A ROBUST MACHINE LEARNING METHOD FOR CELL-LOAD APPROXIMATION IN WIRELESS NETWORKS

Daniyal Amir Awan[⋆] *Renato L.G. Cavalcante*[⋆†] *Slawomir Stanczak*[⋆†]

[⋆] Technical University Of Berlin, Einsteinufer 27, 10587 Berlin, Germany
[†]Fraunhofer Institute for Telecommunications, HHI, Einsteinufer 37, 10587 Berlin, Germany

ABSTRACT

We propose a learning algorithm for cell-load approximation in wireless networks. The proposed algorithm is robust in the sense that it is designed to cope with the uncertainty arising from a small number of training samples. This scenario is highly relevant in wireless networks where training has to be performed on short time scales because of a fast time-varying communication environment. The first part of this work studies the set of feasible rates and shows that this set is compact. We then prove that the mapping relating a feasible rate vector to the unique fixed point of the non-linear cell-load mapping is monotone and uniformly continuous. Utilizing these properties, we apply an approximation framework that achieves the best worst-case performance. Furthermore, the approximation preserves the monotonicity and continuity properties. Simulations show that the proposed method exhibits better robustness and accuracy for small training sets in comparison with standard approximation techniques for multivariate data.

Index Terms— machine learning, 5G, multivariate scattered data, data interpolation, minimax approximation

1. INTRODUCTION

The load-coupling model [1, 2, 3] is widely used when designing networks according to the long-term evolution (LTE) standard and it has also attracted attention in the context of fifth-generation (5G) networks. More specifically, the load-coupling model has been used in various optimization frameworks dealing with different aspects of network design including, but not limited to, data offloading [4], proportional fairness [5], energy optimization [6, 7], and load balancing [8].

The radio resource management (RRM) in future 5G networks is expected to be similar to the RRM in orthogonal frequency-division multiple access (OFDMA)-based networks such as LTE. Unfortunately many of the RRM problem formulations, such as small-scale optimal assignment of time-frequency resource blocks (RBs) to users, have been shown to be NP-hard [9]. As a result, interference models that are able to capture the long-term behavior of OFDMA-like networks while giving rise to tractable problem formulations have been the focus of recent research. The aforementioned non-linear load-coupling model is such a network-layer model that considers long-term average RB consumption and it has been shown to be sufficiently accurate [2]. In this model, the cell-load at a base station (BS) is the proportion of RBs scheduled to support a particular rate demand. Therefore, given some power budget that can be used for transmission, each BS needs to calculate the cell-load required to serve given rate demands. In [4], the authors present an intuitive result showing that cell-load is monotonic in user rate demand, and the non-linear coupling between cells implies that an increase in the rate demand in the network results in an increase in the cell-load at each BS. Therefore before serving a higher rate demand, it is important for a BS to have a reliable estimation of the impact of this increase to the neighboring BSs in terms of cell-load and interference. This estimation can be used to make RRM and self-organizing-network (SON) algorithms more reliable and efficient. The difficulty in performing these management tasks lies in the need for calculating the expected value of induced cell-load at BSs for given user rates. This is because such a computation typically uses iterative methods requiring a large amount of network information such as pathlosses and user rates, to name a few. To address this challenge, we propose a robust and optimal machine learning technique that allows BSs to approximate cell-load values induced for any given rate demand vector. Moreover, the complexity of the proposed method is low and the algorithm can be implemented in parallel at each BS.

The contributions of this study are as follows. We first study the feasible rate region and properties of the cell-load as a function of rate demand vector. To the best of our knowledge, not much attention has been devoted previously to the structure of the feasible rate region. In particular, we show that the feasible rate region is compact. In addition, we prove that the cell-load is a uniformly continuous mapping over the set of feasible rates.

Based on the above results, as our second contribution, we address the problem of cell-load approximation for a given rate demand vector. Previous studies dealing with this problem (for example, in the context of data offloading [4] and maximizing the scaling-up factor of traffic demand [10]) have used model-based methods that require full information about channel gains, pathlosses etc. In contrast, we approach this problem from a machine learning perspective, in which case no channel information is required. Owing to the dynamicity of dense wireless networks, any machine learning algorithm has to train the network within a relatively small time period. As a consequence, the training sample set is small and the information about the unknown function to be approximated is scarce. In this challenging setting, we propose

This research was supported by Grant STA 864/9-1 from German Research Foundation (DFG).

