An Analysis of the t-SNE Algorithm for Data Visualization

Sanjeev Arora ARORA@CS.PRINCETON.EDU
Princeton University

Wei Hu HUWEI@CS.PRINCETON.EDU
Princeton University

Pravesh K. Kothari KOTHARI@CS.PRINCETON.EDU
Princeton University & Institute for Advanced Study

Editors: Sébastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

A first line of attack in exploratory data analysis is *data visualization*, i.e., generating a 2-dimensional representation of data that makes *clusters* of similar points visually identifiable. Standard Johnson-Lindenstrauss dimensionality reduction does not produce data visualizations. The *t-SNE* heuristic of van der Maaten and Hinton, which is based on non-convex optimization, has become the *de facto* standard for visualization in a wide range of applications.

This work gives a formal framework for the problem of data visualization – finding a 2-dimensional embedding of clusterable data that correctly separates individual clusters to make them visually identifiable. We then give a rigorous analysis of the performance of t-SNE under a natural, deterministic condition on the "ground-truth" clusters (similar to conditions assumed in earlier analyses of clustering) in the underlying data. These are the first provable guarantees on t-SNE for constructing good data visualizations.

We show that our deterministic condition is satisfied by considerably general probabilistic generative models for clusterable data such as mixtures of well-separated log-concave distributions. Finally, we give theoretical evidence that t-SNE provably succeeds in *partially* recovering cluster structure even when the above deterministic condition is not met.

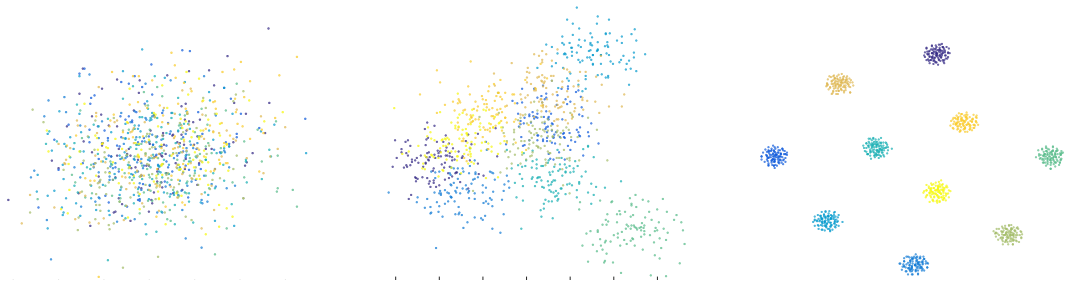
Keywords: Clustering, t-SNE, Visualization

1. Introduction

Many scientific applications, especially those involving exploratory data analysis, rely on visually identifying high-level qualitative structures in the data, such as clusters or groups of similar points. This is not easy since the data of interest is usually high-dimensional and it is unclear how to capture the qualitative cluster structure in a 2-dimensional visualization. For example, linear dimensionality reduction techniques (e.g., data oblivious Johnson-Lindenstrauss (JL) embedding or data-dependent embedding using PCA) are incapable of reducing dimension down to 2 in any meaningful way (see Figure 1) - they merge distinct clusters into a uniform-looking sea of points.

In 2008, van der Maaten and Hinton (2008) introduced a nonlinear algorithm, *t-Distributed Stochastic Neighbor Embedding or t-SNE* (an improvement over the earlier SNE algorithm of Hinton and Roweis (2002)) for this task, which has become the de facto standard (see Figure 1(*c*)) for

. Extended abstract. Full version appears as arXiv:1803.01768.



(*a*) (*b*) (*c*)

Figure 1: 2D embeddings of a mixture of 10 Gaussians with pairwise center separation $0.5\times$radius via: (a) random projection (JL), (b) projection to the subspace of top 2 singular vectors (PCA), (c) t-SNE.

visualizing high-dimensional datasets with diverse applications such as computer security (Gashi et al., 2009), music analysis (Hamel and Eck, 2010), cancer biology (Abdelmoula et al., 2016) and bioinformatics (Wallach and Lilien, 2009).

