Efficient coding explains neural response homeostasis and stimulus-specific adaptation

Edward James Young
Computational and Biological Learning Lab
Department of Engineering
University of Cambridge
ey245@cam.ac.uk
&
Yashar Ahmadian
Computational and Biological Learning Lab
Department of Engineering
University of Cambridge
ya311@cam.ac.uk

Abstract

Neurons are typically sensitive to a small fraction of stimulus space. If the environment changes, making certain stimuli more prevalent, neurons sensitive to those stimuli would respond more often and therefore, if the stimulus-response mapping remains fixed, have a higher average firing rate. However, sufficiently prolonged exposure to the new environment typically causes such neurons to adapt by responding less vigorously. If adaptation consistently returns the average firing rate of neurons, or populations of similarly tuned neurons, to its value prior to environmental shift, it is termed firing rate homeostasis. In sensory cortex adaptation is typically also stimulus specific, with neurons maintaining or even increasing their responsiveness to stimuli far from over-represented ones. Here, we present a normative explanation of firing rate homeostasis grounded in the efficient coding principle. Unlike previous theories based on efficient coding, we formulate the problem in a computation-agnostic manner, enabling our framework to apply far from the sensory periphery. We show that homeostasis can provide an optimal solution to a trade-off between coding fidelity and the metabolic cost of neural firing. We provide quantitative conditions necessary for the optimality of firing rate homeostasis, and predict how adaptation should deviate from homeostasis when these conditions are violated. Based on biological estimates of relevant parameters, we show that these conditions do hold in areas of cortex where homeostatic adaptation has been observed. Finally, we apply our framework to distributed distributional codes, a specific computational theory of neural representations serving Bayesian inference. We show that the resultant coding scheme can be accomplished by divisive normalisation with adaptive weights. We further demonstrate how homeostatic coding, coupled with such Bayesian neural representations, explains stimulus-specific adaptation, as observed, *e.g.*, in the primary visual cortex.

1 Introduction

Neurons act as noisy feature detectors. Moreover, when the responsiveness of a neuron increases, the strength of its response noise, or trial-to-trial variability, typically grows sublinearly and more slowly than its average response; for Poisson-like firing, for example, response noise grows as the square root of the mean response. The neuron can therefore increase its signal-to-noise ratio by increasing its responsiveness or response gain. However, this increase in coding fidelity comes at the cost of an

elevated firing rate, and thus higher metabolic energy expenditure. From a normative perspective, coding fidelity and metabolic cost are thus two conflicting forces. In this paper, we ask how neurons should adjust their response gains in order to optimally balance these two forces, *irrespective* of the computations they perform or the complexity of the features they detect. In particular, we address the following question: given a neural population with a set of *arbitrary* tuning curve shapes, how should neurons optimally adjust their gains depending on the stimulus statistics prevailing in the environment?

Suppose an environmental shift makes a feature to which a group of neurons is sensitive more prevalent. Over time, that feature will be presented more often, in turn increasing the average firing rate of these neurons. Typically, the neurons will then adapt by reducing their responsiveness (Solomon and Kohn, 2014; Clifford et al., 2007; Benucci et al., 2013). A special case of such adaptation is *firing rate homeostasis* (Desai, 2003; Turrigiano and Nelson, 2004; Maffei and Turrigiano, 2008; Hengen et al., 2013), in which adaptation brings the average firing rate back to the rate prior to the environmental shift. Thus, under firing rate homeostasis, neuronal populations maintain a constant stimulus-averaged firing rate in the face of changes to the environment (Benucci et al., 2013; Hengen et al., 2013; Maffei and Turrigiano, 2008).

There are multiple levels at which homeostatic results can be observed. Firstly, there is population homeostasis, in which the stimulus-average firing rate of an entire population of neurons remains constant, without individual neurons necessarily holding their rates constant (Slomowitz et al., 2015; Sanzeni et al., 2023). Secondly, there is what we term *cluster homeostasis*. In this form of homeostasis, stimulus-average firing rate of groups or clusters of neurons with similar stimulus preferences remains stable under environmental shifts, but the firing rate of individual neurons within a cluster can change (Benucci et al., 2013). Lastly, homeostasis can occur at the level of individual neurons, in which case the firing rate of each individual neuron is kept constant under shifts in the environment (Marder and Prinz, 2003). Note that these three forms are progressively stronger.

Previous normative explanations for firing rate homeostasis often focus on the necessity of avoiding hypo- or hyper-excited states (Turrigiano and Nelson, 2004; Maffei and Turrigiano, 2008; Keck et al., 2013; Hengen et al., 2013). Although this might explain homeostasis at the population level, it does not adequately explain homeostasis at a more fine-grained level. Our task in this paper is therefore to provide a normative account of homeostasis at levels below that of an entire population with heterogeneous stimulus tuning.

Specifically, our theory provides a normative account of firing rate homeostasis at the level of clusters of similarly tuned neurons. Furthermore, as we will show, when coupled with a specific computational theory of sensory representations (which determines tuning curve shapes), it further predicts the typical finding that such adaptations are stimulus specific and shift tuning curves away from more prevalent stimuli (Kohn, 2007; Schwartz et al., 2007). In stimulus specific adaptation, the suppression of response is greater for test stimuli that are closer to an over-represented adaptor stimulus than for stimuli further away, to which the neuron may even respond more strongly. This causes a repulsion of tuning curves away from the adaptor (Movshon and Lennie, 1979; Müller et al., 1999; Dragoi et al., 2000, 2002; Benucci et al., 2013).

To address our normative question regarding the optimal adjustment of responsiveness, we work within the framework of efficient coding theory (Attneave, 1954; Nadal and Parga, 1999; Laughlin, 1981; Barlow, 2012; Linsker, 1988; Ganguli and Simoncelli, 2014; Wei and Stocker, 2015; Atick and Redlich, 1990). Efficient coding theory begins by asking two questions: What is this neural system attempting to encode, and what features of biology constrain the quality of its encoding? We can then compute (analytically or numerically) the optimal encoding subject to the posited biological constraints and costs. Under the assumption that natural selection has acted to optimise performance, this solution serves as a prediction for encoding properties of real neurons.

Concretely, the *Infomax Principle* (Atick and Redlich, 1990; Ganguli and Simoncelli, 2014; Wei and Stocker, 2015) states that sensory systems optimise the mutual information between a noisy neural representation and an external stimulus, subject to metabolic constraints. In order to do so, the neural system should exploit the statistics of its local environment (Simoncelli and Olshausen, 2001). A non-adaptive system that performs well in one environment may perform poorly in an environment with

different stimulus statistics. Efficient coding theories can therefore be used, in particular, to predict how sensory systems should *adapt* to changes in the environment statistics (Wei and Stocker, 2015).

We use the mutual information between the stimulus and the neural response as a measure of coding fidelity, and the expected number of spikes fired as a measure of metabolic cost. We show that when neurons are highly selective and population responses have a high-dimensional geometry, optimal gains yield firing rate homeostasis across environments; in other words, average firing rates are maintained due to optimal gain adjustments, despite changes in the environmental statistics. Our work complements and extends previous theoretical work, in particular that of Ganguli and Simoncelli (2014). A novelty of our framework is that we formulate the problem in a way that is agnostic of the computations that the population is designed to perform, which dictate the shapes and arrangement of its tuning curves. We thus allow the population to have tuning curves of arbitrary configuration and shape, possibly defined on a high-dimensional stimulus, but do not optimise their shapes or placements. Ganguli and Simoncelli, on the other hand, did optimise tuning curve widths and placements, but only for the case of a one-dimensional stimulus and a homogeneous (up to stimulus reparametrisation) population of unimodal or sigmoidal tuning curves. Additionally, our framework allows us to consider more general noise models than simple Poissonian firing, in particular correlated and power-law noise.

Having shown the optimality of homeostatic codes without making assumptions on the nature of the neural representation and computation, we apply our framework to a Bayesian theory of neural computation, namely the *Distributed Distributional Code (DDC)* (Vertes and Sahani, 2018). The theory of DDCs assumes that the brain possesses an internal generative model of sensory input arising from unobserved latent variables. The firing rates of neurons are then assumed to directly represent the Bayesian posterior expectation of a rich set of (fixed) functions of the latent variables. Combining the theory of DDCs with our normative results yields what we call a *homeostatic DDC*. Homeostatic DDCs are able to account for stimulus specific adaptation effects which cannot be fully accounted for in previous efficient coding frameworks (Wei and Stocker, 2015; Ganguli and Simoncelli, 2014; Snow et al., 2016). A special case of a homeostatic DDC is a form of representation we term *Bayes-ratio coding*. We show that Bayes-ratio coding has attractive computational properties: it can be propagated between populations without synaptic weight adjustments, and it can be achieved by divisive normalisation with adaptive weights (Carandini and Heeger, 2012; Westrick et al., 2016).

We start the next section by introducing the basic mathematical framework we use to address our normative question. This includes formulating the notion of neuron clustering based on stimulus selectivity. We then show that our framework predicts that the firing rate of each cluster should remain constant subject to shifts in environmental statistics, with individual neurons free to shuffle their average rates. We additionally demonstrate that our framework can account for a wide distribution of mean single-neuron firing rates as observed in cortex. We show that the conditions necessary for our theory to apply are expected to hold in cortical areas, in particular in the primary visual cortex (V1). We then extend our analysis to apply to the cases of correlated and power-law noise, and numerically validate the quality of our homeostatic solutions. Lastly, we apply our theory to DDC representational codes, showing how homeostatic DDC's can account for stimulus specific adaptation effects observed experimentally.

2 Results

2.1 Theoretical framework

We consider a population of N neurons, responding to the (possibly high-dimensional) stimulus $\boldsymbol{s}$, with marginal distribution $P(\boldsymbol{s})$. The distribution $P(\boldsymbol{s})$ is determined by the environment; accordingly, if the environment changes (making certain stimuli more or less prevalent), so will P. We assume that our population engages in rate coding using time bins of a fixed duration, and denote the vector of joint population spike counts in a coding interval by $\boldsymbol{n} = (n_1, \ldots, n_N)$.

We are interested in how changes in the stimulus distribution, P, affect the responsiveness of neurons.

We therefore adopt a shape-amplitude decomposition of the neural tuning curves. The tuning curve of the i-th neuron, $h_i(\boldsymbol{s})$, is factorised into a *representational curve*, $\Omega_i(\boldsymbol{s})$, and a *gain*, g_i:

$$h_i(\boldsymbol{s}) = g_i \Omega_i(\boldsymbol{s}) = \text{gain} \times \text{representational curve}.$$

Fig. 1a demonstrates the effect of changing the gain while keeping the representational curve constant. Importantly, we do not make any assumptions on the shape of the representational curve: Ω_i can be any complex (*e.g.* multi-modal, discontinuous) function of the possibly high-dimensional stimulus, and can thus represent *any* computation. This makes our treatment more general than other efficient coding frameworks (*e.g.* Ganguli and Simoncelli (2014)), which place tight constraints on the shape and configuration of the tuning curves. In particular, this generality enables our theory to apply to populations located deep in the processing pathway, and not just to primary sensory neurons.

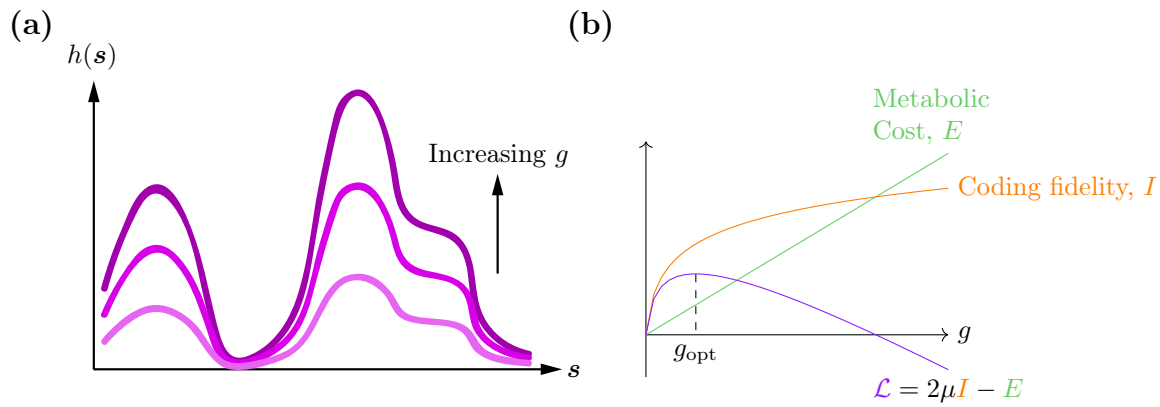

Figure 1: The effect of changinf the gain on tuning curves and components of the objective function. (a) As the gain, g, of a neuron is increased, the shape of its tuning curve remains the same, but all firing rates are scaled upwards. The cartoon shows a one-dimensional example, but our theory applies to general tuning curve shapes and joint configurations of population tuning curves, on high-dimensional stimulus spaces. (b) Cartoon representation of our objective function. Neural gains will be chosen to maximise an objective function, $\mathcal{L}$, which is the weighted difference between mutual information, I, which captures coding fidelity, and metabolic cost, E, given by average population firing rate.

Neural responses $\boldsymbol{n}$ are taken to be noisy encoding of neural tuning curves $\boldsymbol{h}(\boldsymbol{s})$, $\boldsymbol{n}|\boldsymbol{s} \sim P_{\text{noise}}(\boldsymbol{n}|\boldsymbol{h}(\boldsymbol{s}))$ with $\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}] = \boldsymbol{h}(\boldsymbol{s})$. We consider a number of different noise models, including correlated and power-law noise, but begin with the simplest case of uncorrelated Gaussian noise with Poisson-like scaling of variance. Our framework is summarised in Fig. 2.

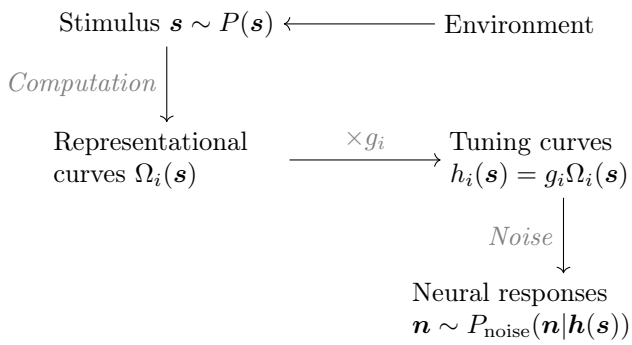

Figure 2: The basic framework for our analysis. The environment gives rise to a stimulus distribution P from which the stimulus $\boldsymbol{s}$ is drawn. Conceptually, the brain performs some computation on $\boldsymbol{s}$ yielding representational curves $\Omega_i(\boldsymbol{s})$. These are multiplied by adaptive gains, g_i, to yields the tuning curves $h_i(\boldsymbol{s})$. The actual, single-trial neural responses are noisy emissions based on $h_i(\boldsymbol{s})$.

We theorise that neurons adapt their gains to maximise an objective function, $\mathcal{L}^0(\boldsymbol{g})$, that trades-off the metabolic cost of neural activity with the information conveyed by the responses (Levy and Baxter, 1996; Ganguli and Simoncelli, 2014):

$$\mathcal{L}^0(\boldsymbol{g}) := 2\mu I(\boldsymbol{n}; \boldsymbol{s}) - \sum_{i=1}^{N} \mathbb{E}[n_i] = \text{Information} - \text{Metabolic cost}, \tag{1}$$

$$\boldsymbol{g}^{\text{opt}} = \arg\max_{\boldsymbol{g}} \mathcal{L}^0(\boldsymbol{g}). \tag{2}$$

Here, $I(\boldsymbol{n}; \boldsymbol{s})$ is the mutual information between the stimulus and response. The second term penalises the average population spike count, as a measure of metabolic energy cost, and $\mu > 0$ controls the information-energy trade-off. This trade-off is illustrated in Fig. 1b. In the following sections we will optimise an approximation to $\mathcal{L}^0$ to make predictions about the behaviour of the optimal gains. Here we make a few remarks on our theoretical framework and its assumptions.

First, note that, because Ω_i depend deterministically on the stimulus $\boldsymbol{s}$, and in turn fully determine the statistics of n_i, we have the identity $I(\boldsymbol{n}; \boldsymbol{s}) = I(\boldsymbol{n}; \boldsymbol{\Omega})$. This means that the coding fidelity term in the objective function can equivalently be interpreted, without any reference to the low-level stimulus, as the mutual information between the population's noisy responses, $\boldsymbol{n}$, and its ideal, noise-free outputs (up to scale), $\boldsymbol{\Omega}$, as determined by the computational goals of the circuit.

Next, note that adjusting gains does not restrict the computations which can be performed by downstream populations, as adjustments in gains can be compensated for by reciprocal adjustments in readout weights. Consider a downstream neuron which receives synaptic input from the population with synaptic weights w_i. The response of this neuron is then a nonlinear function of its total input $\sum_j w_j h_j(\boldsymbol{s})$, which, given our decomposition, can be written as $\sum_j w_j g_j \Omega_j(\boldsymbol{s})$. We therefore see that any modification of the gains, g_i, can be compensated for by a modification of the synaptic weights, w_i, by keeping the product of the two the same. This demonstrates that fixing Ω_i and optimising $\mathcal{L}^0$ in the gains provides no additional constraints on what downstream neurons encode in their responses. Thus, we not only do not constrain the representations of the population in consideration, but also do not constrain the representations of downstream read-out populations.

Finally, note that reciprocal adjustment of synaptic weights with the gains (in such a way to keep $w_j g_j$ constant) will keep constant the synaptic currents and therefore also the metabolic cost of synaptic transmission. This justifies our choice not to explicitly include synaptic transmission costs in our objective function, and take the metabolic cost term to be linear in the firing rates of the population.

2.2 Poisson-like noise model and upper bound to mutual information

We start by considering the simplest case of an uncorrelated Gaussian noise model with Poisson-like scaling of variance (*i.e.*, with unit Fano factor). According to this noise models individual neuron responses, n_i, are conditionally independent with conditional distribution

$$n_i|\boldsymbol{s} \sim \mathcal{N}(h_i(\boldsymbol{s}), h_i(\boldsymbol{s})). \tag{3}$$

Later in Sec. 2.7, we show that, to a good approximation, the solutions for this model extend to the case of true Poisson noise.

We consider the problem of calculating the optimal gains $\{g_j\}_{j=1}^N$ by maximising our target functional $\mathcal{L}^0$ (Eq. (1)). Unfortunately, analytic maximisation of mutual information is intractable. In the tradition of efficient coding theory (Brunel and Nadal, 1998; Ganguli and Simoncelli, 2014; Linsker, 1988), we will therefore consider optimising a bound on $\mathcal{L}^0$. We decompose the mutual information as $I(\boldsymbol{s};\boldsymbol{n}) = H[\boldsymbol{n}] - H[\boldsymbol{n}|\boldsymbol{s}]$. The marginal entropy term $H[\boldsymbol{n}]$ can be upper bounded by the entropy of a Gaussian with the same covariance,

$$H[\boldsymbol{n}] \leq \frac{1}{2}\ln\det(2\pi e\,\mathrm{Cov}[\boldsymbol{n}]). \tag{4}$$

Making this replacement gives us a new objective function

$$\mathcal{L}(\boldsymbol{g}) = 2\mu\left(\frac{1}{2}\ln\det(2\pi e\,\mathrm{Cov}[\boldsymbol{n}]) - H[\boldsymbol{n}|\boldsymbol{s}]\right) - \sum_{j=1}^{N}\mathbb{E}[n_j] \geq \mathcal{L}^0(\boldsymbol{g}), \tag{5}$$

and we assume the optimal gains $\boldsymbol{g}^{\mathrm{opt}}$ maximize $\mathcal{L}(\boldsymbol{g})$. Note that $\mathcal{L}$ depends tacitly on $P(\boldsymbol{s})$ and the tuning curves $\boldsymbol{\Omega}(\boldsymbol{s})$. However, one can show (see App. A.1) that $\mathcal{L}$ depends on P and $\boldsymbol{\Omega}$ only through the following collection of parameters:

1. The neurons' pre-modulated average response (*i.e.*, the average spike count before multiplication by the adaptive gain g_j)
$$\omega_j := \mathbb{E}[\Omega_j(\boldsymbol{s})] = \int \Omega_j(\boldsymbol{s})P(\boldsymbol{s})d\boldsymbol{s}, \tag{6}$$
2. The coefficient of variation of $\Omega_j(\boldsymbol{s})$, or equivalently of $h_j(\boldsymbol{s})$:
$$\mathrm{CV}_j := \frac{\sqrt{\mathrm{Var}[\Omega_j(\boldsymbol{s})]}}{\mathbb{E}[\Omega_j(\boldsymbol{s})]} = \frac{\sqrt{\mathrm{Var}[h_j(\boldsymbol{s})]}}{\mathbb{E}[h_j(\boldsymbol{s})]}, \tag{7}$$
3. The pairwise signal correlations, *i.e.*, the matrix (denoted by ρ) of Pearson correlation coefficients of the vector $\boldsymbol{\Omega}(\boldsymbol{s})$ (or equivalently $\boldsymbol{h}(\boldsymbol{s})$)
$$\rho_{ij} := \frac{\mathrm{Cov}[\Omega_i(\boldsymbol{s}),\Omega_j(\boldsymbol{s})]}{\sqrt{\mathrm{Var}[\Omega_i(\boldsymbol{s})]\ \mathrm{Var}[\Omega_j(\boldsymbol{s})]}} = \frac{\mathrm{Cov}[h_i(\boldsymbol{s}),h_j(\boldsymbol{s})]}{\sqrt{\mathrm{Var}[h_i(\boldsymbol{s})]\ \mathrm{Var}[h_j(\boldsymbol{s})]}}. \tag{8}$$

As we show in App. A.1, in terms of these parameters, $\mathcal{L}(\boldsymbol{g})$ is given by (up to an additive constant):

$$\mathcal{L}(\boldsymbol{g}) = \mu\,\log\det(I_N + P\,\omega g) - \boldsymbol{g}\cdot\boldsymbol{\omega}, \tag{9}$$

where $\cdot$ denotes the inner product of vectors, I_N is the $N\times N$ identity matrix, and we defined the $N\times N$ matrix

$$P := \mathrm{CV}\,\rho\,\mathrm{CV}. \tag{10}$$

Here and below, we have adopted the convention that the non-bold version (with no indices) of a bold symbol representing a vector denotes a diagonal matrix made from that vector; thus, *e.g.*, g, ω, and CV represent, diagonal matrices with the elements of the vectors $\boldsymbol{g}$, $\boldsymbol{\omega}$, and $\mathbf{CV}$ on their diagonal, respectively.

In the next section we will study the properties of the optimal gains, maximizing Equation (9), in clustered neural populations where neurons belong to clusters or types with similar stimulus tuning.

2.3 Clustering neurons based on similarity of response

In cortex, it is found that many neurons located spatially nearby to one another also have very similar response properties (Obermayer and Blasdel, 1993). We will call a collection of neurons which have very similar representational curves (but not necessarily similar gains) a *cluster*. To gain intuition and allow for analytic solutions, we will henceforth adopt a toy model in which the neurons in the population are sorted into K such clusters, each containing k neurons (thus $N = Kk$). In this toy model, neurons within a cluster have very similar stimulus tuning, but neurons in different clusters are tuned differently. This approximates a more realistic regime in which there is a more gradual transition from similar to dissimilar tuning curves within the entire population.

More concretely, we treat the representational curves of the neurons in cluster a as small perturbations to a cluster-wide representational curve $\hat{\Omega}_a(\boldsymbol{s})$. Signal correlations between neurons within a cluster are therefore close to 1. Similarly, signal correlations between neurons in distinct clusters a and b are approximately given by the correlation between $\hat{\Omega}_a(\boldsymbol{s})$ and $\hat{\Omega}_b(\boldsymbol{s})$, which we denote by $\hat{\rho}_{ab}$ (which jointly form a $K \times K$ matrix, $\hat{\rho}$). More precisely, we assume the correlation between two neurons i and j, belonging to clusters a and b, respectively, is given by

$$\rho_{ij} = \hat{\rho}_{ab} - \epsilon\, T_{ij}, \tag{11}$$

where ϵ is a small parameter controlling the deviation from perfect clusters, and T is an arbitrary symmetric $N \times N$ matrix (subject only to the requirement that the full matrix, ρ, remains a valid correlation matrix). We will analogously denote the mean and coefficient of variation of $\hat{\Omega}_a(\boldsymbol{s})$ by $\hat{\omega}_a$ and $\hat{\mathrm{CV}}_a$ (defined as in Eqs. (6), (7), but with Ω's replaced with $\hat{\Omega}$'s), and assume that the coefficients of variation and stimulus averages of the representational curves of all neurons in cluster a are close to $\hat{\mathrm{CV}}_a$ and $\hat{\omega}_a$, respectively.

