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

Index Terms— machine learning, 5G, ultra-dense networks (UDNs), multivariate scattered data, optimal approximation

1. INTRODUCTION

The load-coupling model (see, e.g., [1, 2] and [3]) is widely used when designing networks according to the long-term evolution (LTE) standard and 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. While the RRM problems in OFDMA-based networks deal with small-scale optimal assignment of time-frequency resource blocks (RBs) to users1, the non-linear load-coupling model is a network-layer model that considers long-term average RB consumption. Moreover, in [2], the load-coupling model has been verified to be sufficiently accurate. 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 budgets that can be used for transmissions, each BS needs to calculate the cell-load which is necessary to serve given rate demands. In [4], the authors presented an intuitive result showing that cell-load is monotonic in user rate demand and the non-linear coupling between cells implies that increasing an arbitrary rate demand in the network increases 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 such 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, user rates etc. We provide a robust and optimal machine learning technique which 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 obtain some properties of the cell-load as a function of rate demand vector. The feasible rate region is defined as the set of rate demand vectors for which the cell-load at each BS is less than or equal to 1 [4]. To the best of our knowledge, not much attention has been paid so far to the structure of the feasible rate region and this paper provides some initial results. In particular, we obtain the result that the feasible rate region is compact. Moreover, we prove that the cell-load is a uniformly continuous mapping over the set of feasible rate region.

Based on these 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 highly relevant setting, we propose a learning algorithm which achieves best approximation in the minimax sense, i.e., where the worst possible error under uncertainty is minimized. Yet another difficulty lies in obtaining a shape preserving approximation. It is known that the cell-load based on the underlying load-coupling model is monotonic in rate demand [4]. Therefore, the approximation should preserve this monotonicity. However, incorporating monotonicity in machine learning algorithms for multivariate data with arbitrary dimensions is difficult, and most of the well-known machine learning algorithms do not preserve the shape 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 algorithms 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 $\mathbb{R}^{N+}$ denote the set of non-negative and positive reals, respectively. We denote by $C(X)$ the space of vector-valued continuous functions defined on $X \subset \mathbb{R}^{N+}$. Likewise, we denote by $g \in C(X)$ a vector-valued function whose values at a point $x \in X$ are given by $g(x) = [g_1(x), g_2(x), \ldots, g_M(x)]^T$, where each $g_i : X \rightarrow [0, 1], i = 1, \ldots, M$, is a continuous function. The norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are the standard Euclidean norm and the $l_\infty$ norm, respectively. For a compact set $X$ and a vector-valued continuous function $g \in C(X)$, we define the supremum or uniform norm $\|g\|_{C(X)}$ as

$$\|g\|_{C(X)} = \sup_{x \in X} \max_{1 \leq i \leq M} g_i(x),$$ \hspace{1cm} (1)

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

**Definition 1 (Monotone Function).** Let $f : \mathbb{R}^N_+ \rightarrow [0, 1]^M$ be a vector-valued function. Then $f$ is called monotone on $X \subset \mathbb{R}^N_+$ if $\forall x \in X (\forall y \in X) x \leq y \Rightarrow f(x) \leq f(y)$.

**Definition 2 (Lipschitz Monotone Functions).** Let $f : \mathbb{R}^N_+ \rightarrow [0, 1]^M$ be a vector-valued function with the $i$th component $f_i : \mathbb{R}^N_+ \rightarrow [0, 1], i = 1, \ldots, M$. We say that $f$ belongs to the class of Lipschitz Monotone Functions (LMF) if $f$ is monotone on $X$ and each component $f_i$ is Lipschitz on $X \subset \mathbb{R}^N_+$.

3. SYSTEM MODEL

3.1. 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) (see [1, 10], and [7]). The concept of test points provides a network-layer view of quality-of-service (QoS) and large-scale channel conditions in a network. Within each TP, users are assumed to experience uniform signal propagation conditions. 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 QoS requirements of all users in the network are also met. Under this assumption, small-scale fluctuations in individual user rates are averaged out and are compensated by the lower layers of the protocol stack [7]. 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 $p \in \mathbb{R}_+^M$. Throughout this study, the power and user assignment to BSs is assumed to be fixed. We collect the rate demand of TPs in a vector $r = [r_1, r_2, \ldots, r_N]^T$.

