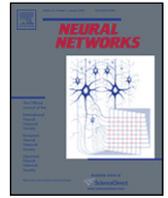




ELSEVIER

Contents lists available at ScienceDirect

Neural Networks

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

2010 Special Issue

Optimization of population decoding with distance metrics

Sonja B. Hofer^a, Thomas D. Mrsic-Flogel^a, Domonkos Horvath^{b,c}, Benedikt Grothe^b,
Nicholas A. Lesica^{b,*}^a Department of Neuroscience, Physiology, and Pharmacology, University College London, London WC1 6JJ, UK^b Department of Biology, Ludwig-Maximilians-University Munich, Martinsried 02152, Germany^c Faculty of Information Technology, Péter Pázmány Catholic University, Budapest, 1083, Hungary

ARTICLE INFO

Article history:

Received 1 October 2009

Received in revised form 12 April 2010

Accepted 27 April 2010

Keywords:

Populations

Decoding

Optimization

ABSTRACT

Recent advances in multi-electrode recording and imaging techniques have made it possible to observe the activity of large populations of neurons. However, to take full advantage of these techniques, new methods for the analysis of population responses must be developed. In this paper, we present an algorithm for optimizing population decoding with distance metrics. To demonstrate the utility of this algorithm under experimental conditions, we evaluate its performance in decoding both population spike trains and calcium signals with different correlation structures. Our results demonstrate that the optimized decoder outperforms other simple population decoders and suggest that optimization could serve as a tool for quantifying the potential contribution of individual cells to the population code.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

With the development of new multi-electrode recording and imaging techniques, it is now possible to record the activity of large neuronal populations simultaneously. While these techniques have already facilitated many important discoveries, the development of new methods for analyzing population activity is required before the full potential of these techniques can be realized. In sensory neuroscience, one of the most powerful methods for analyzing neuronal responses is decoding – using the responses to infer which stimulus evoked them – to determine how different stimulus features are represented in the responses (for a recent review, see [Quiñero and Panzeri \(2009\)](#)). The classical approach to decoding involves estimating the probability distribution of the stimulus conditioned on the observed responses to determine which stimulus was most likely. However, the explicit estimation of this distribution may be difficult in experimental situations where data are limited and the responses are high dimensional (i.e. for example, a large population of cells for which the relevant feature of the responses are the spike times rather than simply the spike rates).

For decoding responses from single cells, several methods have overcome the dimensionality problem by using distance metrics ([van Rossum, 2001](#); [Victor & Purpura, 1996](#)) and these methods have been used successfully in a number of experimental contexts

([Victor, 2005](#)). Distance metrics not only provide an intuitive method of decoding (responses evoked by the same stimulus should be ‘close’ to each other, while responses evoked by different stimuli should be ‘far’ from each other), but can also be related to the classical approach under simplifying assumptions: If each response is represented as a point in multi-dimensional space and the distribution of the responses evoked by repetitions of the same stimulus within that space is assumed to be Gaussian, then the log likelihood that a response was evoked by a particular stimulus (assuming that all stimuli are equally likely) is proportional to the square of its distance from the average of all responses evoked by that stimulus.¹

Previous attempts to extend decoding with distance metrics to population responses have focused on the extent to which decoding performance is dependent on cell identity, i.e. whether performance differs if all spikes are assumed to come from a single neuron ([Aronov, Reich, Mechler, & Victor, 2003](#); [Houghton & Sen, 2008](#)). In this study, we investigate how varying the influence, or weight, of each cell affects population decoding performance. The choice of weight for each cell is a complex problem, and should be based not only on how informative the response of each cell is individually, but also on the correlations between the responses of each cell and the others in the population. This problem can be illustrated through a simple example of averaging: For a series

* Corresponding address: Ear Institute, University College London, London W1CX 8EE, UK.

E-mail address: n.lesica@ucl.ac.uk (A. Lesica).

