Optimal Design of Experiments by Combining Coarse and Fine Measurements

Alpha A. Lee, Michael P. Brenner, and Lucy J. Colwell

Phys. Rev. Lett. **119**, 208101 — Published 16 November 2017

DOI: 10.1103/PhysRevLett.119.208101

Optimal design of experiments by combining coarse and fine measurements

Alpha A. Lee*
Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK and
School of Engineering and Applied Sciences and Kavli Institute of Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA

Michael P. Brenner
School of Engineering and Applied Sciences and Kavli Institute of Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA

Lucy J. Colwell†
Department of Chemistry, University of Cambridge, CB2 1EW, Cambridge, UK

In many contexts it is extremely costly to perform enough high quality experimental measurements to accurately parameterize a predictive quantitative model. However, it is often much easier to carry out large numbers of experiments that indicate whether each sample is above or below a given threshold. Can many such categorical or "coarse" measurements be combined with a much smaller number of high resolution or "fine" measurements to yield accurate models? Here, we demonstrate an intuitive strategy, inspired by statistical physics, wherein the coarse measurements are used to identify the salient features of the data, while the fine measurements determine the relative importance of these features. A linear model is inferred from the fine measurements, augmented by a quadratic term that captures the correlation structure of the coarse data. We illustrate our strategy by considering the problems of predicting the antimalarial potency and aqueous solubility of small organic molecules from their 2D molecular structure.

A large class of scientific questions asks whether dependent variables can be accurately predicted by using training data to learn the parameters of quantitative models. Classical statistics shows that this is possible if sufficiently many high resolution measurements are available, though the cost of performing these experiments can be prohibitive. On the other hand, in many settings, it can be straightforward to evaluate whether a measurement is above or below a certain threshold, raising the question of how such measurements can be incorporated into the modelling framework.

Examples abound in disparate fields. For instance, predicting the solubility of organic molecules is a fundamental challenge in physical chemistry [1]. Although accurate measurements are extremely difficult to obtain [2], determining whether a molecule is soluble at a particular concentration is comparatively simple. Similarly, in drug discovery, biochemical assays that determine whether a molecule binds to a given receptor are much simpler than measuring protein-ligand binding affinity [3]. In protein biophysics, a key challenge is to predict the effect of amino acid changes on protein phenotype. Here, threshold measurements are naturally provided by homologous sequences from the same protein family [4–8]. In contrast, experimentally measuring the phenotypic change is much more difficult. A related problem is to predict the viral fitness landscape given HIV sequences obtained from patients; again collecting patient samples is much easier than measuring fitness directly [9, 10]. In single-cell RNA sequencing, decomposition methods that extract the correlation structure of shallow gene expression measurements is an ongoing challenge [11, 12].

Despite the ubiquity of this problem, to our knowledge there is no principled method for combining numerous binary/categorical ("coarse") measurements with fewer quantitative ("fine") measurements to produce a predictive model. Although regression approaches can account for a prior estimate of sample error [13], this is not the same as combining two qualitatively distinct forms of data to build a more accurate model.

In this Letter, we introduce an intuitive method that combines coarse and fine measurements. The coarse measurements provide sets of labelled samples – those data above and below some threshold – and the proposed method extracts features from the correlations of the variables in each set. Their contribution to the dependent variable is then determined by using the fine measurements to build a regression model for these features. Our model augments a quantitative linear model with a quadratic term which captures the correlation structure extracted from the coarse data. We illustrate our approach by applying it to solubility prediction, and interpret the approach in the context of the Ising model.

To fix ideas, we assume each sample is characterized by a vector of p properties $\mathbf{f}_i \in \mathbb{R}^p$. The binary data indicates that N_+ (N_-) samples are above (below) some threshold. In addition, we are given $\mathbf{y} \in \mathbb{R}^M$, quantitative measurements for M additional samples. These measurements could be binding affinity, solubility etc. We construct matrices $R_\pm \in \mathbb{R}^{N_\pm \times p}$ for samples above/below the threshold, with columns of R_+ and R_- normalized separately to have zero mean and unit vari-

ance. Intuitively, if there are combinations of the p properties that are always present in either sample set, then these properties should be good predictors of the measurement. Such persistent correlations can be identified from the eigendecomposition of each sample covariance matrix $C_\pm$

