Data-driven diagnosis for compressed sensing with cross validationYoshinori Nakanishi-Ohno^{1,2,3,*} and Koji Hukushima^{1,2,†}¹*Graduate School of Arts and Sciences, The University of Tokyo, Komaba 3-8-1, Meguro, Tokyo 153-8902, Japan*²*Komaba Institute for Science, The University of Tokyo, Komaba 3-8-1, Meguro, Tokyo 153-8902, Japan*³*PRESTO, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan*

(Received 26 January 2018; revised manuscript received 8 September 2018; published 19 November 2018)

We propose a data-driven procedure for diagnosing the results of compressed sensing as success or failure. Compressed sensing is an efficient data-acquisition method widely used in experimental physics. Many previous studies demonstrated that compressed sensing works well if sparse modeling can be applied; that is, physical phenomena of interest are assumed to be described by only a few explanatory factors. However, it is difficult to confirm the assumption of sparse modeling for respective instances in advance because we do not know the true representation or its sparseness of the instance being investigated. To overcome this difficulty, we examined a statistical tool called cross validation, in which all available data are randomly divided into two subsets and the accuracy of describing data in one subset by a sparse representation estimated from only data in the other subset is evaluated as an error. In particular, we focused on the dependence of the cross-validation error on the size ratio of the two subsets. Our analysis, inspired by statistical mechanics, showed that the cross-validation error asymptotically follows a power law of the size ratio when the total amount of available data is exactly at the critical point between success and failure. Hence, compressed-sensing results can be diagnosed in a data-driven manner by comparing the behavior of the cross-validation error with the power law.

DOI: [10.1103/PhysRevE.98.052120](https://doi.org/10.1103/PhysRevE.98.052120)**I. INTRODUCTION**

In today's era of data-driven science, compressed sensing (CS) is becoming increasingly common in experimental physics across disciplines. CS is a statistical method that saves more measurement time and thereby allows experiments to be conducted more efficiently [1,2]. A key idea for finding a correct solution in spite of insufficient data is sparse modeling, in which it is assumed that physical phenomena have a sparse representation; that is, they are described by only a small number of explanatory factors. The research history of CS can be traced back to exploration geophysics in the 1970s [3,4]. Modern theoretical work, in which it was mathematically proved that CS works perfectly under the assumption of sparse modeling [5–8], has expanded the scope of application of CS to magnetic resonance imaging [9,10], ghost imaging [11,12], quantum state tomography [13,14], nuclear magnetic resonance spectroscopy [15,16], and black-hole observation with radio interferometry [17,18]. More recently, the authors and their collaborators applied CS to a quasiparticle interference experiment with scanning tunneling microscopy and scanning tunneling spectroscopy [19].

In this article, we address an oft-overlooked but fundamental problem of detecting false results of CS due to very small datasets. It is well known that CS exhibits a threshold behavior, where if the amount of data is above a critical value it can succeed, and vice versa [5,6]. In statistical mechanics, this phenomenon can be interpreted as a phase transition from failure to success of CS with an increasing amount of data, and its phase diagram and phase boundary have been extensively analyzed [20–25]. The phase boundary indicates the minimum amount of data necessary for the success of CS. In practice, however, no method is available to determine whether the given data are sufficient because the minimum amount of data depends on the degree of sparseness of the phenomena being estimated. Thus, it is important to diagnose the results of CS in a data-driven manner, which does not require any prior knowledge on the true representation or its sparseness of the phenomenon. To develop a data-driven diagnosis method, we examined an orthodox statistical technique called cross validation (CV), in which all available data are divided into training data and validation data, and the accuracy of describing the validation data by a sparse representation estimated from only the training data is evaluated as an error [26,27]. We show that, when the total amount of data is at the critical value between success and failure of CS, the CV error (CVE) asymptotically decreases according to a power law with respect to the ratio of the sizes of the training and validation datasets. We thus propose CV-based diagnostics to assess the results of CS without prior knowledge of the true sparseness of the phenomenon, in which the asymptotic behavior of CVE is monitored and compared with the power law.

*Author to whom all correspondence should be addressed: nakanishi@phys.c.u-tokyo.ac.jp

†hukushima@phys.c.u-tokyo.ac.jp

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

The rest of the article is organized as follows: In Sec. II, two main concepts—compressed sensing and cross validation—are mathematically formulated. In Sec. III, we describe how to analyze cross-validation error using statistical mechanics. In Sec. IV, our method is proposed and its performance is discussed with numerical experiment. Section V concludes the article.