At a high level, t-SNE (like SNE) chooses two similarity measures between pairs of points - one for the high dimensional data and one for the 2-dimensional embedding. It then attempts to construct a 2-dimensional embedding that minimizes the KL divergence between the vector of similarities between pairs of points in the original dataset and the similarities between pairs of points in the embedding. This is a non-convex optimization problem and t-SNE employs gradient descent with random initialization (along with other tricks such as *early exaggeration*) to compute a reasonable solution to it. See the full version of this paper for details.

Of course, non-convex optimization drives much of today's progress in machine learning and data science, and thus poses a rich set of theoretical questions. Researchers have managed to rigorously analyze non-convex optimization algorithms in a host of settings (Dasgupta, 1999; Arora et al., 2012, 2014; Bhojanapalli et al., 2016; Ge et al., 2015a; Sun et al., 2017; Ge et al., 2017, 2016; Park et al., 2017). These analyses usually involve making clean assumptions about the structure of data, usually with a generative model. The goal of the current paper is to rigorously analyze t-SNE in a similar vein.

At the outset such a project runs into definitional issues about what a good *visualization* of clustering is. Many such issues are inherited from well-known issues in formalizing the goals of clustering (Kleinberg, 2002). In theoretical studies of clustering, such issues were sidestepped by going with a standard clustering formalization and assuming that data come with an (unknown) ground-truth clustering (for instance, mixtures of Gaussians, k-means, etc.). We make similar assumptions and assume that our goal is to produce a 2-dimensional embedding such that the points in the same clusters are noticeably closer together compared with points in different clusters. Under some of these standard models we show that t-SNE provably succeeds in computing a good visualization.

We emphasize that the focus of this paper is on formalizing the notion of visualization and providing a theoretical analysis of t-SNE. We do not advocate for t-SNE over other visualization methods.

We now begin by describing our formalization of the visualization problem followed by describing our results that give the first provable guarantees on t-SNE for computing visualization of clusterable data.

Formalizing Visualization. We assume that we are given a collection of points $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ and that there exists a "ground-truth" clustering described by a partition $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ of $[n]$ into k clusters.

A visualization is described by a 2-dimensional embedding $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\} \subseteq \mathbb{R}^2$ of $\mathcal{X}$, where each $x_i \in \mathcal{X}$ is mapped to the corresponding $y_i \in \mathcal{Y}$. Intuitively, a cluster $\mathcal{C}_\ell$ in the original data is *visualized* if the corresponding points in the 2-dimensional embedding $\mathcal{Y}$ are well-separated from all the rest. The following definition formalizes this idea.

Definition 1.1 (Visible cluster) *Let $\mathcal{Y}$ be a 2-dimensional embedding of a dataset $\mathcal{X}$ with ground-truth clustering $\mathcal{C}_1, \ldots, \mathcal{C}_k$. Given $\epsilon \geq 0$, a cluster $\mathcal{C}_\ell$ in $\mathcal{X}$ is said to be* $(1-\epsilon)$-visible *in $\mathcal{Y}$ if there exist $\mathcal{P}, \mathcal{P}_{\text{err}} \subseteq [n]$ such that:*

1. $|(\mathcal{P} \setminus \mathcal{C}_\ell) \cup (\mathcal{C}_\ell \setminus \mathcal{P})| \leq \epsilon \cdot |\mathcal{C}_\ell|$, $|\mathcal{P}_{\text{err}}| \leq \epsilon n$, *and*
2. *for every $i, i' \in \mathcal{P}$ and $j \in [n] \setminus (\mathcal{P} \cup \mathcal{P}_{\text{err}})$, $\|y_i - y_{i'}\| \leq \frac{1}{2}\|y_i - y_j\|$.*

In such a case, we say that $\mathcal{P}$ $(1-\epsilon)$-visualizes *$\mathcal{C}_i$ in $\mathcal{Y}$.*

It is now easy to define when $\mathcal{Y}$ is a good *visualization* - we ask that every cluster $\mathcal{C}_\ell$ in the dataset $\mathcal{X}$ is visualized in $\mathcal{Y}$.

