Improving Channel Charting with Representation-Constrained Autoencoders

Pengzhi Huang[1], Oscar Castañeda[1], Emre Gönültaş[1], Saïd Medjkouh[1], Olav Tirkkonen[2], Tom Goldstein[3], and Christoph Studer[1]

[1]*School of Electrical and Computer Engineering, Cornell University, Ithaca, NY; email: studer@cornell.edu*
[2]*School of Electrical Engineering, Aalto University, Finland; e-mail: olav.tirkkonen@aalto.fi*
[3]*Department of Computer Science, University of Maryland, College Park, MD; e-mail: tomg@cs.umd.edu*

***Abstract*—Channel charting (CC) has been proposed recently to enable logical positioning of user equipments (UEs) in the neighborhood of a multi-antenna base-station solely from channel-state information (CSI). CC relies on dimensionality reduction of high-dimensional CSI features in order to construct a channel chart that captures spatial and radio geometries so that UEs close in space are close in the channel chart. In this paper, we demonstrate that autoencoder (AE)-based CC can be augmented with side information that is obtained during the CSI acquisition process. More specifically, we propose to include pairwise representation constraints into AEs with the goal of improving the quality of the learned channel charts. We show that such representation-constrained AEs recover the global geometry of the learned channel charts, which enables CC to perform approximate positioning without global navigation satellite systems or supervised learning methods that rely on extensive and expensive measurement campaigns.**

I. Introduction

Autoencoders (AEs) are single- or multi-layer neural networks widely used for dimensionality-reduction tasks [1]–[4]. AEs learn low-dimensional representations (embeddings) of a given high-dimensional dataset and have been shown to accurately preserve spatial relationships in both high- and low-dimensional space for a broad range of synthetic and real-world datasets [2]. With the success of deep neural networks, AEs are also gaining increased attention for unsupervised learning tasks [5]. Notable application examples of AEs include learning word embeddings [6], image compression [7], generative models [8], and channel charting [9], [10]. AEs are typically trained in an unsupervised manner, i.e., no labels are used, while potential side information on the training data is routinely ignored or application-specific representation structure is not imposed during training.

AEs that impose structural constraints on the latent variables include sparse AEs [11] and variational AEs [12]. Sparse AEs enforce sparsity on the representations, which enables one to learn embeddings with low effective dimensionality. Variational AEs learn an embedding drawn from a distribution that represents the high-dimensional input [12]; such AEs have been shown to be able to generate complex data, such as handwritten digits [13], faces [14], or physical scenes [14].

The work of PH, OC, EG, SM, and CS was supported by Xilinx, Inc. and by the US NSF grants ECCS-1408006, CCF-1535897, CCF-1652065, CNS-1717559, and ECCS-1824379. The work of TG was supported by the US NSF under grant CCF-1535902 and by the US Office of Naval Research grant N00014-17-1-2078. The work of OT was funded in part by the Academy of Finland (grant 319484).

A. Representation-Constrained Autoencoders for Positioning

A range of dimensionality-reduction applications provide valuable side information that can be imposed on the low-dimensional representations. Such side information may stem either from the dataset itself or from the way data was collected. One example arises when data is acquired over time, where it may be natural to enforce constraints *between* representations by exploiting the fact that, for temporally correlated datapoints, the associated low-dimensional representations should be similar. Another example arises when a subset of the representations are known a-priori, e.g., when a small part of the training data has been annotated. The obtained information can then be translated into representation constraints, which leads to semi-supervised training of AEs.

Representation constraints are important for positioning users in wireless systems using *channel charting* (CC) [9], [10]. CC measures high-dimensional channel-state information (CSI) of user equipments (UEs) transmitting data to an access point or cell tower. By collecting CSI at multiple spatial locations over time, one can train an AE for which the low-dimensional representations reflect relative UE positions. While the original CC method [9] enables relative localization without access to global navigation satellite systems (GNSS) and without dedicated measurement campaigns [15], [16], valuable side information should not be ignored when available. As UEs move with finite velocity, one could consider this information when training an AE to ensure that temporally correlated datapoints are nearby in the representation space. One could also imagine that certain points in space with known location (e.g., a coffee shop) can be associated with measured CSI; this helps to pin down a subset of spatial locations in the representation space, which enables absolute positioning. Put simply, enforcing constraints *between* representations may improve the efficacy of AE-based CC.

