ARTICLE **Communicated by Yan Karklin**

A Theory of Slow Feature Analysis for Transformation-Based Input Signals with an Application to Complex Cells

Henning Sprekeler
henning.sprekeler@epfl.ch
Laurenz Wiskott
wiskott@ini.rub.de
Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Berlin, Germany

We develop a group-theoretical analysis of slow feature analysis for the case where the input data are generated by applying a set of continuous transformations to static templates. As an application of the theory, we analytically derive nonlinear visual receptive fields and show that their optimal stimuli, as well as the orientation and frequency tuning, are in good agreement with previous simulations of complex cells in primary visual cortex (Berkes and Wiskott, 2005). The theory suggests that side and end stopping can be interpreted as a weak breaking of translation invariance. Direction selectivity is also discussed.

1 Introduction

More than half a century has passed since the first characterization of the response behavior of cells in primary visual cortex (V1). Despite extensive research, both experimental and theoretical, the processes that shape the structure of their receptive fields are still a matter of debate. Several mechanisms have been proposed, ranging from optimal coding strategies claiming that the receptive fields are matched to the statistics of natural stimuli (Olshausen & Field, 1996), to genetically determined, "hard-wired" (McLaughlin & O'Leary, 2005) or statistical connectivity patterns (Ringach, 2007). Although both approaches can explain aspects of V1 receptive fields, both suffer from a basic dilemma. On the one hand, the idea that receptive fields are learned from natural stimuli is seriously challenged by the experimental finding that in some species, simple cell receptive fields are largely developed before the animal first opens its eyes (Hubel & Wiesel, 1963). On the other hand, the notion that the early visual system is mostly hard-wired is problematic, because it has been shown that it remains plastic and can adapt to the statistics of artificial stimuli on a timescale of minutes

H. S. is now at Laboratory for Computational Neuroscience, Ecole Polytechnique Fédérale de Lausanne, Switzerland. L. W. is now at Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany.

Neural Computation **23, 303–335** (2011)

(Yao & Dan, 2001). Thus, the receptive fields must at least be compatible with natural stimuli because they would otherwise be unlearned quickly. One possible way of establishing this compatibility is that the morphology of the early visual system has adapted to natural stimuli on an evolutionary timescale. A different possibility is that spontaneous retinal activity occurring before eye opening is used for learning the receptive fields and that there are intrinsic similarities between the statistics of retinal waves and natural stimuli. But what is the nature of these similarities, and what type of learning rule could exploit them?

Recently it has been shown that the unsupervised learning paradigm of slowness can reproduce many aspects of complex cell receptive fields (Berkes & Wiskott, 2005). For training, the authors used quasi-natural image sequences that were generated from static natural images by applying transformations such as translation, rotation, and zoom. The simulations yielded a set of quadratic functions that generated slowly varying output signals under the constraint of being uncorrelated. The resulting functions shared several properties with complex cells in V1, including grating-shaped optimal stimuli and different types of selectivity to orientation and frequency. What makes this study interesting in the context of the debate above is that the authors performed test simulations to evaluate which aspects of the training data were responsible for the structure of the resulting receptive fields. They found that although higher-order image statistics were accessible to the learning paradigm, they were immaterial and that the same receptive fields could be learned with colored noise images. If the transformations that were used to generate the image sequences were changed, however, the properties of the receptive fields changed drastically. It is thus tempting to speculate that receptive fields of V1 complex cells are not adapted to higher-order statistics of natural stimuli but rather to transformations that typically occur in natural stimuli. Intriguingly, these transformations could also be present in retinal waves, as one could interpret propagating or rotating waves as an imitation of translation or rotation in natural stimuli.

Based on the observation that the structures of the receptive fields of Berkes and Wiskott (2005) were dominated by the transformations in the image sequences, we present a mathematical analysis of slow feature analysis for the case where the input data are generated by a set of continuous transformations and develop a group-theoretical framework that cumulates in an eigenvalue equation for the optimal functions. One of the main results of the analysis is that under the assumption that the input statistics are invariant under the transformations used to generate their time dependence, the optimal functions are purely determined by the transformations and are, counterintuitively, independent of all other aspects of the input statistics.

We then apply this framework to the scenario simulated by Berkes and Wiskott (2005) and show that several of the observed receptive field properties can be derived analytically. The theory provides an intuitive understanding for the structure of the receptive fields found in the simulations: (fast) translation is responsible for the optimal stimuli, and rotation and zoom underlie the orientation and frequency dependence, respectively. The tuning properties of the simulated cells can be understood as a way of generating harmonic oscillations as output signals if the transformations are applied with constant velocity, in line with previous analytical results (Wiskott, 2003). In addition, the analysis raises a link to previous group-theoretical approaches that learn the dynamical structure of image sequences in terms of the generators of the underlying transformations (Rao & Ruderman, 1998; Miao & Rao, 2007).

In section 2 we introduce slow feature analysis, the learning paradigm that Berkes and Wiskott (2005) used for their simulations. In section 3 we develop the mathematical framework that leads to the eigenvalue equations for the optimal functions. In section 4 we apply this framework to the simulations of Berkes and Wiskott (2005) and show that under certain assumptions, a closed-form solution can be found and that it can account for the optimal stimuli, as well as the orientation and frequency selectivity, of the simulated cells.

2 Slow Feature Analysis

Slow feature analysis (SFA) aims at minimizing temporal variations in a set of output signals $y_j(t) = g_j(\mathbf{x}(t))$ generated from a given time-dependent, vectorial input signal $\mathbf{x}(t)$. The optimization is performed on the functions g_j, which are constrained to lie within a given function space $\mathcal{F}$. How quickly an output signal $y(t)$ varies in time is quantified by the Δ-value, which is defined as the temporal average of the squared temporal derivative $\Delta(y) = \langle \dot{y}^2 \rangle_t$. To avoid the trivial constant solution and degeneracies arising from possible additions of arbitrary constants, the output signals are constrained to have zero mean and unit variance. The optimization is performed sequentially with an asymmetric decorrelation constraint: the function g_1 is optimized first, yielding the slowest possible signal y_1. Next, the second function g_2 is optimized under the constraint that its output signal y_2 is decorrelated from the first output signal y_1. The third function g_3 is the one that generates the slowest possible output signal y_3 under the constraint of being decorrelated from y_1 and y_2, and so on. Iterating this scheme yields a set of functions g_j that are ordered by slowness (i.e., by their Δ-value).

Mathematically, this problem can be formulated as a sequential optimization problem:

Given a function space $\mathcal{F}$ and an N-dimensional input signal $\mathbf{x}(t)$ find a sequence of J real-valued input-output functions $g_j(\mathbf{x})$ such that the output signal $y_j(t) := g_j(\mathbf{x}(t))$ minimizes

$$\Delta(y_j) = \langle \dot{y}_j^2 \rangle_t \tag{2.1}$$

under the constraints

$$\langle y_j \rangle_t = 0 \quad \textit{(zero mean)}, \tag{2.2}$$

$$\langle y_j^2 \rangle_t = 1 \quad \textit{(unit variance)}, \tag{2.3}$$

$$\forall i < j : \langle y_i y_j \rangle_t = 0 \quad \textit{(decorrelation and order)}, \tag{2.4}$$

with $\langle \cdot \rangle_t$ *and* $\dot{y}$ *indicating temporal averaging and the derivative of y, respectively.*

The decorrelation constraint, equation 2.4, ensures that different functions g_j code for different aspects of the input. A more complete constraint would be that the mutual information vanishes for different output signals. Unfortunately, the step from decorrelation to statistical independence is nontrivial in terms of algorithmic implementation. Decorrelation is an approximation of statistical independence up to second-order statistics.

It is important to note that although the objective is the slowness of the output signal, the functions g_j are instantaneous functions of the input, so that slowness cannot be enforced by low-pass filtering. Slow output signals can be obtained only if the input signal contains slowly varying features that can be extracted by the functions g_j.

Depending on the dimensionality of the function space $\mathcal{F}$, the solution of the optimization problem requires different techniques. If $\mathcal{F}$ is finite-dimensional, the problem can be reduced to a (generalized) eigenvalue problem (Wiskott & Sejnowski, 2002; Berkes & Wiskott, 2005). Here, we will consider the case of an infinite-dimensional function space $\mathcal{F}$ that can be solved using standard techniques of functional analysis.

3 Theory

In this section, we develop a theory of slow feature analysis for input data that have the structure that Berkes and Wiskott (2005) used. We assume that the input signals consist of a sequence of trials, each generated by applying time-dependent continuous transformations to a static template $\mathbf{x}^{\mu}$, which is generally different for every trial.

3.1 Assumptions and Notation

3.1.1 Input Data Generation: Transformation Group. We make two assumptions for the transformations that are used to generate the input data: (1) they should be invertible (i.e., for every transformation there is an inverse transformation, which is also allowed in the paradigm used), and (2) the transformations should be continuous. The latter assumption arises naturally, because SFA requires temporally continuous input data. Mathematically, these assumptions imply that the transformations form a Lie group, that is, a continuous group.

The generation of one trial of the training data can be written as

$$\mathbf{x}(t) = T_{\mathbf{x}}(t)\mathbf{x}^{\mu}\,, \tag{3.1}$$

where $T_{\mathbf{x}}(t)$ is an operator that maps the input signal at time $t = 0$ (i.e. the template) to the input signal at time t. The set of all possible operators $T_{\mathbf{x}}$ forms a representation of the transformation group on the vector space that contains the input data.

A different representation of the transformation group can be constructed by defining operators T_g that act on the functions in $\mathcal{F}$ such that

$$(T_g g)(\mathbf{x}) := g(T_{\mathbf{x}}\mathbf{x}) \tag{3.2}$$

is fulfilled for all functions $g \in \mathcal{F}$ and all input signals $\mathbf{x}$. Note that this definition immediately implies that the operators T are linear operators on the function space $\mathcal{F}$, since

$$\begin{aligned}\left(T_g(g_1 + g_2)\right)(\mathbf{x}) &\overset{(6)}{=} (g_1 + g_2)(T_{\mathbf{x}}\mathbf{x}) = g_1(T_{\mathbf{x}}\mathbf{x}) + g_2(T_{\mathbf{x}}\mathbf{x})\\ &= \left(T_g g_1\right)(\mathbf{x}) + \left(T_g g_2\right)(\mathbf{x}).\end{aligned} \tag{3.3}$$

Intuitively, the representation change, equation 3.2, corresponds to a change of the coordinate system. Think of the function g as a measurement device that extracts certain aspects of the input signal. Then instead of changing the input signal that the function g acts on (this is the effect of $T_{\mathbf{x}}$), one may also change the function in the "opposite direction" (this is the effect of T_g). In the following, we skip the subscript g, because all transformation operators T will act functions, not input signals.