In App. A.2, we show that for such clustered populations, the objective function Eq. (9) is given by

$$\mathcal{L}(\boldsymbol{g}) = \hat{\mathcal{L}}(\hat{\boldsymbol{g}}) + \epsilon \mathcal{L}_\epsilon(\boldsymbol{g}) + \mathcal{O}(\epsilon^2), \tag{12}$$

where the cluster gain $\hat{g}_a \equiv \sum_{i \in C_a} g_i$ is the sum of the gains of the neurons belonging to cluster a, and

$$\hat{\mathcal{L}}(\hat{\boldsymbol{g}}) = \mu \ln\det\Big(I_K + \hat{P}\hat{\omega}\hat{g}\Big) - \sum_{a=1}^{K} \hat{\omega}_a \hat{g}_a. \tag{13}$$

Here, I_K denotes the $K \times K$ identity matrix, and we defined

$$\hat{P} = \hat{\mathrm{CV}}\, \hat{\rho}\, \hat{\mathrm{CV}}. \tag{14}$$

Note that $\hat{\mathcal{L}}(\boldsymbol{g})$ is identical in form to the expression in Eq. (9), with all quantities relating to individual neurons replaced by the corresponding quantities for a cluster. This is because for clusters of identically-tuned and independently-firing Poisson neurons (corresponding to $\epsilon = 0$), the cluster can be considered as a single coherent unit whose firing rate is given by the sum of the firing rates of its neurons. Furthermore, in this approximation, the efficient coding objective function is indifferent to the precise distribution of single-cell firing rates (controlled by individual g_i) among the neurons of a cluster, as long as the total rate of the cluster (controlled by $\hat{g}_a$) is held fixed.

We therefore approximate the maximisation of the total objective function $\mathcal{L}$ as follows. We first specify the cluster firing rates by maximising $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$. We then specify the firing rates of individual neurons within a cluster by maximising $\mathcal{L}_\epsilon(\boldsymbol{g})$, subject to the constraint that the neuron firing rates in each cluster must sum to that cluster's previously optimised total firing rate. Although this is not in general the same as maximising the total objective function $\mathcal{L}$, this approximation is justified since terms of order ϵ and above have negligible effect on determining the cluster firing rates, but (as we will see in Sec. 2.5) are crucial for breaking the degeneracy within an individual cluster.

2.4 Optimal cluster gains display homeostasis across environments

To obtain an analytic approximation for the cluster gains, we apply an additional condition, discussed further in Sec. 2.6, that

$$\Delta \equiv (\mu \hat{\mathrm{CV}}\, \hat{\rho}\, \hat{\mathrm{CV}})^{-1} \ll 1. \tag{15}$$

The matrix Δ^{-1} characterises the structure of signal-to-noise ratio (SNR) in the space of population responses, and so our condition can be interpreted as the requirement that signal-to-noise ratio is strong. In App. A.3 we derive perturbative solutions for the cluster gains that maximise Eq. (13). In the first approximation, *i.e.*, to zeroth order in Δ, we find

$$\hat{g}_a \approx \frac{\mu}{\hat{\omega}_a} := \hat{g}_a^{(0)}. \tag{16}$$

To see the significance of this result, consider the average response of cluster a, *i.e.*, the stimulus-averaged total spike count of that cluster. Letting C_a denote the set of indices for neurons belonging to cluster a, this average response is given by

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in C_a} n_i\right] &= \sum_{i\in C_a} \mathbb{E}[h_i(\boldsymbol{s})] = \sum_{i\in C_a} \mathbb{E}[g_i \Omega_i(\boldsymbol{s})] \\ &= \sum_{i\in C_a} g_i \omega_i \approx \hat{g}_a \hat{\omega}_a \approx \mu.\end{aligned}$$

Thus, in the leading approximation, the stimulus-average spike count of cluster a is given by the constant μ. Importantly, this is constant across *both* clusters and environments. In particular, the zeroth order solution yields *homeostasis* of cluster firing rates. This is depicted in Fig. 3a.

We also note that, as long as the SNR structure, Δ, remains fixed between two environments, optimal gains imply exact homeostasis of each cluster's average firing rate between environments, irrespective of the size of Δ and how pre-adaptation mean rates, $\hat{\omega}_a$, change between the two environments. This is true regardless of whether condition Eq. (15) holds; however, unless Δ is small, firing rates are not equalised across clusters in the optimal solution, in general.

We also calculated the first order correction to the homeostatic solution. This is given by

$$\hat{g}_a \approx \frac{\mu}{\hat{\omega}_a}\left(1 - \frac{[\hat{\rho}^{-1}]_{aa}}{\mu \hat{\mathrm{CV}}_a^2}\right) = \frac{\mu}{\hat{\omega}_a}(1 - \Delta_{aa}) := \hat{g}_a^{(1)}. \tag{17}$$

which (assuming Δ_{aa} are small) yields approximate homeostasis, with small deviations depending on the correlational structure between clusters and how that structure shifts between environments.

So far, we have not assumed any kind of homeostasis in order to arrive at these results; rather, the approximate homeostatic solution has arisen from our optimisation problem. In 2.8, we additionally investigate solutions which, by construction, yield homeostasis both across environments and clusters. Such solutions, which we denote by $\hat{g}_a^{hom}$, must thus be of the form Eq. (16), but with the constant μ replaced by an *a priori* unknown constant scalar, χ. In other words,

$$\hat{g}_a^{hom} = \frac{\chi}{\hat{\omega}_a}, \tag{18}$$

such that χ is equal to the average spike-count per cluster per coding interval. The variable χ is then chosen to optimise the expectation of $\hat{\mathcal{L}}$, Eq. (13), across a family of environments. In App. A.9, we further demonstrate that, with the optimal χ, the average spike-count of the entire population of clusters is approximately the same under this solution as it is under $\hat{g}^{(1)}$, and that this average is strictly smaller than μ. In Sec. 2.8, we numerically compare the performance of $\hat{g}^{(0)}, \hat{g}^{(1)}$, and $\hat{g}^{hom}$, and find that in many cases, the homeostatic $\hat{g}^{hom}$ performs close to the true (numerically computed) optimal gains.

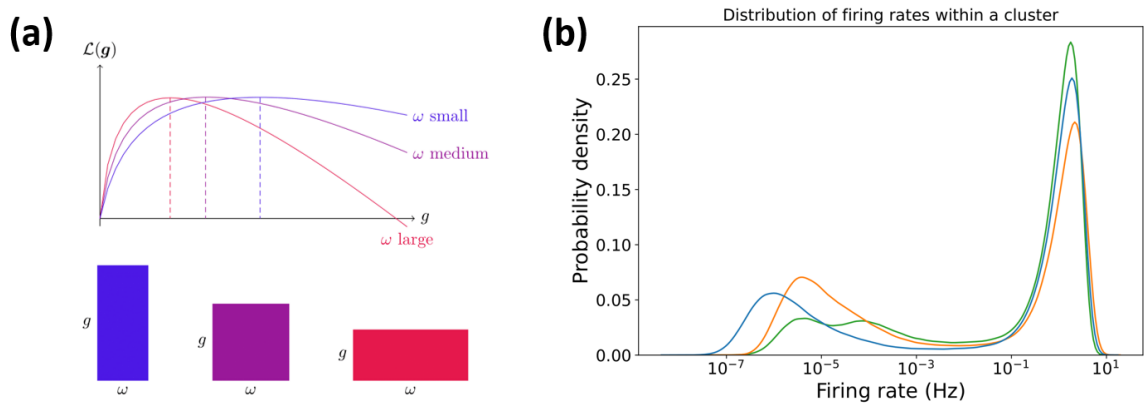

Figure 3: (a) Firing rate homeostasis at the level of clusters of similarly tuned neurons. We consider three environments, indexed by colour. Across these environments, the pre-modulated cluster rate $\hat{\omega}$ is different, as is the functional $\hat{\mathcal{L}}(\hat{g})$. However, the argmax of the target function is so as to keep the product $\hat{\omega}\hat{g}$ approximately constant (here depicted as area of the corresponding rectangle). Since this product is the cluster firing rate, this strategy yields homeostasis. (b) The distribution of log firing rates within a cluster of neurons. The densities obtained here were obtained by averaging over $10,000$ random draws of the within-cluster correlation matrix ρ using the method described in Sec. 3.1. This method uses two parameters, (α, q). The values of (α, q) were as follows: blue $(10, 0.7)$; orange $(10, 0.9)$; green $(20, 0.9)$. We fix $\hat{g} = 10$ and $k = 100$ throughout.

2.5 Distribution of firing rates within neuronal clusters

Having derived an analytic approximation to the cluster firing rates $\hat{\boldsymbol{g}}$, we next maximise the perturbation term $\mathcal{L}_\epsilon(\boldsymbol{g})$ in Eq. (12), subject to the constraint that the sum of gains within a cluster equals the obtained cluster solution, *i.e.*, $\sum_{i \in C_a} g_i = \hat{g}_a$. Since the term $\hat{\mathcal{L}}$ depends only on the sum of the gains in a cluster, there is a great deal of redundancy in potential optimal solutions of single-neuron gains. The term $\mathcal{L}_\epsilon(\boldsymbol{g})$ acts to break symmetry within the cluster and yield a unique optimal solution. However, since $\mathcal{L}_\epsilon(\boldsymbol{g})$ is small and relatively flat, we expected this symmetry breaking yields a solution with broadly heterogeneous mean firing rates across the neurons in a cluster, despite approximately equal total firing rate across clusters.

To first order in ϵ and zeroth order in Δ, individual neural gains in the a-th cluster are given by the following constrained maximisation problem (see derivation in App. A.4), which is a quadratic program in $\boldsymbol{g}$:

$$\min_{\boldsymbol{g}} \sum_{i,j \in C_a} g_i \rho_{ij} g_j \tag{19}$$

subject to the constraints

$$\sum_{i \in C_a} g_i = \hat{g}_a, \qquad g_i \geq 0. \tag{20}$$

Just as the correlation between clusters was the dominant force in determining the firing rates for each cluster, the dominant force in determining firing rates within a cluster is the correlational structure of the neurons *within* the cluster, and not correlations *between* clusters. Note that we can interpret this quadratic program as aligning the vector of mean rates within a cluster with the direction of minimal signal correlation, thereby maximising the information conveyed by the population.

We solved Eq. (19) numerically for correlation matrices ρ which were generated to have a high level of correlation within clusters; specifically, within each cluster ρ has a large eigenvalue with eigenvector closely aligned to the vector with all components equal to 1 (see Sec. 3.1). The firing rate distributions resulting from the numerically optimised gains are shown in Fig. 3b, with the three curves corresponding

to different parameter values governing the within-cluster correlation matrix ρ. Firstly, we see that maximising $\mathcal{L}_\epsilon$ leads to a broad range of firing rates within each cluster. In particular, the distribution of rates spans multiple orders of magnitude, in agreement with cortical observations (*e.g.* Slomowitz et al. (2015); Hengen et al. (2013); Maffei and Turrigiano (2008)). Furthermore, the distributions we found are bimodal. The more prominent peak yields an approximately log-normal distribution consistent with empirical findings in cortex (Buzsáki and Mizuseki, 2014). Given that the smaller mode occurs between extremely small firing rates, 10^{-3} to 10^{-7} Hz, it is reasonable to expect neurons in this mode to be essentially silent during experimental observations and remain undetected. Our theory therefore predicts that a significant but variable fraction of cortical neurons are silent at any given sensory environment, and that which neurons are silent will shuffle following shifts in the environment (and subsequently changes in ρ).

2.6 Conditions for the validity of the homeostatic solution hold in cortex

In this section, we consider conditions under which the homeostatic solution, Eq. (16), provides a good approximation to the optimal gains. We obtained this solution, and the first order correction to it, Eq. (17), by expanding in Δ (Eq. (15)). The rigorous mathematical condition for the convergence of the full series expansion requires that Δ is small in norm (*i.e.*, that its largest eigenvalue is small). Here, we instead consider conditions under which the leading (first order in Δ) correction to the homeostatic solution is small in comparison. This is equivalent to the smallness of the mean eigenvalue of Δ, given by $\mathrm{tr}(\Delta)/K$: note that to first order $g_a^{(1)} = g_a^{(0)} - g_a^{(0)}\Delta_{aa}$, and if the eigenbasis of Δ is unaligned with the standard basis, we expect the diagonal elements Δ_{aa} to be clustered around $\mathrm{tr}(\Delta)/K$. Thus, smallness of $\mathrm{tr}(\Delta)/K$ guarantees that the first order correction terms will be small, such that $g_a^{(1)} \approx g_a^{(0)}$. Using Eq. (15), we can express this condition as

$$\frac{1}{K}\mathrm{tr}(\Delta) = \frac{1}{K}\sum_{a=1}^{K} \frac{\hat{\rho}_{aa}^{-1}}{\mu \hat{\mathrm{CV}}_a^2} \ll 1 \tag{21}$$

As we now show, the smallness of the expression in Eq. (21) requires that firing rates of each cluster are sufficiently high, each cluster is highly selective in its response to stimuli, and neural responses correspond to a high dimensional geometry. We will also discuss biological estimates for $\mu, \hat{\mathrm{CV}}$, and $\hat{\rho}$, and whether they support the above conditions.

- We see from the zeroth order solution (Eq. (16)) that μ is approximately equal to the average total spike count of the cluster over the rate-coding time interval. Condition (15) thus requires that the stimulus-averaged firing rate of all clusters are sufficiently high. A wide range of mean firing rates for individual neurons have been reported in cortex. Here we focus on firing rates in rodent V1 during free behaviour. Reported values tend to range from 0.4 Hz (Greenberg et al., 2008) to 14 Hz (Parker et al., 2022) with other values lying in a tighter range of 4-7 Hz (Hengen et al., 2013; Szuts et al., 2011; Torrado Pacheco et al., 2019). Therefore, for rodent V1, we take a mean firing rate of 5 Hz to be a reasonable rough estimate. Assuming a coding interval of 0.1 seconds, and a cluster size of $k = 20$ (as a guess for the number of similarly coding V1 neurons), we obtain $\mu \approx 10$. Note that for lower or higher single neuron firing rates, a similar value of μ could be achieved by correspondingly scaling cluster sizes up or down.

- The coefficient of variation $\hat{\mathrm{CV}}$ can be seen as a measure of sparseness of responses. To see this, consider a toy model in which the cluster responds at fixed level to a fraction p_a of stimuli and is silent otherwise. In this case, $\hat{\mathrm{CV}}_a^2 = (1-p_a)/p_a \approx 1/p_a$ for small p_a. Our condition therefore requires that neurons are sufficiently selective in their responses and respond only to a small fraction of stimuli. Lennie (2003) places the fraction of simultaneously active neurons (which we use as a proxy for the response probability of a single cluster) at under 5%. For our toy model, this yields an estimate of $\hat{\mathrm{CV}}^2 \approx 20$. The Bernoulli distribution in the toy model is particularly sparse, and so we take $\hat{\mathrm{CV}}^2 \approx 10$ as a more conservative estimate.

- $\hat{\rho}$ is the signal correlation matrix of neuron clusters. It can be shown that $[\hat{\rho}^{-1}]_{aa} \geq 1$, with equality if and only if cluster a has zero signal correlation with every other cluster. $[\hat{\rho}^{-1}]_{aa}$ can thus be seen as a measure of the extent to which cluster a shares its representation with other clusters, *i.e.*, as a measure of representational redundancy. Many traditional efficient coding accounts predict zero signal correlation between neurons, at least in the low noise limit (Barlow, 2012; Nadal and Parga, 1994), providing an additional normative justification for low signal correlations. This is known as a factorial code (Nadal and Parga, 1999).

  However, complete lack of signal correlations is not necessary for our condition to hold; we merely require a sufficiently slow decay of the eigenvalue spectrum of the signal correlation matrix. This condition is geometrically equivalent to neural responses forming a high-dimensional representation of stimuli. Stringer *et al.* (Stringer et al., 2019) found that, for high-dimensional (ecological) stimuli, the signal correlation matrix possesses a $1/n$ spectrum. Again, assuming that the eigenbasis of $\hat{\rho}$ does not align with the standard basis, we expect the values of $\hat{\rho}^{-1}_{aa}$ to be clustered around $\mathrm{tr}(\hat{\rho}^{-1})/K$. In this case, for K large, we obtain the estimate $\hat{\rho}^{-1}_{aa} \approx \mathrm{tr}(\hat{\rho}^{-1})/K \approx \ln(K)/2$ (see App. A.5). Given the slow logarithmic increase with K, we expect that our condition can hold even for very large neural populations. For example, suppose we take the entire human V1 as our neural population, to obtain a reasonable upper bound for K. This contains roughly 1.5×10^8 neurons (Wandell, 1995), leading to 7.5×10^6 clusters of 20 neurons, and therefore an average value of $[\hat{\rho}^{-1}]_{aa}$ just under 8.

The above analysis makes it clear when we should expect homeostasis in general – when cluster firing rates are not too *low*, responses are highly *selective*, and signal correlation structure corresponds to a *high-dimensional geometry* (Stringer et al., 2019). On the other hand, when these conditions are violated, the optimal gain configuration can deviate strongly from the homeostatic solution.

Moreover, given the above estimate for μ, $\hat{\mathrm{C}}\mathrm{V}$, and $[\hat{\rho}^{-1}]_{aa}$, we thus see that in V1 and possibly other cortical areas, we can expect the leading corrections to the homeostatic solution (see Eq. (17)) to not exceed 10%, in relative terms. Indeed, as we will show below using numerical validations (Sec. 2.8), for values of the above parameters in their estimated biological range, the homeostatic solution does provide a very good approximation.

2.7 Generalisations to alternative noise models

Here we move beyond the simple case of unit Fano factor Gaussian noise to other noise models. In particular, we consider first Poissonian noise, followed by correlated noise with possibly non-Poissonian (but power-law) scaling of signal to noise ratio.

In Sec. 2.4 we chose the cluster gains $\hat{\boldsymbol{g}}$ to maximise $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$. Therefore, to extend our analysis, the definition of the functional $\hat{\mathcal{L}}$ will need to be generalised. To do so, we recall that in Sec. 2.3, we demonstrated that (for Gaussian noise) $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$ arises from applying the Gaussian upper bound to the information-energy trade-off applied to the cluster responses $\hat{\boldsymbol{n}}$. Therefore, for other noise models, we take this as the definition of $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$.

Poisson noise model. The toy model of Gaussian noise is unrealistic for two main reasons: spike counts are discrete, while Gaussian variables are continuous; and spike counts are non-negative, while Gaussian variables can be negative. A more realistic noise model is given by Poisson spike counts,

$$n_j|\boldsymbol{s} \sim \mathrm{Poisson}\left(h_j(\boldsymbol{s})\right). \tag{22}$$

Since the sum of independent Poisson random variables is also Poisson, we arrive at the following distribution for cluster spike counts:

$$\hat{n}_a|\boldsymbol{s} \sim \mathrm{Poisson}\left(\hat{g}_a\hat{\Omega}_a(\boldsymbol{s})\right). \tag{23}$$

To find $\hat{\mathcal{L}}$, we consider the Gaussian upper bound to the information-energy trade-off applied to the cluster responses. A slight subtlety arises from the fact that the entropy for a Poisson random variable is discrete and the Gaussian upper bound is for a continuous random variable. However, we show in App. A.7 that in this case the upper bound Eq. (4) can be replaced by the similar bound

$$H[\hat{\boldsymbol{n}}] \leq \frac{1}{2} \ln \left((2\pi e)^K \det \left(\frac{1}{12} I_K + \mathrm{Cov}(\hat{\boldsymbol{n}}) \right) \right). \tag{24}$$

We will further make an approximation by neglecting the $I_K/12$ term, which under the relevant conditions is negligible compared to the covariance term. Specifically, the variance for each cluster (*i.e.*, a diagonal element of $\mathrm{Cov}(\hat{\boldsymbol{n}})$) is at least μ, which under our biological estimates for μ, dominates 1/12 (Sec. 2.6). We make one additional assumption on the representational curves $\hat{\Omega}_a$. Specifically, we assume that these representational curves have a baseline, and never fall down to zero (see App. A.7, Eq. (67) for the precise form of this assumption). We show in App. A.7 that under this condition, the functional $\hat{\mathcal{L}}(\hat{g})$ becomes identical to that obtained in the Gaussian noise case (see Sec. 2.4, Eq. (9)). Therefore we have the same optimal gains, and in particular the leading-order homeostatic approximation generalises to the Poissonian noise model, under the above assumptions.

This analysis reveals why the Gaussian noise model assumed in earlier sections is far less restrictive than it originally appears. Qualitatively, Poisson random variables can be approximated as sums of independent Bernoulli random variables. According to the Central Limit Theorem, for large enough rates, these sums behave like a normal random variable, giving rise to approximately Gaussian noise. Likewise, the spike count of a cluster of neurons can be considered as the sum of a large collection of (mostly) independent random variables (*i.e.* the spike count of individual neurons). Therefore, whatever the conditional distribution (conditioned on the stimulus) of individual neuronal spike counts, we expect the conditional distribution of a cluster spike count to be approximately Gaussian, at least in the case of low enough noise correlations and large enough clusters. This demonstrates that, as long as noise correlations are small, the Gaussian noise case for clusters covers all well-behaved individual neuron distributions, at least approximately.

Power-law and correlated noise Here we extend our analysis to the case of correlated noise with a general, power-law scaling of response variance with trial-average response. We now adopt the following noise model for cluster responses[1]

$$\hat{\boldsymbol{n}} | \boldsymbol{s} \sim \mathcal{N} \left(\hat{\boldsymbol{h}}(\boldsymbol{s}), \sigma^2 \hat{h}(\boldsymbol{s})^\alpha \, \Sigma(\boldsymbol{s}) \, \hat{h}(\boldsymbol{s})^\alpha \right), \tag{25}$$

where $\Sigma(\boldsymbol{s})$ is the stimulus-dependent noise correlation matrix, and $0 < \alpha < 1$ is a scaling parameter. We define $\beta = 2(1-\alpha)$. Note that $\sigma^2 = 1$, $\alpha = \frac{1}{2}$ yields the previously considered special case of unit Fano factor noise scaling, and $\Sigma(\boldsymbol{s}) = I$ yields uncorrelated noise. On the other hand, the $\alpha \to 1^-$ limit, the model becomes pathological, as the noise scales at the same rate as the signal, and the information term becomes degenerate and independent of neural gains; the energetic term then forces all neural firing rates to collapse down to zero. We therefore consider this case unrealistic; indeed empirical data is consistent with sublinear scaling of noise strength with signal strength even in super-Poisson conditions (Goris et al., 2014).

We will once again expand to zeroth and first order in a parameter that scales with inverse signal-to-noise ratio, which we require is small. In this case, we have the condition

$$\Delta := \sigma^2 \left((\beta\mu)^\beta \, \hat{\mathrm{CV}}(\alpha) \, \hat{\rho} \, \hat{\mathrm{CV}}(\alpha) \right)^{-1} W \ll 1, \tag{26}$$

[1]Note that we make this assumption directly for the noise in the total firing rate of a cluster. While in the case of Poisson noise and Gaussian noise with unit Fano factor, such a noise model for clusters follows from a single-neuron model with similar statistics (*i.e.*, Poission or unit-Fano factor Gaussian, respectively), this is not true for the general case of power-law noise that we consider here.

where the matrix W, with elements

$$W_{ab} = \frac{\mathbb{E}\left[\hat{\Omega}_a(\boldsymbol{s})^\alpha \Sigma_{ab}(\boldsymbol{s}) \hat{\Omega}_b(\boldsymbol{s})^\alpha\right]}{\sqrt{\mathbb{E}\left[\hat{\Omega}_a(\boldsymbol{s})^{2\alpha}\right] \mathbb{E}\left[\hat{\Omega}_b(\boldsymbol{s})^{2\alpha}\right]}} = \frac{\mathbb{E}\left[\hat{h}_a(\boldsymbol{s})^\alpha \Sigma_{ab}(\boldsymbol{s}) \hat{h}_b(\boldsymbol{s})^\alpha\right]}{\sqrt{\mathbb{E}\left[\hat{h}_a(\boldsymbol{s})^{2\alpha}\right] \mathbb{E}\left[\hat{h}_b(\boldsymbol{s})^{2\alpha}\right]}}, \tag{27}$$

is an effective stimulus-averaged noise correlation matrix, and $\hat{\mathrm{CV}}_a(\alpha)$ is the "α-coefficient of variation" defined by

$$\hat{\mathrm{CV}}_a(\alpha) = \hat{\mathrm{CV}}_a \frac{\hat{\omega}_a^\alpha}{\sqrt{\mathbb{E}\left[\hat{\Omega}_a(\boldsymbol{s})^{2\alpha}\right]}} = \hat{\mathrm{CV}}_a \frac{\mathbb{E}\left[\hat{h}_a(\boldsymbol{s})\right]^\alpha}{\sqrt{\mathbb{E}\left[\hat{h}_a(\boldsymbol{s})^{2\alpha}\right]}}. \tag{28}$$

Note that W is truly a correlation matrix, in the sense of being positive definite and having unit diagonal elements. Note also, that $\hat{\mathrm{CV}}(\alpha)$ and W are independent of neural gains. Moreover, $\hat{\mathrm{CV}}(\alpha)$ is a decreasing function of α, with $\hat{\mathrm{CV}}(1/2) = \hat{\mathrm{CV}}$.

Expanding in Δ as before (see appendix A.8) we arrive at the following zeroth and first order expressions:

$$\hat{g}_a^{(0)} = \frac{\beta\mu}{\hat{\omega}_a}, \tag{29}$$

and

$$\hat{g}_a^{(1)} = \frac{\beta\mu}{\hat{\omega}_a}\left(1 - \Delta_{aa}\right). \tag{30}$$

Note that, since $\hat{g}_a^{(0)}\hat{\omega}_a = \beta\mu$, we once again have homeostasis of firing rates at zeroth order. This constitutes a strong generalisation of our previous result. In particular, we have demonstrated that that constant Fano factor scaling of noise and uncorrelated noise are not necessary to yield homeostasis as an optimal solution; correlated, power-law scaled noise also suffices, provided conditions similar to those discussed in Sec. 2.6 hold.

