Flexible Cue Integration by Line Attraction Dynamics and Divisive Normalization

Conference Paper · September 2014

DOI: 10.1007/978-3-319-11179-7_87

3 authors:


Mohsen Firouzi
Ludwig-Maximilians-University of Munich


Stefan Glasauer
Brandenburg University of Technology Cottbus - Senftenberg


Jorg Conradt
KTH Royal Institute of Technology

Flexible Cue Integration by Line Attraction Dynamics and Divisive Normalization

Mohsen Firouzi[1,2,3], Stefan Glasauer[2,3,4], and Jörg Conradt[1,2,3]

[1] Neuroscientific System Theory, Technische Universität München, München, Germany
[2] Bernstein Center for Computational Neuroscience, München, Germany
[3] Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, München, Germany
[4] Center for Sensorimotor Research, Ludwig-Maximilians-Universität München, Germany
{mohsen.firouzi,conradt}@tum.de, sglasauer@lmu.de

Abstract. One of the key computations performed in human brain is multi-sensory cue integration, through which humans are able to estimate the current state of the world to discover relative reliabilities and relations between observed cues. Mammalian cortex consists of highly distributed and interconnected populations of neurons, each providing a specific type of information about the state of the world. Connections between areas seemingly realize functional relationships amongst them and computation occurs by each area trying to be consistent with the areas it is connected to. In this paper using line-attraction dynamics and divisive normalization, we present a computational framework which is able to learn arbitrary non-linear relations between multiple cues using a simple Hebbian Learning principle. After learning, the network dynamics converges to the stable state so to satisfy the relation between connected populations. This network can perform several principle computational tasks such as inference, de-noising and cue-integration. By applying a real world multi-sensory integrating scenario, we demonstrate that the network can encode relative reliabilities of cues in different areas of the state space, over distributed population vectors. This reliability based encoding biases the network's dynamics in favor of more reliable cues and realizes a near optimal sensory integration mechanism. Additional important features of the network are its scalability to cases with higher order of modalities and its flexibility to learn smooth functions of relations which is necessary for a system to operate in a dynamic environment.

Keywords: Multi-sensory Cue Integration, Line Attraction Dynamics, Divisive Normalization, Associative Hebbian Learning, Heading estimation.

1 Introduction

A key requirement for any system, including biological or man-made systems is their capability to estimate physical properties of the real world through partially reliable observations to interact properly with their environment. For instance, to reach an object by hand, one must configure the arm joints with respect to the visual location

S. Wermter et al. (Eds.): ICANN 2014, LNCS 8681, pp. 691–698, 2014.

of the object and proprioceptive cues [1]. Apart from intrinsic variability of neural activity in the brain, accessible sensory cues are often uncertain and ambiguous. The human brain can combine these noisy and partially reliable pieces of information to optimally estimate the state of the world and consequently handle cognitive tasks efficiently [2].

Despite decades of research the underlying cortical processing that enables us to optimally operate in ambiguous environments is not well understood yet; what the processing consists of, or even how the processed data is represented [3]. Some computational frameworks using probabilistic population code with hand-crafted connectivity have shown how de-noising, inference (estimation) and sensor perception can possibly be performed by cortical and sub-cortical circuits [1][4]. Recently an unsupervised framework of relation learning between two interacting populations of neurons has been proposed, which allows the network to learn arbitrary relations between two encoded variables [5]. However, a flexible computational framework which could learn relationships between cues rather than using fixed networks is still addressed as a challenge, especially in the presence of higher order modalities [5].

Another issue in multi-sensory integration which is less investigated is how to encode and learn reliability of cues into spatially registered form of neural activity. In fact sensory cues do not have equal distribution of reliability over sensory space. For instance, the location of visual stimuli near fovea is more reliable and identifiable than periphery ones [6].

