Vision Research 50 (2010) 2200–2212

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Encoding natural scenes with neural circuits with random thresholds

Aurel A. Lazar *, Eftychios A. Pnevmatikakis, Yiyin Zhou

Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 16 October 2009
Received in revised form 20 March 2010

Keywords:
Neural encoding of natural scenes
Receptive fields
Neural circuits with random thresholds
Reconstruction of visual stimuli
Reproducing Kernel Hilbert Spaces
Regularization

ABSTRACT

We present a general framework for the reconstruction of natural video scenes encoded with a population of spiking neural circuits with random thresholds. The natural scenes are modeled as space-time functions that belong to a space of trigonometric polynomials. The visual encoding system consists of a bank of filters, modeling the visual receptive fields, in cascade with a population of neural circuits, modeling encoding in the early visual system. The neuron models considered include integrate-and-fire neurons and ON–OFF neuron pairs with threshold-and-fire spiking mechanisms. All thresholds are assumed to be random. We demonstrate that neural spiking is akin to taking noisy measurements on the stimulus both for time-varying and space-time-varying stimuli. We formulate the reconstruction problem as the minimization of a suitable cost functional in a finite-dimensional vector space and provide an explicit algorithm for stimulus recovery. We also present a general solution using the theory of smoothing splines in Reproducing Kernel Hilbert Spaces. We provide examples of both synthetic video as well as for natural scenes and demonstrate that the quality of the reconstruction degrades gracefully as the threshold variability of the neurons increases.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

In the recent years the increasing availability of multi-electrode recordings and functional imaging methods has led to the application of neural decoding techniques to the recovery of complex stimuli such as natural video scenes. An algorithm based on the optimal linear decoder derived in Warland, Reinagel, and Meister (1997) for a rate model was presented in Stanley, Li, and Dan (1999) for the reconstruction of natural video scenes with recognizable moving objects from recordings of a neural population of the cat's lateral geniculate nucleus (LGN). Visual image reconstruction from fMRI data was examined in Miyawaki et al. (2008), whereas in Kay, Naselaris, Prenger, and Gallant (2008) fMRI data was used to identify natural images. The above works suggest that the visual information is preserved along the different layers of the visual system and call for the development of novel algorithms for neural decoding algorithms that are based on spike times.

In this paper we present a formal mathematical, model based approach, for coding and reconstruction in the early visual system. Our neural architecture consists of a population of N spatial filters that model the classical receptive fields, in cascade with an equal number of spiking neural circuits. The neural circuits considered are either integrate-and-fire neurons or ON–OFF neuron pairs with thresholding and feedback. In our architecture the neuronal variability is not attributed to a probabilistic code (Ma, Beck, Latham, & Pouget, 2006); rather the neural circuits are assumed to have random thresholds with known a priori distribution. Neurons with random thresholds have been used to model the observed spike variability of biological neurons of the fly visual system (Gestri, Mastebroek, & Zaagman, 1980), as well as neurons in the early visual system of the cat (Reich, Victor, Knight, Ozaki, & Kaplan, 1997).

We show that neural spiking with these neural circuits represents noisy and independent (Knight, 1972) (generalized) measurements of the input visual stimulus. Based on these measurements, we construct regularized cost functionals and identify the reconstructed stimulus as its minimizer. For simplicity, we assume that the input visual space belongs to a finite-dimensional Hilbert space and use standard optimization techniques to find the reconstructed stimulus. However, as it will be discussed, the results can be directly extended to infinite-dimensional spaces, using the theory of smoothing splines (Wahba, 1990) in Reproducing Kernel Hilbert Spaces (Berlinet & Thomas-Agnan, 2004).

The work presented here builds and extends upon previous work on the representation of stimuli with *deterministic* spiking neurons. Assuming that the input signal is bandlimited and the bandwidth is known, a perfect recovery of the stimulus based upon the spike times can be achieved provided that the spike density is above the Nyquist rate of the stimulus. These results hold for a wide variety of sensory stimuli, including audio (Lazar & Pnevmatikakis, 2008b) and video streams (Lazar & Pnevmatikakis, 2008a; Lazar & Pnevmatikakis, submitted for publication) encoded with a population of spiking neurons. The model of stimuli considered

* Corresponding author. Fax: +1 212 932 9421.
E-mail address: aurel@ee.columbia.edu (A.A. Lazar).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.03.015

in this paper are defined on a discretized version of a band-limited signal space, known as the space of *trigonometric polynomials*. Such spaces are suitable for modeling since they have all the desirable properties of band-limited signal spaces with the added benefit of being finite-dimensional and thus numerically tractable (Lazar, Simonyi, & Tóth, 2008). Moreover, as it will be demonstrated, the finite-dimensionality of the space determines to a first order the complexity of the reconstruction algorithm. Consequently, data recorded from additional neurons can be included into the recovery algorithm at a very moderate computational cost.

Since the encoding neural circuits have random thresholds, a perfect recovery of the input stimulus is not possible. In order to derive an optimal recovery algorithm, we setup the stimulus recovery as a regularized optimization problem. Signal representation using regularization techniques has been discussed in the computational vision (Poggio, Torre, & Koch, 1985) and neural networks (Girosi, Jones, & Poggio, 1995) literature. In this paper we present a formal model for stimulus reconstruction from spike timing using a method of regularization that, as we will show, can approximate complex visual streams, such as natural scenes, in a very efficient way. Using regularization to reconstruct signals encoded with neurons with random thresholds was first presented in Lazar and Pnevmatikakis (2009) in the context of time-varying stimuli belonging to Sobolev spaces encoded with a population of leaky integrate-and-fire neurons.

We explore the recovery of natural scenes and synthetic video streams as a function of the variability of the random thresholds. Variability is quantified as the ratio between the variance and the mean of the threshold. We also explore the modeling of natural scenes with the sample functions that are defined in the space of trigonometric functions. Finally, we present for the first time video sequences of visual stimuli encoded with neural circuit architectures based on neurons with random thresholds. We evaluate the recovery using both traditional measures of signal-to-noise ratio (SNR) as well structural similarity index (SSIM) (Wang, Bovik, Sheikh, & Simoncelli, 2004). The latter more closely relates to perceptual quality of visual stimuli. Rather than focusing on modeling a specific region of the early visual system, we show that the methodology presented here is general and can be applied to arbitrary combinations of receptive fields and neural spiking mechanisms. These include classic models of the early visual pathway (retina, LGN and V1).

The paper is organized as follows. Section 2 deals with the problem of encoding and reconstruction of time-varying stimuli. In Section 2.1 we give a short overview of the spaces of trigonometric functions and discuss how these constitute a natural discretization of spaces of bandlimited functions. In Section 2.2 we present how time-varying stimuli can be encoded with ON–OFF neuron pairs with random thresholds and present their reconstruction by finding the minimizer of an appropriate quadratic cost functional. In Section 2.3 integrate-and-fire neurons with random thresholds encode time-varying stimuli; their recovery is presented in the same section. Examples are given in Section 2.4 that explore the quality of the reconstruction as a function of threshold variability. In Section 3 we introduce the full model for video encoding and reconstruction with a population of spiking neurons with random thresholds. We discuss how video streams can be modeled as space-time trigonometric polynomials and discuss their representation and reconstruction based on this working assumption. Section 4 presents examples of both synthetic and natural video scenes, encoded with neural circuits build with classic models of receptive fields and spiking neurons arising in the retina, LGN and V1. The examples demonstrate the effectiveness of our algorithm by measuring various different quality metrics (Peak SNR, and SSIM) for two different choices of random threshold (Gaussian, Gamma). Actual videos can be found in the Supplementary material. Section 5 discusses various extensions of our work to the recovery of infinite-dimensional stimuli. Finally, Section 6 provides the context for our research and its relation to Bayesian estimates, as well as approaches to globally optimal reconstructions. Section 7 concludes our work and discusses potential future directions.

2. Representation and recovery of time-varying stimuli

Encoding of space-time visual stimuli with neural circuits leads to a fairly complex neural architecture. Since our goal is to present in this paper a rigorous framework for both representation and recovery of visual information, we will first introduce the simpler case of encoding time-varying signals. In this way the reader can develop the needed intuition to deal with the more general encoding of space-time stimuli. As will be clear in Section 3, the key neural building blocks of the encoding architecture for visual stimuli require the careful treatment described below.

Following a short introduction to the space of trigonometric functions, we present a general framework for the representation and recovery of time-varying functions with spiking neuron models. The neuron models considered are of integrate-and-fire and threshold-and-fire type and arise as spiking neuron models in early vision.

2.1. Modeling stimuli as trigonometric functions

In this section we briefly introduce the spaces of trigonometric polynomials and discuss how they can be used for modeling sensory stimuli of interest. We show that trigonometric polynomials are natural discretizations of bandlimited functions, suitable for applications.

In the univariate case, the space of trigonometric polynomials consists of functions that are simultaneously bandlimited with bandwidth Ω (in rad/sec) and periodic with period T. The period and bandwidth are related with each other by the relation

$$T = \frac{2\pi M}{\Omega}, \tag{1}$$

where M is a positive integer that denotes the order of the space. Let $\mathscr{H}$ denote this space. Then $\mathscr{H}$ consists of all the functions $u = u(t), t \in \mathbb{R}$, of the form

$$u(t) = \sum_{m=-M}^{M} a_m \exp(jm\omega_M t), \tag{2}$$

where $\omega_M = \Omega/M$. Note that the space of trigonometric polynomials of order M is a natural discretization of the space of bandlimited functions. The discretization is best viewed in the frequency domain. The exponentials in (2) have a line Fourier spectrum at the points $m\omega_M$ with $m = -M, \ldots, M$. By letting $M \to \infty$, this spectrum becomes $[-\Omega, \Omega]$.

Remark 1. The stimuli defined in (2) are in general complex valued functions. To obtain real valued functions, we require $u = \bar{u} \Rightarrow a_0 \in \mathbb{R}$ and $a_{-m} = \overline{a_m}, m = 1, \ldots, M$, where $\bar{u}$ denotes the complex conjugate of u.

The sesquilinear form $\langle \cdot, \cdot \rangle : \mathscr{H} \times \mathscr{H} \mapsto \mathbb{C}$ defined by

$$\langle u, v \rangle = \int_{-T/2}^{T/2} u(s)\overline{v(s)}ds, \tag{3}$$

is an inner product for $\mathscr{H}$ and thus the space $(\mathscr{H}, \langle \cdot, \cdot \rangle)$ is a well defined Hilbert space. It is easy to see that under the inner product (3), the set of functions $(e_m), m = -M, \ldots, M$ defined as

$$e_m(t) = \frac{1}{\sqrt{T}} \exp(jm\omega_M t), \tag{4}$$

constitutes an orthonormal basis for $\mathscr{H}$.

