Unsupervised adaptive optimization of motion-sensitive systems guided by measurement uncertainty

Peter Jurica[1], Sergei Gepshtein[2], Ivan Tyukin[2,3], Danil Prokhorov[4], Cees van Leeuwen[5]
[1] *Lab. for Perceptual Dynamics, RIKEN Brain Science Institute*
2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan, pjurica@brain.riken.jp
[2] *Lab. for Perceptual Dynamics, RIKEN Brain Science Institute, sergei.gepshtein@gmail.com*
[3] *Dept. of Mathematics, University of Leicester*
University Road, Leicester, LE1 7RH, UK, I.Tyukin@le.ac.uk
[4] *Toyota Technical Center, Ann Arbor 48105, USA, dvprokhorov@gmail.com*
[5] *Lab. for Perceptual Dynamics, RIKEN Brain Science Institute, ceesvl@brain.riken.jp*

Abstract

We propose a design for adaptive optimization of sensory systems. We consider a network of sensors that measure stimulus parameters as well as the uncertainties associated with these measurements. No prior assumptions about the stimulation and measurement uncertainties are built into the system, and properties of stimulation are allowed to vary with time. We present two approaches: one is based on estimation of the local gradient of uncertainty, and the other on random adjustment of cell tuning. Either approach steers the network towards its optimal state.

1. INTRODUCTION

Intelligent systems, artificial or biological, can greatly benefit from the ability to estimate parameters of time-varying optical information. In manipulator robotics, for instance, this ability can help controlling the actuators. In automotive applications, optical information can be used to estimate the speed of moving objects and thus improve traffic control. Measurement of optical information, however, presents a number of difficulties. In particular, the quality of measurements is marred by intrinsic measurement uncertainty [2] that depends on both the properties of sensors and the properties of stimulation.

Performance of sensory systems in face of the measurement uncertainty also depends on their ability to effectively use limited computational resources, whether the "resources" is a pool of processing units that receives signals from light-sensitive elements in a CCD camera of a given resolution, or a pool of motion-sensitive cells in a biological neural network. A systematic approach to how limited resources ought to be used for effective measurement of time-varying optical information was recently proposed in [4], where speed estimation in biological vision was studied using a new normative-economic framework. The authors showed that the normative prescriptions were consistent with experimental evidence in human vision [3] [5].

The normative-economic framework in [4] was used to investigate how a pool of motion-sensitive cells should be allocated in an environment whose statistics was known and stationary. Yet the real systems may have no such *a priori* information. Also, such systems would benefit from reallocating resources as the environmental statistics changes. In either case, the systems should be able to discover properties of the environment autonomously. We presently address this issue by asking how a visual system can tune itself to properties of time-varying environments for efficient motion measurement.

We propose a simple algorithm for unsupervised optimal self-tuning of elements in a motion-sensitive system that faces a time-varying environment. The proposed tuning mechanism is local and probabilistic in nature. We demonstrate that it steers the system toward the optimal configurations derived analytically in [4].

In the following, we first summarize an analytic solution to the problem of optimal resource allocation for motion measurement (Section 2). Then we introduce our approach to adaptive optimization and describe how we implement it (Section 3). We conclude by illustrating results of our computational experiments (Section 4).

2. PRELIMINARIES

It is convenient to describe the analytic results [4] and results of present computational experiments using a graphical convention called the Distance Plot (Fig. 1). Each point in the plot represents the tuning parameters of a standard motion sensor [1], [7], [8]. The two parameters can be thought of as the temporal span and delay of the motion sensors, or as the spatial and temporal wavelengths of the signals that stimulate the sensors optimally.