¹ Note that, under these assumptions, the log likelihood is proportional to the square of the distance only when the distance metric is Euclidean. However, for some non-Euclidean metrics, a similar relationship may be achievable by introducing a different exponent ([Aronov & Victor, 2004](#)).

of measurements in which the noise in each measurement is independent and of equal magnitude, standard averaging yields the optimal estimate of the underlying signal. If the magnitude of the noise varies across measurements, then some measurements will be more reliable than others, and a weighted average based on this reliability will yield the optimal estimate. However, if the noise in a fraction of the measurements is correlated, then averaging across those measurements will be less effective in reducing the noise (in the extreme of identical noise, averaging has no effect) and a weighted average that favors the uncorrelated measurements may provide the optimal estimate, even if the correlated measurements are individually more reliable. As with averaging, the optimal weights for population decoding with distance metrics are dependent on both the individual reliability of cells in the population and the correlations between them. In this study, we describe an algorithm to find these optimal weights and demonstrate its utility by decoding experimental responses with a variety of correlation structures.

2. Decoding with distance metrics

We define the set of responses from cell p in response to I trials of S different stimuli as r^{pSI} , where $\mathbf{S} = \{1, 2, \dots, S\}$ and $\mathbf{I} = \{1, 2, \dots, I\}$. To decode the response evoked by trial i of stimulus s , r^{pSi} , we remove it from the set and infer which stimulus evoked it, $\hat{s}(r^{pSi})$. Assuming we have a metric for quantifying the distance between two responses, $d(r^{pSi}, r^{pS'i'})$, then we can determine the average distance from r^{pSi} to the responses evoked by all trials of a given stimulus s' , $\bar{d}_s(r^{pSi}) = \left\langle d(r^{pSi}, r^{pS'i'}) \right\rangle_{i'}$, with trial i excluded to avoid overfitting when $s = s'$, and choose the stimulus for which this average distance is minimal, $\hat{s}(r^{pSi}) = \arg \min_{s' \in \mathbf{S}} \bar{d}_{s'}(r^{pSi})$ [note that in the equation for \bar{d} , an exponent can be introduced inside the expectation to bias the result toward larger or smaller values]. This approach is easily extended to decode the responses r^{PSi} from a population of cells $\mathbf{P} = \{1, 2, \dots, P\}$. To decode the responses from the population of cells \mathbf{P} in response to trial i of stimulus s , r^{PSi} , we choose the stimulus for which a weighted sum of the average distances for each cell is minimal, $\hat{s}(r^{PSi}) = \arg \min_{s' \in \mathbf{S}} \sum_{p \in \mathbf{P}} w^p \bar{d}_{s'}(r^{pSi})$. The central question in this study is how to choose the weights $w = [w^1, w^2, \dots, w^P]$ so as to maximize decoding performance.

2.1. Optimization of decoder weights

After decoding the spike trains for every trial of every stimulus as described above, we measure overall performance as the percent of spike trains that were correctly decoded and denote this quantity as PC_p for a single cell p , and $PC_p(w)$ for the population \mathbf{P} with weights w . The standard approach to finding the optimal set of weights, i.e. the set of weights that maximize $PC_p(w)$, is to calculate the gradient $dPC_p(w)/dw$ and use it as a guide toward a local, and hopefully global, maximum. However, for the particular problem considered here, analytical specification of the gradient was not possible and algorithms that calculated the gradient numerically performed very poorly. Fortunately, there is another class of algorithms known as ‘evolutionary’ that do not require knowledge of the gradient. These algorithms operate iteratively, choosing the best of several candidate solutions on each iteration until performance saturates. While there are many evolutionary algorithms that may be suitable for this particular problem, we chose to implement two of the most common, genetic and particle swarm. As illustrated in the examples below, the performance of these algorithms was similar. However, the genetic algorithm was superior in that it required less computation time and is easily implemented via the Genetic Algorithm and Direct Search Toolbox

in Matlab (The Mathworks, USA), and, thus, we describe only its implementation in detail here. Details of the particle swarm algorithm can be found in Kennedy, Eberhart, and Shi (2001).