2.2. ON–OFF neuron models

In this section we analyze a single-input two-output time encoding machine (Lazar & Tóth, 2004) with feedback (Fig. 1a). The circuit consists of two interconnected neurons (ON–OFF pair) each with its own feedback. Each neuron is endowed with a level crossing detection mechanisms with a threshold that takes a positive value for the ON component (upper branch) and a negative value for the OFF component (lower branch), respectively. The thresholds are assumed to be i.i.d. Gaussian random variables with normal distributions $\mathcal{N}(\delta^1,(\sigma^1)^2)$ and $\mathcal{N}(-\delta^2,(\sigma^2)^2)$, respectively.

Whenever a spike is generated, for the example in the ON component at time t_k^1, the corresponding membrane potential v^1 is reset by the feedback mechanism. The feedback consists of a causal filter with impulse response $h^{11}(t)$, i.e., $v^1(t_k^1)=\delta_k^1-h^{11}(0)$. Moreover, a new threshold value δ_{k+1}^1 is drawn from the normal distribution. Finally, the spike is also communicated to the OFF component through a cross-feedback (coupling) term that is modeled with a causal filter with impulse response $h^{12}(t)$, i.e., $v^2(t_{k+}^1)=v^2(t_k^1)-h^{12}(0)$. Note that in general, it is required that the new threshold is above the reseted membrane potential, i.e., we need $\mathbb{P}(\delta_{k+1}^1>\delta_k^1-h^{11}(0))\ll 1$. For normal distributions this is satisfied if $h^{11}(0)\gg\sigma^1$, so that resetting is stronger than the threshold noise component. Similarly we require $h^{22}(0)\gg\sigma^2$.

The neural above circuit, first presented in a deterministic setup in Lazar and Pnevmatikakis (submitted for publication), generalizes its single neuron counterpart proposed in Keat, Reinagel, Reid, and Meister (2001) as a flexible model for fitting the responses of RGCs and neurons in the LGN. Its spiking mechanism can be viewed as a simplified version of the spike response model (Gerstner & Kistler, 2002). The pairs of coupled neurons in Fig. 1a arise as models of ON and OFF bipolar cells in the retina and their connections through the non-spiking horizontal cells (Masland, 2001).

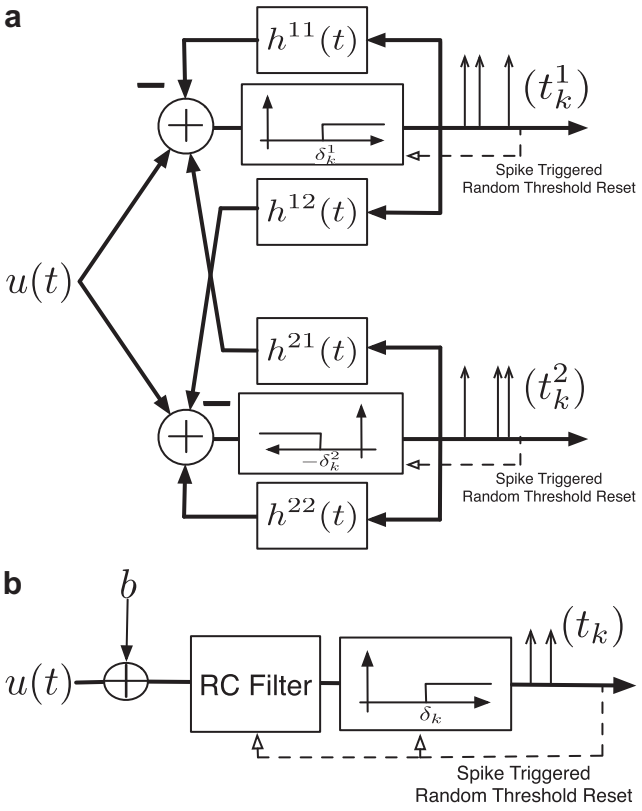

Fig. 1. Canonical neural encoding circuits.

Similar models have also been proposed for various modeling tasks, e.g., (Truccolo, Eden, Fellows, Donoghue, & Brown, 2005; Pillow et al., 2008).

2.2.1. Stimulus encoding and the t-transform

Let $(t_k^j), k=1,2,\ldots,n_j$, be the set of spike times of the neuron j, $j=1,2$. Then the value of the input stimulus can be inferred at the spike times from the equations, formally known as the t-transform (Lazar & Tóth, 2004). Intuitively, the t-transform shows how the neural spike train is associated with a set of linear measurements of the stimulus.

$$\begin{aligned} u(t_k^1) &= +\delta^1+\sum_{l<k}h^{11}(t_k^1-t_l^1)\\ &\quad-\sum_l h^{21}(t_k^1-t_l^2)1_{\{t_l^2<t_k^1\}}\varepsilon_k^1=q_k^1+\varepsilon_k^1\\ u(t_k^2) &= -\delta^2-\sum_{l<k}h^{22}(t_k^2-t_l^2)\\ &\quad+\sum_l h^{12}(t_k^2-t_l^1)1_{\{t_l^1<t_k^2\}}+\varepsilon_k^2=q_k^2+\varepsilon_k^2, \end{aligned} \tag{5}$$

for all $k,k=1,2,\ldots,n_j$, where $\varepsilon_k^1\sim\mathcal{N}(0,(\sigma^1)^2)$ and $\varepsilon_k^2\sim\mathcal{N}(0,(\sigma^2)^2)$. The Eq. (5) show that neural spiking in this circuit is equivalent with the point evaluation of the input stimulus u at the spike times, and it can be rewritten as a bounded linear functional $L_k^i:\mathscr{H}\mapsto\mathbb{R}$:

$$L_k^i u=u(t_k^i)=q_k^i+\varepsilon_k^i, \tag{6}$$

for all $k,k=1,2,\ldots,n_i$, and i, $i=1,2$. From Riesz representation theorem, there is a unique element χ_k^i in $\mathscr{H}$ such that the above linear functional can be written in inner product form as

$$\begin{aligned} \langle u,\chi_k^i\rangle &= L_k^i u=q_k^i+\varepsilon_k^i\\ \left\langle u,\frac{1}{\sigma^i}\chi_k^i\right\rangle &= \frac{1}{\sigma^i}q_k^i+\tilde{\varepsilon}_k^i \end{aligned} \tag{7}$$

for all $k,k=1,2,\ldots,n_i$, and i, $i=1,2$, where $\tilde{\varepsilon}_k^i\sim\mathcal{N}(0,1)$. In order to become operationally significant, we express the *sampling* functions χ_k^i in $\mathscr{H}$ using the standard basis (e_m) as

$$\chi_k^i(t)=\sum_{m=-M}^{M}b_{m,k}^i e_m(t), \tag{8}$$

with

$$b_{m,k}^i=\langle\chi_k^i,e_m\rangle=\overline{\langle e_m,\chi_k^i\rangle}=L_k^i\overline{e_m}=\overline{e_m(t_k^i)}=e_{-m}(t_k^i). \tag{9}$$

The sampling functions χ_k^j and the projections q_k^j are determined by the parameters of the neurons and the spike times. Thus, the t-transform maps the amplitude information of the stimulus into the time information carried by the spike trains.

2.2.2. Stimulus reconstruction

To derive the reconstructed stimulus, we seek a stimulus that minimizes the following regularized cost functional $J:\mathscr{H}\mapsto\mathbb{R}$ defined by

$$\begin{aligned} J(u) &= \frac{1}{(\sigma^1)^2}\sum_{k=1}^{n_1}\left(q_k^1-\langle u,\chi_k^1\rangle\right)^2+\frac{1}{(\sigma^2)^2}\sum_{k=1}^{n_2}(q_k^2-\langle u,\chi_k^2\rangle)^2\\ &\quad+(n_1+n_2)\lambda\|u\|^2. \end{aligned} \tag{10}$$

The cost functional consists of three terms. The first two represent the faithfulness of the reconstructed error with respect to the original noisy measurements, normalized so that they all have the same variance 1. The third term is a regularization term, used to prevent overfiting, due to the noisy data. Finally, λ is a positive smoothing parameter that regulates the tradeoff between faithfulness to the measurements and smoothness. We have the following result:

Proposition 1. *The minimizer* $\hat{u} = \underset{u\in\mathscr{H}}{\operatorname{argmin}}(J(u))$ *is of the form*

$$\hat{u} = \sum_{m=-M}^{M} c_m e_m, \tag{11}$$

where $c_m, m = -M, -M+1, \ldots, M$, *are appropriate coefficients given by*

$$\mathbf{c} = (\mathbf{G}^H\mathbf{G} + (n_1+n_2)\lambda\mathbf{I})^{-1}\mathbf{G}^H\begin{bmatrix}\frac{1}{\sigma^1}\mathbf{q^1}\\ \frac{1}{\sigma^2}\mathbf{q^2}\end{bmatrix}, \tag{12}$$

with $\mathbf{I}$ *the identity matrix with dimension* n_1+n_2, $[\mathbf{q^j}]_k = q_k^j$ *and* $c = [\mathbf{c}]_m$. *In addition* $\mathbf{G} = \begin{bmatrix}\mathbf{G^1}\\ \mathbf{G^2}\end{bmatrix}$, *where* $\mathbf{G^i}$ *is a matrix of dimension* $n_i \times (2M+1)$ *with entries* $[\mathbf{G^i}]_{km} = \frac{1}{\sigma^i}\overline{b_{m,k}}$, *and* $\mathbf{G}^H$ *denotes the hermitian of matrix* $\mathbf{G}$.

Proof. Since the minimizer lies in the same space $\mathscr{H}$ it can be written as in (11). The system of equations (12) is obtained by plugging (11) into (10) and solving the set of equations $\frac{\partial J}{\partial c_m} = 0$ in terms of $c_m, m = -M, \ldots, M$ (see also Appendix A). □

Remark 2. The matrix $\mathbf{G}^H\mathbf{G}$ has dimensions $(2M+1)\times(2M+1)$, independent of the number of spikes. This shows that setting up the problem in a finite-dimensional space, leads to a recovery with complexity determined by M and not by the number of spikes as in Lazar and Pnevmatikakis (2009) or Lazar and Pnevmatikakis (2010). Moreover, a simple calculation shows that the matrix $\mathbf{G}^H\mathbf{G}$ is Toeplitz and Hermitian with entries given by

$$[\mathbf{G}^H\mathbf{G}]_{m,n} = \frac{1}{(\sigma^1)^2}\sum_{k=1}^{n_1} e_{m-n}(t_k^1) + \frac{1}{(\sigma^2)^2}\sum_{k=1}^{n_2} e_{m-n}(t_k^2).$$

2.3. Integrate-and-fire neuron models

The second neuron model that we examine is a leaky integrate-and-fire (LIF) with random threshold (see Fig. 1b). The stimulus u biased by a constant background current b is fed into a LIF neuron with resistance R and capacitance C. Furthermore, the neuron has a random threshold with mean δ and variance σ^2. The value of the threshold changes only at spike times, i.e., it is constant between two consecutive spikes. Assume that after each spike the neuron is reset to the initial value zero. Integrate-and-fire (IAF) neuron models have been used to model the responses of neurons in the early visual system (Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005). Note that an ON–OFF formulation for IAF models is also possible (Lazar & Pnevmatikakis, 2010; Lazar & Pnevmatikakis, submitted for publication), but is omitted here for simplicity.