B. Contributions

This paper investigates representation constraints for AEs and provides a framework for including these during training. We propose constraints on pairs of representations in which either the absolute or relative distance among (a subset of)

representations is enforced. We formulate these constraints as nonconvex regularizers, which can easily be included into well-established deep-learning frameworks. We highlight the efficacy of representation constraints for the application of CC-based positioning in wireless systems: in particular, we demonstrate that by combining partially-annotated locations with temporal UE constraints, the positioning performance of CC [9] can be improved significantly.

C. Relevant Prior Art

Autoencoders provide excellent dimensionality-reduction performance with real-world datasets [2]. Despite their success, AEs do not aim at preserving geometric properties on the representations (such as distances between points); this is in stark contrast to, e.g., Sammon's mapping [17] or multidimensional scaling [18]. Furthermore, AEs commonly ignore side information that stems from the application at hand. We will extend AEs with *pairwise representation constraints* that can improve performance of dimensionality reduction tasks.

The efficacy of deep learning has been explored recently for wireless positioning [19]–[21]. The methods in these papers rely on extensive measurement campaigns and require CSI measurements annotated with exact position information. To avoid the drawback of such supervised methods, CC, as put forward in [9], uses dimensionality reduction to extract *relative* position information of users without the necessity of costly measurement campaigns. CC exploits the fact that CSI is high-dimensional, but strongly depends on UE position, which is low-dimensional. Dimensionality reduction applied to CSI measurements learns a *channel chart*, in which nearby points represent nearby locations in true space—exact position information is not available. However, CC naturally provides representation constraints originating from the acquisition process, which have been ignored in prior work. We will include such constraints into AEs to significantly improve CC.

II. Representation-Constrained Autoencoders

We briefly summarize the basics of AEs and then introduce the concept of pairwise representation constraints.

A. Autoencoders in a Nutshell

AEs take a high-dimensional dataset consisting of N datapoints (vectors) $\mathbf{x}_n \in \mathbb{R}^D$, $n = 1, \ldots, N$, of dimension D, and learn two functions: the encoder $f^e : \mathbb{R}^D \to \mathbb{R}^{D'}$ and the decoder $f^d : \mathbb{R}^{D'} \to \mathbb{R}^D$. The encoder maps datapoints onto low-dimensional *representations* $\mathbf{y}_n \in \mathbb{R}^{D'}$, $n = 1, \ldots, N$, where $D' \ll D$ is the dimension of the representation, and the decoder maps representations back to datapoints:

$$\mathbf{y}_n = f^e(\mathbf{x}_n) \quad \text{and} \quad \mathbf{x}_n = f^d(\mathbf{y}_n), \quad n = 1, \ldots, N.$$

The encoder and decoder functions of AEs are implemented as multilayer (shallow or deep) feed-forward neural networks that are trained to minimize the mean-square error (MSE) between the input and the output of the network, specifically:

$$L(\mathcal{W}^e, \mathcal{W}^d) = \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{x}_n - f^d(f^e(\mathbf{x}_n))\|^2. \tag{1}$$

The parameters to be learned from $\{\mathbf{x}_n\}_{n=1}^N$ are the weights and bias terms contained in the sets $\mathcal{W}^e$ and $\mathcal{W}^d$, which define the encoder and decoder neural networks, respectively.

The D'-dimensional output of the encoder f^e is typically of lower dimension than the intrinsic dimension of the manifold embedding the inputs $\mathbf{x}_i$ in D dimensions. Hence, we have that $\mathbf{x}_n \approx f^d(f^e(\mathbf{x}_n))$, $n = 1, \ldots, N$, unless the dataset $\{\mathbf{x}_n\}_{n=1}^N$ was D'-dimensional and we were able to learn the underlying structure. Nevertheless, AEs often find low-dimensional representations $\{\mathbf{y}_n\}_{n=1}^N$ with small loss that capture the intrinsic dimensionality of the input datapoints.