We consider a system that consists of N such sensors, each tuned to a combination of spatial and temporal parameters, S and T in Fig. 1. Every such sensor is thus also tuned to speed, $v = S/T$ (or $v = dS/DT$), which is the fundamental dimension of biological motion estimation [5], [6]. According

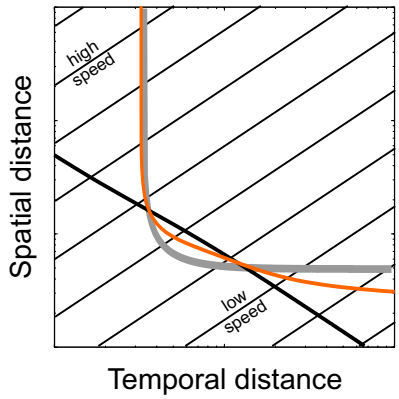

Fig. 1: *Optimal sets for speed estimation in space-time Distance Plot.* The oblique lines (speed lines) correspond to different speeds; these lines are parallel to each other in the logarithmic coordinates. The three curves indicate sets of optimal conditions for speed estimation. The black curve represents conditions ideal for measuring every speed (*local optimal set*); it can be achieved only in a system with infinite resources. The grey hyperbola is the *integral optimal set* in an excessively frugal system whose sensors are stimulated by a a very broad band of speeds, and which should tune to the most likely speed — expected speed v_e - of the environment. The red line is a compromise (*weighed optimal set*) between the two extremes: a compromise between precision and frugality. The compromise is optimal when it is based on the statistics of stimulation.

to [4], the "measurement uncertainty" of a sensor is:

$$U(S,T) = \frac{\lambda_1}{S} + \frac{\lambda_2}{T} + \lambda_3 S + \lambda_4 T, \qquad (1)$$

where $\lambda_i > 0$ are positive constants. This uncertainty function is derived in [4] from the uncertainty principle of measurement [2].

The analysis in [4] showed that there exists an optimal condition for estimation of every speed. A set of such conditions across speeds ("optimal set") has a shape that depends on the properties and the number of sensors. In a realistic motion-sensitive system, the optimal set must lie between the optimal sets of two theoretical extremes:

- a system with unrealistically large amount of resources, which affords infinitely high precision of speed estimation, and which can afford what we call below *local optimization*.
- a system with unrealistically scarce resources, which affords only very low precision for estimating most speeds, and whose optimization is described below as *total integration*.

The theory in [4] finds an optimal compromise between the two extremes (Fig. 1).

A. Local optimization

A system that has infinite resources could dedicate a sensor for every speed and tune the sensor to that speed with infinite precision. In such a system the optimal condition for estimating a speed is a point on the corresponding speed line where the measurement uncertainty of Eq. 1 is minimal. We represent the optimal set for such as system by the black curve in Fig. 1.

This model is called "local" because it is derived for a system that can afford optimization for every speed independent of other speeds. The local optimal set is an idealization, to which any system seeking precise speed estimation ought to aspire, but which most systems (including biological vision) cannot reach for most speeds, for reasons we outline next.

B. Integral optimization

In real motion-sensitive systems the number of sensors and computing elements is limited, which implies that sensitivity of every sensor must span an interval of speeds. In other words, speed-selective sensors in real systems make their measurements by integrating contributions of multiple speeds (which is why this framework is called integral).

1) Optimization by total integration: At the extreme, consider sensors whose sensitivity spans all the measurable speeds. The optimal strategy for this system is to tune its sensors (perhaps a single sensor, at the extreme) to the most likely speed - mathematical expectation of speed v_e:

$$v_e = \int_0^\infty p(v)\, v\, dv, \qquad (2)$$

where $p(v)$ is the distribution of speeds in the stimulation. Whereas the optimal set by local optimization can have different shapes, depending on the parameters λ_i of the uncertainty function in Eq. 1, the shape of the integral optimal set is universally a hyperbola [4]. In Fig. 1, the optimal set found by integrating contributions of *all the speeds* is plotted as the thick grey curve.

