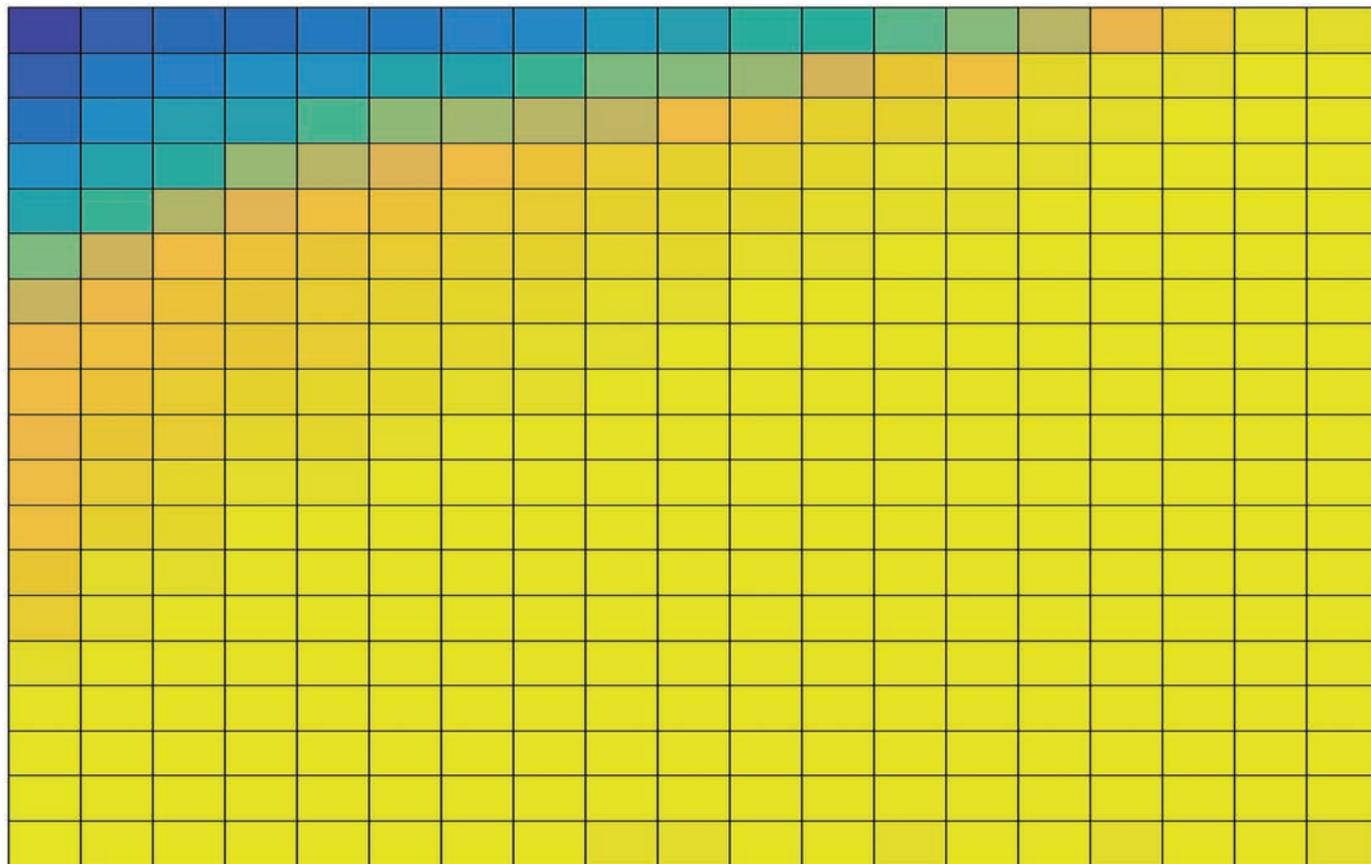

Large Data Analysis and Lyme Disease



Deanna Needell

Introduction

Recent advances in technology have led to a monumental increase in large-scale data across many platforms. One may think that more data means more *information*, but the large-scale nature of modern data actually ends up choking classical analytical methods, making information extraction more challenging than ever before. One mathematical model that has gained a lot of recent attention is the use of sparsity. Sparsity captures the idea that high-dimensional signals often contain a very small amount of

intrinsic information. Using this notion, one may design efficient low-dimensional representations of large-scale data as well as robust reconstruction methods for those representations. Moreover, in many applications one does not desire to reconstruct the full signal but rather perform some data analysis task such as classification, clustering, parameter estimation, and feature selection.

Organization. In this article, we will discuss some key mathematical ingredients that can be used for large-scale data analysis. We begin in the section “Signal Reconstruction” with a background to compressive signal processing (CSP), which is used as a foundation and motivation for more recent work in large-scale data analysis. In the section “1-Bit CSP” we discuss the extension of CSP to the

Deanna Needell is a professor of mathematics at the University of California, Los Angeles. Her email address is deanna@math.ucla.edu.