Much as before, condition (26) is a requirement that we are in a high signal-to-noise condition. Note that we recover the previous expression for Δ when $\beta = 1$, $\sigma^2 = 1$ and $W = I_K$. More generally, the qualitative dependence of Δ on α (or equivalently, β), σ, and W is as follows. Firstly, note that Δ scales linearly with σ^2; thus, if noise is scaled up, so is Δ. Next, consider fixing σ^2 and W and increasing the noise scaling α. This decreases both $\hat{\mathrm{CV}}(\alpha)$ and $(\beta\mu)^\beta$ provided $\alpha \leq 1 - e/2\mu \approx 0.87$. Therefore, in this regime, increasing α also increases Δ. Lastly, the dependency of Δ on W is complex, and depends on the spectra of both $\hat{\rho}$ and W as well as the alignment of their eigenbases (which represent the structure of signal and noise in the neural activity space, respectively). In Sec. 2.8, we will investigate these dependencies quantitatively.

As with the uncorrelated, unit-Fano-factor noise case discussed in Sec. 2.4, we additionally define a solution $\hat{g}_a^{hom}$ for which homeostasis across environments and clusters is enforced. As before, such gains take the form given in Eq. (18), *i.e.*, $\hat{g}_a^{hom} = \chi/\hat{\omega}_a$, where χ is chosen to maximise the expected value of $\hat{\mathcal{L}}$ across a family of environments. In Sec. 2.8, we compare the performance of the optimal homeostatic gains $\hat{g}^{hom}$ to both $\hat{g}^{(0)}$ and $\hat{g}^{(1)}$ using numerical simulations.

2.8 Numerical comparison of homeostatic gains to optimal gains

We used numerical simulations to test and compare the performance of the homeostatic gains $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{hom}$ (as well as the first-order correction $\hat{\boldsymbol{g}}^{(1)}$), as approximate analytical solutions to our efficient coding problem, with the numerically calculated optimal solution $\hat{\boldsymbol{g}}^{\mathrm{opt}}$. To capture the notion of adaptation under environmental shift, we consider a family of environments parameterised by $\epsilon \in [0, 1]$. As noted in Sec. 2.2, an environment (via the stimulus density $P(\boldsymbol{s})$) is specified by $\hat{\omega}_a$, $\hat{\rho}$, and $\hat{\mathrm{CV}}_a(\alpha)$. For simplicity, we took the coefficients of variation to be constant between across both clusters and environments; specifically we fixed $\hat{\mathrm{CV}}_a^2(\alpha) = 10$, independently of ϵ. We also fix $\mu = 10$ and $\sigma^2 = 1$,

and fixed the population size to $K = 10,000$ clusters. (Note that the values used here for μ and $\hat{\text{CV}}$ are consistent with our conservative biological estimates in Sec. 2.6.) Thus, below, variations across simulations occur only in our choice of $\hat{\boldsymbol{\omega}}$, $\hat{\rho}$, and W.

We obtain $\hat{\omega}_a$ and $\hat{\rho}$ by interpolating between the endpoints at $\epsilon = 0$ and $\epsilon = 1$. For each $a = 1, \ldots, K$, $\hat{\omega}_a(\epsilon = 0)$ and $\hat{\omega}_a(\epsilon = 1)$ are drawn independently from a wide distribution (specifically the beta distribution $\text{Beta}(6, 1)$). The value of $\hat{\omega}_a(\epsilon)$ at intermediate values of ϵ was then obtained by linear interpolation between the independently sampled boundary values. This method for sampling environments is illustrated in Fig. 4a. Similarly, to construct $\hat{\rho}(\epsilon)$, we first constructed $\hat{\rho}(\epsilon = 0)$ and $\hat{\rho}(\epsilon = 1)$ with independent random structures and smoothly interpolated between them (see Sec. 3.2 for the details), in such a way that, in line with the results of Stringer et al. (2019), the correlation matrices, $\hat{\rho}(\epsilon)$ (for all ϵ), were derived from covariance matrices with a $1/n$ eigenspectrum.

To provide further intuition about the model of environmental shift we have adopted, we examine the distribution of cluster firing rates within the population following an environmental shift directly from the $\epsilon = 0$ environment to the $\epsilon = 1$ environment. Fig. 4b shows the average cluster firing rates before and after adaptation to the environmental shift. Here, we have used an uncorrelated Poissonian noise model. Pre-adaptation gains are numerically optimised to the pre-shift stimulus statistics; post-adaptation gains are numerical optimised to the post-shift stimulus statistics. In general, we expect any reasonable adaptation scheme, even if sub-optimal, to be effective when environmental shifts are radical. Fig. 4b demonstrates that the model of environmental shift adopted here is not particularly extreme, with the bulk of firing rates changing by less than 10%. Thus, our simulations below (see Figs. 5–6) show that our homeostatic adaptation methods work well even for moderate (ecologically realistic) changes in the environment statistics.

Fig. 4c shows histograms of cluster firing rates before and following an environmental shift from $\epsilon = 0$ to $\epsilon = 1$. Immediately following the shift, and before adaptation, a broad distribution of cluster firing rates results (Fig. 4c middle). By contrast, the distribution of cluster firing rates both before the environmental change and after adaptation to the new environment (Fig. 4c top and bottom panels) are tightly concentrated at a value just below 100 Hz (assuming 20 neurons per cluster this translates to an average rate of 5 Hz for individual neurons). In particular, the sharp concentrations (around the same value) implies that, for any individual cluster, the adapted firing rates in both environments are very close, thereby exhibiting that the optimal gains yield firing rate homeostasis to a very good approximation (note that the cluster firing rates in these histograms and in Fig. 4b are calculated based on the numerically optimised gains, maximising Eq. (13), and not based on the analytical approximations $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{hom}$).

Below, for different classes of noise models, we numerically compare the performance of the zeroth order homeostatic code $\hat{g}_a^{(0)} = (\beta\mu)/\hat{\omega}_a$ (Eq. (29)), the first order correction $\hat{g}_a^{(1)}$ (Eq. (30)), and the homeostatic gains $\hat{g}_a^{hom} = \chi/\hat{\omega}_a$ (where χ is chosen as detailed in App. A.9) against gains $\hat{g}_a^{\text{opt}}$ which have been numerically optimised by performing gradient ascent on the objective $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$ (see Eq. (69), App. A.8). To compare the performance of each approximate solution, $\hat{\boldsymbol{g}}^{\text{app}}(\epsilon)$, we use the following *relative improvement* measure

$$C^{\text{app}}(\epsilon) = \frac{\hat{\mathcal{L}}(\hat{\boldsymbol{g}}^{\text{app}}(\epsilon); \epsilon) - \hat{\mathcal{L}}(\hat{\boldsymbol{g}}^{\text{app}}(0); \epsilon)}{\hat{\mathcal{L}}(\hat{\boldsymbol{g}}^{\text{opt}}(\epsilon); \epsilon) - \hat{\mathcal{L}}(\hat{\boldsymbol{g}}^{\text{app}}(0); \epsilon)}, \tag{31}$$

which can be interpreted as the improvement in $\hat{\mathcal{L}}(\cdot; \epsilon)$ (the efficient coding objective function in environment ϵ, and according to its statistics) achieved by the adaptive gains $\boldsymbol{g}^{\text{app}}(\epsilon)$ over the unadapted gains from the original environment $\boldsymbol{g}^{\text{app}}(0)$, relative to the improvement obtained by the optimally adaptive gains $\boldsymbol{g}^{\text{opt}}(\epsilon)$. The better the approximation the closer $C^{\text{app}}(\epsilon)$ will be to 1. We additionally consider the mean relative errors

$$\frac{1}{K} \sum_{a=1}^{K} \frac{|g_a^{\text{app}}(\epsilon) - g_a^{\text{opt}}(\epsilon)|}{g_a^{\text{opt}}(\epsilon)}, \tag{32}$$

as another quantification of our approximations to the optimal gains. To give a sense of the scale of variation of the optimal gains, we additionally plot the relative change in the optimal gains between

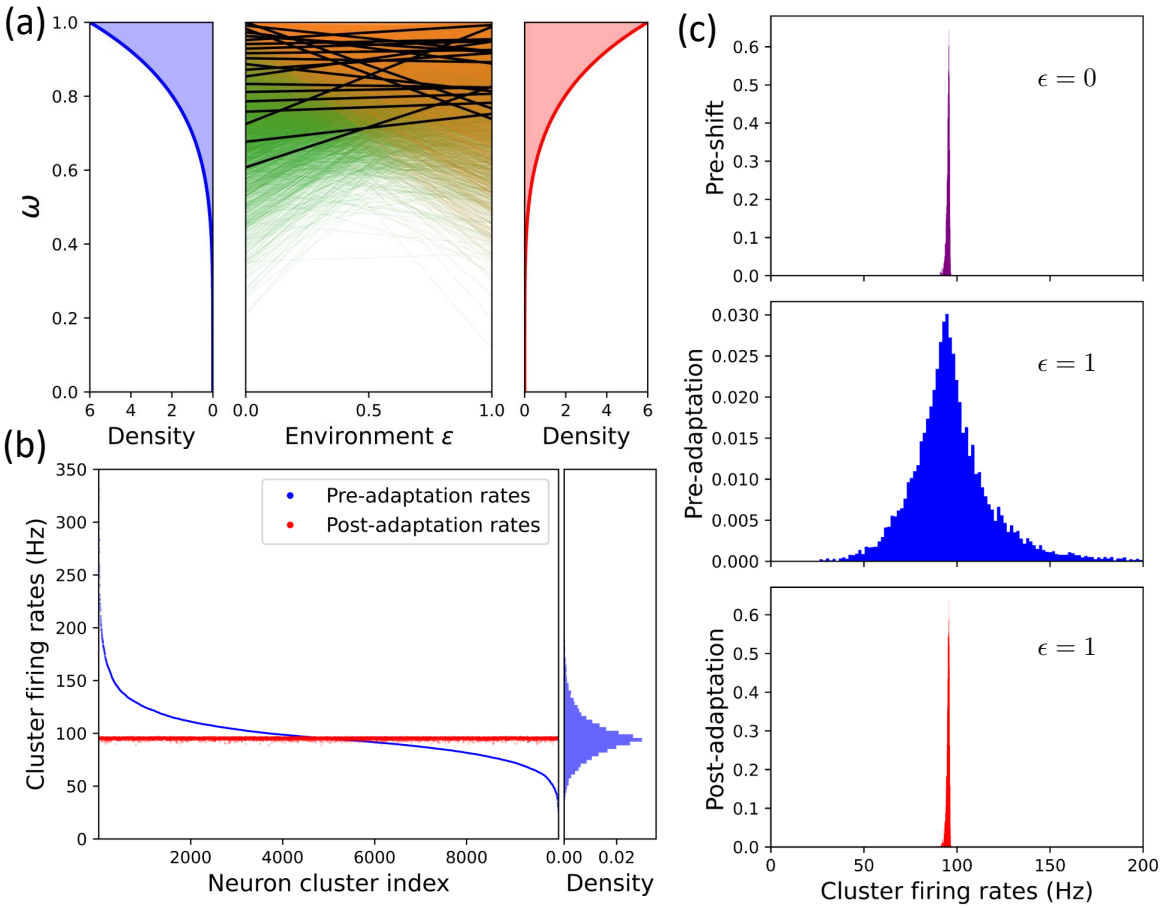

Figure 4: The simulations of environmental change and the effect of optimal gain adaptation on mean firing rates. (a) Model of environments. The pre-modulated average responses, $\hat{\omega}_a$, in a continuous sequence of environments parameterised by ϵ (going from $\epsilon = 0$ to 1) were generated as follows. For each cluster a, the $\hat{\omega}_a$ at $\epsilon = 0$ and 1 are first sampled independently from a Beta$(6, 1)$ distribution (shown in blue and red, respectively, on the left and right plots). These values are then linearly interpolated (middle plot) to obtain $\hat{\omega}_a(\epsilon)$. The lines in the middle panel are plots for $\hat{\omega}_a(\epsilon)$ for the model's 10000 clusters, and are colour coded from orange to green in terms of the rank order of $\hat{\omega}_a(\epsilon = 0)$. Every 500th line is coloured black. Panels (b) and (c) consider an abrupt change or "environment shift" from the $\epsilon = 0$ directly to the $\epsilon = 1$ environment (without traversing intermediate environments). (b) Cluster average firing rates before and after adaptation, following an environmental shift corresponding to the change $\hat{\boldsymbol{\omega}}(\epsilon = 0) \to \hat{\boldsymbol{\omega}}(\epsilon = 1)$ in pre-modulated mean cluster responses. The blue and red points in the left panel show the pre- and post-adaptation average rates in the new environment, $\epsilon = 1$, which are proportional to $\hat{g}_a^{\text{opt}}(\epsilon = 0)\hat{\omega}_a(\epsilon = 1)$ and $\hat{g}_a^{\text{opt}}(\epsilon = 1)\hat{\omega}_a(\epsilon = 1)$, respectively; clusters are ordered by their pre-adaptation firing rate. The right panel shows the histogram of the pre-adaptation firing rates. (c) Distribution of cluster firing rates before and after an environmental shift (directly from from $\epsilon = 0$ to $\epsilon = 1$). Top: the histogram of cluster firing rates adapted to the original environment before the shift ($\epsilon = 0$); these are proportional to $\hat{\omega}_a(\epsilon = 0)\hat{g}_a^{\text{opt}}(\epsilon = 0)$. Middle: the histogram of cluster rates immediately following the environmental shift (to $\epsilon = 1$), but before adaptation; these are proportional to $\hat{\omega}_a(\epsilon = 1)\hat{g}_a^{\text{opt}}(\epsilon = 0)$ (this is the same as that in the right margin of b). Bottom: the histogram of cluster rates after adaptation to the new environment; these rates are proportional to $\hat{\omega}_a(\epsilon = 1)\hat{g}_a^{\text{opt}}(\epsilon = 1)$ (this is therefore the histogram of the red dots in panel b). In (b) and (c), an uncorrelated Poissonian noise model was adopted for computing the optimal gains.

the end points,

$$\frac{1}{K}\sum_{a=1}^{K}\frac{|g_a^{\mathrm{opt}}(0)-g_a^{\mathrm{opt}}(1)|}{g_a^{\mathrm{opt}}(0)}, \tag{33}$$

This can also be interpreted as the effect size of optimal adaptation across environments on gains.

Uncorrelated power-law noise Consider first the case that the noise correlation matrix W is equal to the identity. We examine this scenario for a range of values for the parameter α (or equivalently $\beta = 2(1-\alpha)$) which controls the scaling of the noise (trial-to-trial variability) with the signal (trial average). In Fig. 5a, we plot the average relative errors of the approximate gains as a function of the noise scaling α (see Eq. (32)). (In the plots for the renormalised homeostatic solution, $\hat{\boldsymbol{g}}^{hom}$, Eq. (18), we used an analytical approximation to the optimal χ, derived in App. A.9, which means that the performance measures in the plots of Fig. 5 are a lower-bound for the solution with the truly optimal χ.) We see that the homeostatic solution $\hat{\boldsymbol{g}}^{hom}$, and the first order correction, $\hat{\boldsymbol{g}}^{(1)}$, have very low ($< 3\%$) relative errors for all noise scalings, while the relative error in $\hat{\boldsymbol{g}}^{(0)}$ becomes large for significantly super-Poissonian noise scaling, *i.e.*, $\alpha = 0.75$ (Fig. 5A). The relative improvement measure $C^{hom}(\epsilon)$ for the homeostatic solution is also close to one, *i.e.*, the optimal case, for all α, except for 0.75 (Fig. 5B). For the other approximate solutions too (Fig. 5C-D), the relative improvement decreases with increasing noise scaling. This is consistent with our analysis, since increasing the noise scaling α increases Δ (see Sec. 2.7) and hence the influence of neglected second and higher order terms. In general, the relative improvement worsen when ϵ goes to 0; this follows from the definition of $C(\epsilon)$ and the fact that when environmental change is weak, using gains optimized in the original environment without adapting them to the new environment is superior to adaptive but only approximately optimal gains (also note that $C(\epsilon)$ is a relative measure, and the absolute benefit gained from any adaptation, even optimal, for small environmental changes are of course small). Finally, note that the fact that the relative errors in $\hat{\boldsymbol{g}}^{(1)}$ and $\hat{\boldsymbol{g}}^{hom}$ are comparable, and C^{hom} displays similar performance to $C^{(1)}$ (although slightly worse) indicates that the global suppression of firing rates implemented by $\hat{\boldsymbol{g}}^{hom}$ performs almost as well as the local suppression implemented by $\hat{\boldsymbol{g}}^{(1)}$, even for super-Poissonian noise.

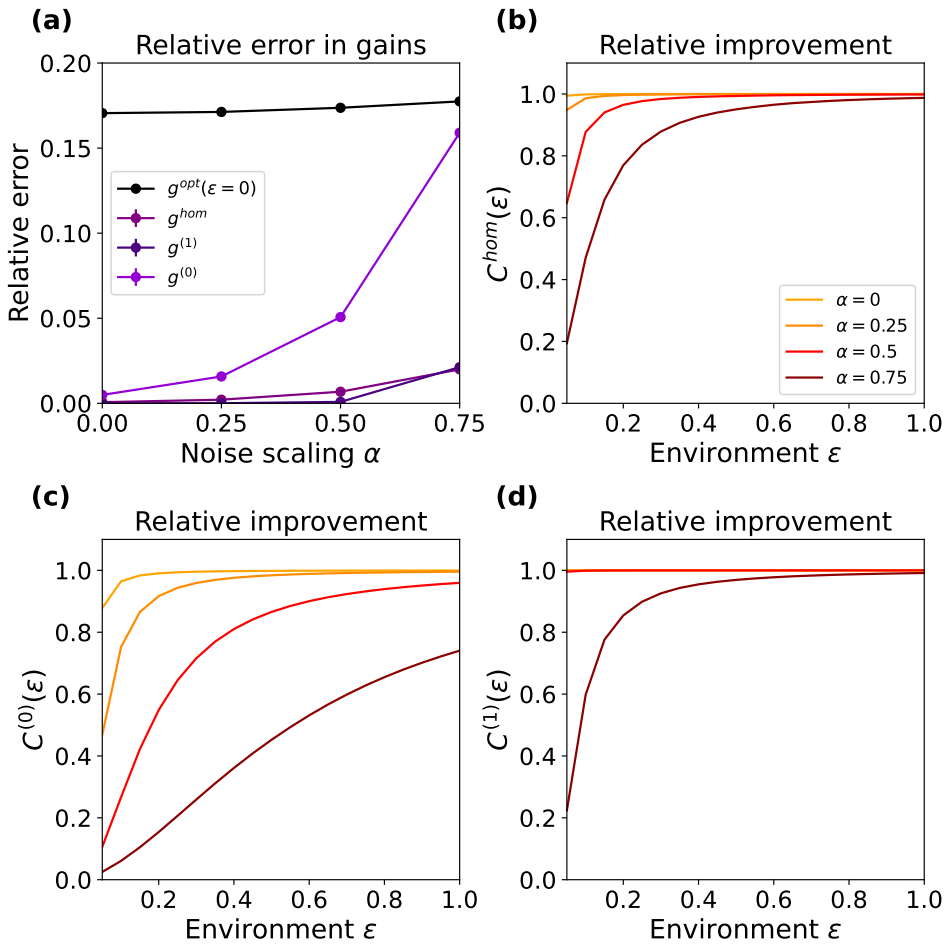

Figure 5: (a) Relative errors of gains. The mean (across ϵ) relative error (see Eq. (32)) of different approximate gains to the optimal gains, with standard deviation error bars, are plotted in purple, as a function of the noise scaling parameter α. In black, we plot the deviation of the optimal gains between the end points (see Eq. (33)). (b) Relative improvement of $\hat{\boldsymbol{g}}^{hom}$. We used the measure given by Eq. (31) to quantify performance of the gains $\hat{\boldsymbol{g}}^{hom}$. Different colours give the curve of C^{hom} for different noise scalings, with darker colours indicating stronger noise. (c-d) Same as (b), but for $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{(1)}$ respectively.

Aligned noise. We now fix the noise to have Poisson-like scaling (*i.e.*, $\alpha = 1/2$), and take the noise correlation matrix W to have approximately the same eigenbasis as $\hat{\rho}$ (hence "aligned noise"), but a different spectrum. In particular, $\hat{\rho}(\epsilon)$ is obtained by normalising a covariance with $1/n$ spectrum, and $W(\epsilon)$ is obtained by normalising a covariance with the same eigenbasis but a $1/n^{\gamma}$ spectrum. See Sec. 3.2 for further details. If we approximate $\hat{\rho}$ and W as having exactly aligned eigenbases with exact power-law eigenspectrum decay, we show in App. A.9 that this implies that χ satisfies the equation:

$$\chi = \mu \left(1 - \frac{1}{K} \sum_{n} \frac{1}{1 + \chi b n^{\gamma - 1}} \right), \tag{34}$$

where b is a constant. Once again, since the ideal case leading to Eq. (34) differs from our numerical simulations, the value of χ found using this equation gives a lower bound on the performance of $\hat{\boldsymbol{g}}^{hom}$.

The case $\gamma = 1$ corresponds to perfect alignment between signal and noise correlations, $\hat{\rho}^{-1}W = I$. $\hat{g}_a^{(1)}\hat{\omega}_a$ becomes uniform in this case (see Eq. (30)), and thus $\hat{\boldsymbol{g}}^{(1)}$ also displays homeostasis between clusters and environments. Furthermore, the optimal gains $\hat{\boldsymbol{g}}^{\text{opt}}$ likewise display homeostasis both between clusters and environments. In fact, in this case $\hat{\boldsymbol{g}}^{\text{opt}} = \hat{\boldsymbol{g}}^{hom}$ exactly, leading to zero relative error (Fig. 6a) and perfect relative improvement C^{hom} (Fig. 6b). This case is that of so-called information-limiting noise correlations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). Our simulations suggest that homeostatic adaptation is particularly optimal in this case. For other cases, we can see that for increasing γ, which corresponds to lower-dimensional noise (as noise power falls off faster as a function of eigenvalue rank), the homeostatic gains $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{hom}$ tend to perform better, while $\hat{\boldsymbol{g}}^{(1)}$ tends to perform worse.

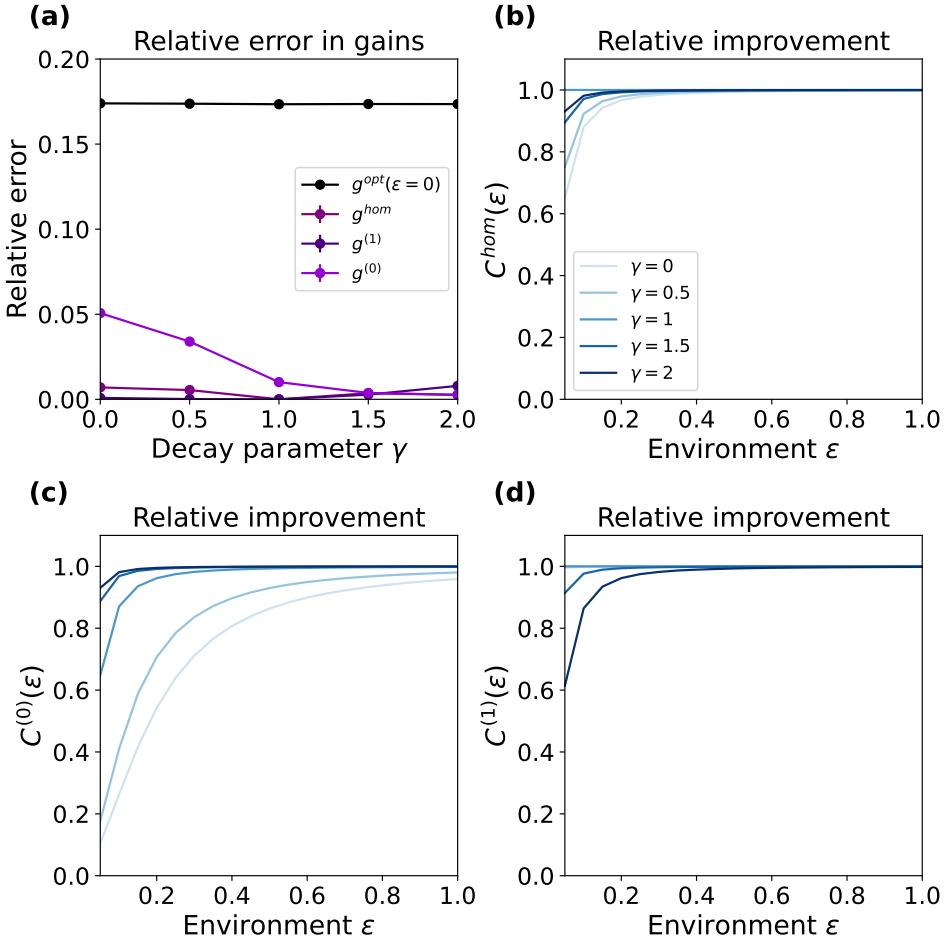

Figure 6: (a) Relative errors of gains. The mean (across ϵ) relative error (see Eq. (32)) of different approximate gains to the optimal gains, with standard deviation error bars, are plotted in purple, as a function of the decay parameter γ. In black, we plot the deviation of the optimal gains between the end points (see Eq. (33)). (b) Relative improvement of $\hat{\boldsymbol{g}}^{hom}$. We used the measure given by Eq. (31) to quantify performance of the gains $\hat{\boldsymbol{g}}^{hom}$. Different colours give the curve of C^{hom} for different noise correlation decay, with darker colours indicating a faster eigenvalue decay for noise correlations. (c-d) Same as (b), but for $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{(1)}$ respectively.

