

Reply to comment

Symphony of high-dimensional brain

Reply to comments on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain”

Alexander N. Gorban^{a,b,*}, Valeri A. Makarov^{b,c}, Ivan Y. Tyukin^{a,b,d}^a Department of Mathematics, University of Leicester, Leicester, LE1 7RH, UK^b Lobachevsky University, Nizhni Novgorod, Russia^c Instituto de Matemática Interdisciplinar, Faculty of Mathematics, Universidad Complutense de Madrid, 28040 Madrid, Spain^d Saint-Petersburg State Electrotechnical University, Saint-Petersburg, Russia

Received 30 May 2019; accepted 21 June 2019

Available online 25 June 2019

Communicated by L. Perlovsky

Keywords: Complexity; Big data; Non-iterative learning; Error correction; Measure concentration; Blessing of dimensionality; Linear discriminant; Memory

This reply is not a ‘comment on comments’ but rather a ‘selective echo’ that resonates with some ideas and questions raised by the commentators [1–9]. We would like to thank all of them. All the individual comments and ideas come together as a symphony of thought, and each individual part changes the way the multidimensional brain is understood. The comments include many deep thoughts and inspiring questions and deserve a close reading.

The flight from independence and uniformity. Kůrková [1] has emphasized that many classical results about measure concentration assume either uniform probability distribution (with respect to the Lebesgue measure in space or the rotationally invariant measures on spheres [10,11]), or independence (‘product measures’ [12]), or both. In classical machine learning theory, data are typically assumed to be i.i.d. samples. On the contrary, in real life data samples are neither i.i.d. nor uniformly distributed.

A similar comment about the difference between the theoretical random distributions and “extremely non-random” real distributions has been received from Geoffrey Hinton (personal communication to ANG). The obvious gap between the current theory and practice is a problem, indeed, and the machine learning theory should be revised to deal with non-independent and extremely non-uniformly distributed data. Kůrková with Sanguineti proposed their ap-

DOI of original article: <https://doi.org/10.1016/j.plrev.2018.09.005>.

DOIs of comments: <https://doi.org/10.1016/j.plrev.2019.02.008>, <https://doi.org/10.1016/j.plrev.2019.03.004>, <https://doi.org/10.1016/j.plrev.2019.03.015>, <https://doi.org/10.1016/j.plrev.2019.04.006>, <https://doi.org/10.1016/j.plrev.2019.04.008>, <https://doi.org/10.1016/j.plrev.2019.02.010>, <https://doi.org/10.1016/j.plrev.2019.03.014>, <https://doi.org/10.1016/j.plrev.2019.03.005>, <https://doi.org/10.1016/j.plrev.2019.02.014>.

* Corresponding author at: Department of Mathematics, University of Leicester, Leicester, LE1 7RH, UK.

E-mail addresses: a.n.gorban@le.ac.uk (A.N. Gorban), vmakarov@ucm.es (V.A. Makarov), i.tyukin@le.ac.uk (I.Y. Tyukin).

<https://doi.org/10.1016/j.plrev.2019.06.003>

1571-0645/© 2019 Elsevier B.V. All rights reserved.

proach to this problem in a very recent publication [13]. Our first two stochastic separation theorems used the classical assumptions about either uniformity of distributions (uniform distribution in balls) or independence (product distributions in cubes) [14]. After addressing these basic cases we started our ‘flight from independence and uniformity’ in the paper “Augmented Artificial Intelligence: a Conceptual Framework” [15] and developed this work further in [16,17].

The intermediate conclusion is as follows. Stochastic separation theorems do not need hypotheses about independence and uniform distribution of data (as well as any other hypothesis about special distributions like the Gaussian one). The main condition we used instead of these simplifications is: sets of small volume should not have a high probability (further specifications on what ‘small’ and ‘large’ mean here can be found, for example in Theorem 1, Definition 3, and Theorem 6 from [17], Theorem 7 and Sec. 6 “Quasiorthogonal sets and Fisher separability of not i.i.d. data” in [16]). In particular, instead of uniform or Gaussian distributions, general log-concave distributions can be used, and this is just an example. Fisher separability of random points holds for more general distributions that satisfy the SmAC (SMeared Absolute Continuity) condition (see [17], Definition 3 or [16], Definition 4). Donoho and Tanner [18] observed ‘surprising’ universality of linear separation in numerical experiments with high-dimensional random datasets. They proved this property for Gaussian distributions and formulated the problem to determine a general definition of distributions with such separability. We now know that explicit and useful Fisher’s separability is also typical, and the general definitions of distributions with such properties are understood much better now.