Communicated by Notices Associate Editor Noah Simon.

*For permission to reprint this article, please contact:
reprint-permission@ams.org.*

DOI: <https://doi.org/10.1090/noti/1760>

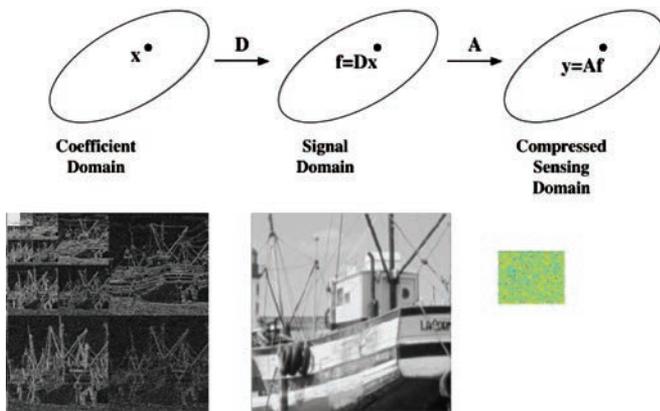


Figure 1. Figure 1. Summary of CSP process.

1-bit version, which utilizes only a single bit per measurement. We relate CSP to the problem of data completion in the section “Data Completion.” Finally, the section “Drawing Conclusions from Large-Scale Data” provides an introduction to classification as a simple problem that is motivated by 1-bit CSP. To showcase important challenges with real-world large-scale data, we will use recently acquired Lyme data as a running example throughout. Lyme disease data is especially challenging because of its high variation in patient symptomatology, diagnosis, and treatment, which leads to large-scale but highly incomplete and varied data.

Signal Reconstruction

In many applications, information about a particular object is acquired in a process called *sampling*. Sampling is used in medical imaging, analog-to-digital converters, and radar, just to name a few examples. The data objects that we wish to sample are called *signals* and could be audio signals, images, data files, or any other type of data. Many signals of interest contain far less information than their ambient dimension implies, and such signals are called *compressible*. Compressive signal processing (CSP) is a recent technology [7, 9] that shows these signals can be accurately represented with far fewer samples than traditionally thought. CSP uses the idea of dimension reduction, which reduces the size of compressible signals while still preserving most of their information. There are certain random linear projections that, when applied, provide this type of dimension reduction. Since the problem of recovering a signal from its compressed form is in general ill-posed (since the resulting mathematical system is highly underdetermined), it is nontrivial to reconstruct a signal from its highly underdetermined measurements; we need sophisticated methods to reconstruct the signal from these samples.

Sparsity Model. We will denote our signal of interest by $f \in \mathbb{C}^n$. We say that f is s -sparse when f has at most s nonzero entries, written $\|f\|_0 \leq s$. Sparsity plays an important role in CSP because compressible signals are those which are approximated well by sparse signals. In general, a signal can be sparse in this sense or with respect to some orthonormal or overcomplete basis, in which case $f = Dx$ for some matrix D and sparse vector x . Given a compressible signal, we acquire samples by applying the sampling matrix A . The sample vector can then be written as $y = Af + e$, where e is an arbitrary noise vector. The CSP problem is to reconstruct an arbitrary compressible signal f from these noisy samples using a tractable algorithm. That is, given knowledge of the measurements y , the measurement matrix A , and the sparsifying dictionary D , one wishes to (approximately) reconstruct the signal vector f . Typically one may assume the sparsity level s is approximately known, although not all methods require such knowledge. The overall process is visualized in Figure 1.

Compressive Signal Processing. Two major approaches for recovery emerged as work in CSP developed. The first method solves an ℓ_1 -minimization program to recover the signal, which, when D is the identity, is simply $\operatorname{argmin}_x \|x\|_1$ s.t. $\|Ax - y\|_2 \leq \epsilon$, where ϵ is the noise tolerance. To analyze this problem, Candès and Tao [6] introduced the restricted isometry property (RIP), which requires the matrix A to be nearly orthonormal on sparse signals [5]. It is now well known that many random $m \times n$ matrices satisfy the RIP with high probability for an optimal number of samples $m \approx s \log n$. Candès et al. show that under the RIP assumption, the ℓ_1 -minimization program robustly and accurately recovers sparse signals. The second approach utilizes greedy algorithms that reconstruct the signal iteratively. Greedy algorithms such as IHT and CoSaMP provide both fast runtime and the same optimal recovery guarantees as the optimization-based approach.

1-Bit CSP

Recovering a signal x from its highly underdetermined image $y = Ax$ is already mathematically challenging enough. However, we now take this problem a step further—we ask to keep only the *first bit* of each entry in y ! One-bit compressed sensing is a new branch of CSP that considers the extreme case when each measurement is quantized to a single bit [2].