This work has been performed in the framework of the Horizon 2020 project ONE5G (ICT-760809) receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project.

a learning algorithm that achieves robust approximation in the minimax sense; i.e., the maximum possible error under uncertainty is minimized. Yet another difficulty lies in obtaining an approximation that preserves the properties of the cell-load, in particular, the monotonicity of the cell-load mapping in rate demand. However, incorporating monotonicity in machine learning algorithms for multivariate data with arbitrary dimensions is very challenging, and most of the well-known machine learning algorithms do not preserve the properties of the true function [11]. The work in [12] shows that monotonicity is also hard to incorporate in popular online learning methods. In contrast to these studies, the author in [11] proposed a shape preserving multivariate approximation. We propose a vector-valued version of this method for cell-load approximation at multiple BSs in parallel. Finally, we compare our method with popular multivariate machine learning techniques through simulations and show that our method outperforms them under the aforementioned restriction of a small training set.

2. PRELIMINARIES

Throughout this study, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the set of non-negative and positive reals, respectively. We denote by $C(\mathcal{X})$ the space of vector-valued continuous functions defined on $\mathcal{X} \subset \mathbb{R}^N_{++}$. Likewise, we denote by $\mathbf{g} \in C(\mathcal{X})$ a vector-valued function whose values at a point $\mathbf{x} \in \mathcal{X}$ are given by $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), ..., g_M(\mathbf{x})]^T$, where each $g_i : \mathcal{X} \to [0,1]$, $i = 1, \ldots, M$, is a continuous function. The norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are the standard Euclidean norm and the l_∞ norm, respectively. We denote by $\overline{\mathcal{X}}$ the closure of the set $\mathcal{X}$. For a compact set $\mathcal{X}$ and a vector-valued continuous function $\mathbf{g} \in C(X)$, we define the supremum or uniform norm $\|\cdot\|_{C(\mathcal{X})}$ as

$$\|\mathbf{g}\|_{C(\mathcal{X})} = \sup_{\mathbf{x}\in\mathcal{X}} \max_{1\le i\le M} g_i(\mathbf{x}), \tag{1}$$

where the $\sup$ is attained because the pointwise maximum of finitely many continuous functions is continuous and $\mathcal{X}$ is compact. We denote by $(\mathbf{x})_+$ the operation $\max\{\mathbf{x}, \mathbf{0}\}$ for a vector $\mathbf{x} \in \mathbb{R}^N$, where in contrast to the $\max$ operation in (1), the $\max$ is taken component-wise and $\mathbf{0}$ is the all-zero vector. The distinction between the two usages of the $\max$ operation will be clear from the context in which they are used. Finally, for two vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \le \mathbf{y}$ should be understood component-wise.

Definition 1 (*Lipschitz Monotone Functions*)**.** *Let* $\mathbf{f} : \mathbb{R}^N_{++} \to [0,1]^M$ *be a vector-valued function with the* i*th component* $f_i : \mathbb{R}^N_{++} \to [0,1], i = 1, ..., M$. *We say that* $\mathbf{f}$ *belongs to the class of Lipschitz Monotone Functions (LIMF) if* $\mathbf{f}$ *is monotone on* $\mathcal{X}$ *and each component* f_i *is Lipschitz on* $\mathcal{X} \subset \mathbb{R}^N_{++}$*, i.e.,* $(\forall i \in \{1, 2, \ldots, M\})(\exists L_i \in \mathbb{R}_+)(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathcal{X})\, |f_i(\mathbf{x}) - f_i(\mathbf{y})| \le L_i \|\mathbf{x} - \mathbf{y}\|$.