In this work, we suggest a recurrent attractor network capable of learning arbitrary relations between one of the encoded sensory variables as a function of other variables using biologically realistic algorithms like Hebbian Learning and Divisive Normalization. In another point of view the attraction surface of the network's dynamics is the same surface (hyper-surface) of the relation function through which the network realizes a relation satisfaction mechanism. We demonstrate that after constructing plastic weights, the network is able to perform inference and reasoning, de-noising, reliability based cue-integration and decision making. This framework is well scalable for scenarios with higher order of modalities and with acceptable flexibility to wide range of smooth functions. Another important feature of the network is the possibility of spatially distributed reliability representation in form of neural encoding. In fact we can strengthen encoded activity of the stimuli according to their relative reliability so that network converges to the point on the relation surface which is closer to initial point of more reliable cue. In better word network dynamics would change more reliable cue slower than the others.

In next section we elaborate the general architecture, encoding, dynamics and learning in the network. In Section 3 some computational abilities of the network e.g. estimation, de-nosing, cue integration and decision making are shown for a linear and a non-linear relation function. In section 4 we demonstrate a practical heading estimation robotic application using a distributed dual-modal version of the proposed network. And finally section 5 summarizes and concludes the paper.

2 Attractor Network Model

2.1 General Architecture and Input Encoding

General architecture of the attractor network for a tri-modal cue integration scenario is shown in Fig.1-left. The network consists of three encoded populations (R^n) and an intermediate layer (A_{lm}). As is shown in Fig.1-right, cues are encoded by activity of the spatially distributed population of neurons with overlapping wrap-around Gaussian tuning curves. Since intrinsic neural activity in brain is governed by Poisson variability, the initial activity or equivalently selectivity of a single neuron r_i (*number of spikes per second*), is drawn from a Poisson distribution with mean firing rate of neuron tuning curves, $\Phi(\kappa, x)$; see equations below where κ and σ are constant showing activity strength and width of neurons tuning curve respectively, $x^c{}_i$ is preferred value of i^{th} neuron, ν is spontaneous activity which is set to 0.1, and finally x is input stimulus.

$$P(r_i|x) = \frac{[\Phi_i(\kappa,x)]^{r_i}}{(r_i)!} e^{-\Phi_i(\kappa,x)} \tag{1}$$

$$\Phi_i(\kappa, x) = \kappa e^{-\frac{|x-x_i^c|}{2\sigma^2}} + \nu \tag{2}$$

All neurons are linear threshold neurons and input neurons are reciprocally connected to intermediate layer A_{lm} ($W^n{}_{RA} = W^n{}_{AR}$). To keep input stimuli into topographically arranged spatial registers and to copy the cues into a common frame of reference, R^1 and R^2 populations (population vectors of x_1, x_2) are projected to the intermediate layer using a fixed *von-Mises* weighting distribution as following equation [4]:

$$W^1_{ilm} = e^{\frac{(\mathrm{i}-\mathrm{l})\left(\cos[\frac{2\pi}{N}]\right)-1}{(\sigma_1)^2}}, \; W^2_{jlm} = e^{\frac{(\mathrm{j}-\mathrm{m})\left(\cos[\frac{2\pi}{N}]\right)-1}{(\sigma_2)^2}} \tag{3}$$

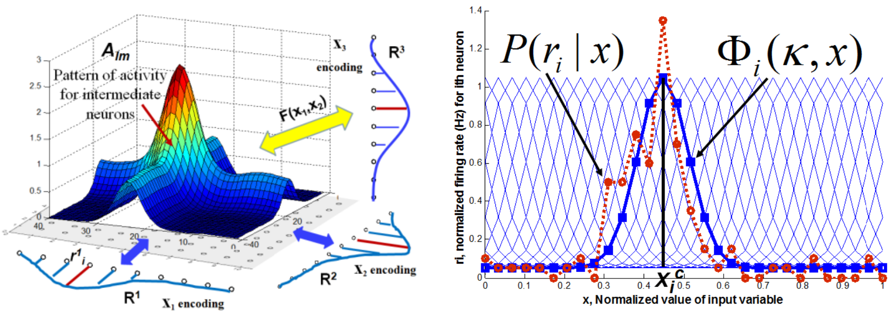


Fig. 1. Left: Network connectivity for three variables encoded by probabilistic population code, R^1 and R^2 are projected to intermediate neurons by *Von-Mises* weight pattern. The connection of third variable "x_3" is plastic so as to realize relation function $F(x_1,x_2)$.
Right: red diagram shows selectivity or the activity of the i^{th} neuron (r_i) in response to normalized stimuli x, governed by *Poisson* variability; Blue: i^{th} neuron tuning curve or equally expected activity (Φ_i), centered at x_i^c as preferred value.

