Letter

Mechanical regularization

Himangsu Bhaumik* and Daniel Hexner†
Faculty of Mechanical Engineering, Technion, 320000 Haifa, Israel

(Received 28 September 2023; revised 29 May 2024; accepted 29 July 2024; published 13 August 2024)

Training materials through periodic drive allows us to endow materials and structures with complex elastic functions. As a result of the driving, the system explores the high-dimensional space of structures, ultimately converging to a structure with the desired response. However, increasing the complexity of the desired response results in ultraslow convergence and degradation. Here, we show that by constraining the search space, we are able to increase robustness, extend the maximal capacity, train responses that previously did not converge, and in some cases accelerate convergence by many orders of magnitude. We identify the geometrical constraints that prevent the formation of spurious low-frequency modes, which are responsible for failure. We argue that these constraints are analogous to regularization used in machine learning. We propose a unified relationship between complexity, degradation, convergence, and robustness.

DOI: 10.1103/PhysRevMaterials.8.L082601

Introduction. The multitude of variable degrees of freedom allows various systems to perform complex tasks, including computations [1], classification [2,3], regulation [4,5], and processing of high-dimensional data. Examples include neural networks (both in vivo and artificial) [6,7], regulatory networks [4,8], flow networks [9], and more recently, mechanical structures [10–19]. Often these systems are overparameterized, having many more degrees of freedom than required for the desired task. As a result, there are many solutions, each with a different set of microscopic parameters, which in principle, could have different properties. Biasing the search algorithms could provide beneficial solutions, which, for example, are more robust, or perhaps more expressive [20,21]. Indeed, machine learning algorithms employ regularization techniques, for example, to prevent overfitting, yielding better generalization. Examples of regularization methods include constraining the set of parameters, biasing the loss function, and early stopping [22,23].

In this paper, we introduce a regularization method for training elastic responses in viscoelastic structures [24]. We build on recent ideas for endowing precise elastic responses in mechanical systems without the aid of a computer [12,13,16,25–27]. A material is trained by applying sequences of strains that produce changes to the structure through plastic deformations. Through repetitive driving, the system may converge to the desired response. The benefit of training materials, as opposed to design and fabrication [28], is that doing so relinquishes the need to manually control a large number of microscopic degrees of freedom.

The notion of material training introduces new considerations [29]. First, the time scales or cycles needed to train new responses are important factors, especially in light of recent findings that convergence can be very slow. Second, training requires repetitive external driving, which has the effect of degrading the material [29,30]. Lastly, it is desirable to find solutions that are robust to small perturbations of the structure.

In the paper, we show that constraining the angles between bonds has a profound effect on training, and acts as a regularizer. While reducing the accessible set of solutions, it selects solutions with beneficial properties. The solutions that are found have an overall larger rigidity, capacity, and increased robustness. Surprisingly, in some cases, the constrained search may accelerate convergence. We propose unified relations [see Fig. 1(a)], based on an analysis of the density of state, between degradation, convergence, complexity, and robustness.

Model and training algorithm. Following Ref. [25,29], we study a disordered bonded network of Maxwell viscoelastic elements, each composed of a spring and a dashpot in series [24] [illustrated in Fig. 1(b)]. On short-time scales the system can be considered an elastic solid, while on large-time scales it undergoes creep. The tension on each bond is given by

$$t_i = k_i(\ell_i - \ell_{i,0}), \tag{1}$$

where k_i is the spring constant, ℓ_i is the bond's length, and $\ell_{i,0}$ is the rest length. Our goal is to control the elastic response of the network by altering the geometry of the network, through changes to the rest lengths ("learning degrees of freedom"). We assume that the rest lengths evolve through plastic deformations that depend linearly on the tension on the bond,

$$\partial_t \ell_{i,0} = \gamma k_i(\ell_i - \ell_{i,0}). \tag{2}$$

We take the quasistatic limit where the energy is minimized to reach force balance after every change to the rest lengths.