$$\begin{aligned} C_\pm &= \frac{1}{N_\pm} R_\pm^T R_\pm \\ &= \sum_{i=1}^{N_\pm} \lambda_i^\pm \mathbf{u}_i^\pm \otimes \mathbf{u}_i^\pm, \end{aligned} \tag{1}$$

where $\{\lambda_i^\pm\}$, $\{\mathbf{u}_i^\pm\}$ are the eigenvalues and eigenvectors (note we perform separate eigendecompositions for the two matrices $C_\pm$). Each $\mathbf{u}_i^\pm$ identifies a particular combination of the p properties, explaining a fraction $\lambda_i^\pm / \sum_i \lambda_i^\pm$ of the variance [13]. Each matrix $C_\pm$ is an unbiased estimator of the corresponding true covariance matrix. The quality of this estimator depends on data sampling. For example, one may inadvertently assay certain samples (easy to obtain, measure etc.), which could distort the estimator by causing an eigenvector with large eigenvalue to be localized on features common to these samples, even though they do not predict the output variable. For protein sequences, a natural source of such spurious correlations is phylogeny [14, 15].

Here we propose that whereas the eigenvectors $\mathbf{u}_i^\pm$ reliably identify data features, their significance as estimated by the corresponding eigenvalues $\lambda_i^\pm$ can be severely corrupted by imperfect sampling. Later we justify this ansatz with ideas from statistical physics, and show that this characterization applies to a large class of problems. This ansatz suggests a strategy to mitigate the corruption by using the additional quantitative measurements to determine the significance of each feature. We posit a general quadratic model

$$y_i = \mathbf{h}^T \mathbf{f}_i + \mathbf{f}_i^T J \mathbf{f}_i + \epsilon_i. \tag{2}$$

Here $\mathbf{h}$ is the variable-specific effect, J captures the coupling between variables, and $\epsilon_i \sim N(0, \sigma)$ models random error. There are p parameters in $\mathbf{h}$ and $p(p-1)/2$ parameters in J. If one had $M \gg p(p+1)/2$ quantitative measurements, these parameters could be estimated using linear least squares regression. However, it is costly to perform many detailed measurements, so we turn instead to the matrices $C_\pm$. We pose the ansatz

$$J = \sum_{k=1}^{\hat{p}_+} c_k^+ \mathbf{u}_k^+ \otimes \mathbf{u}_k^+ + \sum_{k=1}^{\hat{p}_-} c_k^- \mathbf{u}_k^- \otimes \mathbf{u}_k^- . \tag{3}$$

Here $\hat{p}_\pm \leq p$, since some eigenvectors will reflect noise due to finite sampling [16–18]. Our ansatz reflects the hypothesis that the eigendecomposition of $C_\pm$ captures variable-variable correlations. If the number of samples is much smaller than the number of variables, random matrix theory provides a rigorous way to determine $\hat{p}_\pm$ [16–21]; this case will be discussed in detail later. Relaxing this assumption, we include all eigenvectors and determine the parameters $\mathbf{h}$, c^+ and c^- by regressing against the the few quantitative measurements available. We note that the ansatz (3) reduces the number of variables to $p + \hat{p}_+ + \hat{p}_-$. In the case where the coarse measurements yield multiple categories (or a single category), our method generalizes by forming separate correlation matrices for each category, and positing that J is a sum of the outer product of all eigenvectors with coefficients determined by regressing against the quantitative data. Our method generalises to Generalised Linear Models with a link function on the right hand side of Equation (2).

To illustrate our approach, we consider two examples: predicting the potency of chemicals against malaria and the equilibrium aqueous solubility of molecules.

Antimalarials – Developing accurate models that can rank the potency of a library of compounds against a target is an important unsolved challenge in drug discovery. We consider a published antimalarial screening campaign [22]: binary but high throughput assays reported 1528 active compounds against malaria, lower throughput but quantitative assays measured the potency (pIC_{50}) of only 1189 compounds [22, 23]. Figure 1A shows that by combining binary and quantitative measurements, an hitherto unattempted strategy, an order of magnitude less quantitative measurements could have been performed to yield a model with similar predictive accuracy (c.f. Supplementary Information showing the same result for the Pearson correlation coefficient). Moreover, the model with coarse measurements clearly outperforms the linear model without the coarse measurements and the "null" quadratic model where the vectors $\mathbf{u}_k^\pm$ are random orthogonal vectors (i.e. random $R_\pm$). In the Supplemental Information we show that our model also outperforms direct quadratic regression. The compounds are described using the 1024-bit Morgan6 Fingerprint [24] generated with `rdKit` [25].

