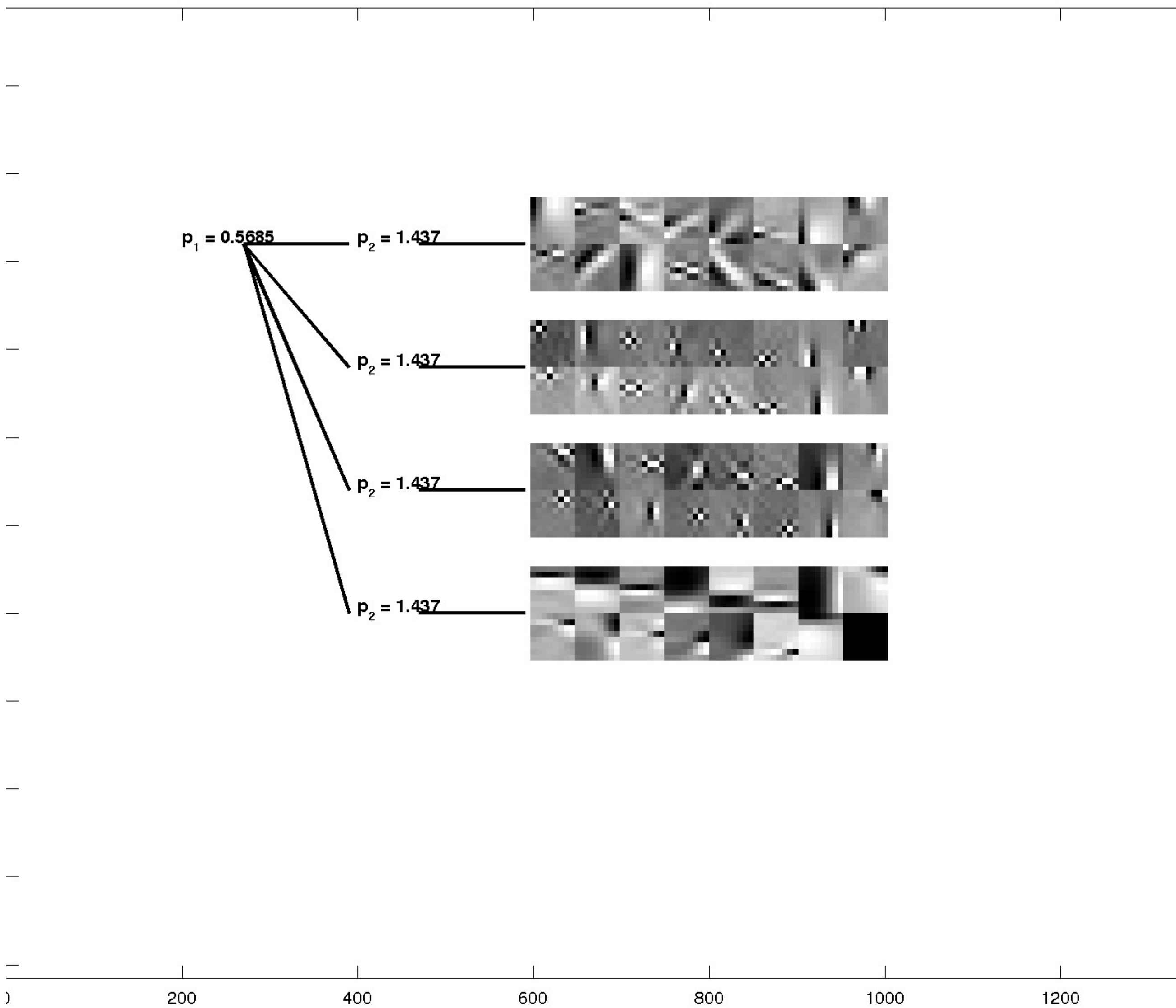
L_p-nested model with PND_{16} tree structure without CGC.