3.1.2 The Hilbert Space of Functions. The function space $\mathcal{F}$ forms a vector space. It is convenient to turn it into a Hilbert space by defining the scalar product

$$(f, g) = \left\langle f(\mathbf{x}(t))g(\mathbf{x}(t))\right\rangle, \tag{3.4}$$

where the average $\langle\cdot\rangle$ is taken over the training data (i.e., over all trials and times within the trials). For simplicity of notation, we omit the average over trials in the following and act as if there was only one trial. All our considerations are valid for an ensemble of trials as well, but many quantities would need additional indices that would only clutter the equations. For the same reason, we often skip the argument of the functions g.

Note that if the functions g_j have zero mean on the training data, this scalar product measures the covariance between the output of the functions f and g. Consequently, the unit variance and decorrelation constraints (see equations 2.3 and 2.4) take the compact form of an orthonormality constraint:

$$(g_i, g_j) = \delta_{ij}\,, \tag{3.5}$$

where δ_{ij} denotes the Kronecker symbol.

For all the derivations that follow, it is assumed that the scalar product exists (i.e., that it is finite) for all functions f and g it acts on. This excludes, for example, functions with infinite variance and thus inflicts constraints on the function space $\mathcal{F}$. Notice that these constraints also depend on the statistics of the training data.

3.1.3 Invariance of the Training Data. In the following, we assume that the statistics of the input data are invariant with respect to the transformations applied. The main argument for this assumption is that the training data are generated by applying these transformations. If we assume, for example, that the temporal derivative of the transformations $T_g(t)$ is independent of the content $\mathbf{x}(t)$ (and therefore the current transformation $T_g(t)$), there is no reason that a transformed version of the image should be less likely than the original one.[1] Of course, this assumption is not fulfilled for all sets of transformation-based training data, so its validity needs to be checked for the application at hand. For example, for the application to receptive fields later in the article, we can argue that natural image statistics are largely translation and rotation invariant (but see Ruderman & Bialek, 1994) and show some degree of scale invariance (Ruderman & Bialek, 1994; Dong & Atick, 1995; Dong, 2001).

The invariance of the input statistics means that if the whole ensemble of input signals used for training is subjected to any of the transformations, the resulting ensemble of input signals has the same statistics. Thus, averages of arbitrary functions remain unchanged by the transformation. In particular, this implies that the scalar product, equation 3.4, is invariant with respect to all operators T in the transformation group:

$$(Tf, Tg) = (f, g). \tag{3.6}$$

[1] In this case, the dynamics of the transformation $T(t)$ is a pure diffusion process, so that the probability density of the operators T on the transformation manifold converges to a uniform distribution. If all transformations are equally likely, all transformed input data are of course also equally likely—the input statistics are invariant under the transformations. This argument is valid only if the duration of the trial is much longer that the mixing time of the stochastic process $T(t)$ and if the transformation manifold is bounded. Otherwise additional assumptions on the statistics of the templates $\mathbf{x}^{\mu}$ have to be made.

In other words, the transformation operators T are orthogonal with respect to the scalar product, equation 3.4, that is, they preserve distances (i.e., the standard deviation of differences of output signals) and angles (which are related to the correlation of output signals) in the function space $\mathcal{F}$ as derived from the scalar product, equation 3.4.

3.1.4 Generators of the Transformations. The transformation operators T form a manifold, embedded in the space of all linear operators on the function space $\mathcal{F}$. Often this manifold is of relatively low dimensionality. For example, if we apply translation, rotation, and zoom to images, the associated operators can be characterized by the translation vector (two degrees of freedom), the rotation angle, and the zoom factor. The operator manifold would thus be four-dimensional. Despite its low dimensionality, the manifold can in principle have a very complicated structure, so that its low dimensionality is not necessarily helpful. Here, we show that the low dimensionality of the manifold, in combination with the invariance assumption introduced above, has important implications for the temporal derivative of time-dependent transformations.

Let us start with a simple example, in which the input signal x is one-dimensional and the transformation is the addition of a constant a to x: $x \to x + a$. The associated transformation on the function space $\mathcal{F}$ is a translation of the functions g: $g(x) \to g(x + a)$. Applying a time-dependent transformation $T(t)$ to a function g amounts to a time-dependent translation by $a(t)$: $[T(t)g](x) = g(x + a(t))$. SFA focuses on the temporal derivative of the output signal $y(t) = g(x + a(t))$, which can readily be calculated using the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}t} y(t) = \frac{\mathrm{d}}{\mathrm{d}t}[T(t)g](x) = \frac{\mathrm{d}}{\mathrm{d}t} g(x + a(t)) \tag{3.7}$$

$$= \left[\dot{x}(t) \frac{\mathrm{d}}{\mathrm{d}x} g\right](x + a(t)) \tag{3.8}$$

$$= \left[T(t) \left(\dot{x}(t) \frac{\mathrm{d}}{\mathrm{d}x}\right) g\right](x) \tag{3.9}$$

$$=: [T(t)Q(t)g](x). \tag{3.10}$$

This shows that the time derivative of the output signal can be calculated by applying an operator $Q(t)$ to the function g before applying the transformation $T(t)$. $Q(t) := \dot{x}(t)\frac{\mathrm{d}}{\mathrm{d}x}$ is a linear operator that consists of a time-dependent scalar $\dot{x}$ and an operator $\frac{\mathrm{d}}{\mathrm{d}x}$ that does not depend on the current translation $T(t)$. In the language of Lie group theory, the differential operator $\frac{\mathrm{d}}{\mathrm{d}x}$ takes the role of the generator of translations. The associated scalar $\dot{x}$ is simply the velocity of the translation. As shown in the following, this scheme can be generalized to the case of arbitrary transformation groups.

Each transformation type has its own generator, which is independent of the transformation itself and associated with a generalized velocity of the transformation.

To show that this generalization holds, let us now consider the general case. In the transformation operator notation, the output signal within one trial is generated by

$$y(t) = (T(t)g)(\mathbf{x}^{\mu}). \tag{3.11}$$

Taking the derivative yields

$$\frac{\mathrm{d}}{\mathrm{d}t} y(t) = \left(\frac{\mathrm{d}}{\mathrm{d}t} T(t) g\right)(\mathbf{x}^{\mu}) \tag{3.12}$$

$$=: [T(t)Q(t)g]\,(\mathbf{x}^{\mu}), \tag{3.13}$$

with $Q(t) := T^{-1}(t)[\frac{\mathrm{d}}{\mathrm{d}t} T(t)]$.

Our goal is to write the operators $Q(t)$ as a sum of products of generators G_α (one for each transformation α) and time-dependent velocities $v_\alpha(t)$:

$$Q(t) = \sum_\alpha v_\alpha(t) G_\alpha. \tag{3.14}$$

As in the example, the generators G_α should be independent of the transformation. Therefore, equation 3.14 implies that the operator $Q(t)$ is an element of a vector space that is spanned by the generators G_α. The following theorem shows that this vector space is the tangent space of the transformation group at the identity element:

Theorem 1. *Let $T(t)$ be a differentiable trajectory of transformations with $T(t)$ element of a Lie transformation group for all t. Then for all t, $Q(t) := T^{-1}(t)[\frac{\mathrm{d}}{\mathrm{d}t} T(t)]$ is an element of the tangent space of the transformation group at the identity element E.*

Proof. It is sufficient to show that for all times t, there is a trajectory $\tilde{T}(s)$ of transformation operators such that $\tilde{T}(s)|_{s=t} = E$ and $\frac{\mathrm{d}}{\mathrm{d}s}\tilde{T}(s)|_{s=t} = Q(t)$. It is easy to see that $\tilde{T}(s) := T^{-1}(t)T(s)$ fulfills these conditions.

This theorem implies that the expansion 3.14 of the operator Q in terms of the generators is correct if we use a set of generators that spans the tangent space of the group at the unit element. Because the tangent space of the group has the same dimensionality as the group itself, it also implies that we need as many generators as the dimensionality of the group, that is, one generator per transformation.

As seen above, the invariance of the training data under the transformation implies that the transformation operators are orthogonal with respect to the scalar product, equation 3.4. A consequence that we will make use of in the following is that the generators G_α are anti-self-adjoint:

Theorem 2. *The generators G_α are anti-self-adjoint with respect to the scalar product, equation 3.4, that is,*

$$(f, G_\alpha g) = -(G_\alpha f, g) \tag{3.15}$$

for all $f, g \in \mathcal{F}$.

Proof. G_α is an element of the tangent space of the transformation group at the identity element. Thus, there is a trajectory $T(s)$ of transformation operators such that $\left[\frac{\mathrm{d}}{\mathrm{d}s}T(s)\right]_{s=0} = G_\alpha$ and $T(0) = E$. Because $T(s)$ is orthogonal for all values of s, $(T(s)f, T(s)g) = (f, g)$ is independent of s for arbitrary f, g. Thus

$$0 = \left[\frac{\mathrm{d}}{\mathrm{d}s}(T(s)f, T(s)g)\right]_{s=0} \tag{3.16}$$

$$= \left[\left(\frac{\mathrm{d}}{\mathrm{d}s}T(s)f, T(s)g\right) + \left(T(s)f, \frac{\mathrm{d}}{\mathrm{d}s}T(s)g\right)\right]_{s=0} \tag{3.17}$$

$$= (G_\alpha f, g) + (f, G_\alpha g), \tag{3.18}$$

which proves the assertion.