The elastic properties of disordered random networks depend on the coordination number, $Z = \frac{2N_B}{N}$, where N_B is the number of bonds, and N is the number of nodes. When $Z > Z_C \approx 2d$, the networks are rigid [31–33]. We consider two limits, small and large excess coordination number

*Present address: Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom.
†Contact author: danielhe@me.technion.ac.il

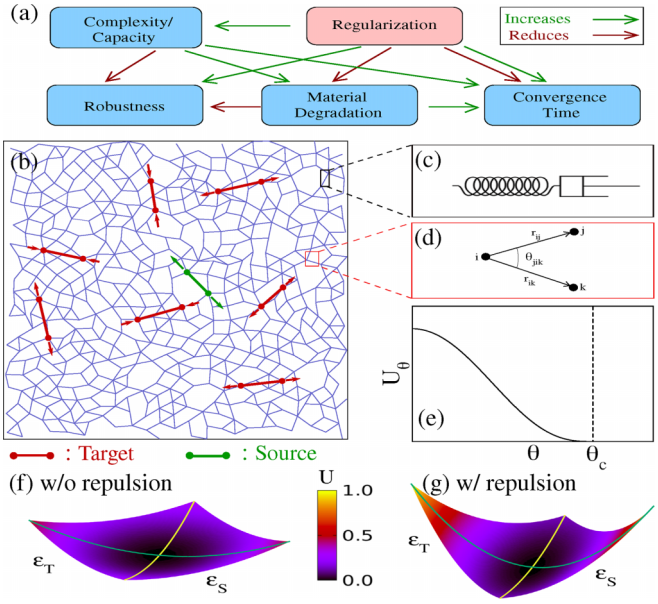

FIG. 1. An illustration of the model: (a) Schematic representation of the effect of regularization and interconnection of various aspects of material training. (b) An example of an elastic network. Each pair of source (green) and target (red) sites are connected by a line, and the arrows represent the phase of the response. (c) Each bond is a spring and a dashpot in series. The dashpot allows for changes in the structure, which alters the elastic properties. (d) In addition, we include an angular repulsive force that prevents the angles between adjacent bonds from becoming small. (e) The angular potential is nonzero only when $\theta < \theta_c$. (f), (g) an illustration of the effect of angular regularization (repulsion). The energy valley with repulsion is deeper and has a larger transverse stiffness, which leads to enhanced properties.

$\Delta Z \equiv Z - Z_c$, since their elastic properties are different. For small ΔZ the network is isostatic and the elastic response is anomalously long-ranged [34,35], whereas, for large ΔZ the Green's function decays quickly [35].

As a test bed for regularization, we consider responses whose difficulty can be tuned, by varying the number of "target" sites whose response we wish to tune and the strain amplitude ϵ_{Age}. Following Ref. [11,29], the desired response is such that an input strain on a single-source site yields a prescribed strain on N_T target sites. For simplicity, the input and output strain amplitude is taken to be the same, ϵ_{Age}, however, the response on each target is chosen with equal probabilities to be either in-phase or out-of-phase.

The source and target site are coupled through an energy "valley," as illustrated in Figs. 1(f) and 1(g) along the $\epsilon_T = \epsilon_S$ direction. That is, the energy as a function of strains on the source and targets is small along the desired trajectory in comparison to the transverse directions ($\epsilon_T = -\epsilon_S$). Periodic drive along the desired response, during training, gradually reduces the energy along that path through plastic deformations that change the rest lengths. Ultimately, training converges to the desired response, provided that the response is not too difficult [29].

With increasing complexity training becomes difficult, which is expressed in the slowdown of convergence [29]. At a critical threshold, the convergence time appears to diverge, marking the limit of trainable responses. As noted, failure occurs through the proliferation of low-frequency modes, which we refer to as degradation. Similar effects have also been seen in the loss landscape of neural networks [36–38]. We have previously found that these can be traced to local geometrical features, where pairs of bonds nearly align [29]. We are therefore motivated to prevent the angles between bonds from becoming small. To this end, we introduce an angular repulsive potential ($U_\theta = k_\theta[1 - (\cos\theta - 1)/(\cos\theta_c - 1)]^3$) that acts when $\theta < \theta_c$, illustrated in Fig. 1(d). Inclusion of an angular modulus is known to increase the rigidity of the network [39–42]. We emphasize that the angular forces have the effect of constraining the parameter space, and after training, these forces vanish. Hence, the form of the potential is not important (For a detailed discussion, see Supplemental Material (SM) [43]), where we consider a quadratic form instead of cubic. Figure 1(g) illustrates the effect of the angular repulsion to increase the transverse stiffnesses.