Assume a case where BS $i \in \mathcal{M}$ is serving a user in TP $j \in \mathcal{N}(i)$, where $\mathcal{N}(i)$ is the set of TPs connected to BS $i$. The bandwidth resources available at each BS are divided into resource blocks (RBs). The load-based SIRN model (see [1], [4]) incorporates the inter-cell interference from a particular BS $k \in \mathcal{M}$ as the product $p_k G_{k,j} \rho_k \geq 0$, where $G_{k,j} \geq 0$ is the channel gain between BS $k$ and TP $j$ and $0 \leq \rho_k \leq 1$ is referred to as the activity level or cell-load at BS $k$ [2]. With this model in hand, for given power levels $p$, cell-load $\rho$ and user assignment, the SINR $\gamma_{ij}$ of the wireless link between BS $i \in \mathcal{M}$ and TP $j \in \mathcal{N}$ is expressed as [2, 1] and [4]

$$\gamma_{ij}(p, \rho) = \frac{p_k G_{k,j} \rho_k + \sigma^2}{\sum_{k \in \mathcal{M} \setminus \{i\}} p_k G_{k,j} \rho_k + \sigma^2},$$ \hspace{1cm} (2)

where $\rho = [\rho_1, \rho_2, ..., \rho_M]^T$ is the vector containing the cell-load levels at all BSs in the network and $\sigma^2$ is the noise power.
Given the value of SINR $\gamma_{ij}(p, \rho)$, a BS can transmit at rate $r^*_{ij} = B \log(1 + \gamma_{ij}(p, \rho))$ bits/s/ RB reliably to a user in TP $j$, where $B$ is the bandwidth of the RB. Therefore, BS $i$ allocates $\rho_{ij} = r_j / r^*_{ij}$ RBs to meet the demand $r_j$. Summing over all TPs connected to BS $i$ we obtain

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

where $R$ is the number of RBs available at each BS. For a fixed rate vector $\mathbf{r} \in \mathbb{R}^N_+$, writing the above equation for each $i \in M$ results in a system of non-linear equations,

$$\mathbf{r} = \mathbf{q}(\mathbf{r}, \mathbf{q}),$$

where $\mathbf{q} : \mathbb{R}^M_+ \times \mathbb{R}^N_+ \rightarrow \mathbb{R}^M_+$ is referred to as load mapping. Given $\mathbf{r} \in \mathbb{R}^N_+$ the mapping $\Gamma_r : \mathbb{R}^M_+ \rightarrow \mathbb{R}^M_+ : \rho \mapsto \mathbf{q}(\rho, \mathbf{r})$ 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 $\text{Fix}(\Gamma_r) := \{\rho \in \mathbb{R}^M_+ | \Gamma_r(\rho) = \rho\}$ contains at most one fixed point. If $\text{Fix}(\Gamma_r) \neq \emptyset$, the unique fixed point is the solution to (4). We define a feasible rate demand as follows:

**Definition 3 (Feasible Rate Demand Vector).** A rate demand vector $\mathbf{r} \in \mathbb{R}^N_+$ is feasible for the network if and only if $\text{Fix}(\Gamma_r) \neq \emptyset$ and $\mathbf{0} \leq \mathbf{r}^* \leq 1$, where $\mathbf{r}^* \in \text{Fix}(\Gamma_r)$.

In the remainder, we denote the set of feasible rate demand vectors (i.e., the feasible rate region) by $\mathcal{X} \subset \mathbb{R}^N_+$ and the set of corresponding fixed points by $\mathcal{Y} := \{\rho \in [0, 1]^M | (\exists \mathbf{r} \in \mathcal{X}) \Gamma_r(\rho) = \rho\}$. We assume that $\mathcal{X} \neq \emptyset$.

In the following section, we proceed to study some properties of this set and the corresponding fixed points. For ease of reference, we first present some important properties of positive concave mappings.

**Fact 1.** [14] Let $\Gamma : \mathbb{R}^M_+ \rightarrow \mathbb{R}^M_+$ be a positive concave mapping. Then each of the following holds:

1. $\text{Fix}(\Gamma) \neq \emptyset$ if and only if there exists $\rho \in \mathbb{R}^M_+$ such that $\Gamma(\rho) \leq \rho$.
2. If it exists, the fixed point can be found by the fixed point iteration $\rho_{n+1} = \Gamma(\rho_n)$, $n \in \mathbb{N}$, with $\rho_1 \in \mathbb{R}^M_+$ chosen arbitrarily.
3. The fixed point iteration is increasing (resp. decreasing) if $\rho_1 \leq \Gamma(\rho_1)$ (resp.$\rho_1 \geq \Gamma(\rho_1)$).

**4. FEASIBLE RATE REGION AND FIXED POINTS**

In this section, we show that the feasible rate region 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 two important conditions in theory of minimax optimal recovery/approximation of functions ([15]), which we use in the proposed learning algorithm in Section 5.

We start this section by a simple result that the fixed point of the load mapping in (4) is monotonic in rate demand. See [4, Theorem 2] for an alternative proof. We provide a simpler proof that suffices for our purposes.

**Lemma 1.** Consider any two rate demand vectors $\mathbf{r}^k, \mathbf{r}^l \in \mathcal{X}$ and the fixed points $\rho^k \in \text{Fix}(\Gamma_{r^k})$ and $\rho^l \in \text{Fix}(\Gamma_{r^l})$. Then $\mathbf{r}^k \geq \mathbf{r}^l \Rightarrow \rho^k \geq \rho^l$.