In 1-bit CSP, the measurements y are of the form $y_i = \operatorname{sign}(\langle a_i, x \rangle)$, where a_i is the i th row of the measurement matrix A , and the sign function returns -1 if the input is negative and 1 otherwise. This gives a 1-bit (per entry) representation of the measurement vector y . Noise

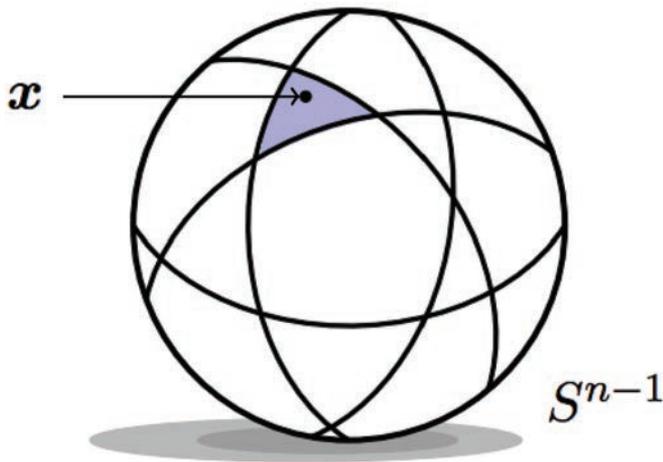


Figure 2. Geometry of 1-bit CSP.

may be added either before or after quantization, the latter type of noise corresponding to a bit-flip. One observation is immediate—there is no way one can ever hope to reconstruct the magnitude of x from such nonlinear measurements. Geometrically, each row of A corresponds to a hyperplane, and the binary measurements y_i simply encode on which side of the hyperplane x lies. For example, given the norm of x , each measurement divides up the sphere of that radius as in Figure 2. Using this geometry, one can obtain lower bounds on the best possible reconstruction error. Indeed, with this 1-bit model, the best reconstruction \hat{x} from y still must satisfy $\|x - \hat{x}\|_2 \gtrsim \frac{1}{\lambda}$, where $\lambda = m/(s \log(n/s))$ denotes the oversampling factor (and recall n and m are the original data dimension and compressed dimension, respectively).

However, all hope is not lost. Both issues (the loss of magnitude information and the lower bound on the recovery rate) can be resolved by using *dithers*, purposefully adding noise to the measurements prior to quantization: $y_i = \text{sign}(Ax + \tau_i)$, where τ_i is some cleverly chosen scalar noise. Dithers are often used in the theory and practice of analog-to-digital conversion, and one sees that in our case, the value of τ_i simply acts as a new threshold on which to quantize—i.e., y_i is now ± 1 depending on whether it is larger or smaller than τ_i . Geometrically, one envisions this as “jiggling” the hyperplanes in such a way as to reveal the norm of the signal x . In addition, Baraniuk et al. [1] and Knudson et al. [10] develop both optimization-based and iterative methods that provably provide an exponentially decaying error guarantee of the form $\|x - \hat{x}\|_2 \lesssim \exp(-\Omega(\lambda))$. In these works, the dithers τ_i are chosen adaptively or set to be i.i.d. Gaussian, depending on the method used for reconstruction.

Data Completion

Although most data is now extremely large-scale, it is also highly *incomplete*. State-of-the-art methods like *matrix completion* allow one to accurately complete a data matrix from observing only a few of its entries, under the assumption that the data is intrinsically low-rank. Generically, the problem is formulated as recovering a matrix M from linear measurements of the form $y_i = \langle A_i, M \rangle := \text{trace}(A_i^* M)$ for $i = 1 \dots m$, where A_i are matrices of the same dimension as M and where A^* denotes the adjoint or conjugate transpose of a matrix A . For example, when A_i are matrices of all zeros and a single 1, this corresponds to the classical matrix completion problem when only a subset of entries are observed. In many applications, the underlying complete matrix is approximately low-rank; for example, in collaborative filtering applications like the now-famous Netflix ratings problem, user preferences are accurately described by a small number of variables, and the full set of theoretical ratings is thus very low-rank. One then seeks a low-rank matrix M consistent with these measurements. Since rank minimization is not computationally tractable, one may solve its semi-definite relaxation: $\min_M \|M\|_*$ s.t. $\langle A_i, M \rangle = y_i \forall i$, where $\|M\|_* := \text{trace}(\sqrt{M^* M})$ denotes the nuclear-norm. It is now well known that when m is on the order of nr , nuclear-norm minimization accurately recovers any rank- r $n \times n$ matrix (from, e.g., Gaussian measurements or uniformly at random chosen observations).