The qualitative essence of these definitions is always the same: *small volume sets should not have a high probability*. If such sets exist then an additional challenge arises: extract the sets of small volume and relatively large probability from the data space. The remaining part will satisfy the stochastic separation theorems with Fisher’s separability, while the extracted set will have effectively lower dimension and can be approximated by the low-dimensional continua [19–21]. The extracted low-dimensional subsets can have complex topology and analysis of this topology is important for many modern applications like single-cell omics in bioinformatics [23]. Here we reveal a special complementarity principle [22]: the data space is split into a low volume (low dimensional) subset, which requires nonlinear methods for data approximation and analysis, and a high dimensional subset, where the linear methods (Fisher discriminants) work well.

Following the Kůrková comment [1] and papers [13,15–17], we invite all readers to join the flight from independence and uniformity. This is a call for revisiting the existing general machine learning theory to bring it closer to real life problems.

Linear algebra vs topology. Tozzi and Peters [2] used the Borsuk–Ulam Theorem (BUT) to explain the blessing of dimensionality. In the topological form (proven by Fet) [24] BUT is: *For every continuous involution $a \mapsto a^*$ ($a^{**} = a$) of a sphere S^n and every continuous map $f : S^n \rightarrow \mathbb{R}^n$ there exists a point $x \in S^n$ such that $f(x) = f(x^*)$.* The most well-known form of BUT uses the central symmetry $a \mapsto -a$ instead of a general continuous involution. The gluing of opposite points is interpreted by Tozzi as a decrease in information in topological sense. Reversely, the dimensionality increase ‘unglues’ points, and information increases. In more detail, this point of view is presented and illustrated with applications in the review paper [25]. It is necessary to add that for a continuous map $f : S^n \rightarrow \mathbb{R}^n$ ‘almost every’ point is glued with another point (just imagine a projection of a 2D sphere on a plane). We should stress, however, that the volume/probability concentration arguments with concentration of the volume near equator are invariant with respect to any orthogonal transformation that preserves the whitened distribution with unit covariance matrix. This property will be destroyed by a typical homeomorphism. On the contrary, the BUT in the topological form and other topological statements are invariant with respect to any homeomorphism. Both arguments, topological and linear, are important, but further work is needed to clarify the relations between them (at least, for us). Another intriguing problem with interplay of linear algebra and topology is the definition of ‘intrinsic dimension’ of data. Real data never are i.i.d. sample from a regular and fixed distribution. There are correlations between data points, the distribution (assuming that it exists) can change over time, there may exist a hidden “concept drift” and many other complications. We proposed to use the Fisher separability property for definition of intrinsic dimension of data: it is the dimension of a sphere having the same separability properties for i.i.d. samples from a uniform distribution (the ‘equivalent sphere’) [16]. This concept was developed further and tested on biological data [26]. Zinovyev et al. [26] showed how the suggested approach can be used to explore the structure of the data types that are generally considered hard to analyze (mutation and single cell RNA-Seq data). We expect further interesting and useful discoveries at the boundary between topology and geometry of data.

Communication between neuroscience and Artificial Intelligence (AI). Varona [3] discussed relations between machine learning and brain studies, from their common roots to a long period of divergence followed by the modern trend towards convergence. The “bidirectional communication between machine learning and neuroscience” is an important way to enrich both disciplines. Problems arising in the ontogeny of the brain and in machine learning can be similar. Both in the brain and in machine learning, re-training large ensembles of neurons is in fact impossible to realize in many real-life situations/applications. The robust low-dimensional neural dynamics in high dimensional brains is important for understanding implementation of cognitive functions (see, for example, [28]). However, there is a difference in this similarity. We agree with Varona that AI is far from incorporating all highly heterogeneous, dynamically robust, and error-handling elements of real brain. Further bi-directional exchange of ideas and methods between AI and neuroscience is needed.