2.1. Non-linear Load Coupling Model

In this study, we consider a dense urban cellular base station (BS) deployment. The service area is represented by a grid of pixels, each occupying a small area which we refer to as a test point (TP) [1, 10, 7]. We use $r_j > 0$ to denote the aggregated user rate demand within TP j per unit time. It is assumed that if the rate requirement of each TP is met, then the rate requirements of all users in the network are also met. We use $\mathcal{M} = \{1, 2, ..., M\}$ and $\mathcal{N} = \{1, 2, ..., N\}$ to denote the set of BSs and TPs, respectively. We consider a downlink transmission scenario and denote the vector of power levels of all BSs by $\mathbf{p} \in \mathbb{R}^M_{++}$. Throughout this study, the power and user assignment (denoted by $\mathcal{N}(i)$) for each BS i is assumed to be fixed. We denote by $\boldsymbol{\rho} = [\rho_1, \rho_2, ..., \rho_M]^T$ the vector containing the cell-load levels at all BSs. The cell-load ρ_i is given by [1, 10]

$$\rho_i = \frac{1}{RB} \sum_{j\in\mathcal{N}(i)} \frac{r_j}{\log(1 + \gamma_{ij}(\mathbf{p}, \boldsymbol{\rho}))}, \tag{2}$$

where R is the number of time-frequency resource blocks (RBs), B is the bandwidth of each RB, and $\gamma_{ij}(\mathbf{p}, \boldsymbol{\rho}) := p_i G_{i,j} / (\sum_{k\in\mathcal{M}\setminus\{i\}} p_k G_{k,j} \rho_k + \sigma^2)$ (where $G_{i,j} \ge 0$ is the channel gain and σ^2 is the noise power) is the signal-to-noise ratio (SINR) of the link between BS i and TP j. If we collect the rate demand of TPs in a vector $\mathbf{r} = [r_1, r_2, ..., r_N]^T$, then for a fixed rate vector $\mathbf{r} \in \mathbb{R}^N_{++}$, writing (2) for each $i \in \mathcal{M}$ results in a system of non-linear equations,

$$\boldsymbol{\rho} = \mathbf{q}(\boldsymbol{\rho}, \mathbf{r}), \tag{3}$$

where $\mathbf{q} : \mathbb{R}^M_+ \times \mathbb{R}^N_{++} \to \mathbb{R}^M_{++}$ is referred to as the load mapping. Given $\mathbf{r} \in \mathbb{R}^N_{++}$ the mapping $\Gamma_\mathbf{r} : \mathbb{R}^M_+ \to \mathbb{R}^M_{++} : \boldsymbol{\rho} \mapsto \mathbf{q}(\boldsymbol{\rho}, \mathbf{r})$ is a positive concave mapping [13], so it also belongs to the class of standard interference mappings [14]. Therefore, for a given rate demand vector $\mathbf{r} \in \mathbb{R}^N_{++}$, the set $\text{Fix}(\Gamma_\mathbf{r}) := \{\boldsymbol{\rho} \in \mathbb{R}^M_{++} | \Gamma_\mathbf{r}(\boldsymbol{\rho}) = \boldsymbol{\rho}\}$ contains at most one fixed point. If $\text{Fix}(\Gamma_\mathbf{r}) \neq \emptyset$, the unique fixed point is the solution to (3). We define a feasible rate demand as follows:

Definition 2 (*Feasible Rate Demand Vector*)**.** *A rate demand vector* $\mathbf{r} \in \mathbb{R}^N_{++}$ *is feasible for the network if and only if* $\textit{Fix}(\Gamma_\mathbf{r}) \neq \emptyset$ *and* $\mathbf{0} \le \boldsymbol{\rho}^* \le \mathbf{1}$*, where* $\boldsymbol{\rho}^* \in \textit{Fix}(\Gamma_\mathbf{r})$.

Denote by $\mathcal{X}_f$ the set of all feasible rate demand vectors as defined in Definition 2. Denote by $\mathbf{r}_{\min} \in \mathbb{R}^N_{++}$, the minimum rate requirement of users and consider the set $\mathcal{X}_{\min} := \{\mathbf{r} \in \mathbb{R}^N_{++} | \mathbf{r} \ge \mathbf{r}_{\min}\}$. In the remainder, we restrict our attention to the set of feasible rate demand vectors (or rate region) $\mathcal{X} := \mathcal{X}_f \bigcap \mathcal{X}_{\min}$ and the set of fixed points $\mathcal{Y} := \left\{\boldsymbol{\rho} \in [0,1]^M | (\exists \mathbf{r} \in \mathcal{X}) \Gamma_\mathbf{r}(\boldsymbol{\rho}) = \boldsymbol{\rho}\right\}$. Furthermore, we assume that $\mathcal{X}, \mathcal{Y} \neq \emptyset$. In the following section, we proceed to study some properties of $\mathcal{X}$ and the corresponding fixed points.

