

A METHOD TO RECONSTRUCT COVERAGE LOSS MAPS BASED ON MATRIX COMPLETION AND ADAPTIVE SAMPLING

Symeon Chouvardas, Stefan Valentin, Moez Draief and Mathieu Leconte

Mathematical and Algorithmic Sciences Lab,
Huawei France R&D,
Paris, France.

{symeon.chouvardas, stefan.valentin, moez.draief, mathieu.leconte}@huawei.com

ABSTRACT

Accurate coverage maps are an important tool for network planning and operation but it is often impossible to obtain these maps completely from measurements. In this paper we describe two new methods that enable operators to minimize the cost for obtaining a complete coverage map at high accuracy. Our first method applies the Singular Value Thresholding (SVT) algorithm to reconstruct a complete map from a sparse matrix of coverage data. We then use the Query by Committee (QbC) rationale to identify the areas where further measurements would maximize accuracy of the completed map. This second method allows operators to plan their drive tests such that a given budget is spent at highest efficiency. Our numerical examples illustrate that our proposed completion technique outperforms relevant state of the art and that QbC further enhances reconstruction accuracy.

Index Terms— coverage maps, radio measurements, drive tests, matrix completion, adaptive sampling

1. INTRODUCTION

The planning and operation of cellular networks relies to a large degree on accurate coverage maps. For instance, coverage maps are fundamental for:

- **Network planning:** For the initial roll-out and to identify coverage problems during operation.
- **Self-Optimizing Network (SON) functions such as:** Dynamic tilting, optimization of handover, admission control and radio resource management (RRM) parameters.
- **New optimization approaches such as:** Anticipatory resource allocation and pro-active handover.

A coverage map is often given as a set of radio measurements over discrete geographical coordinates and is typically obtained by drive tests. However, drive tests are costly, since they require substantial personal, time and equipment to accurately cover an area of sufficient scale. This cost poorly scales with the area size, which means that it grows vastly, the larger the studied area becomes.

As a consequence, methods were developed to reduce the number of drive tests by collecting measurements from the mobile devices in a cellular network [1]. A relatively cost efficient way to do so is to exploit information from so-called crowd-sourcing applications. With crowd-sourcing, a user installs an application on an off-the-shelf Smartphone and returns measurements to a database. The immediate drawbacks of this approach are:

- Increased signaling overhead, which reduces the users' data budget and the battery lifetime of the mobile device.

- Systematic measurement errors due to the wide fluctuation of the involved Smartphone functions, e.g., radio front-end, different filtering algorithms, as well as methods and chipsets for localization.
- Possible manipulation of the database by sending wrong measurements.

As a result, data coming from crowd-sourcing can be unreliable and can lead to erroneous coverage maps. This increases the need for alternative ways to reduce the reconstruction cost. To this end, we will focus on the cost-efficient reconstruction of pathloss maps in this paper.

Related Work: Accurately characterizing pathloss in wireless networks is an ongoing challenge [2]. Over the past years, pathloss models in various types of networks have been proposed and analyzed. Despite the fact that pathloss modeling is useful in many applications, the deviation between the pathloss, measured in a real propagation environment, and the one given by the model can be large [3]. For this reason, learning-based reconstruction has been extensively studied since it can be tailored to the specific environment under consideration and give more accurate results. Such measurement-based learning approaches, can be divided in two categories:

- **Batch Algorithms:** Assume the complete data to be available before performing the reconstruction algorithm. Reconstruction based on Support Vector Machines, Gaussian processes and Kriging-based techniques have been proposed in [4, 5, 6]. Low rank matrix completion has also been employed in the context of indoor positioning, target detection and data collection in Wireless Sensor Networks, e.g., [7, 8, 9]
- **Online Algorithms:** Do not require the complete data set to be available in advance. They assume that the data is received sequentially, once per iteration step, and update the mapping function dynamically. In [10], the measurements are received sequentially and the predictor is improved via a kernel-based approach.