Uniform noise correlation coefficients. The final noise model we considered is that of noise with uniform correlation coefficient across cluster pairs. We once again fix $\alpha = 1/2$. We now take $W = (1-p)I_K + p\mathbf{1}\mathbf{1}^T$, where p is constant, *i.e.*, the noise correlation between any two distinct clusters is p. In this case, $W^{-1} = \frac{1}{1-p}\left(I_K - \frac{p}{1+p(K-1)}\mathbf{1}\mathbf{1}^T\right)$, which we approximate as $I_K/(1-p)$. Under this approximation, adding constant (positive) correlations across neurons has the same effect as scaling down the base level of noise $\sigma^2 \mapsto (1-p)\sigma^2$. We can obtain the following analytic approximation for χ (see App. A.9):

$$\chi = \mu\frac{\ln(K^q)}{K^q - 1},\ q = \frac{\sigma^2(1-p)}{\mu \mathrm{CV}^2}. \tag{35}$$

Note that in the limit as $q \to 0$, this reduces to $\chi = \mu$. Because of our approximations, the value of χ found this way once again gives only a lower bound on the performance of $\hat{\boldsymbol{g}}^{hom}$.

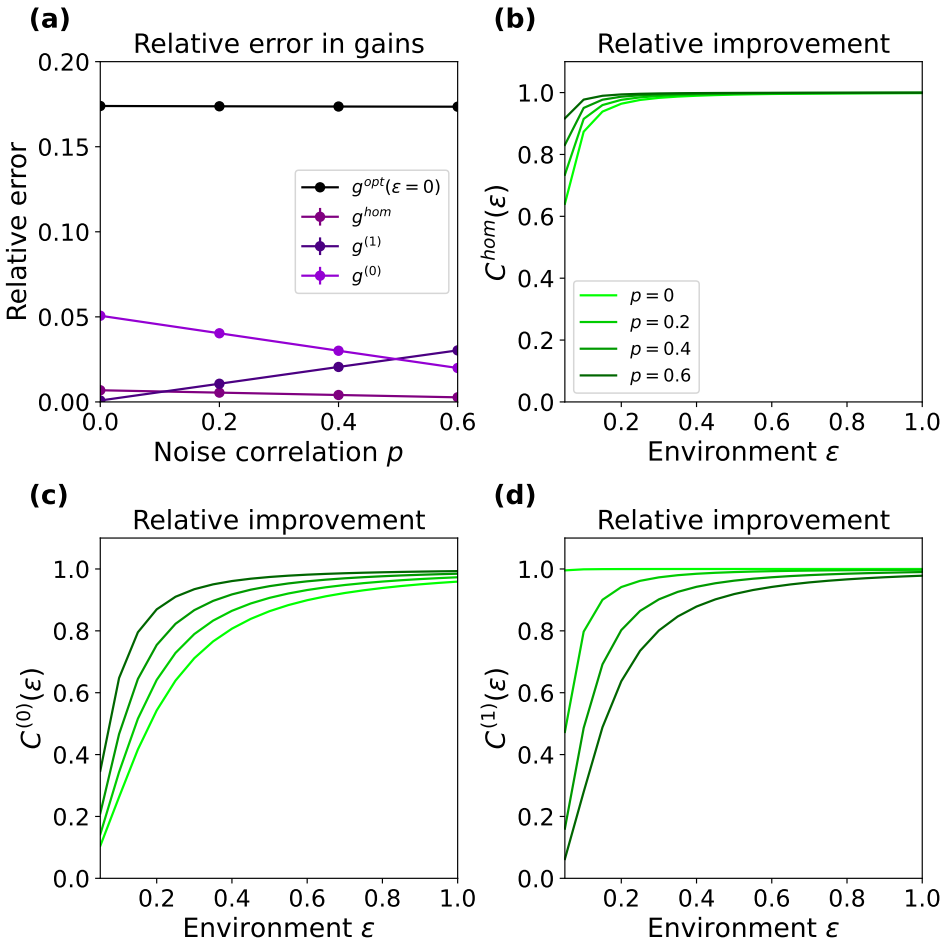

Figure 7: (a) Relative errors of gains. The mean (across ϵ) relative error (see Eq. (32)) of different approximate gains to the optimal gains, with standard deviation error bars, are plotted in purple, as a function of the noise correlation p. In black, we plot the deviation of the optimal gains between the end points (see Eq. (33)). (b) Relative improvement of $\hat{\boldsymbol{g}}^{hom}$. We used the measure given by Eq. (31) to quantify performance of the gains $\hat{\boldsymbol{g}}^{hom}$. Different colours give the curve of C^{hom} for different noise correlations p, with darker colours indicating higher levels of noise correlation. (c-d) Same as (b), but for $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{(1)}$ respectively.

In Fig. 7a we examine the relative error in the approximate gains as a function of the noise correlation p. Note that our results display a linear relationship between the relative error and noise correlation, with $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{hom}$ decreasing and $\hat{\boldsymbol{g}}^{(1)}$ increasing. In a similar fashion, we see (Fig. 7b-d) that $C^{(0)}$ and C^{hom} increase as a function of p, while $C^{(1)}$ decreases. At a noise correlation of $p \approx 0.5$, the relative errors display an intersection between $\hat{\boldsymbol{g}}^{(0)}$ and $\hat{\boldsymbol{g}}^{(1)}$, suggesting that for noise correlations stronger than 0.5 the zeroth order approximation outperforms the first order approximation. This is reinforced by the fact that, for $p = 0.6$, $C^{(0)}$ lies above $C^{(1)}$. Furthermore, for all noise correlations $p \neq 0$, C^{hom} lies

above $C^{(1)}$. Taken together, this indicates that constant positive noise correlations between neurons strongly favours an exact homeostatic strategy.

Our numerical simulations across the three noise models (with varying noise power scaling, noise spectrum, and noise correlations) demonstrate that homeostatic strategies are capable of effectively navigating the trade-off between energetic costs and coding fidelity, while being robust to a wide range of parameters for the noise distribution. This suggests that, regardless of how noise arises in neural systems, homeostasis presents a promising solution to dealing with that noise in an adaptive manner. In particular, we have seen that the solution $\hat{\boldsymbol{g}}^{hom}$ performs very well for both varying noise correlations and power-spectra, even in cases where the zeroth order gains $\hat{\boldsymbol{g}}^{(0)}$ perform less well. Likewise, for most conditions the homeostatic gains perform as well or better than the first order gains $\hat{\boldsymbol{g}}^{(1)}$. The first order solution includes correction terms which depend on the second-order response statistics, such correlation between pairs of neurons, which in general change between environments. In contrast, changes in $\hat{\boldsymbol{g}}^{hom}$ use only information local to each neuron, *i.e.*, its firing rate. This implies that good performance does not require the use of a complex regulation mechanisms which takes into account how the encoding of each cluster relates to the population at large; homeostatic regulation alone is sufficient.

2.9 Synaptic normalisation allows for propagation of homeostasis between populations

So far, we have considered homeostatic coding from a normative perspective, showing with both theoretical arguments and numerical simulations that homeostasis efficiently trades-off between coding fidelity and energetic costs. In this section, we begin to address the question of biological plausibility; specifically, what biological mechanisms are necessary for the homeostatic coding regime to be propagated between neural populations. We will show that there is a deep connection between homeostatic coding and synaptic scaling (Turrigiano, 2008).

Consider two neural populations, with an upstream population providing feedforward input to a downstream population. We will suppose that the upstream population is engaged in homeostatic coding, and ask what is necessary for the downstream population to be also engaged in homeostatic coding. We will denote the tuning and representational curves of the upsteam population by $h_i^{(1)}(\boldsymbol{s})$ and $\Omega_i^{(1)}(\boldsymbol{s})$ respectively. We drop corrections to the zeroth order and work only with the homeostatic coding regime for individual neurons; that is we assume the gains are given by $g_i = \chi/\omega_i$.

The downstream population of tuning curves will be given by $h_m^{(2)}(\boldsymbol{s})$ for $m = 1, \ldots, M$. Let W_{mi} be the synaptic weight[2] from neuron i in the upstream population to neuron m in the downstream population. Working in a linear rate model, this gives us that the tuning curve of neuron m is

$$h_m^{(2)}(\boldsymbol{s}) = \sum_{j=1}^{N} W_{mj} h_j^{(1)}(\boldsymbol{s}). \tag{36}$$

Suppose that the downstream population has representational curves $\Omega_m^{(2)}(\boldsymbol{s})$ for $m = 1, \ldots, M$. We can write each representational curve as a linear combination of the upstream cluster representational curves,

$$\Omega_m^{(2)}(\boldsymbol{s}) = \sum_{j=1}^{N} w_{mj} \Omega_j^{(1)}(\boldsymbol{s}). \tag{37}$$

We call these weights w_{mj} the *computational weights*, since they are determined by the computational goals of the downstream population. In line with our approach so far, we treat these computational goals (and hence also the computational weights) as given.

[2]Synaptic weights here are to be understood as the effect of a pre-synaptic neuron on post-synaptic firing rate, and not as *e.g.* the magnitude of post-synaptic mEPSP. Thus, changes in the intrinsic excitability of a post-synaptic neuron are incorporated into changes in the synaptic weights.

We now address the question of how the synaptic weights W_{mj} should depend on the computational weights w_{mj}, and how they should adapt as stimulus statistics change in order for the downstream population to also engage in homeostatic coding. In App. A.6 we find that homeostatic coding in the downstream population is achieved if the synaptic weights are related to the computational weights by

$$W_{mj} = \frac{w_{mj}\omega_j^{(1)}}{\sum_{i=1}^{N} w_{mi}\omega_i^{(1)}}. \tag{38}$$

where $\omega_j = \mathbb{E}\left[\Omega_j^{(1)}(\boldsymbol{s})\right]$ as before.

Note that this scheme normalises the total synaptic input mass onto neuron m to be 1. Different normalisation factors are equivalent to different values of χ for different populations. Differences in the optimal choice of χ may arise from *e.g.*, different noise correlation statistics or different rate coding intervals.

Thus, we have demonstrated that homeostatic coding, applied sequentially to two populations, provides an additional normative interpretation to the learning rule of synaptic normalisation, in which synapses onto a neuron are jointly scaled to keep total input mass constant. This may be the computational reason for the "synaptic scaling" observed in certain studies of firing rate homeostasis (Turrigiano et al., 1998; Turrigiano, 2008).

2.10 Homeostatic DDCs

Up to this point we have made no assumptions about the nature of cortical representations, beyond rate-coding and the condition described by Eq. (15), and thus our framework has been independent of the computational goals of the circuit. We now apply our framework to a specific theory of neural representation and computation, namely the distributive distributional code (DDC) (Vertes and Sahani, 2018). While there is ample behavioural and psychophysical evidence for the hypothesis that animal perception and decision making approximate optimal Bayesian probabilistic inference (Kersten et al., 2004; Stocker and Simoncelli, 2006; van Bergen et al., 2015; Geurts et al., 2022), our knowledge of the neural implementation of computations serving Bayesian inference (such as computations of posterior distributions and expectations) are rudimentary at best.

The DDC framework is one proposal for neural implementations of Bayesian computations underlying perception. In a DDC neural responses directly encode posterior expectations of the inferred latent causes of sensory inputs based on an internal generative model. This internal model mathematically relates the latent causes or latent variables, which we denote by $\boldsymbol{z}$, to the sensory inputs or stimuli, $\boldsymbol{s}$, via a family of conditional distribution $f(\boldsymbol{s}|\boldsymbol{z})$, as well as a prior distribution, $\pi(\boldsymbol{z})$, over the latent variables. The sensory system is assumed to implement a so-called *recognition model* corresponding to this generative model. More specifically, the task of the sensory system is to invert the generative model by calculating and representing the posterior expectations of the latent variables, or rather, the posterior expectations of a sufficiently rich (and fixed) class of so-called kernel functions of these latent variables, given a stimulus $\boldsymbol{s}$. This is schematically shown in Fig. 8.

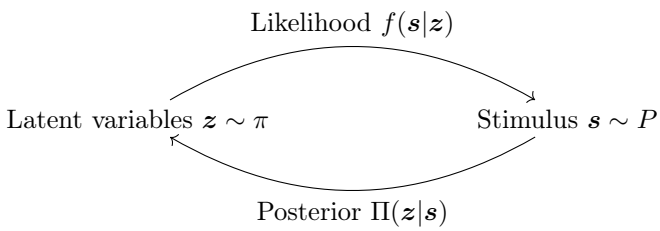

Figure 8: Graphical of the generative and recognition models underlying a DDC. According to the internal generative model, stimulus $\boldsymbol{s}$ is drawn from a distribution conditioned on the value of latent variables or latent causes, $\boldsymbol{z}$, which are themselves drawn from a prior $\pi(\boldsymbol{z})$. The task of the brain is to invert this generative model by computing the posterior Π over latent variables conditioned on the sensory input $\boldsymbol{s}$.

In a DDC, each neuron is thus assigned to a kernel function of the latent variables $\phi_i(\boldsymbol{z})$. The representational curves are then given by the posterior expectation of ϕ_i given the stimulus $\boldsymbol{s}$, $\Omega_i(\boldsymbol{s}) = \mathbb{E}[\phi_i(\boldsymbol{z})|\boldsymbol{s}]$. Our optimal homeostatic solution for the gains $g_i = \chi/\omega_i$, then predicts the following relationship between the neural tuning curves and Bayesian posterior and prior expectations of $\phi_i(\boldsymbol{z})$:

$$h_i(\boldsymbol{s}) = \chi \frac{\mathbb{E}[\phi_i(\boldsymbol{z})|\boldsymbol{s}]}{\mathbb{E}[\phi_i(\boldsymbol{z})]}. \tag{39}$$

Thus this homeostatic DDC framework predicts that the neural tuning curves are the ratio of the posterior to the prior expectations of kernel functions. In deriving this result we have used the tower property to replace the average posterior expectation $\mathbb{E}[\mathbb{E}[\phi_i(\boldsymbol{z})|\boldsymbol{s}]]$ with the prior expectation $\mathbb{E}[\phi_i(\boldsymbol{z})]$ of the internal model; strictly speaking, this is true under an ideal-observer internal model in which the generative internal model (as specified by $f(\boldsymbol{s}|\boldsymbol{z})$ and $\pi(\boldsymbol{z})$) provides a faithful, objective description of the statistics of of sensory stimuli $\boldsymbol{s}$. In Sec. 2.12 below, we will relax this assumption and discuss the resulting consequences.

A special case of the homeostatic DDC arises in the case where the kernel functions are delta functions, $\phi_i(\boldsymbol{z}) = \delta(\boldsymbol{z} - \boldsymbol{z}_i)$ where $\boldsymbol{z}_i$ is a point in latent variable space corresponding to neuron i's preferred value of the latent variable. In this case, the homeostatic DDC becomes

$$h_i(\boldsymbol{s}) = \chi \frac{\Pi(\boldsymbol{z}_i|\boldsymbol{s})}{\pi(\boldsymbol{z}_i)} = \chi \frac{f(\boldsymbol{s}|\boldsymbol{z}_i)}{P(\boldsymbol{s})}, \tag{40}$$

where in writing the last form, we used the Bayes' rule: $\Pi(\boldsymbol{z}|\boldsymbol{s}) = f(\boldsymbol{s}|\boldsymbol{z})\pi(\boldsymbol{z})/P(\boldsymbol{s})$. We call this coding scheme *Bayes-ratio coding*.

2.11 Bayes-ratio coding via divisive normalisation

As we now show, Bayes-ratio coding can be achieved via divisive normalisation with adaptive weights. Divisive normalisation has been dubbed a canonical cortical or brain operation (Carandini and Heeger, 2012; Westrick et al., 2016), as it appears in multiple brain regions serving different computational goals. Given the feedforward inputs, $F_i(\boldsymbol{s})$, the divisive normalisation model computes the response (and thus the tuning curve) of neuron i as

$$h_i(\boldsymbol{s}) = \chi \frac{F_i(\boldsymbol{s})^n}{\sigma^n + \sum_j w_j F_j(\boldsymbol{s})^n}. \tag{41}$$

Here w_j are a collection of normalisation weights, and $\sigma \geq 0$ and $n \geq 1$ are constants. Bayes-ratio coding can be achieved naturally by a divisive normalisation model in which $n = 1$ and the feed-forward inputs, $F_j(\boldsymbol{s})$, are given by the likelihood function, $f(\boldsymbol{s}|\boldsymbol{z}_i)$, of the DDC's internal generative model. We

then choose the adaptive normalisation weights w_i to encode prior probabilities, $w_i = \pi(\boldsymbol{z}_i)\delta\boldsymbol{z}_i$, where the volume element $\delta\boldsymbol{z}_i$ is chosen such that the latent variable space is the disjoint union of volumes of size $\delta\boldsymbol{z}_i$ each containing their corresponding sample point $\boldsymbol{z}_i$. Note that these normalisation weights are adaptive, and vary between environments as π varies. Then we obtain:

$$\begin{aligned}\sum_j w_j F_j(\boldsymbol{s}) &\approx \int f(\boldsymbol{s}|\boldsymbol{z})\pi(\boldsymbol{z})d\boldsymbol{z} \\ &= P(\boldsymbol{s}),\end{aligned}$$

and hence

$$h_i(\boldsymbol{s}) \approx \chi \frac{f(\boldsymbol{s}|\boldsymbol{z}_i)}{\sigma + P(\boldsymbol{s})}. \tag{42}$$

Taking the limit as $\sigma \to 0$, we obtain Bayes-ratio coding Eq. (40). Therefore, provided σ is small compared to $P(\boldsymbol{s})$, divisive normalisation can be used to approximate Bayes-ratio coding.

Not only does this show that implementing Bayes-ratio coding is biologically plausible, this framework gives a normative computational interpretation to both the feedforward inputs and the normalisation weights, as the generative model's likelihoods and prior probabilities, respectively.

Additionally, consider now a hierarchical generative model $\boldsymbol{z}^{(2)} \to \boldsymbol{z}^{(1)} \to \boldsymbol{s}$, specified by the generative likelihood functions $f^{(2)}\left(\boldsymbol{z}^{(1)}|\boldsymbol{z}^{(2)}\right)$ and $f^{(1)}\left(\boldsymbol{s}|\boldsymbol{z}^{(1)}\right)$ (replacing $f(\boldsymbol{s}|\boldsymbol{z})$ above), and prior distribution $\pi^{(2)}\left(\boldsymbol{z}^{(2)}\right)$. Consider now a feedforward network recognition model, with two populations or "layers" that implement a Bayes-ratio encoding of the posterior distributions of these hierarchical latent variable. In particular, we would like the downstream population (the second layer) represents the posterior probabilities of the higher-level latent variables, $\boldsymbol{z}^{(2)}$, conditional on $\boldsymbol{s}$ by means of Bayes-ratio coding. Working in the notation and framework of Sec. 2.9, this means the representational curve of neuron m in the downstream population is given by

$$\Omega_m^{(2)}(\boldsymbol{s}) = \Pi\left(\boldsymbol{z}^{(2)} = \boldsymbol{z}_m^{(2)}|\boldsymbol{s}\right). \tag{43}$$

In App. A.10, we show that the corresponding synaptic weights between the two populations are given by

$$W_{mj} = f^{(2)}\left(\boldsymbol{z}_j^{(1)}|\boldsymbol{z}_m^{(2)}\right)\delta\boldsymbol{z}_j^{(1)}, \tag{44}$$

where $\delta\boldsymbol{z}_j^{(1)}$ is the size of the volume element to which $\boldsymbol{z}_j^{(1)}$ belongs, and $f^{(2)}$ is the conditional probability distribution of $\boldsymbol{z}^{(1)}$ given $\boldsymbol{z}^{(2)}$. This result is significant for two reasons. Firstly, the synaptic weights make no reference to the prior distribution, $\pi^{(2)}$, over $\boldsymbol{z}^{(2)}$. There is therefore no need to adapt the weights when the environment changes, if the change only affects the statistics of the higher level causes $\boldsymbol{z}^{(2)}$. Secondly, the downstream population represents a posterior distribution $\Pi\left(\boldsymbol{z}^{(2)} = \boldsymbol{z}_m^{(2)}|\boldsymbol{s}\right)$, and therefore acts as part of the *recognition* model. However, the synaptic weights are proportional to $f^{(2)}\left(\boldsymbol{z}_j^{(1)}|\boldsymbol{z}_m^{(2)}\right)$, and therefore require only knowledge of the *generative* probabilities. Finally, we can see by induction that this scheme can be propagated through many layers, each representing posterior probabilities of higher-level latent variables further up a generative hierarchy.

2.12 Homeostatic DDCs account for stimulus specific adaptation

We found that homeostatic DDCs can provide a normative account of stimulus specific adaptation (SSA). Here, we start by showing this in the special case of Bayes-ratio coding, which is mathematically simpler and provides good intuitions for more general homeostatic DDCs. We then discuss the general case and build a homeostatic DDCs model of empirically observed SSA in the primary visual cortex (V1) (Benucci et al., 2013).

According to Eq. (40), for Bayes-ratio codes the tuning curve of neuron i is given by $h_i(\boldsymbol{s}) = \chi f(\boldsymbol{s}|\boldsymbol{z}_i)/P(\boldsymbol{s})$. Suppose now that the stimulus distribution in the environment, $P(\boldsymbol{s})$, changes due

to a change in the statistics of the latent causes, $\pi(\boldsymbol{z})$, while the causal processes linking the latent causes and observed stimuli, as captured by $f(\boldsymbol{s}|\boldsymbol{z})$, remains stable. In this case, the tuning curve of *all* neurons are modified by a multiplicative factor given by the ratio of the $P(\boldsymbol{s})$ in the old environment to that in the new environment. This will lead to a multiplicative suppression of the response of all neurons to stimuli that are over-represented in the new environment relative to the old one, in a way that is independent of the neurons' preferred stimulus. Such a suppression thus constitutes a pure form of stimulus specific adaptation. In particular, it leads to a repulsion of tuning curves from an over-represented stimuli, which is a typical manifestation of stimulus specific adaptation. In V1, for example, orientation tuning curves display a repulsion from an over-represented orientation (Movshon and Lennie, 1979; Müller et al., 1999; Dragoi et al., 2000, 2002; Benucci et al., 2013).

Heretofore we have implicitly made a so-called ideal observer assumption by assuming that the internal generative model underlying the DDC or Bayes-ratio code is a veridical model of the stimulus distribution in the environment. We now provide a significant generalization of the result above to the more general and realistic case of a non-ideal-observer internal model. In this case there will be a discrepancy between the environment's true stimulus distribution, $P^E(\boldsymbol{s})$, and the marginal stimulus distribution predicted by the internal generative model

$$P^I(\boldsymbol{s}) \equiv \int f(\boldsymbol{s}|\boldsymbol{z})\pi(\boldsymbol{z})d\boldsymbol{z}. \tag{45}$$

In App. A.11, we demonstrate that the Bayes-ratio tuning curves in this case are given by

$$h_i(\boldsymbol{s}) = \chi \frac{f(\boldsymbol{s}|\boldsymbol{z}_i)}{P^I(\boldsymbol{s})F_i}, \tag{46}$$

where

$$F_i = \int \frac{P^E(\boldsymbol{s})}{P^I(\boldsymbol{s})} f(\boldsymbol{s}|\boldsymbol{z}_i)d\boldsymbol{s}. \tag{47}$$

(For ideal-observer internal models $P^E = P^I$, in which case, $F_i = 1$ due to the normalization of $f(\boldsymbol{s}|\boldsymbol{z})$, and we go back to Eq. (40).) As we can see, the tuning curves neatly decompose into a base curve $\chi f(\boldsymbol{s}|\boldsymbol{z}_i)$ multiplied by a stimulus specific factor $P^I(\boldsymbol{s})^{-1}$ and a neuron specific factor F_i^{-1}. Consequently, if we consider a pair of environments indexed by $v = 0, 1$, and assume the likelihood function of the internal model does not change between environments, the transformation of the tuning curves is given by Eq. (48),

$$h_i(\boldsymbol{s}; v=1) = \frac{P^I(\boldsymbol{s}; v=0)}{P^I(\boldsymbol{s}; v=1)} \times h_i(\boldsymbol{s}; v=0) \times \frac{F_i(v=0)}{F_i(v=1)}. \tag{48}$$

Thus, adaptation from environment $v = 0$ to environment $v = 1$ causes the tuning curves to transform via multiplication by a *stimulus specific factor* $P^I(\boldsymbol{s}; v=0)/P^I(\boldsymbol{s}; v=1)$ and a *neuron specific factor* $F_i(v=0)/F_i(v=1)$. Now suppose an environmental change results in an increase in the prevalence of an adaptor stimulus. As long as the internal model is a sufficiently good (if not perfect) model of $P^E(\boldsymbol{s})$, it should adapt its prior distribution, $\pi(\boldsymbol{z})$, in a way that results in an increase of its predicted stimulus distribution, $P^I(\boldsymbol{s})$, around the adaptor stimulus. This results in a stimulus specific factor that goes below 1 around the adaptor $\boldsymbol{s}$, thus causing suppression and repulsion of tuning curves away from the adaptor. Now consider a neuron whose preferred stimulus (roughly represented by $\boldsymbol{z}_i$) is close to the adaptor stimulus. We further make the assumption that the change in the internal model, in response to the environmental change, is sufficiently conservative such that it leads to a smaller increase in $P^I(\boldsymbol{s})$ around the adaptor, compared with the increase in the veridical $P^E(\boldsymbol{s})$ (this is a reasonable assumption especially for artificially extreme adaptation protocols used in experiments). In this case, we expect the ratio $P^E(\boldsymbol{s})/P^I(\boldsymbol{s})$ to increase in the support of $f(\boldsymbol{s}|\boldsymbol{z}_i)$, which (by our assumption about the preferred stimulus $\boldsymbol{z}_i$) is near the adaptor. According to Eq. (47) this will lead to an increase in F_i, and hence a suppressive neuron-specific factor $F_i(v=0)/F_i(v=1)$ for neurons with preferred stimulus

near the adaptor. Benucci et al. (2013) found that homeostatic and stimulus-specific adaptation in V1 can be modelled via separable stimulus-specific and neuron-specific

Note that the assumption of constancy of the likelihood across the two environments need not be framed as a veridical assumption about the objective environmental change, but rather as an assumption about the inductive biases of the internal generative model, according to which it tends to model changes in the observed stimulus distribution, $P^E(\boldsymbol{s})$, as primarily resulting from changes in the statistics of the latent causes in the environment (which the model captures in $\pi(\boldsymbol{z})$), rather than in the causal process itself, as modelled by $f(\boldsymbol{s}|\boldsymbol{z})$. As long as this is good enough an assumption to result in a $P^I(\boldsymbol{s})$ which changes in similar ways to $P^E(\boldsymbol{s})$ (and in particular exhibits an increase around adaptor stimuli), the conclusions reached above will hold.