B. Pairwise Representation Constraints

In Table I, we propose four distinct pairwise representation constraints, where the underlined quantities represent *constant* scalars or vectors that are known a-priori and used during AE training; non-underlined quantities are optimization variables.

1) Fixed Distance Constraints: The fixed absolute distance (FAD) and fixed relative distance (FRD) constraints enforce a known distance $\underline{d}_{i,j}$ on a pair of representations according to $\|\mathbf{y}_i - \mathbf{y}_j\| = \underline{d}_{i,j}$. The difference between FAD and FRD is that, for FAD, one of the two representations, e.g., $\underline{\mathbf{y}}_j$, is a constant known prior to AE learning; for FRD, both representations $\mathbf{y}_i$ and $\mathbf{y}_j$ are optimization variables. To facilitate the inclusion of these constraints in deep learning frameworks, we propose to use regularizers (see Table I) for which generalized gradients exist. Concretely, the generalized gradient of the FAD and FRD constraints with respect to the representation $\mathbf{y}_i$ is

$$\nabla_{\mathbf{y}_i} (\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j})^2 = 2(\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j}) \frac{\mathbf{y}_i - \mathbf{y}_j}{\|\mathbf{y}_i - \mathbf{y}_j\|}, \tag{2}$$

where the representation $\underline{\mathbf{y}}_j$ is known. If $\underline{d}_{i,j} = 0$, then the FRD regularizer promotes equality among $\mathbf{y}_i$ and $\mathbf{y}_j$, whereas the FAD regularizer will learn a representation $\mathbf{y}_i$ that is close to the constant vector $\underline{\mathbf{y}}_j$. Intuitively, the FAD constraint for $\underline{d}_{i,j} = 0$ acts as a semi-supervised extension in which a subset of representations are known a-priori.

2) Maximum Distance Constraints: The maximum absolute distance (MAD) and maximum relative distance (MRD) constraints enforce a maximum a-priori known distance $\underline{d}_{i,j}$ between a pair of representations according to $\|\mathbf{y}_i - \mathbf{y}_j\| \leq \underline{d}_{i,j}$. For MAD, one of the two vectors in the constraint, e.g., $\underline{\mathbf{y}}_j$, is known a-priori; for MRD, both representations are learned. We include these constraints as regularizers (see Table I) with the generalized gradient

$$\begin{aligned} &\nabla_{\mathbf{y}_i} \max\{\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j}, 0\}^2 \\ &\qquad = 2 \max\{\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j}, 0\} \frac{\mathbf{y}_i - \mathbf{y}_j}{\|\mathbf{y}_i - \mathbf{y}_j\|}, \end{aligned} \tag{3}$$

where $\underline{\mathbf{y}}_j$ is known for MAD. Note, when $\underline{d}_{i,j} = 0$ the FAD, MAD, FRD, and MRD are all equivalent.

3) Practical Considerations: We implemented a stochastic optimizer to minimize the sum of the AE fidelity term (1) and the regularized constraint penalties using the Keras and TensorFlow frameworks [22], [23]. Because penalty terms may represent pairwise constraints that involve two data points, the

TABLE I
SUMMARY OF PROPOSED REPRESENTATION CONSTRAINTS FOR AEs (KNOWN QUANTITIES ARE UNDERLINED).

Name	Constraint	Regularizer
Fixed absolute distance (FAD)	$\|\mathbf{y}_i - \underline{\mathbf{y}}_j\| = \underline{d}_{i,j}$	$(\|\mathbf{y}_i - \underline{\mathbf{y}}_j\| - \underline{d}_{i,j})^2$
Fixed relative distance (FRD)	$\|\mathbf{y}_i - \mathbf{y}_j\| = \underline{d}_{i,j}$	$(\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j})^2$
Maximum absolute distance (MAD)	$\|\mathbf{y}_i - \underline{\mathbf{y}}_j\| \leq \underline{d}_{i,j}$	$\max\{\|\mathbf{y}_i - \underline{\mathbf{y}}_j\| - \underline{d}_{i,j}, 0\}^2$
Maximum relative distance (MRD)	$\|\mathbf{y}_i - \mathbf{y}_j\| \leq \underline{d}_{i,j}$	$\max\{\|\mathbf{y}_i - \mathbf{y}_j\| - \underline{d}_{i,j}, 0\}^2$

stochastic approximation of the regularizers was formed by randomly sampling constraints rather than datapoints.