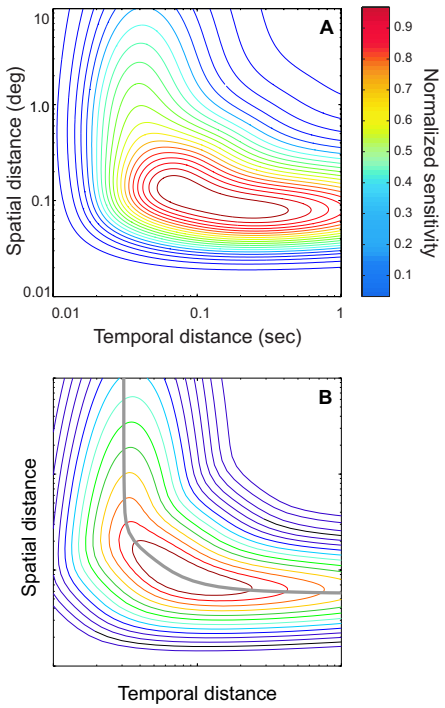

Fig. 2: *Equivalent conditions for motion estimation.* A. Human isosensitivity contours [5]. B. Theoretical equivalence sets for optimal motion estimation derived in [4] reproduce the "bent loaf of bread" shape of human isosensitivity contours in A.

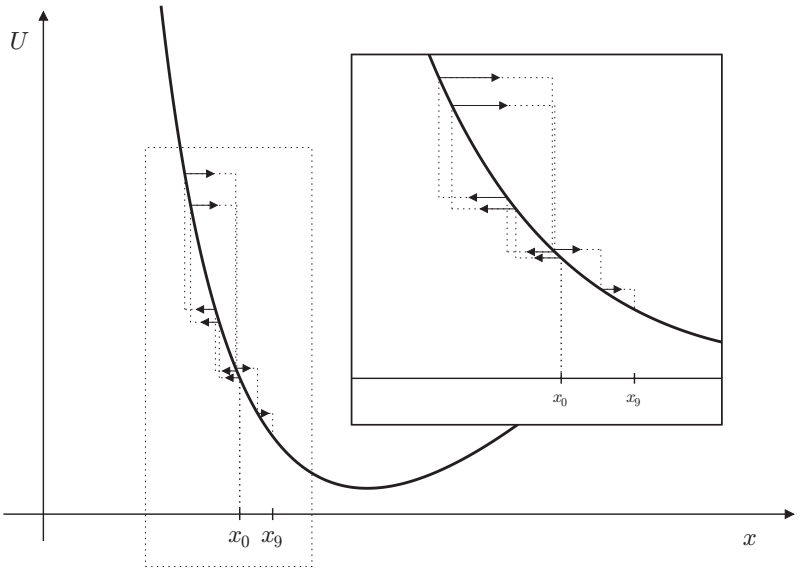

Fig. 3: *Illustration of the basic mechanism of adaptive optimization in one dimension.* The curve is uncertainty function $U(x)$ — where x represents a spatial or temporal parameter — is similar to the one defined in Eq. 5. The shape of $U(x)$ is derived in [4] from the uncertainty principle of measurement [2]: The high values of $U(x)$ at low x represent the high error rates in localizing stimuli in spatial or temporal frequency domain, and the high values of $U(x)$ at high x represent the high error rates in localizing stimuli in space or time.

Each horizontal arrow represents a change of tuning of an individual element of the network for one iteration of simulation. The arrows are scaled linearly with the uncertainty. The inset is an enlargement of the area marked by the dotted rectangle in the main panel. The figure shows eight iterations: From the initial value, x_0, the parameter changes the same number of times in the two directions on x, which brings the cell closer to the minimum of $U(x)$.

2) Weighted optimization: The conditions of realistic speed estimation must lie between the local optimal set and the optimal set by total integration. As proposed by [4], the compromise between the extremes for every speeds should depend on how important the speed is for the system. Thus, the degree to which the realistic optimal set approaches the ideal optimal set (i.e., the optimal set by local optimization) depends on the propensity of that speed in the stimulation: The more likely a speed is in the environment the closer its optimal point approaches the ideal point predicted by the local model. Because of the aforementioned integration across speeds by realistic sensors, the degree to which a realistic optimal point approaches the ideal point also depends on how narrowly the receptive fields are tuned to that speed.