3.2 Reformulation of the Slowness Objective. The conventions introduced in the last section allow us to rewrite the slowness objective 2.1:

$$\Delta(g) \overset{(2.1)}{:=} \langle \dot{y}(t)^2 \rangle \tag{3.19}$$

$$\overset{(3.13)}{=} \langle ([T(t)Q(t)g]\,(\mathbf{x}^\mu))^2 \rangle \tag{3.20}$$

$$\overset{(3.14)}{=} \left\langle \left(\left[\sum_\alpha v_\alpha(t) T(t) G_\alpha g \right] (\mathbf{x}^\mu) \right)^2 \right\rangle \tag{3.21}$$

$$= \sum_{\alpha,\beta} \left\langle v_\alpha(t)\,[T(t)G_\alpha g]\,(\mathbf{x}^\mu) v_\beta(t)\,\left[T(t)G_\beta g\right](\mathbf{x}^\mu)\right\rangle. \tag{3.22}$$

Assuming that the velocities v_α are statistically independent of the transformation, we can split the average and express the Δ-value in the form of a scalar product:

$$\Delta(g) \overset{(3.22)}{=} \sum_{\alpha,\beta} \left\langle v_\alpha(t) v_\beta(t) \right\rangle \left\langle [T(t)G_\alpha g]\,(\mathbf{x}^\mu)\,\left[T(t)G_\beta g\right](\mathbf{x}^\mu) \right\rangle \tag{3.23}$$

$$\overset{(3.2,3.1)}{=} \sum_{\alpha,\beta} \langle v_\alpha(t) v_\beta(t) \rangle \langle [G_\alpha g]\,(\mathbf{x}(t))\,\left[G_\beta g\right](\mathbf{x}(t)) \rangle \tag{3.24}$$

$$\overset{(3.4)}{=} \sum_{\alpha,\beta} \left\langle v_\alpha(t) v_\beta(t) \right\rangle (G_\alpha g, G_\beta g) \tag{3.25}$$

$$\overset{(3.15)}{=} \Big(g, \underbrace{\Big[-\sum_{\alpha,\beta} \left\langle v_\alpha v_\beta \right\rangle G_\alpha G_\beta \Big]}_{=:\mathcal{D}} g\Big) \tag{3.26}$$

$$= (g, \mathcal{D}g). \tag{3.27}$$

Because the operator $\mathcal{D}$ is a bilinear combination of the anti-self-adjoint generators G_α, it is self-adjoint:

$$(f, \mathcal{D}g) = (\mathcal{D}f, g) \quad \forall f, g \in \mathcal{F}. \tag{3.28}$$

3.3 An Eigenvalue Equation for the Optimal Solutions. The main advantage of this reformulation of the objective function is that the optimization problem that underlies SFA takes a form that is common in other contexts and for which a well-developed theory exists. Most important, it is known that the functions that minimize equation 3.27 are the eigenfunctions of the operator $\mathcal{D}$, while the fact that $\mathcal{D}$ is self-adjoint ensures that the eigenfunctions are orthonormal so that the constraints 3.5 are fulfilled (for the relevant mathematical background (see, e.g., Landau & Lifshitz, 1977, or Courant & Hilbert, 1989)). The eigenvalues Δ_j are the Δ-values of the eigenfunctions. We can thus solve the optimization problem of SFA by finding the J solutions of the eigenvalue equation

$$\mathcal{D}g_j = \Delta_j g_j, \tag{3.29}$$

with the smallest eigenvalues Δ_j.

The first important result of this section is that this equation is independent of the templates $\mathbf{x}^\mu$ underlying the training data. Instead, it depends purely on the nature of the transformations that produce the image sequences (as reflected by the generators G_α) and the second-order moments $\left\langle v_\alpha v_\beta \right\rangle$ of the associated velocities v_α. This explains the finding by Berkes and Wiskott (2005) that the simulated receptive fields for training sequences that were generated from colored noise images and those for sequences generated from natural images were essentially the same. On the other hand, a change in the transformations used to generate the image sequences changes the structure of the operator $\mathcal{D}$ and thus the resulting receptive fields. This is also in agreement with the simulations. It is important to bear in mind that although equation 3.29 appears to be independent of the image statistics, it is valid only if they are invariant with respect to the transformations. If this invariance condition is not fulfilled, higher-order statistics may play a role.

Solving the eigenvalue equation 3.29 requires that we know $\mathcal{D}$, which according to equation 3.26 corresponds to knowing the generators G_α of the transformations and the matrix of the second moments $\langle v_\alpha v_\beta \rangle$ of the associated velocities v_α. For finite-dimensional function spaces, the generators have matrix form. In section 4, we present an application of the theory to continuous images. In this case, the input data, and therefore also the function space, are infinite-dimensional and the generators take the form of differential operators. Consequently, the eigenvalue equation 3.29 becomes a partial differential eigenvalue equation.

4 An Application to Complex Cell Receptive Fields

In this section, we derive and discuss an explicit solution of the eigenvalue problem 3.29 for the special case studied by Berkes and Wiskott (2005). The input data are image sequences generated by applying translation, rotation, and zoom to static template images.

4.1 Assumptions

4.1.1 Function Space: Quadratic Forms. We assume that the input data are continuous gray-scale images, with $x(\mathbf{r})$ denoting the gray value at pixel position $\mathbf{r}$. This implies that the input data are infinite-dimensional, so the input-output functions g_j for SFA are functionals that map images to real numbers.

Berkes and Wiskott (2005) have used quadratic functions: sums of monomials of first and second order in the pixel values. We neglect the linear component mainly for the reason that we later focus on functions that are translation invariant. The only linear function that is translation invariant is the mean pixel intensity, which is not very informative about the image. Moreover, Berkes and Wiskott found that in their simulations, the linear contribution to almost all optimal functions was negligible compared to the quadratic contribution. Therefore, we restrict the function space $\mathcal{F}$ to the space of quadratic functions of the images:

$$g[x(\mathbf{r})] = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} g(\mathbf{r}, \mathbf{r}')x(\mathbf{r})x(\mathbf{r}')\, \mathrm{d}^2r \ \mathrm{d}^2r', \tag{4.1}$$

where $g(\mathbf{r}, \mathbf{r}') = g(\mathbf{r}', \mathbf{r})$ is a symmetric function. For mathematical convenience, we assume that the images are infinitely large, so the integrals extend over $\mathbb{R}^2$. Note that $g(\mathbf{r}, \mathbf{r}')$ can be understood as coefficients of the representation of the function $g[x(\mathbf{r})]$ in terms of the basis functions $x(\mathbf{r})x(\mathbf{r}')$.

The analysis presented does not crucially rely on the particular choice of the function space. In particular, a generalization to polynomial mappings of arbitrary order is straightforward.

4.1.2 Transformations. Just like Berkes and Wiskott (2005), we restrict the transformations to translation, rotation, and zoom. One argument for choosing these particular transformations is that they are part of a systematic expansion of the image dynamics. To see this, we assume that image dynamics can be described in terms of a flow field $\mathbf{v}(\mathbf{r})$ that denotes the velocity at which the pixel at position $\mathbf{r}$ is moving, so that

$$\frac{\mathrm{d}}{\mathrm{d}t}x(\mathbf{r}, t) = \mathbf{v}(\mathbf{r}, t) \cdot \nabla_{\mathbf{r}} x(\mathbf{r}, t). \tag{4.2}$$

Assuming that the flow field is smooth, we can get a first-order approximation by a Taylor expansion in $\mathbf{r}$:

$$\mathbf{v}(\mathbf{r}, t) = \mathbf{v}_0(t) + \mathbf{A}(t)\mathbf{r}. \tag{4.3}$$

The spatially constant component $\mathbf{v}_0$ of this expansion corresponds to uniform translation of the image. To get an intuition for the linear component, we split the matrix $\mathbf{A}$ into four components: (1) a multiple of the unit matrix, corresponding to zoom, (2) an antisymmetric component, corresponding to rotation, and (3) two components that correspond to area-preserving combinations of compression or expansion:

$$\mathbf{A} = \underbrace{\dot{z}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\text{zoom}} + \underbrace{\omega\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{\text{rotation}} + \underbrace{c_1\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + c_2\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{\text{area-preserving compression/expansion}} . \tag{4.4}$$

Restricting the transformations to translation, rotation, and zoom therefore corresponds to a first-order Taylor expansion of the flow field, where the components that correspond to area-preserving combinations of compression and expansion are neglected.

Note that the description of the image dynamics in terms of a flow field is not complete, because objects in natural images can occlude each other, which cannot be fully captured using flow fields.

Table 1: Generators of the Transformations Used to Generate the Image Sequences.

Transformation	Generator	Velocity
Translation	$\nabla_{\mathbf{r}} + \nabla_{\mathbf{r}'}$	$\mathbf{v}_0$
Rotation	$r_1\partial_{r_2} - r_2\partial_{r_1} + r'_1\partial_{r'_2} - r'_2\partial_{r'_1}$	ω
Zoom	$\nabla_{\mathbf{r}} \cdot \mathbf{r} + \nabla_{\mathbf{r}'} \cdot \mathbf{r}' = \mathbf{r} \cdot \nabla_{\mathbf{r}} + \mathbf{r}' \cdot \nabla_{\mathbf{r}'} + 4$	$\zeta = \dot{z}/z$

Notes: ∂_{r_1} denotes the derivative with respect to the first component of $\mathbf{r}$. $\nabla_{\mathbf{r}}$ denotes the vector-valued operator $(\partial_{r_1}, \partial_{r_2})^T$.

4.1.3 Generators. To apply the theory developed in section 3 to the specific problem at hand, we need to know the representation of the generators for the transformations on the function space of quadratic forms. For reasons of compactness, we defer the derivation of the generators to the appendix. Suffice to say that we represent G_α such that they act on the kernel $g(\mathbf{r}, \mathbf{r}')$ of the quadratic functionals, equation 4.1. In this representation they become the differential operators listed in Table 1. The associated velocities are the translation velocity $\mathbf{v}_0$, the angular velocity ω for rotation, and a zoom velocity ζ for zoom. We define the zoom velocity as the factor by which the size of the image increases per time unit. Let z denote the factor by which an image has been zoomed relative to its original size. Then constant zoom velocity implies that the zoom factor z grows exponentially in time, so that not $\dot{z}$ is constant but rather $\frac{\dot{z}}{z} =: \zeta$.

With these generators the eigenvalue equation becomes a partial differential eigenvalue equation for $g(\mathbf{r}, \mathbf{r}')$. Finding a closed-form general solution is difficult, mainly because the resulting image depends on the order in which translation and rotation or zoom are applied. A mathematical implication is that the generators for translation and those for rotation and zoom do not commute, so they do not possess a common set of eigenfunctions (which would simplify the analysis significantly). However, in the special case of translation-invariant functions, it is possible to find a closed-form solution that explains the orientation and frequency dependence of the simulated receptive fields in Berkes and Wiskott (2005), as well as aspects of their optimal stimuli.

4.2 Translation-Invariant Solutions. But why translation invariance? There are two reasons. First, the control experiments performed in Berkes and Wiskott (2005) suggest that translation is a necessary and sufficient condition for the optimal functions to resemble complex cells. In simulations where translation was present in the training data, the functions became phase invariant and had optimal stimuli that resemble Gabor wavelets. The invariance of the units to spatial phase corresponds to a certain degree of translation invariance. Second, for the case of translation-invariant functions, the eigenvalue equation 3.29, can be solved analytically.

Mathematically, translation invariance implies that the functions $g(\mathbf{r}, \mathbf{r}')$ depend on the difference $\mathbf{r} - \mathbf{r}'$ only: $g(\mathbf{r}, \mathbf{r}') = \tilde{g}(\mathbf{r} - \mathbf{r}')$. In this case, the output signal depends on the power spectrum of the image only, because

$$g[x(\mathbf{r})] = \int \tilde{g}(\mathbf{r} - \mathbf{r}')x(\mathbf{r})x(\mathbf{r}')\, \mathrm{d}^2r\, \mathrm{d}^2r' \tag{4.5}$$

$$= \int \tilde{g}(\mathbf{k})|x(\mathbf{k})|^2\, \mathrm{d}^2k. \tag{4.6}$$

Here $x(\mathbf{k}) := \frac{1}{2\pi}\int x(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\, \mathrm{d}^2r$ and $\tilde{g}(\mathbf{k})$ denote the Fourier transforms of the image and the function $\tilde{g}(\mathbf{r})$, respectively. The value $|x(\mathbf{k})|^2$ of the power spectrum of an image x is calculated by summing the squares of the sin-Fourier transform and the cos-Fourier transform. Therefore, equation 4.6 implies that the function g is a weighted sum of quadrature filter pairs with filters that are plane waves. Note that quadrature filter pairs are the key element of the standard "energy" model of complex cells (Adelsen & Bergen, 1985).