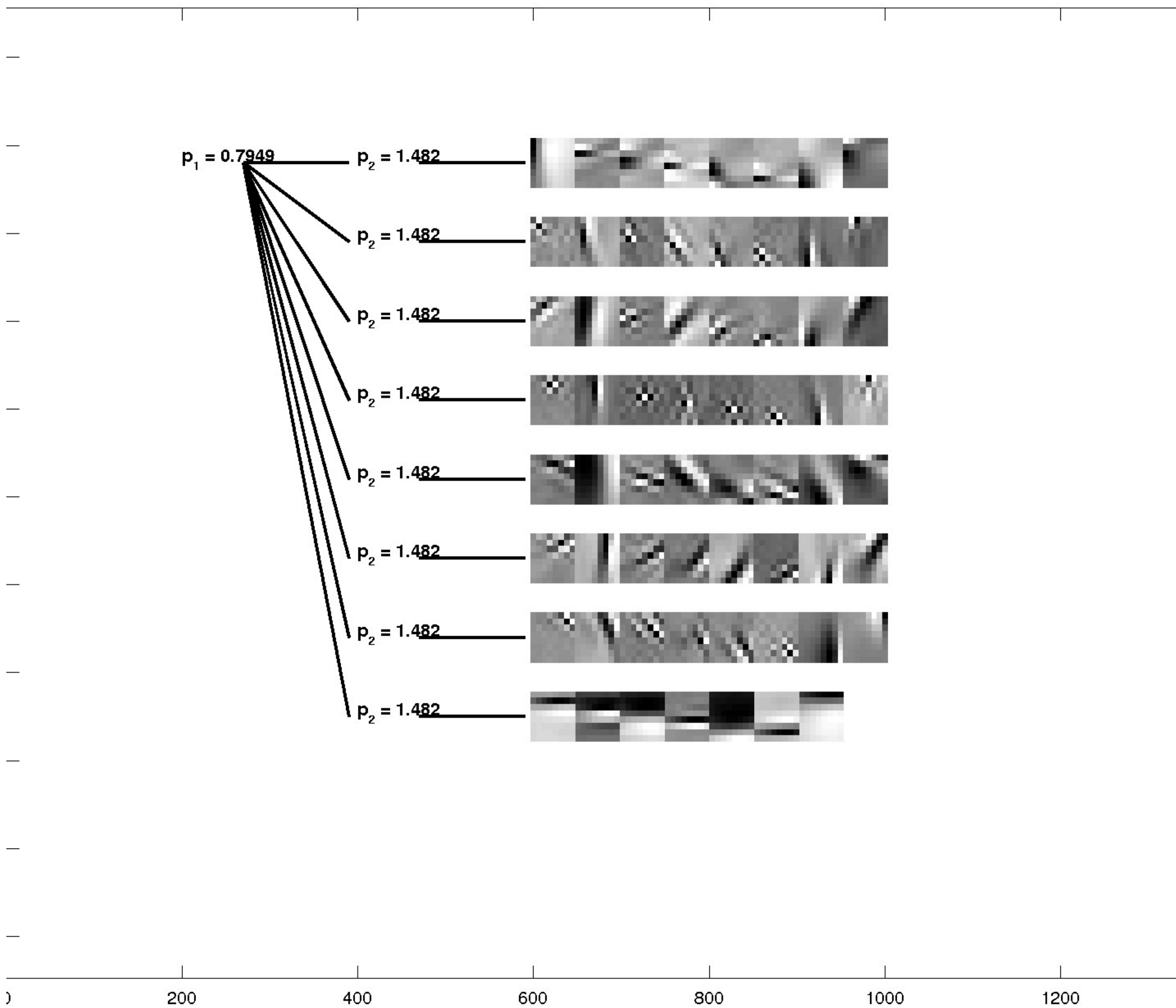
L_p-nested model with PND_2 tree structure with CGC.



L_p-nested model with PND_4 tree structure with CGC.



L_p-nested model with PND_8 tree structure with CGC.



L_p-nested model with PND_{16} tree structure with CGC.