Definition 1.2 (Visualization) *Let $\mathcal{Y}$ be a 2-dimensional embedding of a dataset $\mathcal{X}$ with ground-truth clustering $\mathcal{C}_1, \ldots, \mathcal{C}_k$. Given $\epsilon \geq 0$, we say that $\mathcal{Y}$ is a* $(1-\epsilon)$-visualization *of $\mathcal{X}$ if there exists a partition $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k, \mathcal{P}_{\text{err}}$ of $[n]$ such that:*

(i) For each $i \in [k]$, $\mathcal{P}_i$ $(1-\epsilon)$-visualizes $\mathcal{C}_i$ in $\mathcal{Y}$, and

(ii) $|\mathcal{P}_{\text{err}}| \leq \epsilon n$.

In particular, when $\epsilon = 0$, we say that $\mathcal{Y}$ is a full visualization *of $\mathcal{X}$.*

Remark *Note that this formalization of visualization should be considered a first cut, since ultimately human psychology must come into play. For instance, humans may reasonably visualize two parallel lines as two clusters, but these violate our definition.*

A natural question is whether clustering inferred from a visualization is unique. Our definition above does not guarantee this. Indeed, this is inherently impossible and relates to the ambiguity in the definition of clustering: for example, it can be impossible to determine whether a given set of points should be viewed as one cluster or two different smaller clusters. See Figure 2 for an example.

It is, however, not hard to establish that under an additional assumption that the size (number of points) of any cluster is smaller than twice the size of any other, full visualization as defined in Definition 1.2 uniquely determines a clustering.

In order to study fine-grained behaviors of t-SNE, we also define a weaker variant of visualization where at least one cluster is visualized.

Definition 1.3 (Partial visualization) *Given $\epsilon \geq 0$, we say that $\mathcal{Y}$ is a* $(1-\epsilon)$-partial visualization *of $\mathcal{X}$ if there exists a subset $\mathcal{P} \subseteq [n]$ such that $\mathcal{P}$ $(1-\epsilon)$-visualizes $\mathcal{C}_\ell$ for some $\ell \in [k]$.*

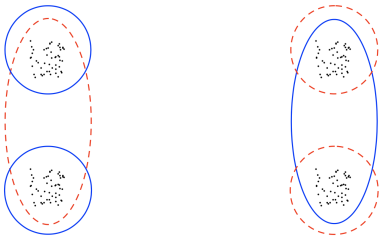


Figure 2: If we knew that there are 3 clusters in the original data, the blue and red outlines denote equally valid guesses for the underlying clustering based on the above visualization.

1.1. Our Results

Our main result identifies a simple deterministic condition on the clusterable data under which t-SNE provably succeeds in computing a full visualization.

Definition 1.4 (Well-separated, spherical data) *Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ be clusterable data with $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ defining the individual clusters such that for each $\ell \in [k]$, $|\mathcal{C}_\ell| \geq 0.1(n/k)$. We say that $\mathcal{X}$ is γ-spherical and γ-well-separated if for some $b_1, b_2, \ldots, b_k > 0$, we have:*

1. *γ-**Spherical:** For any $\ell \in [k]$ and $i, j \in \mathcal{C}_\ell$ $(i \neq j)$, we have $\|x_i - x_j\|^2 \geq \frac{b_\ell}{1+\gamma}$, and for any $i \in \mathcal{C}_\ell$ we have $\left|\left\{j \in \mathcal{C}_\ell \setminus \{i\} : \|x_i - x_j\|^2 \leq b_\ell\right\}\right| \geq 0.51|C_\ell|$.*
2. *γ-**Well-separated:** For any $\ell, \ell' \in [k]$ $(\ell \neq \ell')$, $i \in \mathcal{C}_\ell$ and $j \in \mathcal{C}_{\ell'}$ we have $\|x_i - x_j\|^2 \geq (1 + \gamma \log n) \max\{b_\ell, b_{\ell'}\}$.*

The first condition asks for the distances between points in the same cluster ("intra-cluster distances") to be concentrated around a single value (with γ controlling the "amount" of concentration). The second condition requires that the distances between two points from different clusters should be somewhat larger than the intra-cluster distances for each of the two clusters involved. In addition, we require that none of the clusters has too few points. Such assumptions are satisfied by well-studied probabilistic generative models for clusterable data such as mixture of Gaussians and more generally, mixture of log-concave distributions, and have been used in previous work (Dasgupta, 1999; Arora and Kannan, 2005) studying "distance-based" clustering algorithms.