Solubility – Predicting the aqueous solubility of molecules is a fundamental problem in physical chemistry important to a plethora of chemical industries. However, accurate solubility assays are low throughput (~ 1 hour/compound [26]). Figure 1B shows that one could obtain an accurate solubility model ($r^2 = 0.85$, MAE $= 0.61$) if one were to combine the outcome of a coarse solubility assay that could only tell whether a compound is soluble ($< 10^{-4}$ mol/L) or not ($> 10^{-2}$ mol/L) with much fewer quantitive solubility data. We use a standard dataset of the solubility of 1144 organic molecules [27], and describe the molecule by concatenating the Avalon Fingerprint [28], the MACCS Fingerprint [29], and the 1024-bit Morgan6 Fingerprint [24]. Our result compares favourably with other models that also use binary molecular fingerprints, e.g. kernel partial least squares regres-

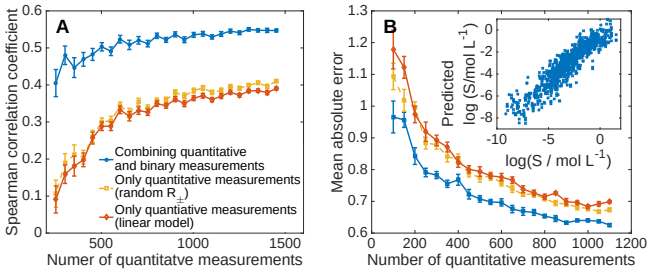

FIG. 1. Combining coarse and fine measurements accurately predicts antimalarial activity and solubility. The predictive accuracy of (A) pIC_{50} against malaria and (B) solubility as a function of the number of quantitative measurements given to the model with coarse measurements (blue line), and without (red line). Including random quadratic terms (orange line) is not effective; error bars obtained over 30 random partitions of data into training (90%) and verification (10%) sets. (Inset) Out-of-sample solubility prediction with 90% of the full dataset has a mean absolute error of 0.61 ($r^2 = 0.85$). The estimate is arrived at by analysing 10 random partitions of the data into training and verification sets.

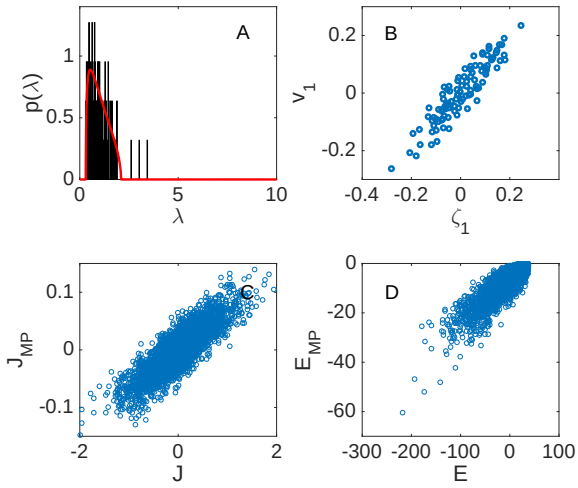

FIG. 2. Hopfield patterns can be recovered from threshold sampling. (A) Histogram of eigenvalues agrees with the Marčenko-Pastur distribution (red curve) save for three significant eigenvalues. (B) The top eigenvector is the lowest energy Hopfield pattern; the other eigenvectors are shown in Supplemental Information. Random matrix cleaning allows us to successfully (C) recover the coupling matrix J and (D) predict Hopfield energies.

sion achieves $r^2 = 0.83$ [30].