C. Performance Metrics

In order to measure the performance of dimensionality reduction, we use two standard metrics for the local neighborhood-preserving performance: *trustworthiness* (TW) and *continuity* (CT). TW measures whether mapping high-dimensional datapoints to the representation space introduces new (false) neighbors. TW is defined as

$$TW(K) = 1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^{N} \sum_{j \in \mathcal{U}_i^K} (r(i,j) - K),$$

where $r(i,j)$ represents the rank of the representation $\mathbf{y}_i$ among the pairwise distances between the other representations. The set $\mathcal{U}_i^K$ contains the points that are among the K nearest neighbors in representation space, but not in the high-dimensional space. CT measures if similar datapoints in original space remain similar in the representation space, and is defined as

$$CT(K) = 1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^{N} \sum_{j \in \mathcal{V}_i^K} (\hat{r}(i,j) - K),$$

where $\hat{r}(i,j)$ represents the rank of the datapoint $\mathbf{x}_i$ among the pairwise distances between the other datapoints. The set $\mathcal{V}_i^K$ contains the points that are among the K nearest neighbors in the high-dimensional space, but not in the representation space. Both TW and CT have values in the range $[0,1]$ and large values indicate that neighborhoods are better preserved.

Besides measuring the local-neighborhood-preservation properties via TW and CT, we also consider Kruskal's stress (KS) [24], [25], which measures how well the global structure in the high-dimensional dataset $\{\mathbf{x}_n\}_{n=1}^N$ is mapped to the low-dimensional embedding $\{\mathbf{y}_n\}_{n=1}^N$. KS is defined as

$$KS = \sqrt{\frac{\sum_{n,m} (\delta_{n,m} - \beta \hat{\delta}_{n,m})^2}{\sum_{n,m} \delta_{n,m}^2}},$$

where $\delta_{n,m} = \|\mathbf{x}_n - \mathbf{x}_m\|$, $\hat{\delta}_{n,m} = \|\mathbf{y}_n - \mathbf{y}_m\|$, and $\beta = \sum_{n,m} \delta_{n,m} \hat{\delta}_{n,m} / \sum_{n,m} \delta_{n,m}^2$ is the optimal distance scaling factor. KS is in the range $[0,1]$ and smaller values indicate that global geometrical structure is preserved better. If $KS = 0$, then the geometry is perfectly preserved.

III. CHANNEL CHARTING WITH REPRESENTATION-CONSTRAINED AUTOENCODERS

We now augment the original CC framework [9] with representation constraints that naturally arise from the application. We start by outlining the concept of CC and then explain how representation constraints are included. We then demonstrate the efficacy in comparison to the original CC framework.

A. Channel Charting in a Nutshell

CC measures CSI from users at different spatial locations and learns a low-dimensional *channel chart* that preserves *locally* the original spatial geometry. Put simply, users that are physically nearby will be placed nearby in the channel chart and vice versa—global geometry is typically not preserved. In this framework, high-dimensional features are extracted from CSI, then processed with dimensionality-reduction methods to obtain the low-dimensional channel chart. CC operates in an unsupervised manner, i.e., learning is only based on CSI that is passively collected at an infrastructure base-station (BS), but from multiple user locations in the service area over time. CC opens up many location-based applications as it provides BS providers with relative user location information without access to GNSS or fingerprinting methods [15].

