


RESEARCH

Open Access



Estimating Fisher discriminant error in a linear integrator model of neural population activity

Matias Calderini¹ and Jean-Philippe Thivierge^{1,2*} 

*Correspondence:
jthivier@uottawa.ca

¹School of Psychology, University of Ottawa, 136 Jean Jacques Lussier, Ottawa, ON K1N 6N5, Canada

²Brain and Mind Research Institute, University of Ottawa, Ottawa, ON K1N 6N5, Canada

Abstract

Decoding approaches provide a useful means of estimating the information contained in neuronal circuits. In this work, we analyze the expected classification error of a decoder based on Fisher linear discriminant analysis. We provide expressions that relate decoding error to the specific parameters of a population model that performs linear integration of sensory input. Results show conditions that lead to beneficial and detrimental effects of noise correlation on decoding. Further, the proposed framework sheds light on the contribution of neuronal noise, highlighting cases where, counter-intuitively, increased noise may lead to improved decoding performance. Finally, we examined the impact of dynamical parameters, including neuronal leak and integration time constant, on decoding. Overall, this work presents a fruitful approach to the study of decoding using a comprehensive theoretical framework that merges dynamical parameters with estimates of readout error.

Keywords: Linear model; Fisher linear discriminant analysis; Noise correlation

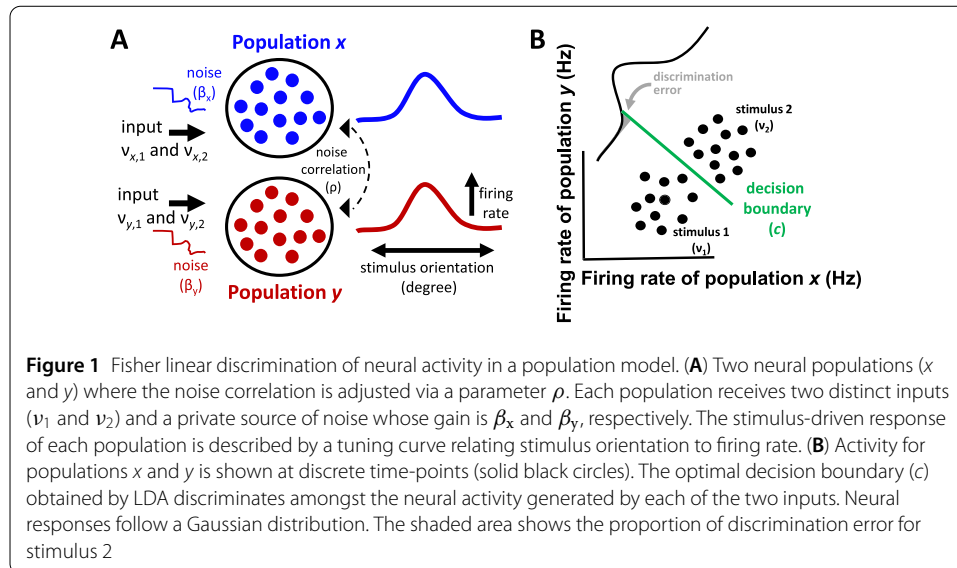
1 Introduction

In recent years, neuronal decoding has emerged as a key aspect of understanding the neural code [1]. The aim of decoding algorithms is to read out the sensory-driven responses of a neuronal population and classify them following a given criterion. Popular criteria include Fisher information [2, 3], mutual information [4], and machine learning approaches [5, 6]. While many types of decoders exist [7], a linear readout of neural activity has often been employed to perform sensory classification [8, 9] and predict motor decisions [10, 11]. Further, different classes of linear readouts are amenable to mathematical analysis and capture biological learning rules such as Hebbian learning [12].

In this work, we formally analyze the optimal decoding error of a linear decoder based on Fisher linear discriminant analysis (LDA). Assuming discrete classes of stimuli, LDA provides an upper bound on linear decoding capacity [13]. In addition, LDA shows good agreement with decision-making tasks and offers a bridge between cortical activity and behavioral performance [14, 15].

Importantly, most theoretical approaches based on neural decoding are not concerned with how linear decoders would be influenced by specific dynamical parameters of mod-

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



eled neural systems [16]. Here, we address this concern by providing expressions that relate decoding error to the adjustable parameters of a rate-based population model with a linear neural integrator [17, 18]. This model captures the average spiking activity of neuronal populations [19–21] and the quasi-linear responses of neurons found in many experimental contexts [22]. Preliminary results have been presented in previous work [13, 14], yet the full analytical solution had remained incomplete and limited to positive noise correlation; we now present the complete solution.

The framework relies on the simplifying assumption that signal and noise correlations originate from independent sources. While this assumption does not hold in biological circuits, where signal and noise are related [1], it allows us to systematically explore a wide range of scenarios that describe the impact of neuronal inputs, noise, correlations, and dynamical parameters on linear decoding, where the contribution of each parameter can be examined independently.

This paper begins by describing the neural integrator model and the LDA readout. Then, we provide expressions for LDA error that rely on the parameters of the integrator model. Finally, we consider the effect of correlation, noise, and dynamical parameters on neuronal decoding using both analytical expressions and numerical simulations.

2 Linear population model

As a starting point, we assume two independent neuronal populations, each projecting in a feedforward manner to a readout discriminating amongst two inputs, v_1 and v_2 , that are constant over time (Fig. 1(A)). Each population's mean firing rate in response to stimuli is conceptualized by a tuning curve where a stimulus feature, for instance visual orientation, generates a graded response. This scenario is analogous to analyses that examine population responses after performing a dimensionality reduction to generate a "population tuning curve" [23]. While a more complex model could account for a heterogeneity of responses within each population, we choose to limit our model to two homogeneous populations in order for the classification problem to remain tractable.

The activity of each population is described by a linear neural integrator

$$\begin{aligned} \tau_x \frac{dx_i}{dt} &= -\alpha_x x_i + v_{i,x} + \beta_x \xi_x(t), \\ \tau_y \frac{dy_i}{dt} &= -\alpha_y y_i + v_{i,y} + \beta_y \xi_y(t), \end{aligned} \tag{1}$$

where x_i and y_i are the firing rates of each population in response to a given stimulus i , τ is a time constant, α is a leak term, $\xi(t)$ is Gaussian white noise ($\mathcal{N}(0, 1)$), and β is the gain of the noise. Network parameters τ , α , and β are bound to $\mathbb{R}_{>0}$. We make no distinction between noise induced by stimuli and noise generated through intrinsic neural activity. While their effect on mean rate activity is similar [24], their impact on noise correlations differs [1]; in the model, we explicitly separate the effect of firing rate and noise correlation. This will be done by controlling noise correlation through a tunable parameter, as detailed in Sect. 4. An advantage of this formalism is that the effect of noise correlation can be systematically isolated from changes in firing rates and signal correlation that would be induced through reciprocal connections between the two populations. Further, depending on the choice of parameters, the addition of recurrent weights to Eq. (1) may prevent the system from reaching a stationary state, which is a fundamental assumption of LDA.

3 Fisher linear discriminant decoder

A linear decoder based on LDA reads out the activity of the population model in order to perform a binary discrimination (Fig. 1(B)). Discrimination error generated by LDA provides an estimate of the statistical confidence in distinguishing pairs of stimuli based on network activity. We focus on pairwise discrimination given that error rates obtained from more than two stimuli are well approximated by values obtained from all combinations of pairwise comparisons [25].