**Proof.** Assume that $\rho^k = \mathbf{q}(\rho^k, \mathbf{r}^k) = \Gamma_{r^k}(\rho^k)$ and consider a demand vector $\mathbf{r}^l \geq \mathbf{r}^k$. Then we have from (4) that $\rho^k = \mathbf{q}(\rho^k, \mathbf{r}^k) \leq \mathbf{q}(\rho^k, \mathbf{r}^l)$. Starting with the vector $\rho^k$, we employ the standard fixed point iteration (see Fact 1) and obtain a monotonically increasing sequence $(\rho_{n+1} = \mathbf{q}(\rho_n, \mathbf{r}^l))_{n \in \mathbb{N}}$, with $\rho_1 = \rho^k$, which converges to the uniquely existing fixed point $\rho^l = \mathbf{q}(\rho^l, \mathbf{r}^l) = \Gamma_{r^l}(\rho^l) \geq \rho^k$.

The next lemma shows that the feasible rate region is bounded.

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

**Proof.** The set $\mathcal{X}$ is clearly bounded from below. Suppose the set $\mathcal{X}$ is unbounded from above, then there exists at least one unbounded sequence $(\mathbf{r}_n)_{n \in \mathbb{N}} \subset \mathcal{X}$. This implies that at least one component of the vector $\mathbf{r}_n$ grows unboundedly. Let us denote the sequence of this component by $(\mathbf{r}_{l,n})_{n \in \mathbb{N}}$ and let the corresponding TP be connected to BS $i \in M$. Every unbounded sequence has an increasing subsequence that diverges to $+\infty$. Let us extract such a subsequence and denote it by $(\mathbf{r}_{l,k})_{k \in \mathbb{N}}$. Likewise, denote by $(\rho_{l,k})_{k \in \mathbb{N}}$ the $k$th component of the subsequence $\rho_{l,k} \in \text{Fix}(\Gamma_{r_{l,k}})$. It can be easily verified that, for fixed $p$ and bandwidth resources (RB) in (3), we have that $\rho_{l,k} = \frac{1}{RB} \sum_{j \in N(i)} \frac{r_{j,k}}{\log(1 + \gamma_{ij}(p, \rho_{l,k}))} \geq \frac{1}{RB} \sum_{j \in N(i)} \frac{r_{j,k}}{\log(1 + \gamma_{ij}(p, \rho_{l,0}))} > 0$, where $\rho_{l,k}$ and $r_{j,k}$ are the $k$th and $j$th vector of components $\rho_{l,k}$ and $\mathbf{r}_{l,k}$, respectively. Now, note that the lower bound $\frac{1}{RB} \log(1 + \gamma_{ij}(p, \rho_{l,0}))$ grows unboundedly as $\rho_{l,k} \rightarrow +\infty$, which in particular implies that $\lim_{k \rightarrow \infty} \rho_{l,k} = +\infty$. However, this contradicts the fact that, by our definition of feasibility, $(\forall k \in \mathbb{R}) 0 \leq \rho_{l,k} \leq 1$. Therefore, we conclude that the feasible set $\mathcal{X}$ does not contain any unbounded sequence, so $\mathcal{X}$ is bounded.

Now we proceed to characterize the fixed point solution of (4) by exploiting its connection with the conditional eigenvalue problem based on concave Perron-Frobenius theory [16]. First, we prove a general result which is valid for positive concave mappings arising in various contexts beyond the scope of this paper.

\[\text{The same result also holds for a more general class of standard interference mappings but we do not consider these in this study [17].}\]
**Proposition 1.** Let $\Gamma : \mathbb{R}^M_+ \to \mathbb{R}^M_+$ be an arbitrary positive concave eigenvalue mapping and consider the following conditional eigenvalue problem which always has a unique solution $(\rho^*, \lambda^*) \in \mathbb{R}^M_+ \times \mathbb{R}_+$ [16]:

**Problem 1.** Find $(\rho^*, \lambda^*) \in \mathbb{R}^M_+ \times \mathbb{R}_+$ such that $\Gamma(\rho^*) = \lambda^* \rho^*$ and $\|\rho^*\|_\infty = 1$.

Then $\text{Fix}(\Gamma) \neq \emptyset$ and $1 \geq \rho^* \in \text{Fix}(\Gamma)$ if and only if $\lambda^* \leq 1$.

**Proof.** If $\lambda^* \leq 1$, then $\Gamma(\rho^*) \leq \rho^*$ and, by Fact 1.1, it follows that $\text{Fix}(\Gamma) \neq \emptyset$ and the fixed point $\rho^* \in \text{Fix}(\Gamma)$ exists and satisfies $0 \leq \rho^* \leq 1$. This shows one direction of the claim.