2.3.1. Stimulus encoding and the t-transform

Let $(t_k), k = 1, 2, \ldots, n+1$, denote the output spike train of the LIF neuron. Between two consecutive spike times the operation of the neuron is described by the t-transform equations

$$\int_{t_k}^{t_{k+1}} \exp\left(-\frac{t_{k+1}-s}{RC}\right)(b+u(s))\,ds = C\delta_k, \tag{13}$$

where δ_k is the value of the random threshold during the interspike interval $[(t_k, t_{k+1})]$. The t-transform can also be rewritten as

$$L_k u = q_k + \varepsilon_k, \tag{14}$$

where $L_k : \mathscr{H} \mapsto \mathbb{R}$ is a linear functional given by

$$\begin{aligned} L_k u &= \int_{t_k}^{t_{k+1}} \exp\left(-\frac{t_{k+1}-s}{RC}\right)u(s)ds\\ q_k &= C\delta - bRC\left(1-\exp\left(-\frac{t_{k+1}-t_k}{RC}\right)\right)\\ \varepsilon_k &= C(\delta_k - \delta), \end{aligned} \tag{15}$$

where the ε_k's are i.i.d. random variables with mean zero and variance $(C\sigma)^2$ for all $k = 1, 2, \ldots, n$. The sequence $(L_k), k = 1, 2, \ldots, n$, has a simple interpretation: it represents the set of n generalized measurements performed on the stimulus u.

By using the Riesz representation theorem, the measurements of (14) can be given in the inner product form

$$\langle u, \chi_k\rangle = q_k + \varepsilon_k, \tag{16}$$

where the sampling functions $\chi_k, k = 1, 2, \ldots, n$, can be expressed in the standard form as

$$\chi_k = \sum_{m=-M}^{M} b_{m,k} e_m, \tag{17}$$

where similarly to (9), we have

$$\begin{aligned} b_{m,k} = L_k e_{-m} &= \frac{1}{\sqrt{T}}\int_{t_k}^{t_{k+1}} e^{\frac{t_{k+1}-t}{RC}} e_{-m}(t)\,dt\\ &= \frac{RCe_{-m}(t_{k+1}) + (y_k - RC)e_{-m}(t_k)}{\sqrt{T}(1 - jm\omega_M RC)}, \end{aligned} \tag{18}$$

where

$$y_k = RC\left(1-\exp\left(-\frac{t_{k+1}-t_k}{RC}\right)\right), \quad k = 1, 2, \ldots, n.$$

2.3.2. Stimulus reconstruction

Similarly to the previous case, we seek a stimulus $\hat{u} \in \mathscr{H}$ that satisfies

$$\hat{u} = \underset{u\in\mathscr{H}}{\operatorname{argmin}}\left(\sum_{j=1}^{n}(q_k - \langle u, \chi_k\rangle)^2 + n\lambda\|u\|^2\right). \tag{19}$$

The minimizer is given in the Proposition below, whose proof is similar to the one of Proposition 1.

Proposition 2. *The minimizer* $\hat{u}$ *is of the form*

$$\hat{u} = \sum_{m=-M}^{M} c_m e_m, \tag{20}$$

where $c_m, m = -M, -M+1, \ldots, M$, *are appropriate coefficients given by the solution of the system of equations*

$$\mathbf{c} = \left(\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I}\right)^{-1}\mathbf{G}^H\mathbf{q}, \tag{21}$$

where $\mathbf{I}$ *is the identity matrix with dimension* n, $[\mathbf{q}]_k = q_k$ *and* $c = [\mathbf{c}]_m$ *and* $\mathbf{G}$ *is a matrix of dimension* $n\times(2M+1)$ *and entries* $[\mathbf{G}]_{km} = \overline{b_{m,k}}$, *where* $b_{m,k}$ *are given by* (18).

2.4. Examples

In this section we present a detailed example to test the performance of the recovery algorithms presented above. The input space is a space of trigonometric polynomials with signals that are bandlimited with maximum frequency of 50 Hz and periodic with period 0.5 s. In order to avoid the periodic boundary effects that do not appear in practice, the tested signals were restricted to a time interval of length of 0.25 s.

First the signal was encoded with a pair of ON–OFF neurons with random thresholds for 10 different noise levels. At each noise level

the reconstruction algorithm of Proposition 1 was applied for 50 different values of the smoothing parameter λ. Note that the noise levels of the two branches were equal and on average each branch produced roughly 40 spikes. The exact parameters of the neuron pair were $\delta^1 = -\delta^2 = 0.1$, $h^{11}(t) = h^{22}(t) = 0.15\exp(-t/0.01)1_{\{t>0\}}$, and $h^{12} = h^{21}(t) = 0.01\exp(-t/0.015)1_{\{t>0\}}$.

Fig. 2a shows the performance of the algorithm in terms of the signal-to-noise ratio (SNR) averaged over 10 repetitions. Note that the standard error of the mean (SEM) was always below 1 dB (not shown). It can be seen that as the variance of the thresholds decreases, the quality of reconstruction improves and practically reaches excellent recovery (50 dB SNR) for low noise levels. Moreover it can be observed that the smoothing parameter that gives the optimal reconstruction slowly increases with the variance of the thresholds. This is also expected as increased threshold variability essentially increases the noise level in the t-transform and thus calls for more smoothing (larger λ) during reconstruction.

Fig. 2b shows a similar figure for the case when the stimulus is encoded with an LIF neuron. The parameters of the neuron were $b = 2.5$, $\delta = 0.8$, $R = 30$, and $C = 0.01$ and the neuron produced an average of roughly 75 spikes per trial. The qualitative behavior of the SNR is the same as in the previous case of the ON–OFF neural circuit and exhibits a graceful degradation of the SNR as the threshold variability increases.

A close observation of the two figures shows that for the same level of noise power, the stimulus encoded with the ON–OFF neural circuit can be reconstructed with a substantially higher SNR than the one encoded with the LIF neuron. An explanation for this comes from the observation of the t-transform Eqs. (5) and (13). In the case of the ON–OFF neural circuit, the spikes of the circuit correspond to irregular samples of the signal at values that are related to the thresholds. Therefore each random measurement has a mean that is in general away from zero and thus the effect of the threshold variability is limited. The situation is different for the LIF neuron. Due to the existence of the bias b, the neuron fires even if the contribution of the stimulus is minimal. Moreover the integrator averages out the contribution of zero mean signals. Therefore, the mean-to-standard deviation for the corresponding random samples is much lower and consequently the effect of threshold variability much larger.

3. Encoding and decoding of visual stimuli

In this section we extend the formalism presented above to space-time varying visual stimuli. The signals belong again to a space of trigonometric polynomials with appropriate parameters. The neural encoding architecture consists of a population of spiking neural circuits with spatial receptive fields, such as center-surround and Gabor, that are selective to certain features of the input stimulus. These have been widely used to model receptive fields in the retina, LGN and V1. The spiking mechanisms of the circuits are either integrate-and-fire or ON–OFF with thresholding and feedback, as analyzed in the previous section, and are assumed to have random thresholds. We note that the methodology employed here is very general and allows for an arbitrary combination of the receptive fields and spiking neuron models.

By establishing the t-transform of the encoding architecture, we show how the population of spike trains is equivalent with a noisy inner product representation of the input visual signal. We then derive an optimal reconstruction algorithm based on the theory of smoothing splines. We test the algorithm for both the relatively simple case of synthetic video streams as well as for the case of natural scenes.

3.1. The space of trigonometric visual stimuli

We denote by $\mathscr{V}$ the space of trigonometric video sequences with spatial bandwidths Ω_x and Ω_y, temporal bandwidth Ω_t, and order (resolution) M_x, M_y, M_t, respectively. The video sequences $I \in \mathscr{V}$ are periodic and can be completely defined on the grid $\mathbb{D} = \mathbb{S}_x \times \mathbb{S}_y \times \mathbb{T}$, where

$$\begin{aligned} \mathbb{S}_x &= [-S_x/2, S_x/2], \quad S_x = 2\pi M_x/\Omega_x, \\ \mathbb{S}_y &= [-S_y/2, S_y/2], \quad S_y = 2\pi M_y/\Omega_y, \\ \mathbb{T} &= [-T/2, T/2], \quad T = 2\pi M_t/\Omega_t. \end{aligned} \tag{22}$$

With $\omega_{M_t} = \Omega_t/M_t$, ω_{M_x} and ω_{M_y} similarly defined, the space $\mathscr{V}$ consists of all the functions of the form

$$I(x,y,t) = \sum_{m_x=-M_x}^{M_x} \sum_{m_y=-M_y}^{M_y} \sum_{m_t=-M_t}^{M_t} a_{m_x,m_y,m_t} \exp(jm_x\omega_{M_x}x + jm_y\omega_{M_y}y + jm_t\omega_{M_t}t). \tag{23}$$

The space endowed with the inner product $\langle\cdot,\cdot\rangle : \mathscr{V} \times \mathscr{V} \mapsto \mathbb{C}$

$$\langle I_1, I_2\rangle = \int_{\mathbb{D}} I_1(x,y,t)\overline{I_2(x,y,t)}\,dx\,dy\,dt, \tag{24}$$

is a Hilbert space with dimension $(2M_t+1)(2M_x+1)(2M_y+1)$ and the set of functions

$$e_{m_x,m_y,m_t}(x,y,t) = \frac{1}{\sqrt{S_xS_yT}} \exp(jm_x\omega_{M_x}x + jm_y\omega_{M_y}y + jm_t\omega_{M_t}t), \tag{25}$$

constitutes an orthonormal basis for $\mathscr{V}$. It is clear that the functions above can be written into the form

$$e_{m_x,m_y,m_t}(x,y,t) = e_{m_x}(x)e_{m_y}(y)e_{m_t}(t).$$

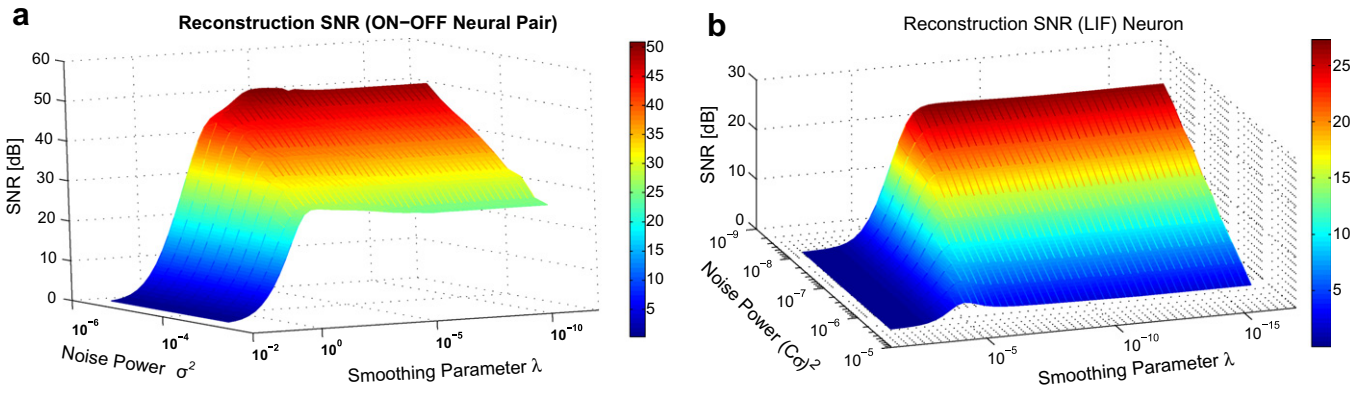

Fig. 2. Mean reconstruction signal-to-noise ratio for a stimulus encoded with: (a) a pair of ON–OFF neurons and (b) an IAF neuron.

