DATA CLUSTERING USING MATRIX FACTORIZATION TECHNIQUES FOR WIRELESS PROPAGATION MAP RECONSTRUCTION

Junting Chen and Urbashi Mitra

Ming Hsieh Department of Electrical Engineering, University of Southern California
Los Angeles, CA 90089, USA, email:{juntingc, ubli}@usc.edu

ABSTRACT

This paper develops an efficient data clustering technique by transforming and compressing the measurement data to a low-dimensional feature matrix, based on which, matrix factorization techniques can be applied to extract the key parameters for data clustering. For the application of wireless propagation map reconstruction, a theoretical result is developed to justify that the feature matrix is a composite of several unimodal matrices, each containing key parameters for an individual propagation region. As a result, instead of iterating with N data points at each step, the proposed scheme provides a low complexity online solution for data clustering based on the feature matrix with dimension much smaller than N.

Index Terms— Clustering, matrix factorization, unimodal, nonparametric estimation

1. INTRODUCTION

This paper studies the problem of predicting the propagation from an autonomous vehicle to a static ground device in an unknown environment. As the terrain structure may block the propagation signal and divide the activity area of the autonomous vehicle into several propagation regions, it is crucial to build a propagation map that both recovers the pattern of the propagation region and makes a good prediction on the propagation channel gain. In general, propagation map reconstruction finds applications in various domains, such as sensor deployment for target monitoring in obstructive environment, relay placement for establishing dynamic communication infrastructure [1–3], and millimeter-wave communication, which desires a reliable line-of-sight (LOS) link [4–6].

There is some related work using machine learning techniques for the propagation map prediction, such as kernel regression for power map learning [7], support vector regression [8] and artificial neural network methods for path loss prediction [9]. These approaches are quite general and thus did not exploit the geometric characteristics of the propagation environment. As a result, they are not optimized for recovering the propagation pattern. The approach in [9] required labeled training data with manually classified propagation scenarios. Our prior work [10, 11] developed a segmented regression approach based on perfect knowledge of the ground terminal position. Realizing such a condition is challenging in practice.[1] More broadly, in [12, 13], a stochastic propagation map is learnt from a stochastic terrain model, which does not capture the actual propagation conditions.

This paper focuses on the problem of data clustering for several propagation scenarios based on a number of energy measurements taken randomly at different locations by an autonomous vehicle. Neither the terrain structure, the signal characteristics, nor the location of the ground device are known. Such a clustering problem is important because once the propagation scenarios are identified, classical regression techniques or other supervised learning techniques can be applied to learn and reconstruct the propagation map of the entire area. However, this problem is also challenging, as a brute-force solution from a classical regression approach may get stuck at a local optimum. Mathematically, it is a joint clustering (a combinatorial problem) and nonlinear regression problem.

In this paper, we develop an efficient data clustering technique that maps the measurement data to a low-dimensional *feature* matrix, where each entry corresponds to an estimate of a set of propagation parameters, and the value of the entry corresponds to the likelihood of the estimate. We prove that under some mild conditions, the feature matrix is a composite of several *unimodal* matrices. The peak location of each of those unimodal matrices corresponds to a close approximation of the propagation parameters in a corresponding propagation region. With this result in hand, our recent work on unimodal-constrained matrix factorization techniques [14] can be applied to extract the local peaks of each composite matrix, following which, the data clustering result can be easily obtained. As a result, instead of performing iterations to re-cluster N data points at each step as occurs in existing work, we employ a one-off pre-processing step to transform and compress the data into an $M \times M$ matrix, where the ma-

This research has been funded in part by one or more of the following grants: ONR N00014-15-1-2550, NSF CNS-1213128, NSF CCF-1718560, NSF CCF-1410009, NSF CPS-1446901, and AFOSR FA9550-12-1-0215.

[1] In practical cellular systems, for example, the network operator does not have direct access to the GPS data from a mobile.

trix stores sufficient information for data clustering. Therefore, the proposed method provides an efficient solution for online implementation, where N may grow unboundedly as sensors keep collecting measurements. Moreover, as the proposed data identification and clustering algorithm is based on simple signal characteristics, *i.e.*, aggregate energy rather than a multi-path power-delay profile, the method can be implemented using low-cost sensors.