The latter model is called "weighted optimization" because it was derived in [4] by weighting the contributions of the local and the total-integration models into the optimization process for every speed. The resulting optimal set is shown in Fig. 1 as a red curve. The curve approaches the optimal set by local optimization more at low than high speeds because low speeds prevail in the natural stimulation.

C. Optimal distribution of sensors

In Fig. 2B we plot predictions of [4] for the optimal allocation of sensors with different tuning properties across the entire space of parameters. The predicted distribution is similar to the human distribution of sensitivity as measured by [5] (Fig. 2A): it has the characteristic "bent loaf of bread" shape with the maximal sensitivity set having a hyperbolic shape. As the normative theory [4] suggests, this shape is close to hyperbolic because motion sensors integrate stimulation across speeds (Section 2-B and Fig. 1) and because the visual system seeks to optimize the precision of speed estimation.

The distribution of sensitivity shown in Fig. 2B is an emergent invariant of an optimal motion-sensitive system. Each element in the system responds to stimuli locally. In the following we show that a sensory system made of elements that estimate only local stimulation can evolve toward an optimal state similar to the one shown in Fig. 2B.

3. Adaptive optimization steered by measurement uncertainty

In this section we describe general properties of the algorithm of adaptive optimization.

A. Rationale

Consider a network of sensors each of which can localize stimulus $\mathcal{M}(S,T)$ in space-time and in the spatiotemporal frequency domain, and also estimate the errors associated with the estimates.

Localization. Localizing $\mathcal{M}$ in space-time amounts to measuring the mean values of stimulation by every sensor:

$$\begin{aligned} \bar{S}(T,\mathcal{M}) &= \frac{1}{S_0}\int_0^{S_0} \mathcal{M}(S,T)dS, \\ \bar{T}(S,\mathcal{M}) &= \frac{1}{T_0}\int_0^{T_0} \mathcal{M}(S,T)dT. \end{aligned} \tag{3}$$

where S_0 and T_0 are the spatial and temporal extents of the sensor's receptive field and $\bar{S}(T,\mathcal{M})$ and $\bar{T}(S,\mathcal{M})$ are the estimates of spatial and temporal parameters of stimulation, respectively. In general, the values of $\bar{S}(T,\mathcal{M})$ and $\bar{T}(S,\mathcal{M})$ are functions of S and T. Here, we assume quasi-stationarity of stimulation, i.e., we assume that the spatial and temporal properties of stimulation do not change significantly during the measurement.

Uncertainty estimation. We assume that the system can also estimate variances $\bar{S}^1(T)$ and $\bar{T}^1(S)$ associated with estima-

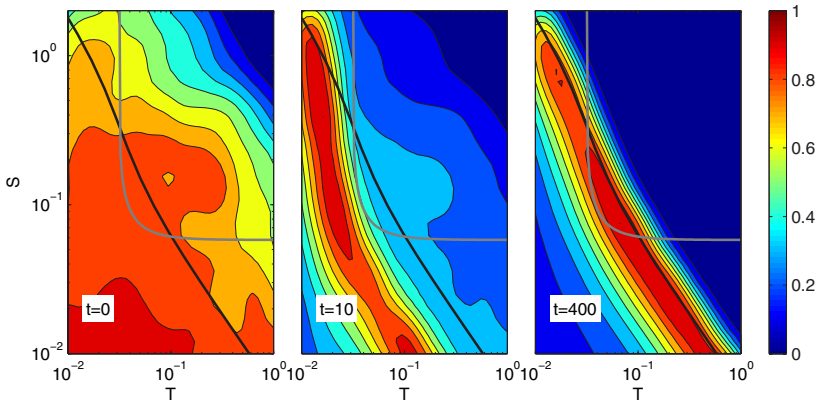

Fig. 4: Illustration of evolution by gradient of implicit uncertainties in the local optimization framework. The three panels represent the states of the network before the simulation begun (t=0) and after iterations 10 and 400. The color represents the "sensitivity" (Eq. 8): the warmer the color the higher the sensitivity. (We normalized sensitivity within every map.) As the network evolved, the region of highest sensitivity was gradually aligned with the *analytic* prediction of local optimization, represented by the black curve.