II. FORMULATION

A. Compressed sensing

The purpose of CS is to accurately obtain the true representation of the target phenomenon $\mathbf{x}^0 = (x_1^0, \dots, x_N^0) \in \mathbb{R}^N$, which contains a few nonzero components, from a small set of data samples $Y = \{y_\mu | \mu \in \mathcal{M}\}$, where $\mathcal{M} = \{1, 2, \dots, M\}$ and $M < N$. Each component of the data is acquired in the form of an inner product, $y_\mu = \langle \mathbf{a}_\mu, \mathbf{x}^0 \rangle \equiv \sum_i a_{\mu i} x_i^0$, where $\mathbf{a}_\mu = (a_{\mu 1}, \dots, a_{\mu N}) \in \mathbb{R}^N$ is a measurement vector. The set of measurement vectors is denoted by $A = \{\mathbf{a}_\mu | \mu \in \mathcal{M}\}$. Basis pursuit is a representative algorithm of estimation in noise-free CS [28]. The estimator of basis pursuit $\hat{\mathbf{x}}_{\text{BP}}$ is defined as follows:

$$\hat{\mathbf{x}}_{\text{BP}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to } \langle \mathbf{a}_\mu, \mathbf{x} \rangle = y_\mu \text{ for all } \mu \in \mathcal{M}, \quad (1)$$

where $\|\cdot\|_1$ is called the l_1 -norm of vectors and is defined by $\|\mathbf{x}\|_1 = \sum_i |x_i|$. The optimization problem in Eq. (1) can be solved with linear programming. Note that no unique solution is derived from only the equality constraints because the size of dataset Y is smaller than the dimension of representation $\mathbf{x}$. However, the l_1 -norm minimization leads to the true representation, that is, $\hat{\mathbf{x}}_{\text{BP}} = \mathbf{x}^0$, if it is sufficiently sparse.

Many theoretical studies have been conducted on the performance of CS with basis pursuit. In an influential study, it was proved using geometry that in the large-size limit, $N \rightarrow +\infty$, basis pursuit can obtain the true representation accurately if the ratio of the number of samples to the dimension of representation, that is, $\alpha = M/N$, exceeds a critical value [5,6]. The critical value α_c is given by

$$\alpha_c = \rho + (1 - \rho)\Psi(\theta_c), \quad (2)$$

where ρ represents the degree of sparseness of the true representation, θ_c satisfies

$$\frac{2\phi(\theta_c)}{\theta_c} - \Psi(\theta_c) = \frac{\rho}{1 - \rho}, \quad (3)$$

$\phi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}$, and $\Psi(\theta) = 2 \int_\theta^{+\infty} dz \phi(z)$ in the case of random Gaussian measurement vectors A [29], which will be analyzed later in this study. These results, indicating a phase transition of CS with respect to the amount of data, were discussed and confirmed from the viewpoint of statistical mechanics [20–22]. However, the formula for α_c , Eq. (2), is of little practical use because it depends on ρ . In practical situations, the sparseness of the true representation, ρ , is not known in advance; furthermore, it cannot be judged whether the data already acquired are sufficient for CS.

In other representative work, the performance of CS was analyzed using a mathematical concept called the restricted isometry property [7,30]. For each integer K , the

restricted isometry constant δ_K of a measurement matrix $A = (\mathbf{a}_1^T, \dots, \mathbf{a}_M^T)^T$ is defined as the smallest number such that

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2 \quad (4)$$

holds for all K -sparse vectors $\mathbf{x}$. The transpose of matrix A is denoted by A^T . A vector is said to be K -sparse if it has at most K nonzero components. Using the restricted isometry constant δ_{2K} of A , it has been shown that if $\delta_{2K} < \sqrt{2} - 1$, the K -sparse vectors are guaranteed to be recovered using the basis pursuit [30]. This sufficiency condition has subsequently been improved several times [31–35]. However, these conditions are no more useful than the aforementioned formula for α_c because whether the true representation $\mathbf{x}^0$ is K -sparse is not known in advance.