LDA assumes that neural activity is sampled from a multivariate Gaussian distribution with class covariance matrix Σ_i and class mean vector μ_i . Further, LDA assumes equal class covariance, therefore $\Sigma_1 = \Sigma_2 = \Sigma$. LDA attempts to find a projection line w , perpendicular to the decision boundary, onto which the input space is projected. The optimal projection line maximizes the Fisher criterion $J(w)$ defined as the ratio of the projected between- to within-class variance:

$$J(w) = \frac{w \cdot (\mu_2 - \mu_1)^2}{w^T \cdot \Sigma_W \cdot w}.$$

Given the assumption of equal class covariance, we set $\Sigma_W = 2\Sigma$. By taking the derivative of $J(w)$ with respect to w and setting it to zero, one finds the closed-form solution for the optimal projection line to be

$$W = (2\Sigma)^{-1}(\mu_2 - \mu_1). \tag{2}$$

4 Formulating a model-based linear decoder

To analytically derive means (μ_1 and μ_2) and covariance (Σ) from the neural population model, we rearrange Eq. (1) as follows, using population x as example:

$$dx_i = \frac{\alpha_x}{\tau_x} \left(\frac{v_{ix}}{\alpha_x} - x_i \right) dt + \frac{\beta_x}{\tau_x} \xi_x(t) dt. \tag{3}$$

Given that a white noise process is by definition the time derivative of a Weiner process, $\xi(t) = dW_t/dt$, we can rewrite Eq. (3) as

$$dx_i = \theta_x(\mu_{ix} - x_i) dt + \lambda_x dW_{x,t}, \tag{4}$$

with $\theta_x = \alpha_x/\tau_x$, $\mu_{ix} = v_{ix}/\alpha_x$, and $\lambda_x = \beta_x/\tau_x$. Equation (4) is an Orstein–Uhlenbeck process with known solution

$$x_i(t) = \mu_{ix} + (x_{i0} - \mu_{ix})e^{-\theta_x t} + \lambda \int_0^t e^{-\theta_x(t-s)} dB(s). \tag{5}$$

Equation (5) is a mean reverting process whose stable state follows a Gaussian distribution. A full derivation of this process is found in Sections A.1–A.2. To summarize this derivation, the expected mean and variance are

$$E[x_i(t)] = \mu_{ix} + (x_{i0} - \mu_{ix})e^{-\theta_x t},$$

$$\text{Var}(x_i(t)) = \frac{\lambda_x^2}{2\theta_x} (1 - e^{-2\theta_x t}).$$

The stationary mean and variance of Eq. (5) are

$$\lim_{t \rightarrow \infty} E[x_i] = \mu_{ix} = \frac{v_{x,i}}{\alpha_x},$$

$$\lim_{t \rightarrow \infty} \text{Var}(x_i) = \frac{\lambda_x^2}{2\theta_x} = \frac{\beta_x^2}{2\tau_x \alpha_x} = \sigma^2.$$

With the assumption that the mean of x is much larger than the variance, there is negligible probability that x would fall below zero. Imposing strictly positive values of x could be achieved by the addition of a constant and would not alter the results obtained from the linear classifier.

The readout of neural activity depends on the following feature space:

$$Z \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}),$$

$$\boldsymbol{\mu}_i = [\mu_{xi}, \mu_{yi}]^T,$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix},$$

where Z is obtained from the probability distribution of a multivariate Gaussian with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}$. Setting the parameter $\rho = 0$ would be equivalent to a so-called “diagonal decoder” where off-diagonal elements of the covariance matrix are neglected, thus ignoring noise correlations altogether [16].

The closed form solution of LDA (Eq. (2)) can be expressed using the parameters of the population model (Eq. (1)) as follows. First, the total within-class scatter S_w is

$$S_w = 2\Sigma$$

$$= 2 \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

$$= 2 \begin{bmatrix} \frac{\beta_x^2}{2\tau_x\alpha_x} & \rho\sqrt{\frac{\beta_x^2}{2\tau_x\alpha_x} \frac{\beta_y^2}{2\tau_y\alpha_y}} \\ \rho\sqrt{\frac{\beta_x^2}{2\tau_x\alpha_x} \frac{\beta_y^2}{2\tau_y\alpha_y}} & \frac{\beta_y^2}{2\tau_y\alpha_y} \end{bmatrix}.$$

To alleviate the notation, we define $\Delta\boldsymbol{\mu}^T = [\Delta\mu_x, \Delta\mu_y]^T = \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1$, where $\Delta\mu_u = \Delta v_u/\alpha_u$, and Δv_u is the absolute difference between the two stimuli, given an index u that stands for either population x or y . In this way, Eq. (2) becomes

$$\begin{aligned} W &= (2\Sigma)^{-1} \Delta\boldsymbol{\mu} \\ &= \frac{1}{2(1-\rho^2)} \begin{bmatrix} \frac{1}{\sigma_x^2} \Delta\mu_x - \rho \frac{1}{\sqrt{\sigma_x^2\sigma_y^2}} \Delta\mu_y \\ \frac{1}{\sigma_y^2} \Delta\mu_y - \rho \frac{1}{\sqrt{\sigma_x^2\sigma_y^2}} \Delta\mu_x \end{bmatrix} \\ &= \frac{1}{2(1-\rho^2)} \begin{bmatrix} \frac{\tau_x}{\beta_x^2} \Delta v_x - \rho \sqrt{\frac{\tau_x\alpha_x}{\beta_x^2} \frac{\tau_y}{\beta_y^2\alpha_y}} \Delta v_y \\ \frac{\tau_y}{\beta_y^2} \Delta v_y - \rho \sqrt{\frac{\tau_y\alpha_y}{\beta_y^2} \frac{\tau_x}{\beta_x^2\alpha_x}} \Delta v_x \end{bmatrix}. \end{aligned}$$

From the law of total probability, the error rate of classification is given by

$$\varepsilon = P[y = 0|k = 1]P[k = 1] + P[y = 1|k = 0]P[k = 0], \tag{6}$$

where $P[k = 1]$ is the probability that a randomly sampled point from any distribution belongs to class j and $P[y = i|k = j]$ is the probability that a point is classified as belonging to class i when it belongs to class j . Given that the classifier is unbiased towards each of the two neural populations, $P[k = 0] = P[k = 1] = 0.5$. To calculate conditional probabilities $P[y = i|k = j]$, one must define a threshold c that serves as a boundary between the two distributions. The value of c is chosen to be the midpoint between the means of the projected distributions.

We calculate $P[y = i|k = j]$ as the area under the curve of the density function for j in the region where i is the correct class. As a first step, we shift the projected distributions by a factor c , so that the threshold becomes zero to simplify the integration. More specifically, the unshifted threshold c , the means of the shifted distributions η_i , and their variance ζ^2 are

$$c = W \cdot \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) + b,$$

$$\eta_i = W \cdot \boldsymbol{\mu}_i + b - c,$$

$$\zeta^2 = W^T \Sigma W,$$

with bias term b . The error rate from Eq. (6) then becomes

$$\varepsilon = \frac{1}{2} \int_{-\infty}^0 \frac{1}{\sqrt{2\zeta^2\pi}} e^{-\frac{(w-\eta_1)^2}{2\zeta^2}} dw + \frac{1}{2} \int_0^{\infty} \frac{1}{\sqrt{2\zeta^2\pi}} e^{-\frac{(w-\eta_0)^2}{2\zeta^2}} dw.$$

Details of the full integration of error can be found in Section A.3. The final expression is

$$\varepsilon = \frac{1}{2} \operatorname{erfc} \left(\frac{\eta_1}{\sqrt{2\zeta^2}} \right).$$

This expression is further simplified by introducing the squared Mahalanobis distance d^2

$$\varepsilon = \frac{1}{2} \operatorname{erfc}\left(\frac{1}{2\sqrt{2}}\sqrt{d^2}\right), \tag{7}$$

where

$$d^2 = \Delta\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \Delta\boldsymbol{\mu}. \tag{8}$$

Because of equal class covariance, the above expression has the property that

$$d(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0) = d.$$

Using Eq. (8), we rewrite d^2 from the network parameters:

$$\begin{aligned} d^2 &= \frac{1}{1-\rho^2} \left[\frac{1}{\sigma_x^2} \Delta\mu_x^2 + \frac{1}{\sigma_y^2} \Delta\mu_y^2 - 2\rho \frac{1}{\sqrt{\sigma_x^2 \sigma_y^2}} \Delta\mu_x \Delta\mu_y \right], \\ &= \frac{2}{1-\rho^2} \left[\frac{\tau_x}{\beta_x^2 \alpha_x} \Delta v_x^2 + \frac{\tau_y}{\beta_y^2 \alpha_y} \Delta v_y^2 - 2\rho \sqrt{\frac{\tau_x}{\beta_x^2 \alpha_x} \frac{\tau_y}{\beta_y^2 \alpha_y}} \Delta v_x \Delta v_y \right]. \end{aligned}$$

As the ratio $\Delta\mu_u / \sqrt{\sigma_u^2}$ appears often in the above solution, we simplify our notation by introducing

$$r_u = \frac{\Delta\mu_u}{\sqrt{\sigma_u^2}} = \Delta v_u \sqrt{\frac{\tau_u}{\beta_u^2 \alpha_u}}.$$