Another implication of the translation invariance of g is that it is an eigenfunction of the generator of translations with the eigenvalue 0:

$$(\nabla_\mathbf{r} + \nabla_{\mathbf{r}'})g(\mathbf{r}, \mathbf{r}') = (\nabla_\mathbf{r} + \nabla_{\mathbf{r}'})\tilde{g}(\mathbf{r} - \mathbf{r}') = 0. \tag{4.7}$$

We can thus neglect the contribution of the translation generator in the eigenvalue equation 3.29.

In the simulations, the transformation velocities (i.e., the differences in position, angle, and zoom factor between successive frames) were chosen independently and from gaussian distributions centered at zero. The matrix $\langle v_\alpha v_\beta \rangle$ is then diagonal and contains the mean squares of the velocities on the diagonal. If we neglect terms arising from translation, the eigenvalue equation 3.29 then takes the form

$$-\left[\langle\omega^2\rangle(G^{\mathrm{rot}})^2 + \langle\zeta^2\rangle(G^{\mathrm{zoom}})^2\right]g_j = \Delta_j\, g_j. \tag{4.8}$$

Because the behavior of the functions g is easier to discuss in the Fourier representation, equation 4.6, it is convenient to solve the eigenvalue equation for the Fourier transform $\tilde{g}_j(\mathbf{k})$ directly. Transferring the eigenvalue equation into Fourier space requires a long, but not very illustrative, derivation. Essentially we have to insert the generators stated in Table 1 and the definition of the Fourier transform of $\tilde{g}$ into equation 4.8 and use the property of the Fourier transform that multiplications with $\tilde{\mathbf{r}}$ correspond to derivatives with respect to $\mathbf{k}$ in Fourier space and that derivatives with respect to $\tilde{\mathbf{r}}$ become multiplications with $\mathbf{k}$. For brevity, we skip the details and simply state the resulting eigenvalue equation:

$$-[\langle\omega^2\rangle(k_1\partial_{k_2} - k_2\partial_{k_1})^2 + \langle\zeta^2\rangle(\mathbf{k}\cdot\nabla_{\mathbf{k}} - 2)^2]\tilde{g}_j(\mathbf{k}) = \Delta_j\tilde{g}_j(\mathbf{k}). \tag{4.9}$$

It is easier to solve this equation in polar coordinates $(k, \phi) \in \mathbb{R}^+ \times [0, 2\pi]$, because then the operators for translation and rotation separate:

$$-\left[\langle\omega^2\rangle\,\partial_\phi^2 + \langle\zeta^2\rangle\,(k\partial_k - 2)^2\right]\tilde{g}_j(k, \phi) = \Delta_j\tilde{g}_j(k, \phi). \tag{4.10}$$

The eigenfunctions to this equation are given by

$$\tilde{g}_{q,m}(k, \phi) = A_{q,m}k^2 Q_q(k)M_m(\phi) \tag{4.11}$$

$$\text{with} \quad Q_q(k) = \begin{cases} \cos(q\ln k) & \text{for } q \geq 0 \\ \sin(q\ln k) & \text{for } q < 0 \end{cases} \tag{4.12}$$

$$\text{and} \quad M_m(\phi) = \begin{cases} \cos(m\phi) & \text{for } m \text{ even} \\ \sin((m+1)\phi) & \text{for } m \text{ odd} \end{cases}, \tag{4.13}$$

and with the associated eigenvalues

$$\Delta_{q,m} = \langle\zeta^2\rangle q^2 + \begin{cases} \langle\omega^2\rangle m^2 & \text{for } m \text{ even} \\ \langle\omega^2\rangle(m+1)^2 & \text{for } m \text{ odd} \end{cases}. \tag{4.14}$$

$A_{q,m}$ denotes a normalization constant that ensures that the unit variance constraint is fulfilled for the training data at hand and $q \in \mathbb{R}$ and $m \in \mathbb{N}^0$ are indices that label the solution. Notice that the oscillation in the angular direction contains only even frequencies (m for m even and $m+1$ for m odd), because $\tilde{g}(\tilde{\mathbf{r}})$ is real valued and symmetric, so its Fourier transform has to be symmetric: $\tilde{g}(k, \phi) = \tilde{g}(k, \phi + \pi)$. For each Δ-value, there are four solutions, corresponding to all possible combinations of sine and cosine in k and ϕ. Moreover, at least for large m, an increase of the rotation-dependent contribution to the Δ-value (increasing m) can be compensated by a decrease of the frequency-dependent contribution (decreasing q), which leads to additional degeneracies.

Notice that in addition to those given in equation 4.11, there are also solutions that have negative eigenvalues. These solutions have a frequency dependence that follows $\tilde{g} \approx k^2 e^{q\ln(k)} = k^{2+q}$ with $q \in \mathbb{R}$. They are qualitatively different from the solutions with positive eigenvalue. If an image is zoomed at constant velocity, the output signals of these functions show an exponential divergence. The output signals of the solutions with positive eigenvalues show harmonic oscillations (cf. section 5), and they remain bounded. Therefore, we exclude solutions with negative eigenvalues.

4.3 Optimal Stimuli. We define the optimal excitatory stimulus of a function $g[x(\mathbf{r})]$ as the image $S^+(\mathbf{r})$ that maximizes $g[x(\mathbf{r})]$ under the constraint of fixed total image power

$$\int S^+(\mathbf{r})^2 \, \mathrm{d}^2 r = \int |S^+(\mathbf{k})|^2 \, \mathrm{d}^2 k = \text{const.} \tag{4.15}$$

Similarly, the optimal inhibitory stimulus $S^-(\mathbf{r})$ is the image that minimizes $g[x(\mathbf{r})]$ with fixed power. According to equation 4.6, translation-invariant quadratic functionals are linear functionals of the power spectrum $|x(\mathbf{k})|^2$, so it is intuitively clear that the optimal excitatory/inhibitory stimulus concentrates all its power to those frequencies where $\tilde{g}_{q,m}(\mathbf{k})$ is maximal/minimal. This has several implications:

Plane wave optimal stimuli. The optimal excitatory and inhibitory stimuli are (possibly linear combinations of) plane waves $S^{\pm}(\mathbf{r}) = \cos(\mathbf{k} \cdot \mathbf{r} + \text{phase shift})$ with wave vectors $\mathbf{k}$ for which $\tilde{g}_{q,m}(\mathbf{k})$ is maximal or minimal. In practice, $\mathbf{k}$ is restricted to a finite domain, in particular because the finite resolution introduces a frequency cutoff. $\tilde{g}_{q,m}$ has at least one maximum within this domain. For large m, $\tilde{g}_{q,m}$ has many maxima of equal value, but in practice, one of these maxima is usually slightly higher, so that the optimal stimulus is a single plane wave. This agrees with the observations by Berkes and Wiskott (2005).

Phase invariance. The phase of the plane waves is arbitrary because the functions g depend on only the power spectrum of the images, not on its phase structure. This is in line with the notion of complex cells as being invariant with respect to the phase of their optimal stimulus and is also consistent with the results of Berkes and Wiskott (2005).

Frequency dependence. Since all functions $\tilde{g}_{q,m}(\mathbf{k})$ rise quadratically with the frequency $k = |\mathbf{k}|$, high spatial frequencies are favored. This may appear counterintuitive for a paradigm that is based on slowness, but due to the quadrature filter pair property of the receptive field, high spatial frequencies do not result in quickly varying output signals, because translation invariance is ensured. In experiments with real data, there is, of course, a frequency cutoff due to the finite resolution of the images, so that the optimal stimuli cannot have arbitrarily high frequencies. Berkes and Wiskott (2005) used principal component analysis (PCA) to reduce the input dimensionality from two pictures with 16×16 pixels each to a total of 100 dimensions (approximately 50 dimensions per image). It is known that PCA on natural images concentrates on low spatial frequencies while neglecting high frequencies. The highest spatial frequency that is possible after this preprocessing is then on the order of $\sqrt{50}/2 \approx 3.5$ cycles per side length of the

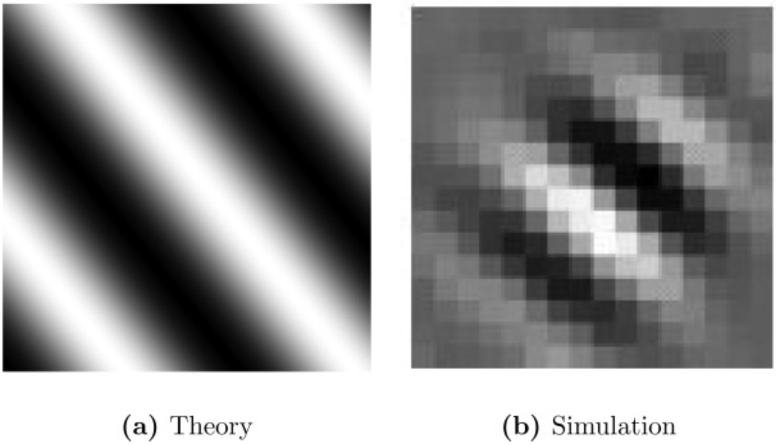

Figure 1: Optimal stimuli. (a) The theoretically predicted optimal stimuli are delocalized plane waves. (b) Typical optimal stimulus for the simulated SFA units in Berkes and Wiskott (2005). As theoretically predicted, the unit responds most strongly to a grating with a specific orientation. The observed decay of the simulated optimal stimulus toward the boundary of the image patch, however, is not captured by the theory.

image patch. This is a rather accurate estimate of the frequency of the optimal stimuli found by Berkes and Wiskott (2005).

Localization. Unlike in physiological findings, the optimal stimuli are not localized. Intuitively, this can be understood as follows. In the case of image sequences that are generated by continuous transformations (i.e., where image content stays within the vicinity of its original position for a certain time), spatial integration effectively acts as a low-pass filter, with spatial integration over larger areas corresponding to low-pass filtering with longer timescales. Because low-pass filtering with longer timescales generally leads to slower signals, optimal functions for SFA always try to integrate over the largest area possible—the full image. This is reflected by the delocalized optimal stimuli. Notice that SFA does not allow low-pass filtering, but requires the functions to process the input instantaneously. The low-pass filtering discussed here is purely spatial in nature but has the same effect as a temporal low-pass filter due to the spatiotemporal correlation structure of the input signals. Note also that the apparent localization of the simulated optimal stimuli (see Figure 1B) is not a real localization as found, for example, by Olshausen and Field (1996). The optimal functions decay toward the boundary of the image patch in order to reduce the abrupt influence of new image structure that enters the image patch. The optimal stimuli should thus vanish on the boundary as well. The optimal stimuli found by Berkes and Wiskott (2005) are as delocalized as this constraint allows them to be.