3. FEASIBLE RATE REGION AND FIXED POINTS

In this section we show that the feasible rate region $\mathcal{X}$ is compact and the fixed points are generated by a uniformly continuous monotonic mapping on this set. The compactness of the domain set and continuity of the function to be approximated are necessary conditions for the proposed learning algorithm in Section 4.

We start this section with a simple result stating that the fixed point of the load mapping in (3) is monotonic in rate demand. Owing to the space restrictions, the proofs for most of the results are provided in the extended version [15].

Lemma 1. *Consider any two rate demand vectors* $\mathbf{r}^k, \mathbf{r}^j \in \mathcal{X}$

and the fixed points $\boldsymbol{\rho}^j \in Fix(\Gamma_{\mathbf{r}^j})$ and $\boldsymbol{\rho}^k \in Fix(\Gamma_{\mathbf{r}^k})$. Then $\mathbf{r}^j \geq \mathbf{r}^k \Rightarrow \boldsymbol{\rho}^j \geq \boldsymbol{\rho}^k$.

Lemma 2. *The feasible rate region $\mathcal{X}$ is bounded.*

We are now in a position to present the main results of this section. These results enable us to apply robust and optimal approximation methods in Section 4.

Proposition 1. *The feasible rate region $\mathcal{X} \subset \mathbb{R}^N_{++}$ is compact.*

In what follows, we denote by $\mathbf{f} : \mathcal{X} \to \overline{\mathcal{Y}} : \mathbf{r} \mapsto \boldsymbol{\rho} \in \text{Fix}(\Gamma_{\mathbf{r}})$ the function that maps each $\mathbf{r}$ to the unique fixed point of the mapping $\Gamma_{\mathbf{r}}$ in (3). The following theorem shows that $\mathbf{f}$ is uniformly continuous over the compact set $\mathcal{X}$.

Theorem 1. *The function $\mathbf{f} : \mathcal{X} \to \overline{\mathcal{Y}}$ is uniformly continuous over $\mathcal{X}$.*

Proof. Since $\mathcal{X}$ is compact, every infinite sequence in $\mathcal{X}$ has a convergent subsequence whose limit is in $\mathcal{X}$. Let $(\mathbf{r}_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ be an arbitrary convergent sequence, and let the point $\mathbf{r}^* \in \mathcal{X}$ be its limit. To prove that $\mathbf{f}$ is continuous, we need to show that $\lim_{n\to\infty} \mathbf{f}(\mathbf{r}_n) = \mathbf{f}(\mathbf{r}^*)$. To this end, let $(\boldsymbol{\rho}_n)_{n\in\mathbb{N}} \subset \overline{\mathcal{Y}}$ be the corresponding sequence of $(\mathbf{f}(\mathbf{r}_n))_{n\in\mathbb{N}}$. Since $\overline{\mathcal{Y}}$ is compact, such a sequence has a convergent subsequence $(\boldsymbol{\rho}_n)_{n\in K_1\subset\mathbb{N}}$ whose limit $\boldsymbol{\rho}^*$ exists and belongs to $\overline{\mathcal{Y}}$. The corresponding sequence of rate vectors $(\mathbf{r}_n)_{n\in K_1\subset\mathbb{N}}$ is a subsequence of the convergent sequence $(\mathbf{r}_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ and therefore also convergent. Now consider the function $\mathbf{g}(\boldsymbol{\rho}, \mathbf{r}) = \boldsymbol{\rho} - \mathbf{q}(\boldsymbol{\rho}, \mathbf{r})$, where $\mathbf{q}$ is the load mapping, and note that this function is continuous and $(\forall n \in K_1 \subset \mathbb{N})\ \mathbf{g}(\boldsymbol{\rho}_n, \mathbf{r}_n) = \mathbf{0}$. It follows from the definition of a continuous function that $\lim_{n\in K_1} \mathbf{g}(\boldsymbol{\rho}_n, \mathbf{r}_n) = \mathbf{g}(\boldsymbol{\rho}^*, \mathbf{r}^*) = \mathbf{0}$. Therefore, the limit of the subsequence $(\boldsymbol{\rho}_n)_{n\in K_1\subset\mathbb{N}}$ is the unique fixed point $\mathbf{f}(\mathbf{r}^*) = \boldsymbol{\rho}^*$ and $\lim_{n\in K_1} \mathbf{f}(\mathbf{r}_n) = \mathbf{f}(\mathbf{r}^*)$. Since $\mathcal{X}$ is compact, $\mathbf{f}$ is uniformly continuous on $\mathcal{X}$. □

4. THE LEARNING PROBLEM

In this section we present a learning algorithm which is not only robust and optimal in a challenging machine learning scenario, but also preserves the monotonicity and continuity of the function to be approximated.