Where W^n_{ilm} is the synaptic weight between i^{th} neuron of n^{th} input population (r_i^n) and lm^{th} intermediate neuron (a_{lm}), N is the number of neurons in each population and σ_n tunes width of projection. Synaptic connectivity between R^3 neurons and intermediate layer, W^3_{klm} (*yellow arrow in Fig.1-left*) is modifiable so as to construct the relation F by means of associative Hebbian Learning. In order to perform integration over more than three spatial cues, intermediate layer can be simply organized as a cubic or hyper-cubic topographically arranged population of neurons. Furthermore the way of encoding and line-attraction dynamics of the network, enable us to initialize input cues, based on their relative reliabilities.

2.2 Network Dynamics

Through dynamics of the network, population activities or equivalently encoded cues would be shifted so to satisfy relation function. In other word during the network's dynamics, input cues follow a trajectory to be converged toward surface of attraction in steady-state. In each time step the activity of single intermediate neuron is weighted sum of momentary activity of connected input neuron which is normalized by Divisive Normalization to keep single bumps of activities and eliminate the effect of ridge-like pattern of activities (see Fig.1-left). Equations (4)-(5) represent the dynamics of intermediate neurons:

$$A_{lm}(t+1) = \frac{(d_{lm}(t))^{\alpha}}{\beta + s\sum_p \sum_q (d_{pq}(t))^{\alpha}} \tag{4}$$

$$d_{lm}(t) = \sum_{k=1}^{N} W^1_{klm} r^1_k(t) + \sum_{k=1}^{N} W^2_{klm} r^2_k(t) + \sum_{k=1}^{N} W^3_{klm} r^3_k(t) \tag{5}$$

Where α is divisive power which tunes the sharpness of normalization, β is a constant bias to prevent division by zero and W^n_{klm} synaptic weight between k^{th} input neuron of n^{th} input population and $l_{lm}{}^{th}$ intermediate neuron. After updating the activity of intermediate layer, activity of input populations should be updated by feedback connections and DN similar to intermediate neurons. See equation (6):

$$r_i^{n\{=1,2,3\}}(t+1) = \frac{[\sum_l \sum_m W^n_{ilm} A_{lm}(t+1)]^{\alpha}}{\beta + s\sum_{k=1}^{N} [\sum_l \sum_m W^n_{klm} A_{lm}(t+1)]^{\alpha}} \tag{6}$$

It is worth to notice that for non-invertible functions, DN is not enough to elicit bumps of activity in intermediate layer, so in addition to DN an additive inhibition using a global inhibition neuron has been used to inhibit irrelevant pattern of activities in intermediate layer.

2.3 Relation Learning

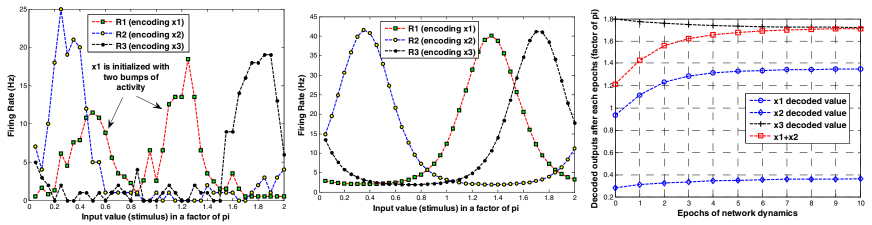
As is mentioned in previous section, to construct an arbitrary relation function $F(x_1,x_2)$ between input cues, synaptic connection of third input population with intermediate layer, W^3_{klm} can be modified by a simple associative Hebbian learning. In learning phase, after projection of R^1 and R^2 into intermediate layer followed by DN and additive inhibition, a single bump of activity would emerge, and then plastic connections would be modified as following equation (δ is learning rate):