4.4 Orientation and Frequency Tuning. The typical approach for testing the orientation and the frequency tuning of a cortical cell is to plot its response to a grating as a function of the orientation and the frequency of the grating (see, e.g., De Valois, Yund, & Hepler, 1982). We represent the grating by a plane wave with frequency k_0 and orientation ϕ_0. As the power spectrum of this function is δ-shaped, the output of the function $g_{q,m}[x(\mathbf{r})]$ for a plane wave is given by $\tilde{g}_{q,m}(k_0, \phi_0)$.

Figure 2 shows a comparison of the orientation and frequency tuning of the analytical solutions and the simulations. They are in good agreement apart from a frequency cutoff in the simulations that arises from the finite resolution of the images and the preprocessing (see the discussion in section 4.3). The fact that the analytical solutions agree with the simulations indicates that the orientation and frequency tuning as observed in the simulations are an effect of the transformations used to generate the image sequences.

4.5 An Intuition for the Orientation and Frequency Tuning. The key to getting an intuitive understanding for why the optimal functions show the observed orientation and frequency tuning is the earlier result by Wiskott (2003) that the optimal output signals for SFA are harmonic oscillations. It is obvious that the output signal of the functions $\tilde{g}_{q,m}$ when applied to an image that rotates with constant velocity is sinusoidal. Similarly, the frequency dependence is such that the output signal is sinusoidal if the image is subjected to zoom with constant velocity. Remember that constant zoom velocity ζ implies that the zoom factor $z(t) = \exp(\zeta t)$ increases exponentially. As the image is zoomed by a factor z, the frequency decreases as $1/z$, so with an exponentially increasing zoom factor, the frequencies also decrease exponentially. In combination with the logarithmic dependence of $Q_q(k)$ on the frequency k, this yields a harmonic oscillation.

The reason for the quadratic rise of the oscillation amplitude of $\tilde{g}_{q,m}$ as a function of the frequency k is more subtle. When the image is zoomed by a factor z (i.e., $x(\mathbf{r}) \rightarrow x(\mathbf{r}/z)$), the total image power P increases by a factor z^2. This can be seen by means of a coordinate transformation $\mathbf{r}' = \mathbf{r}/z$:

$$\begin{aligned} \text{power zoomed} &= \int |x(\mathbf{r}/z)|^2 \, \mathrm{d}^2 r = \int |x(\mathbf{r}')|^2 z^2 \, \mathrm{d}^2 r' \\ &= z^2 \times \text{power unzoomed}. \end{aligned} \tag{4.16}$$

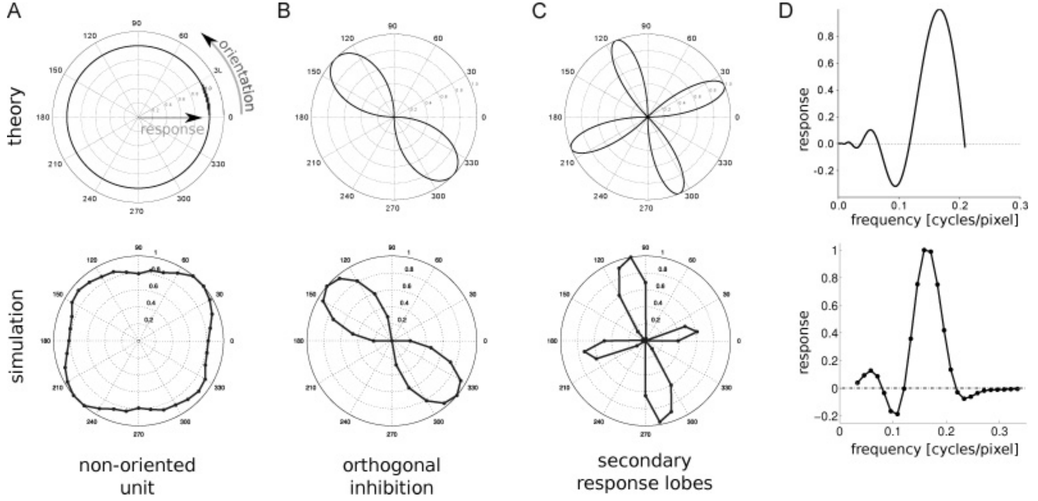

Figure 2: Orientation and frequency tuning. Comparison of analytical results (top row) and simulations (bottom row; reprinted with permission from (Berkes & Wiskott, 2005)). (A–C): Polar plot of the orientation-dependent component $M_m(\phi)$ of the analytical solutions $\tilde{g}_{q,m}$ for $m = 0, 2$, and 4, compared with the orientation tuning of three units from the simulations of Berkes and Wiskott (2005). The radius represents the output signal as a function of the orientation of a sinusoidal grating. Negative responses were truncated. Cells with similar orientation tuning were also observed in primary visual cortex of the macaque (De Valois et al., 1982; for a comparison of experimental and simulation results, see Berkes and Wiskott, 2005). The deviations of the simulation results from the theoretical predictions are due to random correlations in the input data that lead to a weak mixing of the theoretical solutions. For example, the simulated orientation tuning in A is a combination of the theoretical predictions in A ($m = 0$) and C ($m = 4$). The amplitude difference of the lobes of the simulation results in C can be explained by a mixture of the theoretical solutions in B ($m = 2$) and C ($m = 4$). (D): Frequency tuning. The theoretical curve in the upper row shows the frequency-dependent component $Q_q(k)$ of the analytical solutions $\tilde{g}_{q,m}$. To emphasize the structural similarity of the theoretical and the simulation results (bottom row), we adjusted the parameter q, the phase of the oscillation (which is arbitrary due to the degeneracy of the sine and cosine solutions), and the cutoff frequency to match the simulation results.

The additional factor k^2 counterbalances the increase in the output signal that would normally result from the increase in power, so that the amplitude of the harmonic oscillation remains constant.

4.6 Toward Side and End Inhibition. Some of the functions that Berkes and Wiskott (2005) learned showed higher-order properties of complex cells like side and end inhibition. Those units show a strong response to a grating presented to a subregion of the receptive field, which is gradually suppressed as the grating is shifted either along the bars of the grating (end inhibition) or perpendicular to the bars (side inhibition). Because these properties are inherently not translation invariant, they cannot be reproduced by the optimal functions we derived above. It is possible, however, to interpret them as the result of a weakly broken translation invariance.

To this end, let us first consider the classical quadrature filter pair model of a complex cell:

$$g(\mathbf{r}, \mathbf{r}') = \cos(\mathbf{k} \cdot \mathbf{r}) \cos(\mathbf{k} \cdot \mathbf{r}') + \sin(\mathbf{k} \cdot \mathbf{r}) \sin(\mathbf{k} \cdot \mathbf{r}') \tag{4.17}$$

$$= \cos(\mathbf{k}(\mathbf{r} - \mathbf{r}')). \tag{4.18}$$

Translation invariance is ensured by using the same vector $\mathbf{k}$ in both the dependence on $\mathbf{r}$ and $\mathbf{r}'$. One way of breaking the invariance is to use two different vectors $\mathbf{k}$ and $\mathbf{k}'$:

$$g_{\mathbf{k},\mathbf{k}'}(\mathbf{r}, \mathbf{r}') = \cos(\mathbf{k} \cdot \mathbf{r}) \cos(\mathbf{k}' \cdot \mathbf{r}') + \sin(\mathbf{k} \cdot \mathbf{r}) \sin(\mathbf{k}' \cdot \mathbf{r}') + \dots \tag{4.19}$$

$$= \cos(\mathbf{k} \cdot \mathbf{r} - \mathbf{k}' \cdot \mathbf{r}') + \dots, \tag{4.20}$$

where the ellipses stand for similar terms, just with $\mathbf{r}$ and $\mathbf{r}'$ swapped, to ensure that $g_{\mathbf{k},\mathbf{k}'}$ is symmetric in $\mathbf{r}$ and $\mathbf{r}'$.

If we ignore rotation and zoom and concentrate on (isotropic) translation, the operator $\mathcal{D}$ in the eigenvalue equation 3.29 is essentially the squared generator of translation. A brief calculation shows that the function $g_{\mathbf{k},\mathbf{k}'}$ is then an eigenfunction of $\mathcal{D}$, with an eigenvalue that is determined by the difference of $\mathbf{k}$ and $\mathbf{k}'$:

$$\mathcal{D} g_{\mathbf{k},\mathbf{k}'}(\mathbf{r}, \mathbf{r}') = -\langle v^2 \rangle (\nabla_{\mathbf{r}} + \nabla_{\mathbf{r}'})^2 g_{\mathbf{k},\mathbf{k}'}(\mathbf{r}, \mathbf{r}') \tag{4.21}$$

$$= \langle v^2 \rangle ||\mathbf{k} - \mathbf{k}'||^2 g_{\mathbf{k},\mathbf{k}'}(\mathbf{r}, \mathbf{r}') \tag{4.22}$$

$$= \Delta_{\mathbf{k},\mathbf{k}'} g_{\mathbf{k},\mathbf{k}'}(\mathbf{r}, \mathbf{r}'). \tag{4.23}$$

Therefore, the functions $g_{\mathbf{k},\mathbf{k}'}$ are optimal functions for SFA in the case of pure translation. Translation-invariant functions arise as a special case with $\mathbf{k} = \mathbf{k}'$. The Δ-value increases with the mismatch between $\mathbf{k}$ and $\mathbf{k}'$. Therefore, the slow functions found by SFA are those with $\mathbf{k} \approx \mathbf{k}'$.

To see that the functions $g_{\mathbf{k},\mathbf{k}'}$ can have properties similar to side and end inhibition, let us consider their output signal in response to an image $x(\mathbf{r} + \mathbf{R})$ as a function of the position $\mathbf{R}$ of the image:

$$g_{\mathbf{k},\mathbf{k}'}[x(\mathbf{r}+\mathbf{R})] = \iint g_{\mathbf{k},\mathbf{k}'}(\mathbf{r},\mathbf{r}')x(\mathbf{r}+\mathbf{R})x(\mathbf{r}'+\mathbf{R})\,\mathrm{d}^2r\,\mathrm{d}^2r' \tag{4.24}$$

$$= [x(\mathbf{k})\overline{x(\mathbf{k}')} + \overline{x(\mathbf{k})}x(\mathbf{k}')]\cos((\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}), \tag{4.25}$$

where $x(\mathbf{k})$ again denotes the Fourier transform of the image (for $\mathbf{R} = 0$) and $\overline{x(\mathbf{k})}$ denotes its complex conjugate.