In the case of general DDC codes, we arrive at an analogous transformation, given by Eq. (88) in App. A.11 in which neuron specific and stimulus specific effects are mixed. In general, more mixing occurs the wider the DDC kernels are, with the separation seen in Eq. (48) arising in the zero-width limit. However, when the kernels are unimodal and sufficiently narrow, we would expect an approximate factorization of the effect of adaptation into a stimulus-specific and a neuron-specific suppression factor.

We now apply this result to the experiments performed by Benucci et al. (2013). These experiments examined the effects of adaptation on orientation tuned neurons in primary visual cortex of anaesthetised cats. Orientation tuned cells in V1 have tuning curves which are decreasing functions of the difference between the orientation of a grating placed in their visual field and a preferred orientation. Anaesthetised cats were shown a distribution of such gratings that had an increased prevalence (3 to 5 times more likely) of one particular orientation (arbitrarily defined to be to 0°). A control group was exposed to a uniform distribution of gratings. After 2 seconds (or approximately 50 stimulus presentations), V1 tuning curves had adapted, and exhibited both the suppressive and repulsive effects mentioned above.

To model these findings we constructed a homeostatic DDC model as follows, and fit it to the data from Benucci et al. (2013). We took the model's stimulus and latent variable spaces to be the circular orientation space $[-90, 90)$, and we took the model's generative likelihoods to be given by translation invariant functions $f(s|z) = f(s - z)$, proportional to the normal density with standard deviation σ_f, normalised over the circle[3]. Likewise, we take the DDC kernel functions $\phi_i(z) = \phi\left(z - z^{(i)}\right)$ proportional to the normal density with standard deviation σ_ϕ.

The distribution over stimuli orientations used in the experiment is highly artificial, in the way it jumps discontinuously near the adaptor orientation. The internal prior distribution employed by the brain is, on the other hand, more likely to be smooth, reflecting the brain's inductive bias adapted to natural environments. Thus in our model we chose the internal prior to be a smooth distribution. Specifically, we constrained it to be a mixture of a uniform and a Gaussian distribution centered at the adaptor orientation, and having standard deviation σ_π. The mixing proportion was not a free (fitted) parameter, but was chosen to enforce that the density of the Gaussian component, plus the density from the uniform component over the adaptor stimulus region, is equal to the experimental adaptor probability. Our model therefore has only three free parameters, σ_f, σ_ϕ and σ_π.

[3]More rigorously, we ought to use a wrapped (periodic) normal distribution. However, since the standard deviations are small compared to the length of the circle, the renormalisation constant is approximately 1, and we can treat the density as a standard normal.

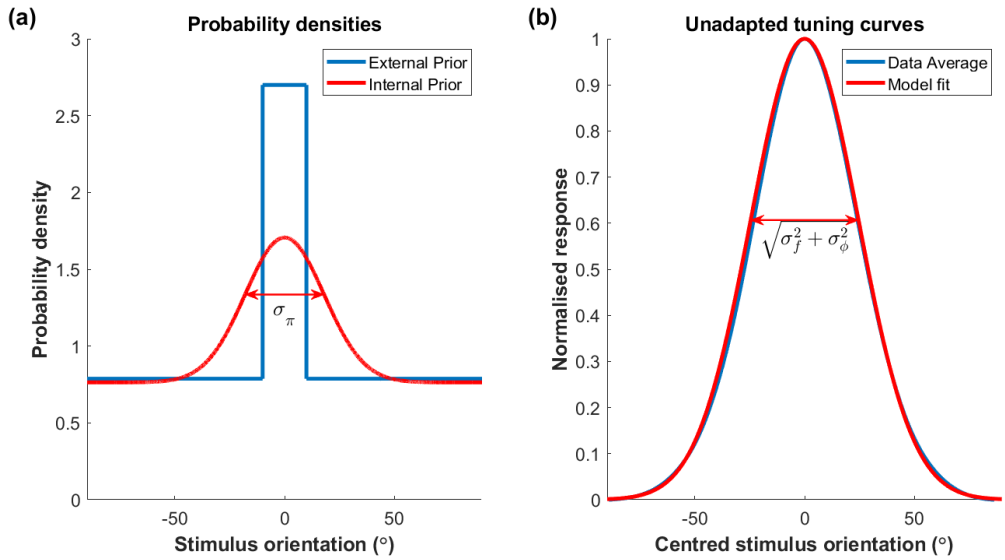

Figure 9: (a) The density of orientations used in the experiments (Benucci et al., 2013), for an adaptor probability of 30% and 8 stimulus orientations is shown in red. In blue is the Gaussian-uniform mixture distribution used in our model. (b) Unadapted tuning curves, normalised to have response of 1 at the preferred orientation and 0 at orthogonal orientations. The blue curve is obtained from neural recordings, the red curve is our model fit.

Using the dataset from Benucci et al. (2013), we found the tuning curves of the unadapted population by assuming that neural response is a function only of the difference between the preferred orientation and stimulus orientation (see Sec. 3.3). We then calculate the changes in preferred orientation of neurons for each of the experimental conditions. We then fit our model's three parameters to both the unadapted tuning curves (Fig. 9b) and the changes in preferred orientation (see Sec. 3.3 for further details). The best fit values for the model parameters were found to be $\sigma_\pi = 18°$, $\sigma_f = 14.2°$, and $\sigma_\phi = 20°$. The predictions of our model are compared to experimental data in Fig. 10.

We found that the adaptive tuning curves within our model display suppression and repulsion away from the adaptor stimulus (Fig. 10a,d). The experimental data and our model both show increased firing rate of the adapted population around the adaptor stimulus (Fig. 10b,e). Additionally, we see homeostasis of firing rates, demonstrated by the uniformity of the red curves. Lastly, our model recapitulates the repulsion of preferred orientation found experimentally (Fig. 10c,f). Repulsion here is reflected by the change in preferred orientation having the same sign as the pre-adaptation preferred orientation. As seen, repulsion is stronger nearer the adaptor, with the largest magnitude of repulsion occurring 30° from the adaptor in both the experiment and the model.

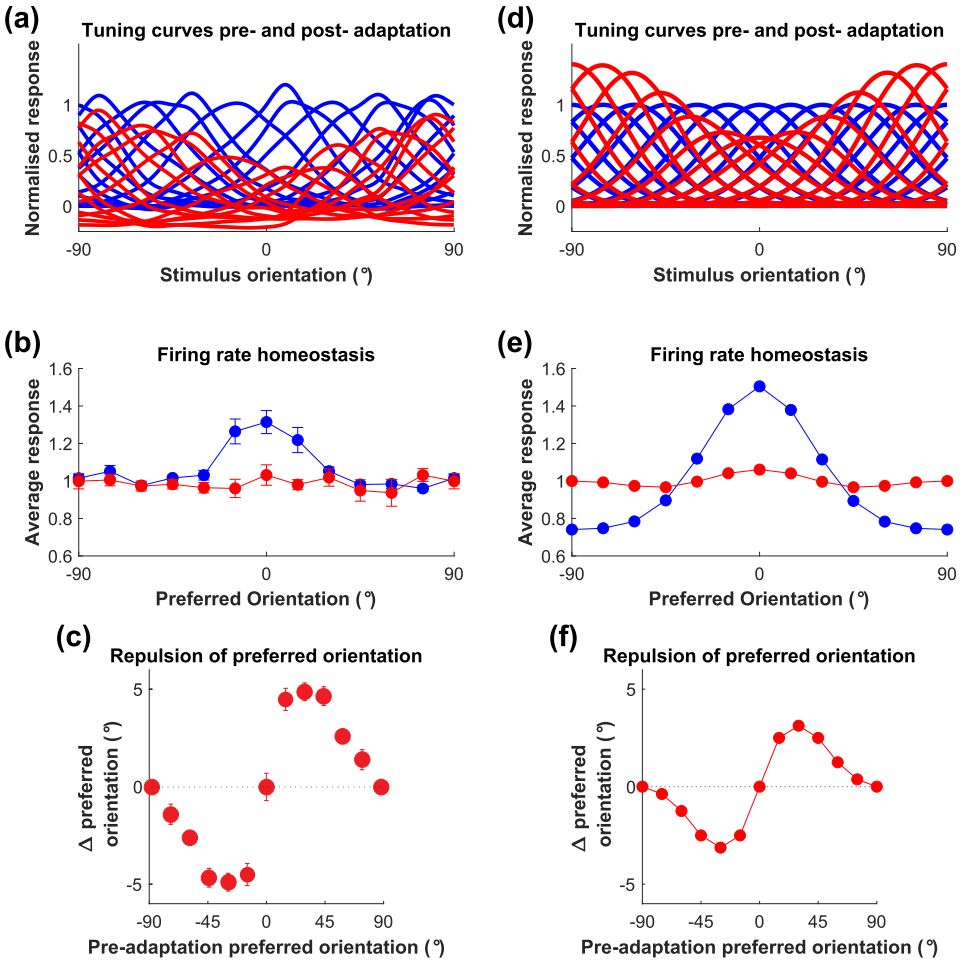

Figure 10: A comparison of the results obtained by Benucci et al. 2013 (a-c) and our model's fit (e-f). (a) The tuning curves for the adapted (red) and unadapted (blue) populations, averaged across experimental conditions. (b) The trial-averaged firing rates of the adapted (red) and unadapted (blue) populations exposed to the non-uniform stimulus ensemble. The time average of the population responses to uniform stimulus ensembles was normalised to 1. (c) The repulsion of preferred orientations, obtained from the average tuning curves in panel a. (e-f) The same as (a-c), but for our model.

3 Methods

3.1 Randomly generating high-correlation matrices

In our numerical simulations of firing rates within neuronal clusters, we randomly generated correlation matrices. Since we are focusing on behaviour within a cluster, such correlation matrices must have large and positive off-diagonal elements.

Our method is parameterised by α and q. q can be approximately interpreted as the average off-diagonal element (*i.e.* the correlation coefficient between different neurons within a cluster), while α controls the spread (*i.e.* variance) of correlations coefficients. The matrix ρ within each cluster is

derived by normalising a covariance matrix Σ, given by

$$\Sigma = q\mathbf{1}_k\mathbf{1}_k^T + (1-q)U\xi U^T$$

U is an orthogonal matrix drawn uniformly (*i.e.* according to the Haar measure), and $\boldsymbol{\xi}$ is a vector of independent $\Gamma(\alpha, 1/\alpha)$ random variables.

3.2 Specification of environments for simulations

To generate $\hat{\rho}(\epsilon)$ in our numerical simulations (Sec. 2.8) we first generate a covariance matrix $\Sigma(\epsilon)$, and let $\hat{\rho}(\epsilon)$ be the corresponding correlation matrix. The procedure for generating $\Sigma(\epsilon)$ is as follows.

We randomly and independently sample two $K \times K$ random Gaussian matrices $R_0, R_1 \sim \mathcal{N}_{K\times K}(0, I_{K\times K})$ and obtain symmetric matrices $S_0 = R_0 + R_0^T$ and $S_1 = R_1 + R_1^T$. As is well known, the eigen-basis (represented by an orthogonal matrix) of such a random Gaussian matrix is distributed uniformly (*i.e.*, according to the corresponding Haar measure) over the orthogonal group $O(K)$.

Now for $\epsilon \in (0,1)$, set $S(\epsilon) = (1-\epsilon)S_0 + \epsilon S_1$. Almost surely these matrices have a non-degenerate spectrum, and hence a unique representation as $S(\epsilon) = U(\epsilon)D(\epsilon)U(\epsilon)^T$ where $D(\epsilon)$ is diagonal with strictly decreasing eigenvalues and $U(\epsilon)$ is orthogonal. Moreover, $U(\epsilon)$ depends continuously on ϵ. Finally, we define $\Sigma(\epsilon) = U(\epsilon)\Lambda_1 U(\epsilon)^T$, where $\Lambda_1 = \mathrm{diag}(1, 1/2, 1/3, \ldots, 1/K)$.

In the case of simulations for aligned noise, we also generate a noise correlation matrix $W(\epsilon)$ which has an approximately $1/n^\gamma$ spectrum. Ideally, W and $\hat{\rho}$ would have the same eigen-basis. However, this is impossible, since W and $\hat{\rho}$ are both correlation matrices. Instead, we generate $W(\epsilon)$ be normalising the positive definite matrix $U(\epsilon)\Lambda_\gamma U(\epsilon)^T$ where $\Lambda_\gamma = \mathrm{diag}(1, 1/2^\gamma, 1/3^\gamma, \ldots, 1/K^\gamma)$

3.3 Fitting a homeostatic DDC to Benucci *et al.* 2013

The data we obtained from the Benucci *et al.* 2013 paper (Benucci et al., 2013) comprised 11 data sets. In each data set, the neural population was clustered into 12 groups based on preferred orientation. These neural populations were then exposed (on different occasions) to a distribution of 6-12 oriented gratings. This distribution was either uniform or biased, with one particular grating (arbitrarily set to $0°$) having higher prevalence, either $30\%, 35\%, 40\%$, or 50%. We discarded all data sets with a 50% prevalence, since (Benucci et al., 2013) report that homeostasis was not obtained in this case. After exposure to the distribution of gratings, the responses (*i.e.*, spike count) of each cluster were then measured to a test set of 20 oriented gratings. These responses were then normalised via an affine transform so that the after exposure to the uniform stimulus ensemble, the normalised response ranged from 0 to 1. See their paper (Benucci et al., 2013) for further details.

To obtain the data average tuning curve after exposure to the uniform ensemble (see Fig. 9b, blue) we used a cubic spline to interpolate all the tuning curves of the populations after exposure to the uniform ensemble in each data set. These were then translated to center the preferred orientation at $0°$ and averaged. The width of the resultant curve at height e^{-2} was found, and divided by 4 to obtain the standard deviation of a Gaussian fit to the unadapted tuning curve (see Fig. 9b, red). In subsequent fitting, $\sqrt{\sigma_f^2 + \sigma_\phi^2}$ was constrained to equal this value.

This reduces the number of free parameters to 2, which we choose to be σ_π and σ_f. We fit these parameters to the change in preferred orientation of each data set. For each data set, cluster (*i.e.*, group of neurons with a similar preferred orientation) and condition (either exposure to uniform or biased ensemble) a smoothing kernel was applied before using cubic interpolation to generate tuning curves. The smoothing kernel was applied to ensure that none of the tuning curves were multimodal. The argmax of these tuning curves was found to give the preferred orientation. The preferred orientation was then compared across conditions (uniform ensemble vs biased ensemble) to give a change in preferred orientation. These were then averaged across data sets with the same adaptor probability to give change in preferred orientation curves for each adaptor probability.

We then performed a grid search across values of σ_π and σ_f to compare these to the change in preferred orientation curves generated by our model. We chose the pair of values which minimised the sum of the absolute value of the difference between the curves obtained by our model and those in their data.

4 Discussion

We developed a theory of optimal gain modulation for combating noise in neural representations serving arbitrary computations. We demonstrated that, when mean neural firing rates are not too small, neurons are sufficiently selective and sparsely activated, and their responses form a high-dimensional geometry, the trade-off between coding fidelity and metabolic cost is optimised by gains that react to shifts in environmental stimulus statistics to yield homeostasis of the firing rates of clusters of similarly tuned neurons. We further demonstrated that our framework could account for the optimality of a broad distribution of firing rates within such clusters. By examining empirical estimates of parameters characterising cortical representations, we argued that the conditions necessary for the homeostatic adaptation to be optimal are expected to hold in the visual cortex and potentially other cortical areas. We further validated our approximation by demonstrating numerically that it performs well compared to the numerically calculated optimal gains. We then demonstrated that our results are robust with respect to the precise model of noise and generalized them to a large family of noise models, including Poissonian noise as well as correlated noise with general power-law scaling. Having developed a normative theory of neural homeostasis divorced from the computational aims of the neural population, we next showed how our theory can account for stimulus specific adaptation, when it is coupled with distributed distributional codes (DDC), a computational theory of neural representations serving Bayesian inference. In particular, we focused on a special case of homeostatic DDC codes which we termed Bayes-ratio coding and showed that this model can account for stimulus specific adaptation effects empirically observed in the primary visual cortex. In the following we will situate our work within the broader context of efficient coding theory, Bayesian neural representations, and alternative theories of stimulus specific adaptation. We will also discuss the shortcomings of the present work and possible directions for extension.

Our framework utilized an upper bound on the mutual information which becomes tight when the marginal distribution (*i.e.*, not conditional on a specific stimulus) of neural responses can be approximated by a Gaussian distribution. Another commonly used bound for mutual information is the Fisher Information lower bound first derived by Burnel and Nadal (Brunel and Nadal, 1998). In App. A.12, we illustrate that there is a close connection between these two bounds, in that both scale as the logarithm of the signal to noise ratio. A consequence of this is that the mutual information scales as the logarithm of the population firing rate, since this scaling holds for both an upper and a lower bound. Our choice to use the Gaussian upper bound was motivated by increased analytic tractability in the multidimensional stimulus case.

Within efficient coding theory, our work is closely related to that of Ganguli and Simoncelli (2014), but differs from it in important ways. Ganguli and Simoncelli (2014) considered a population of tuning curves, defined on a one-dimensional stimulus space, which was homogeneous and translationally invariant up to a warping (*i.e.*, reparametrization) of the stimulus space and possible variations in gain. Their normative framework then optimised the neural gains as well as the stimulus warping (*i.e.*, the reparametrization mapping from the original stimulus parameter to one over which the tuning curve population is homogeneous) to maximise the same objective function as ours, except for using the Fisher Information lower bound on the mutual information (Brunel and Nadal, 1998), instead of the upper bound we employed. Our work also differs in that we only optimised the neural gains. They showed that (in the case of unimodal tuning curves) the optimal warping amounts to a concentration of tuning curves around stimulus values over-represented in the environment, while optimal gains stay constant and do not adapt. As a consequence, their solution exhibits homeostasis of single-neuron firing rate.

Our framework extends and generalises that of Ganguli and Simoncelli (2014) in several ways. In the latter framework, the tuning curves are one-dimensional, and rigid constraints (most importantly, either unimodal or sigmoidal tuning curves, and homogeneity up to stimulus reparametrization) are placed on the shape and arrangement of the tuning curve population. In contrast, our framework is applicable to tuning curves with heterogeneous (in particular multi-modal) shapes and arbitrary arrangement over a high-dimensional stimulus space; our homeostatic solution only requires sufficiently high signal-to-noise ratio (which is also the condition for the tightness of the mutual information lower-bound employed in Ganguli and Simoncelli (2014)). Additionally, Ganguli and Simoncelli (2014) only considered the case of uncorrelated Poissonian noise. In Sec. 2.7, we showed how our framework can handle correlated noise with general (non-Poissonian) power-law scaling. Furthermore, we analytically calculated first order corrections to the homeostatic solution that arises in the high signal-to-noise ratio limit.

As noted above, the theory of Ganguli and Simoncelli (2014) predicts that tuning curves optimised to provide an efficient code under a stimulus distribution with over-represented stimuli cluster around those stimulus. Empirically, long-term exposure (on the order of few minutes) to adaptor stimuli has been found to result in attraction of tuning curves (Ghisovan et al., 2009). This is in contrast to the repulsive effects seen in short term adaptation (Movshon and Lennie, 1979; Müller et al., 1999; Dragoi et al., 2000, 2002; Benucci et al., 2013). It is likely that the effects of long-term exposure are the results of mechanisms which operate in parallel to those underlying short-term adaptation, with both types co-existing at different timescales. We can therefore interpret Ganguli and Simoncelli (2014) as a model of adaptation at long timescales. Our results obtained here are applicable to adaptation on shorter timescales (Müller et al., 1999; Dragoi et al., 2000, 2002; Benucci et al., 2013). In particular, when married with Bayesian theories of neural representation, our framework predicts repulsive effects around an adaptor stimulus, as shown in Sec. 2.12.

Westrick et al. (2016) also propose a model of the experimental findings of Benucci et al. (2013). Their model (which is not explicitly normative or based on efficient coding) uses divisive normalisation (Carandini and Heeger, 2012) with adaptive weights to achieve homeostasis and stimulus specific repulsion. As discussed above in Sec. 2.11, Bayes-ratio coding (a special case of homeostatic DDCs) can be accomplished by such a divisive normalisation scheme. Our framework therefore yields a normative interpretation of the model of Westrick et al. (2016), and links divisive normalisation with Bayesian representations and efficient coding theory.

Snow et al. (2016) give two alternative models of stimulus specific adaptation in V1. In their models divisive normalisation also plays a key role, but is given a different normative interpretation. They consider generative models of dynamic visual scenes (movies) in the class of mixtures of Gaussian scale mixture models (MGSM). In each of these, the response of a collection of linear filters applied to an image patch are assumed to be independently Gaussian, multiplied by a scaling variable shared across space and time. The task of V1 is then to infer which filters share a common scaling variable, and then "factor out" this scaling variable via divisive normalisation to compute the latent Gaussian variables. Their models are primarily designed to capture adaptation effects over a very short, sub-second timescale. Although these models are able to capture effects such as classical receptive field (CRF) orientation adaptation effects well, they were not designed to account for longer timescale homeostatic adaptation. Accordingly, their models did not successfully reproduce the results of Benucci et al. (2013), in which adaptation occurs on the order of seconds. In particular, each of their two models could only account for either stimulus specific adaptation or neuron specific adaptation, but not both simultaneously. However, as we showed in Sec. 2.12, our homeostatic solution combined with the DDC framework, robustly accounts for stimulus specific adaptation, and furthermore when there is a discrepancy between the internal model's stimulus prior and the true environmental stimulus distribution, it also accounts for neuron specific adaptation. It is possible that Bayesian homeostatic coding coupled with a more sophisticated generative model could account also for very short timescale CRF adaptation effects as well as homeostatic effects. We believe that this would be a fruitful direction for future research.

In order to obtain the analytic results throughout the paper, we have been forced to make a number of assumptions and simplifications. Firstly, we maximise a proxy for mutual information in the form of an upper bound obtained by replacing the marginal entropy with the entropy of a

corresponding Gaussian (see Sec. 2.2). The practice of maximising bounds on mutual information for analytic tractability is standard within efficient coding theory (Ganguli and Simoncelli, 2014; Brunel and Nadal, 1998). A bound will serve as a good proxy for the mutual information itself provided that the bound is maximised in the same region of parameter space as the mutual information. In particular, the tightness of the bound is not in and of itself important. In App. A.12, we argue that both the mutual information and our bound, at least as functions of the cluster firing rates, display the same asymptotic behaviour in the high signal-to-noise regime. This partly justifies our use of the bound as a proxy.

Secondly, our toy model (see section 2.3) imposed a cluster structure on the neural tuning curves, whereby the population is composed of clusters of similarly tuned neurons with small signal cross-correlations between the clusters. We found that total firings rates are equalized across clusters, but the optimal rates of individual neurons within clusters span a broad range. This results from the fact that the efficient coding objective function depends sharply on total cluster firing rates, but changes only slightly when the partition of the cluster firing rate among the cluster's neurons changes. In reality there is not such a sharp distinction between neurons with low and neurons with high signal correlations. As this distinction weakens, higher order terms in Eq. (12) cannot be neglected and hence the justification for dividing the problem into two separate, cluster-level vs. within-cluster, problems, breaks down. However, arguably, in a more realistic model with a more graded variation in neural signal correlations, the optimization landscape will still contain shallow directions (*i.e.*, directions over which the objective changes slowly), corresponding to partitions of firing across similarly tuned neurons, and non-shallow directions corresponding to total firings of such groups of neurons; in which case, we may expect a similar solution to emerge, with firing rate homeostasis at the cluster level, accompanied by heterogeneous firing rates within clusters.