$$W^3_{klm}(t+1) = W^3_{klm}(t) + \delta\, r^3_k\, A_{lm} \tag{7}$$

In each learning epoch, synaptic weights are normalized to maintain relative strength of connections and regulate overall synaptic drive received by a single neuron similar to Synaptic Scaling in biological neurons [7].

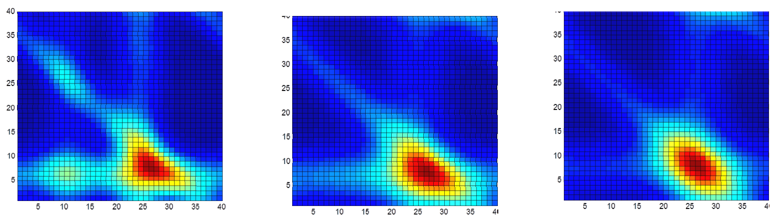
3 De-noising, Inference and Cue-Integration

In this section we will validate attractor network in some computational principles. The network is first trained to learn a simple linear relation function: $x_3 = x_2 + x_1$. After learning, network is initialized by noisy patterns of activity as is depicted in Fig.2a. Also R^1 has been initialized by two peaks of activity or equivalently two different stimuli located in different position in uni-sensory state space; one which is totally inconsistent with other cues according to relation and another is more consistent with other cues but not perfectly satisfies the relation. In the equilibrium state of the network's dynamics (after 10 epochs), activity of intermediate neurons will converge to a single bump of activity (Fig.3c). This bump would generate final stabilized population vectors (Fig.2b). As is shown in Fig.2b the network is able to remove internal noise perfectly. More interestingly the stimulus which is not consistent with the other stimuli has been totally removed, and the more consistent stimulus (more spatially correlated) has been strengthened (R^1 or square-red dash curve in Fig.2a & b). The hills of activities (or equally encoded variables) are moving towards being in equilibrium point where three encoded variables perfectly satisfy the relation (Fig.2c). In this network N is set to 40, $\beta = 0.1$, $s = 0.001$, $\alpha = 2$ and $\sigma = 0.45$.



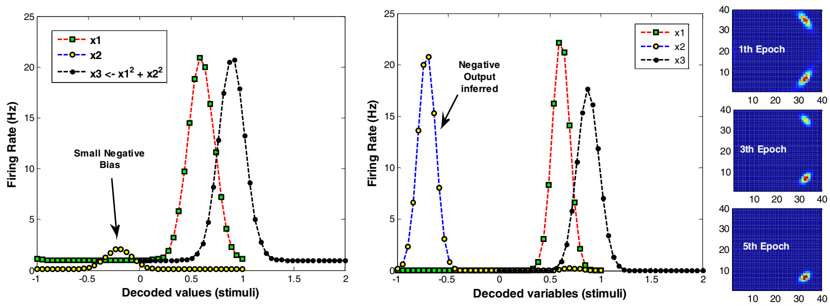
(a) (b) (c)

Fig. 2. (a) Initial population, (b) Population vectors after 10 epochs, (c) Decoded values in each epoch



(a) (b) (c)

Fig. 3. Momentary transient activity of intermediate neurons emerged as a single bump of activity in stable state of network dynamic, (a) epoch=1, (b) epoch=5, c) epoch=10



(a) (b) (c)

Fig. 4. (a) Initial populations, (b) Final populations, (c) Intermediate activity in 5-epochs

By initializing one of the population vectors with zero (shutting all neurons), the network can infer and retrieve the value for unknown variable that is consistent with the other initialized variables (consistency in terms of relation). Another important feature of the network is demonstrated in Fig.2c; the less reliable cue (x_1) tends to move faster (steeper trajectory) compared to the other cues. Similarly if one of the modalities is encoded by a smaller peak of activity (smaller κ in (2)) compared with the others, the attractor dynamics weights that cue as less confident cue and it would be changed faster toward being coherent with other cues with respect to relation (weighted cue integration). In section 4 by showing a realistic scenario, we will show if we perform weighted encoding or equivalently weighted projection to intermediate layer, according to relative reliability of cues (e.g. reverse of Gaussian noise power in each sensory modality), the network can simply follow a near optimal cue integration.