B. Cross validation

Our proposed method for data-driven diagnosis is based on CV. CV is one of the most frequently used statistical tools for parameter tuning and model selection [26,27]. For example, the regularization parameter for the least absolute shrinkage and selection operator (LASSO) used in noisy CS, which is not a topic tackled in this study, is often determined by applying CV [36]. The procedure of CV is formulated as follows: The dataset Y is divided into a validation dataset $Y_V = \{y_\mu | \mu \in \mathcal{V}\}$ and a training dataset $Y_T = \{y_\mu | \mu \notin \mathcal{V}\}$, where $\mathcal{V}$ is a nonempty proper subset of $\mathcal{M}$. An algorithm such as basis pursuit is then applied to the training dataset Y_T to obtain an estimate of the representation $\hat{\mathbf{x}}_T$. Note that CV is a versatile tool applicable to other estimation algorithms, though only the use of basis pursuit is discussed hereafter. The CVE is then evaluated by comparing $\hat{\mathbf{x}}_T$ with Y_V . In this study, CVE is defined by the mean-squared error as follows:

$$\text{CVE} = \frac{1}{2M_V} \sum_{\mu \in \mathcal{V}} (\langle \mathbf{a}_\mu, \hat{\mathbf{x}}_T \rangle - y_\mu)^2, \quad (5)$$

where M_V is the size of Y_V . Considering the fluctuation in values of the CVE, the CV procedure is repeated many times with different partitions of the same dataset into training and validation subsets. We should stress that the calculation of CVE does not need the true representation or its sparseness. Therefore, CV-based methods satisfy the demand of being data-driven.

III. ANALYSIS

We analyzed the CVE in the large-size limit, $N \rightarrow +\infty$, when basis pursuit is used as the estimation algorithm. The typical value of CVE is expressed as an expected value

$$\text{CVE}_{\text{BP}} = \left[\left\langle \frac{1}{2M_V} \sum_{\mu \in \mathcal{V}} (\langle \mathbf{a}_\mu, \mathbf{x}_T \rangle - y_\mu)^2 \right\rangle_{\mathbf{x}_T} \right]_{Y, A, \mathbf{x}^0}. \quad (6)$$

The angular brackets $\langle \cdot \rangle_{\mathbf{x}_T}$ denote thermal averaging with respect to

$$P_{\text{BP}}(\mathbf{x}_T) \propto e^{-\beta \|\mathbf{x}_T\|_1} \prod_{\mu \notin \mathcal{V}} \delta(\langle \mathbf{a}_\mu, \mathbf{x}_T \rangle - y_\mu), \quad (7)$$

where, compared to the original optimization in Eq. (1), the δ functions represent the equality constraints and the Boltzmann factor represents the l_1 -norm minimization considering the low-temperature limit of $\beta \rightarrow +\infty$. The square

brackets $[\cdot]_{Y,A,\mathbf{x}^0}$ denote configurational averaging with respect to $\{Y, A, \mathbf{x}^0\}$. Our analysis deals with a simple noise-free case in which each component of measurement vectors, $a_{\mu i}$, is an independent and identically distributed Gaussian random variable of mean zero and variance N^{-1} and each component of the true representation, x_i^0 , is independent and identically generated from a sparse prior distribution $P(x_i^0) = \rho \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2\sigma_x^2}(x_i^0)^2} + (1-\rho)\delta(x_i^0)$. The parameters ρ and σ_x^2 represent the degree of sparseness and magnitude of nonzero components, respectively.

For convenience, we introduce a free-energy function,

$$f = -\frac{1}{N\beta} [\ln Z]_{Y,A,\mathbf{x}^0}. \quad (8)$$

The function Z is a partition function defined by

$$Z = \text{Tr}_{\mathbf{x}} e^{-\beta \left\{ \frac{1}{2\lambda} \sum_{\mu \in \mathcal{V}} ((\mathbf{a}_\mu, \mathbf{x}) - y_\mu)^2 + \frac{\gamma}{2} \sum_{\mu \in \mathcal{V}} ((\mathbf{a}_\mu, \mathbf{x}) - y_\mu)^2 \right\}}, \quad (9)$$

where $\text{Tr}_{\mathbf{x}} = \int d\mathbf{x} e^{-\beta \|\mathbf{x}\|_1}$. Then, the CVE is also expressed as a partial derivative of the free energy,

$$\text{CVE}_{\text{BP}} = \lim_{\lambda \rightarrow +0} \lim_{\beta \rightarrow +\infty} \lim_{\gamma \rightarrow +0} \frac{N}{M_V} \frac{\partial f}{\partial \gamma}. \quad (10)$$