Contributions: We propose a new method which aims at reducing the costs to reconstruct coverage maps, while maintaining high accuracy. This is achieved by: a) reconstructing accurately the pathloss map, by choosing an effective solution from the algorithmic family of matrix completion, b) identifying areas where measurements maximize the accuracy of the map reconstruction so that to obtain new samples from these. For the reconstruction, we represent the map as a sparse matrix having a small number of observed entries and we complete it using the Singular Value Thresholding (SVT) algorithm. For the identification of the informative areas we follow the QbC rationale. Numerical examples illustrate that the proposed

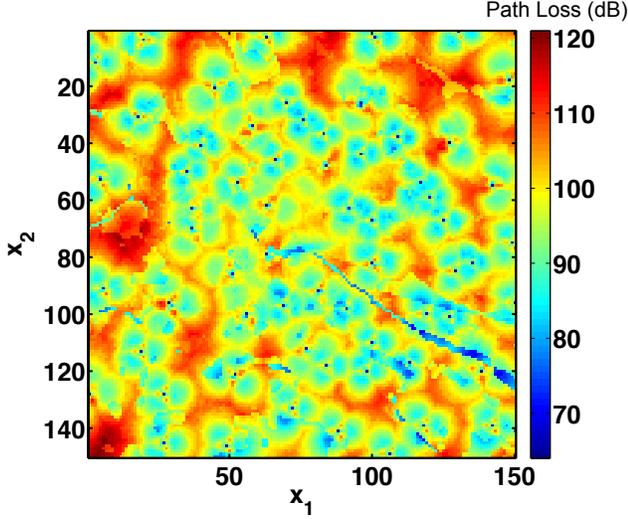


Fig. 1. Example of a coverage map for 7500 m² of downtown Berlin, Germany [11]: Pathloss of strongest serving cell in dB, where x_1 and x_2 stand for the discrete geographical coordinates.

reconstruction technique outperforms a relevant state of the art technique and that the QbC enhances the reconstruction results.

2. THE MATRIX COMPLETION PROBLEM

We consider a geographical area, in which a set of measurements in discrete locations is available to a mobile operator. These measurements could possibly represent received signal strength, interference, data rate, loss rate, anomalous events, quality of service indicators, etc. Here, we focus on pathloss measurements and we are interested in the problem of reconstructing pathloss maps. An example of such a map is illustrated in Fig. 1. We pose the aforementioned problem as a matrix completion task, where we represent the area under consideration as a matrix, the entries of which correspond to the pathloss, and the cells corresponding to the locations.

The task of matrix completion (MC), e.g., [12, 13], is the recovery of a data matrix from a sample of its entries. Formally, given a matrix \mathbf{P} of dimension $m \times n$ we have access to $k \ll m \cdot n$ entries and the goal is the prediction of the rest unobserved ones. It has been shown that under certain conditions this can be achieved [14, 12]. Intuitively, MC builds upon the observation that if a certain matrix is structured, in the sense that it is of low rank or of approximate low rank, then it can be recovered exactly, under some mild assumptions regarding the positions of the observed entries. This observation will be our starting point for the reconstruction of the pathloss map.

Due to the regular propagation of a radio wave in unobstructed environments, pathloss maps exhibit spatial correlation and smooth patterns¹. Hence, they can be well approximated by low rank matrices and a natural choice is to resort to the family of MC algorithms. So, in our problem we first represent the coverage map, of the geographical area in which we are interested, as a matrix, say $\mathbf{P} \in \mathbb{R}^{m \times n}$, with missing entries. This matrix, is used to represent the physical space, where each cell corresponds to a physical position of the spatial space. Furthermore, the value of the coefficient can

¹In practice, there are some exception to this rule especially in obstructed environments, such as cities. These situations will be discussed in Section 3.

be either zero, if the pathloss is unobserved at this position, or equal to the value of the pathloss measurement when this is available, e.g., via a drive test. The problem of estimating the unobserved entries of the matrix can be summarized as follows: Compute a matrix, \mathbf{A} , which will be of low rank and equal to the observation matrix \mathbf{P} in the set of observed entries, say Ω ; that is $A_{ij} = P_{ij}, \forall i, j \in \Omega$, where P_{ij} , A_{ij} is the i, j -th entry of \mathbf{P} and \mathbf{A} respectively. A way to do so is to solve the following problem:

$$\min_{\mathbf{A}} \text{rank}(\mathbf{A}) \quad (1)$$

$$s.t. A_{ij} = P_{ij}, \forall i, j \in \Omega. \quad (2)$$

Unfortunately, the rank minimization problem described previously cannot be solved efficiently, since it is NP-hard [12]. However, it has been shown, [14], that this problem can be relaxed, solved efficiently via convex optimization and the resulting solution will be close to the optimum one. The relaxation of the initial problem can be written as follows:

$$\min_{\mathbf{A}} \|\mathbf{A}\|_* \quad (3)$$

$$s.t. A_{ij} = P_{ij}, \forall i, j \in \Omega, \quad (4)$$

where $\|\mathbf{A}\|_*$ denotes the nuclear norm of the matrix \mathbf{A} with definition: $\|\mathbf{A}\|_* = \sum_{k=1}^{\min(m,n)} \sigma_k(\mathbf{A})$, with $\sigma_k(\cdot)$ being the k -th larger singular value. Several techniques have been proposed in the literature to solve the optimization (3), (4), including Semi Definite Programming [12], projection based techniques [15], just to name a few. Here we follow an iterative MC algorithm, which is known as *Singular Value Thresholding*, originally proposed in [16]. Starting from an initial zero matrix $\mathbf{Y}_0 = \mathbf{O} \in \mathbb{R}^{m \times n}$ the following steps take place in each iteration of the algorithm:

$$\mathbf{A}_i = \text{shrink}(\mathbf{Y}_{i-1}, \tau) \quad (5)$$

$$\mathbf{Y}_i = \mathbf{Y}_{i-1} + \mu P_{\Omega}(\mathbf{P} - \mathbf{A}_i) \quad (6)$$

where μ is a non-negative step size, $P_{\Omega}(\mathbf{X}), \forall \mathbf{X} \in \mathbb{R}^{m \times n}$ is a sampling operator $\mathbb{R}^{m \times n} \mapsto \mathbb{R}^{m \times n}$ associated with the set Ω , which assigns entries of its matrix argument $\notin \Omega$ equal to zero, and keeps the rest unchanged. Finally, $\text{shrink}(\cdot, \tau)$ is a rank reduction thresholding function, setting zero the singular values which are below a certain threshold determined by the parameter $\tau > 0$. A typical example of a thresholding function is the soft thresholding one with definition: is:

$$\forall \mathbf{X} \in \mathbb{R}^{m \times n} : \text{shrink}(\mathbf{X}, \tau) = \mathbf{U}\mathcal{T}(\boldsymbol{\Sigma})\mathbf{V}^T$$

where $\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}$ is the output of the regular Singular Value Decomposition algorithm and the operator $\mathcal{T}(\cdot)$ applied to the diagonal matrix $\boldsymbol{\Sigma}$ applies the following operation to each component of the diagonal: $\max(0, \sigma_k - \tau), k = 1, \dots, \min\{m, n\}$.

The intuition behind the SVT algorithm can be summarized as follows. The thresholding operator forces some singular values to zero and the resulting matrix is of low rank. In the sequel, step (4) “brings close” the estimated matrix and the matrix \mathbf{P} in the set of observed entries. The steps of the pathloss reconstruction algorithm are summarized in Table 1. The parameter ϵ is a user defined threshold for the termination of the algorithm.

3. IDENTIFYING THE MOST INFORMATIVE AREAS

Let us now focus on the estimation of the informative areas; that is areas from which we want to have samples, since such knowledge

Table 1. SVT algorithm for completing the pathloss map

Initialize: Incomplete matrix of pathloss measurements \mathbf{P} , $\mathbf{Y}_0, \tau > 0, \mu > 0, \epsilon > 0$
WHILE $\|\mathbf{A}_i - \mathbf{A}_{i-1}\| > \epsilon$
DO
 Update \mathbf{A}_i via (5)
 Update \mathbf{Y}_i via (6)
ENDWHILE

can improve the pathloss reconstruction. Note that some regions can be non-smooth. This is the consequence of large buildings, obstacles, tunnels that can abruptly attenuate the propagating radio wave. Due to these, the pathloss in such areas, exhibits low spatial correlation and this can lead to poor reconstruction effects. Consequently, in this case targeted measurements are required. In order to identify these regions, which in our case correspond to entries of the matrix, we resort to the family of the *active learning* algorithms, e.g., [17, 18], and in particular we employ the *Query by Committee* rationale, e.g., [18].