Linear analysis. Within the linear response, the energy landscape can be characterized by the eigenmodes and frequencies of the system. The relation between an applied force, f and the resulting displacement δx depends on the Hessian, H, which can be decomposed in the eigenmodes, $\vec{e_\omega}$:

$$\vec{\delta x} = H^{-1}\vec{f} = \sum_i \frac{\vec{e_{\omega_i}} \cdot \vec{f}}{\omega_i^2} \vec{e_{\omega_i}}. \tag{3}$$

Here, $\vec{f}$ corresponds to the force acting on the source, and its amplitude is set to provide a desired strain.

Since the training rule reduces the energy along the desired trajectory, the lowest frequency mode ω_1 after training corresponds to the trained response. The remaining transverse modes compete with the desired response. Since the contribution of each mode is inversely proportional to its frequency squared, the response is predominantly influenced by the two lowest frequency modes, ω_1 and ω_2. Consequently, the magnitude of the error can be approximated by the ratio of the squared frequencies of these two modes: $\delta\epsilon \sim \frac{\omega_1^2}{\omega_2^2}$ [29]. Training is successful when $\omega_1 \ll \omega_2$.

Accelerated convergence. We begin discussing the convergence by measuring the error between the desired and measured response on the targets,

$$\delta\epsilon^2 = \frac{1}{N_T} \sum_T \frac{\left(\epsilon_T - \epsilon_T^{\text{desired}}\right)^2}{\epsilon_{\text{Age}}^2}. \tag{4}$$

Here, ϵ_T is the strain on a target. Previously, it was found that the error approximately decays as a power-law $\tau^{-\alpha}$ and that α depends on the number of targets per node, denoted by, $\Delta = \frac{N_T}{N}$. With increased complexity, α decreases and vanishes at a critical point. The transition marks the maximal number of sites that can be trained [29].

In Fig. 2, we compare the results with and without the angular constraints. The difficulty of the task is tuned by modifying either ϵ_{Age} or Δ. Figures 2(a) and 2(c) show the training error $\delta\epsilon$ as a function of the number of cycles, τ. Figures 2(b) and 2(d) show α as a function of ϵ_{Age} and Δ, characterizing the speed of convergence.

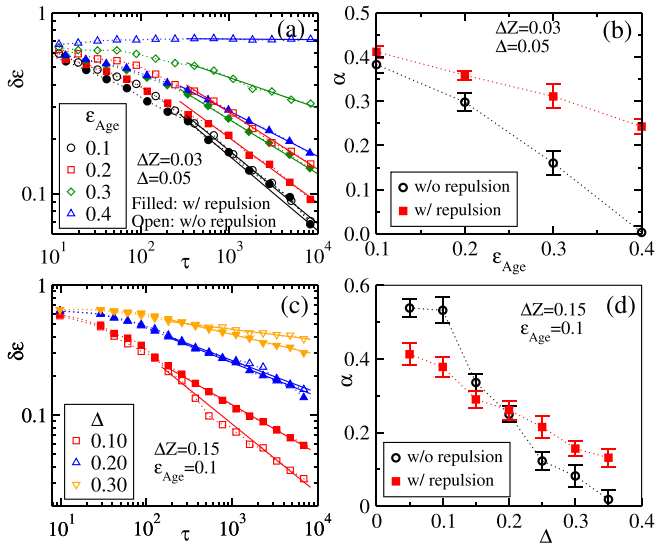

FIG. 2. Convergence of training error: The error against τ with repulsion (filled symbols) and without repulsion (open symbols) for different values of ϵ_{Age} (a) and Δ (c). The exponent α as a function of ϵ_{Age} (b) and Δ (d). Convergence is accelerated with angular repulsion in (b) and at large values of Δ in (d). Note that for both (b) and (d) there is a larger capacity when angular repulsion is present. Here, $N = 200$.

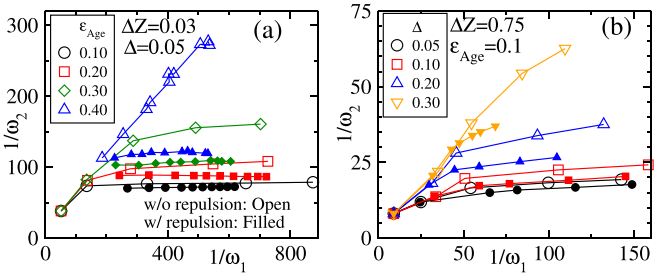

FIG. 3. Suppressed degradation: $1/\omega_2$ against $1/\omega_1$, where ω_1 is the lowest nontrivial frequency, corresponding to the trained response and ω_2 is the second smallest frequency. (a) $\Delta Z \approx 0.03$ and we vary ϵ_{Age} and in (b) $\Delta Z \approx 0.75$ and we vary Δ. Note that solutions with repulsion have a smaller ω_2 per value of ω_1, which leads to enhanced properties. Here, $N = 200$.