For small differences $\mathbf{k} - \mathbf{k}'$, the function $g_{\mathbf{k},\mathbf{k}'}$ basically extracts the power of the image at the spatial frequency $\mathbf{k} \approx \mathbf{k}'$ and multiplies it with a slow sinusoidal dependence on the position $\mathbf{R}$. Note that this position dependence can again be understood as a way of generating harmonic oscillations under translations with constant speed. To generate a large output signal (again under an energy constraint for the image), we have to present an image that (1) contains most of its power at the spatial frequency $\mathbf{k}$ (and $\mathbf{k}' \approx \mathbf{k}$) and is (2) located at the right position, so that $\cos((\mathbf{k} - \mathbf{k}')\cdot\mathbf{R}) = 1$. The important point in the context of side and end inhibition is that a shift of such an image by $\Delta\mathbf{R} = \pi\,\frac{\mathbf{k}-\mathbf{k}'}{|\mathbf{k}-\mathbf{k}'|^2}$ leads to a large negative output signal, that is, an "inhibition."

We can now distinguish two special cases:

- **Side inhibition**. $\mathbf{k}$ and $\mathbf{k}'$ are parallel, with slightly different length. For a strong response, the image should contain a lot of power at $\mathbf{k}$; it should resemble a plane wave with wave vector $\mathbf{k}$. If this image is shifted in the direction $\mathbf{k} - \mathbf{k}'$, we get inhibition. Because $\mathbf{k} - \mathbf{k}'$ is parallel to the wave vector $\mathbf{k}$, this corresponds to shifting the bars of the grating "sideways," so the effect resembles a side inhibition.
- **End inhibition.** $\mathbf{k}$ and $\mathbf{k}'$ have the same length but slightly different directions. In this case, $\mathbf{k} - \mathbf{k}'$ is orthogonal to $\mathbf{k}$, so that shifts of the grating "along" the bars lead to inhibition, just as for end inhibition.

4.7 Toward Direction Selectivity. Some of the units in Berkes and Wiskott (2005) also show a selectivity to the direction of motion of the image. Because velocities cannot be calculated from the image alone, the authors used input signals that consisted of images at different moments in time: $x(\mathbf{r}, t)$ and $x(\mathbf{r}, t + \tau)$, where τ is a small time difference. The most general quadratic form in these two images can be written as

$$\begin{aligned} g[x(\mathbf{r},t),x(\mathbf{r},t+\tau)] &= g_1[x(\mathbf{r},t)] + g_2[x(\mathbf{r},t+\tau)] \\ &\quad + g_3[x(\mathbf{r},t),x(\mathbf{r},t+\tau)] \qquad (4.26) \\ &= \iint g_1(\mathbf{r},\mathbf{r}')x(\mathbf{r},t)x(\mathbf{r}',t)\,\mathrm{d}^2r\,\mathrm{d}^2r' \\ &\quad + \iint g_2(\mathbf{r},\mathbf{r}')x(\mathbf{r},t+\tau)x(\mathbf{r}',t+\tau)\,\mathrm{d}^2r\,\mathrm{d}^2r' \\ &\quad + \iint g_3(\mathbf{r},\mathbf{r}')x(\mathbf{r},t)x(\mathbf{r}',t+\tau)\,\mathrm{d}^2r\,\mathrm{d}^2r'. \qquad (4.27) \end{aligned}$$

The only term from which direction selectivity can arise is the third term g_3, because it compares the images at different moments in time. Let us for simplicity neglect the other terms. The term of interest is again a quadratic form in the images, so we expect that the generators for the transformations are essentially the same.[2] Thus, if we again concentrate on translation-invariant functions $g_3(\mathbf{r},\mathbf{r}') = \tilde{g}(\mathbf{r}-\mathbf{r}')$, we get the same expressions 4.11 for the frequency and orientation dependence of the optimal functions $\tilde{g}_3$. There is a subtle yet important difference to our previous calculations, though: because the arguments for the function g_3 are the images at two different moments in time, the kernel $g_3(\mathbf{r}-\mathbf{r}')$ need not be symmetric in $\mathbf{r}-\mathbf{r}'$, as was the case for the instantaneous functions. Consequently, additional solutions become available, which are antisymmetric in $\mathbf{r}-\mathbf{r}'$. These are the odd values for the oscillation frequency m in the orientation dependence,

$$M_m(\phi) = \begin{cases} i\cos(m\phi) & \text{for } m \text{ odd} \\ i\sin((m+1)\phi) & \text{for } m \text{ even} \end{cases}, \qquad (4.28)$$

where the imaginary factor i ensures that $\tilde{g}(\mathbf{r}-\mathbf{r}')$ is real valued.

To show that these solutions are direction selective, let us assume that the image moves at velocity $\mathbf{v}$, so that the image at time $t+\tau$ is given by the image at time t, shifted by $\mathbf{v}\tau$: $x(\mathbf{r},t+\tau) = x(\mathbf{r}+\mathbf{v}\tau,t)$. If τ is small, the

[2] A detailed analysis reveals that the generators contain correction terms that contain the accelerations of the transformations. This reflects the fact that direction selectivity in the presence of time-dependent translation velocity contributes to the speed of variation of the functions—the Δ-value. In a strict sense, our arguments are therefore true only for uniform transformation speed.

function g_3 can be approximated by a Taylor expansion:

$$\begin{aligned}
g_3[x(\mathbf{r},t), x(\mathbf{r},t+\tau)] &= \iint \tilde{g}_3(\mathbf{r}-\mathbf{r}')x(\mathbf{r},t)x(\mathbf{r}'+\mathbf{v}\tau,t)\,\mathrm{d}^2r\,\mathrm{d}^2r' \qquad (4.29)\\
&= \iint \tilde{g}_3(\mathbf{r}-\mathbf{r}'+\mathbf{v}\tau)x(\mathbf{r},t)x(\mathbf{r}',t)\,\mathrm{d}^2r\,\mathrm{d}^2r'\\
&\approx \iint \left(\tilde{g}_3(\mathbf{r}-\mathbf{r}') + \tau\mathbf{v}\cdot\nabla\tilde{g}_3(\mathbf{r}-\mathbf{r}')\right)\\
&\quad \times x(\mathbf{r},t)x(\mathbf{r}',t)\,\mathrm{d}^2r\,\mathrm{d}^2r'\\
&= \iint \tilde{g}_3(\mathbf{r}-\mathbf{r}')x(\mathbf{r},t)x(\mathbf{r}',t)\,\mathrm{d}^2r\,\mathrm{d}^2r' \qquad (4.30)\\
&\quad + \tau\mathbf{v}\cdot\iint \nabla\tilde{g}_3(\mathbf{r}-\mathbf{r}')x(\mathbf{r},t)x(\mathbf{r}',t)\,\mathrm{d}^2r\,\mathrm{d}^2r' \qquad (4.31)
\end{aligned}$$

The output signal can therefore be split into two terms, equations 4.30 and 4.31, the second of which depends linearly on velocity $\mathbf{v}$ and is therefore direction selective. Depending on the symmetry of $\tilde{g}_3$, we can now distinguish two cases. If $\tilde{g}_3$ is symmetric, the first term, equation 4.30, is nonzero, while the second term vanishes for symmetry reasons ($\nabla\tilde{g}_3$ is antisymmetric in $\mathbf{r}-\mathbf{r}'$, while $x(\mathbf{r})x(\mathbf{r}')$ is symmetric). Symmetric functions $\tilde{g}_3$ therefore show no direction selectivity. If $\tilde{g}_3$ is antisymmetric, the first term vanishes for symmetry reasons, while the second term is nonzero, because the derivative $\nabla\tilde{g}_3$ is symmetric in $\mathbf{r}-\mathbf{r}'$. Therefore, antisymmetric functions $\tilde{g}_3$ are direction selective: they change their sign when the image moves in the opposite direction.

This analysis also reveals an interaction between the orientation tuning and the direction selectivity of the cell: the number m that specifies the frequency of the orientation dependence also specifies the orientation dependence of the gradient and therefore the direction selectivity of the unit.

Finally, the spatiotemporal optimal stimuli for direction-selective units can be calculated analytically. To this end, we use a Lagrange multiplier approach to maximize $g_3[x_1(\mathbf{r}), x_2(\mathbf{r})]$ under an energy constraint for its two arguments $x_{1/2}(\mathbf{r})$. The Lagrange function is given by

$$\begin{aligned}
\mathcal{L} &= g_3[x_1(\mathbf{r}), x_2(\mathbf{r})] - \sum_i \lambda_i \int x_i(\mathbf{r})^2\,\mathrm{d}^2r \qquad (4.32)\\
&= \int \tilde{g}_3(\mathbf{k})\overline{x_1(\mathbf{k})}x_2(\mathbf{k})\,\mathrm{d}^2k - \sum_i \lambda_i \int |x_i(\mathbf{k})|^2\,\mathrm{d}^2k, \qquad (4.33)
\end{aligned}$$

where we changed to the Fourier representation for convenience. Taking the variational derivative with respect to x_1 and x_2 yields two necessary conditions for the images:

$$\tilde{g}_3(\mathbf{k})x_2(\mathbf{k}) = \lambda_1 x_1(\mathbf{k}) \tag{4.34}$$

$$\overline{\tilde{g}_3(\mathbf{k})}x_1(\mathbf{k}) = \lambda_2 x_2(\mathbf{k}), \tag{4.35}$$

which can be combined to:

$$|\tilde{g}_3(\mathbf{k})|^2 x_i(\mathbf{k}) = \lambda_1 \lambda_2 x_i(\mathbf{k}). \tag{4.36}$$

These equations can be fulfilled only if the optimal stimuli $x_i(\mathbf{r})$ are superpositions of plane waves with spatial frequencies that fulfill $|\tilde{g}_3(\mathbf{k})|^2 = \lambda_1 \lambda_2$. Moreover, since $\tilde{g}_3(\mathbf{r})$ is antisymmetric for direction-selective units, $\tilde{g}_3(\mathbf{k})$ is purely imaginary. A multiplication with an imaginary number in Fourier space corresponds to a phase shift of $\pi/2$ in real space. Equation 4.34 therefore shows that x_1 and x_2 are plane waves with the same frequency and a phase shift of $\pi/2$. This is exactly what Berkes and Wiskott (2005) observed.

5 Discussion

We have presented a mathematical framework for SFA for the case that the input data are generated by applying a set of continuous transformations to a set of static templates. The theory is based on a group-theoretical approach and culminates in an eigenvalue problem for an operator that contains the generators of the transformation group.

Applying the framework to the simulation setup in Berkes and Wiskott (2005), we have shown that the eigenvalue equation becomes a partial differential equation that can be solved analytically for the special case of translation invariant receptive fields. The assumption of translation invariance implies that the optimal functions are invariant to phase shifts, similar to the simulated receptive fields in Berkes and Wiskott (2005). The orientation and frequency tuning of the analytical solutions are in good agreement with the simulation results. Moreover, the optimal stimuli of the analytical solutions are plane waves, similar to the gratings that were found in the simulations and are also common in physiological studies of cells in V1.