3.2. Encoding of visual stimuli and the t-transform

The general encoding architecture is shown in Fig. 3. The input video I is filtered by a set of spatial receptive fields D^j, $j = 1, 2, \ldots, N$. The resulting dendritic currents v^j, $j = 1, 2, \ldots, N$, serve as the input to N spiking neural circuits realized with IAF neurons or ON–OFF threshold-and-fire neuron pairs.

Filtering the video stream with the receptive field of the neural circuit j gives the receptive field output $v^j(t)$ which amounts to

$$v^j(t) = \int_{\mathbb{S}} D^j(x,y) I(x,y,t) dx dy, \tag{26}$$

where $\mathbb{S} = \mathbb{S}_x \times \mathbb{S}_y$.

Neural spiking is interpreted here as a series of linear functionals acting on the input stimulus. In what follows the left superscripts ${}^{\mathbb{T}}, {}^{\mathbb{S}}$ and ${}^{\mathbb{D}}$ indicate action (of functionals) on the temporal, spatial and spatiotemporal domain, respectively. Following the discussion of Section 2, the t-transform of ith branch of the jth neural circuit is described by

$${}^{\mathbb{T}}L_k^i v^j = q_k^{ji} + \varepsilon_k^{ji}, \tag{27}$$

where ${}^{\mathbb{T}}L_k^i : \mathscr{H} \mapsto \mathbb{R}$ is a linear functional. This functional is an instantiation of the evaluation functional at time t_k^{ji} defined in (6) for the case of the ON–OFF neuron pair or the linear functional defined in (15) for the case of the LIF neuron, and $\varepsilon_k^{ji} \sim \mathcal{N}(0, (\sigma^{ji})^2)$. Note that (26) can be rewritten in a functional form as

$${}^{\mathbb{S}}L^j I = \int_{\mathbb{S}} D^j(x,y) I(x,y,t) dx dy = v^j(t). \tag{28}$$

Combining (27) and (28) we obtain

$${}^{\mathbb{D}}L_k^{ji} I = q_k^{ji} + \varepsilon_k^{ji}, \tag{29}$$

where ${}^{\mathbb{D}}L_k^{ji} : \mathscr{V} \mapsto \mathbb{R}$ is a linear functional given by ${}^{\mathbb{D}}L_k^{ji} = {}^{\mathbb{T}}L_k^i {}^{\mathbb{S}}L^j$. Therefore with each spike (or spike pair) we can associate a linear functional acting on the input visual stimulus. We seek again to express these functionals in an inner product form. The following lemma provides the needed representation.

Lemma 1. *The t-transform can be written in inner product form as*

$$\langle I, \phi_k^{ji} \rangle = q_k^{ji} + \varepsilon_k^{ji}, \tag{30}$$

where ϕ_k^{ji} is of the form of the right-hand-side of (23) with

$$a_{m_x,m_y,m_t,k}^{ji} = ({}^{\mathbb{S}}L^j e_{-m_x,-m_y})({}^{\mathbb{T}}L_k^i e_{-m_t}). \tag{31}$$

Proof. The representation result holds because of the Riesz representation theorem. To find the coefficients, we have that

$$\begin{aligned} a_{m_x,m_y,m_t,k}^{ji} &= \langle \phi_k^{ji}, e_{m_x,m_y,m_t} \rangle = \overline{\langle e_{m_x,m_y,m_t}, \phi_k^{ji} \rangle} = {}^{\mathbb{D}}L_k^{ji}(e_{-m_x,-m_y,-m_t}) \\ &= ({}^{\mathbb{S}}L^j e_{-m_x,-m_y})({}^{\mathbb{T}}L_k^i e_{-m_t}). \end{aligned}$$

The first term of (31) is independent of the spiking mechanism and equals to

$$\begin{aligned} {}^{\mathbb{S}}L^j e_{-m_x,-m_y} &= \frac{1}{\sqrt{S_x S_y}} \int_{\mathbb{S}} D^j(x,y) \exp(-jm_x\omega_{M_x}x - jm_y\omega_{M_y}y) dx dy : \\ &= d_{m_x,m_y}^j, \end{aligned} \tag{32}$$

whereas the second term is equals to $b_{m_t,k}^i$, given by (9) for the ON–OFF neuron case and (18) for the case of the LIF neuron. Therefore

$$a_{m_x,m_y,m_t,k}^{ji} = d_{m_x,m_y}^j b_{m_t,k}^i. \tag{33}$$

□

3.3. Visual stimulus decoding

As before, an estimate of the visual stimulus I based on the set of t-transform equations, as imposed by spike trains, satisfies

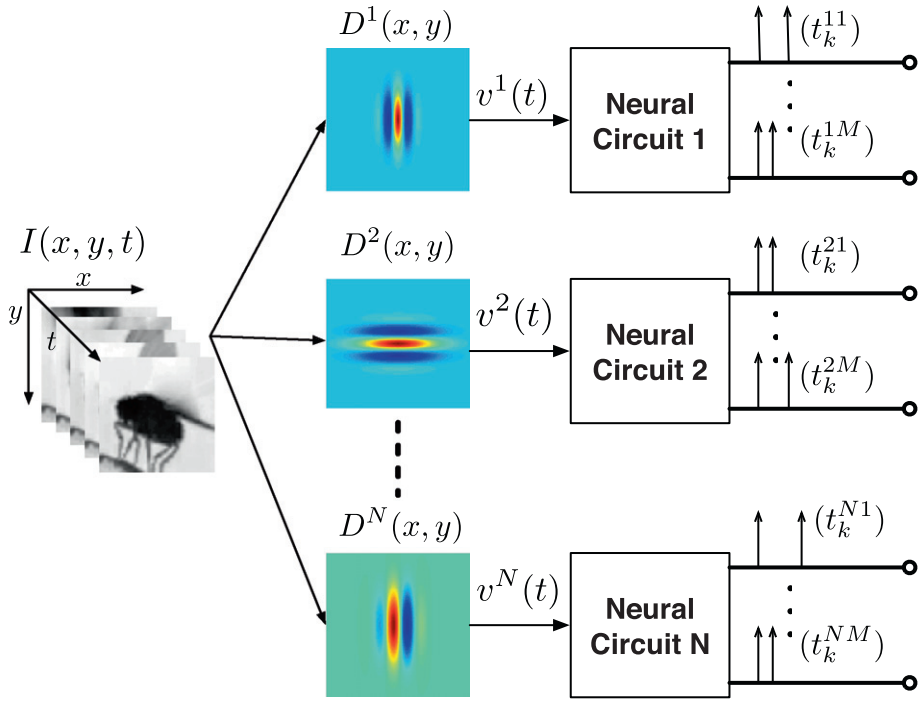

Fig. 3. Architecture for encoding visual stimuli.

$$\widehat{I} = \underset{I\in\mathscr{H}}{\operatorname{argmin}} \left(\sum_{j=1}^{N} \sum_{i=1}^{M} \frac{1}{(\sigma^{ji})^2} \left(\sum_{k=1}^{n_{ij}} \left(q_k^{ji} - \langle I, \phi_k^{ji} \rangle \right)^2 \right) + \left(\sum_{j=1}^{N} \sum_{i=1}^{M} n_{ij} \right) \lambda \|I\|^2 \right). \tag{34}$$

We have the following theorem

Theorem 1. *The minimizer $\widehat{I}$ is given by*

$$I(x,y,t) = \sum_{m_x=-M_x}^{M_x} \sum_{m_y=-M_y}^{M_y} \sum_{m_t=-M_t}^{M_t} c_{m_x,m_y,m_t} e_{m_x,m_y,x_t}(x,y,t), \tag{35}$$

where c_{m_x,m_y,m_t} are suitable coefficients given by the solution of the system of equations

$$\mathbf{c} = (\mathbf{G^H G} + n\lambda \mathbf{I})^{-1} \mathbf{G^H q}, \tag{36}$$

where $n = \sum_{j=1}^{N} \sum_{i=1}^{M} n_{ji}$, $\mathbf{q} = [\mathbf{q}^1, \mathbf{q}^2, \ldots, \mathbf{q}^N]^T$, $\mathbf{q}^j = \left[\frac{1}{\sigma^{j1}} \mathbf{q}^{j1}, \frac{1}{\sigma^{j2}} \mathbf{q}^{j2}, \ldots, \frac{1}{\sigma^{jM}} \mathbf{q}^{jM} \right]^T$ *and* $[\mathbf{q}^{ji}]_k = q_k^{ji}$, $\mathbf{c}$ *is a column vector containing* $(2M_x+1)(2M_y+1)(2M_t+1)$ *entries traversing all possible subscript combination of ordered indices m_x, m_y, m_t for c_{m_x,m_y,m_t}. The entries of the matrix* $\mathbf{G}$ *are given by*

$$\mathbf{G} = \left[\frac{1}{\sigma^{11}} \mathbf{G^{11}}, \frac{1}{\sigma^{12}} \mathbf{G^{12}}, \ldots, \frac{1}{\sigma^{1M}} \mathbf{G^{1M}}, \ldots, \frac{1}{\sigma^{N1}} \mathbf{G^{N1}}, \frac{1}{\sigma^{N2}} \mathbf{G^{N2}}, \ldots, \frac{1}{\sigma^{NM}} \mathbf{G^{NM}} \right]^H,$$

where

$$\mathbf{G^{ji}} = \left[\mathbf{A}_1^{ji}, \mathbf{A}_2^{ji}, \ldots, \mathbf{A}_{n^{ji}}^{ji} \right]^H, \tag{37}$$

$\mathbf{A}_k^{ji}$ *is a row vector containing* $(2M_x+1)(2M_y+1)(2M_t+1)$ *entries traversing all possible subscript combination of* $\overline{a^{ji}_{m_x,m_y,m_t,k}}$ *defined in (33) in the same order as in* $\mathbf{c}$*, for all i = 1,2,...,N; j = 1,2,...,M and* $k = 1, 2, \ldots, n_{ji}$.

