Appendices

Background on large-scale biophysical models

In the literature, the different large-scale modelling approaches can be grouped into three main categories:

- **Data-driven approaches**: these methods usually borrow from digital signal processing, and summarise key aspects of the observed dynamics (*e.g.* spectral profiles) with few parameters. For example, auto-regressive methods [34, 46] formulate explicit lagged dependencies between successive timepoints, typically in a multivariate fashion in order to capture structural interactions between distant brain regions. Although such approaches are designed to accurately reproduce the observed signals (often under linear assumptions), and can reveal structural dependencies in the data [45], they do not provide any insight into the underlying biophysical processes.

- **Neuronal-mass models**: these methods describe the average dynamics of large populations of densely connected neurons, with parametric systems of ordinary differential equations [60, 47]. In comparison with the previous data-driven approaches, these models are typically derived from first principles with a concern to remain faithful to the (relevant) underlying biophysics. Importantly, these models are inherently *local*, because they neglect the spatial dependency of the dynamics (*e.g.* due to axonal conduction delays) in order to work with functions of time only. However, large-scale models can be constructed, by considering *networks* of neuronal masses, with vertices summarising the local activity in different brain regions, and with edges representing (delayed) interactions between these regions.

- **Neural-field models**: these methods extend the previous mass-models, by considering brain activity as a diffusion process over the cortical surface [5, 13]. The temporal dynamics at any given point on the surface are derived from local equations, but the diffusion process allows nearby points to affect each other, and the ensuing propagation across the surface typically leads to oscillatory activity through *travelling wave* mechanisms. Long-range interactions with delays can also be included by modifying the topology of the cortical surface, and multiple layers can be considered to model interactions between different types of neurons [14]. However, these models are theoretically and computationally complex compared to neuronal masses, and less developed to date. Furthermore, although quantitative differences have been demonstrated [48], it is unclear whether this added complexity is *required* in order to produce dynamics of interest in large-scale models.

Additionally, it is worth noting that large-scale *spiking* models considering very large networks of individual neurons have been attempted [38], but their overwhelming complexity precludes any practical exploration of their capabilities.

In this paper, we focused on the use of neuronal-mass models, as a reasonable compromise between complexity and biophysical pertinence. The general structure of these models is best described as a discrete **network**; in particular, the different brain regions are associated with network **units**, which may themselves be composed of several **nodes** associated with different local populations of neurons.

Evaluation of GPSO in dimension 5

Below, we assess the ability of GPSO to converge to the global optimum in the case of a five-dimensional search-space. For reference, we compare the performance achieved to that of a sequential Monte-Carlo method (particle filter). We emphasize, however, that the two methods differ fundamentally in their approach, and that this comparison is not intended as a competition; random sampling methods are built around the property of ergodicity (*i.e.* the probability of sampling any open subspace is positive, hence any region will *eventually* be sampled), whereas space-partitioning methods like GPSO implement a multi-scale "divide-and-conquer" philosophy, relying on the smoothness of the objective function. Therefore, as noted previously in §2.1,

random sampling methods are undesirable in practice in the case of expensive objective functions, because they do not attempt to minimise the number of function evaluations, and yield variable results for small sample sizes.

In order to assess the convergence of GPSO fairly and reliably, we constructed *ten* objective functions as mixtures of (isotropic) Gaussians, with 5 modes (or peaks) each, restricted to the space $(0,1)^5$. We imposed that the modes be far apart from eachother, mainly to ensure that the global optimum would be one of the peaks (allowing multiple peaks to merge makes it difficult to analytically locate the global optimum).

For each peak of each mixture, the width σ and amplitude A were sampled randomly, respectively in $(0.1, 0.2)$ and $(1, 5)$. Then, the location μ was sampled randomly within $(\sigma, 1-\sigma)^5$, and the peak was *rejected* if either:

- the value of the mixture considering previous peaks only was greater than $V = 0.3$;
- or the closest previous peak was at a distance closer than:

$$\sigma\sqrt{2\log\left(\frac{A}{V}\right)}$$

Once the peaks were chosen for each mixture, GPSO was run for each mixture with 800 samples, using the same parameters given in Tab. 3. The performance of each run was measured in terms of:

- whether the global optimum was found;
- if so, with the closest distance to the optimum;
- if not, with the distance to the closest peak.

We found that GPSO converged to the global optimum nine out of ten times (despite the presence of local extrema), and the distance to the optimum in these cases was 0.0096 ± 0.0062. The one time it converged to a local maximum instead, the distance to the closest peak was 0.013, and the difference in height between this peak (4.34) and the global maximum (4.90) was 0.56.

This performance was compared to a sequential Monte-Carlo sampling scheme, namely *particle-filter* optimisation [21, 42]. This scheme proceeded over eight cycles, each with 100 samples (*i.e.* 800 samples in total, as previously with GPSO), as follows:

- the first 100 samples were selected uniformly randomly;
- at each cycle, the last third of the samples were resampled uniformly, and the best two-thirds were resampled following an SIR procedure (sequential importance resampling), using an isotropic Gaussian distribution with standard-deviation 0.1.

Due to the stochastic nature of this procedure, we repeated the optimisation ten times for each mixture function, and estimated the success rate across these ten trials, based on whether the particle with the highest score (across all 800) was closest to the highest peak. Overall, the success-rate was of 69%; the distance to the global optimum in these cases was 0.085 ± 0.027; and otherwise, the distance to the closest optimum was 0.080 ± 0.028.

It is worth noting that the average time GPSO required in order to carry out this optimisation was 11.4 min, whereas the particle sampling scheme completed on average within 10 ms. This illustrates the fact that GPSO is only valuable in the case of *costly* objective functions, for which the time required to carry out the optimisation is negligible compared to the time required to evaluate the various samples.