Lyme Data. Lyme disease is the most common vector-borne disease in the United States, with CDC estimates of over 300,000 people in the US diagnosed each year. A significant proportion of patients with Lyme disease develop chronic debilitating symptoms, often mimicking other illnesses such as multiple sclerosis (MS) and ALS. Founded over twenty-five years ago, LymeDisease.org (LDo) is a national nonprofit dedicated to advocacy, research, and education. In November 2015, LDo announced the launch of MyLymeData, a patient-powered research project. This data has over 10,000 patients enrolled, includes several phases of initial and follow-up survey responses, and asks patients questions about diagnosis, treatment, symptoms, and quality of life. Like many large-scale surveys, this data is noisy, incomplete, and has a tree-like structure that makes it mathematically challenging. The data used in this work was obtained from the LymeDisease.org patient registry, MyLymeData, Phase 1, June 17, 2017.

Lyme data, like most real-world data, has missing entries that are not random, much less uniformly selected at random. Still, one would like to complete such data matrices so that accurate conclusions can be drawn from the

complete data, or so that future surveys can be purposefully designed to have missing data (i.e., by not asking every patient every question) to reduce survey fatigue. For example, in the Lyme data, entries may be missing because the user chose not to answer, the survey structure deemed the question irrelevant (i.e., branching), there were corruptions/errors in acquisition, or the survey was designed to ask each patient only a subset of questions. Lee and Shraibman [11] and Foucart et al. [8] recently obtained results (using other methods) exposing recovery guarantees that depend on appropriate parameters of the observation pattern. For example, the latter group analyzed a low-rank projection with de-biasing scheme using a weight matrix W depending on the observation pattern Ω ,

$$\hat{M} = \operatorname{argmin}_{\operatorname{rank} \hat{M} \leq r} \|M_\Omega - W \odot \hat{M}\|^2, \quad (1)$$

where \odot denotes the Hadamard (entrywise) product. Note that (1) can be solved very easily by simply using a truncated singular value decomposition (SVD) of M_Ω to find its best rank- r approximation and then multiplying the result entrywise by the reciprocals of the entries in the weight matrix W (assuming they are all nonzero). In other words, the recovered matrix \hat{M} is taken to be the rank- r matrix that best matches the observed entries M_Ω , after an appropriate rescaling of the entries. Without such a rescaling, when the sampling pattern is far from uniform, the reconstruction \hat{M} turns out to be very biased, so parts of the matrix need to be scaled by different weights. They also analyze a max-norm ($\|X\|_{\max} := \min_{X=UV^*} \|U\|_{2,\infty} \cdot \|V\|_{2,\infty}$) projection scheme that tolerates arbitrary non-uniform sampling. Both methods can be applied to matrices with *arbitrary* sampling patterns; of course, the methods unsurprisingly cannot guarantee accurate completion when the patterns are far from uniform. The results of the former method on the Lyme data are shown in Figure 3. The left plot shows light areas for observed entries and black for unobserved entries; in particular, note that the observation pattern is far from uniform. The middle and right plots show the actual (coded) survey values for the full data matrix (with missing entries) and the completed matrix. Without ground truth for this real data, we cannot provide errors for the completed matrix, but one can validate the method using other statistical techniques.

Structured Observations. Looking again at Figure 3 (left), one immediately notices that the sampling pattern is far from uniform. Indeed, in this type of data and many other applications, the sampling pattern may be deterministic and/or far from uniform. In the Lyme disease survey, missing data arose from patients skipping questions either by survey design or by patient choice. In collaborative filtering, most users do not rate most items whereas a small number of so-called “super-users” rate a large number of

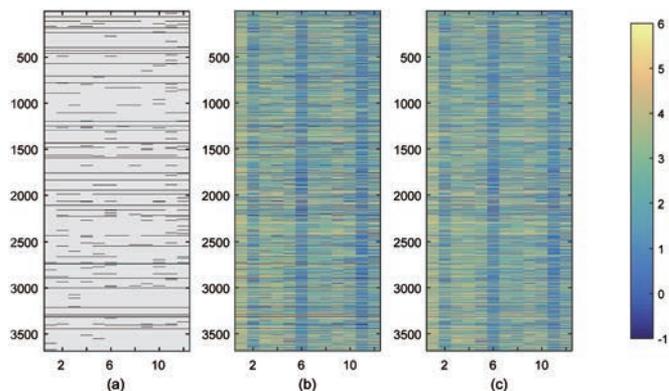


Figure 3. Figure 3. Simple method (1) on the full (incomplete) Lyme symptom data. The actual sampling pattern from the survey for these questions is shown in (a). The full matrix is shown in (b) and the reconstruction in (c).

items. In fact, not only are these sampling patterns non-random, the pattern itself yields a lot of untapped information. In both of these examples, most of the unobserved entries are likely due to participant disinterest or irrelevance. This information is useful! Motivated by this setting, we consider a nuclear-norm matrix completion program with an added regularizer that promotes unobserved entries to have small values.