2. SYSTEM MODEL

We assume that an autonomous vehicle V moves on a 2D plane $\mathcal{A}$ that is perpendicularly away from a static device U with distance $H \geq 0$. Denote $\mathbf{x} \in \mathbb{R}^2$ as the position of V on the 2D plane $\mathcal{A}$ and $\mathbf{x}_\mathrm{u} \in \mathbb{R}^2$ as the position of U *projected* on $\mathcal{A}$. The distance between V and U is defined as $d(\mathbf{x}, \mathbf{x}_\mathrm{u}) = \sqrt{\|\mathbf{x} - \mathbf{x}_\mathrm{u}\|^2 + H^2}$. For simplicity, the dependency on $\mathbf{x}_\mathrm{u}$ is omitted as it is fixed throughout the paper, *i.e.*, we denote $d(\mathbf{x}) = d(\mathbf{x}, \mathbf{x}_\mathrm{u})$.

2.1. Propagation Map

Suppose that there are objects surrounding the device U, and correspondingly, forming several disjoint propagation regions for the signals between V and U. Without loss of generality (w.l.o.g.), assume that there are $K = 2$ propagation regions. Specifically, denote $\mathcal{D}_1 \subseteq \mathcal{A}$ as the set of autonomous vehicle positions $\mathbf{x}$ such that there is a LOS link between V and U, and $\mathcal{D}_2 = \mathcal{A} \backslash \mathcal{D}_1$ as the set of positions $\mathbf{x}$ such that the LOS link is blocked.

In addition, assume that the propagation regions are *regular*, *i.e.*, for any $\mathbf{x} \in \mathcal{D}_k$ and $0 \leq \rho \leq 1$, it holds that $\mathbf{x}_\mathrm{u} + \rho(\mathbf{x} - \mathbf{x}_\mathrm{u}) \in \mathcal{D}_j$, $j \leq k$. The *regularity* condition captures the following physical characteristic: when V moves towards U, it only improves the propagation scenario as the number of objects that would block the signal decreases. On the other hand, when V moves aways from U, the propagation condition deteriorates as it is likely to lose the LOS link.

The propagation map is defined as

$$\gamma(\mathbf{x}) = \sum_{k=1}^{K} h_k(d(\mathbf{x}))\mathbb{I}\{\mathbf{x} \in \mathcal{D}_k\} \tag{1}$$

for every autonomous vehicle position $\mathbf{x} \in \mathcal{A}$, where $h_k(d)$ is the propagation function in terms of distance d in the kth region.

Assume that the propagation functions $h_k(d)$ follow a general parametric form $h(d; \alpha_1^{(k)}, \alpha_2^{(k)}) = \alpha_1^{(k)} p_1(\mathbf{x}) + \alpha_2^{(k)} p_2(\mathbf{x})$, $k = 1, 2$, where $p_k(\mathbf{x})$ are functions from a specific application scenario. For example, in radio signal propagation, a popular model is $h(d; \alpha_1^{(k)}, \alpha_2^{(k)}) = \alpha_1^{(k)} \log_{10}(d(\mathbf{x})) + \alpha_2^{(k)}$, where $p_1(\mathbf{x}) = \log_{10} d(\mathbf{x})$ and $p_2(\mathbf{x}) = 1$. For underwater acoustic signals, one can consider the model $h(d; \alpha_1^{(k)}, \alpha_2^{(k)}) = 1.5 \log_{10}(d) + \alpha_1^{(k)} d + \alpha_2^{(k)}$, where $p_1(\mathbf{x}) = d(\mathbf{x})$ and $p_2(\mathbf{x}) = 1$. For easy elaboration, we consider $p_2(\mathbf{x}) \equiv 1$ in this paper.