The decoding circuit (time decoding machine) is depicted in Fig. 4.

4. Examples

In this section we present two examples that demonstrate the performance of our algorithm and highlight its key features. The first example describes the encoding of an synthetic video stream with a population of ON–OFF neural circuits, with center-surround receptive fields arising in the retina and LGN. The second example deals with the encoding of a natural scene flow with a population of IAF neurons, with receptive fields forming a Gabor wavelet filterbank arising in V1. We provide detailed recovery statistics as well as videos that compare the original natural scenes with the reconstructed ones. We explicitly show the visual error signal and the spectrum of the error signal as a function of time for various random threshold distributions (Gaussian and Gamma) and distribution parameters. The videos are part of the Supplementary material.

4.1. Synthetic video example

A synthetic (real) video stream was constructed based on equation (23), with $M_x = M_y = M_t = 8$ and domain $\mathbb{D} = [-2,2] \times [-2,2] \times [-0.1, 0.1]$. The maximum temporal bandwidth was 20 Hz and the spatial bandwidth 1 Hz in each direction. The video stimulus was encoded with a population of symmetric ON–OFF circuits (presented in Section 2.2), identical to each other with parameters $\delta^1 = \delta^2 = 0.05$, $h^{12}(t) = h^{21}(t) = 0.01 \exp(-t/0.015) 1_{\{t>0\}}$ and $h^{11}(t) = h^{22}(t) = 0.5 \exp(-t/0.01) 1_{\{t>0\}}$ (time in seconds). Each threshold was chosen to be distributed according to a Gaussian distribution, with a threshold variability of 1%, i.e., $\sigma = \delta/100$. The receptive fields formed a filterbank generated from Difference-of-Gaussian (DoG) mother wavelet that has been used to model retinal ganglion cells (RGCs) (Rodieck, 1965; Van Rullen & Thorpe, 2001). The filterbank consisted of five different scalings and suitable number of translations to ensure that in each scaling, the filters extend to the whole spatial domain. We performed eight simulations with different number of neuron pairs. In each simulation, we gradually decreased the distance between the neighboring pairs in each scaling to cover the spatial domain more tightly.

In Fig. 5 we show the performance of the reconstruction algorithm (for fixed $\lambda = 10^{-6}$) as a function of the number of neuron pairs, resulting from different spacing of them. The x-axis corresponds to the number of neuron pairs that actually fired at least one spike. The total number of spikes is also depicted along the same axis. As it can be seen, the quality of the reconstruction (SNR, PSNR) (Lazar & Pnevmatikakis, submitted for publication) improves as more neurons are used to encode the stimulus. These

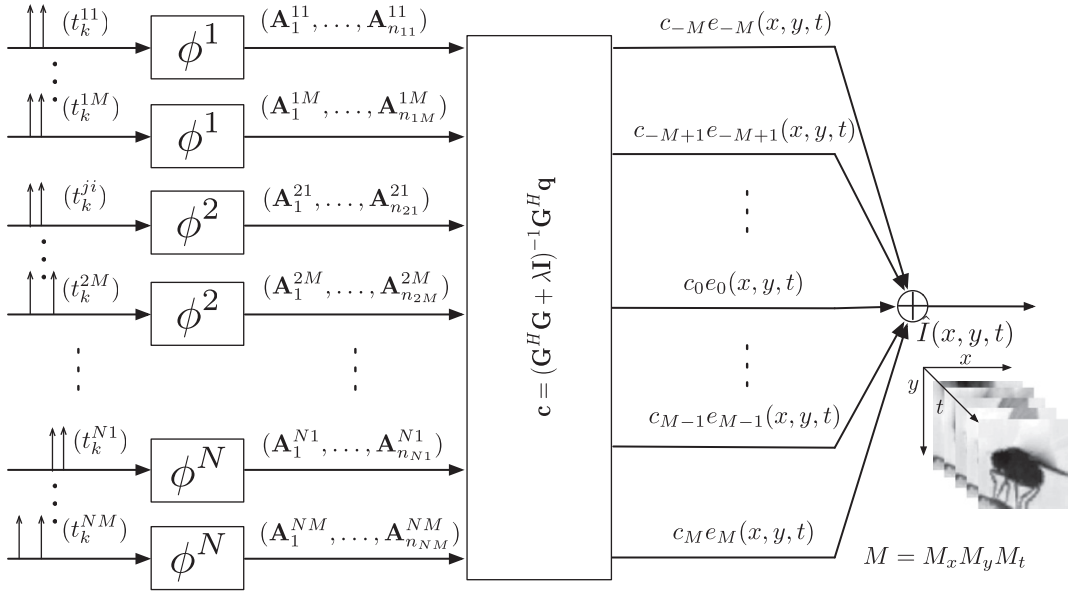

Fig. 4. Architecture for decoding visual stimuli.

results demonstrate that increasing the number of neurons achieves a better encoding of the input stimuli; they are consistent with basic evolutionary thought (Lazar & Pnevmatikakis, 2008b). The percentage of the neuron pairs that fired was in all cases around 70%. No specific increasing or decreasing pattern of this percentage was found as the total number neurons was increased.

4.2. Natural scene example

The second example pertains to a natural video scene, where the flight initiation of a Drosophila was recorded with a high-quality digital camera (Card & Dickinson, 2008). The neural architecture that was used to encode this signal, consisted of a population of IAF neurons with Gabor receptive fields. Although the flight of the fruit fly imposed strong requirements on the encoding architecture the decoding circuit was able to recover the visual stimulus even under noisy conditions (see Supplementary material).

4.2.1. Modeling natural scenes as trigonometric polynomials

The video had a frame rate of 6 kHz (maximum temporal bandwidth of 3 kHz) and a duration of 120 frames (20 ms) and a spatial resolution of 96×96 pixels. For simplicity the video was dilated in the time domain to have length of 1 s (maximum temporal bandwidth 60 Hz). The spatial domain was chosen (arbitrarily) to be $[-3,3] \times [-3,3]$, yielding a maximum spatial bandwidth of 8 Hz in each direction. To avoid the effects of periodicity, the spatial domain was embedded within a space of trigonometric polynomials with domain $[-5,5] \times [-5,5]$, yielding a fundamental frequency of 0.1 Hz in each direction. For similar reasons the temporal domain was embedded into one of 40 ms duration (fundamental frequency 0.5 Hz).

Fourier analysis on the input stimulus indicated that in the frequency domain most of the energy of the input signal was included in the cylinder $\{(f_t,f_x,f_y) : |f_t| \leqslant 4\text{ Hz}, 0 < \sqrt{f_x^2+f_y^2} \leqslant 4\text{ Hz}\}$. Therefore the resolution of the space of trigonometric space-time functions was chosen to be $M_t = 4/0.5 = 8$ and $0 < \sqrt{m_x^2+m_y^2} \leqslant M$, with $M_x = M_y = M = 4/0.1 = 40$. The importance of the choice of M is highlighted in Fig. 6: the left column in Fig. 6 shows frames 10, 50, and 80 of the original visual stimulus. The next three columns show the same frames when the order of the space of trigonometric polynomials used to model the visual input was $M = 40$, $M = 30$, and $M = 20$. The threshold values of the neural circuits of the encoding architecture were deterministic. The Structural Similarity Index (SSIM) (Wang et al., 2004) values are shown in the table below:

	Order		
	$M = 40$	$M = 30$	$M = 20$
Frame			
10	0.9453	0.7860	0.6781
50	0.9403	0.8935	0.7836
80	0.9711	0.9377	0.8918

These values of the SSIM, that we shall investigate in more detail below, are in agreement with the visual perception that the quality of the model is increasing with M. Clearly, the value of the order of the space of trigonometric polynomials depends on the frequency content, and thereby, on the statistics of the visual field. Increasing M leads to improved stimulus recovery. It also leads to an increase in the complexity of the decoding algorithm. Note that in this setting, increasing the order of the space results in an increase of the spatial bandwidths Ω_x and Ω_y as the fundamental frequencies are determined by the input video and are kept fixed. This is different from the case when the bandwidth is fixed and the order increases. In the latter case, the fundamental frequency becomes smaller and the space converges to the limit to the one of bandlimited functions.

These brief considerations further highlight the flexibility of the spaces of trigonometric polynomials to accurately model natural scenes, while taking into consideration their statistics.

4.2.2. Recovery of natural scenes

The video stimulus was encoded with a population of 3408 IAF neurons. The receptive fields of the population formed a spatial Gabor filterbank generated with the same mother wavelet (Jones & Palmer, 1987). The filterbank consisted of combinations between 8 rotations, 5 dilations and 3 to 11 translations in each direction depending on whether the scaling resulted in a wavelet function with coarse (few translations) or fine resolution (many translations). All the IAF neurons were assumed to be ideal ($R \to \infty$) and all had $C = 1$. The bias varied from neuron to neuron with a mean value of 0.39.

Initially we tested neurons with random thresholds drawn from a Gaussian distribution with mean $\delta = 0.03$ and variance σ^2 for all neurons. At every repetition the number of spikes produced was around 46,500. No large deviations were observed as a function of threshold variability.

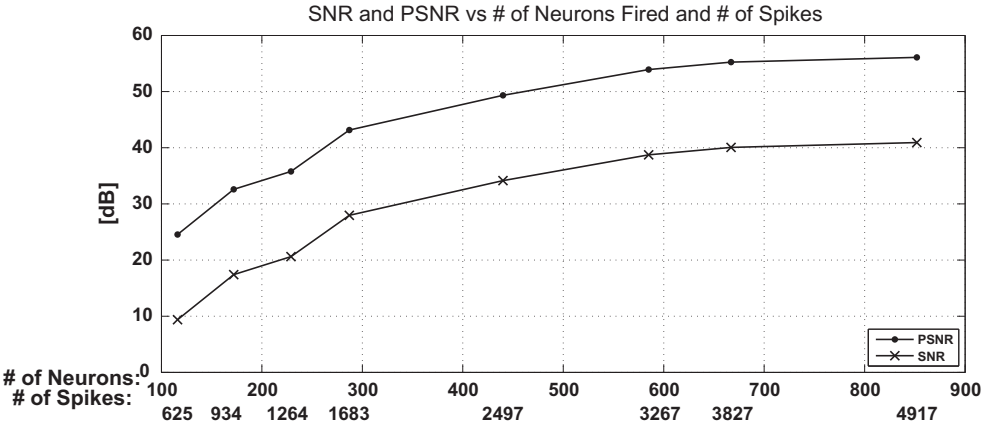

Fig. 5. Performance of the reconstruction algorithm as a function of the number of active circuits/spikes that encode the stimulus.