The genetic algorithm for optimization begins by creating a population of y random vectors of length p , drawn with uniform probability from the interval $[0, 1]$, and computing $PC_p(w)$ for each vector. Next, the population is evolved in three steps: First, e ‘elite’ vectors, those with the highest $PC_p(w)$, are moved to the next generation. Next, x ‘crossover’ vectors are created by random recombination between two ‘parent’ vectors from the current population, with the probability of a particular vector being chosen as a parent proportional to its $PC_p(w)$. Finally, u ‘mutant’ vectors are created by adding random noise $n \sim \mathcal{N}(0, \sigma)$ to a parent vector, with parent vectors chosen as above. The standard deviation of the noise $\sigma = 1$ for the first generation and is decreased linearly with each successive generation such that $\sigma = 0$ if the algorithm runs to completion. The algorithm stops after either completing V evolution generations or when the change in the highest $PC_p(w)$ over the past G generations is less than ε . The set of weights with the highest $PC_p(w)$ after the completion of the algorithm is denoted w_{genetic} . For the examples in this study, $y = 25$, $e = 2$, $x = 18$, $u = 5$, $V = 100$, $G = 25$, and $\varepsilon = 10^{-5}$, in accordance with the suggested default parameters for the *ga* function in Matlab.

As a baseline for comparison with w_{genetic} , we also computed the optimal weights w_{swarm} via particle swarm optimization, and used two other simple weighting schemes: $w_{\text{equal}} = 1$, where all cells are weighted equally, and $w_{\text{percorr}} = [PC_1, PC_2, \dots, PC_p]$, where the weights are determined by the overall performance of each cell when its responses are decoded individually. To prevent overfitting, it is important to exclude the responses to be decoded when optimizing the weights. For all optimizations, we split the responses into successive training sets (95% of responses) and testing sets (5% of responses) such that all responses were included in the testing set exactly once.

3. Decoding experimental spike trains

To illustrate the utility of the optimization algorithm under experimental conditions, we decoded spike trains from $P = 34$ cells recorded in the inferior colliculus of anesthetized gerbils in response to the presentation of $I = 20$ trials of $S = 8$ different sounds (different instances of Gaussian white noise). The details of the experimental procedures can be found in Lesica and Grothe (2008a, 2008b).

We decoded the spike trains using the metric introduced by Victor and colleagues (Victor & Purpura, 1996), which measures the distance between two spike trains as the overall cost of the set of operations required to transform one spike train into the other, with possible operations including the insertion of a spike, the deletion of a spike, and the time-shift of a spike (software available at <http://neuroanalysis.org/toolkit>). By changing the cost of time-shifting a spike relative to deleting the spike at one time and inserting it at another, the metric can be used to evaluate the distance between spike trains at different timescales. The details of the implementation of the metric are not given here, but can be found in Aronov (2003) and Victor and Purpura (1996).

A 10 ms segment of the set of spike trains r^{PSi} for a typical cell is shown in Fig. 1(a) along with the decoding performance for the individual responses of each cell as a function of the response timescale parameter of the decoder (decoded responses were 100 ms in duration). The median best timescale, i.e. the timescale that yielded the best decoding performance, across the sample of cells was 2 ms (black arrow) and, for simplicity, we fixed the response timescale at this value for all decoding of these responses.

The distribution of significant pairwise correlation coefficients for the population (computed at a timescale of 2 ms) is shown in Fig. 1(b). The total correlation (ρ_{total}) was significant between approximately half of the cell pairs (262 of 561). Correlations