2.2. Measurement Model

Note that neither the parameters $\boldsymbol{\alpha}^{(k)} = (\alpha_1^{(k)}, \alpha_2^{(k)})$ of the functions $h_k(d)$, the propagation regions $\mathcal{D}_k$, nor the device position $\mathbf{x}_\mathrm{u}$ are known to the system. To estimate the propagation map $\gamma(\mathbf{x})$, $\mathbf{x} \in \mathcal{A}$, the autonomous vehicle V randomly samples N locations $\mathbf{x}_l$, where the corresponding the measurements y_l are modeled as

$$y_l = \gamma(\mathbf{x}_l) + n_l$$

where n_l is a zero-mean random variable with variance $\sigma^2 = \sum_{k=1}^{K} \sigma_k^2 \mathbb{I}\{\mathbf{x} \in \mathcal{D}_k\}$ depending on the propagation regions. The ultimate goal of this paper is to reconstruct the propagation map $\hat{\gamma}(\mathbf{x})$ based on the dataset $\{\mathbf{x}_l, y_l\}$.

3. CLUSTERING VIA MATRIX FACTORIZATION

Let $\hat{\mathbf{x}}_\mathrm{u} = \mathbf{x}_{\hat{m}}$ be an initial estimate of the position of the device U, where $\hat{m} = \arg\max_l\{y_l\}$, *i.e.*, the measurement location that observes the highest energy. The estimated distance between V and U is thus $\hat{d}(\mathbf{x}) \triangleq d(\mathbf{x}, \hat{\mathbf{x}}_\mathrm{u}) = \sqrt{\|\mathbf{x} - \hat{\mathbf{x}}_\mathrm{u}\|^2 + H^2}$. In addition, let $\mathbf{a}$ be an estimate of the parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ in the propagation model $h(d; \boldsymbol{\alpha})$. As a result, a predicted propagation channel gain at position $\mathbf{x}$ is given by $\hat{\gamma}(\mathbf{x}) = h(\hat{d}(\mathbf{x}); \mathbf{a}) = a_1 \hat{p}_1(\mathbf{x}) + a_2$, where $\hat{p}_1(\mathbf{x})$ is a function based on the estimated distance $\hat{d}(\mathbf{x})$.

3.1. Matrix Model for the Clustering Problem

Denote the prediction error on the training data $\{\mathbf{x}_l, y_l\}$ as

$$\begin{aligned} \triangle_l(a_1, a_2) &= y_l - h(\hat{d}(\mathbf{x}_l); \mathbf{a}) \\ &= \delta_l(a_1, a_2) + n_l \end{aligned}$$

in which, $\delta_l(a_1, a_2) \triangleq (\alpha_1 - a_1)p_1(\mathbf{x}) + (\alpha_2 - a_2) + a_1(p_1(\mathbf{x}) - \hat{p}_1(\mathbf{x}))$, where the first two terms represent the prediction error due to parameter estimation, and the third term represents the bias due to localization error on the device U.

To formulate a data clustering problem, define a value function as

$$v(\mathbf{a}) = \sum_{l=1}^{N} L(\triangle_l(a_1, a_2)) \tag{2}$$

in which, the *similarity* metric $L(z)$ is chosen as $L(z) = e^{-\lambda z^2}$, where the choice of parameter $\lambda > 0$ will become clear in Proposition 1 below. In the special case of one propagation region $K = 1$ (LOS link exists everywhere), one can obtain the best parameter $\mathbf{a}$ by solving a least-squares regression problem using $L(z) = -z^2$. However, the case for

$K \geq 2$ regions is highly non-trivial, especially when the bias term $a_1(p_1(\mathbf{x}) - \hat{p}_1(\mathbf{x}))$ appears in $\delta_l(a_1, a_2)$.

Let $\mathcal{C}_t = \{-\infty < c_{t1} < c_{t2} < \cdots < c_{tM_t} < \infty\}$, $t = 1, 2$, be two sequences of candidate values to approximate the parameters $\alpha_1^{(k)}$ and $\alpha_2^{(k)}$, for $k = 1, 2$. Construct a *feature matrix* $\mathbf{V} \in \mathbb{R}^{M_1 \times M_2}$ by specifying its (i, j)th entry as

$$V_{ij} = \sum_{l=1}^{N} L(\triangle_l(c_{1i}, c_{2j})). \tag{3}$$

Intuitively, the entry V_{ij} has a large value if the pair (c_{1i}, c_{2j}) closely approximates the unknown parameters $(\alpha_1^{(k)}, \alpha_2^{(k)})$ for either $k = 1, 2$.