4.1. Minimax Optimal Aprroximation

Let the training data set be denoted by $\mathcal{D} = \{(\mathbf{r}^k, \boldsymbol{\rho}^k)\}_{k=1}^K$, $(\mathbf{r}^k, \boldsymbol{\rho}^k) \in (\mathcal{X} \times \overline{\mathcal{Y}})\}$, where $\boldsymbol{\rho}^k := \mathbf{f}(\mathbf{r}^k)$ are the measured cell-load values generated by the underlying function $\mathbf{f} : \mathcal{X} \to \overline{\mathcal{Y}}$. Our objective is to approximate the value $\mathbf{f}(\mathbf{r})$ for any $\mathbf{r} \in \mathcal{X}$, i.e., our objective is to solve an interpolation problem given $\mathcal{D}$. In the classical approximation theory [16, 17, 18], the data interpolation problem entails computing an approximation $\mathbf{g}$ of the function $\mathbf{f}$ by observing the values in the set $\mathcal{D}$, and then replacing future evaluations of $\mathbf{f}(\mathbf{r})$ with $\mathbf{g}(\mathbf{r})$ for $\mathbf{r} \in \mathcal{X}$. We have shown by Theorem 1 that $\mathbf{f}$ is a uniformly continuous function on the compact set $\mathcal{X}$. Clearly there are infinitely many functions in the space $C(\mathcal{X})$ that interpolate $\mathcal{D}$. Since we are interested in a robust approximation of the unknown $\mathbf{f}^* \in C(\mathcal{X})$, we aim at minimizing the worst-case error [17, 19],

$$\mathrm{E}_w(\mathbf{g}) = \sup_{\mathbf{f}\in C(\mathcal{X})} \|\mathbf{f} - \mathbf{g}\|_{C(\mathcal{X})}, \tag{4}$$

where $\mathbf{g} \in C(\mathcal{X})$ is confined to a class of functions such that $\mathbf{f}^*(\mathbf{r}^k) = \mathbf{g}(\mathbf{r}^k)$, $k = 1, \ldots, K$.

Unfortunately, if the only information about $\mathbf{f}^*$ are the observations in $\mathcal{D}$ and the fact that $\mathbf{f}^* \in C(\mathcal{X})$, the worst-case error can be arbitrarily large for some appropriately chosen $\mathbf{g} \in C(\mathcal{X})$. However if $\mathbf{f}^*$ belongs to a compact subset of $C(\mathcal{X})$, the sup in (4) is attained and we can guarantee a finite worst-case error. A sufficient condition for compactness of a subset in the space $\mathcal{C}(\mathcal{X})$ is that all functions in the subset are Lipschitz continuous with the same Lipschitz constant [11]. Moreover, Lipschitz continuity imposes a nonlinear restriction on the function class. In this case it has been shown in [16], that for any given $\mathbf{r} \in \mathcal{X}$, the values $\mathbf{f}^*(\mathbf{r})$ belong to a closed interval, and the optimal approximation to $\mathbf{f}^*(\mathbf{r})$ is the midpoint of this interval. This means that no matter how inconvenient the machine learning scenario is (for example, a small sample set and fast changing statistics), we are guaranteed a certain finite worst-case error. Therefore, in addition to the monotonicity and uniform continuity properties of $\mathbf{f}$ we presented in the previous section, we make the following assumption:

Assumption 1. The function $\mathbf{f} : \mathcal{X} \to \overline{\mathcal{Y}} : \mathbf{r} \mapsto \boldsymbol{\rho} \in \text{Fix}(\Gamma_{\mathbf{r}})$ is a (component-wise) Lipschitz monotone function (LIMF) on the set $\mathcal{X}$ (see Definition 1).

We can now state the optimal approximation as an optimization problem.