The technical concepts behind channel charting are as follows. Suppose that we have N single-antenna users located in real space with coordinates $\mathbf{z}_n \in \mathbb{R}^3$. If the nth user at location $\mathbf{z}_n$ is transmitting data to a BS, then the BS first extracts high-dimensional CSI in the form of a high-dimensional vector $\mathbf{h}_n \in \mathbb{C}^D$, which represents multi-path scattering and path loss of the wireless channel. From the CSI vector $\mathbf{h}_n$, one can extract features $\mathbf{x}_n \in \mathbb{R}^D$ that represent large-scale fading properties of the wireless channel. The main assumption of CC is that large-scale fading properties are mostly static and are strongly tied to user location. Specifically, due to the underlying physics of electromagnetic wave propagation, each CSI feature is a (noisy) function of the user position, a function that represents the effect of the (unknown) physical environment on the transmitted signal. One can then learn the channel chart from the set of channel features $\{\mathbf{x}_n\}_{n=1}^N$ in an unsupervised manner by means of dimensionality-reduction methods. If AEs are used in this procedure, the encoder f^e corresponds to the forward charting function that maps CSI features $\mathbf{x}_n$ to relative position information $\mathbf{y}_n$ in the representation space.

Reference [9] proposed the use of Sammon's mapping (SM) and AEs to learn the channel charts. While SM exhibited good performance, AEs scale well to large problem sizes and provide a parametric mapping that enables one to map new, unseen CSI features to a relative location. Despite the advantages of AEs, valuable side information that arises from the application itself has been ignored. First, in contrast to SM, conventional AEs do not enforce any geometric structure

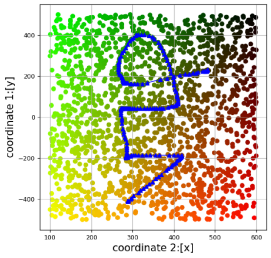


Fig. 1. Channel charting scenario as in [9]. A 32-antenna BS located at the origin extracts 2048 CSI vectors from several UEs in space. Each point represents a UE position; the points on the "vip" curve model a user that moves in time. The color gradients and "vip" curve enable an easy visual comparison with the learned channel charts shown in Figure 2.

on their representations. Second, by tracking a user's CSI over time, the corresponding low-dimensional representations that reflect their position should be similar as velocity is finite.

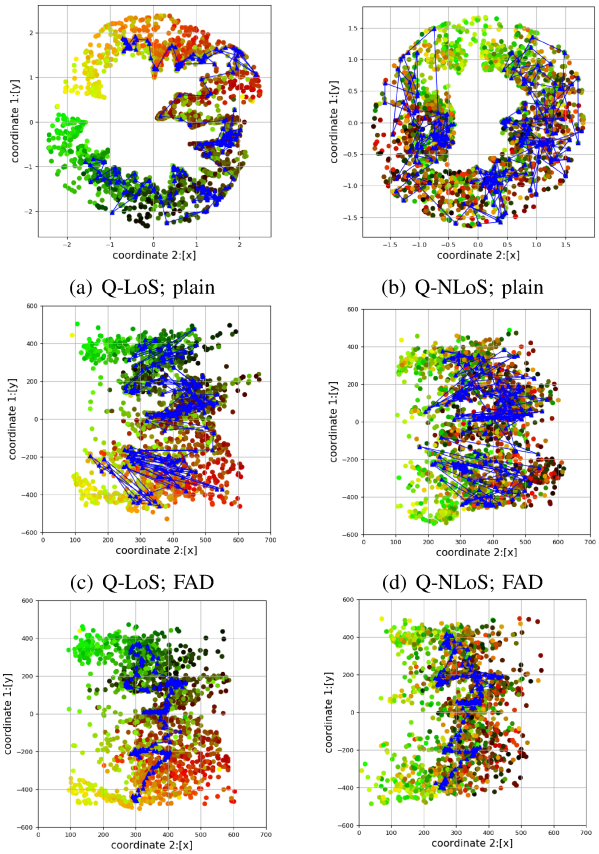
B. Channel Charting with Representation Constraints

We propose the inclusion of representation constraints to overcome the limitations of the original CC framework. We impose MRD constraints on pairs of representations from a user over time, to ensure nearby representations for nearby spatial locations. We can estimate an upper limit on the maximum distance in the representation space based on the measurement CSI acquisition rate. Note that this information comes directly from the CSI measurement process and the fact that we know how data was collected.