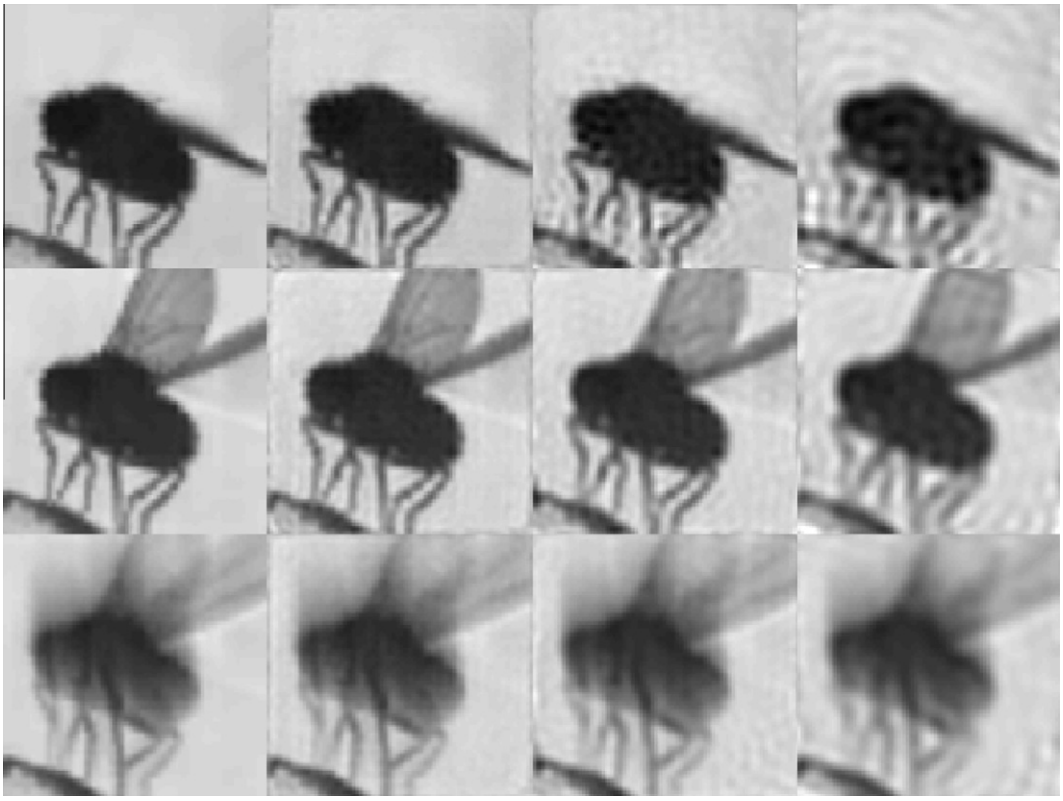

Fig. 6. Modeling a flow of natural scenes.

Fig. 7a shows the performance of the recovery algorithm for various noise levels and various values of the smoothing parameter λ. The threshold variability is defined as the coefficient of variation of the thresholds, i.e., σ/δ. Note that for large threshold variability, the Gaussian distribution was truncated in order to impose positive threshold values. However, the mean and the variance of these Gaussian distributions were adjusted such that the truncated Normal distributions have the same mean $\delta = 0.03$ and the same threshold variability. In the left column the SNR of the recovered natural scenes is plotted, whereas in the right column the SSIM is shown. As it can be seen, the reconstruction improves as the threshold variability decreases and it can reach quite high values, e.g., SSIM > 0.9. For extreme values of the threshold variability (e.g., $\sigma/\delta = 1$), we see that the quality of the reconstruction is poor, e.g., SSIM $\approx$ 0.1.

The same experiment was also performed with neurons with random thresholds drawn from a Gamma distribution. The results are depicted in Fig. 7b. For small threshold variability values, e.g., $\sigma/\delta < 0.1$, the Gamma distribution "resembles" a Gaussian distribution. It starts to visibly differ from the (truncated) Gaussian distribution at higher threshold variability levels. For example for $\sigma/\delta = 1$, the Gamma distribution is exactly an exponential and is significantly different from the (truncated) Gaussian. The maximum difference of the recovery results when using the two distributions was 0.91dB for SNR and 0.0239 for SSIM and was mostly observed when the threshold variability was high. Overall, our simulation results shown in Fig. 7 indicate that the quality of the recovered stimulus displays small differences when encoding with neural circuits with random thresholds drawn from these two distributions.

Finally, Fig. 8 shows the original visual stimulus, the recovered stimulus, the error and the spectrum of the error for frame 10, 50, and 80, respectively. The coefficient of variation was set to 1% around the mean (Gaussian thresholds). As can be seen in Fig. 8, the quality of stimulus recovery is very high. Moreover, the noise of the recovered natural scenes is white when restricted to the frequency support of the input space. Real-time videos exploring the behavior of the encoding architecture with neural circuits with random thresholds drawn from both Gaussian and Gamma distributions, the nature of the recovery error, as well stimulus recovery for deterministic threshold values are shown in the Supplement.

5. Reconstruction of infinite-dimensional stimuli

The results presented so far, can be easily extended to the case of infinite-dimensional stimuli. The tools required are provided by the theory of smoothing splines (Wahba, 1990) in Reproducing Kernel Hilbert Spaces (RKHS) (Berlinet & Thomas-Agnan, 2004). In essence, a Hilbert space $(\mathscr{H}, \langle\cdot,\cdot\rangle)$ defined on a domain $\mathbb{T}$ is called a RKHS if it has the property that the evaluation functional at every point $t \in \mathbb{T}$ is bounded. If $\mathscr{H}$ is a RKHS then there exists a unique function $K : \mathbb{T} \times \mathbb{T} \mapsto \mathbb{C}$, called the *reproducing kernel* (RK) such that $K(\cdot, t) \in \mathscr{H}$ and for any $u \in \mathscr{H}$ and any $t \in \mathscr{T}$ the so called reproducing property

$$\langle u, K(\cdot, t)\rangle = u(t),$$

holds. It is easy to see that the space of trigonometric polynomials, as well as any finite-dimensional vector space, is a RKHS. It's reproducing kernel, called the Dirichlet kernel, is given by

$$K_M(s,t) = \sum_{m=-M}^{M} e_m(s)\overline{e_m(t)} = \frac{1}{T}\sum_{m=-M}^{M} e^{jm\omega_M(s-t)} = \frac{2M+1}{T}\,\frac{\operatorname{sinc}\left(\frac{(2M+1)\Omega}{2M}(s-t)\right)}{\operatorname{sinc}\left(\frac{\Omega}{2M}(s-t)\right)}, \tag{38}$$

where $\operatorname{sinc}(x) = \sin(x)/x$. By letting $M \to \infty$ it is easy to see that

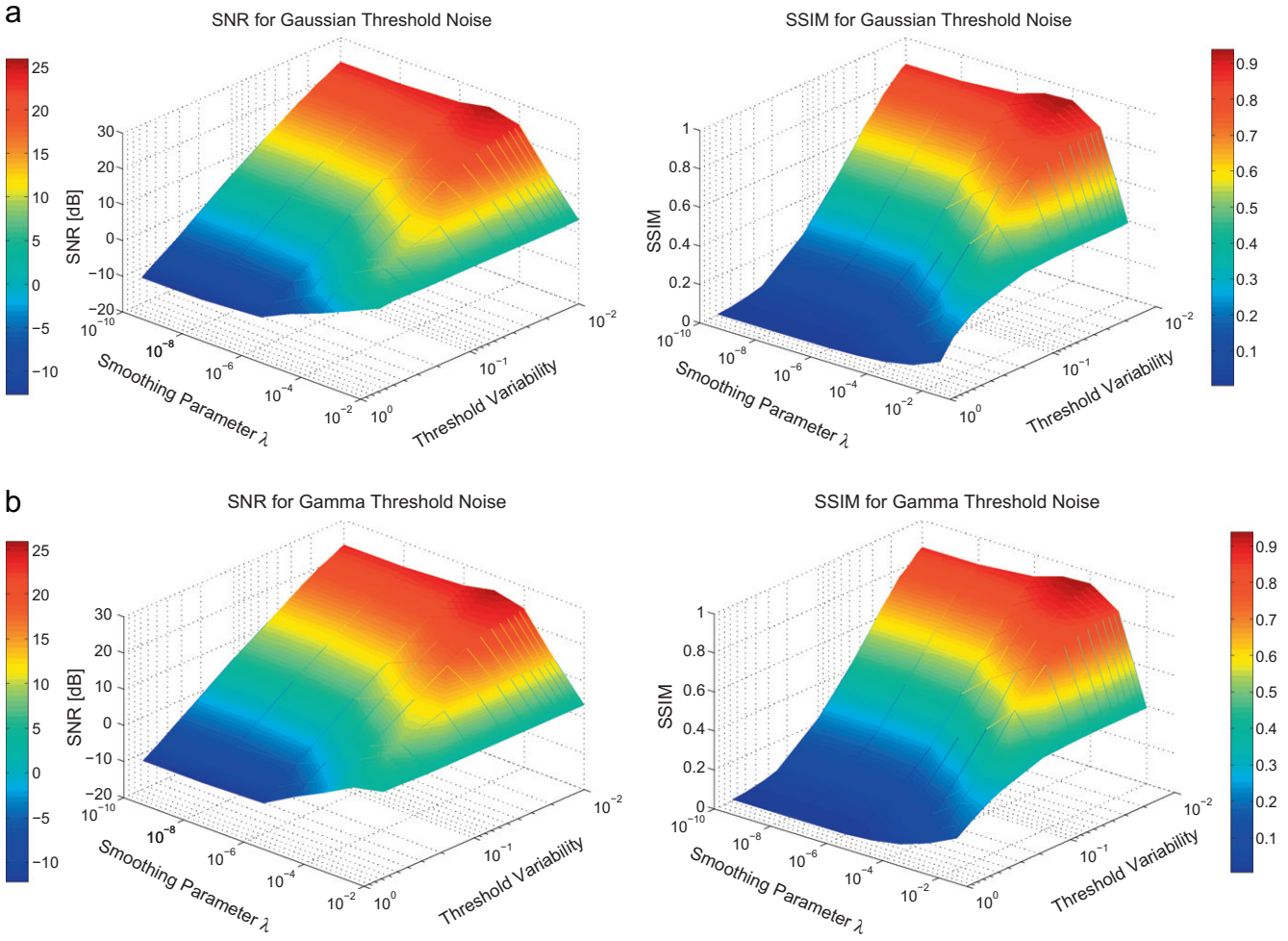

Fig. 7. Performance of the recovery algorithm for Gaussian and Gamma distributed thresholds.

$$\lim_{M\to\infty} K_M(s,t) = \frac{\Omega}{\pi}\operatorname{sinc}(\Omega(s-t)) = \frac{\sin(\Omega(s-t))}{\pi(s-t)},$$

i.e., exactly is the RK for the space of bandlimited functions. Therefore trigonometric polynomials are a natural, finite, discretization of bandlimited functions.