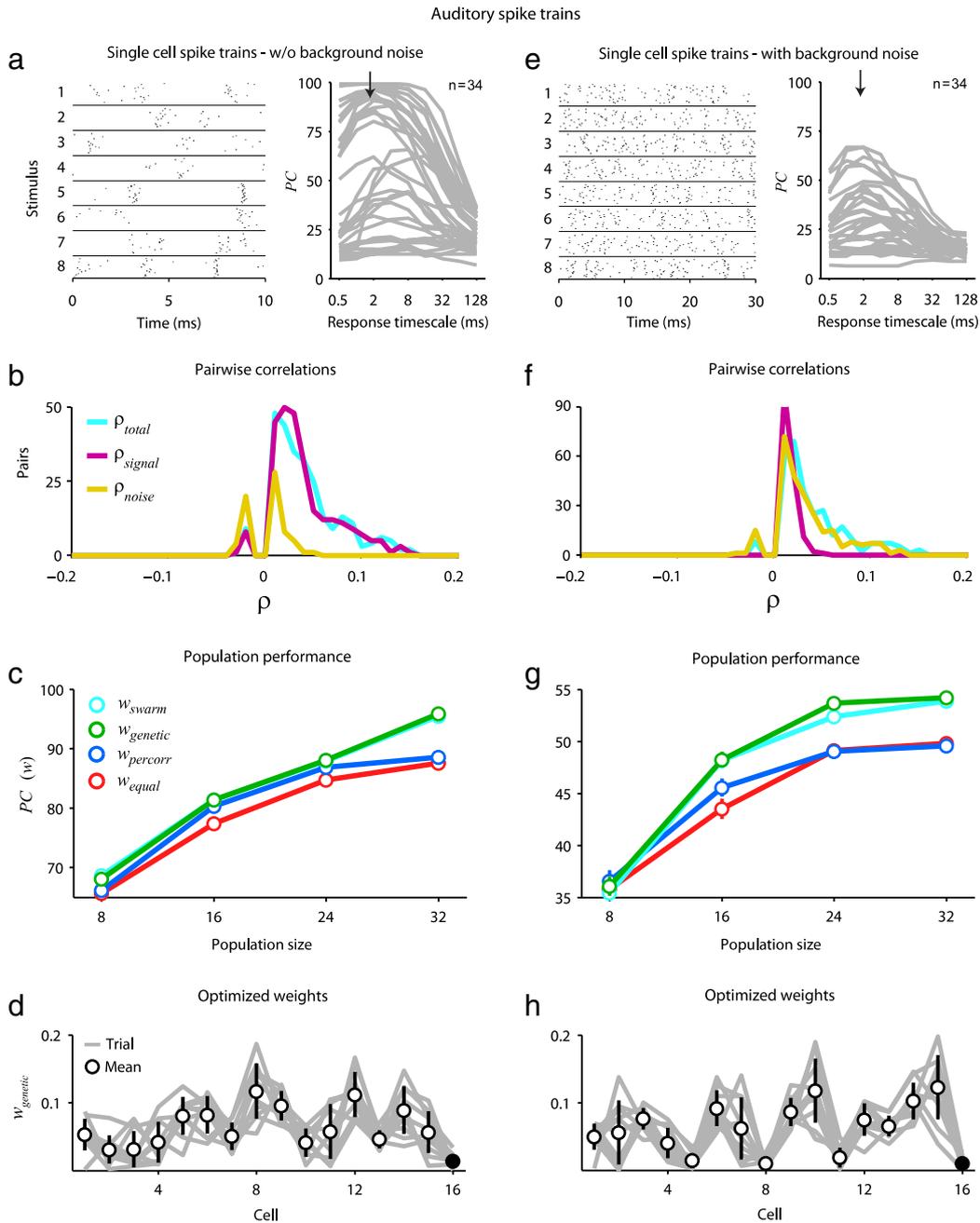


Fig. 1. Decoding population spike trains. (a) A raster plot showing the spike trains of a neuron in the inferior colliculus of an anesthetized gerbil in response to $I = 20$ trials of $S = 8$ different sounds and decoder performance (percent correct, PC) as a function of decoder response timescale for the responses of each individual cell. The black arrow indicates the population median best timescale. (b) Histograms of the correlation coefficients between pairs of cells ($P = 34$). Correlation coefficients were estimated after converting the spike trains to binary vectors with a temporal resolution of 2 ms. Only significant correlations ($p < 0.05$) are shown. The total correlation (ρ_{total}) was computed directly from the responses, the signal correlation (ρ_{signal}) was computed from the responses after randomizing the trial order, and the noise correlation (ρ_{noise}) was computed as the difference between ρ_{total} and ρ_{signal} . (c) The decoder performance $PC_P(w)$ for subpopulations of increasing size for four sets of weights: w_{equal} (all weights equal), $w_{percorr}$ (weights based on individual performance), $w_{genetic}$ (weights optimized with genetic algorithm), and w_{swarm} (weights optimized with particle swarm algorithm). The circles and bars indicate the mean and standard error of the performance for 100 different random subpopulations. (d) The weights $w_{genetic}$ resulting from 10 different optimizations with random initial values in the population y for a particular subpopulation of 16 cells. The lines indicate the weights for each individual optimization and the circles and bars indicate the mean and standard deviation. The dummy cell is indicated by the filled black circle. (e)–(h) Results for a second set of responses of the same cells to the same sounds, with a different random background noise added to each sound on each trial, presented as in (a)–(d).