In this illustration, as in the other illustrations (Figs. 5–7), the simulations started from the sensors distributed across 60 speed lines according to a uniform random distribution, with 60 sensors allocated to uniformly random positions on every speed line (3600 sensors in total).

tion of the mean values of stimulation:

$$\begin{aligned}\bar{S}^1(T) &= \frac{1}{S_0}\int_0^{S_0}(\mathcal{M}(S,T)-\bar{S})^2 dS,\\ \bar{T}^1(S) &= \frac{1}{T_0}\int_0^{T_0}(\mathcal{M}(S,T)-\bar{T})^2 dT.\end{aligned} \tag{4}$$

For the quasi-stationary stimulation, the variance functions $\bar{S}^1(T,\mathcal{M})$ and $\bar{T}^1(S,\mathcal{M})$ are independent on S and T.

Frequency domain. Similarly, the cells are assumed to estimate mean spatial and temporal frequencies in the stimulation $\bar{f}_S(\mathcal{M})$, $\bar{f}_T(\mathcal{M})$ and the corresponding variances $\bar{f}^1_S(\mathcal{M})$, $\bar{f}^1_T(\mathcal{M})$.

Thus, the estimates of composite uncertainty u_i of the i-th sensor is assumed to depend on the values of $\bar{S}^1(\mathcal{M})$, $\bar{T}^1(\mathcal{M})$, $\bar{f}^1_S(\mathcal{M})$, $\bar{f}^1_T(\mathcal{M})$. According to [2], in a system that uses optimal sensors, measurements errors by i-th sensor are defined as

$$\begin{aligned}u_i(\mathcal{M}) =& \lambda_1\bar{f}^1_S(\mathcal{M}) + \lambda_2\bar{f}^1_T(\mathcal{M})\\ &+ \lambda_3\bar{S}^1(\mathcal{M}) + \lambda_4\bar{T}^1(\mathcal{M})\end{aligned} \tag{5}$$

which is similar to the explicit uncertainty function in Eq. 1. Note that the explicit uncertainty in Eq. 1 is derived for the complete set of stimuli (spanning the entire space of parameters), whereas Eq. 5 applies to a single stimulus $\mathcal{M}$ at a time. Hence, Eq. 5 can serve as an *implicit estimate* of measurement uncertainty for current stimulation $\mathcal{M}$ in evolutionary simulations.

Next we describe the algorithms we used to explore how a network of sensors, each estimating the implicit uncertainty associated with its measurements (Eq. 5), can optimize itself in face of on-going stimulation.

B. Implementation

Initialization. Each sensor was assigned an initial parameter value, i.e., a location (S, T) in the space-time Distance Plot. We investigated different deployments of neurons across the space of parameters and found that, provided sufficient evolution time, the initial deployment did not affect the results of evolution.

Iterative evolution. We implemented two approaches to evolution of sensitivity in our networks. In both approaches, cells were allowed to change their locations on the corresponding speed lines, based on the estimates of measurements uncertainty at cell current locations. Measurement uncertainties were computed as in Eq. 5.

1) Gradient approach: This approach was an application of the standard gradient descent algorithm. Parameters of every cell were adjusted according to the local gradient of measurement uncertainty. We assumed that every i-th cell had a preferred velocity v_i: the ratio of cell's preferred temporal and spatial parameters, S_i and T_i. Although the preferred T_i and S_i were allowed to vary from one measurement to another, we assumed that cell tuning to speed did not change significantly over a large number of measurements. In other words, cell tuning was allowed to drift along its preferred speed line. In the framework of optimization by gradient descent, these assumptions imply that the values of T_i and S_i were updated according to the following rule:

$$\begin{aligned}\Delta T_i &= -\gamma\cos\left((T_i, v_iT_i),(\partial u_{T(i)},\partial u_{S(i)})\right)\\ \Delta S_i &= -v_i\Delta T_i,\ \gamma>0,\end{aligned} \tag{6}$$

where γ is a positive constant, and $\partial u_{T(i)} = \partial u_i(S_i,T_i)/\partial T_i$ and $\partial u_{S(i)} = \partial u_i(S_i,T_i)/\partial S_i$ are the spatial and temporal partial derivatives of measurement uncertainty at the tuning parameters of the cell. Term $\cos(\cdot)$ in Eq. 6 represents the difference between two orientations: orientation of the local gradient of uncertainty

$$(\partial u_{T(i)},\partial u_{S(i)})$$

and orientation of the speed line

$$(T_i, v_iT_i)$$

for the speed that dominates optimization of i-th cell. Parameter γ controls susceptibility of cell tuning to change and thus it determines how quickly the network settles into a stable state for given conditions of stimulation.

2) Random search approach: We illustrate this approach schematically in Fig. 3, on a single dimension. In the actual simulations at every step and for every cell, a direction on the speed line was chosen randomly from a uniform distribution.

The step for each cell was then scaled by the amount of uncertainty measured at that location:

$$\begin{aligned} \Delta T_i &= \gamma R u_i(S_i, T_i) \\ \Delta S_i &= v_i \Delta T_i, \end{aligned} \tag{7}$$

where $R \in \{-1, 1\}$ represents the (random) direction on the speed line and γ is defined as in Eq. 6.

We assumed that the tuning functions of individual cells were bivariate Gaussian functions, whose means were cell's preferred parameters (T_i, S_i), and whose temporal and spatial sizes (dispersions) were, respectively, $C_T T_i$ and $C_S S_i$. Positive constants C_T and C_S were chosen so that the tuning functions covered the entire space of parameters. We calculated an additive performance characteristic $\mathcal{P}$ of the system as a whole as follows:

$$\mathcal{P}(S, T) = \sum_{i=1}^{N} e^{-\frac{(S-S_i)^2}{(C_S S_i)^2} - \frac{(T-T_i)^2}{(C_T T_i)^2}}. \tag{8}$$

We called this characteristic "sensitivity" and plotted it across the space of parameters (T, S) as a "map of sensitivity," to capture the dynamics of evolution in our networks.

4. Results

Simulations by different optimization models yielded results consistent with the analytic predictions in [4]. This result demonstrated that local estimates of measurement uncertainty are sufficient to steer sensory system toward globally optimal states. We illustrate results of the computational experiments in Figs. 4–7.

In Fig. 4 we plot results of evolution by a method that approximated the local optimization model. Here we assumed that each cell estimated uncertainties associated only with measurements of a single speed: the speed preferred by the cell (tuning speed). As we show in Fig. 4, after a sufficient number of iterations, the region of highest sensitivity grew aligned with the analytic prediction of local optimization: the local optimal set, which is represented in Fig. 4 by a black curve. This result did not depend on statistics of stimulation, as long as the statistics allowed all the cells in the system to be stimulated, to insure optimization for all the represented speeds.

In Fig. 5 we plot results of evolution by a method that approximated the total-integration model. Here sensors tuning was very broad and the optimization process was governed by a single speed: expected speed v_e of stimulation (Eq. 2). In agreement with the analytic prediction, now the region of highest sensitivity evolved to resemble a hyperbola in the Distance Plot. Contrary to the simulations of local optimization, results of present simulations did depend on statistics of stimulation. It is because now every cell was stimulated by every stimulus, so that the average estimated speed by every cell was the expected speed v_e.