Under the assumption that the statistics of the input data are invariant with respect to the transformations used for their generation, the equations that determine the optimal functions are independent of the input statistics. Instead, they depend solely on the transformations as reflected by their group-theoretical generators. This purely mathematical statement agrees with control experiments performed by Berkes and Wiskott (2005), which showed that the simulation results were qualitatively the same when using colored noise instead of natural images. Which transformations were used, however, had a drastic influence on the structure of the receptive fields. For example, a lack of translation abolished the grating structure of the optimal stimuli. This is in agreement with the theory, because the optimal stimuli were plane waves only because the functions were assumed to be translation invariant. The assumption of translation invariance, however, is valid only when translation is the dominant transformation in the image sequences, so that any dependence on position would yield quickly varying output signals and would thus be unfavorable for the slowness objective.

5.1 The Harmonic Oscillation Argument. The theory shows that each of the properties of the optimal functions can be understood as an effect of one particular transformation: Translation leads to optimal stimuli that are plane waves, rotation causes a sinusoidal dependence of the output on the orientation, and zoom is responsible for the frequency tuning of the cells. Intuitively, both the orientation and the frequency tuning can be understood as a way of generating harmonically oscillating output signals when the associated transformation is applied with constant velocity. This interpretation is in line with earlier results indicating that the optimal output signals for SFA are harmonic oscillations (Wiskott, 2003).

Moreover, the "harmonic oscillation argument" suggests that for the given stimulus paradigm, the orientation and frequency tuning of the learned functions are not only optimal in the space of quadratic functions, but rather optimal in general. Therefore, we expect that an increase in the complexity of the function space will not lead to changes in orientation and frequency tuning. However, more complex function spaces may contain a degenerate set of translation-invariant functions with the same orientation and frequency tuning. These functions will generate output signals with the same Δ-value on the given training data. In that case, the optimal function set is no longer uniquely determined by the transformations in the training data, because an arbitrary mixture of these degenerate solutions is also valid. The resulting additional degrees of freedom can be used either to encode the identity of the templates $\mathbf{x}^{\mu}$ (depending on the application, this is either dangerously close to overfitting or useful for object recognition) or learn additional transformations in the images. This insight may be particularly relevant for hierarchical SFA networks (Franzius, Sprekeler, & Wiskott, 2007). If the complexity of the input statistics is not sufficiently high, higher layers in the hierarchy will essentially reproduce the output signals on the lower layers, which have already achieved the optimal harmonic oscillation response.

5.2 Localized Receptive Fields. One property of complex cells in visual cortex is not captured by the simulations or the theory: receptive fields of cells in primary visual cortex are localized. According to the discussion in section 4.3, however, this cannot be expected from the slowness principle alone, because—at least in the presence of large-scale transformations like global translation—larger receptive fields allow slower responses, so that localization is not favorable from the perspective of the slowness principle.

In this context, the question arises if localized receptive fields are learned by SFA if the image dynamics are dominated by local transformations. The following argument shows that this is not necessarily the case, at least if the image statistics are translation invariant. Consider two regions in a natural image sequence, located at a sufficiently large distance from another so that we can treat the associated image patches as statistically independent. The image sequences in each of these regions contain a set of slow features with associated Δ-values. The assumption of translation invariance implies that the slow features and their Δ-values should be the same for both regions. Consequently, the solutions for SFA are not unique: arbitrary (orthogonal) linear combinations of slow features with the same Δ-value are also valid solutions for SFA. We therefore expect that SFA learns arbitrary linear combinations of local features at different locations, which will not necessarily be localized. Unfortunately, this shows that for SFA, purely local image statistics are not a guarantee for local receptive fields. The underlying reason is that decorrelation is a rather crude approximation of statistical independence. Localized receptive fields probably require a stronger constraint like statistical independence or additional objectives like sparseness. An (efficiently solvable) extension of SFA in this direction would be desirable, in particular because statistical independence and sparseness have been proposed as principles for the unsupervised learning of localized receptive fields of simple cells in V1 (Bell & Sejnowski, 1995; Olshausen & Field, 1996, 1997).

The optimal stimuli found in the simulation seem to possess at least some kind of localization, since they decay toward the borders of the images patch. A similar decay of the optimal functions toward the boundaries was also observed in an earlier one-dimensional model of visual processing (Wiskott & Sejnowski, 2002). These results suggest that for the case where the input images are not infinitely large, the differential equation for the optimal functions has to be complemented by a boundary condition that requires the kernel of the optimal functionals to vanish on the boundary. Such a boundary condition would weaken the effect of a new image structure that enters the receptive field at its border and thus ensures a smoothly varying output signal. Unfortunately, so far we have not managed to find a mathematical proof for this boundary condition.

5.3 Future Work. A future direction would be to study the properties of the optimal solutions in other function spaces. An extension of the calculations to polynomials of arbitrary order is straightforward and might allow predictions for more complicated response properties of cells in (possibly higher-order) visual cortices. As already discussed, we expect that richer function spaces require a richer set of image transformations to resolve possible degeneracies.

A question that cannot be answered within the mathematical framework presented here is what happens if the statistics of the input is not invariant with respect to the transformations at hand. Would the optimal functions for SFA show a different orientation tuning if the orientation dependence of natural image statistics were taken into account, for example, by using natural videos as training data? Slowness-based learning of complex cells from natural videos has been done (Körding, Kayser, Einhäuser, & König, 2004) but to our knowledge not been systematically analyzed from this perspective. Experimentally it has been shown for cats that an extreme dependence of image statistics on orientation during rearing has a strong impact on the orientation tuning properties of cells in V1 (Hirsch & Spinelli, 1971). More research is necessary to assess if these influences can be explained in terms of slowness learning.

From the theoretical perspective, it would be interesting if there is a "unified theory" for SFA that captures both the finite-dimensional case with an unrestricted function space (Franzius et al., 2007; Sprekeler, Zito, & Wiskott, 2010) and the case considered here. Such a theory could describe the effects of input statistics that are not invariant with respect to the transformations but still capture the restrictions on the function space.

In the light of the introductory discussion, one could argue that any learning rule that aims at explaining the response properties of cells in V1 should, given the maturity of these properties shortly after birth, be able to establish the same receptive field structure from natural images and retinal waves. In this line, it was recently shown that sparse coding and independent component analysis on a model of retinal waves lead to Gabor-shaped simple cell receptive fields (Albert, Schnabel, & Field, 2008). Since propagating waves could be interpreted as a "prenatal imitation" of translation in visual scenes, it is likely that slowness learning on these patterns can lead to translation-invariant units with similar response properties as complex cells. Preliminary results indicate that this is indeed the case (Dähne, Wilbert, & Wiskott, 2009).

5.4 Experimental Predictions. According to the theory, the orientation and frequency tuning of visual neurons should depend on the relative velocities of rotation and zoom in the visual stimuli. This leads to the following experimental prediction. Assume that an animal is confronted with rotating and zooming Gabor wavelets at two different retinal locations. At one retinal location, rotation is faster than zoom, and at the other one, zoom dominates over rotation. It should be straightforward to design the stimuli such that the ensembles of images at the two locations are identical and differ only in their dynamics. If visual neurons adjust to the stimuli to generate more slowly varying output signals, one should observe differential alterations in the degree of orientation and frequency tuning of V1 neurons at the two retinal locations—sharper orientation and wider frequency tuning for slow rotation and fast zoom and vice versa.

An open question, to which we have no clear prediction is, at what age should this experiment be done? Do we interpret slowness in learning as a developmental principle that acts prenatally or during critical periods or as a principle that is also valid in adults? The proposition that complex cells could be established on the basis of retinal waves is based on the former assumption, while the experiment decribed above could be conducted in both young animals (e.g., during critical periods) and in adulthood.

Currently experimental evidence if, in which areas, and at which developmental stages slowness is an appropriate principle for describing sensory learning is still scarce. Findings in higher visual areas of adult monkeys suggest that position-invariant object representations are subject to a dynamic learning process in agreement with the slowness principle and ongoing in adulthood (Miyashita, 1988; Li & DiCarlo, 2008). The proposed experiment could reveal if this finding generalizes to earlier visual areas.

Appendix A: Derivation of the Generators

A.1 Translation. We use the convention that the effect of a translation T_x of an image $x(\mathbf{r})$ by a vector $\mathbf{R}$ is the replacement of the pixel value at position $\mathbf{r}$ by the pixel value of the original image at the position $\mathbf{r} - \mathbf{R}$:

$$[T_x x](\mathbf{r}) = x(\mathbf{r} - \mathbf{R}). \tag{A.1}$$

What is the corresponding representation of translation on the quadratic functions, equation 4.1? This can be seen immediately by means of a variable substitution:

$$\begin{aligned}
\left[T_g g\right][x(\mathbf{r})] &\overset{(3.2)}{:=} \int \tilde{g}(\mathbf{r}, \mathbf{r}')\,[T_x x](\mathbf{r})\,[T_x x](\mathbf{r}')\ \mathrm{d}^2 r\ \mathrm{d}^2 r' && \text{(A.2)}\\
&\overset{(A.1)}{=} \int \tilde{g}(\mathbf{r}, \mathbf{r}') x(\mathbf{r} - \mathbf{R}) x(\mathbf{r}' - \mathbf{R})\ \mathrm{d}^2 r\ \mathrm{d}^2 r' && \text{(A.3)}\\
&= \int \tilde{g}(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R}) x(\mathbf{r}) x(\mathbf{r}')\ \mathrm{d}^2 r\ \mathrm{d}^2 r'. && \text{(A.4)}
\end{aligned}$$

Thus, the effect of the translation operator on the functional g is the replacement of the kernel $\tilde{g}(\mathbf{r}, \mathbf{r}')$ by $\tilde{g}(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R})$:

$$\left[T_g \tilde{g}\right](\mathbf{r}, \mathbf{r}') = \tilde{g}(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R}). \tag{A.5}$$

Remember that we represent the functionals in terms of the basis functions $x(\mathbf{r})x(\mathbf{r}')$. In this basis, the functional g is represented by the "coefficient function" $\tilde{g}(\mathbf{r}, \mathbf{r}')$. Equation A.5 is the representation of the translation operator in this basis.