The general idea is to quantify the uncertainty of prediction of each missing entry in the matrix and to obtain, afterward, measurements from the most uncertain ones. To achieve this, we apply the QbC method. Initially, we reconstruct the area or equivalently, we complete the sparse matrix, using several different algorithms and using a subset of the available measurements. In our application, assuming that our available budget corresponds to k measurements, coming from drive tests, we first complete the matrix using a number of $l < k$ observed entries. Subsequently, having access to a number of reconstructed matrices we find the top $K := k - l$ entries with the largest “disagreement” according to a certain criterion and we obtain measurements from them. Finally, we perform drive tests to obtain the K samples indicated by the previous step and we reconstruct the pathloss map exploiting the newly obtained information.

In general, one can employ any number of MC algorithms to reconstruct the matrix. Here, we employ the SVT and two more schemes, i.e., the K Nearest Neighbors (KNN) and the Kernel Adaptive Projected Subgradient Method (KAPSM), which are described next.

- **KNN:** The KNN method identifies the K columns, which are closest to the one containing a missing value and uses the average of them as a guess for the missing entry, [19].
- **KAPSM:** This algorithm, proposed in [10], attempts to fit a nonlinear function between the geographical location and the pathloss. Next, this function is used to predict the channel gain in a certain position. The KAPSM algorithm is online, i.e., the data is received sequentially, once per iteration step, and the unknown function is obtained via kernel adaptive filtering.

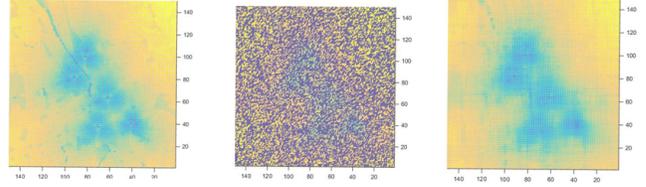
These three algorithms run in parallel using the same set of measurements as an input. Their output is a complete matrix. After the estimation of the missing entries we obtain the entries with the largest disagreement according to the following simple rule. Denoting by $a_{ij}^{(\zeta)}$, $\zeta = 1, 2, 3$, the predicted entry ij of algorithm ζ the disagreement equals to

$$d_{ij} = (a_{ij}^{(1)} - a_{ij}^{(2)})^2 + (a_{ij}^{(2)} - a_{ij}^{(3)})^2 + (a_{ij}^{(1)} - a_{ij}^{(3)})^2. \quad (7)$$

The K entries, which score the largest disagreement, are chosen and we perform drive tests to obtain the pathloss. The steps of the algorithm are summarized in Table 2.

Table 2. The QbC method for identifying informative areas

Initialize: Incomplete matrix of pathloss measurements \mathbf{P} , available budget k , number of entries for initial reconstruction l
1: Employ the SVT, the KNN and the KAPSM to compute the missing entries $a_{ij}^{(\zeta)}$, $\zeta = 1, 2, 3$ using $k - l$ measurements.
2: Compute the K entries with the largest disagreement w.r.t. (7)
3: Obtain the pathloss in the areas corresponding to the K entries
4: Reconstruct the pathloss map using the newly identified entries

**Fig. 2.** Illustration of a) the original pathloss map, b) the sparse one with the missing entries, c) the reconstructed one via the SVT

4. SIMULATIONS

In this section, we present numerical examples in order to test the SVT algorithm as well as the QbC method, in the context of pathloss reconstruction, in a real cellular network. Specifically, we consider a pathloss map of Berlin, Germany, originating from the data of the MOMENTUM project [11] and we reconstruct pathloss within an area of 7500 m². The size of each pixel, i.e., entry of the matrix, equals to 50 × 50 m and, consequently, the dimension of the matrix we want to reconstruct is 150 × 150. The total number of Base Stations (BSs) collecting pathloss measurements from the users equals to 187 and we take into consideration the BS with the strongest signal to fill the observed entries of the matrix. Furthermore, the adopted performance metric is the Normalized MSE with the following definition:

$$\text{NMSE} = \frac{\|\hat{\mathbf{A}} - \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2} \quad (8)$$

where \mathbf{H} stands for the Berlin pathloss matrix comprising all the coefficients, $\hat{\mathbf{A}}$ is the reconstructed matrix and $\|\cdot\|_F$ is the Frobenius norm.