Moving from “brainland” to “flatland” and backwards. Barrio [4] has directed our attention to the intriguing interplay between the low and high-dimensional worlds, between the high-dimensional brainland (the whole world of brain) and low-dimensional flatland of small neural ensembles. A flexible walk along the stairs of models of different dimensions is necessary to understand the brain dynamics. In data analysis, we can find a partial realization of this idea in our complementarity principle [22]: extraction of a lower dimensional subset for non-linear analysis [19] and application of simple linear methods to essentially high-dimensional part of data [16]. Some open access software for these problems is available on github [27].

Jewels of low-dimensional representation in complexity of high dimensional systems. Kreiman [5] analyzed three forms of the curse of dimensionality and high-dimensional pitfalls in neuroscience: (1) Huge diversity of possible stimuli and tasks; (2) Huge number of neurons; (3) No adequate tools – most modeling and analysis tools are developed for small dimensional worlds. Nevertheless, there are many examples in neuroscience, where small dimensional models explain a large fraction of a cognitive behavior. Most of them are manifestations of the concentration of measure phenomena. Kreiman attracted attention to the mechanism of removing the curse of dimensionality by projection on random bases. This approach is based on the Johnson and Lindenstrauss lemma and related topics [29,31,30]. According to this lemma and its probabilistic proof, a set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that ratios of distances between the points are nearly preserved. The map used for the embedding can even be taken to be a random orthogonal projection. This method of random projections was used in a recent publication [32] to explain a striking simplicity underlying multi-neuronal data. It is worth to mention that the Johnson and Lindenstrauss lemma and the method of random projections are based on the measure concentration phenomena [30]. The ‘brainland’ can turn out to be a small dimensional world after all.

Coupling of stochastic separation theory with nonlinear dynamics techniques. Fortuna [6] proposed to combine the data analysis methods based on statistical physics of data and advanced experimental based nonlinear dynamics techniques. He demonstrated the power of dynamical models in neuroscience by three examples: (i) a multi jump resonance behavior of an ensemble of neurons that explains selectivity, (ii) stochastic resonance in a group of neurons (when the signal to noise ratio at the output of the neuron ensemble is greater than the input signal to noise ratio), and (iii) synchronization effect in networks in the brain. The challenge is in effective combination of the stochastic separation technique and topological data approximation with low-dimensional dynamical models and model reduction methods.

Models differ from reality. van Leeuwen [7] presented an anti-cybernetic manifesto. He sought to show that the classical cybernetic understanding of the brain had run its course. According to his comment, cybernetic approach to brain study is a misleading metaphor. For example, it does not take into account non-local field interaction of neurons. The brain is actively generating predictions about what will happen and engages in controlled hallucination. The brain, unlike legacy software, is constantly updating itself. These and other important properties of the brain may deplete the cybernetic approach and make it useless. Such a serious gap between modern AI and real brain can ruin our hope for the efficient communication between neuroscience and AI. According to van Leeuwen, we have to modify our mathematical and computer science approaches so that this communication becomes useful for both parts. What can we say here? First of all, models always differ from reality. The question is not, however, whether there exist important properties of the brain that are neglected by a model. They always exist. The question is whether the model reflects any important aspects of reality. A model is an intelligent device that should perform useful work. Simple models, ranged from stochastic separation, concept cells, and Hebb’s rule described in our review [17] to direct simulation of

large networks of relatively simple neurons [34] and rule based models of development neuronal network generating behavior [35], have not yet exhausted their capabilities. Nevertheless, we are ready to accept the expert opinion of van Leeuwen and agree that a new conceptual basis of brain modeling is needed. Let us indeed hope that a new theoretical framework will be eventually developed in neuroscience that will appeal to mathematician and computer scientists, and there will be a decent place for simple basic models.

Akaki Akakievich and the modern point of view on Concept Cells. Quian Quiroga [8] clarified and commented on modern ideas about Concept Cells and sparse coding in the form of a revised Lettvin story about grandmothers cells. He stressed that Concept Cells are not involved in identifying a particular stimulus or concept. They are rather involved in creating and retrieving associations and can be seen as the “building blocks of episodic memory”. We analyzed basic dynamical models of such associations [33] (see also [17], Sec. 4. “Encoding and rapid learning of memories by single neurons”) but this analysis should be extended and compared quantitatively to experimental data. The professional and clearly presented explanation of the basic concepts in Quian Quiroga’s commentary is a very useful addition to our review, and we are happy to recommend this commentary to all readers.