In Fig. 6 we illustrate results of evolution by a method that approximated the weighted model, which is a case of integral optimization. Here, the receptive field of every cell was stimulated by a limited range of speeds, as opposed to the entire range of speeds in the total-integration model. Sizes of the receptive fields across speeds were set to grow linearly as a function of speed (in agreement with the principle of uncertainty [4]), i.e., the range of integration across speeds was wider for cells tuned to high speeds than for cells tuned to low speeds. As a result, optimization of i-th cell was governed by the mean speed over its range of integration, $\Omega_v(i)$:

$$v_{e(i)} = \int_{\Omega_v(i)} p(v)\, v\, dv.$$

Now measurement uncertainty steered cells such that the region of highest sensitivity occupied an intermediate position between the optimal set by the local-optimization model and the optimal set by the total-integration model.

In Fig. 7 we illustrate an approximation of local optimization using the method of random search (explained for one dimension in Fig. 3). As in the gradient descent simulation, this optimization steered the cells toward the predicted optimal set. Notice, however, that now the region of highest sensitivity was not concentrated around the optimal set as tightly as in the gradient descent simulation. The reason for this difference is that in the gradient descent simulations the cells stopped changing their tuning parameters when the conditions of optimization were met; for example, when they reached the minimum of the measurement uncertainty function in the local optimization framework. In contrast, tuning parameters of cells in the random-search simulations never ceased to change: the search algorithm brought the cells toward the optimal conditions, but it forced the cells to wander in the vicinity of these conditions, which is why the distribution of cells in the optimized network in Fig. 7 is wider than in Figs. 4–6.

References

[1] A. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, Series A*, 2(2):284–299, 1985.

[2] D. Gabor. Theory of communication. *Institution of Electrical Engineers*, 93 (Part III):429–457, 1946.

[3] S. Gepshtein and M. Kubovy. The lawful perception of apparent motion. *Journal of Vision*, 7(9):1–15, 2007. http://journalofvision.org/7/8/9/

[4] S. Gepshtein, I. Tyukin, and M. Kubovy. The economics of motion perception and invariants of visual sensitivity. *Journal of Vision*, 7(8):1–18. http://journalofvision.org/7/8/8/.

[5] D. Kelly. Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the optical society of America*, 69:1340–1349, 1979.

[6] D. H. Kelly. Eye movements and contrast sensitivity. In D. H. Kelly, editor, *Visual Science and Engineering. (Models and Applications)*, pages 93–114. Marcel Dekker, Inc., New York, USA, 1994.

[7] W. Reichardt. Autocorrelation, a principle of evaluation of sensory information by the central nervous system. In W. A. Rosenbluth, editor, *Sensory Communication*, pages 303–318. MIT Press, Cambridge, MA, USA, 1961.

[8] J. P. H. van Santen and G. Sperling. Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2(2):300–321, 1985.

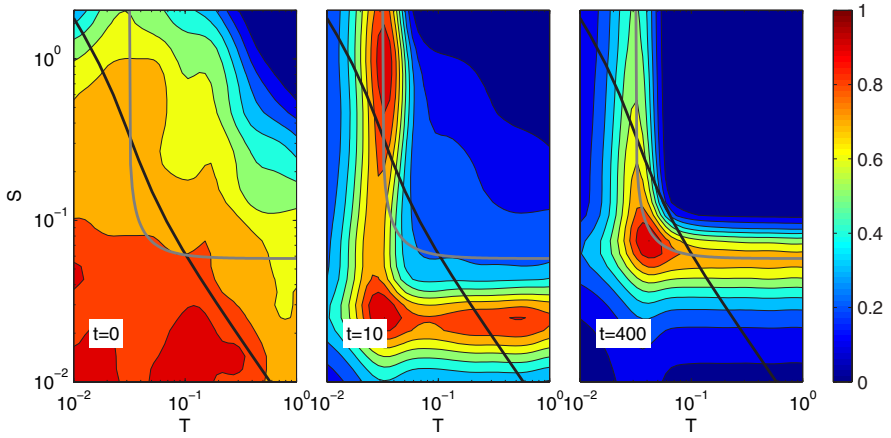

Fig. 5: Illustration of evolution by gradient of implicit uncertainties in optimization by total integration. The graphic conventions are as in Fig. 4. The grey hyperbola represents the analytic optimal set by the total-integration model. We simulated optimization by total integration by making optimization of every cell depend on the mean speed of stimulation v_e (Eq. 2). As the network evolved, the region of highest sensitivity was gradually aligned with the analytic prediction of optimization by total integration.