For spherical and well-separated data, our main theorem below shows that t-SNE with early exaggeration succeeds in finding a full visualization.

Theorem 1.5 (Informal) *Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ be γ-spherical and γ-well-separated clusterable data with $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ defining the individual clusters. Then, t-SNE with early exaggeration on input $\mathcal{X}$ outputs a full visualization of $\mathcal{X}$ with high probability.*

Proof Technique. At a high level, t-SNE starts with a randomly initialized embedding and makes iterative gradient updates to it. The analysis thus demands understanding the effect of this update rule to the embedding of the high-dimensional points as a function of whether they lie in the same cluster or not. In a recent work, Linderman and Steinerberger (2017) established a "shrinkage" result for this update rule - they showed that points in the same cluster move towards each other under some mild conditions, that is, the embedding of any cluster "shrinks" as the iterations proceed.

This result, however, is insufficient to establish that t-SNE succeeds in finding a full visualization as it does not rule out multiple clusters merging into each other.

We resort to a more fine-grained analysis built on the one by Linderman and Steinerberger (2017) and obtain an update rule for the *centroids* of the embeddings of all underlying clusters. This allows us to track the changes to the positions of the centroids and show that the distance between distinct centroids remains *lower-bounded* whenever the data is γ-spherical and γ-well-separated. Combined with the shrinkage result for points in the same cluster, this implies that t-SNE outputs a full visualization of the data.

Our analysis implicitly relies on the update rule in t-SNE closely mimicking those appearing in the well-studied *noisy power method* (with non-random noise). We make this connection explicit and show that the behavior of t-SNE (with early exaggeration) on γ-spherical and well-separated data can in fact be closely approximated by power method run on a natural matrix of pairwise similarities.

Application to Visualizing Mixture Models. Mixture of Gaussians and more generally, mixture of log-concave distributions, are well-studied probabilistic generative models for clusterable data. As an immediate application of our main theorem above, we show that t-SNE produces a full visualization for data generated according to such models. Before describing the result, we quickly recall the definition of mixture of log-concave distributions.

A distribution $\mathcal{D}$ with density function f on $\mathbb{R}^d$ is said to be *log-concave* if $\log(f)$ is a concave function. $\mathcal{D}$ is said to be *isotropic* if its covariance is I. Many natural distributions including Gaussian distributions and the uniform distribution on any convex set are log-concave.

A mixture of k log-concave distributions is described by k positive mixing weights $w_1, w_2, \ldots, w_k$ ($\sum_{\ell=1}^{k} w_\ell = 1$) and k log-concave distributions $\mathcal{D}_1, \ldots \mathcal{D}_k$ in $\mathbb{R}^d$. To sample a point from this model, we pick cluster ℓ with probability w_ℓ and draw x from $\mathcal{D}_\ell$.

Theorem 1.5 immediately implies that t-SNE constructs a full visualization for data generated from a mixture of isotropic Gaussians or log-concave distributions with well-separated means. For isotropic Gaussians, the required pairwise separation between means is $\tilde{\Omega}(d^{1/4})$. For more general isotropic log-concave distributions, we require that the means be separated by $\tilde{\Omega}(d^{5/12})$.

Observe that the radius of samples from an isotropic log-concave distribution is $\approx d^{1/2}$ - thus, t-SNE succeeds in constructing 2D visualizations for clustering models far below the separation at which the clusters are non-overlapping. This is in stark contrast to standard linear dimensionality reduction techniques such as the Johnson-Lindenstrauss embedding that require mean separation of $\Omega(d^{1/2})$ to construct 2D visualizations that correctly separate 99% of points.