In the case of finite-dimensional spaces, we can express any linear measurement (functional) $L_k u$ in an inner product form $\langle u, \chi_k\rangle$. The sampling function is evaluated in terms of the space basis with appropriate coefficients (see Eqs. (8),(9)). In the general RKHS the sampling functions are computed using the reproducing property (Lazar & Pnevmatikakis, 2009):

$$\chi_k(t) = \langle \chi_k, K(\cdot,t)\rangle = \overline{\langle K(\cdot,t),\ \chi_k\rangle} = L_k\overline{K(\cdot,t)}.$$

Suppose now that a receiver reads the following noisy measurements

$$q_k = \langle u,\ \chi_k\rangle + \varepsilon_k, \tag{39}$$

where ε_k are i.i.d. Gaussian random variables. The following theorem is a special case of a very general result in the theory of smoothing splines, proven in (Wahba, 1990).

Theorem 2. *The minimizer $\hat{u}$ of*

$$\sum_{k=1}^{n}(q_k - \langle u,\ \chi_k\rangle)^2 + n\lambda\|u\|^2, \tag{40}$$

is given by

$$\hat{u} = \sum_{k=1}^{n} d_k\chi_k. \tag{41}$$

Furthermore, the optimal coefficients $[\mathbf{d}]_k = d_k$ satisfy the matrix equation

$$\mathbf{d} = (\mathbf{F} + n\lambda\mathbf{I})^{-1}\mathbf{q}, \tag{42}$$

where $[\mathbf{F}]_{kl} = \langle \chi_k, \chi_l\rangle$, $d_k = [\mathbf{d}]_k$ and $[\mathbf{q}]_k = q_k$, for all k, $l = 1,2,\ldots,n$.

The above theorem states that the minimizer of the cost functional is a linear combination of the sampling functions. Since the sampling functions can be obtained from spike times the decoding problem becomes tractable.

For a finite-dimensional space (41) and (21) are equivalent. To see that, note that $\chi_k = \sum_m b_{k,m}e_m$ and (21) becomes

$$\hat{u} = \sum_{k=1}^{n} d_k \sum_m b_{k,m}e_m = \sum_m \left(\sum_{k=1}^{n} b_{k,m}d_k\right)e_m.$$

Moreover, since $\mathbf{F} = \mathbf{G}\mathbf{G}^H$, it suffices to prove that $\mathbf{G}^H\mathbf{d} = \mathbf{c}$. Since

$$(\mathbf{G}\mathbf{G}^H + n\lambda\mathbf{I})\mathbf{d} = \mathbf{q},$$
$$(\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})\mathbf{c} = \mathbf{G}^H\mathbf{q},$$

and therefore,

$$(\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})\mathbf{c} = \mathbf{G}^H(\mathbf{G}\mathbf{G}^H + n\lambda\mathbf{I})\mathbf{d} = (\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})\mathbf{G}^H\mathbf{d},$$

and the result follows.

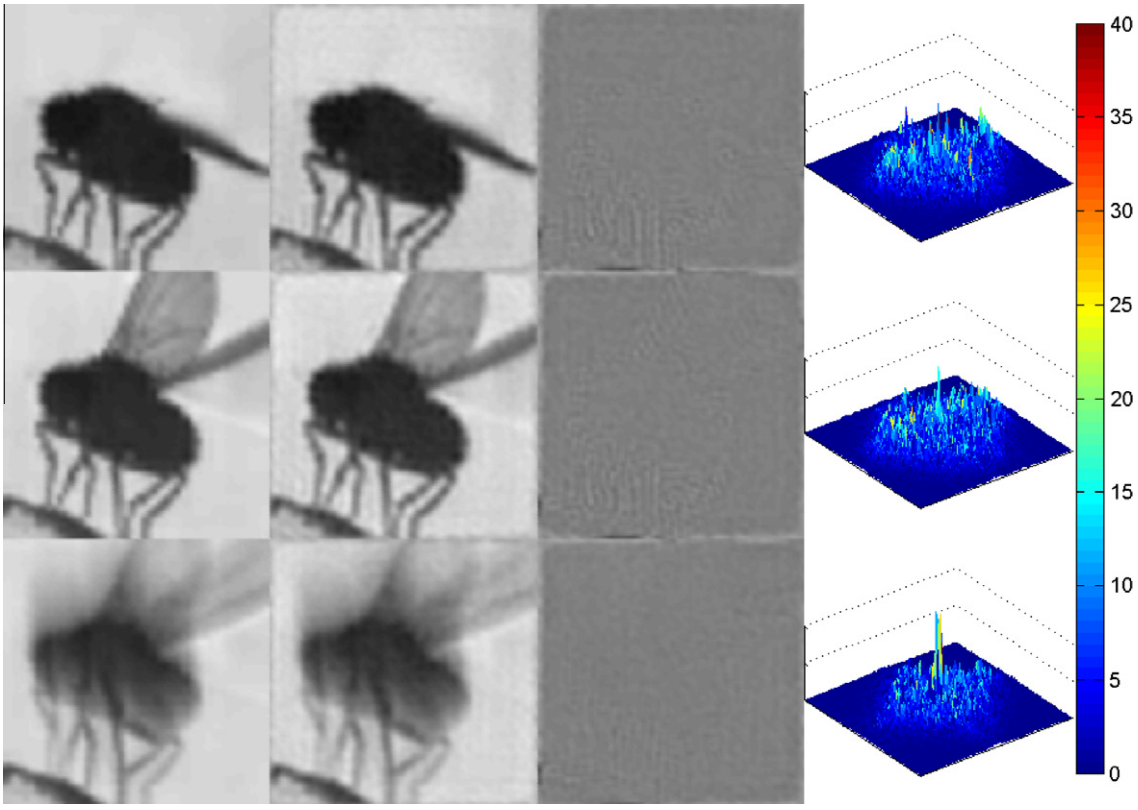

Fig. 8. Recovery of the natural scenes flow.

6. Discussion – related work

Our decoding approach is based on two steps: first, each interspike interval is associated with a generalized measurement of the input stimulus, in the form of an inner product operation. Second, based on these measurements, the reconstructed stimulus minimizes a certain cost functional. As it was seen from the examples, this methodology provides excellent stimulus recovery for highly complex stimuli, such as natural scenes. However, there are two questions that naturally arise.

First, the t-transform equations do not include information about the membrane potential. If t_k, t_{k+1} are two consecutive spikes, then $V(t) < \delta_{k+1}$ for all $t \in [t_k, t_{k+1}]$, where δ_{k+1} is the threshold of the neuron in the same interspike interval. Such *inequality* constraints were considered in Paninski, Pillow, and Simoncelli (2004) in the context of maximum likelihood estimation of the parameters of a LIF neuron. In our problem setting they can be introduced as additional hard constraints for stimulus recovery solved using quadratic programming methods (Boyd & Vandenberghe, 2004). However, the incorporation of inequality constraints in simulations did not show a marked improvement in the reconstructed stimulus. In general, the equality constraints of the t-transform equations appear to be much more informative than the inequality constraints. Note that in the noiseless case, the t-transform completely determines the input stimulus under certain spike density conditions (Lazar & Pnevmatikakis, 2008b). Intuitively, the inequality constraints ensure that the reconstructed stimulus does not fire additional spikes in the interval $[t_k, t_{k+1}]$. Assuming that an additional spike occurs, the reconstructed stimulus oscillates fast on the newly formed interspike intervals thereby resulting in a high energy signal. However, even without the inequality constraints, such high energy stimuli are prevented by the regularizer. In the random threshold case, the inequality constraints hold in a probabilistic sense and call for tools from stochastic programming (Birge & Louveaux, 1997). For high threshold variance values, such constraints may be helpful for stimulus reconstruction and need to be thoroughly examined.

Second, what is the best choice of the cost functional? Our approach here follows the classical regularization approach (Tikhonov & Arsenin, 1977). Such regularized cost functionals appear in stochastic filtering as they lead to minimum variance unbiased estimators (MVUE) (Berlinet & Thomas-Agnan, 2004). For inputs modeled as trigonometric functions with Gaussian i.i.d. coefficients the methodology employed here gives an optimal solution. The regularizer controls the energy of the stimulus by giving a uniform penalty across all the stimulus frequencies (basis functions).

Our model encoding architecture combines the following, desirable, characteristics: use of temporal codes, receptive fields with operational significance and neural circuits with feedback for encoding in the presence of noise. It builds upon results obtained previously in the field. We shall focus in the following only on a narrow subset of the vast literature.

Recordings of cell responses to visual stimuli exhibit sub-millisecond precision for many different cell types of the early visual system, including retinal ganglion cells (RGCs) and lateral geniculate nucleus (LGN) neurons (Keat et al., 2001; Uzzell & Chichilnisky, 2004; Reinagel & Reid, 2000). Such recordings suggest that precision contributes fundamentally to the neural code (Butts et al., 2007). A number of computational spiking neuron models have been published (Keat et al., 2001; Pillow et al., 2005; Pillow et al., 2008) that show a certain degree of fit to neural recordings. In this paper, we used spiking neuron models inspired from the aforementioned ones. By showing that these models constitute tractable neural circuit building blocks, we constructed a large scale model architecture for the encoding of natural video scenes. The spatiotemporal neural encoding architecture turned out to be analytically tractable as well.

Individual neurons in the early visual system exhibit remarkable selectivity to various characteristics of the input stimuli (scale, position, orientation, direction of movement, etc.). This selectivity is inherited from the spatiotemporal receptive fields (Ringach, 2004) of the neurons that filter the input. A widely accepted model for the population of receptive fields is the one of space-time wavelet filterbank (Jones & Palmer, 1987; Field & Chichilnisky, 2007) which highlights the encoding properties and capabilities of the visual system, and can reproduce many properties of the ensemble response, orientation and direction selectivity (Hubel & Wiesel, 1962), etc. Such structures have also been shown to lead to optimal coding, in terms of sparsity, of natural scenes (Olshausen, 2002). A few computational models that exploit the structure of the receptive field population exist in the literature, for example the deterministic models in (Lee, 1996; Rozell, Johnson, Baraniuk, & Olshausen, 2008). These models however operate under the rate assumption and represent video streams on a frame-by-frame basis. A stochastic model appeared in (Pillow, Ahmadian, & Paninski, submitted for publication), where the maximum-a-posteriori (MAP) decoder for images encoded with a population neurons with center-surround receptive fields was derived.

In our model, the receptive fields are integrated with the spiking mechanism of the neurons and appear explicitly in the t-transform of the encoder. Hence the action of the receptive fields on the stimulus is fed directly into the neural spiking and consequently used by the optimal stimulus reconstruction algorithm. Our model assumes prior knowledge of the receptive fields and, naturally, the quality of the decoding depends on quality of knowledge of these receptive fields. In the case where these are unknown, similar methods can be used to identify these, as it was shown in Pillow and Simoncelli (2003) for determining the parameters of a LIF neuron. As our examples demonstrated, the receptive fields can have many different shapes (mother-wavelet). What is critical, however, is the number (or density) of filters. As the results in Section 4.1 suggested, there is a density threshold upon which minimal improvement can be made. In essence this is achieved when the receptive fields cover completely the spatial domain, and depends on the spike density of the neurons that respond to the time-varying stimuli (Lazar & Pnevmatikakis, submitted for publication).