The angular constraints may accelerate convergence (larger α) or slow convergence (smaller α). Acceleration is more common for smaller coordination numbers [Figs. 2(a) and 2(b)] and at larger difficulty, where the slow convergence is due to the low-frequency spurious modes. Figure 2(d) shows a transition from a slow-down to acceleration when varying Δ. Below, we discuss how this may arise from two competing effects.

We argue that a speedup in convergence is surprising. Imagine we minimize a loss function through gradient descent. Adding a constraint to the loss function changes the gradient. Since the unconstrained gradient is the direction of greatest change to the function, following the constrained gradient will typically result in a slower convergence. Indeed, here, we find that the energy along the trained trajectory decreases at a slower rate (see SM), yet results in an overall faster convergence. We also note that regularization and metadynamics have previously been used to accelerate convergence in other systems [44–46].

The angular constraints also affect the capacity, which is defined by the value of Δ or ϵ_{Age} where the exponent α vanishes. Figures 2(b) and 2(d) show that the angular constraints increase the capacity, allowing responses with a larger Δ and ϵ_{Age}. Additional data on convergence is provided in the SM.

Reduced degradation. To track down the cause of the speedup in convergence, we study the low-frequency spectrum. While training reduces the stiffness along the desired path, we have also found [29] that it has the unintended effect of reducing the stiffness of the transverse modes, particularly as large Δ and large ϵ_{Age}. Here, we show that our regularized training method arrests these spurious low-frequency modes from creeping to lower frequencies.

To characterize the evolution of spurious mode we measure ω_2, the second-lowest nontrivial eigenfrequency, per given value of lowest frequency ω_1. This compares the transverse stiffness of the spurious modes to the longitudinal stiffness along the desired motion. For convenience, we plot $1/\omega_2$ as a function of $1/\omega_1$ in Fig. 3. Since ω_1 decreases with training, $1/\omega_1$ measures the progress of training. Ideally, $1/\omega_2$ should be small in comparison to $1/\omega_1$.

Figure 3 shows $1/\omega_2$ vs. $1/\omega_1$ for different ϵ_{Age} (in (a))and for different Δ (in (b)). For small ϵ_{Age} (or Δ) $1/\omega_2$, grows very slowly for both training protocols. This is the regime with low difficulty, where the training error decays faster. However, for larger values of ϵ_{Age} [or Δ in (b)], a striking difference is observed. In the presence of repulsion, $1/\omega_2$ grows at a slower rate. Namely, the transverse modes are suppressed.

In summary, the angular repulsion allows us to find "solutions" with deeper energy valleys, whose transverse stiffnesses are larger. The arrested degradation also allows us to understand the accelerated convergence. Recall that the error scales as $\delta\epsilon \propto \omega_1^2/\omega_2^2$. The angular constraints have two competing effects: 1. Suppressing spurious modes (ω_2). 2. Decreasing the rate of the formation of the valley, i.e., ω_1 decreases at a slower rate (see also the SM).

The result of these two effects is that regularization either speeds up or slows down convergence. Acceleration occurs if the transverse modes (ω_2) are suppressed, while the longitudinal mode (ω_1) is weakly affected. Acceleration appears to occur at large difficulty (large Δ or large ϵ_{Age}) where the excess low-frequency modes hinder convergence, as well as at a small ΔZ where the density of states extends to lower frequencies, even before training.

Coupling distant sites. Here, we show that regularization enables training responses that previously did not converge. We focus on the allostery-inspired response, where pinching a source site results in motion on a *single* faraway target site [10,47,48].