It is technical to evaluate $[\ln Z]_{Y,A,\mathbf{x}^0}$ with the help of the replica method developed in the field of spin-glass theory [37]. The replica method is based on an identity called the replica trick:

$$[\ln Z]_{Y,A,\mathbf{x}^0} = \lim_{n \rightarrow +0} \frac{\partial}{\partial n} \ln[Z^n]_{Y,A,\mathbf{x}^0}. \quad (11)$$

The replica trick replaces $[\ln Z]_{Y,A,\mathbf{x}^0}$ with $[\ln[Z^n]_{Y,A,\mathbf{x}^0}]$, and the latter is easier to calculate than the former when n is a positive integer. The value in the limit, $n \rightarrow +0$, is extrapolated from the resultant values at positive integers n . The extrapolation, however, assumes that the functional relation of $[\ln[Z^n]_{Y,A,\mathbf{x}^0}]$ to positive integers n is valid for any real numbers n in the same manner, though it is not always mathematically correct. This is why the replica trick is called a trick.

A lengthy but straightforward calculation, which is explained in Sec. S-I of the supplemental material [38], gives an analytic result of CVE,

$$\text{CVE}_{\text{BP}} = \frac{Q}{2}, \quad (12)$$

and saddle-point equations (SPEs) with respect to $(Q, \chi, \hat{Q}, \hat{\chi})$,

$$\begin{aligned} Q &= \rho \frac{\hat{\chi} + \sigma_x^2 \hat{Q}^2}{\hat{Q}^2} \{ (1 + \theta_1^2) \Psi(\theta_1) - 2\theta_1 \phi(\theta_1) \} \\ &+ (1 - \rho) \frac{\hat{\chi}}{\hat{Q}^2} \{ (1 + \theta_0^2) \Psi(\theta_0) - 2\theta_0 \phi(\theta_0) \} \\ &+ \rho \sigma_x^2 \{ 1 - 2\Psi(\theta_1) \} \end{aligned} \quad (13a)$$

$$\chi = \rho \frac{1}{\hat{Q}} \Psi(\theta_1) + (1 - \rho) \frac{1}{\hat{Q}} \Psi(\theta_0), \quad (13b)$$

$$\hat{Q} = \left(1 - \frac{1}{k} \right) \alpha \frac{1}{\chi}, \quad (13c)$$

$$\hat{\chi} = \left(1 - \frac{1}{k} \right) \alpha \frac{Q}{\chi^2}, \quad (13d)$$

FIG. 1. Typical values of cross-validation error plotted against k when $\rho = 0.1$. The solid line (magenta), indicating an asymptotic power law for large k , represents a critical case, $\alpha = \alpha_c \simeq 0.3288$. Each dashed line (red) having a nonvanishing value at the limit of large k represents a failure case below the criticality, $\alpha = 0.30, 0.31, 0.32$. Each dashed line (blue) exponentially decaying to zero represents a successful case above the criticality, $\alpha = 0.33, 0.34, 0.35$.

where $k = M/M_V$, $\phi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}$, $\Psi(\theta) = 2 \int_{\theta}^{+\infty} dz \phi(z)$, $\theta_1 = (\hat{\chi} + \sigma_x^2 \hat{Q}^2)^{-\frac{1}{2}}$, and $\theta_0 = \hat{\chi}^{-\frac{1}{2}}$. In our analysis, the replica-symmetry ansatz was assumed, and the validity of the analytic result will be confirmed later through numerical analysis.

IV. DISCUSSION

We examined the dependence of the CVE on the parameter k , which is derived from Eqs. (12) and (13). Figure 1 plots CVE values against k in log-log scale. At a glance, regardless of the value of α , a larger k corresponds to a smaller CVE, and this trend is natural because a larger k implies that the training dataset Y_T has a larger size and the estimate $\hat{\mathbf{x}}_T$ in the procedure of CV is thus expected to be closer to the true representation $\mathbf{x}^0$. To develop data-driven diagnostics, the manner in which the CVE decreases with k should be focused on. In the case of $\alpha > \alpha_c$, the CVE exponentially converges to zero with respect to k , and for $\alpha < \alpha_c$ it remains a finite value for large values of k .