Definition 3 (*Optimal Approximation*)**.** *Let $\mathcal{D} = \{(\mathbf{x}^k, \boldsymbol{\rho}^k) \in (\mathcal{X} \times \mathcal{Y}\}_{k=1}^K$, with $\mathcal{X}$ compact, be a data set and assume that $\boldsymbol{\rho}^k := \mathbf{f}(\mathbf{x}^k), k = 1, \ldots, K$, are values generated by an unknown function $(\mathcal{F} \ni)\mathbf{f} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{F} \subset C(\mathcal{X})$ is a set of LIMF functions. The minimax optimal approximation problem is stated as follows:*

Problem 1. [17, 11, 19] Find $\mathbf{g} : \mathcal{X} \to \mathcal{Y}$, such that

$$\mathbf{g} \in \arg\min_{\mathbf{g}\in S} \mathrm{E}_{\max}(\mathbf{g}) \tag{5}$$

where $S := \{\mathbf{g} \in C(\mathcal{X}) | \mathbf{g}(\mathbf{x}^k) = \mathbf{f}(\mathbf{x}^k), \forall k \in \{1, \ldots, K\}\}$, and $\mathrm{E}_{\max}(\mathbf{g}) := \max_{\mathbf{f}\in\mathcal{F}} \|\mathbf{f} - \mathbf{g}\|_{C(\mathcal{X})}$ is the worst-case error from (4) computed over the set $\mathcal{F}$.

In [11], the author provides a framework for monotone interpolation of Lipschitz functions defined over a compact set by using a *central scheme* [17, 18], that can be used to construct an optimal solution for Problem 1. The following Fact summarizes the important properties of an optimal solution constructed using this framework.

Fact 1. Let $\mathcal{D} = \{(\mathbf{x}^k, \boldsymbol{\rho}^k) \in (\mathcal{X} \times \mathcal{Y})\}_{k=1}^K$, be a dataset generated by an unknown function $\mathbf{f} \in \mathcal{F}$, where $\mathcal{F}$ is a set of LIMF functions (see Definition 1). Then, we have the following:

1. An optimal minimax approximation $\mathbf{g} \in C(\mathcal{X})$ of $\mathbf{f} \in \mathcal{F}$ is given by

$$(\forall i \in \mathcal{M})(\forall \mathbf{x} \in \mathcal{X})\ g_i(\mathbf{x}) = \frac{\sigma_l^i(\mathbf{x}) + \sigma_u^i(\mathbf{x})}{2}, \tag{6}$$

where $\sigma_l^i(\mathbf{x}) = \max_k\{f_i(\mathbf{x}^k) - L_i\|(\mathbf{x} - \mathbf{x}^k)_+\|\}$, $\sigma_u^i(\mathbf{x}) = \min_k\{f_i(\mathbf{x}^k) + L_i\|(\mathbf{x} - \mathbf{x}^k)_+\|\}$, $f_i(\mathbf{x}^k) = \rho_i^k$, and L_i is the Lipschitz constant of the ith component f_i.

2. $\mathbf{g} \in \mathcal{F} \subset C(\mathcal{X})$.

Algorithm 1 Cell-Load Estimation At Each BS

1: **Training:** (a) Collect the training data set $\mathcal{D}^{\text{noise}} = \{(\mathbf{r}^k, y^k = f(\mathbf{r}^k) + \epsilon(\mathbf{r}^k))\}_{k=1}^K$. (b) Perform the estimation of L and data smoothing to construct the compatible $\mathcal{D} = \{(\mathbf{r}^k, \rho^k)\}_{k=1}^K$ as described in Section 5.1.

2: **Online Prediction:** Given a new rate demand vector $\mathbf{r} \in \mathcal{X}$, perform the direct computation (6) in Fact 1

$$g(\mathbf{r}) = \frac{1}{2}(\max_k\{\rho^k - L\|(\mathbf{r} - \mathbf{r}^k)_+\|\}) + \frac{1}{2}(\min_k\{\rho^k + L\|(\mathbf{r} - \mathbf{r}^k)_+\|\}).$$

5. ALGORITHM AND SIMULATION

We consider a network with $M = 10$ BS sites and $N = 50$ TPs placed randomly. Each TP is connected to exactly one BS with the best received SNR. The pathloss for links between BSs and TPs follows the 3GPP ITU propagation model for urban macro cell environments. In the following, we restrict our attention to a single BS and omit the index i because the cell-load approximation (g_i in (6)) is computed independently at each BS.