between cells can have both signal and noise components: signal correlations arise from correlations in the stimulus itself and/or similarity in preferred stimulus features (frequency, orientation, etc.), while noise correlations arise from shared inputs that contribute to the trial to trial variability in responses. In this set of responses, most of the total correlation was due to signal correlations (in fact, because most of the cells were not recorded simultaneously, the population is expected to contain few noise correlations).

The decoder performance $PC_P(w)$ for subpopulations of increasing size is shown in Fig. 1(c) for four sets of weights: w_{equal} (all weights equal), $w_{percorr}$ (weights based on individual performance), $w_{genetic}$ (weights optimized with genetic algorithm), and w_{swarm} (weights optimized with particle swarm algorithm). For population decoding, only 10 ms segments of the responses were used in order to increase the difficulty of the decoding task. While the decoder performance was similar for all sets of weights for small population sizes, genetic and particle swarm optimization

provided a performance increase of approximately 10% for large populations.

To determine whether optimization produced global optima, we used the genetic algorithm to find the optimal weights for a given set of responses using 10 different initial populations y . In each set of responses, we also included a 'dummy cell' for which the stimulus identity associated with each response was randomized. As illustrated in Fig. 1(d) for a particular subpopulation of $P = 16$ cells, the genetic algorithm converges to approximately the same set of optimal weights w_{genetic} independent of the initial values in the population y and the weights associated with the dummy cell (filled black circle) were always near zero. These results suggest that the set of weights produced by the genetic algorithm is indeed the global optimum and that the algorithm is successful in minimizing the contribution of uninformative cells.

For the same cells, we also recorded responses to the same sounds in the presence of background noise. Because a different background noise was added on each trial, and this noise was the same for all cells, the background noise served to reduce signal correlations and introduce noise correlations. The set of spike trains r^{PSI} for a typical cell is shown in Fig. 1(e) along with the decoding performance for the individual responses of each cell as a function of the response timescale (decoded responses were 100 ms in duration). While the background noise resulted in considerable variability in the spike trains evoked by the same stimulus, the median best timescale was again 2 ms and we fixed the response timescale at this value for all decoding of these responses.

The distribution of significant pairwise correlation coefficients for the population is shown in Fig. 1(f). Again, the total correlation (ρ_{total}) was significant between approximately half of the cell pairs (299 of 561), but for this set of responses, most of the total correlation was due to noise correlations.

The decoder performance for the population $PC_p(w)$ for subpopulations of increasing size is shown in Fig. 1(g) for w_{equal} , w_{percorr} , and w_{genetic} , w_{swarm} . For population decoding, only 30 ms segments of the responses were used in order to increase the difficulty of the decoding task. As in the previous example, performance was similar for all sets of weights for small population sizes, but optimization provided a substantial performance increase for large populations and, as shown in Fig. 1(h), optimizations with different initial values in the population y produced similar sets of weights with values near zero for the dummy cell. Taken together, the results in Fig. 1 demonstrate that the optimization algorithm was effective for decoding population spike trains under experimental conditions when the responses contained both signal and noise correlations.

4. Decoding experimental calcium signals

To further illustrate the utility of optimization under experimental conditions, we also decoded calcium signals (relative change in indicator fluorescence) from $P = 37$ cells recorded in the visual cortex of anesthetized mice in response to the presentation of $I = 18$ trials of $S = 8$ different oriented sinusoidal gratings (each grating was displayed for 2 s at 50% contrast and drifted at a rate of 2 Hz; calcium signals were sampled at 15 Hz). The details of the experimental procedures can be found in Mrcsic-Flogel et al. (2007).