To understand why our heuristic strategy is successful, we consider a model problem where data is generated according to Eqn. (2), which is the maximum entropy model [31–33], analogous to the Ising model. We thus interpret the dependent variable as an "energy", noting that the logarithm of the solubility is proportional to the solvation energy. The interaction matrix J can be decomposed into a sum of outer products of eigenvectors $\boldsymbol{\zeta}_i$ (Hopfield patterns [34]), and eigenvalues E_i (Hopfield energies) as

$$J = \sum_{i=1}^{m} E_i \boldsymbol{\zeta}_i \otimes \boldsymbol{\zeta}_i. \tag{4}$$

Furthermore, to model the binary features used in solubility prediction, we make the assumption that the independent variable is a vector of ± 1.

To simulate binary measurements we randomly draw samples from the uniform distribution, evaluate Eqn. (2) to determine the energy of each sample, and retain those samples that fall below a certain energy. Consider an interaction matrix J with $p = 100$, and $m = 3$ randomly generated patterns. To fix ideas, henceforth let all patterns be attractive with $(E_1, E_2, E_3) = (-30, -25, -20)$, $\mathbf{h} = 0$ and $\epsilon_i = 0$. Using this model, we generate 5000 random vectors, consider the $N = 500$ samples with lowest energy as above threshold samples, and compute the eigendecomposition of the resulting correlation matrix.

Figure 2A shows that the eigenvalue distribution of the sample correlation matrix C_+ follows the Marčenko-Pastur distribution expected for a random matrix [35],

$$\rho(\lambda) = \frac{\sqrt{\left[\left(1+\sqrt{\gamma}\right)^2 - \lambda\right]_+ \left[\lambda - \left(1-\sqrt{\gamma}\right)^2\right]_+}}{2\pi\gamma\lambda} \tag{5}$$

where $(\cdot)_+ = \max(\cdot, 0)$, $\gamma = p/N$, with the exception of three distinct eigenvalues. Figure 2B shows that their corresponding eigenvectors are indeed the Hopfield patterns that we put in. Therefore, the large eigenvectors of C_+ correspond to eigenvectors of J. Note that the random matrix theory framework applies because $m \ll p$, i.e. the signal is low rank compared to the noise. If the signal was high rank, all eigenvectors should be included and their significance determined by regression against fine measurements, as in the examples discussed above.

Turning to eigenvalues, in this model, which features *uniform* sampling, we find that the eigenvalues are proportional to the Hopfield energy E_i. This allows us to "clean" the correlation matrix, by using the q eigenvectors above the Marčenko-Pastur threshold to construct a rank q approximation J_{MP} of the correlation matrix (here $q = 3$). Figure 2C shows that J_{MP} accurately reconstructs J, and allows accurate prediction of the energy of particular states (Figure 2D). Analogously, the eigendecomposition of C_- allows one to recover *repulsive* patterns with positive Hopfield energies (see Supplemental Information).

Since the Hopfield patterns are energy minima, taken together Figure 2A-D imply that the probability of the

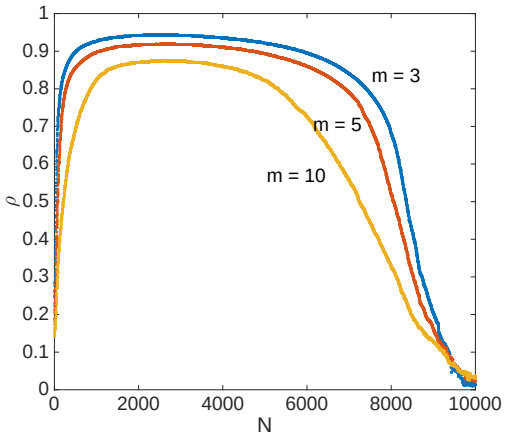

FIG. 3. Hopfield inference with random matrix cleaning is robust to the energy threshold. Here ρ is the Pearson correlation coefficient between the entries in J and J_{MP}. The results are computed by averaging over 50 realizations.