Let $M \in \mathbb{R}^{n_1 \times n_2}$ be the unknown matrix we would like to recover, and let Ω be the set of indices of the observed entries. Let $\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$, where

$$[\mathcal{P}_\Omega]_{ij} = \begin{cases} M_{ij} & (i, j) \in \Omega \\ 0 & (i, j) \notin \Omega \end{cases}$$

as in [3].

Recall the nuclear-norm minimization,

$$\hat{M} = \operatorname{argmin}_A \|A\|_* \text{ s.t. } \mathcal{P}_\Omega(A) = \mathcal{P}_\Omega(M). \quad (2)$$

Motivated by applications in which the unobserved entries tend to have small values, we instead solve

$$\begin{aligned} \tilde{M} = \operatorname{argmin}_A \|A\|_* + \alpha \|\mathcal{P}_\Omega^c(A)\|_1 \\ \text{s.t. } \mathcal{P}_\Omega(A) = \mathcal{P}_\Omega(M), \end{aligned} \quad (3)$$

where $\alpha > 0$ and the entrywise L_1 -norm $\|M\|_1 = \sum_{ij} |M_{ij}|$.

From the Lyme data, we consider a subset of 2,126 patients responding to sixty-five particular questions that gives a fully complete matrix (so we have ground truth for testing purposes). Question responses are integer values between zero and four. Due to computational constraints, for each of the ten trials executed, we randomly sample fifty of these patient surveys to generate a 50 x 65 matrix. We then subsample from the zero and nonzero entries of the data matrix at various rates to generate a matrix

with missing entries. We complete this subsampled matrix with both (2) and (3) using L_1 regularization on the unobserved entries and report $\|\tilde{M} - M\|_F / \|\hat{M} - M\|_F$, averaged over ten trials in Figure 4. The parameter α , for the regularization term, is chosen to be optimal among $\alpha \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. We see, as expected, that when most of the unobserved entries are small and most of the observed entries are large, there is the most improvement in using the regularizer. Preliminary theoretical results can be found in [12], which are motivated by work in robust principal component analysis [4], but future work is needed to clearly quantify the theoretical gains as a function of the sampling rates.

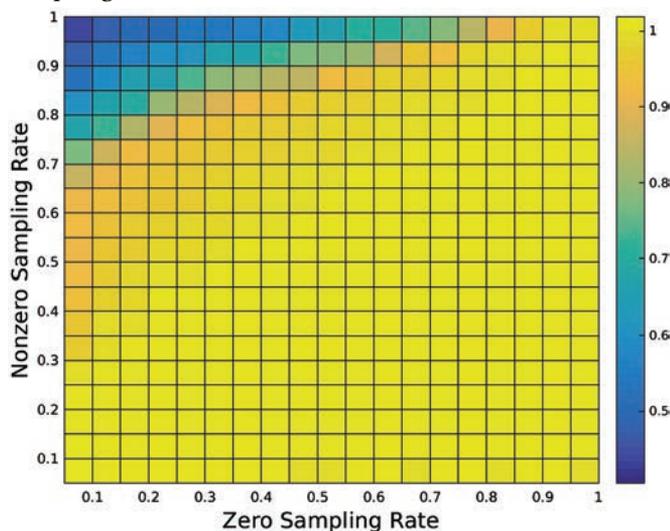


Figure 4. Figure 4. For \tilde{M} and \hat{M} given by (3) and (2), respectively, with L_1 regularization on the recovered values for the unobserved entries, we plot $\|\tilde{M} - M\|_F / \|\hat{M} - M\|_F$. We consider fifty patient surveys with sixty-five responses each chosen randomly from 2,126 patient surveys. We average results over ten trials and with α optimal among $\alpha \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$.

Drawing Conclusions from Large-Scale Data

Background. In many high-dimensional data applications, one is not interested in just representing the data efficiently or completing missing data, but also in drawing analytical conclusions from that data. Such analytical techniques can be performed on the large-scale data itself or from the efficient representations. Although there are many important data analysis problems, and many sophisticated methods to address them, we focus here on the problem of classification and discuss a simple method motivated by 1-bit CSP that exhibits extremely efficient computation and storage as well as interpretability and mathematical theoretical support.

Classification. Given labeled training data, the classification problem asks to accurately label new unlabeled data. Classification is a canonical problem considered in many areas of statistics and machine learning. Classical approaches include support vector machines (SVM) and logistic regression, and recently there has been a surge in so-called deep learning methods, which rely on learned hidden features through a neural network design. Here, we describe a recent approach put forth by Needell et al. [13] that performs simple and efficient classification from binary data. It will use random hyperplanes and the corresponding *binary sign information* that provides information about which side of each hyperplane each data point lies. For a new query point, this information will be compared to the same information for all training data, and then a label that best matches the new query point will be assigned. Although motivated by other approaches like SVM and deep learning, this approach is not very comparable, since, e.g., SVM *searches* for a separating hyperplane as a rule, whereas this approach will utilize and aggregate information from many random hyperplanes.