3.1 Decision Making in Non-invertible Relations

In case of symmetrical or non-invertible relations like parabola function ($x_3 = x_1^2 + x_2^2$), to infer one of the x_1 or x_2 variables, it is probable to emerge two possible peaks of activity as inferred value. One solution is evaluating network dynamics and updating neuron activities using an asynchronous dynamics [8]. Another simple solution is violating the symmetry in support of one possible stimulus for unknown variable. For instance if the network is initialized with a tiny negative bias (Fig.4 a) for the unknown cue, this negative bias helps the network to retrieve the negative peak for hidden variable (Fig.4 b). Consequently the bump corresponding to the positive value in the intermediate layer has been removed during network dynamics (Fig.4 c). In this network N is set to 40, β = 0.1, s = 0.002, σ = 0.38, and finally α = 3 to achieve a sharper DN inhibition for irrelevant patterns of activity.

4 Cue Weighting, Heading Estimation in a Mobile Robot

As a practical case study for multi-sensory cue integration, we have evaluated a distributed architecture of dual-modal attractor networks for head estimation in an Omni-direction mobile robot [9]. The robot is equipped with an IMU unit including on board Gyroscope and Compass sensor. The robot exploring the space through a closed trajectory and an efferent copy of motor command driving wheels (*odometry*) is provided to estimate angle of heading [9]. Consequently we have three sensory readings; each is supposed to estimate the angle of heading of the robot with respect to room coordinates. We have assumed that external noise has Gaussian distribution, so simply using EM algorithm variance of noise process for a single sensor can recursively be estimated and updated by exploring around the space (from 0° to 360°). Since we want to evaluate how possibly optimal, Line Attractor Network can operate in noisy environment, we have compared the network's outcome with Maximum Likelihood Estimator as a statistically optimal estimator [2]. Let's assume sensory measurements are statistically independent, so MLE optimally combines uni-sensory estimates $\{x_k\}$ by a simple weighted average computation in which weights are reversely related to noise power (variance). See equation bellow (σ_k^2 is noise power of k^{th} sensor):

$$x_{MLE} = \frac{1}{\sum \frac{1}{\sigma_k^2}} \sum \frac{1}{\sigma_k^2} x_k \tag{8}$$

We have evaluated two way of cue weighting in LAN network. First way which does not need any information about noise process is voting-based method [9] [10]. Simplified underlying idea of this method is that the most reliable cue is the one which is closest to Center of Gravity of all sensory estimates. In better word the best sensor is the one which is more coherent with the others. The second method is weighting the initial peak of population activities (κ in (2)) with a normalized value similar to gain filed tuning in cortical circuits and in accordance with relative reliability [11]. The normalized weight is proportional to reverse of sensory variance over exploring space. In Fig.5-down this reliability map has been shown for 1780 sample points of the state space from 0° to 360°. It is clear that Compass sensor is much noisier and less reliable than Gyro and odometry. It is worth to mention that in this scenario a dual-modal version of the network with three input populations is used.