To obtain an analytical insight on the feature matrix $\mathbf{V}$, consider the (unknown) subset of data $\mathcal{L}_k = \{l : \mathbb{I}\{\mathbf{x}_l \in \mathcal{D}_k\} = 1\}$ that belongs to the kth propagation region. Suppose that a matrix $\mathbf{V}^{(k)} \in \mathbb{R}^{M_1 \times M_2}$ is formed by assigning its (i, j)th entry with $V_{ij}^{(k)} = \sum_{l \in \mathcal{L}_k} L(\triangle_l(c_{1i}, c_{2j}))$. It turns out that the matrix $\mathbf{V}^{(k)}$ has a *unimodal* property that is shown as follows.

A vector $\mathbf{u} = (u_1, u_2, \ldots, u_M)$ is *unimodal* if it is non-negative and has a single peak, *i.e.*, $0 \leq u_1 \leq u_2 \leq \cdots \leq u_s$ and $u_s \geq u_{s+1} \geq \cdots \geq u_M$ for some integer $1 \leq s \leq M$. A matrix is unimodal if all its row vectors and column vectors are unimodal. With these notions, we establish the following result.

Proposition 1 (Unimodality)**.** *Suppose that the distribution of* $\mathsf{z} = p_1(\mathsf{x})$ *has a non-negative and bounded support, where* x *is a random variable of the position distributed over* $\mathcal{D}_k$*. In addition, assume that the measurement noise is Gaussian distributed. Then, under a large data set* $\mathcal{L}_k$ *and a small enough parameter* $\lambda > 0$ *in* (2)*, the matrix* $\mathbf{V}^{(k)}$ *is unimodal.*

Proof. See the appendix. □

It has been shown in [14] that if a matrix is unimodal, then its dominant left and right singular vectors are also unimodal, and their peaks appear at the locations that approximate the underlined parameters to be estimated. Proposition 1 suggests that one may choose a not-so-large parameter λ, such that $\mathbf{V}^{(k)}$ are unimodal so that their peaks appear at the desire locations approximating $(\alpha_1^{(k)}, \alpha_2^{(k)})$. On the other hand, one may choose a not-so-small parameter λ, such that the data points $l \in \mathcal{L}_k$ have negligible values outside the neighborhood of the peak of the matrix $\mathbf{V}^{(k)}$.

As a result, the feature matrix $\mathbf{V}$ can be modeled as

$$\mathbf{V} = \gamma_1 \mathbf{u}_1 \mathbf{w}_1^{\mathrm{T}} + \gamma_2 \mathbf{u}_2 \mathbf{w}_2^{\mathrm{T}} + \mathbf{N} \tag{4}$$

where $\gamma_1, \gamma_2 > 0$, $\mathbf{u}_k \mathbf{w}_k^{\mathrm{T}}$ are rank-1 approximations of $\mathbf{V}^{(k)}$, and the vectors $\mathbf{u}_k$ and $\mathbf{w}_k$ are unit-normed *signature* vectors. Denote m_1 and m_2 as the peak positions of $\mathbf{u}_k$ and $\mathbf{w}_k$, respectively. Then, we obtain the estimates $\hat{\alpha}_1^{(k)} = c_{1,m_1}$ and $\hat{\alpha}_2^{(k)} = c_{2,m_2}$. The matrix $\mathbf{N}$ captures the residual noise, which is believed to be small under a good choice of parameter λ and the sets of sequences $\mathcal{C}_t$, $t = 1, 2$.