5.1. Noisy Training Data

Practical systems are subject to noise during measurement, so that instead of a data set $\mathcal{D} = \{(\mathbf{r}^k, f(\mathbf{r}^k))\}_{k=1}^K$, a noisy training data set $\mathcal{D}^{\text{noise}} = \{(\mathbf{r}^k, y^k = f(\mathbf{r}^k)) + \epsilon(\mathbf{r}^k))\}$ is available, where $\epsilon(\mathbf{r}^k)$ is the measurement noise assumed to be bounded. As a consequence, y^k might not be compatible with the monotonicity property of f and must be smoothed to obtain a compatible set. In more detail, we first estimate the Lipschitz constant L by $L := \max_{k \neq j} \frac{|y^k - y^j| - 2\epsilon}{\|\mathbf{r}^k - \mathbf{r}^j\|}$, where $\epsilon := \sup_k |\epsilon(\mathbf{r}^k)|$ [20]. The monotone-smoothing problem is given by a linear program (LP) [11]

$$\min_{q_+^k, q_-^k \geq 0} \sum_{k=1}^K |q^k|$$
$$\text{s.t. } q^k - q^j \leq y^j - y^k + L\|(\mathbf{r}^k - \mathbf{r}^j)_+\|,$$

where $k, j \in \{1, 2, \ldots, K\}$, $q^k = q_+^k - q_-^k$, $|q^k| = q_+^k + q_-^k$; and, $q_+^k, q_-^k \geq 0$ are the optimization variables. The smoothed compatible values can be calculated as $\rho^k := y^k + q^k$. An LP is a convex optimization problem and can be solved by a standard convex solver.

5.2. Implementation and Complexity

The cell-load estimation algorithm is shown in Algorithm 1. Note that, for a given $\mathbf{r} \in \mathcal{X}$, each BS $i \in \mathcal{M}$ can calculate the component $g_i(\mathbf{r})$ independently of other BSs using (6) . Therefore, Algorithm 1 is scalable to a larger dense network and is amenable to distributed implementation. The *training* step can be performed by standard convex solvers whereas the complexity of the *online prediction* step is linear in sample size K, i.e, $\mathcal{O}(K)$. Therefore for small sample sizes considered in this study, Algorithm 1 exhibits a fast computational speed.

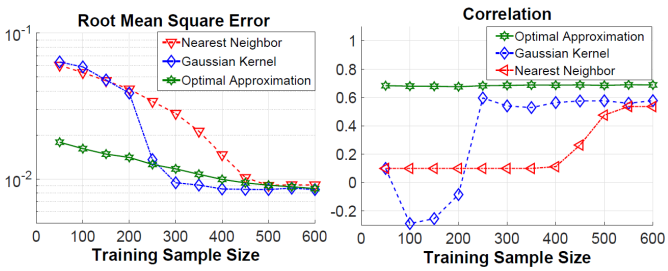

Fig. 1. RMSE and (Pearson's) Correlation coefficient

5.3. Results

We train the network over the set $\mathcal{X} = \{\mathbf{r} \in \mathbb{R}_{++}^{50} | \mathbf{r}_{\min} \leq \mathbf{r} \leq \mathbf{r}_{\max}\}$, where $\mathbf{r}_{\min} = (10^6)\,\mathbf{1}$ and $\mathbf{r}_{\max} = (10^7)\,\mathbf{1}$ is the pre-configured range (in bits/s) of rate vectors and $\mathbf{1} \in \mathbb{R}_{++}^{50}$ is the ones vector. We calculate the cell-load values using the fixed point iterative method [14] with the cell-load mapping (3). Other important parameters are: $RB = 20\,\text{MHz}$, $(\forall i \in \mathcal{M}) p_i = 1\,\text{W}$, $\sigma^2 = 1.38 \times 10^-23 \times 300/20 \times 10^5$. Normally distributed noise with $\epsilon = 0.05$ is added to the data. The *training step* is performed by a standard convex solver.

We compare the performance of Algorithm 1 and two other standard machine learning techniques, namely the standard Gaussian kernel regression and the 2-nearest neighbor interpolation. Note that neither of these two techniques are in general shape preserving. We use these two techniques because they are able to handle problems involving high-dimensional multivariate scattered data such as the case in this study [11]. For brevity we compare the quality (in terms of *Pearson's correlation coefficient*) and accuracy (in terms of *root mean square error* (RMSE)) for cell-load predictions at a single BS. Similar results were obtained for each BS. We simulate increasing sample size K and make 100 000 test predictions at random values of rate demand vectors in $\mathcal{X}$ for each value of K to gather reliable statistics.