This expression simplifies the Mahalanobis distance to

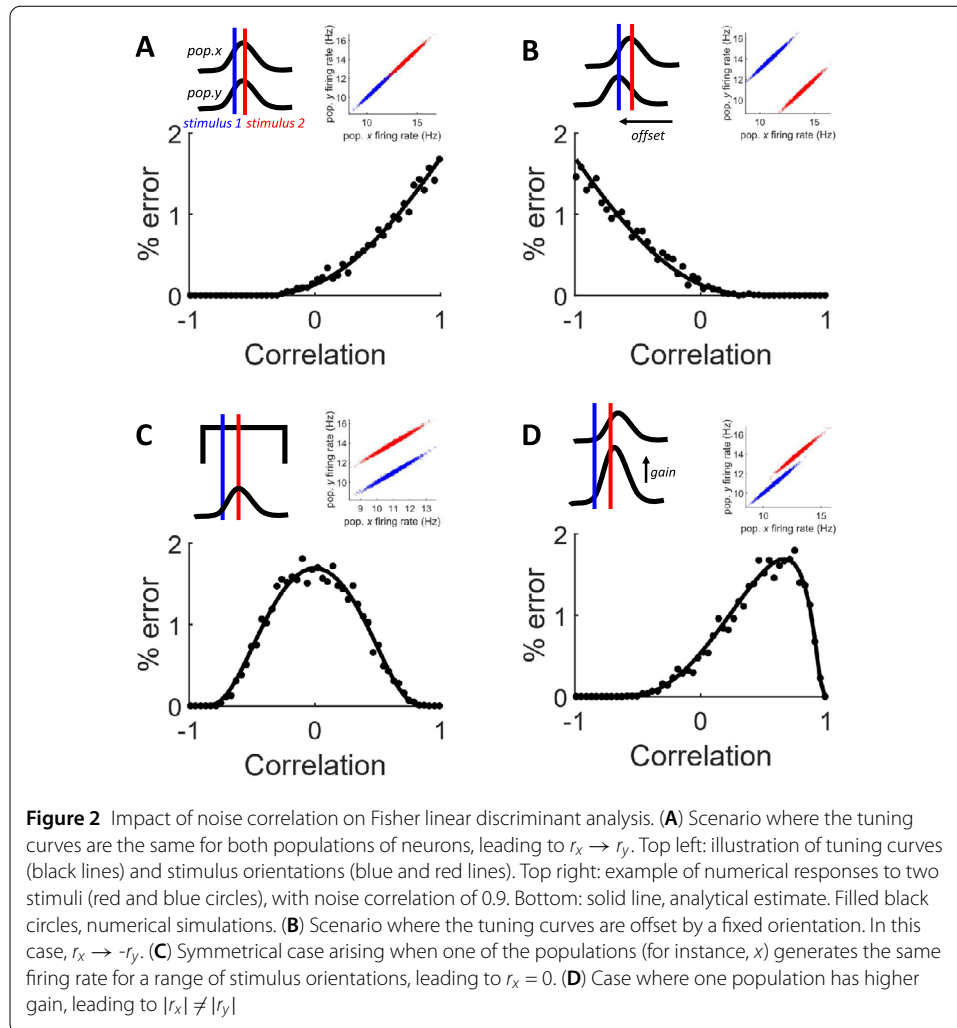
$$d^2 = \frac{1}{1-\rho^2} [r_x^2 + r_y^2 - 2\rho r_x r_y].$$

The full derivation of expected error using Mahalanobis distance is found in Sections A.3–A.4. The above analysis provides a relationship between classification error and the network parameters of the population model. In the sections to follow, we explore the various links between these quantities.

5 Noise correlation

Neurons that are in close physical proximity exhibit correlations in their activity. An extensive body of work has examined the impact of these noise correlations on behavioral tasks [26] and the activity of brain circuits [27–35]. Noise correlations may be advantageous or detrimental to cognitive and sensory processing; however, the specific network-level properties that give rise to these effects have not been fully elucidated.

In the proposed model, the effect of noise correlation on classification error is highly dependent upon the sensory inputs (v_1 and v_2). We distinguish four main cases that lead to qualitatively different conclusions on the impact of noise correlations. Details of these analyses are provided in Sections A.5–A.6.



A first case arises when the tuning curves of populations x and y are identical in terms of both their orientation preference and their gain (Fig. 2(A)). In this case, $r_x \rightarrow r_y$, leading to monotonically increasing error as a function of correlation. Intuitively, this happens because correlation forces the firing rate distributions to “stretch” towards each other. We verified the analytical solution by comparing it to numerical estimates of the error rate as a function of noise correlation. These numerical estimates were obtained with Eq. (1), where populations x and y both received inputs $v_1 = 11$ and $v_2 = 14$ in order for the model to mimick a scenario where the two populations have identical tuning properties. The goal here is not to capture the model’s response to a continuum of stimulus values along the tuning curves, but rather to illustrate the behavior of the model using discrete stimuli. We set $\tau = 1$, $\beta = 1$, and $\alpha = 1$ for both populations. We then numerically generated 5000 points per stimulus class. A subset of 80% of the total number of data points were randomly selected to train the LDA classifier. The proportion of misclassified points was calculated based on the remaining data points. We found good agreement between the numerical estimates and analytical solution (Fig. 2(A)).

Note that the range of error may be increased by moving the firing rate distributions closer to each other without altering the overall shape of the function relating error and

noise correlation. While the goal here was to show the distribution of readout error across a broad range of correlation values, we acknowledge that not all combinations of tuning curves and noise correlations are physiologically plausible. In fact, while noise correlations in cortex vary across experimental conditions, regions, and behavioral states, they are typically reported to be on the order of 0.1–0.3 for nearby cells [26]. Therefore, extreme values (both positive and negative) are unlikely to arise in living circuits.

In a second scenario, the two populations are offset in terms of their orientation preference (Fig. 2(B)). We examined classification error in this scenario by setting the input of population x to $v_1 = 11$ and $v_2 = 14$, while population y was set to $v_1 = 14$ and $v_2 = 11$. Analytically, this scenario leads to $r_x \rightarrow -r_y$, resulting in a monotonically decreasing error as correlation increases from -1 to 1 . Intuitively, this scenario arises because correlation stretches the distributions of responses along parallel lines, decreasing the overlap between them.

A third case arises when the tuning curve of one of the two populations yields the same response for two stimuli (Fig. 2(C)). This happens if the tuning curve of population x exhibits a broad region where firing rate remains constant despite changes in stimulus orientation. Analytically, this would lead to $r_x = 0$. We illustrate this scenario by setting $v_1 = 11$ and $v_2 = 11$ for population x , and $v_1 = 11$ and $v_2 = 14$ for population y . This case yields a “symmetrical” effect of correlation on readout error, where maximum error is found at $\rho_* = 0$ and error tends towards zero as ρ approaches either 1 or -1 .

Finally, a fourth scenario occurs when the two populations have tuning curves that are aligned in terms of orientation preference, but where one population has higher response gain (Fig. 2(D)). This case is defined by $|r_x| \neq |r_y|$. Error tends to zero as noise correlation (ρ) goes to either -1 or 1 . The correlation associated with maximum error is found somewhere in between these extremes and is given by

$$\rho_* = \frac{\min(r_x^2, r_y^2)}{r_x r_y}. \quad (9)$$

To illustrate this scenario, we set $v_1 = 11$ and $v_2 = 13$ for population x , and $v_1 = 11$ and $v_2 = 14$ for population y . Graphically, this scenario arises when noise correlation “stretches” the distribution of responses along parallel lines and their centroids do not align on either dimension. Starting from a correlation of zero, as correlation increases, the distributions will stretch towards each other, thus increasing overlap and error. After a maximum overlap defined by ρ_* , further stretching of the distributions will force them to spread too thinly for them to overlap, until the extreme case of a correlation of one, where both distributions would appear as perfectly parallel lines, leading to zero error.

A continuum of cases exists between the different scenarios illustrated in Fig. 2(A)–(D). For instance, the peak error (ρ_*) in Fig. 2(D) can shift to lower correlation values by offsetting one of the tuning curves, yielding a curve closer to Fig. 2(B).

In sum, the above results show that, depending upon the structure of the input delivered to the two neural populations, noise correlations produce widely different effects on classification error. While insights into these results can be obtained without the full formalism described here [34], such formalism becomes pivotal when examining the effect of specific network parameters, as described next.