3.2. Matrix Factorization

The signature vectors $\mathbf{u}_k$ and $\mathbf{w}_k$, $k = 1, 2$, can be obtained from the solutions $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$ and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$ of the following problem

$$\begin{aligned} \mathscr{P}: \quad & \underset{\mathbf{U},\mathbf{W}}{\text{minimize}} \quad \|\mathbf{S} \odot (\mathbf{V} - \mathbf{U}\mathbf{W}^{\mathrm{T}})\|_{\mathrm{F}}^2 \\ & \text{subject to} \quad \mathbf{U} \in \mathcal{U}^{M_1 \times 2}, \mathbf{W} \in \mathcal{U}^{M_2 \times 2} \end{aligned}$$

where $\mathcal{U}^{M_k \times 2}$ denotes the set of $M_k \times 2$ matrices with unimodal columns; $\mathbf{S}$ is an $M_1 \times M_2$ indicator matrix with $S_{ij} = 1$ if the (i, j)th entry of $\mathbf{V}$ is available, and $S_{ij} = 0$, otherwise. The symbol $\odot$ denotes the Hadamard product. Other formulations and solvers, such as the l_p-norm minimization [15, 16], can also be applied.

Problem $\mathscr{P}$ can be solved using projected gradient methods [14] and the parameter estimates $\hat{\alpha}_t^{(k)}$, $t, k = 1, 2$, are obtained from peak localization [14] of $\mathbf{u}_k$ and $\mathbf{w}_k$.

3.3. Data Clustering and Localization

From (3) and (4), the signature vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ capture the empirical distribution of $L(\mathsf{y} - h(\hat{d}(\mathbf{x}); \hat{\boldsymbol{\alpha}}^{(k)}))$. Therefore, $\mathbf{w}_1$ and $\mathbf{w}_2$ can be used for non-parametric hypothesis testing for data clustering.

Specifically, let $w_1(c)$ and $w_2(c)$ be a linear interpolation of $\mathbf{w}_1$ and $\mathbf{w}_2$, respectively, over $c \in (-\infty, \infty)$. Then, a data point $\{\mathbf{x}_l, y_l\}$ is clustered to region $\mathcal{D}_1$ if

$$w_1\Big(L\big(y_l - h(\hat{d}(\mathbf{x}_l); \hat{\boldsymbol{\alpha}}^{(1)})\big)\Big) > w_2\Big(L\big(y_l - h(\hat{d}(\mathbf{x}_l); \hat{\boldsymbol{\alpha}}^{(2)})\big)\Big) \tag{5}$$

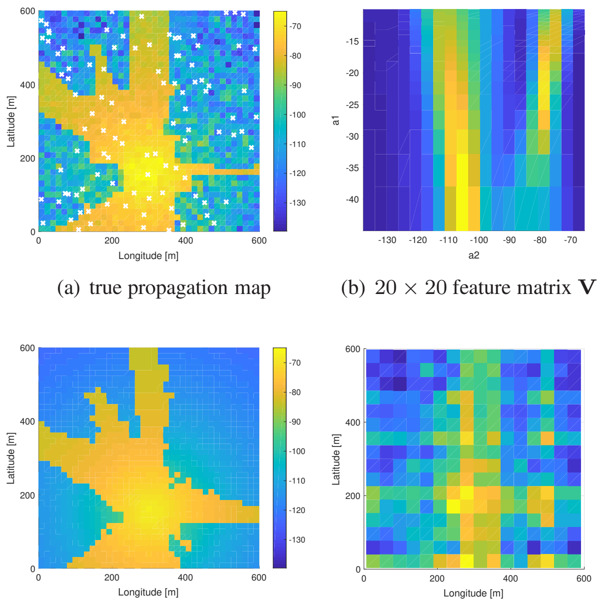
and it is clustered to region $\mathcal{D}_2$, otherwise.

3.4. Comparison with Classical Methods

The joint data clustering and parameter regression problem can also be solved by formulating a maximum likelihood estimation problem and obtaining the solution using expectation-maximization (EM) algorithms with randomized initializations. However, each EM iteration needs to compute a new clustering for all N data points, where N is usually very large. In the proposed technique, the clustering is performed via factorizing a $M_1 \times M_2$ feature matrix that compresses all the data points, and the complexity scales with $M_1^2 + M_2^2$ which is much smaller than N. The data clustering that involves all N data points is performed only twice (before and after factorizing the feature matrix $\mathbf{V}$).