The analytic results we obtained throughout have been in the high signal-to-noise regime for cluster responses. In general, these results fail in the low signal-to-noise ratio limit, as correction terms to our solutions become large. However, as we argued in Sec. 2.6, based on empirical values we expect the signal-to-noise ratio within cortex to be sufficiently high for our results to hold. There is currently no framework within efficient coding theory for analytically predicting response properties in the low signal-to-noise ratio limit using the Infomax principle; in particular, the Fisher Information Lower Bound discussed above is not valid in this regime, and can be violated (Yarrow et al., 2012; Bethge et al., 2002). The development of such a framework presents an important opportunity for future research.

Our analysis here only considers a limited class of noise models. However, these noise models include those of particular biological relevance, such as information-limiting noise correlations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015), signatures of which have been found in the cortex (Rumyantsev et al., 2020), and power-law variability (Goris et al., 2014) (see Sec. 2.8). We note that our homeostatic solution is particularly well suited to the case of information-limiting noise correlations, performing optimally. The core ideas of our framework and derivation could potentially be applied to an even wider class of noise models, such as exponential families (Beck et al., 2007).

Up to section 2.10, we make no assumptions about what the nervous system is attempting to represent via the representational curves. We then apply our framework to a popular theory of Bayesian encoding (namely DDCs), and develop the new idea of Bayes-ratio coding. There is therefore an opportunity to apply our framework to other theories of representation, for example Probabilistic Population Codes (PPCs) (Beck et al., 2007).

In summary, we have introduced a new analytic framework which extends the applicability of the Infomax principle to a broad range of coding schemes. We have shown that this framework provides a normative account for homeostatic regulation of firing rates within a neural population. Based off this homeostatic coding scheme, we derive a novel implementation of probabilistic inference in neural populations, known as Bayes-ratio coding, which can be achieved by divisive normalisation and can invert generative probabilistic models in an adaptive manner. This coding scheme accounts for adaptation effects that are otherwise hard to explain within efficient coding theory. These contributions provide important connections between Bayesian representation, efficient coding theory, and adaptation.

5 Acknowledgements

We thank Máté Lengyel for helpful discussions and Guillaume Hennequin for comments on the manuscript. EY was supported by the UKRI Engineering and Physical Sciences Research Council Doctoral Training Program grant EP/T517847/1. YA was supported by UKRI Biotechnology and Biological Sciences Research Council research grant BB/X013235/1.

A Appendix

A.1 Analytic expression for $\mathcal{L}$

The functional $\mathcal{L}^0(\boldsymbol{g}) = 2\mu I(\boldsymbol{n};\boldsymbol{s}) - \sum_{j=1}^N \mathbb{E}[n_j]$ can be upper bounded by $\mathcal{L}$ using

$$H[\boldsymbol{n}] \leq H[\mathcal{N}(\boldsymbol{h}(\boldsymbol{s}), \mathrm{Cov}[\boldsymbol{n}])] = \frac{1}{2}\ln\left((2\pi e)^N \det\left(\mathrm{Cov}[\boldsymbol{n}]\right)\right)$$

giving

$$\mathcal{L}^0(\boldsymbol{g}) \leq \mathcal{L}(\boldsymbol{g}) = 2\mu\left(\frac{1}{2}\ln\left((2\pi e)^N \det(\mathrm{Cov}[\boldsymbol{n}])\right) - H[\boldsymbol{n}|\boldsymbol{s}]\right) - \sum_{j=1}^N \mathbb{E}[h_j(\boldsymbol{s})]$$

In this section of the appendix we demonstrate that, up to an additive constant,

$$\mathcal{L}(\boldsymbol{g}) = \mu \ln\det\left(I_N + \mathrm{CV}\rho\mathrm{CV}\omega g\right) - \sum_{j=1}^N g_j\omega_j.$$

Here and below, we are using the conventions that the non-bold version of a vector (with no indices) represents the diagonal matrix of that vector. Thus, *e.g.*, , $g = \mathrm{diag}(\boldsymbol{g})$.

Firstly, we demonstrate that $\mathbb{E}[n_j] = g_j\omega_j$.

$$\begin{aligned}
\mathbb{E}[n_j] &= \mathbb{E}[\mathbb{E}[n_j|\boldsymbol{s}]] \\
&= \mathbb{E}[h_j(\boldsymbol{s})] \\
&= g_j\mathbb{E}[\Omega_j(\boldsymbol{s})] \\
&= g_j\omega_j
\end{aligned}$$

Next, we derive an expression for $H[\boldsymbol{n}|\boldsymbol{s}]$. Recall that $n_i|\boldsymbol{s} \sim N(h_i(\boldsymbol{s}), h_i(\boldsymbol{s}))$ independently. In this case,

$$\begin{aligned}
H[\boldsymbol{n}|\boldsymbol{s}] &= \sum_{j=1}^N H[n_j|\boldsymbol{s}] \\
H[n_j|\boldsymbol{s}] &= \int H[N(h_j(\boldsymbol{s}), h_j(\boldsymbol{s}))]P(\boldsymbol{s})d\boldsymbol{s} \\
&= \int \frac{1}{2}\ln(2\pi e h_j(\boldsymbol{s}))P(\boldsymbol{s})d\boldsymbol{s} \\
&= \frac{1}{2}\ln(2\pi e) + \frac{1}{2}\ln(g_j) + \frac{1}{2}\mathbb{E}[\ln(\Omega_j(\boldsymbol{s}))] \\
H[\boldsymbol{n}|\boldsymbol{s}] &= \frac{N}{2}\ln(2\pi e) + \frac{1}{2}\sum_{j=1}^N \ln(g_i) + \sum_{j=1}^N \frac{1}{2}\mathbb{E}\left[\ln(\Omega_j(\boldsymbol{s}))\right]
\end{aligned}$$

Lastly, we find an expression for $\mathrm{Cov}[\boldsymbol{n}]$. We use the decomposition $\mathrm{Cov}[\boldsymbol{n}] = \mathbb{E}[\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}]] + \mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}])$.

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}] &= \mathrm{Var}(\boldsymbol{n}|\boldsymbol{s}) \\
&= h(\boldsymbol{s}) \\
\mathbb{E}[\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}]] &= \mathbb{E}[h(\boldsymbol{s})] \\
&= g\omega \\
\mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}]) &= \mathrm{Cov}(\boldsymbol{h}(\boldsymbol{s})) \\
&= g\mathrm{Cov}(\boldsymbol{\Omega}(\boldsymbol{s}))g \\
\mathrm{Cov}(\boldsymbol{\Omega}(\boldsymbol{s}))_{ij} &= \\
&= \sqrt{\mathrm{Var}[\Omega_i(\boldsymbol{s})]}\frac{\mathrm{Cov}(\Omega_i(\boldsymbol{s}), \Omega_j(\boldsymbol{s}))}{\sqrt{\mathrm{Var}[\Omega_i(\boldsymbol{s})]\ \mathrm{Var}[\Omega_j(\boldsymbol{s})]})}\sqrt{\mathrm{Var}[\Omega_j(\boldsymbol{s})]} \\
&= \omega_i \frac{\sqrt{\mathrm{Var}[\Omega_i(\boldsymbol{s})]}}{\omega_i}\rho_{ij}\frac{\sqrt{\mathrm{Var}[\Omega_j(\boldsymbol{s})]}}{\omega_j}\omega_j \\
&= \omega_i \mathrm{CV}_i \rho_{ij} \mathrm{CV}_j \omega_j \\
\mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}]) &= g\omega\mathrm{CV}\rho\mathrm{CV}\omega g
\end{aligned}$$

Substituting these into the covariance, we obtain

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{n}] &= \mathbb{E}[\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}]] + \mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}]) \\
&= g\omega + g\omega\mathrm{CV}\rho\mathrm{CV}\omega g \\
&= g\omega\left(I_N + \mathrm{CV}\rho\mathrm{CV}\omega g\right) \\
\ln\det\left(\mathrm{Cov}[\boldsymbol{n}]\right) &= \sum_{j=1}^{N}\ln(g_j) + \sum_{j=1}^{N}\ln(\omega_j) + \ln\det\left(I_N + \mathrm{CV}\rho\mathrm{CV}\omega g\right)
\end{aligned}$$

Putting all these parts together gives us

$$\begin{aligned}
\ln\left((2\pi e)^N \det(\mathrm{Cov}[\boldsymbol{n}])\right) - 2H[\boldsymbol{n}|\boldsymbol{s}] &= N\ln(2\pi e) + \ln\det(\mathrm{Cov}[\boldsymbol{n}]) \\
&\quad - N\ln(2\pi e) - \sum_{j=1}^{N}\ln(g_i) - \sum_{j=1}^{N}\mathbb{E}\left[\ln(\Omega_j(\boldsymbol{s}))\right] \\
&= \ln\det(\mathrm{Cov}[\boldsymbol{n}]) - \sum_{j=1}^{N}\ln(g_j) - \sum_{j=1}^{N}\mathbb{E}\left[\ln(\Omega_j(\boldsymbol{s}))\right] \\
&= \ln\det\left(I_N + \mathrm{CV}\rho\mathrm{CV}\omega g\right) \\
&\quad - \sum_{j=1}^{N}\mathbb{E}\left[\ln\left(\frac{\Omega_j(\boldsymbol{s})}{\omega_j}\right)\right]
\end{aligned}$$

Plugging this in, we obtain

$$\mathcal{L}(\boldsymbol{g}) = \mu\ln\det\left(I_N + \mathrm{CV}\rho\mathrm{CV}\omega g\right) - \sum_{j=1}^{N} g_j\omega_j - \mu\sum_{j=1}^{N}\mathbb{E}\left[\ln\left(\frac{\Omega_j(\boldsymbol{s})}{\omega_j}\right)\right]$$

Dropping the last term which does not depend on the gains (and hence the optimisation), we obtain Eq. (9). Note that the remaining two terms depend on g_i only via the combination $r_i \equiv \omega_i g_i$, where

r_i is the average rate of neuron i. This means that, mathematically, optimisation in terms of $\boldsymbol{g}$ is equivalent to optmisation in terms of $\boldsymbol{r}$ (unless for silent neurons with $\omega_i = 0$; but those constitute degenerate cases for which the gain has no real effect and is thus left undetermined by the optimisation). We will thus write most mathematical derivations in the following appendices in terms of $\boldsymbol{r}$ instead of $\boldsymbol{g}$. To this end, we rewrite the above equation as

$$\mathcal{L}(\boldsymbol{g}) = \mu \ln\det\left(I_N + Pr\right) - \sum_{j=1}^{N} r_j + \text{const.}$$

In line with our general notational conventions, r will denote a diagonal matrix with components of $\boldsymbol{r}$ on the diagonal, and $\hat{r}$ denotes the corresponding diagonal matrix for cluster mean rates, as opposed to single-neuron mean rates.

A.2 Expanding $\hat{\mathcal{L}}$ in ϵ

We now take $\rho = \hat{\rho} \otimes \mathbf{1}_k \mathbf{1}_k^T - \epsilon T$ as in Sec. 2.3. Analogously to Eq. (10), we define

$$\hat{P} = \hat{\mathrm{CV}}\, \hat{\rho}\, \hat{\mathrm{CV}} \tag{49}$$

Substituting ρ into Eq. (9), $\mathcal{L}$ becomes (up to an additive constant)

$$\mathcal{L}(\boldsymbol{g}) = \mu \ln\det\left(I_N + \mathrm{CV}[\hat{\rho} \otimes \mathbf{1}_k \mathbf{1}_k^T - \epsilon T]\mathrm{CV}\, r\right) - \sum_{j=1}^{N} r_j \tag{50}$$

$$= \mu \ln\det\left(I_N + \mathrm{CV}[\hat{\rho} \otimes \mathbf{1}_k \mathbf{1}_k^T]\mathrm{CV}\, r - \epsilon \mathrm{CV} T \mathrm{CV}\, r\right) - \sum_{j=1}^{N} r_j \tag{51}$$

$$= \mu \ln\det\Big(I_N + [\hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T] r - \epsilon \mathrm{CV} T \mathrm{CV}\, r\Big) - \sum_{j=1}^{N} r_j \tag{52}$$

$$= \mu \ln\det\Big(I_N + \hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T\ r\Big) - \sum_{j=1}^{N} r_j \tag{53}$$

$$- \epsilon\, \mathrm{Tr}\Bigg(\Big(I_N + \hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T\ r\Big)^{-1} \mathrm{CV}\, T\, \mathrm{CV}\, r\Bigg) + \mathcal{O}(\epsilon^2) \tag{54}$$

We call the zeroth and first order term $\hat{\mathcal{L}}$ and $\mathcal{L}_\epsilon$ respectively, thereby obtaining

$$\mathcal{L} = \hat{\mathcal{L}} + \epsilon \mathcal{L}_\epsilon + \mathcal{O}(\epsilon^2) \tag{55}$$

We now show that the zeroth order term depends only on the cluster rates $\hat{r}_a = \sum_{i \in C_a} r_i$. We start by showing that

$$\ln\det\Big(I_N + \hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T r\Big)$$

can be re-written as

$$\ln\det\Big(I_K + \hat{P}\hat{r}\Big)$$

Consider the following basis for $\mathbb{R}^N$. For each cluster $a = 1, \ldots, K$ define $\boldsymbol{v}^a$ given by

$$\boldsymbol{v}^a = \boldsymbol{e}_a \otimes \mathbf{1}_k$$

where $\boldsymbol{e}_a$ is the ath standard basis element on $\mathbb{R}^K$. Then

$$\hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T r \boldsymbol{v}^a = (\hat{P}\hat{r}\boldsymbol{e}_a) \otimes 1_k = \sum_{b=1}^{K} \Big[\hat{P}\hat{r}\Big]_{ba} \boldsymbol{v}^b$$

We then extend this to a basis of $\mathbb{R}^N$ as follows. For each cluster a, we add an additional $k-1$ vectors which have support only within cluster a. Specifically, for each index $i \neq (a-1)k+1$, we add the vector which is equal to $(-1/r_{(a-1)k+1}, 0, \ldots, 0, 1/r_i, 0, \ldots, 0)$ on cluster a and zero elsewhere. Note that all such vectors are in the kernel of $\hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T r$. In this basis of $\mathbb{R}^N$, $\hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T r$ is represented by the block matrix

$$\begin{pmatrix} \hat{P}\hat{r} & 0 \\ 0 & 0 \end{pmatrix}$$

Accordingly,

$$\begin{aligned} \ln\det(I_N + \hat{P} \otimes \mathbf{1}_k \mathbf{1}_k^T r) &= \ln\det \begin{pmatrix} I_K + \hat{P}\hat{r} & 0 \\ 0 & I_{N-K} \end{pmatrix} \\ &= \ln\Big(\det(I_K + \hat{P}\hat{r})\det(I_{N-K})\Big) \\ &= \ln\det(I_K + \hat{P}\hat{r}) + \text{const.} \end{aligned}$$

Next, note that

$$\begin{aligned} \sum_{i=1}^N r_i = \sum_{i=1}^N g_i \omega_i &\approx \sum_{a=1}^K \sum_{i \in C_a} g_i \hat{\omega}_a \\ &= \sum_{a=1}^K \hat{g}_a \hat{\omega}_a = \sum_{a=1}^K \hat{r}_a \end{aligned}$$

This therefore allows us to write that

$$\hat{\mathcal{L}} = \ln\det(I_K + \hat{P}\hat{r}) - \sum_{a=1}^K \hat{r}_a$$

which is the same as Eq. (13) in Sec. 2.3.

A.3 First order maximisation of $\hat{\mathcal{L}}$

We now consider optimising

$$\hat{\mathcal{L}}(\hat{\boldsymbol{r}}) = \mu \ln\det\Big(I_K + \hat{P}\hat{r}\Big) - \sum_{a=1}^K \hat{r}_a$$

Taking the derivative with respect to $\hat{r}_a$, and setting to zero, we obtain the condition

$$\mu\left[(I_K + \hat{P}\hat{r})^{-1}\hat{P}\right]_{aa} = 1$$

This can also equivalently be written as

$$\mu(I_K + \hat{r}^{-1}\hat{P}^{-1})^{-1}_{aa} = \hat{r}_a$$

or, using the definition Eq. (15) (and hence $\hat{P}^{-1} = \mu\Delta$), as

$$\mu(I_K + \mu\,\hat{r}^{-1}\Delta)^{-1}_{aa} = \hat{r}_a \tag{56}$$

Thus, when Δ is small, we can use a first order Neumann expansion for $\big(I_K + \mu\hat{g}^{-1}\hat{\omega}^{-1}\Delta\big)^{-1}$. To first order, this expansion yields:

$$\big(I_K + \mu\hat{r}^{-1}\Delta\big)^{-1} \approx I_K - \mu\hat{r}^{-1}\Delta$$

Plugging this in to Eq. (56), we obtain

$$\hat{r}_a \approx \mu \left(1 - \frac{\mu}{\hat{r}_a}\Delta_{aa}\right), \tag{57}$$

which can be rearranged to the quadratic equation

$$\left(\frac{\hat{r}_a}{\mu}\right)^2 = \left(\frac{\hat{r}_a}{\mu}\right) - \Delta_{aa}. \tag{58}$$

To zeroth order in Δ, this has the (homeostatic) solution

$$\frac{\hat{r}_a}{\mu} = 1,$$

while to first order, the solution is

$$\frac{\hat{r}_a}{\mu} = 1 - \Delta_{aa}.$$

A.4 Manipulating $\mathcal{L}_\epsilon$

Recall from App. A.2 that

$$\mathcal{L}(\boldsymbol{g}) = \hat{\mathcal{L}}(\hat{\boldsymbol{g}}) + \epsilon\mathcal{L}_\epsilon(\boldsymbol{g}) + \mathcal{O}(\epsilon^2).$$

In this section, we derive a simplified form of $\mathcal{L}_\epsilon$. We begin with the expression Eq. (54) from App. A.2. This is

$$\mathcal{L}_\epsilon(\boldsymbol{g}) = -\mathrm{tr}\left(\left[I_N + (\hat{P}\otimes \mathbf{1}_k\mathbf{1}_k^T)r\right]^{-1}\mathrm{CV}T\mathrm{CV}r\right).$$

We will write $\hat{P}\otimes\mathbf{1}_k\mathbf{1}_k^T = (I_K\otimes\mathbf{1}_k)(\hat{P}\otimes\mathbf{1}_k^T)$ and apply the Woodbury matrix identity to obtain that

$$\begin{aligned}
\left[I_N + (\hat{P}\otimes \mathbf{1}_k\mathbf{1}_k^T)r\right]^{-1} &= I_N - (I_K\otimes\mathbf{1}_k)\left(I_K + (\hat{P}\otimes\mathbf{1}_k^T)r(I_K\otimes\mathbf{1}_k)\right)^{-1}(\hat{P}\otimes\mathbf{1}_k^T)r \\
(\hat{P}\otimes\mathbf{1}_k^T)r(I_K\otimes\mathbf{1}_k) &= \hat{P}\hat{r} \\
\left[I_N + (\hat{P}\otimes \mathbf{1}_k\mathbf{1}_k^T)r\right]^{-1} &= I_N - \left(\left(I_K + \hat{P}\hat{r}\right)^{-1}\hat{P}\right)\otimes\mathbf{1}_k\mathbf{1}_k^T r
\end{aligned}$$

Substituting in and using linearity of the trace, we obtain

$$\begin{aligned}
\mathcal{L}_\epsilon &= -\mathrm{tr}\left(\mathrm{CV}T\mathrm{CV}r\right) \\
&\quad + \mathrm{tr}\left(\left[\left(\left(I_K + \hat{P}\hat{r}\right)^{-1}\hat{P}\right)\otimes\mathbf{1}_k\mathbf{1}_k^T\right]r\mathrm{CV}T\mathrm{CV}r\right).
\end{aligned}$$

The first term is equal to

$$-\sum_{i=1}^{N}\mathrm{CV}_i^2 r_i T_{ii}.$$

But recall that $\rho = \hat{\rho}\otimes\mathbf{1}_k\mathbf{1}_k^T - \epsilon T$. Since $\hat{\rho}$ and ρ are both correlation matrices, their diagonal elements are both equal to 1, and therefore the diagonal elements of T are all zero. So the first term vanishes. This leaves us with the expression

$$\mathcal{L}_\epsilon = \mathrm{tr}\left(\left[\left(\left(I_K + \hat{P}\hat{r}\right)^{-1}\hat{P}\right)\otimes\mathbf{1}_k\mathbf{1}_k^T\right]r\mathrm{CV}T\mathrm{CV}r\right).$$

Note that this is exact, and does not require taking any truncations. We now simplify this functional by truncating to zeroth order in an expansion in

$$\Delta = \left(\mu \hat{\mathrm{CV}} \hat{\rho} \hat{\mathrm{CV}}\right)^{-1}$$

In App. A.3, we demonstrated that to zeroth order, $\hat{r}_a = \mu$, and therefore $\hat{P}\hat{r} = \Delta^{-1}$. Therefore

$$\begin{aligned}
\left(I_K + \hat{P}\hat{r}\right)^{-1} \hat{P} &\approx \left(I_K + \Delta^{-1}\right)^{-1} \hat{P} \\
&= (\Delta + I_K)^{-1} \Delta \hat{P} \\
&= (\Delta + I_K)^{-1} \mu I_K \\
&\approx \mu I_K
\end{aligned}$$

where we have taken a zeroth order expansion in Δ. We substitute this into our functional. We will use $a(i)$ to denote the index of the cluster to which neuron i belongs. We can then write

$$\begin{aligned}
\mathcal{L}_\epsilon &= \mu \operatorname{tr}\left(\left[I_K \otimes \mathbf{1}_k \mathbf{1}_k^T\right] r \mathrm{CV} T \mathrm{CV} r\right) \\
&= \mu \sum_{i=1}^{N} \sum_{j:a(j)=a(i)} r_j \mathrm{CV}_j T_{ji} \mathrm{CV}_i r_i \\
&= \mu \sum_{i=1}^{N} \sum_{j:a(j)=a(i)} r_j \hat{\mathrm{CV}}_{a(j)} T_{ji} \hat{\mathrm{CV}}_{a(i)} r_i \\
&= \sum_{i=1}^{N} \mu \hat{\mathrm{CV}}_{a(i)}^2 \sum_{j:a(j)=a(i)} r_j T_{ji} r_i
\end{aligned}$$

where in deriving the third line we used our assumption that the neurons in a cluster have the same CV. Note that this has decoupled the problem between clusters. In particular, the problem now can be written as a collection the following K problems:

$$\{r_i\}_{i \in C_a} \in \arg\max \left\{ \sum_{i,j \in C_a} r_i T_{ij} r_j \right\} \text{ s.t. } \sum_{i \in C_a} r_i = \hat{r}_a,\ r_i \geq 0$$

However, recall that for $i, j \in C_a$, $T_{ij} = (\hat{\rho}_{aa} - \rho_{ij})/\epsilon$. Substituting this in gives the problems

$$\{r_i\}_{i \in C_a} \in \arg\max \left\{ \hat{\rho}_{aa} \sum_{i,j \in C_a} r_i r_j - \sum_{i,j \in C_a} r_i \rho_{ij} r_j \right\} \text{ s.t. } \sum_{i \in C_a} r_i = \hat{r}_a,\ r_i \geq 0$$

We can use that $\sum_{i \in C_a} r_i = \hat{r}_a$ to rewrite this as

$$\{r_i\}_{i \in C_a} \in \arg\min \left\{ \sum_{i,j \in C_a} r_i \rho_{ij} r_j \right\} \text{ s.t. } \sum_{i \in C_a} r_i = \hat{r}_a,\ r_i \geq 0 \tag{59}$$

As we are assuming that neurons within a cluster have representational curves that are approximately equal, so are the ω_i's within a cluster. To obtain the first order (in ϵ) solution for optimal within-cluster gains, we can ignore small variations of ω_i within the cluster, in which case the optimisation problem Eq. (59) becomes equivalent to Eq. (19).

A.5 Estimates for $\mathrm{tr}(\hat{\rho}^{-1})$

In this appendix, we give estimates for $[\hat{\rho}^{-1}]_{aa}$ through the average value $\mathrm{tr}\left(\hat{\rho}^{-1}\right)/K$. Stringer *et al.* (Stringer et al., 2019) found that in response to a ecological visual stimuli, signal correlations of neural responses display a $1/n$ power-law decay. The eigenvalues of $\hat{\rho}$ are therefore $\lambda_n = A_1/n$ where A_1 normalises the trace to be K. In particular,

$$\begin{aligned}
A_1 &= \frac{K}{\sum_{n=1}^{K} \frac{1}{n}} \\
\frac{K}{A_1} &= \sum_{n=1}^{K} \frac{1}{n} \\
&\approx \int_1^K \frac{1}{x} dx \\
&= \ln(K) \\
A_1 &= \frac{K}{\ln(K)}
\end{aligned}$$

We then calculate

$$\begin{aligned}
\mathrm{tr}\left(\hat{\rho}^{-1}\right) &= \sum_{n=1}^{K} \frac{1}{\lambda_n} \\
&= \frac{1}{A_1} \sum_{n=1}^{K} n \\
&= \frac{1}{A_1} \frac{K(K+1)}{2} \\
\frac{\mathrm{tr}\left(\hat{\rho}^{-1}\right)}{K} &= \frac{1}{A_1} \frac{(K+1)}{2} \\
&\approx \frac{\ln(K)}{2}
\end{aligned}$$

A.6 Homeostatic propagation

In this appendix, we derive how the weights between neural populations must change in order for homeostasis to be propagated between them. We work in a linear rate model.