6 Impact of noise gain on classification error

To explore the effect of network parameters on error, we first modify Eq. (9) as follows:

$$\rho_* = \frac{\min(r_x^2, r_y^2)}{r_x r_y} = \begin{cases} \frac{r_x}{r_y} & \text{if } |r_x| < |r_y|, \\ \frac{r_y}{r_x} & \text{if } |r_x| > |r_y|, \end{cases} \tag{10}$$

where the ratio r_x/r_y can be expressed using network parameters

$$\frac{r_x}{r_y} = \frac{\Delta v_x \beta_y \sqrt{\tau_x \alpha_y}}{\Delta v_y \beta_x \sqrt{\tau_y \alpha_x}}$$

We define a set containing all network parameters $G_u = \{\alpha_u, \tau_u, \beta_u, \Delta v_u\}$. If g is a subset of these parameters, we can manipulate them using a function $f(g)$ while setting the other parameters to a constant c_g . In this way, we can rewrite Eq. (10) as

$$\rho_* = \begin{cases} f(g)c_g & \text{if } |r_x| < |r_y|, \\ f(g)^{-1}c_g^{-1} & \text{if } |r_x| > |r_y|. \end{cases}$$

We can investigate the effect of network parameters on ρ_* . For example, the effect of noise gain (β_x and β_y) on ρ_* when keeping all other parameters constant except for the input is expressed as

$$\rho_* = f(\beta_y, \beta_x)c_{\beta_y, \beta_x} = \frac{\beta_y}{\beta_x} \left(\frac{\Delta v_x \sqrt{\tau_x \alpha_y}}{\Delta v_y \sqrt{\tau_y \alpha_x}} \right)$$

for $|r_x| < |r_y|$.

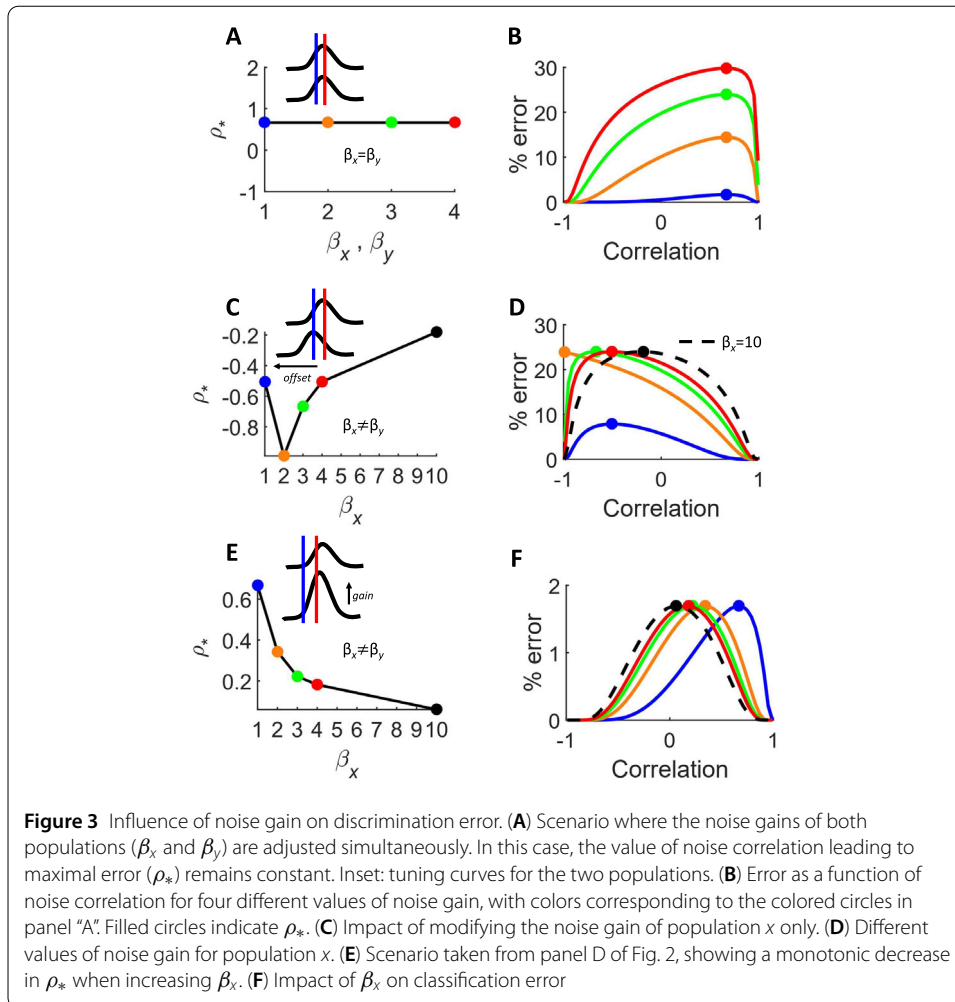
For illustration purposes, we explored the scenario described in Fig. 2(A), where two populations have equivalent tuning properties. Keeping all parameters constant while altering both β_x and β_y simultaneously has no effect on ρ_* (Fig. 3(A)). The main impact is an increase in the amount of classification error (Fig. 3(B)). This result is not surprising: increasing the gain of the noise worsens readout performance.

However, markedly different results emerge in a scenario where tuning curves are offset (Fig. 2(B)) and β_x is altered while keeping β_y unchanged. In this case, $\rho_* = f(\beta_x)c_{\beta_x}$ with c_{β_x} given by

$$c_{\beta_x} = \frac{\beta_y \Delta v_x \sqrt{\tau_x \alpha_y}}{\Delta v_y \sqrt{\tau_y \alpha_x}},$$

and $f(\beta_x) = 1/\beta_x$. Alterations in β_x impact ρ_* in a non-monotonic fashion (Fig. 3(C)). A small increase from $\beta_x = 1$ to $\beta_x = 2$ shifts ρ_* towards a more negative value. However, further increasing to $\beta_x = 3$ and $\beta_x = 4$ increases ρ_* and alters the relationship between correlation and readout error (Fig. 3(D)).

Hidden in these results is a counter-intuitive finding: under certain circumstances, increasing β_x leads to a decrease in classification error. This can be seen with $\beta_x = 10$ (Fig. 3(D), dashed line), leading to lower error than $\beta_x = 3$ (green line) and $\beta_x = 4$ (red line) for negative correlations. Intuitively, this can happen when increasing β_x stretches the distribution of activity for population x along a single dimension away from the classification



boundary [13]. Similar findings are borne out of graphical explanations where noise covariance stretches the distribution of firing rates [36].

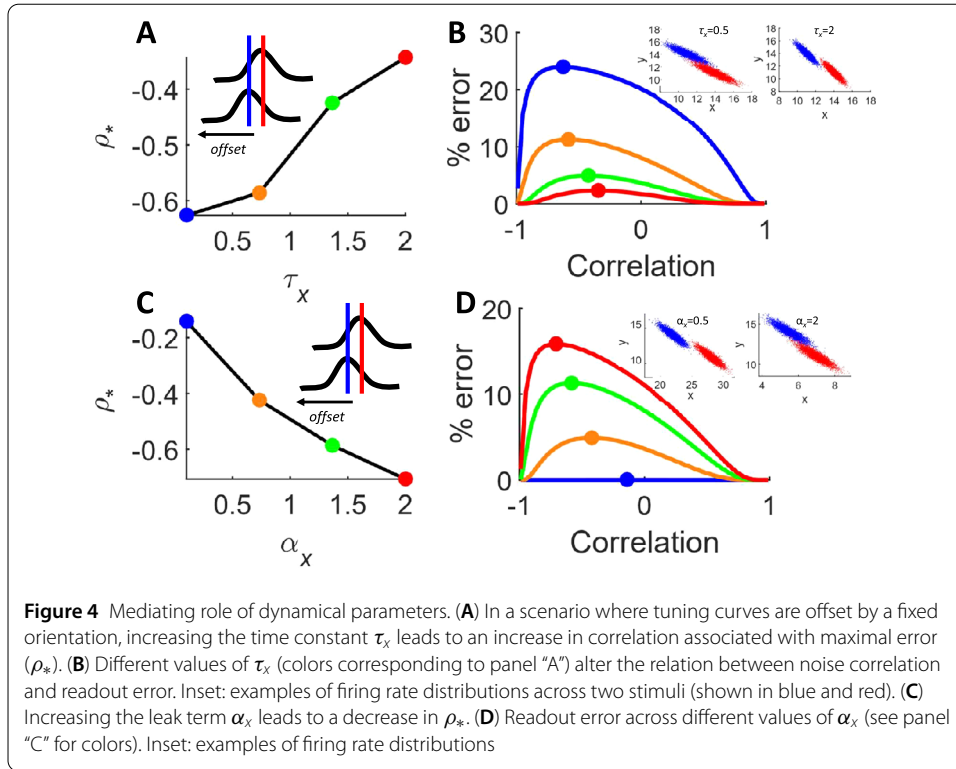
The benefits of noise gain are even more pronounced in a scenario where one population has higher gain than the other, as in Fig. 2(D). In this case, β_x monotonically shifts ρ_* towards decreasing values (Fig. 3(E)). For a broad range of positive correlation values, a high noise gain ($\beta_x > 1$) leads to lower classification error (Fig. 3(F)).