4. NUMERICAL RESULTS

Consider a unmanned aerial vehicle (UAV) flies in a 600 [m] $\times$ 600 [m] area with 50 meter height above ground. It mea-



(a) true propagation map (b) 20×20 feature matrix $\mathbf{V}$

(c) proposed reconstruction (d) matrix completion

Fig. 1. Reconstruction of the noise corrupted radio map in (a) from 100 random samples (white crosses): The proposed method recovers the propagation details in (c), while a matrix completion baseline fails to recover the propagation structure, shown in (d).

sures the received signal from a ground device with unknown position, and the received signal energy y in dB at UAV position $\mathbf{x}$ is modeled as $y = \alpha_1 \log_{10} d(\mathbf{x}) + \alpha_2 + n$, where $\boldsymbol{\alpha}^{(1)} = (-22, -28)$, $n \sim \mathcal{N}(0, 1)$, if the link is in LOS condition, and $\boldsymbol{\alpha}^{(2)} = (-36, -22)$, $n \sim \mathcal{N}(0, 8^2)$ otherwise. The full propagation map and a random sample pattern with 100 measurements are shown in Fig. 1 (a). The corresponding feature matrix in (3) is demonstrated in Fig. 1 (b). The clustering $\hat{z}_l \in \{1, 2\}$ of the sample data $\{\mathbf{x}_l, y_l, l = 1, 2, \dots, 100\}$ is obtained by solving $\mathscr{P}$ followed by peak localization of the signature vectors in (4) and hypothesis testing in (5).

For propagation map reconstruction, we employ a k-nearest neighbor (KNN) algorithm to classify every position $\mathbf{x}$ to one of the two propagation regions, $Z(\mathbf{x}) \in \{1, 2\}$, based on the clustered dataset $\{\mathbf{x}_l, \hat{z}_l\}$. We then reconstruct the propagation map using $\hat{\gamma}(\mathbf{x}) = \hat{\alpha}_1^{(k)} \log_{10} \hat{d}(\mathbf{x}) + \hat{\alpha}_2^{(k)}$, $k = Z(\mathbf{x})$. As a benchmark, a baseline scheme first arranges the measurements to a 16×16 matrix $\mathbf{Y}$, and completes the matrix by minimizing the nuclear norm of a matrix variable $\mathbf{X}$ subject to $\sum_{(i,j)\in\Omega} |X_{ij} - Y_{ij}|^2 \le \epsilon^2$, where Ω is the set of entries in $\mathbf{Y}$ that are observed from measurement samples and ϵ is chosen to minimize the reconstruction error (*i.e.*, a genius-aided approach) [17]. Fig. 1 (c) and (d) show that the proposed scheme succeeds in recovering the propagation pattern while the baseline fails to do so due to large measurement noise.

5. CONCLUSIONS

This paper developed an efficient data clustering technique by transforming the measurement data to a feature matrix with a focus on the application of propagation map reconstruction. It was proven that under some mild conditions, the feature matrix is a composite of several unimodal matrices, each containing key parameters of an individual propagation region. With that, recent matrix factorization techniques can be applied to extract the sets of propagation parameters without processing the whole dataset. Numerical experiments demonstrated that the proposed method succeeded in recovering the propagation pattern, whereas, a naive matrix completion baseline failed to do so due to large measurement noise.

Appendix: Proof of Proposition 1

Ergodicity implies that $\frac{1}{N_k} V_{ij}^{(k)} \to \mathbb{E}_{\mathbf{x},n}\{L(\triangle_l(a_1, a_2))\}$ in probability, as $N_k = |\mathcal{L}_k| \to \infty$, where the expectation is taken over both the measurement noise and random position $\mathbf{x}$ in $\mathcal{D}_k$. On the other hand,

$$\bar{v}_k(a_1, a_2) \triangleq \frac{1}{N_k} \sum_{l \in \mathcal{L}_k} \mathbb{E}\{L(\triangle_l(a_1, a_2))\}$$
$$\to \mathbb{E}_{\mathbf{x},n}\{L(\triangle_l(a_1, a_2))\} \to \frac{1}{N_k} V_{ij}^{(k)}$$

in probability, where $\mathbb{E}\{\cdot\}$ is taken over measurement noise only. Lemma 1 below shows that $\bar{v}_k(a_1, a_2)$ is concave, and hence, there is a single local maximum over a_1 (under a fixed a_2) and a single local maximum over a_2 (under a fixed a_1). As $\mathbf{V}^{(k)}$ is a discretization of $\bar{v}_k(a_1, a_2)$ over grid points specified by the sets $\mathcal{C}_t$, $t = 1, 2$, $\mathbf{V}^{(k)}$ is unimodal.