In the first experiment we apply the SVT algorithm to the Berlin Map (Fig. 2). In particular, we reconstruct the matrix having access to 5000 entries out of the total number of 22500. The step-size equals to $\mu = 1.5$, and the parameter τ is computed via cross-validation, e.g., [20]. Finally, the parameter ϵ is set equal to 10^{-5} . Fig. 2 illustrates the original pathloss map, the one with the missing entries and the reconstructed one. It can be readily seen that the reconstructed map approaches the original one, despite the large number of unobserved entries.

In the second experiment, we compare the performance of the SVT with that of the KAPSM [10] for different values of observed entries. For the former algorithm the parameters are the same as in the previous experiment and for the latter the involved parameters are optimized in the sense that the algorithm reaches the lowest error floor after convergence. Fig. 3 illustrates the performance of the two reconstruction algorithms. The horizontal axis represents the num-

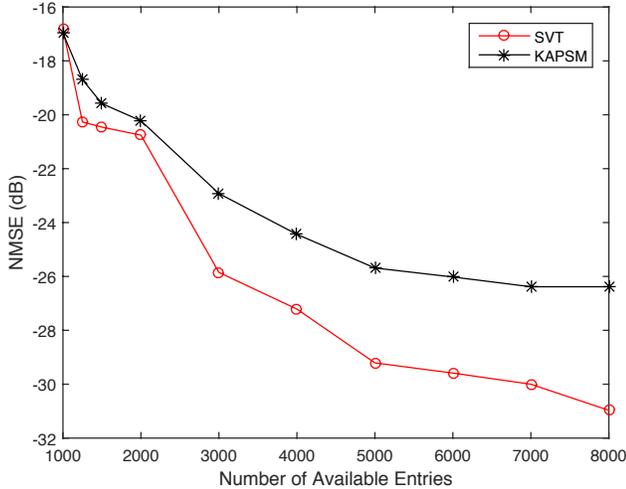


Fig. 3. Comparison of the SVT and the KAPSM for the pathloss reconstruction

ber of available measurements and the vertical axis corresponds to the NMSE (in dB). It can be readily seen that the SVT outperforms significantly the kernel based algorithm. Note that, at 7500 measurements, the accuracy with the MC algorithm is approximately 4.5 dB. On top of that, apart from the NMSE performance plots, the SVT is computationally efficient, as the runtime of our MATLAB implementation is 4 s for 5000 measurements while for the KAPSM the runtime is approximately 30 s. However, it should be pointed out that the KAPSM algorithm is online and, consequently, saves memory by storing fewer measurements.

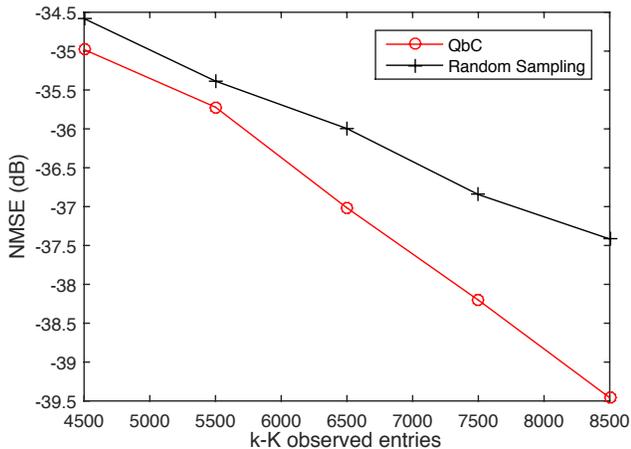


Fig. 4. Evaluation of the QbC method in pathloss reconstruction

In the third experiment, our goal is to validate the benefit of sampling the most informative areas by employing the QbC rationale. The experimental setup can be summarized as follows. We consider that k measurements are available and we compare the reconstruction accuracy, in terms of NMSE, in the following scenarios:

- The pathloss is reconstructed using all the available measurements. This scheme is referred to as Random Sampling.
- The pathloss is reconstructed using $k - K$ measurements. In

the sequel, we apply the QbC method (Table 2) in order to find the K entries with the largest disagreement according to (7) and sample them. Finally, we apply the SVT using the initial $k - K$ entries and the k entries computed previously. This scheme is referred to as QbC.