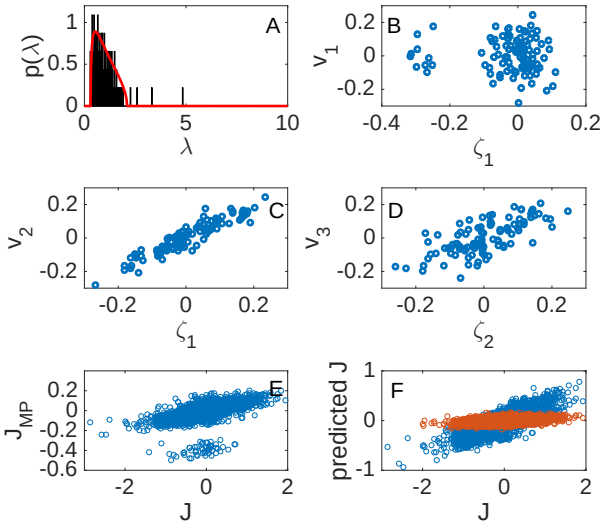

FIG. 4. Few quantitative measurements enable J to be inferred accurately for stratified datasets. (A) There are now four significant eigenvectors but still only three Hopfield patterns in the model. (B)-(D) The top eigenvector is uncorrelated with all Hopfield patterns, and the Hopfield patterns are demoted to the second to fourth significant eigenvectors. (E) Random matrix cleaning does not recover the coupling matrix. (F) Blue data points: The elements of the coupling matrix recovered by incorporating quantitative data and using Eqn. (3); Orange data points: the elements of the coupling matrix recovered using the quantitative data only and ridge regression (with J_{ij} being the coefficients).

system visiting a particular basin under uniform sampling is proportional to the energy of that minima. Therefore, the hypervolume of each energy basin is proportional to the basin depth. This fact can be derived by noting that Eqn. (2) is a quadratic form, so the Hessian matrix is a constant. Therefore, all local minima have the same mean curvature. Given that low lying energy minima are wide, we can extract the position of energy minima in the space of input variables by studying the correlation structure of the binary dataset. We note that the correlation between basin hypervolume and basin depth appears in many complex physical systems beyond the Ising model [36–38].

A lingering question is whether our inference procedure is robust to the choice of threshold. To test this, we consider m Hopfield patterns, chosen as eigenvectors of a symmetrized $p \times p$ Gaussian random matrix, with the Hopfield energy chosen to be Gaussian distributed with mean 10 and unit variance. We draw 10000 samples randomly and compute the correlation matrix with the lowest energy N samples. Figure 3 shows that our method is robust: the correlation coefficient between J and J_{MP} is large and constant for a wide range of thresholds and number of Hopfield patterns. The question of how many energy minima can be recovered from the binary data and a thermodynamic interpretation is discussed in the Supplemental Information.

We now turn to consider two common scenarios that break the assumptions made so far, stratified sampling and more complex energy landscapes.

Stratified sampling: Thus far we assumed that the sampling is uniform before thresholding. However, in many settings, the sampling is biased. To model this effect, we draw 5000 random samples, but freeze the first 5 variables to +1 for the first 2500 samples and the last 5 variables to −1 for the remaining 2500 samples. We then evaluate the energy, and take again the lowest 10^{th} percentile. Figure 4A-D shows that the frozen variables introduce sample-sample correlations, and now there are 4 significant eigenvectors with the first Hopfield pattern demoted to the second largest eigenvector (Figure 4C). As such, the informative eigenvectors are still present in J_{MP}, but the eigenvalues are misplaced.

In this case, naïve random matrix cleaning does not recover J (Figure 4E), since there is no *a priori* reason to discard the first eigenvector unless we know the Hopfield patterns beforehand. We need additional information – for which we turn to the quantitative measurements – to accurately recover the Hopfield energies. Figure 4F shows that an additional 500 quantitative measurement allow us to recover the coupling matrix (MAE = 0.12) using ridge regression and the ansatz Eqn. (3). The error is significantly larger (MAE = 0.31) if only the quantitative measurements are used.