The set of calcium signals r^{PSI} for a typical cell is shown in Fig. 2(a). The top image gives an overview of the dynamics and reproducibility of the signals as the orientation of the grating changed (the order of the orientations was the same on each trial), while the lower plots show the signals in detail for two particular orientations. The distribution of significant pairwise correlation coefficients for the population (computed at

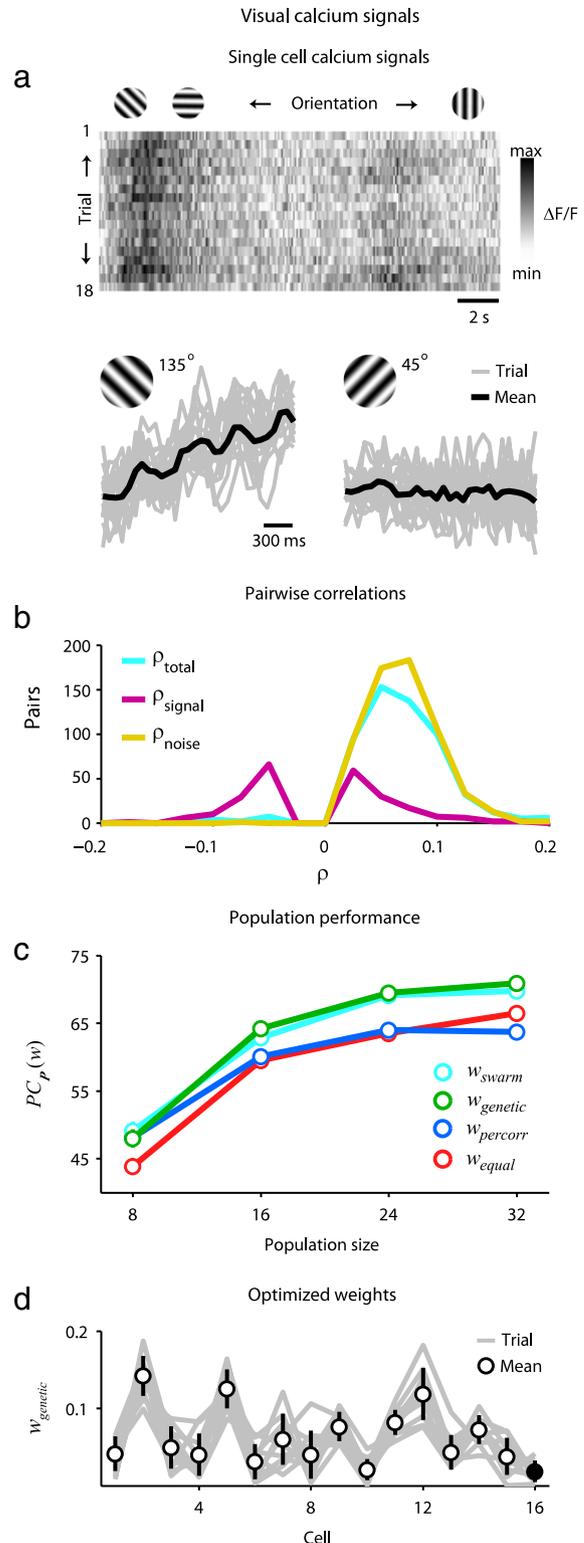


Fig. 2. Decoding population calcium signals. (a) Top: an image showing the calcium signal (relative change in indicator fluorescence) of a neuron in the visual cortex of an anesthetized mouse in response to $I = 18$ trials of $S = 8$ different oriented gratings. Bottom: the calcium signals for the same neuron in response to two particular orientations. Gray lines indicate the signal for each trial and the black line indicates the mean. (b)–(d) Histograms of the correlation coefficients between pairs of cells ($P = 37$), decoder performance $PC_p(w)$ for subpopulations of increasing size for four sets of weights: w_{equal} , w_{percorr} , w_{genetic} , and w_{swarm} , and the weights w_{genetic} resulting from 10 different optimizations with random initial values in the population y for a particular subpopulation of 16 cells, presented as in Fig. 1.

a timescale of 66 ms) is shown in Fig. 2(b). The total correlation (ρ_{total}) was significant between most of the cell pairs (560 of 666) and contained both signal and noise components (all cells were recorded simultaneously).