Corollary 1.6 (Informal) *Let $\mathcal{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ be i.i.d. samples from an equal-weighted mixture of k isotropic Gaussians in $\mathbb{R}^d$ with every pair of distinct means separated by $\tilde{\Omega}(d^{1/4})$. Then, with high probability t-SNE with early exaggeration on input $\mathcal{X}$ outputs a full visualization of $\mathcal{X}$. Moreover, the same result holds for mixture of isotropic log-concave distributions with every pair of distinct means separated by $\tilde{\Omega}(d^{5/12})$.*

Remark *Our result actually holds for a larger subclass of mixtures of* non-isotropic *log-concave distributions that may not be equal-weighted. See the full version of this paper for details. Mixture of log-concave distributions is among the weakest assumptions under which clustering algorithms with provable guarantees have been designed (Arora and Kannan, 2005; Vempala and Wang, 2004).*

We show that the t-SNE heuristic can visualize clusters under assumptions similar to the more sophisticated methods in previous theoretical work.

Finally, we show that even when the conditions in Definition 1.4 are not met, t-SNE can still provably visualize at least one cluster in the original data in some cases. As an example, using a more fine-grained analysis, we show that t-SNE computes a partial visualization for data obtained from a mixture of two *concentric* (thus, no mean separation at all!) spherical Gaussians with variances differing by a constant factor.

Theorem 1.7 (Informal) *Let $\mathcal{X}$ be generated from an equal-weighted mixture of two Gaussians $\mathcal{N}(0, \sigma_1^2)$ and $\mathcal{N}(0, \sigma_2^2)$ such that $1.5 \leq \sigma_2/\sigma_1 \leq 10$. Then t-SNE with early exaggeration on input $\mathcal{X}$ outputs a $(1 - d^{-\Omega(1)})$-partial visualization of $\mathcal{X}$ where $\mathcal{C}_1$ is $(1 - d^{-\Omega(1)})$-visible.*

1.2. Related Work

This paper continues the line of work focused on analyzing gradient descent and related heuristics for non-convex optimization problems, examples of which we have discussed before. Theoretically analyzing t-SNE, in particular, was recently considered in a work of Linderman and Steinerberger (2017) who showed that running t-SNE with early exaggeration causes points from the same cluster to move towards each other (i.e., embedding of any cluster shrinks). As discussed before, however, this does not imply that t-SNE ends up with a visualization as all the clusters could potentially collapse into each other. Another work by Shaham and Steinerberger (2017) derived a theoretical property of SNE, but their result is only nontrivial when the number of clusters is significantly larger than the number of points per cluster, which is an unrealistic assumption.

Mixture models are natural average-case generative models for clusterable data which have been studied as benchmarks for analyzing various clustering algorithms and have a long history of theoretical work. By now, a sequence of results (Dasgupta et al., 2007, 2006; Arora and Kannan, 2005; Vempala and Wang, 2004; Achlioptas and McSherry, 2005; Kannan et al., 2005; Vempala, 2007; Hsu and Kakade, 2013; Ge et al., 2015b; Kalai et al., 2012; Belkin and Sinha, 2010; Kalai et al., 2010; Kothari and Steinhardt, 2017; Hopkins and Li, 2017; Diakonikolas et al., 2017) have identified efficient algorithms for clustering data from such models under various natural assumptions.

Acknowledgments

This research was done with support from NSF, ONR, Simons Foundation, Mozilla Research, and Schmidt Foundation.

References

Walid M. Abdelmoula, Benjamin Balluff, Sonja Englert, Jouke Dijkstra, Marcel J. T. Reinders, Axel Walch, Liam A. McDonnell, and Boudewijn P. F. Lelieveldt. Data-driven identification of prognostic tumor subpopulations using spatially mapped t-sne of mass spectrometry imaging data. *Proceedings of the National Academy of Sciences*, 113(43):12244–12249, 2016.

Dimitris Achlioptas and Frank McSherry. On spectral learning of mixtures of distributions. In *18th Annual Conference on Learning Theory*, pages 458–469, 2005.

Sanjeev Arora and Ravi Kannan. Learning mixtures of separated nonspherical gaussians. *The Annals of Applied Probability*, 15(1A):69–92, 2005.

Sanjeev Arora, Rong Ge, and Ankur Moitra. Learning topic models—going beyond SVD. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, pages 1–10. IEEE Computer Soc., Los Alamitos, CA, 2012.

Sanjeev Arora, Rong Ge, and Ankur Moitra. New algorithms for learning incoherent and overcomplete dictionaries. In *COLT*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 779–806. JMLR.org, 2014.