We can now calculate the associated generator by applying a time-dependent translation $T_g(t)$ by a vector $\mathbf{R}(t)$ and calculating the temporal derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[T_g(t)\tilde{g}\right](\mathbf{r}, \mathbf{r}') \overset{(A.5)}{=} \frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}(\mathbf{r} + \mathbf{R}(t), \mathbf{r}' + \mathbf{R}(t)) \tag{A.6}$$

$$= \left[\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}(t) \cdot [\nabla_{\mathbf{r}} + \nabla_{\mathbf{r}'}]\,\tilde{g}\right](\mathbf{r} + \mathbf{R}(t), \mathbf{r}' + \mathbf{R}(t)) \tag{A.7}$$

$$= \left[T_g(t)\, Q^{\text{trans}}(t)\tilde{g}\right](\mathbf{r}, \mathbf{r}') \tag{A.8}$$

with

$$Q^{\text{trans}}(t) := \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}(t) \cdot [\nabla_{\mathbf{r}} + \nabla_{\mathbf{r}'}]. \tag{A.9}$$

Clearly, the translation velocity $\mathbf{v} := \mathrm{d}\mathbf{R}/\,\mathrm{d}t$ plays the role of the velocity in equation 3.14, while the sum of the gradients is the generator of translations as stated in Table 1.

A.2 Rotation. A rotation of an image $x(\mathbf{r})$ by an angle ϕ corresponds to the application of an orthogonal matrix $\mathbf{O}^{-1} = \mathbf{O}^T$ to the pixel positions:

$$[T_x x](\mathbf{r}) = x(\mathbf{O}^T\mathbf{r}), \tag{A.10}$$

where

$$\mathbf{O} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ \sin(-\phi) & \cos(\phi) \end{pmatrix}. \tag{A.11}$$

The effect of the related rotation operator T_g on the integral kernel $g(\mathbf{r}, \mathbf{r}')$ can again be derived by a variable substitution:

$$\left[T_g g\right][x(\mathbf{r})] \overset{(3.2)}{:=} \int \tilde{g}(\mathbf{r}, \mathbf{r}')\,[T_x x](\mathbf{r})\,[T_x x](\mathbf{r}')\;\mathrm{d}^2r\;\mathrm{d}^2r' \tag{A.12}$$

$$\overset{(A.10)}{=} \int \tilde{g}(\mathbf{r}, \mathbf{r}')x(\mathbf{O}^T\mathbf{r})x(\mathbf{O}^T\mathbf{r}')\;\mathrm{d}^2r\;\mathrm{d}^2r' \tag{A.13}$$

$$= \int \tilde{g}(\mathbf{O}\mathbf{r}, \mathbf{O}\mathbf{r}')x(\mathbf{r})x(\mathbf{r}')\;\mathrm{d}^2r\;\mathrm{d}^2r'. \tag{A.14}$$

Thus, in the basis $x(\mathbf{r})x(\mathbf{r})'$, rotations are represented by

$$[T_g\tilde{g}](\mathbf{r},\mathbf{r}') = \tilde{g}(\mathbf{Or},\mathbf{Or}'). \tag{A.15}$$

Again, we can calculate the generator by taking the temporal derivative of a time-dependent rotation $T_g(t)$ by a matrix $\mathbf{O}(t)$. To keep the notation short, we omit the time dependence of the rotation matrix $\mathbf{O}$ and use the short notation $\dot{\mathbf{O}} := \frac{\mathrm{do}}{\mathrm{d}t}$ for its temporal derivative:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left[T_g(t)\tilde{g}\right](\mathbf{r},\mathbf{r}') &\overset{(A.13)}{=} \frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}(\mathbf{Or},\mathbf{Or}') && \text{(A.16)}\\
&= \left[((\dot{\mathbf{O}}\mathbf{r})\cdot\nabla_{\mathbf{r}} + (\dot{\mathbf{O}}\mathbf{r}')\cdot\nabla_{\mathbf{r}'})\tilde{g}\right](\mathbf{Or},\mathbf{Or}') && \text{(A.17)}\\
&= \Big[T_g(t)\underbrace{((\dot{\mathbf{O}}\mathbf{O}^T\mathbf{r})\cdot\nabla_{\mathbf{r}} + (\dot{\mathbf{O}}\mathbf{O}^T\mathbf{r}')\cdot\nabla_{\mathbf{r}'})}_{=:Q^{\mathrm{rot}}(t)}\tilde{g}\Big](\mathbf{r},\mathbf{r}') && \text{(A.18)}\\
&= \left[T_g(t)\,Q^{\mathrm{rot}}(t)\tilde{g}\right](\mathbf{r},\mathbf{r}'). && \text{(A.19)}
\end{aligned}$$

The matrix $\dot{\mathbf{O}}\mathbf{O}^T$ is antisymmetric, because

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{I} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{O}\mathbf{O}^T) = \dot{\mathbf{O}}\mathbf{O}^T + \mathbf{O}\dot{\mathbf{O}}^T = \dot{\mathbf{O}}\mathbf{O}^T + (\dot{\mathbf{O}}\mathbf{O}^T)^T. \tag{A.20}$$

Here, $\mathbf{I}$ denotes the unit matrix. Because of the antisymmetry, $\dot{\mathbf{O}}\mathbf{O}^T$ can be written as

$$\dot{\mathbf{O}}\mathbf{O}^T = \begin{pmatrix} 0 & -\omega(t) \\ \omega(t) & 0 \end{pmatrix}. \tag{A.21}$$

It can be shown that $\omega(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t}$ is the angular velocity of the rotation. In this notation, $Q^{\mathrm{rot}}(t)$ becomes

$$Q^{\mathrm{rot}}(t) = \omega(t)\left[r_1\partial_{r_2} - r_2\partial_{r_1} + r_1'\partial_{r_2'} - r_2'\partial_{r_1'}\right], \tag{A.22}$$

which leaves us with the generator and the associated velocity ω given in Table 1.

A.3 Zoom. Zooming an image by a zoom factor z around the origin corresponds to replacing the pixel value at position $\mathbf{r}$ by the pixel value of the original image at position $\mathbf{r}/z$. Using similar considerations as above, this leads to the following representation of the zoom operator:

$$\left[T_g\tilde{g}\right](\mathbf{r},\mathbf{r}') = z^4\tilde{g}(z\mathbf{r},z\mathbf{r}'). \tag{A.23}$$

The factor z^4 is the Jacobian determinant that arises from the coordinate changes $\mathbf{r} \to \mathbf{r}/z$ and $\mathbf{r}' \to \mathbf{r}'/z$ in the integration for $g[x(\mathbf{r})]$.

The generator can again be calculated by introducing a time-dependent zoom factor $z(t)$ and taking the temporal derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[T_g(t)\tilde{g}\right](\mathbf{r}, \mathbf{r}') = \frac{\mathrm{d}}{\mathrm{d}t} z^4 \tilde{g}(z\mathbf{r}, z\mathbf{r}') \tag{A.24}$$

$$= \left[\left(z^4\left[\dot{z}\mathbf{r} \cdot \nabla_{\mathbf{r}} + \dot{z}\mathbf{r}' \cdot \nabla_{\mathbf{r}'}\right] + 4z^3\dot{z}\right)\tilde{g}\right](z\mathbf{r}, z\mathbf{r}') \tag{A.25}$$

$$= z^4\frac{\dot{z}}{z}\left[\left((z\mathbf{r}) \cdot \nabla_{\mathbf{r}} + (z\mathbf{r}') \cdot \nabla_{\mathbf{r}'} + 4\right)\tilde{g}\right](z\mathbf{r}, z\mathbf{r}') \tag{A.26}$$

$$= \left[T_g(t)\, Q^{\mathrm{zoom}}(t)\tilde{g}\right](\mathbf{r}, \mathbf{r}'), \tag{A.27}$$

with an operator $Q^{\mathrm{zoom}}(t)$ that contains the generator and the velocity $\zeta := \frac{\dot{z}}{z}$ for zoom as given in Table 1:

$$Q^{\mathrm{zoom}}(t) = \frac{\dot{z}}{z}[\mathbf{r} \cdot \nabla_{\mathbf{r}} + \mathbf{r}' \cdot \nabla_{\mathbf{r}'} + 4]. \tag{A.28}$$

Acknowledgments

We thank the anonymous reviewers for their constructive comments. This work was supported by the Volkswagen Foundation through a junior research group to L.W.

References

Adelsen, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal Optical Society of America A, 2*(2), 284–299.

Albert, M. V., Schnabel, A., and Field, D. J. (2008). Innate visual learning through spontaneous activity patterns. *PLoS Computational Biology, 4*(8), e1000137.

Bell, A. J., & Sejnowski, T. J. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision, 5*(6), 579–602.

Courant, R., & Hilbert, D. (1989). *Methods of mathematical physics Part I.* Hoboken, NJ: Wiley.

Dähne, S., Wilbert, N., & Wiskott, L. (2009). Learning complex cell units from simulated prenatal retinal waves with slow feature analysis. In *Eighteenth Annual Computational Neuroscience Meeting CNS*2009* (Vol. 10, p. 129). London: BioMed Central.

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22*, 531–544.

Dong, D. W. (2001). Spatiotemporal inseparability of natural images and visual sensitivities. In J. M. Zanker & J. Zeil (Eds.), *Computational, neural and ecological constraints of visual motion processing*. New York: Springer.

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems, 6*(3), 345–358.

Franzius, M., Sprekeler, H., & Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Computationl Biology, 3*(8), e166.

Hirsch, H., & Spinelli, D. (1971). Modification of the distribution of receptive field orientation in cats by selective visual exposure during development. *Experimental Brain Research, 12*(5), 509–527.

Hubel, D., & Wiesel, T. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology, 26*(6), 994–1002.

Körding, K., Kayser, C., Einhäuser, W., & König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli? *Journal of Neurophysiology, 91*(1), 206–212.

Landau, L. D., & Lifshitz, E. M. (1977). *Quantum mechanics: Non-relativistic theory.* London: Pergamon Press.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321*(5895), 1502–1507.

McLaughlin, T., & O'Leary, D. (2005). Molecular gradients and development of retinotopic maps. *Annual Reviews of Neuroscience*, 28, 327–355.

Miao, X., & Rao, R. P. N. (2007). Learning the Lie groups of visual invariance. *Neural Computation, 19*(10), 2665–2693.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature, 335*(6193), 817–820.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*, 3311–3325.

Rao, R., & Ruderman, D. (1998). Learning Lie groups for invariant visual perception. In M. S. Kears, S. A. Sola, & D. N. Cohn (Eds.), *Neural information processing systems, 11* (pp. 810–816). Cambridge, MA: MIT Press.

Ringach, D. (2007). On the origin of the functional architecture of the cortex. *PLoS ONE, 2*(2), e251.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*(6), 814–817.

Sprekeler, H., Zito, T., & Wiskott, L. (2010). An extension of slow feature analysis for nonlinear blind source separation. Cognitive Sciences EPrint Archive (CogPrints), 7056.

Wiskott, L. (2003). Slow feature analysis: A theoretical analysis of optimal free responses. *Neural Computation, 15*(9), 2147–2177.

Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Yao, H., & Dan, Y. (2001). Stimulus timing-dependent plasticity in cortical processing of orientation. *Neuron, 32*(2), 315–323.

Received January 26, 2009; accepted March 17, 2010.