The parameter K is fixed and equal to $K = 6500$ and we let the parameter k vary. The results are summarized in Fig. 4. It can be clearly seen that the QbC enhances the accuracy of the reconstruction. Obviously, this comes at the expense of an increased computational complexity since the map is reconstructed using more than one algorithm. Nevertheless, there is an important gain, since with the same number of measurements we improve the reconstruction accuracy, which is very important if we take into consideration how costly drive tests are.

5. CONCLUSIONS AND FUTURE WORK

In this paper, an algorithm for pathloss map reconstruction based on matrix completion was proposed. Furthermore, we present a method to identify areas where further measurements maximize the accuracy of the reconstruction. This method is based on the QbC rationale and provides operators with a powerful tool to plan cost-efficient drive tests. The paper also presents a performance evaluation and comparative experiments of our algorithms. Future work focuses on deriving online algorithms for pathloss maps reconstruction and for adaptive identification of the informative areas.

6. REFERENCES

- [1] Wuri Hapsari, Anil Umesh, Mikio Iwamura, Małgorzata Tomala, Bódog Gyula, Benoist Sebire, et al., “Minimization of drive tests solution in 3gpp,” *Communications Magazine, IEEE*, vol. 50, no. 6, pp. 28–36, 2012.
- [2] Chris Phillips, Douglas Sicker, and Dirk Grunwald, “A survey of wireless path loss prediction and coverage mapping methods,” *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 1, pp. 255–270, 2013.
- [3] Caleb Phillips, Scott Raynel, Jamie Curtis, Sam Bartels, Douglas Sicker, Dirk Grunwald, and Tony McGregor, “The efficacy of path loss models for fixed rural wireless links,” in *Passive and Active Measurement*. Springer, 2011, pp. 42–51.
- [4] Mauro Piacentini and Francesco Rinaldi, “Path loss prediction in urban environment using learning machines and dimensionality reduction techniques,” *Computational Management Science*, vol. 8, no. 4, pp. 371–385, 2011.
- [5] David M Gutierrez-Estevéz, Ian F Akyildiz, and Etimad Fadel, “Spatial coverage cross-tier correlation analysis for heterogeneous cellular networks,” *Vehicular Technology, IEEE Transactions on*, vol. 63, no. 8, pp. 3917–3926, 2014.
- [6] Emiliano Dall Anese, Seung-Jun Kim, and Georgios B Giannakis, “Channel gain map tracking via distributed kriging,” *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 3, pp. 1205–1211, 2011.
- [7] Sofia Nikitaki, Grigorios Tsagakatakis, and Panagiotis Tsakalides, “Efficient training for fingerprint based positioning using matrix completion,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 195–199.

- [8] Shobhit Choudhary, Narendra Kumar, Shrikanth Narayanan, and Urbashi Mitra, "Active target detection with mobile agents," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4185–4189.
- [9] Jie Cheng, Hongbo Jiang, Xiaoqiang Ma, Lanchao Liu, Lijun Qian, Chen Tian, and Wenyu Liu, "Efficient data collection with sampling in wsns: making use of matrix completion techniques," in *IEEE Global Telecommunications Conference (GLOBECOM)*,. IEEE, 2010, pp. 1–5.
- [10] Martin Kasparick, Renato L.G. Cavalcante, Stefan Valentin, Slawomir Stanczak, and Masahiro Yukawa, "Kernel-based adaptive online reconstruction of coverage maps with side information," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.
- [11] "Momentum, "MOdels and siMulations for nEtwork plaNning and conTrol of UMts"," 2004, [Online]. Available: <http://momentum.zib.de>.
- [12] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [13] Sergios Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, Academic Press, 2015.
- [14] Emmanuel J Candès and Terence Tao, "The power of convex relaxation: Near-optimal matrix completion," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2053–2080, 2010.
- [15] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.
- [16] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [17] Rong Jin and Luo Si, "A Bayesian approach toward active learning for collaborative filtering," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 278–285.
- [18] Shiladri Chakraborty, Jiayu Zhou, Vineeth Balasubramanian, Sethuraman Panchanathan, Ian Davidson, and Jieping Ye, "Active matrix completion," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 81–90.
- [19] Trevor Hastie, Robert Tibshirani, Gavin Sherlock, Michael Eisen, Patrick Brown, and David Botstein, "Imputing missing data for gene expression arrays," 1999.
- [20] Christopher M Bishop, *Pattern recognition and machine learning*, springer, 2006.