At large ΔZ, when the source and target sites are distant, training does not converge. In this regime, the response to a localized perturbation decays quickly as a function of distance (e.g., the tension decays as r^{-d}, where d is the dimension), and as a result, distant sites do not couple [25]. Instead of a single mode that couples the source and target, the stiffness for

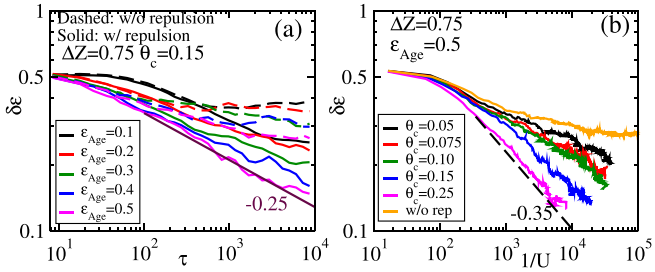

FIG. 4. Training allostery-inspired response:(a) The error as a function of the number of training cycles for different amplitudes, with (solid line) and without (dashed line) angular repulsion for an allosteric response (single source and single target site). Without repulsive regularization, the response does not converge. (b) The training error against the inverse of longitudinal energy $1/U$ for different θ_c. Here, $N = 200$, and $\Delta Z = 0.75$.

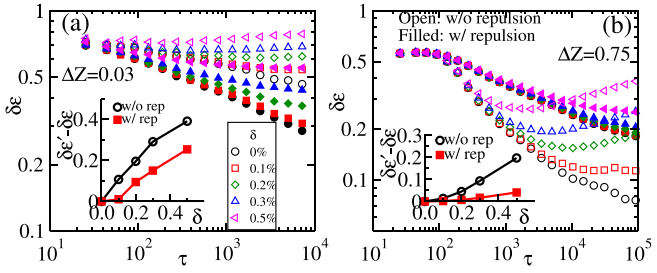

FIG. 5. Robustness to small perturbations: Training error after randomly perturbing the rest lengths with an amplitude δ. In (a) $\Delta Z = 0.03$ and in (b) $\Delta Z = 0.75$. Dashed lines indicate training without repulsion while solids indicate training with repulsion. Inset shows the change in error due to the perturbations, as a function of δ. Note that the repulsion leads to enhanced robustness (a smaller change in the error). Here, $N = 200$, $\epsilon_{\text{Age}} = 0.1$, and $\Delta = 0.2$.

actuating the source and target individually and independently becomes small.

Figure 4(a) shows the error when training for the allostery-inspired response at different strain amplitudes. Without the angular forces, the error saturates at a finite value. With the angular repulsion, the error continually decreases, even after 10^4 training cycles. Thus, the angular constraints enable us to train responses that do not usually converge. Interestingly, the effectiveness of the angular constraints is most pronounced at large strain amplitudes, yielding the fastest convergence. In Fig. 4(b) we also show the error as a function of the inverse energy, which measures the progress in creating a low-energy valley. Here, larger θ_c results in a smaller error.

Enhanced robustness. We next discuss the effect of the angular constraints on the robustness of the networks to small perturbations. To characterize the robustness, we alter each rest length randomly $\ell_{i,0} \to \ell_{i,0}(1 + \delta_i)$, where δ_i is uniformly distributed in the range $[-\delta, \delta]$, and measure the change in the error. We emphasize that training occurs as before and that the perturbations are only introduced after training to characterize the robustness.

We begin by working out the predictions of the linear response. As noted, the error scales as ω_1^2/ω_2^2. The small changes to the structure change both frequencies by order δ. Since ω_1 is smaller than ω_2 the relative change to the frequencies is more significant for ω_1. We expect that the random changes to the structure will generically increase ω_1, since it is uncharacteristically small. Therefore, the error after the perturbations scales as

$$\delta\epsilon' \sim \frac{\omega_1^2 + A\delta}{\omega_2^2} \sim \delta\epsilon + \frac{A\delta}{\omega_2^2}. \tag{5}$$

Here, A is a constant. The change in the error is large when the transverse stiffness, ω_2, is small. Namely, a larger gap implies a higher robustness.

In Fig. 5 we present the error as a function of the number of training cycles for different values of perturbation amplitudes, δ, with and without the angular repulsion. Figure 5(a) shows the results for small ΔZ where training is quicker with the angular repulsion. In the inset, we show that the change with repulsion is smaller, and therefore more robust.

For large ΔZ, the effect is more substantial as shown in Fig. 5(b). Before applying the perturbation, $\delta = 0$, the error is smaller without the angular repulsion. However, when the system is perturbed, the error at long times can be smaller with angular repulsion. Thus, the angular regularization leads to more robust responses.