Let us build some intuition for the approach. Consider the two-dimensional data X shown in the top plot of Figure 5, consisting of three labeled classes (green, blue, red). Consider the four hyperplanes shown in the same plot, and suppose we had access only to the binary data $Q = \text{sign}(AX)$, where A contains the normals to each hyperplane as its rows. For the new test point x (which by visual inspection should be labeled blue) and its binary data $q = \text{sign}(Ax)$, one could simply cycle through the hyperplanes and decide which class x matches most often. For example, for the hyperplane colored purple in the plot, x has the same sign (i.e., lies on the same side) as the blue and green classes. For the black hyperplane, x only matches the blue class, and so on. Then for this example, x will clearly match the blue class most often, and we could assign it that label correctly. However, next consider the more complicated geometry given in the bottom plot, where the data consists of only two classes (red and blue), but they are now intermixed. This same strategy will no longer be accurate for the test point x . However, now instead of single hyperplanes, consider hyperplane *pairs*, and ask which class label x most often matches (note that in this context, by “matches” we now mean that points lie in the same cone that the hyperplanes divide the space into). For example, for the pair colored orange and green, x matches both red and blue points, whereas for the pair of hyperplanes colored orange and purple, x only matches the blue class. One could now cycle through all pairs and again ask which class x matches most often.

Let us now describe the approach more formally. Consider a data matrix $X \in \mathbb{R}^{d \times p}$. Let $A \in \mathbb{R}^{m \times d}$ have rows

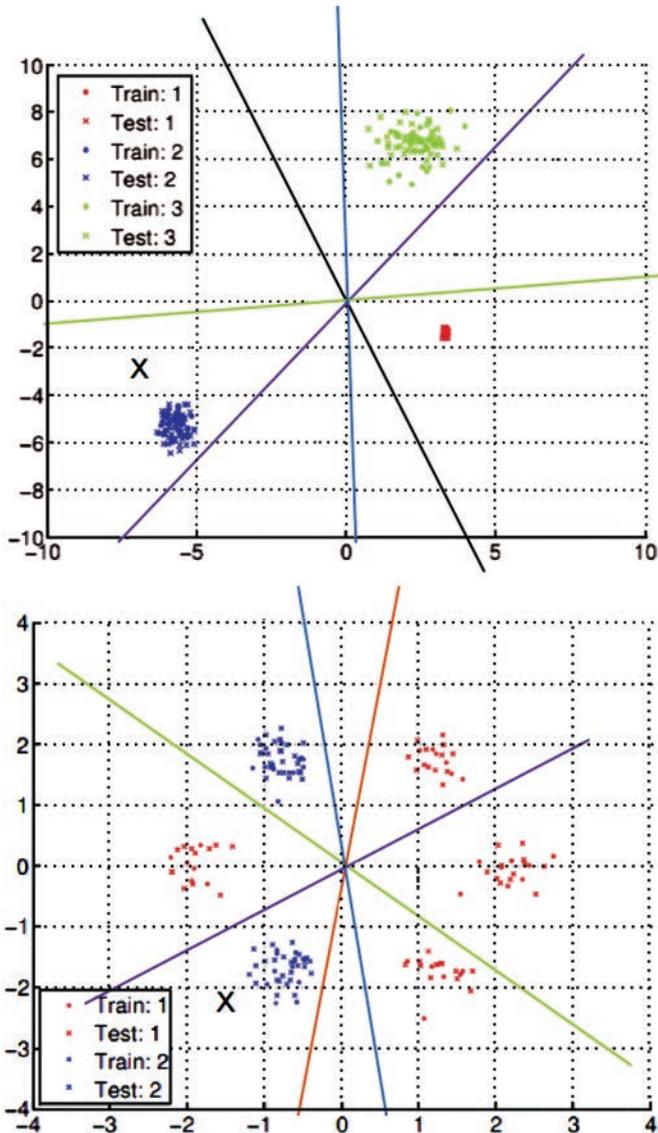


Figure 5. Figure 5. Two motivating examples for the classification method.

corresponding to the normal vectors of m randomly oriented hyperplanes, and let $Q = \text{sign}(AX)$ be the binary sign information. Then, the training algorithm proceeds in L “levels.” In the ℓ th level, m index sets $\Lambda_{\ell,i} \subset [m]$, $|\Lambda_{\ell,i}| = \ell$, $i = 1, \dots, m$, are randomly selected. During the i th “iteration” of the ℓ th level, the rows of Q indexed by $\Lambda_{\ell,i}$ are used to form the $\ell \times p$ submatrix of Q , the columns of which define the sign patterns $\{\pm 1\}^\ell$ observed by the training data. As in 1-bit CSP, these sign patterns contain the information about what sides of the hyperplanes the data points lie on.