Furthermore, to enable CC with true positioning capabilities, we unwrap the channel chart using *anchor vectors*, i.e., points in space for which we know both their CSI as well as their true location. One can imagine measuring CSI at a small set of locations when setting up a new BS (e.g., by knowing the precise location of a fixed access point). With this information, we impose FAD representation constraints on the AE with $\underline{d}_{i,j} = 0$ to enforce the exact anchor positions. The inclusion of such constraints leads to a semi-supervised version of CC (and AEs in general). In contrast to conventional fingerprinting methods that are fully supervised and require training at wavelength resolution in space, we only require a small number of anchor vectors and use the rest of the (unlabeled) data to improve the localization accuracy of the channel chart.

C. Numerical Results

1) Scenario: Figure 1 depicts the scenario. We measure the CSI of $N = 2048$ randomly placed user locations (except for the positions on the "vip" curve) within a rectangular area of $1000\,\text{m} \times 500\,\text{m}$. We model the acquisition of CSI at $0\,\text{dB}$ SNR. At spatial location $(x, y, z) = (0, 0, 10)$ meters, we consider a uniform linear BS antenna array with half-wavelength spacing and 32 antennas. We model narrowband data transmission at $2\,\text{GHz}$ and consider two channel models: (i) Quadriga LoS (Q-LoS; a realistic model for LoS channels that includes scatterers and path loss); and (ii) Quadriga non-LoS (Q-NLoS; a realistic model for channels where there is no direct path between users and the BS antenna array). For both models, we used the "Berlin UMa" scenario, which has



(a) Q-LoS; plain (b) Q-NLoS; plain

(c) Q-LoS; FAD (d) Q-NLoS; FAD

(e) Q-LoS; FAD&MRD (f) Q-NLoS; FAD&MRD

Fig. 2. Channel charts learned from the scenario in Figure 1. The channel charts of plain AEs represent well the local neighborhood structure, but do not capture the global geometric properties. Imposing FAD and MRD constraints recovers the fine aspects of the geometry, enabling approximate positioning.

been calibrated with real-world measurements [26]. At the BS side, we extract the same features proposed in [9], i.e., we apply feature scaling, convert them to the angular domain, and take the entry-wise absolute value. The input has $D = 32$ real dimensions; the representation dimension is $D' = 2$. We use a similar structure of the AE in [9], namely 9 hidden dense layers: 4 dense layers for the encoder and another 4 dense layers for the decoder (the layers consist of 500, 100, 50 and 20 neurons), and an intermediate layer where we extract the channel chart with 2 neurons and linear activation functions.

2) Results: Figure 2 shows the channel charts. The top row (Figures 2(a)–(b)) shows the results of a "plain" AE, i.e., without any representation constraints; these results reproduce those in [9]. The middle row (Figures 2(c)–(d)) shows channel charts from AEs that include FAD constraints, where we randomly selected 10% of the users to be anchor vectors. Clearly, these FAD constraints unwrap the channel chart and lead to a representation with a distance scale comparable to the original scenario in Figure 1. While the global structure is approximately preserved, the points on the "vip" curve are not represented accurately. The bottom row (Figures 2(e)–(f)) shows the combination of FAD and MRD constraints in AEs. Concretely, we also enforce the fact that points on the "vip" curve model a user's motion and we can impose maximum

TABLE II
TW, CT, AND KS RESULTS FOR CHANNEL CHARTING WITH AND WITHOUT REPRESENTATION CONSTRAINTS.

		Q-LoS			Q-NLoS		
		Plain	FAD	FAD&MRD	Plain	FAD	FAD&MRD
TW	$K = 1$	0.8468	0.8516	**0.8576**	0.8480	0.8492	**0.8597**
	$K = 51$	0.8597	0.8570	**0.8651**	0.8502	0.8560	**0.8665**
	$K = 102$	0.8609	0.8642	**0.8736**	0.8546	0.8626	**0.8739**
CT	$K = 1$	**0.9700**	0.9195	0.9321	**0.9281**	0.8924	0.9110
	$K = 51$	**0.9440**	0.9067	0.9215	**0.9168**	0.8928	0.9128
	$K = 102$	**0.9358**	0.9081	0.9216	**0.9155**	0.8964	0.9151
KS		0.3548	0.2652	**0.2598**	0.4096	0.2749	**0.2693**

absolute distance constraints among pairs of representations pertaining to this curve. FAD and MRD combined are able to reproduce the original scenario: It is evident that points in the channel chart approximately represent the true locations. We also observe that the propagation conditions of the wireless channel do not substantially affect the performance.