“Happiness is when you are understood”. The beautiful text of Kreinovich [9] does not need comments. It is our pleasure just to mention that in our internal discussions of the manuscript [17], we have many times quoted Pasternak’s poetry about the “heresy of unheard-of simplicity”. Kreinovich found the same golden words of Pasternak to describe the apparent simplicity in a multidimensional brain. This is an example of an amazing intellectual resonance.

Acknowledgements

The work was supported by the Ministry of Science and Higher Education of the Russian Federation (Project No. 14.Y26.31.0022). Work of ANG and IVT was also supported by Innovate UK (Knowledge Transfer Partnership grants KTP009890; KTP010522) and University of Leicester. VAM acknowledges support from the Spanish Ministry of Economy, Industry, and Competitiveness (grant FIS2017-82900-P).

References

- [1] Kůrková V. Some insights from high-dimensional spheres: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:98–100. <https://doi.org/10.1016/j.plrev.2019.03.014> [in this issue].
- [2] Tozzi A, Peters JF. The Borsuk-Ulam theorem solves the curse of dimensionality: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:89–92. <https://doi.org/10.1016/j.plrev.2019.04.008> [in this issue].
- [3] Varona P. High and low dimensionality in neuroscience and artificial intelligence: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by A.N. Gorban et al. *Phys Life Rev* 2019;29:106–7. <https://doi.org/10.1016/j.plrev.2019.02.008> [in this issue].
- [4] Barrio R. “Brainland” vs. “flatland”: how many dimensions do we need in brain dynamics? Comment on the paper “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:108–10. <https://doi.org/10.1016/j.plrev.2019.02.010>. in this issue.
- [5] Kreiman G. It’s a small dimensional world after all: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:96–7. <https://doi.org/10.1016/j.plrev.2019.03.015> [in this issue].
- [6] Fortuna L. Nonlinear effects for the reinforcement of small neural ensembles in high dimensional brain: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by A.N. Gorban, V.A. Makarov, I.Y. Tyukin. *Phys Life Rev* 2019;29:101–3. <https://doi.org/10.1016/j.plrev.2019.03.004> [in this issue].
- [7] van Leeuwen C. The reasonable ineffectiveness of biological brains in applying the principles of high-dimensional cybernetics: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:104–5. <https://doi.org/10.1016/j.plrev.2019.03.005> [in this issue].
- [8] Quian Quiroga R. Akakievich revisited: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by Alexander N. Gorban et al. *Phys Life Rev* 2019;29:111–4. <https://doi.org/10.1016/j.plrev.2019.02.014> [in this issue].
- [9] Kreinovich V. The heresy of unheard-of simplicity: Comment on “The unreasonable effectiveness of small neural ensembles in high-dimensional brain” by A.N. Gorban, V.A. Makarov, I.Y. Tyukin. *Phys Life Rev* 2019;29:93–5. <https://doi.org/10.1016/j.plrev.2019.04.006> [in this issue].
- [10] Lévy P. *Problèmes concrets d’analyse fonctionnelle*. Paris: Gauthier-Villars; 1951.
- [11] Ball K. *An elementary introduction to modern convex geometry*. Flavors of geometry, vol. 31. Cambridge, UK: MSRI Publications; 1997.
- [12] Talagrand M. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. IHES* 1995;81:73–205.
- [13] Kůrková V, Sanguinetti M. Probabilistic bounds for binary classification of large data sets. In: Oneto L, Navarin N, Sperduti A, Anguita D, editors. *Recent advances in big data and deep learning*. Proceedings of the international neural networks society, vol. 1. Springer; 2019. p. 309–19.