Complex energy landscapes: The geometric property that the depth of an energy minima is related to its hypervolume is not universal to all energy landscapes [39, 40]. A natural question is whether the significant

eigenvectors and eigenvalues of the correlation matrices of samples below/above an energy threshold allow us to infer features of a complex energy landscapes. We consider a landscape that comprises a sum of Gaussians

$$H(\mathbf{f}) = \sum_i E_i \exp\left(-E_i^2(\mathbf{f}\cdot\boldsymbol{\zeta}_i)^2\right). \tag{6}$$

This landscape has the property that the depth of each energy minima, E_i (located at $\boldsymbol{\zeta}_i$), is *inversely* proportional to its width $1/E_i$. As above, we let $(E_1, E_2, E_3) = (-30, -25, -20)$ and generate Hopfield patterns by diagonalising a symmetrized Gaussian random matrix. We draw 5000 samples and threshold to find the 500 lowest energy samples. Figure 5 shows that there are again three significant eigenvectors above the Marčenko-Pastur threshold, but the lowest energy Hopfield pattern is demoted to the third eigenvector, while the highest energy Hopfield pattern is promoted to the top eigenvector. This is expected: the eigenvalue corresponding to each minimum is proportional to the number of samples near that minimum, i.e. the basin volume, which in this case is not proportional to basin depth. However, the eigenvectors still indicate the locations of the energy minima, motivating the approach described in Eqn. (2), where we use these eigenvectors to identify features.

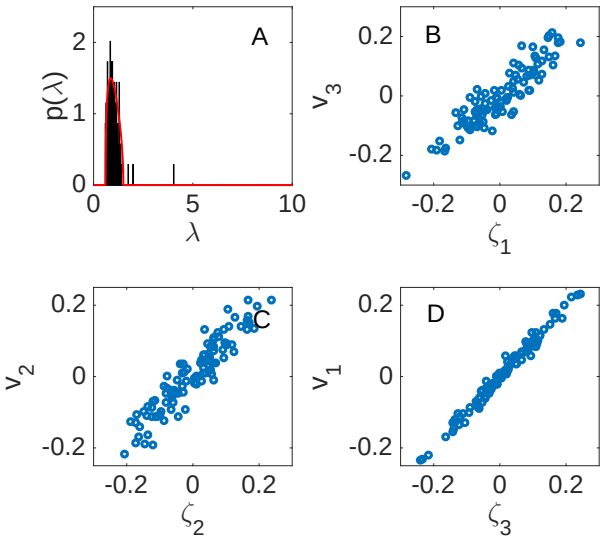

FIG. 5. For energy landscapes where basin depth is not proportional to basin width, the eigenvectors indicate the locations of energy minima but the eigenvalues are awry. (A) There are three significant eigenvectors above the Marčenko-Pastur threshold. (B) - (D) The top eigenvector is correlated with the highest energy minimum, and the last significant eigenvector is correlated with the lowest energy minimum.

In conclusion, we develop a general strategy, grounded in statistical physics, which integrates coarse and fine measurements to yield a predictive model. Since coarse measurements are often significantly less costly to obtain, our strategy provides a new avenue for experiment design. Although our Letter only considered an Ising-type model, the fact that the eigenvectors of the correlation matrix of coarse measurements point toward energy minima suggests a natural way to integrate our result into more complex non-linear models, for example by using $\mathbf{f}\cdot\mathbf{u}_i$, the overlaps between the sample vector and each eigenvector, as inputs to a general nonlinear function such as an artificial neural network.

The authors thank R Monasson for insightful discussions. AAL acknowledges the support of the George F. Carrier Fellowship. LJC acknowledges a Next Generation fellowship, and a Marie Curie CIG [Evo-Couplings, Grant 631609]. MPB is an investigator of the Simons Foundation, and acknowledges support from the National Science Foundation through DMS-1411694.