We analyzed the critical case of $\alpha = \alpha_c$ in more detail. Considering the limit of $k \rightarrow +\infty$, the solution of the SPEs (13) converges to that of the SPEs from which Eqs. (2) and (3) are derived in [20]. Thus, the leading-order terms of the Taylor expansion of Eq. (13a) at $(\theta_1, \theta_0) = (0, \theta_c)$ are relevant to obtain an asymptotic formula of CVE:

$$\text{CVE}_{\text{BP}} \simeq \frac{\alpha_c \sigma_x^2}{2\theta_c^2} \theta_1^2. \quad (14)$$

A similar calculation for Eqs. (13a) and (13b) after substitution of Eqs. (13c) and (13d) shows that θ_1 is asymptotically proportional to $1/k$. Therefore, the CVE has a power-law relation to k with the exponent -2 when the amount of data is on the phase boundary, $\alpha = \alpha_c$. This indicates that the results of CS can be diagnosed as success or failure by applying CV at various k values and comparing the decreasing trend of CVE with the power law.

FIG. 2. Plots of numerical results for $N = 100, 200, \dots, 1000$ and the extrapolated value of $N \rightarrow +\infty$. The cross on the vertical axis represents the theoretical value derived from the replica-symmetry ansatz. The parameters are set to $\rho = 0.1$, $\sigma_x^2 = 1$, $\alpha = 0.3$, and $k = 5, 6, 7.5, 10, 15$, and 30 from (a) to (f).

As mentioned above, we performed numerical analysis to confirm the theoretical results from the replica-symmetry ansatz. The numerical results are shown in Figs. 2 and 3. According to the settings, we generated numerous triplets of

$\{Y, A, x^0\}$ and calculated CVE by applying basis pursuit to $\{Y, A\}$. The median over 5000 instances of $\{Y, A, x^0\}$ was plotted to take account of the expectation denoted by $[\cdot]_{Y,A,x^0}$. Extrapolation to the large-size limit was performed by

FIG. 3. Plots of numerical results for $N = 100, 200, \dots, 1000$ and the extrapolated value of $N \rightarrow +\infty$. The cross on the vertical axis represents the theoretical value derived from the replica-symmetry ansatz. The parameters are set to $\rho = 0.1$, $\sigma_x^2 = 1$, $\alpha = 0.35$, and $k = 5, 5.83, 7, 8.75, 11.7$, and 17.5 from (a) to (f).

FIG. 4. Results of numerical experiments on two instances are shown in the top and bottom rows, which correspond to success and failure cases of compressed sensing, respectively. In the left, filled (green) and open (gray) circles indicate observed and unobserved data, respectively. In the middle, estimated and true sparse representations are shown by filled (red) and open (black) circles, respectively. The overlapping of corresponding pairs of filled and open circles of the top instance indicates a perfectly accurate estimation. Basis pursuit is applied for estimation from the samples. In the right, CVE values are plotted against k in log-log scale. In both instances, the parameters are set to $N = 300$, $\rho = 0.1$, $\sigma_x^2 = 1$, and $\alpha = 0.3$.

linear regression using a simple model, $\text{CVE} \simeq \beta_0 + \beta_1 N^{-1} + \beta_2 N^{-2}$. The linear regression was performed by using the least-squares method. Error bars represent the standard error evaluated using the bootstrap method. We can see that the analytic solution is close to the extrapolated value. Therefore, the results of our analysis with the replica-symmetry ansatz can be considered valid. We conducted numerical experiments to show that our proposed method works well in practical cases of finite N for each instance, while theoretical results by the replica method are obtained for a typical case of infinite N averaged over instances. The results on two instances of $\{Y, A, \mathbf{x}^0\}$ with parameters $N = 300$, $\rho = 0.1$, $\sigma_x^2 = 1$, and $\alpha = 0.3$ are shown in Fig. 4. Each left panel shows a set of data samples Y , from which a sparse representation is estimated using basis pursuit. Judging from the results of CS shown in the middle panels, the top and bottom instances represent success and failure cases, respectively. We need to determine whether our proposed CV-based method can distinguish between success and failure properly in a data-driven manner. According to our proposal, CVE values, which are calculated from only the data, are plotted against k in log-log scale in the right panels. At a glance, the top instance indicates a faster exponential decay of CVE compared to the power law with the exponent -2 , which is consistent with the success of CS. In contrast, as shown in the bottom instance of failure, the CVE remains a finite value at the largest k that can be examined when $M = 90$.