Table II lists the TW, CT, and KS results for the channel charts shown in Figure 2. We see that including representation constraints improves TW while slightly lowering the CT; note that TW and CT are evaluated for $K = 1$, $K = 51$ (2.5% of the dataset), and $K = 102$ (5% of the dataset). Hence, we observe a tradeoff with respect to neighborhood-preserving properties. More concretely, an increase in TW means that we are introducing fewer "fake" near neighbors; a reduction in CT means that neighborhoods in the original space are not as well preserved in the channel chart as before. With respect to the global geometric structure, we see that KS significantly improves for all representation-constrained AEs; this implies that the inclusion of constraints enables us to recover global geometry. We note this is also visible in Figure 2, especially in the AE results that include both FAD constraints (anchor vectors) and MRD constraints (to enforce continuity of a user's motion). Once again, we see that the propagation conditions do not substantially affect the performance.

IV. CONCLUSIONS

We have shown how to incorporate pairwise representation constraints into autoencoders (AEs). To demonstrate effectiveness of representation-constrained AEs, we have shown an improvement to the channel charting (CC) framework in [9], where we use side information on user motion and anchor vectors to improve the positioning performance of CC. Numerical results for this application have shown that the use of representation constraints that are readily available in wireless positioning scenarios yield (often significant) improvements in recovered global geometry.

There are many opportunities for future work. While our methods enable approximate positioning, AEs with representation constraints are still not sufficient to enable GNSS-grade accuracy. Promising research directions towards this goal are the development of improved CSI features as well as the inclusion of additional geometric constraints, e.g., when acquiring CSI from multiple cell-towers or access points.

REFERENCES

[1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[2] L. van der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative review," in *J. Mach. Learn. Res.*, vol. 10, Oct. 2009, pp. 66–71.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[4] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Workshop Unsupervised Transfer Learn.*, vol. 27, Jul. 2012, pp. 37–49.

[5] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, Nov. 2015.

[6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Adv. Neural Inf. Process. Syst.*, Dec. 2013, pp. 3111–3119.

[7] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," *ArXiv*, Mar. 2017.

[8] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising auto-encoders as generative models," in *Adv. Neural Inf. Process. Syst.*, Dec. 2013, pp. 899–907.

[9] C. Studer, S. Medjkouh, E. Gönültaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, Aug. 2018.

[10] S. Medjkouh, C. Studer, E. Gönültaş, T. Goldstein, and O. Tirkkonen, "Unsupervised charting of wireless channels." submitted to GLOBECOM 2018, 2018.

[11] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[12] C. Doersch, "Tutorial on variational autoencoders," *ArXiv*, Jun. 2016.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *ArXiv*, Dec. 2013.

[14] T. D. Kulkarni, W. F. Whitney, P. Kohli, and J. Tenenbaum, "Deep convolutional inverse graphics network," in *Adv. Neural Inf. Process. Syst.*, Dec. 2015, pp. 2539–2547.

[15] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[16] Q. Xu, R. Zheng, W. Saad, and Z. Han, "Device fingerprinting in wireless networks: Challenges and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 94–104, 1st Quarter 2016.

[17] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Trans. Computers*, vol. C-18, no. 5, pp. 401–409, May 1969.

[18] M. L. Davison, *Multidimensional scaling*. Krieger Publishing, 1991.

[19] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor wi-fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[20] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," in *Proc. IEEE Intl. Symp. Personal, Indoor, Mobile Radio Commun.*, Oct. 2017, pp. 1–6.

[21] M. Arnold, S. Dörner, S. Cammerer, and S. ten Brink, "On deep learning-based massive MIMO indoor user localization," *ArXiv*, Apr. 2018.

[22] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org.

[23] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[24] J. A. Lee and M. Verleysen, "Quality assessment of dimensionality reduction: Rank-based criteria," in *Neurocomputing*, vol. 72, no. 7, Mar. 2009, pp. 1431–1443.

[25] R. N. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function. I." *Psychometrika*, vol. 27, no. 2, pp. 125–140, Jun. 1962.

[26] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.