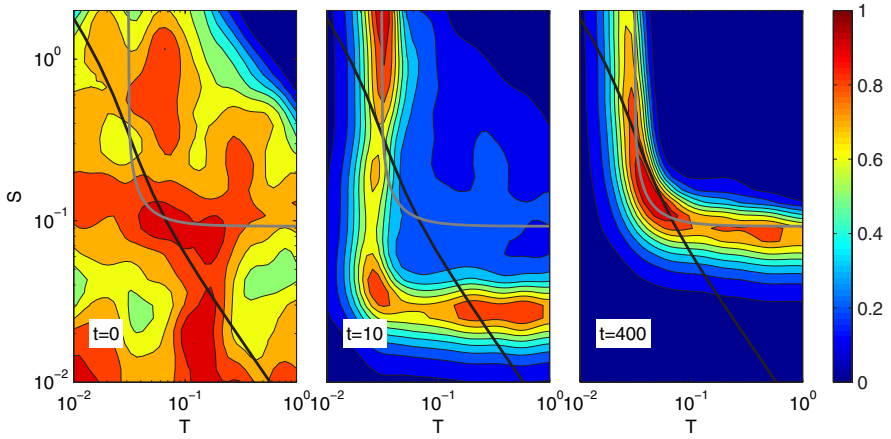

Fig. 6: Illustration of evolution by gradient of implicit uncertainties in the weighted optimization framework. Now every sensor integrated contributions of multiple speeds, such that optimization depended on the mean speed of stimulation that fell within the region of integration. On every iteration, a sensor was presented with 100 stimuli on the same speed line. The speeds were drawn across the iterations from a normal distribution with mean $\mu = 1.0$ °/sec and standard deviation $\sigma = 4.47$ °/sec. Only those cells were updated whose receptive fields were stimulated on a given iteration. As is evident in the figure, the region of highest sensitivity was gradually aligned with a set that was intermediate between the one predicted by local optimization and the one predicted by total integration, as expected from the normative considerations (Fig. 1).

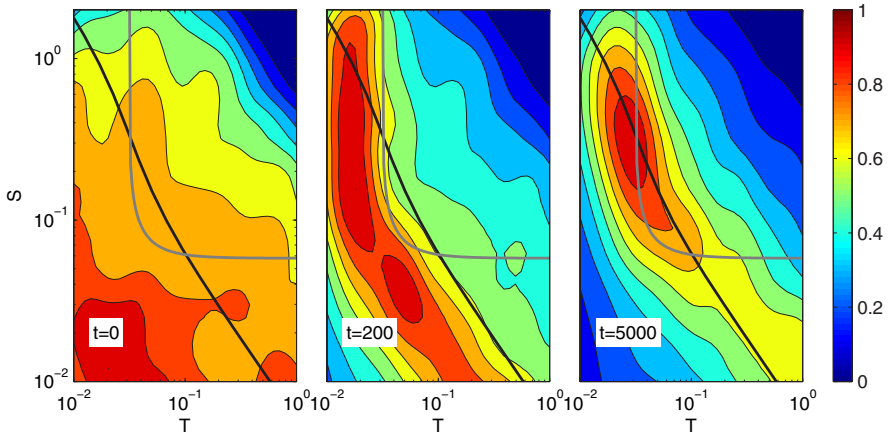

Fig. 7: Illustration of evolution by the random-search algorithm driven by implicit uncertainties in the local optimization framework. This evolution proceeds slower than evolution by gradient descent, which is why we show the states of the system separated by a greater number of iterations than in the previous figures. In the optimized network, high sensitivity tends to concentrate along the local optimal set, as in Fig 4, but the random search algorithm forces the cells to continuously change their tuning parameters, which is why the distribution of sensitivity in a network optimized by random search is more dispersed than in a network optimized by gradient descent.