- [14] Gorban AN, Tyukin IY. Stochastic separation theorems. *Neural Netw* 2017;94:255–9. <https://doi.org/10.1016/j.neunet.2017.07.014>.
- [15] Gorban AN, Grechuk B, Tyukin IY. Augmented artificial intelligence: a conceptual framework. *arXiv preprint*, arXiv:1802.02172 [cs.AI]. <https://arxiv.org/abs/1802.02172>, 2018.
- [16] Gorban AN, Golubkov A, Grechuk B, Mirkes EM, Tyukin IY. Correction of AI systems by linear discriminants: probabilistic foundations. *Inf Sci* 2018;466:303–22. <https://doi.org/10.1016/j.ins.2018.07.040>.
- [17] Gorban AN, Makarov VA, Tyukin IY. The unreasonable effectiveness of small neural ensembles in high-dimensional brain. *Phys Life Rev* 2019;29:55–88. <https://doi.org/10.1016/j.plrev.2018.09.005> [in this issue].
- [18] Donoho D, Tanner J. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos Trans R Soc A* 2009;367:4273–93. <https://doi.org/10.1098/rsta.2009.0152>.
- [19] Gorban AN, Zinovyev A. Principal manifolds and graphs in practice: from molecular biology to dynamical systems. *Int J Neural Syst* 2010;20(03):219–32. <https://doi.org/10.1142/S0129065710002383>.
- [20] Gorban AN, Kégl B, Wunsch D, Zinovyev A, editors. *Principal manifolds for data visualisation and dimension reduction. Lecture notes in computational science and engineering*, vol. 58. Springer: Berlin–Heidelberg–New York; 2008.
- [21] Zinovyev A, Mirkes E. Data complexity measured by principal graphs. *Comput Math Appl* 2013;65(10):1471–82. <https://doi.org/10.1016/j.camwa.2012.12.009>.
- [22] Gorban AN, Tyukin IY. Blessing of dimensionality: mathematical foundations of the statistical physics of data. *Philos Trans R Soc A* 2018;376:20170237. <https://doi.org/10.1098/rsta.2017.0237>.
- [23] Chen H, Albergante L, Hsu JY, Lareau CA, Bosco GL, Guan J, et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun* 2019;10:1903. <https://doi.org/10.1038/s41467-019-09670-4>.
- [24] Fet AI. Generalization of a theorem of Lusternik–Schnirelman on covering of spheres. *Dokl Akad Nauk SSSR* 1954;95:1149–52.
- [25] Tozzi A. The multidimensional brain. *Phys Life Rev* 2019. <https://doi.org/10.1016/j.plrev.2018.12.004>.
- [26] Albergante L, Bac J, Zinovyev A. Estimating the effective dimension of large biological datasets using Fisher separability analysis. *arXiv*: 1901.06328 [cs.LG]. <https://arxiv.org/pdf/1901.06328.pdf>, 2019.
- [27] Laboratory of advanced methods for high-dimensional data analysis. <https://github.com/lamhda>.
- [28] Rabinovich MI, Varona P. Discrete sequential information coding: heteroclinic cognitive dynamics. *Front Comput Neurosci* 2018;12:73. <https://doi.org/10.3389/fncom.2018.00073>.
- [29] Johnson W, Lindenstrauss J. Extensions of Lipschitz maps into a Hilbert space. In: *Conference modern analysis and probability. Contemp Math*, vol. 26. 1984. p. 189–206.
- [30] Gorban AN, Tyukin I, Prokhorov D, Sofeikov K. Approximation with random bases: Pro et contra. *Inf Sci* 2016;364–365:129–45. <https://doi.org/10.1016/j.ins.2015.09.021>.
- [31] Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality. In: *STOC'98 proceedings of the thirtieth annual ACM symposium on theory of computing*. New York, NY: ACM; 1998. p. 604–13.
- [32] Gao P, Trautmann E, Yu B, Santhanam G, Ryu S, Shenoy K, et al. A theory of multineuronal dimensionality, dynamics and measurement, *bioRxiv* 2017, 214262. <https://doi.org/10.1101/214262>.
- [33] Tyukin IY, Gorban AN, Calvo C, Makarova J, Makarov VA. High-dimensional brain. A tool for encoding and rapid learning of memories by single neurons. *Bull Math Biol* 2018. <https://doi.org/10.1007/s11538-018-0415-5>.
- [34] Izhikevich EM, Edelman GM. Large-scale model of mammalian thalamocortical systems. *Proc Natl Acad Sci USA* 2008;105(9):3593–8. <https://doi.org/10.1073/pnas.0712231105>.
- [35] Roberts A, Conte D, Hull M, Merrison-Hort R, al Azad AK, Buhl E, et al. *J Neurosci* 2014;34(2):608–21. <https://doi.org/10.1523/JNEUROSCI.3248-13.2014>.