* aal44@cam.ac.uk
† ljc37@cam.ac.uk

[1] A. Llinas, R. C. Glen, and J. M. Goodman, Journal of Chemical Information and Modeling **48**, 1289 (2008).
[2] D. S. Palmer and J. B. Mitchell, Molecular Pharmaceutics **11**, 2962 (2014).
[3] N. Malo, J. A. Hanley, S. Cerquozzi, J. Pelletier, and R. Nadon, Nature Biotechnology **24**, 167 (2006).
[4] F. Morcos, N. P. Schafer, R. R. Cheng, J. N. Onuchic, and P. G. Wolynes, Proceedings of the National Academy of Sciences **111**, 12408 (2014).
[5] M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, and M. Weigt, Molecular Biology and Evolution **33**, 268 (2015).
[6] T. A. Hopf, J. B. Ingraham, F. J. Poelwijk, M. Springer, C. Sander, and D. S. Marks, Nature Biotechnology (2017).
[7] P. Barrat-Charlaix, M. Figliuzzi, and M. Weigt, Scientific Reports **6** (2016).
[8] R. M. Levy, A. Haldane, and W. F. Flynn, Current Opinion in Structural Biology **43**, 55 (2017).
[9] K. Shekhar, C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, and A. K. Chakraborty, Physical Review E **88**, 062705 (2013).
[10] A. L. Ferguson, J. K. Mann, S. Omarjee, T. Ndungu, B. D. Walker, and A. K. Chakraborty, Immunity **38**, 606 (2013).
[11] A. Dixit, O. Parnas, B. Li, J. Chen, C. P. Fulco, L. Jerby-Arnon, N. D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, *et al.*, Cell **167**, 1853 (2016).
[12] F. Buettner, N. Pratanwanich, J. C. Marioni, and O. Stegle, bioRxiv , 087775 (2016).
[13] C. M. Bishop, *Pattern recognition and Machine Learning*, Vol. 128 (Springer, 2007).
[14] J. Y. Dutheil, Briefings in Bioinformatics **13**, 228 (2012).
[15] B. Obermayer and E. Levine, New Journal of Physics **16**, 123017 (2014).
[16] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Physical Review Letters **83**, 1467 (1999).
[17] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Ama-

ral, and H. E. Stanley, Physical Review Letters **83**, 1471 (1999).
[18] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, Vol. 20 (Springer, 2010).
[19] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, IEEE Transactions on Information Theory **62**, 7475 (2016).
[20] A. A. Lee, M. P. Brenner, and L. J. Colwell, Proceedings of the National Academy of Sciences **113**, 13564 (2016).
[21] J. Bun, J.-P. Bouchaud, and M. Potters, Physics Reports **666**, 1 (2017).
[22] W. A. Guiguemde, A. A. Shelat, D. Bouck, S. Duffy, G. J. Crowther, P. H. Davis, D. C. Smithson, M. Connelly, J. Clark, F. Zhu, *et al.*, Nature **465**, 311 (2010).
[23] S. Riniker, G. A. Landrum, F. Montanari, S. D. Villalba, J. Maier, J. M. Jansen, and W. P. Walters, F1000 Research **6**, 1136 (2017).
[24] D. Rogers and M. Hahn, Journal of Chemical Information and Modeling **50**, 742 (2010).
[25] "RDKit: Open-source cheminformatics," `http://www.rdkit.org`.
[26] K. J. Box, G. Völgyi, E. Baka, M. Stuart, K. Takacs-Novak, and J. E. A. Comer, Journal of Pharmaceutical Sciences **95**, 1298 (2006).
[27] J. S. Delaney, Journal of Chemical Information and Computer Sciences **44**, 1000 (2004).
[28] P. Gedeck, B. Rohde, and C. Bartels, Journal of Chemical Information and Modeling **46**, 1924 (2006).
[29] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, Journal of Chemical Information and Computer Sciences **42**, 1273 (2002).
[30] D. Zhou, Y. Alelyunas, and R. Liu, Journal of chemical information and modeling **48**, 981 (2008).
[31] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006).
[32] S. Cocco, S. Leibler, and R. Monasson, Proceedings of the National Academy of Sciences **106**, 14058 (2009).
[33] E. D. Lee, C. P. Broedersz, and W. Bialek, Journal of Statistical Physics **160**, 275 (2015).
[34] J. J. Hopfield, Proceedings of the National Academy of Sciences **79**, 2554 (1982).
[35] V. A. Marčenko and L. A. Pastur, Mathematics of the USSR-Sbornik **1**, 457 (1967).
[36] J. P. Doye, D. J. Wales, and M. A. Miller, The Journal of Chemical Physics **109**, 8143 (1998).
[37] J. P. Doye and C. P. Massen, The Journal of Chemical Physics **122**, 084105 (2005).
[38] C. J. Pickard and R. Needs, Journal of Physics: Condensed Matter **23**, 053201 (2011).
[39] D. J. Wales, M. A. Miller, and T. R. Walsh, Nature **394**, 758 (1998).
[40] D. Wales, *Energy landscapes: Applications to clusters, biomolecules and glasses* (Cambridge University Press, 2003).