In Sec. S-II of the supplemental material [38], we show the results of other 200 instances with various system sizes from $N = 100$ to 1000. Regarding small-size systems, a few instances were found difficult to diagnose by our method because it is ambiguous as to whether the convergence of

CVE is faster or slower than the power law. Practically in such a case, we should regard the results of CS as failure, and acquire more data so as not to draw a hasty conclusion. In addition, we see that the results of diagnosis were not always consistent with the true results of CS. A closer look reveals that all the inconsistent instances were diagnosed as failure by our method, though CS itself succeeded. Our method is considered to make a conservative diagnosis because in CV, a part of the available data needs to be kept aside for validation data. Conversely, we argue that our method never failed to detect hidden failure of CS and prevents us from a meaningless or misleading discussion derived from insufficient data. As a whole, our method worked well, and in particular, all of the instances with $N \geq 800$ were correctly diagnosed as far as we examined. Therefore, we conclude that our proposed CV-based method performs well in data-driven diagnosis for CS.

V. CONCLUSION

We have proposed a data-driven framework based on CV to diagnose the results of CS as success or failure even before the truth is recognized. The framework notifies us when the amount of data has just exceeded a critical threshold and helps improve experimental design to maximize the utilization of experimental resources. In addition, our framework is expected to provide a criterion based on which an algorithm should be chosen in CS among numerous algorithms such as basis pursuit, orthogonal matching pursuit [39,40], iterative hard thresholding [41,42], and subspace pursuit [43]. As is often the case, different results are obtained using different algorithms even when the same dataset is analyzed, but our framework enables us to distinguish which results can be trusted by evaluating the relation of CVE to the

parameter k for each algorithm [44]. Moreover, our framework is considered widely applicable to many other formulations of CS. A representative formulation is matrix completion [45,46], which is useful for quantum state tomography in quantum information science [13,14] and phase retrieval in x-ray crystallography [47]. Considering computational cost, cross validation intensively used in our proposed method is time-consuming because it involves many repetitions of estimation from various training datasets. Future work on developing efficient methods of cross validation is needed and the approach of generalized cross validation would be promising [36,48–50]. Statistical methodologies such as CS

deserve further discussion from the viewpoint of statistical mechanics including phase transition and critical phenomena. It will be important not only to point out the existence of a phase transition in information processing but also to establish a practical framework of data-driven diagnosis.

ACKNOWLEDGMENTS

We wish to acknowledge the support of JSPS KAKENHI Grants No. JP17K12749 (Y.N.-O.) and No. JP25120010 (K.H.), and JST PRESTO Grant No. JPMJPR1773 (Y.N.-O.).