Interestingly, without the angular repulsion, the error for $\delta > 0$ is nonmonotonic in time; initially, it decreases and then at later times it increases. Thus, an early stopping condition can also yield more robust results. The nonmonotonic trend over time can be interpreted as overfitting to the training data. Often, learning models are trained on a learning dataset and tested on a distinct test data set. A signature of overfitting to the training data is a nonmonotonic evolution of the test error, which initially decreases and then later increases. Here, the unperturbed system represents training data, and the perturbed system acts as the test data. Extensive training of the unperturbed system reduces the "training error" but compromises the robustness, resulting in a larger "test error." The proposed regularization method helps counteract this problem.

Conclusions and discussion. In summary, we have introduced a regularization method for training viscoelastic networks that constrains the angles between bonds through a repulsive force. This regularization has the effect of increasing the capacity, increasing robustness, reducing degradation, and in some cases, accelerating convergence. Furthermore, it allows us to train responses that previously could not be trained in the same manner.

The central problem with the unconstrained training rule is that, while it reduces the energy along the desired trajectory, it has the unwanted effect of also reducing the energy along unintended directions [29,49]. The formation of the spurious low-frequency modes is due to the local structural features of aligning bonds. The angular repulsion, which constrains small angles, prevents the proliferation of soft modes, and therefore degradation. In the SM, we show that this regularizer is also effective in other training rules, suggesting broader generality.

Our analysis suggests a unified understanding of the relations between convergence, degradation, complexity, and robustness, illustrated in Fig. 1(a). Increasing complexity reduces system robustness, prolongs training times, and

exacerbates degradation. Training is slower due to degradation effects. Excess low-frequency modes compete with desired responses, necessitating longer training times. Systems with small transverse stiffnesses are particularly sensitive to perturbations, reducing overall robustness. As a result, complex responses that have a small frequency "gap" are therefore less robust.

Our study is reminiscent of work on energy-based models in machine learning, where information is encoded in an energy landscape [23]. Tailoring the energy landscape requires careful control of the high-dimensional parameter space. Unlike contrastive methods, regularization techniques aim to constrain overall energy, ensuring that lowering it in one area raises it elsewhere. This approach avoids the need for directly manipulating the entire landscape [23] or having duplicate systems [50].

Acknowledgments. We would like to thank Marc Berneman for fruitful discussions. This work was supported by the Israel Science Foundation (Grant No. 2385/20) and the Alon Fellowship.