Lemma 1 (Concavity)**.** *Under the same conditions in Proposition 1, the function* $\bar{v}_k(a_1, a_2) \triangleq \frac{1}{N_k} \mathbb{E}\{\sum_{l \in \mathcal{L}_k} L(\triangle_l)\}$ *is concave in* $a_1 \in A_1$ *and* $a_2 \in A_2$*, respectively, where* A_1 *and* A_2 *are bounded intervals.*

Proof. (Sketch) A computation on the expectation yields

$$\bar{v}_k(a_1, a_2) = \frac{1}{N_k \sqrt{\mu}} \sum_{l \in \mathcal{L}_k} \exp\left\{ -\lambda_0 \delta_l(a_1, a_2)^2 \right\}$$

where $\mu = 2\lambda\sigma_k^2 + 1$ and $\lambda_0 = \lambda / \mu$. The second order derivative is given by

$$\frac{\partial^2 \bar{v}_k}{\partial a_1^2} = \frac{2}{N_k \sqrt{\mu}} \sum_{l \in \mathcal{L}_k} e^{-\lambda_0 \delta_l^2} \lambda_0 p_1(\mathbf{x}_l)^2 (2\lambda_0 \delta_l^2 - 1)$$

which is negative, if $\delta_l^2 < \frac{1}{2\lambda_0} = \frac{1}{2\lambda} + \sigma_k^2$, for all $l \in \mathcal{L}_k$. Note that δ_l^2 is upper bounded with probability 1 due to the assumption on the finite support of $p_1(\mathbf{x})$ and the finite intervals A_1 and A_2. Therefore, there exists a small enough $\lambda > 0$, such that the condition is satisfied for all data points. Similar argument can be applied to $\frac{\partial^2 \bar{v}_k}{\partial a_2^2} < 0$. With that, the concavity result is proven. $\square$

6. REFERENCES

[1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, 2016.

[2] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3949–3963, 2016.

[3] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeter-wave communication: Potentials and approaches," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 66–73, 2016.

[4] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, 2015.

[5] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[6] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, 2014.

[7] D. Romero, S.-J. Kim, G. B. Giannakis, and R. Lopez-Valcarce, "Learning power spectrum maps from quantized power measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2547–2560, 2017.

[8] R. D. Timoteo, D. C. Cunha, and G. D. Cavalcanti, "A proposal for path loss prediction in urban environments using support vector regression," in *Proc. Advanced Int. Conf. Telecommun*, 2014, pp. 1–5.

[9] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, 2010.

[10] J. Chen, U. Yatnalli, and D. Gesbert, "Learning radio maps for UAV-aided wireless networks: A segmented regression approach," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017.

[11] J. Chen, O. Esrafilian, D. Gesbert, and U. Mitra, "Efficient algorithms for air-to-ground channel reconstruction in UAV-aided communications," in *Proc. IEEE Global Telecomm. Conf.*, Dec. 2017, Wi-UAV workshop.

[12] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Proc. IEEE Global Telecomm. Conf.*, 2014, pp. 2898–2904.

[13] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proc. IEEE Global Telecomm. Conf.*, 2015, pp. 1–6.

[14] J. Chen and U. Mitra, "Matrix factorization for nonparametric multi-source localization exploiting unimodal properties," 2017, submitted to *IEEE Trans. Signal Process.*, preprint arXiv:1711.07457.

[15] X. Jiang, Z. Zhong, X. Liu, and H. C. So, "Robust matrix completion via alternating projection," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 579–583, 2017.

[16] W.-J. Zeng and H. C. So, "Outlier-robust matrix completion via l_p-minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1125–1140, 2018.

[17] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Commun. of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.