Our model exhibits stimulus dependent dynamics and attributes neuronal variability to the effect of random thresholds. As a result the measurements provided by neural spiking are independent both across different neurons and within each individual neuron. Consequently every single interspike interval contributes an independent noisy measurement that is included in the regularized cost functional. Thus our model architecture can efficiently reconstruct complex stimuli such as natural scenes, using a relatively small number of spikes and with moderate complexity.

7. Conclusions

We presented a formal model for the encoding and reconstruction of visual stimuli with a spiking neural architecture akin to the neural ensembles of the early visual system. We described how information is encoded in the time domain and worked out in detail a reconstruction algorithm, based on regularization techniques, for the case of integrate-and-fire neurons as well as for the case of ON–OFF neural circuits with thresholding and feedback. We demonstrated the effectiveness of our algorithm by reconstructing video streams as complex as natural scenes, based solely on the spike times and the neuron parameters.

The paper also introduced trigonometric polynomials as a formal modeling tool for stimuli such as natural scenes. We showed that trigonometric polynomials are a natural discretization of bandlimited functions, with added modeling flexibility and thus suitable for applications.

In terms of future directions, we note that the optimization criteria space as well as the stimulus modeling options remain largely unexplored. For example, the right part of (34) is just the energy of the stimulus. Based on the properties of the stimulus or the desired computational task to be performed, other criteria can be used (Poggio et al., 1985) and other spline models can arise (Duchon, 1977). Moreover, the spaces of trigonometric polynomials have great flexibility and can adapt to the statistical properties of the expected inputs (Van der Schaaf & Van Hateren, 1996). These, along with other issues, will be the subject of future research.

Acknowledgments

The work presented here was supported by the AFOSR under Grant No. FA9550-09-1-0350. E.A. Pnevmatikakis was also supported by a fellowship from the Onassis Public Benefit Foundation. The authors wish to thank Lev E. Givon for configuring the programming environment supporting the implementation of the natural video scenes example, as well as the anonymous reviewers for their comments and suggestions about improving the quality of the paper.

Appendix A. Proof of stimulus reconstruction

We have already given the minimizers to the cost functions in (10), (19) and (34), here we present a general proof for minimizeres of such cost functions in finite dimensional Hilbert spaces.

Proposition 3. *Assume $\mathscr{H}$ is a finite-dimensional Hilbert space with orthonormal basis $e_m, m = 1, 2, \ldots, M$. We define the cost function to be*

$$\sum_{k=1}^{n}(q_k - \langle u, \chi_k\rangle)^2 + n\lambda\|u\|^2, \tag{A.1}$$

where $\chi_k \in \mathscr{H}$ are a set of n sampling functions that gives measurements q_k, and λ is the Tikhonov regularization parameter and $\|\cdot\|$ denotes the norm in $\mathscr{H}$. Also, we assume $\chi_k = \sum_{m=1}^{M} b_{m,k}e_m$. The minimizer to this cost function in $\mathscr{H}$ is given by

$$\hat{u} = \sum_{m=1}^{M} c_m e_m, \tag{A.2}$$

where $c_m, m = 1, 2, \ldots, M$ are appropriate coefficients given by the solution of equation

$$\left(\mathbf{G}^{\mathbf{H}}\mathbf{G} + n\lambda\mathbf{I}\right)\mathbf{c} = \mathbf{G}^{\mathbf{H}}\mathbf{q}, \tag{A.3}$$

with $[\mathbf{G}]_{km} = \overline{b_{m,k}}$, $[\mathbf{q}]_k = q_k$ and $c = [\mathbf{c}]_m$, for all k=1,2,...,n, m = 1,2,...,M.

Proof. Since the minimizer is also in $\mathscr{H}$, let it be of the form (A.2), using matrix form (A.1) becomes

$$J(\mathbf{c}) = \|\mathbf{q} - \mathbf{G}\mathbf{c}\|^2 + n\lambda\mathbf{c}^H\mathbf{c}, \tag{A.4}$$

since both terms are strictly convex, we find the $\mathbf{c}$ that minimizes the cost function when gradient of J equals to zero, thus we have

$$\begin{aligned} &\nabla_{\mathbf{c}}(J) = 0 \\ &\nabla_{\mathbf{c}}(\mathbf{q}^H\mathbf{q} - 2\mathbf{q}^H\mathbf{G}\mathbf{c} + \mathbf{c}^H\mathbf{G}^H\mathbf{G}\mathbf{c} + n\lambda\mathbf{c}^H\mathbf{c}) = 0 \\ &-2\mathbf{G}^H\mathbf{q} + (\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})\mathbf{c} = 0 \\ &(\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})\mathbf{c} = \mathbf{G}^H\mathbf{q} \end{aligned} \tag{A.5}$$

Note that $\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I}$ is often nonsingular even when $\lambda = 0$, $\mathbf{c}$ can be solved uniquely as

$$\mathbf{c} = (\mathbf{G}^H\mathbf{G} + n\lambda\mathbf{I})^{-1}\mathbf{G}^H\mathbf{q}. \tag{A.6}$$

□

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2010.03.015.

References

Berlinet, A., & Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers..

Birge, J., & Louveaux, F. (1997). *Introduction to stochastic programming*. Springer Verlag.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge Univ Pr.

Butts, D., Weng, C., Jin, J., Yeh, C., Lesica, N., Alonso, J., et al. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature, 449*(7158), 92–95.

Card, G., & Dickinson, M. (2008). Performance trade-offs in the flight initiation of drosophila. *Journal of Experimental Biology, 211*(3), 341–353.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp & K. Zeller (Eds.), *Constructive theory of functions of several variables* (pp. 85–100). Springer-Verlag.

Field, G. D., & Chichilnisky, E. J. (2007). Information processing in the primate retina: Circuitry and coding. *Annual Review of Neuroscience, 30*, 1–30.

Gerstner, W., & Kistler, W. (2002). *Spiking neuron models: Single neurons, populations, and plasticity*. Cambridge University Press.

Gestri, G., Mastebroek, H., & Zaagman, W. (1980). Stochastic constancy, variability and adaptation of spike generation: Performance of a giant neuron in the visual system of the Fly. *Biological Cybernetics, 38*(1), 31–40.

Girosi, F., Jones, M., & Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation, 7*(2), 219–269.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*(1), 106–154.

Jones, J., & Palmer, L. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1233–1258.

Kay, K., Naselaris, T., Prenger, R., & Gallant, J. (2008). Identifying natural images from human brain activity. *Nature, 452*(March), 352–356.

Keat, J., Reinagel, P., Reid, R., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron, 30*(3), 801–817.

Knight, B. (1972). Dynamics of encoding in a population of neurons. *Journal of General Physiology, 59*(6), 734–766.

Lazar, A.A., & Pnevmatikakis, E.A. (2008a). A video time encoding machine. In *IEEE international conference on image processing* (pp. 717–720). San Diego, CA.

Lazar, A. A., & Pnevmatikakis, E. A. (2008b). Faithful representation of stimuli with a population of integrate-and-fire neurons. *Neural Computation, 20*(11), 2715–2744.

Lazar, A. A., & Pnevmatikakis, E. A. (2009). Reconstruction of sensory stimuli encoded with integrate-and-fire neurons with random thresholds. *EURASIP Journal on Advances in Signal Processing 2009*, 14. article ID:682930. <http://www.hindawi.com/journals/asp/2009/682930.html>.

Lazar, A. A., & Pnevmatikakis, E. A. (2010). Consistent recovery of sensory stimuli encoded with MIMO neural circuits. *Computational Intelligence and Neuroscience 2010*, 13. doi:10.1155/2010/469658. article ID:469658.

Lazar, A.A., & Pnevmatikakis, E.A. (submitted for publication). Video time encoding machines. *IEEE Transactions on Neural Networks*.

Lazar, A. A., & Tóth, L. T. (2004). Perfect recovery and sensitivity analysis of time encoded bandlimited signals. *IEEE Transactions on Circuits and Systems–I: Regular Papers, 51*(10), 2060–2073.

Lazar, A. A., Simonyi, E. K., & Tóth, L. T. (2008). An overcomplete stitching algorithm for time decoding machines. *IEEE Transactions on Circuits and Systems–I: Regular Papers, 55*(9), 2619–2630.

Lee, T. S. (1996). Image representation using 2D gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(10), 959–971.

Ma, W., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432–1438.

Masland, R. H. (2001). The fundamental plan of retina. *Nature Neuroscience, 4*(9), 877–886.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H., et al. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron, 60*, 915–929.

Olshausen, B. A. (2002). Sparse codes and spikes. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 257–272). MIT Press.

Paninski, L., Pillow, J., & Simoncelli, E. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Computation, 16*, 2533–2561.

Pillow, J.W., Ahmadian, Y., Paninski, L. (submitted for publication). *Model-based decoding, information estimation, and point detection techniques for multi-neuron spike trains*.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *The Journal of Neuroscience, 25*(47), 11003–11013.

Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature, 454*(7207), 995–999.

Pillow, J., & Simoncelli, E. (2003). Biases in white noise analysis due to non-Poisson spike generation. *Computational Neuroscience: Trends in Research, 52*(54), 109–115.

Poggio, T., Torre, V., & Koch, C. (1985). Computational Vision and Regularization Theory. *Nature, 317*(6035), 314–319.

Reich, D., Victor, J., Knight, B., Ozaki, T., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *Journal of neurophysiology, 77*(5), 2836–2841.

Reinagel, P., & Reid, R. (2000). Temporal coding of visual information in the thalamus. *Journal of Neuroscience, 20*(14), 5392.

Ringach, D. L. (2004). Mapping receptve fields in primary visual cortex. *Journal of Physiology, 5583*, 717–728.

Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell responses to visual stimuli. *Vision Research, 5*, 583–601.

Rozell, C., Johnson, D., Baraniuk, R., & Olshausen, B. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Computation, 20*(10), 2526–2563.

Stanley, G., Li, F., & Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *Journal of Neuroscience, 19*(18), 8036–8042.

Tikhonov, A., & Arsenin, V. (1977). *Solutions of ill-posed problems*. VH Winston Washington, DC.

Truccolo, W., Eden, U., Fellows, M., Donoghue, J., & Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology, 93*, 1074–1089.

Uzzell, V. J., & Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *Journal of Neurophysiology, 92*, 780–789.

Van der Schaaf, A., & Van Hateren, J. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research, 36*(17), 2759–2770.

Van Rullen, R., & Thorpe, S. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation, 13*, 1255–1283.

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial Mathematics.

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.

Warland, D., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Physiology, 78*, 2336–2350.