At a given level ℓ , for the t th sign pattern (out of the possible 2^ℓ) and g th class, a *membership index* parameter $r(\ell, i, t, g)$ that uses knowledge of the number of training

points in class g having the t th sign pattern is calculated for every $\Lambda_{\ell,i}$. Larger values of $r(\ell, i, t, g)$ suggest that the t th sign pattern is more heavily dominated by class g ; thus, if a signal with an unknown label corresponds to the t th sign pattern, we will be more likely to classify it into the g th class. In [13], the authors define the following membership index parameter $r(\ell, i, t, g)$. Below, $P_{g|t} = P_{g|t}(\Lambda_{\ell,i})$ denotes the number of training points from the g th class with the t th sign pattern at the i th set selection in the ℓ th level:

$$r(\ell, i, t, g) = \frac{P_{g|t}}{\sum_{j=1}^G P_{j|t}} \frac{\sum_{j=1}^G |P_{g|t} - P_{j|t}|}{\sum_{j=1}^G P_{j|t}}. \quad (4)$$

Note that the first fraction in (4) indicates the proportion of training points in class g out of all points with sign pattern t (at the ℓ th level and i th iteration). The second fraction in (4) is a balancing term that gives more weight to group g when that group is much different in size than the others with the same sign pattern. Thus intuitively, these values should be large when the sign pattern t is very popular among data points labeled as class g and not as popular among other classes. With this intuition, we can then assign a label to a new test point x using its binary data $q = \text{sign}(Ax)$. For each class g , we simply sum the membership index function values over all ℓ and i , for those sign patterns t that match the sign pattern of the new test point x (which is known via the data q). The label for x is then decided by simply taking the class g corresponding to the largest sum.

Figure 6 shows classification results for the Lyme data. For this result, we use the survey responses for the symptom-related questions as our data matrix. This matrix consists of 3,686 “unwell” patients and 362 “well” patients (so 4,048 patients in total) who each answered twelve symptom-related questions (the “well” patients were asked about their worst symptoms while being sick). We randomly select a number (ten, fifty, or one hundred, see Figure 6) of those patients from each group to serve as our training data, and the remaining serve as our test data. We then run our method using this training and testing data and compute the correct classification rate, where the “ground truth” is determined from the separately asked question in the survey about whether the patient identifies as being “well” or “unwell.” The top plot of Figure 6 demonstrates the ability to accurately identify well versus unwell patients from the symptoms (current or past) that they report. Since the “well” patients were asked about their worst *prior* symptoms, one might ask whether it is simply the case that “well” patients showcase higher (or lower) symptom levels in general, making classification easy. However, the bottom plot of the figure demonstrates this is not the case, and that perhaps more intricate and complex symptom patterns are at work.

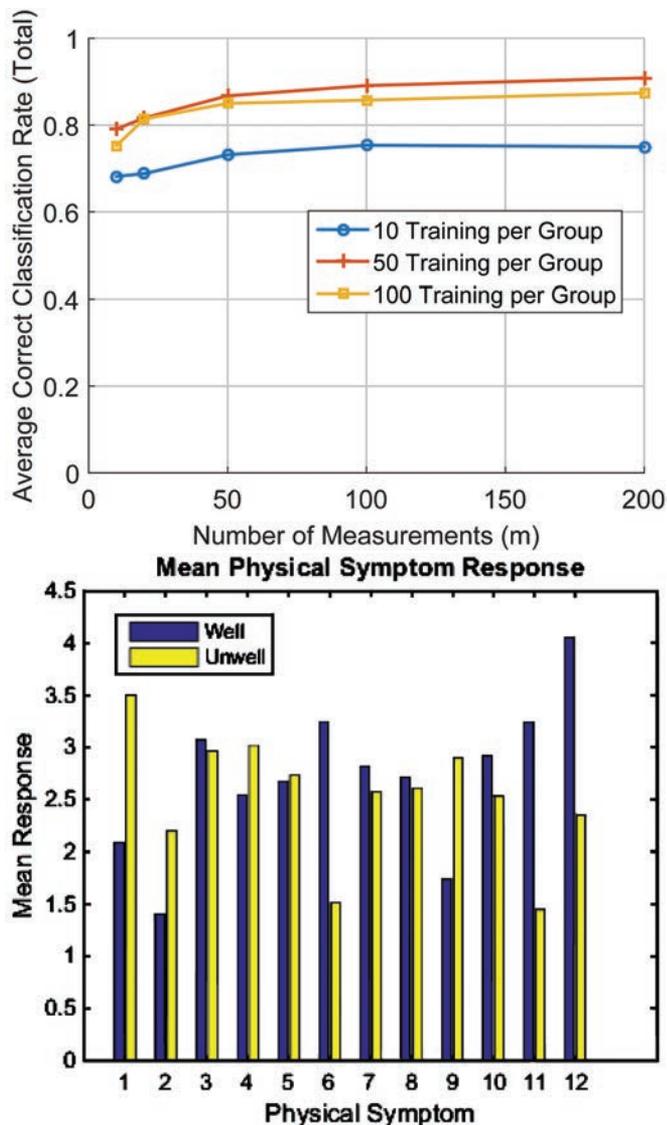


Figure 6. Figure 6. Top: Results from classification approach on symptom data using five layers for various numbers of randomly selected training points (patients). Bottom: Means on the survey questions for these groups.