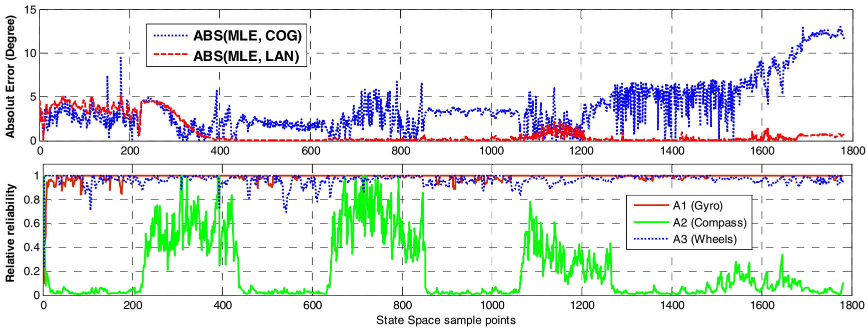


Fig. 5. Upper: Absolute error of MLE and COG voting integration algorithm, and LAN network. **Lower**: normalized relative reliability of cues calculated using recursive EM algorithm.

In Fig.5-up absolute error between MLE as an optimal estimator, and both methods are depicted and it is illustrated that the outcome of LAN network with normalized relative reliability map which is shown in Fig.5-down, is near optimal and close to MLE. Despite of simplicity of COG based weighted encoding, since it does not take into account the noise variability it is less noise robust.

5 Conclusion and Remarks

The idea of retrieving information from perturbed patterns using association networks is not new in machine learning. But the architecture of these networks is a promising and inspiring framework to understanding how cortical circuits can possibly represent, preserve and combine information to establish a coherent and robust representation of the world. On the other hand, seemingly distributed cortical areas implement functional relation between each other through mutual connectivity and correlated neural activity. In this work we have investigated how a simple recurrent attractor network can come up with relation learning amongst multiple sensory cues and how possibly to combine them in an optimal fashion. The network provides a computational framework for relation satisfaction using attraction dynamics and is able to represent cues reliabilities in a distributed form of neural activity.

Results exhibit the capability of the network to perform de-noising, cue integration and inference even for non-invertible and smooth nonlinear functions. A real world sensory integration scenario for heading estimation is investigated and it is observed that by proper encoding of the reliability, based on uni-sensory variability, the network is capable of performing weighted integration in near optimal fashion.

Acknowledgment. This work was supported by the German Federal Ministry of Education and Research, Grant 01GQ0440 (BCCN).

References

1. Pouget, A., Sejnowsky, T.J.: Spatial Transformation in the Parietal Cortex Using Basis Functions. Journal of Cognitive Neuroscience 9(2), 222–237 (1997)
2. Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. Trends on Cognitive Science 8, 162–168 (2004)
3. Simoncelli, E.P.: Optimal estimation in sensory systems. In: Gazzaniga, M. (ed.) The Cognitive Neurosciences, IV, ch. 36, pp. 525–535. MIT Press (2009)
4. Jazayeri, M., Movshon, A.: Optimal representation of sensory information by neural populations. Nature Neuroscience 9, 690–696 (2006)
5. Cook, M., Jug, F., Krautz, C., Steger, A.: Unsupervised Learning of Relations. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) ICANN 2010, Part I. LNCS, vol. 6352, pp. 164–173. Springer, Heidelberg (2010)
6. Weber, C., Triesch, J.: Implementations and implications of foveated vision. Recent Patents on Computer Science 2(1), 75–85 (2009)
7. Turrigiano, G.G., Leslie, K.R., Desai, N.S., Rutherford, L.C., Nelson, S.B.: Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature 391(6670), 892–896 (1998)
8. Rougier, N.P., Hutt, A.: Synchronous and asynchronous evaluation of dynamic neural fields. Journal of Difference Equations and Applications 17(8) (2011)
9. Axenie, C., Conradt, J.: Cortically Inspired Sensor Fusion Network for Mobile Robot Heading Estimation. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E.P., Appollini, B., Kasabov, N. (eds.) ICANN 2013. LNCS, vol. 8131, pp. 240–247. Springer, Heidelberg (2013)
10. Trische, J., Von der Malsburg, C.: Democratic Integration: Self-Organized Integration of Adaptive Cues. Neural Computation 13(9), 2049–2207 (2001)
11. Brostek, L., Büttner, U., Mustari, M.J., Glasauer, S.: Eye Velocity Gain Fields in MSTd during Optokinetic Stimulation. Cerebral Cortex (in press February 20, 2014)