We start by considering a downstream population of tuning curves $h_m^{(2)}(\boldsymbol{s})$ for $m = 1, \ldots, M$, with W_{mi} the synaptic weight from neuron i to neuron m in the downstream population. The tuning curve of neuron m is

$$h_m^{(2)}(\boldsymbol{s}) = \sum_{i=1}^{N} W_{mi} h_i^{(1)}(\boldsymbol{s}) \tag{60}$$

We assume the downstream population has representational curves $\Omega_m^{(2)}(\boldsymbol{s})$ for $m = 1, \ldots, M$ given by

$$\Omega_m^{(2)}(\boldsymbol{s}) = \sum_{j=1}^{N} w_{mj} \Omega_j^{(1)}(\boldsymbol{s}). \tag{61}$$

If the downstream population is also implementing homeostatic coding, we know that

$$h_m^{(2)}(\boldsymbol{s}) = \frac{\chi \Omega_m^{(2)}(\boldsymbol{s})}{\omega_m^{(2)}},$$

where $\omega_m^{(2)} = \mathbb{E}\left[\Omega_m^{(2)}(\boldsymbol{s})\right]$. But then

$$\begin{aligned}
\omega_m^{(2)} &= \mathbb{E}\left[\Omega_m^{(2)}(\boldsymbol{s})\right] \\
&= \sum_{j=1}^{N} w_{mj}\mathbb{E}\left[\Omega_j^{(1)}(\boldsymbol{s})\right] \\
&= \sum_{j=1}^{N} w_{mj}\omega_j^{(1)} \\
\chi\Omega_m^{(2)}(\boldsymbol{s}) &= \sum_{j=1}^{N} w_{mj}\chi\Omega_j^{(1)}(\boldsymbol{s}) \\
&= \sum_{j=1}^{N} w_{mj}\omega_j^{(1)}h_j^{(1)}(\boldsymbol{s}).
\end{aligned}$$

Substituting this in, we get

$$h_m^{(2)}(\boldsymbol{s}) = \frac{\sum_{j=1}^{N} w_{mj}\omega_j^{(1)}h_j^{(1)}(\boldsymbol{s})}{\sum_{i=1}^{N} w_{mi}\omega_i^{(1)}}. \tag{62}$$

Comparing coefficients, we can see that

$$W_{mj} = \frac{w_{mj}\omega_j^{(1)}}{\sum_{i=1}^{N} w_{mi}\omega_i^{(1)}},$$

as claimed.

A.7 Upper bound for Poissonian noise

In this section, we consider the following model for cluster spike counts.

$$\hat{n}_a|\boldsymbol{s} \sim \text{Poisson}\left(\hat{g}_a\hat{\Omega}_a\right). \tag{63}$$

Here we derive a Gaussian upper bound to the mutual information, and show that an approximation to it leads to the same objective $\hat{\mathcal{L}}$ derived in the uncorrelated, unit Fano-factor Gaussian noise case. We start with the objective

$$\hat{\mathcal{L}}^0(\hat{\boldsymbol{g}}) = 2\mu I(\hat{\boldsymbol{n}};\boldsymbol{s}) - \sum_{a=1}^{K}\hat{g}_a\hat{\omega}_a. \tag{64}$$

We decompose $I(\hat{\boldsymbol{n}};\boldsymbol{s}) = H[\hat{\boldsymbol{n}}] - H[\hat{\boldsymbol{n}}|\boldsymbol{s}]$. We will once again upper bound the marginal entropy $H[\hat{\boldsymbol{n}}]$. A difficulty arises from the fact that a Poisson random variable is discrete and the Gaussian upper bound we previously used is for a continuous random variables. We address this problem as follows: Consider a random variable $\boldsymbol{U}$ which is uniformly distributed on $[0,1)^K$, independent of $\hat{\boldsymbol{n}}$. Then $\hat{\boldsymbol{n}}+\boldsymbol{U}$ is a continuous random variable. We apply the Gaussian bound to this. Let p be the p.d.f. of $\hat{\boldsymbol{n}}+\boldsymbol{U}$,

$\mathcal{P}$ the p.m.f. of $\hat{\boldsymbol{n}}$, and u the p.d.f. of $\boldsymbol{U}$.

$$\begin{aligned}
H[\hat{\boldsymbol{n}}+\boldsymbol{U}] &\leq \frac{K}{2}\ln(2\pi e)+\frac{1}{2}\ln\left(\det\left(\mathrm{Cov}(\hat{\boldsymbol{n}}+\boldsymbol{U})\right)\right)\\
\mathrm{Cov}(\hat{\boldsymbol{n}}+\boldsymbol{U}) &= \mathrm{Cov}(\hat{\boldsymbol{n}})+\mathrm{Cov}(\boldsymbol{U})\\
\mathrm{Cov}(\boldsymbol{U}) &= \frac{1}{12}I_K\\
H[\hat{\boldsymbol{n}}+\boldsymbol{U}] &= -\int_{\mathbb{R}^K} p(\boldsymbol{x})\ln(p(\boldsymbol{x}))d\boldsymbol{x}\\
S(\boldsymbol{y}) &= \{\boldsymbol{x}\in\mathbb{R}^K | y_i\leq x_i < y_i+1\}\\
H[\hat{\boldsymbol{n}}+\boldsymbol{U}] &= -\sum_{\boldsymbol{y}\in\mathbb{N}^K}\int_{S(\boldsymbol{y})} p(\boldsymbol{x})\ln(p(\boldsymbol{x}))d\boldsymbol{x}\\
&= -\sum_{\boldsymbol{y}\in\mathbb{N}^K}\int_{S(\boldsymbol{y})} \mathcal{P}(\boldsymbol{y})u(\boldsymbol{x}-\boldsymbol{y})\ln(\mathcal{P}(\boldsymbol{y})u(\boldsymbol{x}-\boldsymbol{y}))d\boldsymbol{x}\\
&= -\sum_{\boldsymbol{y}\in\mathbb{N}^K}\int_{S(\boldsymbol{y})} \mathcal{P}(\boldsymbol{y})\ln(\mathcal{P}(\boldsymbol{y}))d\boldsymbol{x}\\
&= H[\hat{\boldsymbol{n}}]
\end{aligned}$$

Putting this together gives us the upper bound

$$H[\hat{\boldsymbol{n}}] \leq \frac{1}{2}\ln\left((2\pi e)^K\det\left(\frac{1}{12}I_K+\mathrm{Cov}(\hat{\boldsymbol{n}})\right)\right). \tag{65}$$

We now address the problem of the marginal entropy $H[\hat{\boldsymbol{n}}|\boldsymbol{s}]$. By conditional independence, we have that

$$H[\hat{\boldsymbol{n}}|\boldsymbol{s}] = \sum_{a=1}^{K}\int H[\mathrm{Poisson}(\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}))]P(\boldsymbol{s})d\boldsymbol{s}. \tag{66}$$

We now make use of the assumption discussed in Sec. 2.7 that the representational curves $\hat{\Omega}_a$ have a baseline, and in particular

$$\hat{\Omega}_a \gg 5\hat{\omega}_a/\mu \tag{67}$$

everywhere. Under this condition we can use the fact that for fixed $\boldsymbol{s}$,

$$H[\mathrm{Poisson}(\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}))] \approx H[\mathcal{N}(\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}),\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}))]. \tag{68}$$

This means we can obtain the approximate upper bound:

$$\begin{aligned}
\hat{\mathcal{L}}^0(\hat{\boldsymbol{g}}) &\leq 2\mu\left(\frac{1}{2}\ln\left((2\pi e)^K\det\left(\frac{1}{12}I_K+\mathrm{Cov}(\hat{\boldsymbol{n}})\right)\right)-H[\hat{\boldsymbol{n}}|\boldsymbol{s}]\right)-\sum_{a=1}^{K}\hat{g}_a\hat{\omega}_a\\
&\approx 2\mu\left(\frac{1}{2}\ln\left((2\pi e)^K\det\left(\mathrm{Cov}(\hat{\boldsymbol{n}})\right)\right)-\sum_{a=1}^{K}H[\mathcal{N}(\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}),\hat{g}_a\hat{\Omega}_a(\boldsymbol{s}))]\right)-\sum_{a=1}^{K}\hat{g}_a\hat{\omega}_a\\
&= \hat{\mathcal{L}}(\hat{\boldsymbol{g}})+\text{const.},
\end{aligned}$$

where $\hat{\mathcal{L}}(\hat{\boldsymbol{g}})$ is the same functional that we defined earlier for Gaussian random variables, Eq. (13).

A.8 Power-law and correlated noise

In this appendix, we consider the case of power-law, correlated noise for the clusters. We will take cluster spike counts to be given by

$$\hat{\boldsymbol{n}}|\boldsymbol{s} \sim \mathcal{N}\left(\hat{\boldsymbol{h}}(\boldsymbol{s}),\sigma^2\hat{h}(\boldsymbol{s})^\alpha\Sigma(\boldsymbol{s})\hat{h}(\boldsymbol{s})^\alpha\right),$$

where $\Sigma(\boldsymbol{s})$ is the stimulus-dependent noise correlation matrix, and $0 < \alpha < 2$ is a scaling parameter. The gains will once again be taken to maximise

$$\hat{\mathcal{L}}(\hat{\boldsymbol{g}}) = \mu \left(\ln \left((2\pi e)^K \det \left(\mathrm{Cov}(\hat{\boldsymbol{n}}) \right) \right) - 2H[\hat{\boldsymbol{n}}|\boldsymbol{s}] \right) - \sum_{a=1}^{K} \hat{g}_a \hat{\omega}_a.$$

Much as in appendix A.1, we will simplify this expression. To keep our derivation clean, we will use the expression $\hat{r}_a = \hat{g}_a \hat{\omega}_a$.

We start by deriving an expression for $2H[\boldsymbol{n}|\boldsymbol{s}]$.

$$\begin{aligned}
2H[\boldsymbol{n}|\boldsymbol{s}] &= \int 2H \left[\mathcal{N} \left(\hat{\boldsymbol{h}}(\boldsymbol{s}), \sigma^2 \hat{h}(\boldsymbol{s})^\alpha \Sigma(\boldsymbol{s}) \hat{h}(\boldsymbol{s})^\alpha \right) \right] P(\boldsymbol{s}) d\boldsymbol{s} \\
&= \int \ln \left((2\pi e)^K \det \left(\sigma^2 \hat{h}(\boldsymbol{s})^\alpha \Sigma(\boldsymbol{s}) \hat{h}(\boldsymbol{s})^\alpha \right) \right) P(\boldsymbol{s}) d\boldsymbol{s} \\
&= \ln \left(\det \left(\hat{r}^{2\alpha} \right) \right) + K \ln(2\pi e \sigma^2) \\
&+ \int \ln \left(\det \left(\hat{\omega}^{-\alpha} \hat{\Omega}(\boldsymbol{s})^\alpha \Sigma(\boldsymbol{s}) \hat{\Omega}(\boldsymbol{s})^\alpha \hat{\omega}^{-\alpha} \right) \right) P(\boldsymbol{s}) d\boldsymbol{s} \\
&= \ln \left(\det \left(\hat{r}^{2\alpha} \right) \right) + \text{ const.} \\
&= 2\alpha \sum_{a=1}^{K} \ln(\hat{r}_a) + \text{ const.}
\end{aligned}$$

Next, we find an expression for $\mathrm{Cov}(\hat{\boldsymbol{n}})$. We use the decomposition $\mathrm{Cov}(\hat{\boldsymbol{n}}) = \mathbb{E}[\mathrm{Cov}(\hat{\boldsymbol{n}}|\boldsymbol{s})] + \mathrm{Cov}(\mathbb{E}[\hat{\boldsymbol{n}}|\boldsymbol{s}])$.

$$\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{n}}|\boldsymbol{s}) &= \sigma^2 \hat{h}(\boldsymbol{s})^\alpha \Sigma(\boldsymbol{s}) \hat{h}(\boldsymbol{s})^\alpha \\
&= \sigma^2 \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]} \times \frac{\hat{h}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]}} \Sigma(\boldsymbol{s}) \frac{\hat{h}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]}} \times \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]} \\
\mathbb{E}[\mathrm{Cov}(\hat{\boldsymbol{n}}|\boldsymbol{s})] &= \sigma^2 \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]} W \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]}
\end{aligned}$$

Where we have defined

$$W := \mathbb{E} \left[\frac{\hat{h}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]}} \Sigma(\boldsymbol{s}) \frac{\hat{h}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]}} \right] = \mathbb{E} \left[\frac{\hat{\Omega}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}} \Sigma(\boldsymbol{s}) \frac{\hat{\Omega}(\boldsymbol{s})^\alpha}{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}} \right],$$

which is the stimulus-averaged noise correlation matrix. Next, we find an expression for $\mathrm{Cov}(\mathbb{E}[\hat{\boldsymbol{n}}|\boldsymbol{s}])$

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{n}}|\boldsymbol{s}] &= \hat{\boldsymbol{h}}(\boldsymbol{s}) \\
\mathrm{Cov}(\mathbb{E}[\hat{\boldsymbol{n}}|\boldsymbol{s}]) &= \mathrm{Cov} \left(\hat{\boldsymbol{h}}(\boldsymbol{s}) \right) \\
&= \hat{r} \hat{\mathrm{CV}} \hat{\rho} \hat{\mathrm{CV}} \hat{r} \\
\mathrm{Cov}(\hat{\boldsymbol{n}}) &= \mathbb{E}[\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}]] + \mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}]) \\
&= \sigma^2 \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]} W \sqrt{\mathbb{E}[\hat{h}(\boldsymbol{s})^{2\alpha}]} + \hat{r} \hat{\mathrm{CV}} \hat{\rho} \hat{\mathrm{CV}} \hat{r} \\
&= \sigma^2 \hat{r}^\alpha \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} W \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} \hat{r}^\alpha + \hat{r} \hat{\mathrm{CV}} \hat{\rho} \hat{\mathrm{CV}} \hat{r}.
\end{aligned}$$

We now make use of the α-coefficient of variation,

$$\hat{\mathrm{CV}}_a(\alpha) := \hat{\mathrm{CV}}_a \times \frac{\hat{\omega}_a^\alpha}{\sqrt{\mathbb{E}[\hat{\Omega}_a(\boldsymbol{s})^{2\alpha}]}}.$$

Substituting this in, we obtain

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{n}}) &= \sigma^2 \hat{r}^\alpha \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} W \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} \hat{r}^\alpha + \hat{r}\hat{\mathrm{CV}}\hat{\rho}\hat{\mathrm{CV}}\hat{r} \\ &= \hat{r} \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} [\sigma^2 \hat{r}^{\alpha-1} W \hat{r}^{\alpha-1} + \hat{\mathrm{CV}}(\alpha)\hat{\rho}\hat{\mathrm{CV}}(\alpha)] \frac{\sqrt{\mathbb{E}[\hat{\Omega}(\boldsymbol{s})^{2\alpha}]}}{\hat{\omega}^\alpha} \hat{r}.\end{aligned}$$

Therefore, up to an additive constant, we obtain the expression

$$\begin{aligned}\ln\left(\det\left(\mathrm{Cov}(\hat{\boldsymbol{n}})\right)\right) &= 2\sum_{a=1}^{K} \ln(\hat{r}_a) + \ln\left(\det\left(\sigma^2 \hat{r}^{\alpha-1} W \hat{r}^{\alpha-1} + \hat{\mathrm{CV}}(\alpha)\hat{\rho}\hat{\mathrm{CV}}(\alpha)\right)\right) \\ &= 2\sum_{a=1}^{K} \ln(\hat{r}_a) + \ln(\det(\hat{\mathrm{CV}}(\alpha)\hat{\rho}\hat{\mathrm{CV}}(\alpha))) \\ &\quad + \ln\left(\det\left(\sigma^2 \hat{\mathrm{CV}}(\alpha)^{-1}\hat{\rho}^{-1}\hat{\mathrm{CV}}(\alpha)^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1} + I_K\right)\right) \\ &= 2\sum_{a=1}^{K} \ln(\hat{r}_a) + \ln\left(\det\left(I_K + MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)\right) + \text{ const.},\end{aligned}$$

where we have used the substitution

$$M = \sigma^2 \hat{\mathrm{CV}}(\alpha)^{-1}\hat{\rho}^{-1}\hat{\mathrm{CV}}(\alpha)^{-1} W.$$

Putting these results together gives (up to an additive constant)

$$\hat{\mathcal{L}} = \mu\beta \sum_{a=1}^{K} \ln(\hat{r}_a) - \sum_{a=1}^{K} \hat{r}_a + \mu \ln\left(\det\left(I_K + MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)\right). \tag{69}$$

Taking the derivative of this with respect to $\hat{r}_a$ gives the expression

$$\begin{aligned}0 &= \frac{\beta\mu}{\hat{r}_a} - 1 \\ &\quad + \mu \mathrm{tr}\left(\left(I_K + MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)^{-1} \frac{\partial}{\partial \hat{r}_a}\left(MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)\right).\end{aligned}$$

Notice that the trace is order M. Truncating to zeroth order in M, we have $\hat{r}_a = \beta\mu$. We will also find a first order approximation. Write $\hat{r}_a = \beta\mu(1-\gamma_a)$ where γ_a is order M. Then to first order in M, we have

$$0 = \gamma_a + \mu \mathrm{tr}\left(\left(I_K + MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)^{-1} \frac{\partial}{\partial \hat{r}_a}\left(MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)\right)$$

We take the zeroth order Neumann expansion $\left(I_K + MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)^{-1} \approx I_K$, which gives

$$\begin{aligned}-\gamma_a &= \mu \mathrm{tr}\left(\frac{\partial}{\partial \hat{r}_a}\left(MW^{-1}\hat{r}^{\alpha-1} W \hat{r}^{\alpha-1}\right)\right) \\ &= \mu(\alpha-1)\hat{r}_a^{\alpha-2}\left[W\hat{r}^{\alpha-1}MW^{-1} + MW^{-1}\hat{r}^{\alpha-1}W\right]_{aa} \\ &\approx \mu(\alpha-1)[2(1-\alpha)\mu]^{\alpha-2}\left[W[2(1-\alpha)\mu]^{\alpha-1}MW^{-1} + MW^{-1}[2(1-\alpha)\mu]^{\alpha-1}W\right]_{aa} \\ &= -\frac{1}{2}[\beta\mu]^{-\beta}\left[M^T + M\right]_{aa} \\ &= -[\beta\mu]^{-\beta} M_{aa} \\ \gamma_a &= [\beta\mu]^{-\beta} M_{aa}\end{aligned}$$

This gives the following first order expression

$$\hat{g}_a^{(1)} = \frac{\beta\mu}{\hat{\omega}_a}\left(1 - \frac{\sigma^2}{(\beta\mu)^\beta}\sum_{b=1}^{K}\frac{[\hat{\rho}^{-1}]_{ab}W_{ba}}{\hat{\mathrm{CV}}_a(\alpha)\hat{\mathrm{CV}}_b(\alpha)}\right).$$

and corresponding zeroth order expansion

$$\hat{g}_a^{(0)} = \frac{\beta\mu}{\hat{\omega}_a}.$$

A.9 Optimal homeostatic gains

We now consider the case where we enforce homeostasis on the gains. Prior to taking expectations across environments, our objective, considered as a function of $\hat{r}_a = \hat{g}_a\hat{\omega}_a$ is (up to additive constants)

$$\hat{\mathcal{L}} = -\sum_{a=1}^{K}\hat{r}_a + \mu\log\det\left(I_K + \sigma^{-2}\hat{r}^{1-\alpha}W^{-1}\hat{r}^{1-\alpha}\hat{\mathrm{CV}}(\alpha)\hat{\rho}\hat{\mathrm{CV}}(\alpha)\right)$$

Enforcing homeostasis at the cluster level means setting $\hat{r}_a = \chi$ for all clusters. Substituting this in, we obtain the function

$$\hat{\mathcal{L}}(r = \chi I_K) = -K\chi + \mu\log\det\left(I_K + \chi^{2(1-\alpha)}Q\right) \tag{70}$$

$$= -K\chi + \mu\sum_n \log\left(1 + \chi^\beta\lambda_n\right), \tag{71}$$

where the matrix Q is defined by

$$Q = \sigma^{-2}W^{-1}\hat{\mathrm{CV}}(\alpha)\hat{\rho}\hat{\mathrm{CV}}(\alpha),$$

with λ_n its eigenvalues. Note that this is only a function of the spectrum of Q, and not of the eigenbasis. Therefore, under the assumption that the spectrum remains fixed across environments, we can drop the need to take expectations, and work with $\hat{\mathcal{L}}$ as a function only of the spectrum. We work under this assumption going forward.

The optimal χ, within this family of approximate solutions, obeys

$$0 = -K + \mu\sum_n \frac{\lambda_n\beta\chi^{\beta-1}}{1+\chi^\beta\lambda_n}$$

$$K = \frac{\mu\beta}{\chi}\sum_n \frac{\lambda_n\chi^\beta}{1+\chi^\beta\lambda_n}$$

$$\chi = \frac{\mu\beta}{K}\sum_n\left(1 - \frac{1}{1+\chi^\beta\lambda_n}\right)$$

$$\chi = \mu\beta\left(1 - \frac{1}{K}\sum_n\frac{1}{1+\chi^\beta\lambda_n}\right). \tag{72}$$

We now demonstrate that, under appropriate conditions, $\chi \approx (1/K)\sum_a \hat{g}_a^{(1)}\hat{\omega}_a^{(1)}$.

$$\begin{aligned}
\hat{g}_a^{(1)} &= \frac{\mu\beta}{\hat{\omega}_a}\left(1 - \frac{\sigma^2}{(\beta\mu)^\beta}\sum_{b=1}^{K}\frac{[\hat{\rho}^{-1}]_{ab}W_{ba}}{\hat{\mathrm{CV}}_a(\alpha)\hat{\mathrm{CV}}_b(\alpha)}\right) \\
&= \frac{\mu\beta}{\hat{\omega}_a}\left(1 - \frac{1}{(\beta\mu)^\beta}\sum_{b=1}^{K}[Q^{-1}]_{bb}\right) \\
&= \frac{\mu\beta}{\hat{\omega}_a}\left(1 - \frac{\mathrm{tr}(Q^{-1})}{(\beta\mu)^\beta}\right) \\
&= \frac{\mu\beta}{\hat{\omega}_a}\left(1 - \frac{1}{(\beta\mu)^\beta}\sum_{n=1}^{K}\frac{1}{\lambda_n}\right) \\
\frac{1}{K}\sum_a \hat{g}_a^{(1)}\hat{\omega}_a &= \mu\beta\left(1 - \frac{1}{K}\sum_{n=1}^{K}\frac{1}{(\beta\mu)^\beta\lambda_n}\right)
\end{aligned}$$

Now, under the assumption that $(\beta\mu)^\beta\lambda_n \gg 1$ for each n, we can say that $\chi \approx \mu\beta$ and $(\beta\mu)^\beta\lambda_n \approx \chi^\beta\lambda_n + 1$. Substituting this in gives us the required result. Note that this condition is equivalent to the high signal-to-noise ratio condition we have used throughout.

We now consider the special cases discussed in Sec. 2.7 in which more precise solutions can be obtained. Note that in all three cases the analytics correspond to idealisations of the numerical simulations we actually perform in Sec. 2.8.