Mikhail Belkin and Kaushik Sinha. Polynomial learning of distribution families. In *FOCS*, pages 103–112. IEEE Computer Society, 2010.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. In *NIPS*, pages 3873–3881, 2016.

Anirban Dasgupta, John E. Hopcroft, Ravi Kannan, and Pradipta Prometheus Mitra. Spectral clustering by recursive partitioning. In *ESA*, volume 4168 of *Lecture Notes in Computer Science*, pages 256–267. Springer, 2006.

Anirban Dasgupta, John E. Hopcroft, Ravi Kannan, and Pradipta Prometheus Mitra. Spectral clustering with limited independence. In *SODA*, pages 1036–1045. SIAM, 2007.

Sanjoy Dasgupta. Learning mixtures of gaussians. In *FOCS*, pages 634–644, 1999.

Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. List-decodable robust mean estimation and learning mixtures of spherical gaussians. *CoRR*, abs/1711.07211, 2017.

Ilir Gashi, Vladimir Stankovic, Corrado Leita, and Olivier Thonnard. An experimental study of diversity with off-the-shelf antivirus engines. In *Proceedings of The Eighth IEEE International Symposium on Networking Computing and Applications, NCA 2009, July 9-11, 2009, Cambridge, Massachusetts, USA*, pages 4–11, 2009.

Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points - online stochastic gradient for tensor decomposition. In *COLT*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 797–842. JMLR.org, 2015a.

Rong Ge, Qingqing Huang, and Sham M. Kakade. Learning mixtures of gaussians in high dimensions. In *STOC*, pages 761–770. ACM, 2015b.

Rong Ge, Jason D. Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *CoRR*, abs/1605.07272, 2016.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *ICML*, pages 1233–1242, 2017.

Philippe Hamel and Douglas Eck. Learning features from music audio with deep belief networks. In *ISMIR*, pages 339–344. International Society for Music Information Retrieval, 2010.

Geoffrey E. Hinton and Sam T. Roweis. Stochastic neighbor embedding. In *NIPS*, pages 833–840, 2002.

Samuel B. Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs. *CoRR*, abs/1711.07454, 2017.

Daniel Hsu and Sham M. Kakade. Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In *ITCS'13—Proceedings of the 2013 ACM Conference on Innovations in Theoretical Computer Science*, pages 11–19. ACM, New York, 2013.

Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Efficiently learning mixtures of two gaussians. In *STOC*, pages 553–562. ACM, 2010.

Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Disentangling gaussians. *Commun. ACM*, 55(2):113–120, 2012.

Ravindran Kannan, Hadi Salmasian, and Santosh Vempala. The spectral method for general mixture models. In *COLT*, pages 444–457, 2005.

Jon M. Kleinberg. An impossibility theorem for clustering. In *NIPS*, pages 446–453, 2002.

Pravesh K. Kothari and Jacob Steinhardt. Better agnostic clustering via relaxed tensor norms. *CoRR*, abs/1711.07465, 2017.

George C Linderman and Stefan Steinerberger. Clustering with t-sne, provably. *arXiv preprint arXiv:1706.02582*, 2017.

Dohyung Park, Anastasios Kyrillidis, Constantine Caramanis, and Sujay Sanghavi. Non-square matrix sensing without spurious local minima via the burer-monteiro approach. In *AISTATS*, pages 65–74, 2017.

Uri Shaham and Stefan Steinerberger. Stochastic neighbor embedding separates well-separated clusters. *arXiv preprint arXiv:1702.02670*, 2017.

Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere I: overview and the geometric picture. *IEEE Trans. Information Theory*, 63(2):853–884, 2017.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Santosh Vempala. Spectral algorithms for learning and clustering. In *COLT*, volume 4539 of *Lecture Notes in Computer Science*, pages 3–4. Springer, 2007.

Santosh Vempala and Grant Wang. A spectral algorithm for learning mixture models. *Journal of Computer and System Sciences*, 68(4):841–860, 2004.

Izhar Wallach and Ryan Lilien. The protein-small-molecule database (psmdb), a non-redundant structural resource for the analysis of protein-ligand binding. *Bioinformatics, 25(5):615-20*, 2009.