7 Impact of dynamical parameters

The approach described in the previous section can be applied to study the impact of the model's dynamical parameters on readout error. The two parameters of interest are the leak term (α) and the time constant (τ).

The effect of the time constants τ_x and τ_y on ρ_* can be expressed as

$$\rho_* = \frac{\tau_x}{\tau_y} \left(\frac{\Delta v_x \beta_y \sqrt{\alpha_y}}{\Delta v_y \beta_x \sqrt{\alpha_x}} \right)$$



for $|r_x| < |r_y|$. To study the effect of a single term (e.g., τ_x), we set $\rho_* = f(\tau_x)c_{\tau_x}$ with c_{τ_x} given by

$$c_{\tau_x} = \frac{\beta_y \Delta v_x \sqrt{\alpha_y}}{\beta_x \Delta v_y \sqrt{\tau_y \alpha_x}}$$

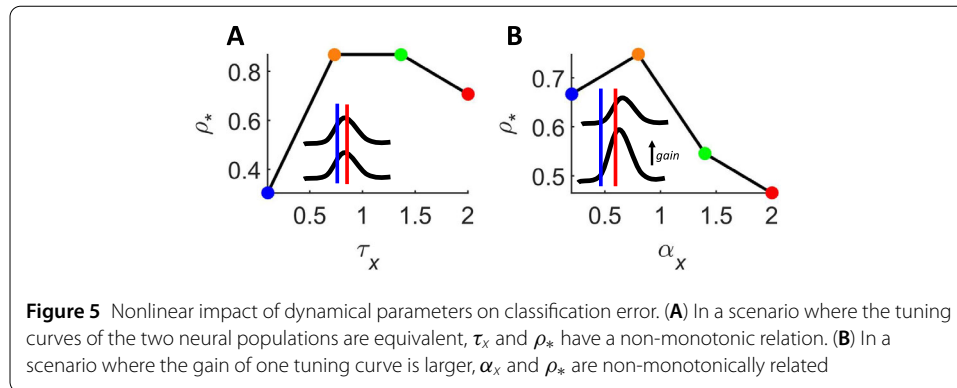
and $f(\tau_x) = \tau_x$. Similarly, the role of leak terms α_x and α_y on ρ_* is

$$\rho_* = \frac{\alpha_y}{\alpha_x} \left(\frac{\Delta v_x \beta_y \sqrt{\tau_x}}{\Delta v_y \beta_x \sqrt{\tau_y}} \right).$$

For a single term (α_x), we have $\rho_* = f(\alpha_x)c_{\alpha_x}$ with

$$c_{\alpha_x} = \frac{\beta_y \Delta v_x \sqrt{\tau_x \alpha_y}}{\beta_x \Delta v_y \sqrt{\tau_y}}$$

and $f(\alpha_x) = 1/\alpha_x$. Taking one scenario as illustration, we examined the case where tuning curves are offset by a fixed orientation ($r_x \rightarrow -r_y$). In this case, the time constant affects the relation between noise correlation and readout error, with larger values of τ_x shifting ρ_* towards smaller negative values of correlation (Fig. 4(A)). The reason for this shift follows from an earlier example (Fig. 2(D)), where an increased correlation resulted in greater overlap between the firing rate distributions, but only up to a point beyond which these distributions became too narrow to overlap. With larger values of τ_x , a given correlation does not create as much overlap as it would for smaller values of τ_x , thus leading to a shift in ρ_* .



The overall impact of a larger time constant is a decrease in classification error (Fig. 4(B)): as τ_x increases, there is less overlap between the distributions of firing rate across stimuli (Fig. 4(B), inset). By contrast, shifting the leak term α_x towards higher values decreases ρ_* (Fig. 4(C)) and increases overall readout error (Fig. 4(D)). The impact of increasing α_x on error is due to an increase in the overlap between firing rate distributions (Fig. 4(D), inset). The inverse effects of τ_x and α_x on these distributions explain their opposite impact on ρ_* .

More complex, non-monotonic relations between ρ_* and values of τ_x and α_x are found in different scenarios where tuning curves of the two populations are aligned (Fig. 5(A)) or when the gain of one population is larger (Fig. 5(B)).

Together, these results show that the integration time constant and leak term of the population model mediate the impact of noise correlation on classification error by shifting the value ρ_* at which correlation reaches maximal error. The impact of network parameters on readout error is therefore not straightforward to describe but is brought to light using a framework that derives error estimates from the dynamical parameters of a population model.

8 Discussion

This work described an analytical framework for performing Fisher linear decoding in a rate-based neural model. With this formalism, we began by capturing well-documented findings on the role of noise gain and correlations on discrimination error. Going further, the framework allowed us to analytically examine the mediating role of dynamical parameters (neuronal leak and time constant) on the relation between noise correlation and error. Overall, this framework suggests that linear decoding is highly sensitive to dynamical model parameters as well as the characteristics of the sensory input.

One surprising finding was the presence of conditions where increased neuronal noise led to reduced classification error. This result was especially prominent when the gain of the two population tuning curves was unmatched (Fig. 3(E)–(F)). Taken together, our findings cover all possible qualitative scenarios where noise correlations have either a beneficial, detrimental, or null effect on decoding [36].

A related approach termed the leaky competing accumulator model was proposed in order to account for perceptual decision making [37]. Some key differences exist between this model and ours. Firstly, our framework assumes a steady-state of neural activity that is characteristic of a decision point and does not capture the time-course of deliberation. Our

framework assumes an optimal bound on decision accuracy given a linear decoder, representing a ceiling in accuracy that would be associated with long response times (typically >500 ms in human subjects). Secondly, the accumulator model provides explicit connections, through lateral inhibition, that modulate correlations. These lateral connections, however, may also impact firing rates. By comparison, our framework isolates analytically the contribution of firing rates and correlations, and examines their relative role on perceptual discrimination.

It would be challenging to speculate on whether the analytical results provided would generalize to other classes of neural network models, particularly those that include a non-linear transfer function [38]. However, our work opens the door to such analyses by describing a framework for linking neuronal readout and dynamical modeling.

Limitations and future work. While the framework described here strived to cover all possible scenarios involving firing rates, noise correlations, and network parameters, it is important to emphasize that not all such scenarios are plausible from a physiological standpoint. In particular, the framework treats firing rates and noise correlations as independent contributors to decoding error and allows for implausible cases where increases in firing rate would lead to an increase, a decrease, or no impact on correlations. Interactions between stimulus and noise correlations are a crucial factor limiting the coding capacity of neural circuits [1, 23] and should be considered alongside the dynamical parameters discussed in this work.

Several future directions based on the proposed framework will be worth exploring. First, the assumption of equal class covariances in LDA is challenged by experimental work showing input-dependent neuronal variance [39]. This assumption could be relaxed by replacing LDA with quadratic discriminant analysis, albeit at the cost of a more complex solution when relating readout error to model parameters.

An extension of the current framework could consider the impact of pooling more than two neural populations, as well as more than two stimuli, when performing decoding. This extension would be helpful in examining the interactions between several populations of neurons, each with a unique tuning curve. Going further, one could examine decoding error at the limit of a large number of neurons with heterogeneous tuning curves that vary in both orientation preference and gain [2].

Conclusion. In summary, this work described a theoretical framework that merges Fisher linear decoding with a population model of sensory integration. This approach highlighted the role of correlation, neuronal noise, and network parameters, revealing a broad range of potential outcomes where different conditions generated either detrimental, beneficial, or null impacts on classification performance. These results motivate further developments in theoretical work that systematically link neural network models to optimal decoders in order to reveal the impact of key neurophysiological variables on sensory information processing.