- [1] D. L. Donoho, *IEEE Trans. Inf. Theor.* **52**, 1289 (2006).
- [2] E. J. Candes and M. B. Wakin, *IEEE Signal Proc. Mag.* **25**, 21 (2008).
- [3] J. F. Claerbout and F. Muir, *Geophysics* **38**, 826 (1973).
- [4] H. L. Taylor, S. C. Banks, and J. F. McCoy, *Geophysics* **44**, 39 (1979).
- [5] D. L. Donoho and J. Tanner, *Proc. Natl. Acad. Sci. USA* **102**, 9446 (2005).
- [6] D. L. Donoho and J. Tanner, *Proc. Natl. Acad. Sci. USA* **102**, 9452 (2005).
- [7] E. J. Candes and T. Tao, *IEEE Trans. Inf. Theor.* **51**, 4203 (2005).
- [8] E. J. Candes, J. Romberg, and T. Tao, *IEEE Trans. Inf. Theor.* **52**, 489 (2006).
- [9] M. Lustig, D. Donoho, and J. M. Pauly, *Magn. Reson. Med.* **58**, 1182 (2007).
- [10] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, *IEEE Signal Proc. Mag.* **25**, 72 (2008).
- [11] O. Katz, Y. Bromberg, and Y. Silberberg, *Appl. Phys. Lett.* **95**, 131110 (2009).
- [12] C. Zhao, W. Gong, M. Chen, E. Li, H. Wang, W. Xu, and S. Han, *Appl. Phys. Lett.* **101**, 141123 (2012).
- [13] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Phys. Rev. Lett.* **105**, 150401 (2010).
- [14] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, *New J. Phys.* **14**, 095022 (2012).
- [15] K. Kazimierczuk and V. Y. Orekhov, *Angew. Chem. Int. Ed.* **50**, 5556 (2011).
- [16] D. J. Holland, M. J. Bostock, L. F. Gladden, and D. Nietlispach, *Angew. Chem. Int. Ed.* **50**, 6548 (2011).
- [17] M. Honma, K. Akiyama, M. Uemura, and S. Ikeda, *Publ. Astron. Soc. Jpn.* **66**, 95 (2014).
- [18] S. Ikeda, F. Tazaki, K. Akiyama, K. Hada, and M. Honma, *Publ. Astron. Soc. Jpn.* **68**, 45 (2016).
- [19] Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa, and M. Okada, *J. Phys. Soc. Jpn.* **85**, 093702 (2016).
- [20] Y. Kabashima, T. Wadayama, and T. Tanaka, *J. Stat. Mech. Theor. Exp.* (2009) L09003.
- [21] D. L. Donoho, A. Maleki, and A. Montanari, *Proc. Natl. Acad. Sci. USA* **106**, 18914 (2009).
- [22] S. Ganguli and H. Sompolsky, *Phys. Rev. Lett.* **104**, 188701 (2010).
- [23] K. Takeda and Y. Kabashima, *Europhys. Lett.* **95**, 18006 (2011).
- [24] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, *Phys. Rev. X* **2**, 021005 (2012).
- [25] L. Zdeborová and F. Krzakala, *Adv. Phys.* **65**, 453 (2016).
- [26] B. Efron and G. Gong, *The American Statistician* **37**, 36 (1983).
- [27] S. Arlot and A. Celisse, *Stat. Surv.* **4**, 40 (2010).
- [28] S. Chen and D. Donoho, in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers* (IEEE, Piscataway, NJ, 1994), p. 41.
- [29] D. L. Donoho, *Discrete Comput. Geom.* **35**, 617 (2006).
- [30] E. J. Candes, *C. R. Math.* **346**, 589 (2008).
- [31] S. Foucart and M.-J. Lai, *Appl. Comput. Harmon. Anal.* **26**, 395 (2009).
- [32] T. T. Cai, L. Wang, and G. Xu, *IEEE Trans. Signal Proc.* **58**, 1300 (2010).
- [33] S. Foucart, *Appl. Comput. Harmon. Anal.* **29**, 97 (2010).
- [34] Q. Mo and S. Li, *Appl. Comput. Harmon. Anal.* **31**, 460 (2011).
- [35] J. Andersson and J.-O. Stromberg, *IEEE Trans. Inf. Theor.* **60**, 1700 (2014).
- [36] R. Tibshirani, *J. R. Statist. Soc. Series B (Methodol.)* **58**, 267 (1996).
- [37] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Clarendon, Oxford, England, UK, 2001), Vol. 111.
- [38] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.98.052120> for a supplementary explanation of the theoretical analysis of cross-validation error, and for the results of a supplementary experiment.
- [39] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers* (IEEE, Piscataway, NJ, 1993), p. 40.
- [40] J. A. Tropp, *IEEE Trans. Inf. Theor.* **50**, 2231 (2004).
- [41] T. Blumensath and M. E. Davies, *J. Fourier Anal. Appl.* **14**, 629 (2008).
- [42] T. Blumensath and M. E. Davies, *Appl. Comput. Harmon. Anal.* **27**, 265 (2009).
- [43] W. Dai and O. Milenkovic, *IEEE Trans. Inf. Theor.* **55**, 2230 (2009).
- [44] Y. Nakanishi-Ohno and K. Hukushima, *J. Phys. Conf. Ser.* **1036**, 012014 (2018).
- [45] E. J. Candès and B. Recht, *Found. Comput. Math.* **9**, 717 (2009).
- [46] C. Schülke, P. Schniter, and L. Zdeborová, *Phys. Rev. E* **94**, 062136 (2016).
- [47] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *SIAM Rev.* **57**, 225 (2015).
- [48] P. Craven and G. Wahba, *Numer. Math.* **31**, 377 (1978).
- [49] G. H. Golub, M. Heath, and G. Wahba, *Technometrics* **21**, 215 (1979).
- [50] T. Obuchi and Y. Kabashima, *J. Stat. Mech. Theory Exp.* **2016**, 053304 (2016).