We decoded the calcium signals using the Euclidean distance metric $d(r^{psi}, r^{psi'}) = |r^{psi} - r^{psi'}|$. The decoder performance $PC_p(w)$ for subpopulations of increasing size is shown in Fig. 2(c) for w_{equal} , w_{percorr} , and w_{genetic} , w_{swarm} . As with spike trains, performance was similar for all sets of weights for small populations, but optimization improved performance for large populations and, as shown in Fig. 2(d), optimizations with different initial values in the population y produced similar sets of weights with values near zero for the dummy cell. These results demonstrate that the optimization algorithm was effective for decoding not only population spike trains, but also population calcium signals with signal and noise correlations.

5. Conclusions

We have demonstrated that when decoding population spike trains and calcium signals using distance metrics, optimization of the influence of each cell on the overall result can provide an increase in performance relative to simple weighting schemes. Our results demonstrate that for populations of cells in the auditory and visual systems with a variety of signal and noise correlations, the benefit of genetic optimization can be relatively large (up to 10%).

Our results demonstrate that the optimal weights for population decoding cannot be derived simply from the performance of each cell as an individual, suggesting that there may be a relationship between the optimal weights for decoder performance and the correlations between cells in the population. One interesting avenue for further research would be to characterize this relationship, i.e. to explicitly describe the impact of correlations on the optimal weighting scheme when decoding population responses. This relationship also suggests the potential of optimization as a tool for measuring the contribution of individual cells to the population code. For example, as the stimulus and/or correlations in the population change, the corresponding changes in the optimal

weights for different cells or groups of cells could be used to assay the change in the distribution of information across the population.

Acknowledgements

The authors thank Jonathan Victor and Christian Leibold for helpful discussions, and Leila Khouri, Abhinav Singh, and Dmitry Lyamzin for comments on the manuscript. This work was supported by the Bernstein Center for Computational Neuroscience (BG and NL), the Deutsche Forschungsgemeinschaft (NL), the ERASMUS Program (DH), the Wellcome Trust (TMF and NL), and the EMBO Long-term Fellowship (SH).

References

- Aronov, D. (2003). Fast algorithm for the metric-space analysis of simultaneous responses of multiple single neurons. *Journal of Neuroscience Methods*, 124(2), 175–179.
- Aronov, D., Reich, D. S., Mechler, F., & Victor, J. D. (2003). Neural coding of spatial phase in V1 of the macaque monkey. *Journal of Neurophysiology*, 89(6), 3304–3327.
- Aronov, D., & Victor, J. D. (2004). Non-Euclidean properties of spike train metric spaces. *Physical Review E. Statistical, Nonlinear, and Soft Matter Physics*, 69(6 Pt 1), 061905–061905.
- Houghton, C., & Sen, K. (2008). A new multineuron spike train metric. *Neural Computation*, 20(6), 1495–1511.
- Kennedy, J., Eberhart, R., & Shi, Y. (2001). *Swarm intelligence*. Morgan Kaufmann Publishers.
- Lesica, N. A., & Grothe, B. (2008a). Dynamic spectrotemporal feature selectivity in the auditory midbrain. *Journal of Neuroscience*, 28(21), 5412–5421.
- Lesica, N. A., & Grothe, B. (2008b). Efficient temporal processing of naturalistic sounds. *PLoS One*, 3(2), e1655.
- Mrsic-Flogel, T. D., Hofer, S. B., Ohki, K., Reid, R. C., Bonhoeffer, T., & Hubener, M. (2007). Homeostatic regulation of eye-specific responses in visual cortex during ocular dominance plasticity. *Neuron*, 54(6), 961–972.
- Quiari Quiroga, R., & Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3), 173–185.
- van Rossum, M. C. (2001). A novel spike distance. *Neural Computation*, 13(4), 751–763.
- Victor, J. D. (2005). Spike train metrics. *Current Opinion in Neurobiology*, 15(5), 585–592.
- Victor, J. D., & Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of Neurophysiology*, 76(2), 1310–1326.