Appendix

A.1 Solving the integrator model as a linear differential equation

To solve the integrator model, we began by dropping the unit indices to alleviate the notation:

$$\tau \frac{dx}{dt} = -\alpha x + v_i + \beta \xi(t)$$

$$\begin{aligned} \Leftrightarrow \frac{dx}{dt} &= -\frac{\alpha}{\tau}x + \frac{v_i + \beta\xi(t)}{\tau} \\ \Leftrightarrow \frac{dx}{dt} + \frac{\alpha}{\tau}x &= \frac{v_i + \beta\xi(t)}{\tau} \\ \Leftrightarrow \frac{dx}{dt} + p(t)x &= r(t) \end{aligned}$$

with $p(t) = \alpha/\tau$ and $r(t) = (v + \beta\xi(t))/\tau$. We defined

$$\begin{aligned} u(t) &= e^{\int p(t) dt} \\ &= e^{\int \frac{\alpha}{\tau} dt} \\ &= e^{\frac{\alpha}{\tau}t}. \end{aligned}$$

Then,

$$\begin{aligned} u(t)\left(\frac{dx}{dt} + p(t)x\right) &= u(t)r(t) \\ \Leftrightarrow e^{\frac{\alpha}{\tau}t}\frac{dx}{dt} + e^{\frac{\alpha}{\tau}t}\frac{\alpha}{\tau}x &= e^{\frac{\alpha}{\tau}t}r(t). \end{aligned}$$

Applying the chain rule,

$$\begin{aligned} \frac{d}{dt}(u(t)x) &= u(t)r(t) \\ \Leftrightarrow d(u(t)x) &= u(t)r(t) dt \\ \Leftrightarrow \int_0^t d(u(s)x) &= \int_0^t u(s)r(s) ds \\ \Leftrightarrow u(s)x|_0^t &= \int_0^t u(s)r(s) ds \\ \Leftrightarrow u(t)x - u(0)x_0 &= \int_0^t u(s)r(s) ds \\ \Leftrightarrow x = u(t)^{-1} \left(x_0 + \int_0^t u(s)r(s) ds \right) \\ &= e^{-\frac{\alpha}{\tau}t} \left(x_0 + \int_0^t e^{\frac{\alpha}{\tau}s} \frac{v_i + \beta\xi(s)}{\tau} ds \right) \\ &= e^{-\frac{\alpha}{\tau}t} \left(x_0 + \frac{1}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} (v_i + \beta\xi(s)) ds \right) \\ &= e^{-\frac{\alpha}{\tau}t} \left(x_0 + \frac{1}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} v_i ds + \frac{1}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} \beta\xi(s) ds \right) \\ &= e^{-\frac{\alpha}{\tau}t} \left(x_0 + \frac{1}{\tau} \frac{\tau v_i}{\alpha} e^{\frac{\alpha}{\tau}s} \Big|_0^t + \frac{\beta}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} \xi(s) ds \right) \\ &= e^{-\frac{\alpha}{\tau}t} \left(x_0 + \frac{v_i}{\alpha} [1 - e^{-\frac{\alpha}{\tau}t}] + \frac{\beta}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} \xi(s) ds \right) \\ &= x_0 e^{-\frac{\alpha}{\tau}t} - \frac{v_i}{\alpha} [1 - e^{-\frac{\alpha}{\tau}t}] + e^{-\frac{\alpha}{\tau}t} \frac{\beta}{\tau} \int_0^t e^{\frac{\alpha}{\tau}s} \xi(s) ds, \end{aligned}$$

$$x = \frac{v_i}{\alpha} + \left[x_0 - \frac{v_i}{\alpha} \right] e^{-\frac{\alpha}{\tau} t} + \frac{\beta}{\tau} \int_0^t e^{-\frac{\alpha}{\tau}(t-s)} \xi(s) ds.$$

A.2 Expected value and variance

We sought to find the expected mean and variance of the random variable x such that

$$x = \mu + (x_0 - \mu)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dB_s.$$

The expected mean is

$$\begin{aligned} E[x] &= E \left[\mu + (x_0 - \mu)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dB_s \right] \\ &= E[\mu] + E[(x_0 - \mu)e^{-\theta t}] + E \left[\sigma \int_0^t e^{-\theta(t-s)} dB_s \right]. \end{aligned}$$

Given the zero-mean property of Ito integrals,

$$E \left[\sigma \int_0^t e^{-\theta(t-s)} dB_s \right] = 0,$$

we have

$$E[x] = \mu + (x_0 - \mu)e^{-\theta t}. \tag{11}$$

The expected variance is

$$\begin{aligned} \text{var}(x) &= \text{var} \left(\mu + (x_0 - \mu)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dB_s \right) \\ &= \sigma^2 \text{var} \left(\int_0^t e^{-\theta(t-s)} dB_s \right) \\ &= \sigma^2 \left(E \left[\left(\int_0^t e^{-\theta(t-s)} dB_t \right)^2 \right] - E \left[\int_0^t e^{-\theta(t-s)} dB_s \right]^2 \right) \\ &= \sigma^2 E \left[\left(\int_0^t e^{-\theta(t-s)} dB_s \right)^2 \right]. \end{aligned}$$

By Ito isometry,

$$\sigma^2 E \left[\left(\int_0^t e^{-\theta(t-s)} dB_t \right)^2 \right] = \sigma^2 E \left[\int_0^t (e^{-\theta(t-s)})^2 ds \right].$$

Hence, the expected variance can be concisely written as

$$\begin{aligned} \text{var}(x) &= \sigma^2 \int_0^t e^{-2\theta(t-s)} ds \\ &= \sigma^2 \left[-\frac{e^{-2\theta(t-s)}}{-2\theta} \right]_0^t \\ &= \frac{\sigma^2}{2\theta} [e^{-2\theta(t-t)} e^{-2\theta(t-0)}] \end{aligned}$$

$$= \frac{\sigma^2}{2\theta} [1 - e^{-2\theta t}].$$

A.3 Classification error

The classification error as a function of neural activity is given by

$$\begin{aligned} \varepsilon &= \frac{1}{2} \int_{-\infty}^0 \frac{1}{\sqrt{2\zeta^2\pi}} e^{-\frac{(w-\eta_1)^2}{2\zeta^2}} dw + \frac{1}{2} \int_0^{\infty} \frac{1}{\sqrt{2\zeta^2\pi}} e^{-\frac{(w-\eta_0)^2}{2\zeta^2}} dw \\ &= \frac{1}{2} \frac{1}{\sqrt{2\zeta^2\pi}} \left[\sqrt{\frac{\pi\zeta^2}{2}} \operatorname{erfc}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right) + \sqrt{\frac{\pi}{2}} \left(\sqrt{\zeta^2} \operatorname{erf}\left(\frac{\eta_0}{\sqrt{2\zeta^2}}\right) + \sqrt{\zeta^2} \right) \right] \\ &= \frac{1}{2} \frac{1}{\sqrt{2\zeta^2\pi}} \sqrt{\frac{\pi\zeta^2}{2}} \left[\operatorname{erfc}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right) + \operatorname{erf}\left(\frac{\eta_0}{\sqrt{2\zeta^2}}\right) + 1 \right] \\ &= \frac{1}{4} \left[1 - \operatorname{erf}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right) + \operatorname{erf}\left(\frac{\eta_0}{\sqrt{2\zeta^2}}\right) + 1 \right] \\ &= \frac{1}{4} \left[2 - \operatorname{erf}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right) + \operatorname{erf}\left(\frac{-\eta_1}{\sqrt{2\zeta^2}}\right) \right] \\ &= \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right) \right] \\ &= \frac{1}{2} \operatorname{erfc}\left(\frac{\eta_1}{\sqrt{2\zeta^2}}\right). \end{aligned}$$

Substituting the mean and variance from the previous section, this becomes

$$\begin{aligned} \varepsilon &= \frac{1}{2} \operatorname{erfc}\left(2 \frac{d^2}{4\sqrt{2d^2}}\right) \\ &= \frac{1}{2} \operatorname{erfc}\left(\frac{1}{2\sqrt{2}} \sqrt{d^2}\right). \end{aligned}$$

A.4 Mahalanobis distance

We began with the following definitions:

$$\begin{aligned} W &= (2\Sigma)^{-1} \Delta \boldsymbol{\mu}, \\ c &= W \cdot \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) + b, \\ \eta_i &= W \cdot \boldsymbol{\mu}_i + b - c, \\ \zeta^2 &= W^T \Sigma W. \end{aligned}$$

Expanding η_i yields

$$\begin{aligned} \eta_i &= W \cdot \boldsymbol{\mu}_i + b - \left(W \cdot \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) + b \right) \\ &= W \cdot \left(\boldsymbol{\mu}_i - \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) \right). \end{aligned}$$