It can be observed in Figure 1 that our proposed framework shows a more robust and consistent performance both in terms of quality of prediction and accuracy over the range of sample sizes as compared to the other two techniques, especially for small sample sizes, i.e, it is robust under uncertainty. The improvement in RMSE with increasing sample sizes is due to the decrease in uncertainty about the true function $\mathbf{f}$. Note that even though the cost function (4) which Algorithm 1 optimizes is not the same as the RMSE, we can still represent its performance using a standard error measure like RMSE. At values near $K = 600$ the three techniques show comparable performance in terms of RMSE, but in contrast to our framework, the other techniques are not guaranteed to be shape preserving and the predictions might not be compatible with the monotonicity property of the function.

6. REFERENCES

[1] I. Siomina, "Analysis of Cell Load Coupling for LTE Network Planning and Optimization," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2287–2297, June 2012.

[2] A. J. Fehske and G. P. Fettweis, "Aggregation of variables in load models for interference-coupled cellular data networks," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 5102–5107.

[3] K. Majewski and M. Koonert, "Conservative cell load approximation for radio networks with shannon channels and its application to lte network planning," in *2010 Sixth Advanced International Conference on Telecommunications*, May 2010, pp. 219–225.

[4] Chin Keong Ho, Di Yuan, and Sumei Sun, "Data offloading in load coupled networks: A utility maximization framework," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1921–1931, 2014.

[5] M. A. Gutierrez-Estevez, R. L. G. Cavalcante, S. Stanczak, J. Zhang, and H. Zhuang, "A distributed solution for proportional fairness optimization in load coupled ofdma networks," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP' 16)*, 2016.

[6] D.A. Awan, R. L. G. Cavalcante, and Slawomir Stanczak, "Distributed ran and backhaul optimization for energy efficient wireless networks," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP' 16)*.

[7] E. Pollakis, Renato L G Cavalcante, and S. Stanczak, "Traffic demand-aware topology control for enhanced energy-efficiency of cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–17, 2016.

[8] I. Siomina and D. Yuan, "Load balancing in heterogeneous lte: Range optimization via cell offset and load-coupling characterization," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 1357–1361.

[9] I. C. Wong, Zukang Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in ofdma systems," in *IEEE Workshop onSignal Processing Systems, 2004. SIPS 2004.*, Oct 2004, pp. 1–6.

[10] I. Siomina and D. Yuan, "Optimizing small-cell range in heterogeneous and load-coupled lte networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 2169–2174, May 2015.

[11] G. Beliakov, "Monotonicity preserving approximation of multivariate scattered data," *BIT Numerical Mathematics*, vol. 45, no. 4, pp. 653–677, 2005.

[12] Wojciech Kotlowski, "Online isotonic regression," in *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, 2016, pp. 1165–1189.

[13] R. L. G. Cavalcante, Y. Shen, and S. Stanczak, "Elementary properties of positive concave mappings with applications to network planning and optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1774–1783, April 2016.

[14] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, Sep 1995.

[15] D. A. Awan, R. L. G. Cavalcante, and S. Stanczak, "A robust machine learning method for cell-load approximation in wireless networks," *ArXiv e-prints*, Oct. 2017.

[16] M Golomb and H.F. Weinberger, *On Numerical Approximation*, The University of Wisconsin Press, Madison, r.e. langer (ed.) edition, 1959.

[17] Aleksei G. Sukharev, *Minimax Models in the Theory of Numerical Methods*, Kluwer Academic Publishers, Norwell, MA, USA, 1992.

[18] J.F. Traub and H. Woźniakowski, *A general theory of optimal algorithms*, ACM monograph series. Academic Press, 1980.

[19] Geneva G. Belford, "Uniform approximation of vector-valued functions with a constraint," *Mathematics of Computation*, vol. 26, no. 118, pp. 487–492, 1972.

[20] Jan-Peter Calliess, *Conservative decision-making and inference in uncertain dynamical systems*, Ph.D. thesis, Department of Engineering Science, University of Oxford, 2014.