Uncorrelated power-law noise In the first special case under consideration, $W = I_K$ and $\hat{\rho}$ has approximately a A_1/n eigenspectrum, where A_1 is chosen to normalise the trace of $\hat{\rho}$ to be equal to K, *i.e.*,

$$A_1 = \frac{K}{\sum_{n=1}^{K}\frac{1}{n}}$$

The spectrum of Q is therefore $\lambda_n = b/n$ where

$$b = \sigma^{-2}A_1\mathrm{CV}(\alpha)^2 \tag{73}$$

In this case, (defining $u = n/K$) we can make the following approximation to the right hand side of Eq. (72):

$$\begin{aligned}
1 - \frac{1}{K}\sum_{n=1}^{K}\frac{1}{1+\chi^\beta\lambda_n} &= \frac{1}{K}\sum_{n=1}^{K}\frac{\chi^\beta\lambda_n}{1+\chi^\beta\lambda_n} \\
&= \frac{1}{K}\sum_{n=1}^{K}\frac{\frac{b\chi^\beta}{n}}{1+\frac{b\chi^\beta}{n}} \\
&\approx \int_{1/K}^{1}\frac{\frac{b\chi^\beta}{Ku}}{1+\frac{b\chi^\beta}{Ku}}du \\
&\approx \frac{b}{K}\chi^\beta\int_0^1\frac{1}{u+\frac{b}{K}\chi^\beta}du \\
&= \frac{b}{K}\chi^\beta\ln\left(\frac{1+\frac{b}{K}\chi^\beta}{\frac{b}{K}\chi^\beta}\right) \\
&= \frac{b}{K}\chi^\beta\ln\left(1+\frac{K}{b\chi^\beta}\right).
\end{aligned}$$

This gives the new equation

$$\chi = \beta\mu \frac{b}{K}\chi^{\beta} \ln\left(1 + \frac{K}{b\chi^{\beta}}\right). \tag{74}$$

Approximating $A_1 \approx K/\ln(K)$ (see App. A.5) and using Eq. (73), we obtain that

$$\frac{K}{b} \approx \frac{\sigma^2 \ln(K)}{\hat{\mathrm{CV}}(\alpha)^2}. \tag{75}$$

Substituting Eq. (75) into Eq. (74) gives us

$$\frac{\chi}{\beta\mu}\left(\frac{\sigma^2 \ln(K)}{\chi^{\beta}\hat{\mathrm{CV}}(\alpha)^2}\right) = \ln\left(1 + \frac{\sigma^2 \ln(K)}{\chi^{\beta}\hat{\mathrm{CV}}(\alpha)^2}\right). \tag{76}$$

Aligned noise In the aligned noise case, we approximate $\hat{\rho}$ and W as having the same eigenbasis. $\hat{\rho}$ still has an eigenspectrum of A_1/n, and we take W to have an eigenspectrum of A_γ/n^γ where A_γ normalises the trace of W to be equal to K,

$$A_\gamma = \frac{K}{\sum_{n=1}^{K} \frac{1}{n^\gamma}}.$$

The matrix Q therefore has eigenspectrum $bn^{\gamma-1}$ where $b = \sigma^{-2}A_1\hat{\mathrm{CV}}^2/A_\gamma$. Inserting this into Eq. (72) gives us

$$\chi = \mu\beta\left(1 - \frac{1}{K}\sum_n \frac{1}{1 + b\chi^{\beta}n^{\gamma-1}}\right). \tag{77}$$

Constant correlation noise The next special case occurs when $\beta = 1$, and $W = (1-p)I_K + p\mathbf{1}\mathbf{1}^T$. Using the Sherman-Morrison formula, we obtain

$$W^{-1} = \frac{1}{1-p}\left(I_K - \frac{p}{1 + p(K-1)}\mathbf{1}\mathbf{1}^T\right).$$

Since $\frac{p}{1+p(K-1)} \leq \frac{1}{K-1}$ we neglect this term and approximate $W^{-1} \approx \frac{1}{1-p}I_K$. This therefore has the same effect as making the replacement $\sigma^2 \mapsto (1-p)\sigma^2$. Substituting this into Eq. (76), and using $\beta = 1$, we obtain

$$\frac{\sigma^2(1-p)\ln(K)}{\mu\hat{\mathrm{CV}}^2} = \ln\left(1 + \frac{\sigma^2(1-p)\ln(K)}{\chi\hat{\mathrm{CV}}^2}\right). \tag{78}$$

We define

$$q = \frac{\sigma^2(1-p)}{\mu\hat{\mathrm{CV}}^2},$$

and rearrange to get

$$\begin{aligned} q\ln(K) &= \ln\left(1 + \frac{\mu q \ln(K)}{\chi}\right) \\ K^q - 1 &= \frac{\mu q \ln(K)}{\chi} \\ \chi &= \mu\frac{q\ln(K)}{K^q - 1}. \end{aligned}$$

A.10 Hierarchical Bayes-ratio coding

In this appendix, we calculate synaptic weights for propagation of Bayes-ratio coding between populations. We start with a generative model $\boldsymbol{z}^{(2)} \to \boldsymbol{z}^{(1)} \to \boldsymbol{s}$. The downstream representational curves are

$$\Omega_m^{(2)}(\boldsymbol{s}) = \Pi\left(\boldsymbol{z}^{(2)} = \boldsymbol{z}_m^{(2)}|\boldsymbol{s}\right).$$

We recall that the synaptic weights W_{mj} are given by the Eq. (38), repeated here:

$$W_{mj} = \frac{w_{mj}\omega_j^{(1)}}{\sum_{i=1}^N w_{mi}\omega_i^{(1)}},$$

where w_{mi} are the coefficients of the expansion $\Omega_m^{(2)}(\boldsymbol{s}) = \sum_i w_{mi}\Omega_i^{(1)}(\boldsymbol{s})$.

We start by calculating w_{mi}.

$$\begin{aligned}
\Omega_m^{(2)}(\boldsymbol{s}) &= \Pi\left(\boldsymbol{z}^{(2)} = \boldsymbol{z}_m^{(2)}|\boldsymbol{s}\right) \\
&= \int g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}^{(1)}\right)\Pi\left(\boldsymbol{z}^{(1)}|\boldsymbol{s}\right)d\boldsymbol{z}^{(1)} \\
&\approx \sum_{j=1}^N g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_j^{(1)}\right)\Pi\left(\boldsymbol{z}^{(1)} = \boldsymbol{z}_j^{(1)}|\boldsymbol{s}\right)\delta\boldsymbol{z}_j^{(1)} \\
&= \sum_{j=1}^N g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_j^{(1)}\right)\delta\boldsymbol{z}_j^{(1)}\Omega_j^{(1)}(\boldsymbol{s}) \\
w_{mj} &= g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_j^{(1)}\right)\delta\boldsymbol{z}_j^{(1)}.
\end{aligned}$$

We plug this into our formula, and use additionally that $\omega_j^{(1)} = \pi\left(\boldsymbol{z}_j^{(1)}\right)$. This gives us

$$\begin{aligned}
\sum_{j=1}^N w_{mj}\omega_j^{(1)} &= \sum_{j=1}^N g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_j^{(1)}\right)\delta\boldsymbol{z}_j^{(1)}\pi\left(\boldsymbol{z}_j^{(1)}\right) \\
&= \int g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_j^{(1)}\right)\pi\left(\boldsymbol{z}_j^{(1)}\right)d\boldsymbol{z}^{(1)} \\
&= \pi\left(\boldsymbol{z}_m^{(2)}\right) \\
W_{mi} &= \frac{w_{mi}\omega_i^{(1)}}{\sum_{j=1}^N w_{mj}\omega_j^{(1)}} \\
&= \frac{g\left(\boldsymbol{z}_m^{(2)}|\boldsymbol{z}_i^{(1)}\right)\delta\boldsymbol{z}_i^{(1)}\pi\left(\boldsymbol{z}_i^{(1)}\right)}{\pi\left(\boldsymbol{z}_m^{(2)}\right)} \\
&= g\left(\boldsymbol{z}_i^{(1)}|\boldsymbol{z}_m^{(2)}\right)\delta\boldsymbol{z}_i^{(1)},
\end{aligned}$$

which gives the result as required.

A.11 Stimulus specific adaptation for non-ideal-observer models

In this appendix, we show how discrepancies between the internal model and the external environment lead to simultaneous stimulus specific and neuron specific adaptation effects in a homeostatic DDC code.

We begin with the special case of a Bayes-ratio code, which is a DDC in which the kernel functions are delta functions (as discussed in Sec. 2.11). In a Bayes-ratio code, representational curves are given by

$$\Omega_i(\boldsymbol{s}) = \Pi(\boldsymbol{z}_i|\boldsymbol{s}), \tag{79}$$

where Π is the posterior distribution under the internal generative model. We consider the case of a non-ideal-observer generative model, in which case there will be a discrepancy between the marginal stimulus distribution predicted by the model, $P^I(\boldsymbol{s})$ and the external environment stimulus marginal $P^E(\boldsymbol{s})$. We reason as follows:

$$\begin{aligned}\Omega_i(\boldsymbol{s}) &= \Pi^I(\boldsymbol{z}_i|\boldsymbol{s}) \\ &= \frac{f(\boldsymbol{s}|\boldsymbol{z}_i)\pi^I(\boldsymbol{z}_i)}{P^I(\boldsymbol{s})}\end{aligned} \tag{80}$$

where

$$\begin{aligned}P^I(\boldsymbol{s}) &\equiv \int f(\boldsymbol{s}|\boldsymbol{z})\pi^I(\boldsymbol{z})d\boldsymbol{z} \\ &\approx \sum_i f(\boldsymbol{s}|\boldsymbol{z}_i)\pi^I(\boldsymbol{z}_i)\Delta\boldsymbol{z}\end{aligned} \tag{81}$$

According to homeostatic coding, the tuning curve is given by

$$h_i(\boldsymbol{s}) = \chi\frac{\Omega_i(\boldsymbol{s})}{\omega_i} \tag{82}$$

where ω_i is the temporal average of the representational curve, or equivalently, the expectation of $\Omega_i(\boldsymbol{s})$ under the *true* environmental stimulus distribution. Thus

$$\begin{aligned}\omega_i &= \mathbb{E}^E[\Omega_i(\boldsymbol{s})] \\ &= \int \frac{f(\boldsymbol{s}|\boldsymbol{z}_i)\pi^I(\boldsymbol{z}_i)}{P^I(\boldsymbol{s})}P^E(\boldsymbol{s})d\boldsymbol{s} \\ &= \pi^I(\boldsymbol{z}_i)F_i\end{aligned} \tag{83}$$

where we defined

$$F_i := \int f(\boldsymbol{s}|\boldsymbol{z}_i)\frac{P^E(\boldsymbol{s})}{P^I(\boldsymbol{s})}d\boldsymbol{s}. \tag{84}$$

Notice that when $P^E = P^I$, as would be the case for an ideal-observer, we will have $F_i = 1$, due to the normalization of $f(\boldsymbol{s}|\boldsymbol{z})$. Substituting Eq. (80) and Eq. (83) into Eq. (82) we obtain

$$h_i(\boldsymbol{s}) = \chi\frac{f(\boldsymbol{s}|\boldsymbol{z}_i)}{P^I(\boldsymbol{s})F_i} \tag{85}$$

as claimed in Eq. (46).

We now consider the more general case of a DDC code. We define

$$F(\boldsymbol{z}) = \int f(\boldsymbol{s}|\boldsymbol{z})\frac{P^E(\boldsymbol{s})}{P^I(\boldsymbol{s})}d\boldsymbol{s}, \tag{86}$$

so that F_i in the above is equal to $F(\boldsymbol{z}_i)$. An analogous derivation yields:

$$h_i(\boldsymbol{s}) = \chi\frac{\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z})f(\boldsymbol{s}|\boldsymbol{z})d\boldsymbol{z}}{P^I(\boldsymbol{s})\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z})F(\boldsymbol{z})d\boldsymbol{z}}. \tag{87}$$

We can see that in this case there is no clean separation of neuron specific and stimulus specific factors. In particular, the generalization of Eq. (48) takes the form:

$$\frac{h_i(\boldsymbol{s}; v=1)}{h_i(\boldsymbol{s}; v=0)} = \frac{P^I(\boldsymbol{s}; v=0)}{P^I(\boldsymbol{s}; v=1)} \times \frac{\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z}; v=1) f(\boldsymbol{s}|\boldsymbol{z}) d\boldsymbol{z}}{\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z}; v=0) f(\boldsymbol{s}|\boldsymbol{z}) d\boldsymbol{z}} \times \frac{\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z}; v=0) F(\boldsymbol{z}; v=0) d\boldsymbol{z}}{\int \phi_i(\boldsymbol{z})\pi^I(\boldsymbol{z}; v=1) F(\boldsymbol{z}; v=1) d\boldsymbol{z}}. \quad (88)$$

In the limit as the width of the kernels becomes small, the integrals involving kernels collapse to sampling at a single point, as occurs with the Bayes-ratio code. When this happens, we get full separation to stimulus and neuron specific effects; otherwise these effects are mixed together, with more mixing occurring the larger the width of the kernels. However, when the kernels are unimodal and sufficiently narrow, we would expect an approximate factorization of the effect of adaptation into a stimulus-specific and a neuron-specific suppression factor.

A.12 Relationship between Gaussian upper bound and Fisher Information lower bound

In this appendix, we elaborate on the relationship between the Gaussian upper bound we have used in our theoretical arguments and the Fisher Information lower bound popularly used within the efficient coding literature. The Fisher Information lower bound (FILB), first derived by Bruner and Nadal (Brunel and Nadal, 1998), states that

$$I(\boldsymbol{n}; \boldsymbol{S}) \geq I_{\text{Fisher}} = H[\boldsymbol{s}] + \frac{1}{2} \int \ln \left(\frac{\det(J(\boldsymbol{s}))}{(2\pi e)^N} \right) P(\boldsymbol{s}) d\boldsymbol{s}, \quad (89)$$

where J is the Fisher Information of the neural response as a function of the stimulus. Note that this lower bound is only valid in the high signal-to-noise ratio case, when J is large; for lower signal-to-noise ratios, the bound does not hold (Yarrow et al., 2012; Bethge et al., 2002).

For a Poissonian noise model, the Fisher Information is given by

$$J(\boldsymbol{s}) = \sum_{i=1}^{N} \frac{\nabla h_j \nabla^T h_j}{h_j} = \sum_{i=1}^{N} g_j \frac{\nabla \Omega_j \nabla^T \Omega_j}{\Omega_j}. \quad (90)$$

In the framework we have been working in, neuron tuning curves are well approximated by their cluster tuning curves. This gives

$$J(\boldsymbol{s}) \approx \sum_{a=1}^{K} \left(\sum_{i \in C_a} g_i \right) \frac{\nabla \hat{\Omega}_a \nabla^T \hat{\Omega}_a}{\hat{\Omega}_a} = \sum_{a=1}^{K} \hat{g}_a \frac{\nabla \hat{\Omega}_a \nabla^T \hat{\Omega}_a}{\hat{\Omega}_a}. \quad (91)$$

Therefore, we see that (neglecting corrections to the cluster tuning curves) both the Fisher information lower bound and the Gaussian upper bound depend on the cluster firing rates.

Moreover, both bounds display the same asymptotic behaviour in the magnitide of the gains, namely that $I(\hat{\boldsymbol{n}}; \boldsymbol{I}) \sim \frac{1}{2} \ln(\hat{g})$ as $\hat{g} \to \infty$. This asymptotic behaviour is crucial to understanding the core of the results obtained above. We now examine where this behaviour arises from.

We first claim that mutual information scales like 1/2 log of the signal-to-noise ratio squared of the neural responses. The Gaussian upper bound approximates neural information flow as a Gaussian channel. In general, if signal $\boldsymbol{X} \sim N(0, \Sigma_s)$ is perturbed with Gaussian noise $\boldsymbol{Z} \sim N(0, \Sigma_n)$, then the resultant mutual information is

$$I(\boldsymbol{X}, \boldsymbol{X} + \boldsymbol{Z}) = \frac{1}{2} \ln \det \left(I + \Sigma_n^{-1} \Sigma_s \right) = \frac{1}{2} \ln \det(I + \text{SNR}^2) \quad (92)$$

Up to an additive constant, this is the same expression we derive for the information term in $\hat{\mathcal{L}}$. In the Fisher Information Lower Bound (Eq. (91)) interpreting J as the signal-to-noise ratio gives the

same qualitative behaviour. For Poissonian neurons, with Fisher Information given by Eq. (90), this interpretation comes from interpreting the numerator $\nabla \hat{h}_a \nabla^T \hat{h}_a$ as a squared signal term (quantifying the sensitivity of cluster a to changes in stimuli) and the denominator $\hat{h}_a$ as a squared noise term (a baseline level of spike firing which obscures changes in firing rate).

Next, we assume that the signal scales linearly with the gain, and the noise scales sublinearly, specifically as $\sqrt{\hat{g}}$. It follows that the squared signal-to-noise ratio scales like $\hat{g}$. In the Gaussian channel case, the squared signal is the firing rate covariance $\mathrm{Cov}(\mathbb{E}[\boldsymbol{n}|\boldsymbol{s}])$ and the squared noise is the mean spike count variance $\mathbb{E}[\mathrm{Cov}[\boldsymbol{n}|\boldsymbol{s}]]$. The squared signal term scales quadratically with the firing rate, and therefore the gain. For constant Fano factor neurons, the squared noise term scales linearly with the mean, and therefore linearly with the gain, giving the required result. Taking the inverse of the squared noise gives us the expression

$$\frac{1}{2} \ln \det \left(I_K + \hat{P} \hat{\omega} \hat{g} \right) \tag{93}$$

which should be compared to the Gaussian Channel result Eq. (92) above. For the Fisher Information expression (Eq. (91)), the squared "signal" term $\nabla \hat{h}_a \nabla^T \hat{h}_a = \hat{g}_a^2 \nabla \hat{\Omega}_a \nabla^T \hat{\Omega}_a$ scales quadratically, and the squared noise $\hat{h}_a = \hat{g}_a \hat{\Omega}_a$ scales linearly.

We have therefore shown that for both the upper and lower bounds, the mutual information behaves approximately as 1/2 log of the gains. This requires that we are in a high signal-to-noise regime. For the Fisher Information, this requires a large number of neurons and a long coding interval, with the signal vectors $\nabla \hat{\Omega}_a$ not too cluster together in stimulus space. Our condition on the matrix Δ is an analogous high signal-to-noise condition.

Since metabolic cost scales linearly with the magnitude of the gains, our target function $\hat{\mathcal{L}}$ has the asymptotic behaviour as a function of the magnitude of the gains

$$\hat{\mathcal{L}} \sim \mu \ln(\hat{g}) - \hat{g} \hat{\omega}. \tag{94}$$

This function is maximised when $\hat{g} = \mu / \hat{\omega}$, which yields homeostasis.

Although the arguments that we have pursued in earlier sections revolve around use of one particular upper bound, this gives us reason to believe that our results are robust to modelling assumptions. We have demonstrated in this section that the mutual information is upper and lower bounded by terms that scale as 1/2 log gain, and therefore must also display the same scaling. Moreover, we have shown that this scaling, combined with a linear metabolic cost term, yields homeostasis as an optimal solution. There is therefore reason to believe that other approximations to mutual information will give similar results, provided we are in a high signal-to-noise ratio regime.

References

Atick, J. J. and Redlich, A. N. (1990). Towards a Theory of Early Visual Processing. *Neural Computation*, 2(3):308–320.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193.

Barlow, H. B. (2012). Possible Principles Underlying the Transformations of Sensory Messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 216–234. The MIT Press.

Beck, J., Ma, W. J., Latham, P. E., and Pouget, A. (2007). Probabilistic population codes and the exponential family of distributions. In Cisek, P., Drew, T., and Kalaska, J. F., editors, *Progress in Brain Research*, volume 165 of *Computational Neuroscience: Theoretical Insights into Brain Function*, pages 509–519. Elsevier.

Benucci, A., Saleem, A. B., and Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature Neuroscience*, 16(6):724–729.

Bethge, M., Rotermund, D., and Pawelzik, K. (2002). Optimal Short-Term Population Coding: When Fisher Information Fails. *Neural Computation*, 14(10):2317–2351.

Brunel, N. and Nadal, J. P. (1998). Mutual information, Fisher information, and population coding. *Neural Computation*, 10(7):1731–1757.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews. Neuroscience*, 15(4):264–278.

Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62. Number: 1 Publisher: Nature Publishing Group.

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., and Schwartz, O. (2007). Visual adaptation: neural, psychological and computational aspects. *Vision Research*, 47(25):3125–3131.

Desai, N. (2003). Homeostatic plasticity in the CNS: Synaptic and intrinsic forms. *Journal of physiology, Paris*, 97:391–402.

Dragoi, V., Sharma, J., Miller, E. K., and Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature Neuroscience*, 5(9):883–891.

Dragoi, V., Sharma, J., and Sur, M. (2000). Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. *Neuron*, 28(1):287–298.

Ganguli, D. and Simoncelli, E. P. (2014). Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations. *Neural Computation*, 26(10):2103–2134.

Geurts, L. S., Cooke, J. R. H., van Bergen, R. S., and Jehee, J. F. M. (2022). Subjective confidence reflects representation of Bayesian probability in cortex. *Nature Human Behaviour*, 6(2):294–305. Number: 2 Publisher: Nature Publishing Group.

Ghisovan, N., Nemri, A., Shumikhina, S., and Molotchnikoff, S. (2009). Long adaptation reveals mostly attractive shifts of orientation tuning in cat primary visual cortex. *Neuroscience*, 164(3):1274–1283. Publisher: Elsevier.

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865.

Greenberg, D. S., Houweling, A. R., and Kerr, J. N. D. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature Neuroscience*, 11(7):749–751.

Hengen, K. B., Lambo, M. E., Van Hooser, S. D., Katz, D. B., and Turrigiano, G. G. (2013). Firing Rate Homeostasis in Visual Cortex of Freely Behaving Rodents. *Neuron*, 80(2):335–342.

Kanitscheider, I., Coen-Cagli, R., and Pouget, A. (2015). Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences of the United States of America*, 112(50):E6973–6982.

Keck, T., Keller, G. B., Jacobsen, R. I., Eysel, U. T., Bonhoeffer, T., and Hübener, M. (2013). Synaptic Scaling and Homeostatic Plasticity in the Mouse Visual Cortex In Vivo. *Neuron*, 80(2):327–334.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55:271–304.

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, 97(5):3155–3164.

Laughlin, S. (1981). A Simple Coding Procedure Enhances a Neuron's Information Capacity. *Zeitschrift für Naturforschung C*, 36(9-10):910–912. Publisher: De Gruyter.

Lennie, P. (2003). The cost of cortical computation. *Current biology: CB*, 13(6):493–497.

Levy, W. and Baxter, R. (1996). Energy Efficient Neural Codes. *Neural computation*, 8:531–43.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, 21(3):105–117. Conference Name: Computer.

Maffei, A. and Turrigiano, G. G. (2008). Multiple Modes of Network Homeostasis in Visual Cortical Layer 2/3. *The Journal of Neuroscience*, 28(17):4377–4384.

Marder, E. and Prinz, A. A. (2003). Current compensation in neuronal homeostasis. *Neuron*, 37(1):2–4.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*, 17(10):1410–1417.

Movshon, J. A. and Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurones. *Nature*, 278(5707):850–852.

Müller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P. (1999). Rapid Adaptation in Visual Cortex to the Structure of Images. *Science*, 285(5432):1405–1408.

Nadal, J.-P. and Parga, N. (1994). Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. *Network: Computation in Neural Systems*, 5(4):565–581. Publisher: Taylor & Francis _eprint: https://doi.org/10.1088/0954-898X_5_4_008.

Nadal, J.-P. and Parga, N. (1999). Sensory coding: information maximization and redundancy reduction. In *Neuronal Information Processing*, volume Volume 7 of *Series in Mathematical Biology and Medicine*, pages 164–171. WORLD SCIENTIFIC.

Obermayer, K. and Blasdel, G. G. (1993). Geometry of orientation and ocular dominance columns in monkey striate cortex. *Journal of Neuroscience*, 13(10):4114–4129. Publisher: Society for Neuroscience Section: Articles.

Parker, P. R. L., Abe, E. T. T., Leonard, E. S. P., Martins, D. M., and Niell, C. M. (2022). Joint coding of visual input and eye/head position in V1 of freely moving mice. *Neuron*, 110(23):3897–3906.e5.

Rumyantsev, O. I., Lecoq, J. A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M. J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, 580(7801):100–105. Number: 7801 Publisher: Nature Publishing Group.

Sanzeni, A., Palmigiano, A., Nguyen, T. H., Luo, J., Nassi, J. J., Reynolds, J. H., Histed, M. H., Miller, K. D., and Brunel, N. (2023). Mechanisms underlying reshuffling of visual responses by optogenetic stimulation in mice and monkeys. *Neuron*, 0(0).

Schwartz, O., Hsu, A., and Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, 8(7):522–535.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, 24(1):1193–1216.

Slomowitz, E., Styr, B., Vertkin, I., Milshtein-Parush, H., Nelken, I., Slutsky, M., and Slutsky, I. (2015). Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLife*, 4:e04378.

Snow, M., Coen-Cagli, R., and Schwartz, O. (2016). Specificity and timescales of cortical adaptation as inferences about natural movie statistics. *Journal of Vision*, 16(13):1.

Solomon, S. G. and Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. *Current biology: CB*, 24(20):R1012–1022.

Stocker, A. A. and Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585. Number: 4 Publisher: Nature Publishing Group.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365. Number: 7765 Publisher: Nature Publishing Group.

Szuts, T. A., Fadeyev, V., Kachiguine, S., Sher, A., Grivich, M. V., Agrochão, M., Hottowy, P., Dabrowski, W., Lubenov, E. V., Siapas, A. G., Uchida, N., Litke, A. M., and Meister, M. (2011). A wireless multi-channel neural amplifier for freely moving animals. *Nature Neuroscience*, 14(2):263–269.

Torrado Pacheco, A., Tilden, E. I., Grutzner, S. M., Lane, B. J., Wu, Y., Hengen, K. B., Gjorgjieva, J., and Turrigiano, G. G. (2019). Rapid and active stabilization of visual cortical firing rates across light–dark transitions. *Proceedings of the National Academy of Sciences*, 116(36):18068–18077.

Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3):422–435.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391(6670):892–896.

Turrigiano, G. G. and Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5(2):97–107. Number: 2 Publisher: Nature Publishing Group.

van Bergen, R. S., Ji Ma, W., Pratte, M. S., and Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, 18(12):1728–1730. Number: 12 Publisher: Nature Publishing Group.

Vertes, E. and Sahani, M. (2018). Flexible and accurate inference and learning for deep generative models. arXiv:1805.11051 [cs, stat].

Wandell, B. A. (1995). *Foundations of vision*. Foundations of vision. Sinauer Associates, Sunderland, MA, US. Pages: xvi, 476.

Wei, X.-X. and Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, 18(10):1509–1517.

Westrick, Z. M., Heeger, D. J., and Landy, M. S. (2016). Pattern Adaptation and Normalization Reweighting. *Journal of Neuroscience*, 36(38):9805–9816. Publisher: Society for Neuroscience Section: Articles.

Yarrow, S., Challis, E., and Seriès, P. (2012). Fisher and Shannon information in finite neural populations. *Neural Computation*, 24(7):1740–1780.