Given

$$\begin{aligned} \eta_1 &= \frac{1}{2} W \cdot (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) \\ &= \frac{1}{2} W \cdot \Delta \boldsymbol{\mu} \\ &= -\eta_0, \end{aligned}$$

we expanded W using the property $u \cdot v = u^T v$,

$$\begin{aligned} \eta_1 &= \frac{1}{2} [(2\Sigma)^{-1} \Delta \boldsymbol{\mu}] \cdot \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} [\Sigma^{-1} \Delta \boldsymbol{\mu}]^T \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} \Delta \boldsymbol{\mu}^T (\Sigma^{-1})^T \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} \Delta \boldsymbol{\mu}^T \Sigma^{-1} \Delta \boldsymbol{\mu}. \end{aligned}$$

Hence, the squared Mahalanobis distance between means is

$$d^2 = \Delta \boldsymbol{\mu}^T \Sigma^{-1} \Delta \boldsymbol{\mu}.$$

We can rewrite η_i as

$$\eta_1 = \frac{1}{4} d^2 = -\eta_0.$$

Similarly, for the variance ζ^2 ,

$$\begin{aligned} \zeta^2 &= [(2\Sigma)^{-1} \Delta \boldsymbol{\mu}]^T \Sigma (2\Sigma)^{-1} \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} \Delta \boldsymbol{\mu}^T \Sigma^{-1} \Sigma \Sigma^{-1} \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} \Delta \boldsymbol{\mu}^T \Sigma^{-1} \Delta \boldsymbol{\mu} \\ &= \frac{1}{4} d^2. \end{aligned}$$

A.5 Derivation of error

We analyzed the extrema of the error function in relation to noise correlation by taking its first derivative through the chain rule

$$\begin{aligned} \frac{d\varepsilon}{d\rho} &= \frac{d}{d\rho} \left(\frac{1}{2} \operatorname{erfc} \left(\frac{1}{2\sqrt{2}} \sqrt{d^2} \right) \right) \\ &= \frac{1}{2} \frac{d}{d\rho} (\operatorname{erfc}(z)) \\ &= \frac{1}{2} \frac{d}{dz} (\operatorname{erfc}(z)) \frac{dz}{dd^2} \frac{dd^2}{d\rho}, \end{aligned} \tag{12}$$

with

$$d^2 = \frac{1}{1 - \rho^2} [r_x^2 + r_y^2 - 2\rho r_x r_y]. \tag{13}$$

The first derivative is given by

$$\begin{aligned} \frac{d}{dz} (\operatorname{erf} c(z)) &= \frac{-2e^{-z^2}}{\sqrt{\pi}} \\ &= \frac{-2e^{-(\frac{1}{2\sqrt{2}}\sqrt{d^2})^2}}{\sqrt{\pi}} \\ &= \frac{-2e^{-\frac{1}{8}d^2}}{\sqrt{\pi}}. \end{aligned} \tag{14}$$

The second derivative is

$$\begin{aligned} \frac{dz}{dd^2} &= \frac{d}{dd^2} \frac{1}{2\sqrt{2}} \sqrt{d^2} \\ &= \frac{1}{4\sqrt{2}d^2}. \end{aligned} \tag{15}$$

The third derivative is

$$\begin{aligned} \frac{dd^2}{d\rho} &= \frac{d}{d\rho} \left(\frac{1}{1 - \rho^2} [r_x^2 + r_y^2 - 2\rho r_x r_y] \right) \\ &= [r_x^2 + r_y^2 - 2\rho r_x r_y] \frac{d}{d\rho} \frac{1}{1 - \rho^2} + \frac{1}{1 - \rho^2} \frac{d}{d\rho} [r_x^2 + r_y^2 - 2\rho r_x r_y] \\ &= [r_x^2 + r_y^2 - 2\rho r_x r_y] \frac{2\rho}{(1 - \rho^2)^2} + \frac{1}{1 - \rho^2} [-2r_x r_y] \\ &= \frac{1}{(1 - \rho^2)^2} [(r_x^2 + r_y^2 - 2\rho r_x r_y)2\rho + (1 - \rho^2)(-2r_x r_y)] \\ &= \frac{1}{(1 - \rho^2)^2} [(2\rho r_x^2 + 2\rho r_y^2 - 2\rho 2\rho r_x r_y) + (-2r_x r_y + 2r_x r_y \rho^2)] \\ &= \frac{-2}{(1 - \rho^2)^2} [\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y]. \end{aligned} \tag{16}$$

A.6 Extrema of error

We evaluated the extrema of error by finding the points where Eqs. (14)–(16) are equal to zero,

$$\begin{aligned} 0 = \frac{d}{dz} (\operatorname{erf} c(z)) &\Leftrightarrow 0 = \frac{-2e^{-\frac{1}{8}d^2}}{\sqrt{\pi}}, \\ d^2 &\rightarrow \infty. \end{aligned} \tag{17}$$

We assumed that the ratios r_x and r_y are finite and the Euclidean distance between the distribution means is finite and non-null. In other words, if $d^2 \rightarrow \infty$ it is exclusively due to the correlation coefficient. Then,

$$d^2 \rightarrow \infty \Leftrightarrow |\rho| \rightarrow 1. \tag{18}$$

We proceeded in a similar fashion for the second derivative (Eq. (15)):

$$\begin{aligned}
 0 = \frac{dz}{dd^2} &\Leftrightarrow 0 = \frac{1}{4\sqrt{2d^2}} \\
 d^2 \rightarrow \infty &\Leftrightarrow |\rho| \rightarrow 1.
 \end{aligned}
 \tag{19}$$

The third derivative (Eq. (16)) is

$$\begin{aligned}
 0 = \frac{dd^2}{d\rho} &\Leftrightarrow 0 = \frac{-2}{(1-\rho^2)^2} [\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y], \\
 &\Leftrightarrow 0 = \rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y.
 \end{aligned}
 \tag{20}$$

Depending on network parameters, two cases are possible. One case arises if one of the ratios, either r_x or r_y , is zero. This happens if the mean activity of one population is equal across inputs. If the mean activity of both units remained unchanged, the resulting multivariate distributions would overlap, thus breaking the basic assumptions justifying the choice of LDA. In this first case,

$$0 = \frac{dd^2}{d\rho} \Leftrightarrow 0 = \rho \quad \text{if } r_x = 0 \text{ or } r_y = 0.
 \tag{21}$$

The second case occurs when neither r_x nor r_y is zero:

$$\begin{aligned}
 0 = \frac{dd^2}{d\rho} &\Leftrightarrow 0 = \rho^2 - \rho \frac{r_x^2 + r_y^2}{r_x r_y} + 1 \\
 \Leftrightarrow \rho &= \frac{\frac{r_x^2 + r_y^2}{r_x r_y} \pm \sqrt{\frac{(r_x^2 + r_y^2)^2}{r_x^2 r_y^2} - 4}}{2} \\
 &= \frac{r_x^2 + r_y^2 \pm \sqrt{(r_x^2 + r_y^2)^2 - 4r_x^2 r_y^2}}{2r_x r_y} \\
 &= \frac{r_x^2 + r_y^2 \pm \sqrt{r_x^4 + r_y^4 - 2r_x^2 r_y^2}}{2r_x r_y} \\
 &= \frac{r_x^2 + r_y^2 \pm \sqrt{(r_x^2 - r_y^2)^2}}{2r_x r_y} \\
 &= \frac{r_x^2 + r_y^2 \pm |r_x^2 - r_y^2|}{2r_x r_y} \\
 &= \frac{r_x^2 + r_y^2 \pm [\max(r_x^2, r_y^2) - \min(r_x^2, r_y^2)]}{2r_x r_y}.
 \end{aligned}$$

The last expression can be decomposed into four distinct cases. First, when $r_x \rightarrow r_y$,

$$\rho \rightarrow \frac{r_y^2 + r_y^2}{2r_y r_y} \rightarrow 1.
 \tag{22}$$

Second, when $r_x \rightarrow -r_y$,

$$\rho \rightarrow \frac{r_y^2 + r_y^2}{-2r_y r_y} \rightarrow -1. \tag{23}$$

Third, when $r_x \neq r_y$, we examined the positive and negative roots of ρ . The positive root is

$$\begin{aligned} \rho_+ &= \frac{r_x^2 + r_y^2 + \max(r_x^2, r_y^2) - \min(r_x^2, r_y^2)}{2r_x r_y} \\ &= \frac{\max(r_x^2, r_y^2)}{r_x r_y}. \end{aligned} \tag{24}$$

Because $|\max(r_x^2, r_y^2)| > |r_x r_y|$ from the assumption that one ratio is smaller than the other (or unequal, non-null), this means that $|\rho_+| > 1 \forall r_x, r_y$. Since the correlation is bound in the range $]-1, 1[$, the positive root must be rejected. The negative root does not suffer from the same problem,

$$\begin{aligned} \rho_- &= \frac{r_x^2 + r_y^2 - \max(r_x^2, r_y^2) + \min(r_x^2, r_y^2)}{2r_x r_y} \\ &= \frac{\min(r_x^2, r_y^2)}{r_x r_y}. \end{aligned} \tag{25}$$

Fourth, when either $r_x = 0$ or $r_y = 0$, $\rho = 0$.