Conclusion

Large-scale data is now ubiquitous, and the magnitude and abundance of data only continue to grow. Without proper mathematical techniques, this data will only bog down symptoms and stall scientific understanding. However, with sophisticated mathematical methods for large-scale data, the ability to acquire, store, complete, and analyze such data is possible. The recently acquired Lyme disease data is a perfect example of complicated data that has the ability to significantly enhance scientific and medical understanding. Unfortunately, clinical research of this illness

has been somewhat stagnant, controversial, and challenging. That makes the use of big data and mathematical analysis even more critical to make progress toward better understanding, diagnosis, and treatment of Lyme disease and on other important scientific fronts.

References

- [1] Baraniuk R, Foucart S, Needell D, Plan Y, Wootters M, Exponential decay of reconstruction error from binary measurements of sparse signals, *IEEE Trans. Info. Theory*, (63) no. 6: 3368–3385, 2017. [MR3677739](#)
- [2] Boufounos PT, Baraniuk RG, 1-bit compressive sensing, *42nd Annual Conf. Info. Sciences and Systems (CISS)*, 16–21, 2008.
- [3] Candès E, Tao T, The power of convex relaxation: Near-optimal matrix completion, *IEEE Trans. Inform. Theory*, (56) no. 5: 2053–2080, 2010. [MR2723472](#)
- [4] Candès EJ, Li X, Ma Y, Wright J, Robust principal component analysis?, *Journal of the ACM*, (58) no. 3: 1–37, 2011. [MR2811000](#)
- [5] Candès EJ, Romberg J, Tao T, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pur. Appl. Math.*, (59) no. 8: 1207–1223, 2006. [MR2230846](#)
- [6] Candès EJ, Tao T, Decoding by linear programming, *IEEE Trans. Info. Theory*, (51): 4203–4215, 2005. [MR2243152](#)
- [7] Eldar YC, Kutyniok G, *Compressed sensing: theory and applications*, Cambridge University Press, 2012. [MR3254193](#)
- [8] Foucart S, Needell D, Plan Y, Wootters M, De-biasing low-rank projection for matrix completion, *Proc. SPIE Optics and Photonics*, 2017.
- [9] Foucart S, Rauhut H, *A mathematical introduction to compressive sensing*, vol. 1, Birkhäuser Basel, 2013. [MR3100033](#)
- [10] Knudson K, Saab R, Ward R, One-bit compressive sensing with norm estimation, *IEEE Trans. Info. Theory*, (62) no. 5: 2748–2758, 2016. [MR3493876](#)
- [11] Lee T, Shraibman A, Matrix completion from any given set of observations, *Adva. Neur. In.*, 1781–1787, 2013.
- [12] Molitor D, Needell D, Matrix completion for structured observations, *Proc. Information Theory and Approximation*, 2018, to appear.
- [13] Needell D, Saab R, Woolf T, Simple classification using binary data, 2017, submitted.



Deanna Needell

Credits

Figure 1 is reprinted from *Applied and Computational Harmonic Analysis*, Volume 31, Issue 1, Emmanuel J. Candes, Yonina C. Eldar, Deanna Needell, and Paige Randall, Compressed Sensing with Coherent and Redundant Dictionaries, Pages 59-73, Copyright 2011, with permission from Elsevier.

Figure 2 is ©2017 IEEE. Reprinted, with permission, from Richard Baraniuk, Simon Foucart, Deanna Needell, Yaniv Plan, and Mary Wootters. Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Transactions on Information Theory*. Volume: 63, Issue: 6, June 2017.

Figures 3–6 are courtesy of the author.

Author photo is by Blake Hunter.

Advertise in the Notices

of the American Mathematical Society



Connect with an audience of approximately 30,000 subscribers through **Classified and Display Advertising** in the *Notices of the American Mathematical Society*, available in print and online.

As the world's most widely read magazine aimed at professional mathematicians, the *Notices* is an excellent medium for announcing publications, products, and services, and for recruiting mathematical scientists in academic, industry, and government.

Learn more about advertising
in the *Notices* at

www.ams.org/noticesadvertising

 **AMS** AMERICAN
MATHEMATICAL
SOCIETY