[1] H. T. Siegelmann, Science **268**, 545 (1995).
[2] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, Artificial Intelligence Review **26**, 159 (2006).
[3] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, in *Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona* (Citeseer, 2011).
[4] U. Alon, Nat. Rev. Genet. **8**, 450 (2007).
[5] A.-L. Barabasi and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
[6] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2005).
[7] J. J. Hopfield, Proc. Natl. Acad. Sci. **79**, 2554 (1982).
[8] E. Davidson and M. Levin, Proc. Natl. Acad. Sci. **102**, 4935 (2005).
[9] K. Bhattacharyya, D. Zwicker, and K. Alim, Phys. Rev. Lett. **129**, 028101 (2022).
[10] J. W. Rocks, N. Pashine, I. Bischofberger, C. P. Goodrich, A. J. Liu, and S. R. Nagel, Proc. Natl. Acad. Sci. **114**, 2520 (2017).
[11] J. W. Rocks, H. Ronellenfitsch, A. J. Liu, S. R. Nagel, and E. Katifori, Proc. Natl. Acad. Sci. **116**, 2506 (2019).
[12] M. Stern, D. Hexner, J. W. Rocks, and A. J. Liu, Phys. Rev. X **11**, 021045 (2021).
[13] N. Pashine, D. Hexner, A. J. Liu, and S. R. Nagel, Sci. Adv. **5**, eaax4215 (2019).
[14] N. Pashine, Phys. Rev. Mater. **5**, 065607 (2021).
[15] N. Pashine, A. M. Nasab, and R. Kramer-Bottiglio, Soft Matter **19**, 1617 (2023).
[16] V. P. Patil, I. Ho, and M. Prakash, arXiv:2304.08711 [cond-mat.soft].
[17] R. H. Lee, E. A. B. Mulder, and J. B. Hopkins, Sci. Rob. **7**, eabq7278 (2022).
[18] V. F. Hagh, S. R. Nagel, A. J. Liu, M. L. Manning, and E. I. Corwin, Proc. Natl. Acad. Sci. **119**, e2117622119 (2022).
[19] S. Kim and S. Hilgenfeldt, Phys. Rev. Lett. **129**, 168001 (2022).
[20] C. M. Bishop, Neural Comput. **7**, 108 (1995).
[21] A. Neelakantan, L. Vilnis, Q. V. Le, I. Sutskever, L. Kaiser, K. Kurach, and J. Martens, arXiv:1511.06807.
[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[23] Y. Lecun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, A tutorial on energy-based learning, in *Predicting Structured Data*, edited by G. Bakir, T. Hofman, B. Scholkopt, A. Smola, and B. Taskar (MIT Press, 2006).
[24] J. C. Maxwell, Philos. Trans. R. Soc. London **157**, 49 (1867).
[25] D. Hexner, A. J. Liu, and S. R. Nagel, Proc. Natl. Acad. Sci. **117**, 31690 (2020).
[26] V. R. Anisetti, B. Scellier, and J. M. Schwarz, Phys. Rev. Res. **5**, 023024 (2023).
[27] C. Arinze, M. Stern, S. R. Nagel, and A. Murugan, Phys. Rev. E **107**, 025001 (2023).
[28] S. H. Thomke, Manag. Sci. **44**, 743 (1998).
[29] H. Bhaumik and D. Hexner, Phys. Rev. Res. **4**, L042044 (2022).
[30] S. Suresh, *Fatigue of Materials* (Cambridge University Press, 1998).
[31] D. J. Durian, Phys. Rev. Lett. **75**, 4780 (1995).
[32] C. S. O'Hern, L. E. Silbert, A. J. Liu, and S. R. Nagel, Phys. Rev. E **68**, 011306 (2003).
[33] A. J. Liu and S. R. Nagel, Annu. Rev. Condens. Matter Phys. **1**, 347 (2010).
[34] W. G. Ellenbroek, E. Somfai, M. van Hecke, and W. van Saarloos, Phys. Rev. Lett. **97**, 258001 (2006).
[35] E. Lerner, E. DeGiuli, G. Düring, and M. Wyart, Soft Matter **10**, 5085 (2014).
[36] G. Blanc, N. Gupta, G. Valiant, and P. Valiant, in *Conference on Learning Theory* (PMLR, 2020), pp. 483–513.
[37] N. Yang, C. Tang, and Y. Tu, Phys. Rev. Lett. **130**, 237101 (2023).
[38] A. Ratzon, D. Derdikman, and O. Barak, Elife **12**, RP90069 (2024).
[39] C. P. Broedersz, X. Mao, T. C. Lubensky, and F. C. MacKintosh, Nat. Phys. **7**, 983 (2011).
[40] C. P. Broedersz, M. Sheinman, and F. C. MacKintosh, Phys. Rev. Lett. **108**, 078102 (2012).
[41] D. B. Liarte, O. Stenull, X. Mao, and T. C. Lubensky, J. Phys.: Condens. Matter **28**, 165402 (2016).
[42] J. Michel, G. von Kessel, T. W. Jackson, L. J. Bonassar, I. Cohen, and M. Das, Phys. Rev. Res. **4**, 043152 (2022).
[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.8.L082601 for details on the implementation of the angular potential, normal mode analysis, additional data on large coordinated networks, the role of critical angle θ_c and the generality of the regularization, which includes Refs. [51–53].
[44] A. Barducci, M. Bonomi, and M. Parrinello, WIREs Comput. Mol. Sci. **1**, 826 (2011).
[45] H. Fu, X. Shao, W. Cai, and C. Chipot, Acc. Chem. Res. **52**, 3254 (2019).
[46] P. Chaudhari and S. Soatto, arXiv:1511.06485 [cs.LG].
[47] L. Yan, R. Ravasio, C. Brito, and M. Wyart, Proc. Natl. Acad. Sci. **114**, 2526 (2017).

[48] J.-P. Eckmann, J. Rougemont, and T. Tlusty, Rev. Mod. Phys. **91**, 031001 (2019).
[49] M. Stern, A. J. Liu, and V. Balasubramanian, Phys. Rev. E **109**, 024311 (2024).
[50] S. Dillavou, M. Stern, A. J. Liu, and D. J. Durian, Phys. Rev. Appl. **18**, 014040 (2022).
[51] B. Scellier and Y. Bengio, Front. Comput. Neurosci. **11**, 24 (2017).
[52] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Phys. Rev. Lett. **97**, 170201 (2006).
[53] C. P. Goodrich, W. G. Ellenbroek, and A. J. Liu, Soft Matter **9**, 10993 (2013).