A.7 Minima and maxima

We determined upward and downward trends of the error curve by calculating the sign of the derivative between the potential maxima (considering that they are mutually exclusive). Taking Eqs. (14)–(16) and substituting into Eq. (12),

$$\begin{aligned} \frac{d\varepsilon}{d\rho} &= \frac{1}{2} \frac{d}{dz} (\operatorname{erf} c(z)) \frac{dz}{dd^2} \frac{dd^2}{d\rho} \\ &= \frac{1}{2} \frac{-2e^{-\frac{1}{8}d^2}}{\sqrt{\pi}} \frac{1}{4\sqrt{2}d^2} \frac{-2}{(1-\rho^2)^2} [\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y] \\ \Leftrightarrow \operatorname{sign}\left(\frac{d\varepsilon}{d\rho}\right) &= \operatorname{sign}(\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y). \end{aligned} \tag{26}$$

For the condition where $r_x \rightarrow r_y$,

$$\begin{aligned} \operatorname{sign}\left(\frac{d\varepsilon}{d\rho}\right) &= \operatorname{sign}(-\rho r_y^2) \\ &= -\operatorname{sign}(\rho). \end{aligned} \tag{27}$$

For the condition where either $r_x = 0$ or $r_y = 0$, we have already found the zeros of $\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y$ to be ρ_- and ρ_+ . To determine $\operatorname{sign}(d\varepsilon/d\rho)$, we need to know whether the

extremum of the parabola is a minimum or a maximum,

$$\frac{d^2}{d\rho^2}(\rho^2 r_x r_y - \rho(r_x^2 + r_y^2) + r_x r_y) = 2 > 0.$$

Given $\rho_+ > 1$,

$$\frac{d\varepsilon}{d\rho} > 0 \Leftrightarrow p \in [-1, \rho_-] \quad (28)$$

$$\frac{d\varepsilon}{d\rho} < 0 \Leftrightarrow p \in [\rho_-, 1]. \quad (29)$$

Regardless of the conditions for r_x and r_y , following Eqs. (27)–(29), the error curve as a function of correlation increases from $\rho = -1$ until its maximum, found at a value of $\rho_* = 0$ or $\rho_* = \rho_-$, and then decreases until $\rho = 1$.

Acknowledgements

This work benefited from discussions with Brent Doiron and Richard Naud.

Funding

This work was supported by a Discovery grant to J.P.T. from the Natural Sciences and Engineering Council of Canada (NSERC Grant No. 210977).

Abbreviations

LDA, Linear discriminant analysis.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' contributions

JPT and MC conceptualized the study, designed the work, performed the analyses, interpreted the data, wrote software, drafted the work, and revised the final version of the manuscript. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 14 April 2020 Accepted: 3 February 2021 Published online: 19 February 2021

References

- Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A. Information-limiting correlations. *Nat Neurosci*. 2014;17:1410–7.
- Ecker AS, Berens P, Tolias AS, Bethge M. The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci*. 2011;31:14272–83.
- Shamir M, Sompolinsky H. Implications of neuronal diversity on population coding. *Neural Comput*. 2006;18:1951–86.
- Quiñero Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*. 2009;10:173–85.
- Wen H, Shi J, Zhang Y, Lu KH, Cao J, Liu Z. Neural encoding and decoding with deep learning for dynamic natural vision. *Cereb Cortex*. 2018;28:4136–60.
- Glasser JI, Benjamin AS, Chowdhury RH, Perich MG, Miller LE, Kording KP. Machine learning for neural decoding. *ENeuro*. 2020;7:1–16.
- Kriegeskorte N, Douglas PK. Interpreting encoding and decoding models. *Curr Opin Neurobiol*. 2019;55:167–79.
- Klompf S, David SV, Yin P, Shamma SA, Maass W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J Neurophysiol*. 2012;108:1366–80.

9. Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol.* 2008;100:1407–19.
10. Nienborg H, Cumming B. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Curr Opin Neurobiol.* 2010;20:376–81.
11. Shadlen MN, Britten KH, Newsome WT, Movshon JA. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci.* 1996;16:1486–510.
12. Buonomano DV, Maass W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat Rev Neurosci.* 2009;10:113–25.
13. Calderini M, Zhang S, Berberian N, Thivierge JP. Optimal readout of correlated neural activity in a decision-making circuit. *Neural Comput.* 2018;30:1573–611.
14. Berberian N, MacPherson A, Giraud E, Richardson L, Thivierge JP. Neuronal pattern separation of motion-relevant input in LIP activity. *J Neurophysiol.* 2017;117:738–55.
15. Rich EL, Wallis JD. Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci.* 2016;19:973–80.
16. Averbeck BB, Lee D. Effects of noise correlations on information encoding and decoding. *J Neurophysiol.* 2006;95:3633–44.
17. Cain N, Barreiro AK, Shadlen M, Shea-Brown E. Neural integrators for decision making: a favorable tradeoff between robustness and sensitivity. *J Neurophysiol.* 2013;109:2542–59.
18. Goldman MS. Memory without feedback in a neural network. *Neuron.* 2009;61:621–34.
19. Ganguli S, Biseley JW, Roitman JD, Shadlen MN, Goldberg ME, Miller KD. One-dimensional dynamics of attention and decision making in LIP. *Neuron.* 2008;58:15–25.
20. Miri A, Daie K, Arrenberg AB, Baier H, Aksay E, Tank DW. Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nat Neurosci.* 2011;14:1150–9.
21. Murphy BK, Miller KD. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron.* 2009;61:635–48.
22. Chance FS, Abbott LF, Reyes AD. Gain modulation from background synaptic input. *Neuron.* 2002;35:773–82.
23. Romyantsev OI, Lecoq JA, Hernandez O, Zhang Y, Savall J, Chrapkiewicz R, Li J, Zeng H, Ganguli S, Schnitzer MJ. Fundamental bounds on the fidelity of sensory cortical coding. *Nature.* 2020;580:100–5.
24. Burak Y, Fiete IR. Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci.* 2012;109:17645–50.
25. Cover TM, Thomas JA. *Elements of information theory.* 2nd ed. New York: Wiley; 2006.
26. Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci.* 2011;14:811–9.
27. Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature.* 1994;370:140–3.
28. Brunel N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J Comput Neurosci.* 2000;8:183–208.
29. Bujan AF, Aertsen A, Kumar A. Role of input correlations in shaping the variability and noise correlations of evoked activity in the neocortex. *J Neurosci.* 2015;35:8611–25.
30. de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A. Correlation between neural spike trains increases with firing rate. *Nature.* 2007;448:802–6.
31. Graupner M, Reyes AD. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *J Neurosci.* 2013;33:15075–85.
32. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD. The asynchronous state in cortical circuits. *Science.* 2010;327:587–90.
33. Salinas E, Sejnowski TJ. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J Neurosci.* 2000;20:6193–209.
34. Hu Y, Zylberberg J, Shea-Brown E. The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes. *PLoS Comput Biol.* 2014;10:e1003469.
35. Yim MY, Kumar A, Aertsen A, Rotter S. Impact of correlated inputs to neurons: modeling observations from in vivo intracellular recordings. *J Comput Neurosci.* 2014;37:293–304.
36. Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat Rev Neurosci.* 2006;7:358–66.
37. Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol Rev.* 2001;108:550–92.
38. Ostojic S, Brunel N. From spiking neuron models to linear-nonlinear models. *PLoS Comput Biol.* 2011;7:e1001056.
39. Churchland AK, Kiani R, Chaudhuri R, Wang X-J, Pouget A, Shadlen MN. Variance as a signature of neural computations during decision making. *Neuron.* 2011;69:818–31.