Conversely, assume $\rho^* \in \text{Fix}(\Gamma)$ and $0 \leq \rho^* \leq 1$. Let $\|\rho^*\|_\infty = c$, where $0 < c \leq 1$, and $\lambda' = 1$. Then $(\rho^*, \lambda')$ solves the following conditional eigenvalue problem: First note that by definition $\Gamma_r(\rho^*) = \lambda' \rho^* \Leftrightarrow q(\rho^*; \frac{1}{\lambda'} r) = \rho^*$. Suppose $\lambda' \geq 1$, then we have that $\frac{1}{\lambda'} r \geq r$. Let $\Gamma = \frac{1}{\lambda'} r$, then it can be seen that $\rho^* = q(\rho^*; \Gamma)$ and $\rho^*$ is the fixed point of the mapping $\Gamma_r$ with $\|\rho^*\|_\infty = 1$. Observing that $\Gamma_r(\rho^*) \leq \Gamma_r(\rho^*) = \rho^*$, it follows from Fact 1 that $0 \leq \text{Fix}(\Gamma_r) \leq \rho^* \leq 1$. Conversely, denote the fixed point of $\Gamma_r$ by $\rho^*$ and suppose that $0 \leq \rho^* \leq 1$. Then $(\rho^*, 1)$ solves the following eigenvalue problem:

**Problem 2.** Find $(\rho, \lambda) \in \mathbb{R}^M_+ \times \mathbb{R}_+$ such that $\Gamma(\rho) = \lambda \rho$ and $\|\rho\|_\infty = c$.

Now consider Problem 1 with this mapping $\Gamma$, and denote its solution by $(\rho^*, \lambda^*) \in \mathbb{R}^M_+ \times \mathbb{R}_+$. To obtain a contradiction suppose that $\lambda^* > 1 = \lambda'$. From [18, Lemma 3.1.3], and the fact that positive concave mappings are a special case of standard interference mappings [13], it follows that $\lambda^* > 1 \Rightarrow \rho^* > \rho^*$. From the monotonicity of the $l_\infty$ norm, we obtain the contradiction that $c = \|\rho^*\|_\infty > \|\rho^*\|_\infty = 1$. This concludes the second direction of the proof.

The immediate consequence of Proposition 1, which we will utilize in Proposition 2, is that a rate vector $r \in \mathbb{R}^N_+$ is feasible if and only if the solution of the conditional eigenvalue Problem 1 with the positive concave mapping $\Gamma_r$ satisfies $\lambda^* \leq 1$. We also obtain the following result from Proposition 1:

**Corollary 1.** Let $\lambda^*$ be the eigenvalue that solves Problem 1 in Proposition 1 with the positive concave mapping $\Gamma_r$. As a function of $r \in \mathcal{X}$, $\lambda^*$ is bounded away from 0, i.e., $\exists \epsilon > 0$ such that $\forall r \in \mathcal{X}$, $\lambda^* \geq \epsilon$.

**Proof.** Consider a sequence $(\rho_n)_{n \in \mathbb{N}} \subset \mathcal{Y}$ and sequence of solutions $(\lambda_n^*, \rho_n^*)_{n \in \mathbb{N}}$ to Problem 1. Now suppose that $\forall n \in \mathbb{N}$, $0 \leq \lambda_n^* \leq 1$, which means that there exists a sequence $(\lambda_n^*)_{n \in \mathbb{N}}$ such that $\lambda_n^* \to 0$. By definition, for a given $n \in \mathbb{N}$, if $(\lambda_n^*, \rho_n^*)$ solves Problem 1, then $\Gamma_r(\rho_n^*) = \lambda_n^* \rho_n^* \Leftrightarrow q(\rho_n^*; \frac{1}{\lambda_n^*} r_n) = \rho_n^*$. Define $r_n = \frac{1}{\lambda_n^*} r_n$. It can be seen that $\rho_n^* = q(\rho_n^*; r_n)$. Hence, $\rho_n^*$ is the fixed point of the mapping $\Gamma_{r_n}$, which implies that $r_n$ is feasible since $\rho_n^* \leq 1$. But note that as $\lambda_n^* \to 0$, $\|r_n\|_\infty \to \infty$ which contradicts Lemma 2. So $\exists \epsilon > 0$ such that $\epsilon \leq \lambda^* \leq 1$.

Let $\mathcal{F} : \mathcal{X} \to \mathcal{Y} : r \mapsto \rho \in \text{Fix}(\Gamma_r)$ be the function which maps each $r$ to the unique fixed point of the mapping $\Gamma_r$ in (4). We are now in a position to present the main results of this section, namely that the feasible rate region is compact and the fixed point mapping defined on this set is continuous. These results enable us to apply robust and optimal approximation methods in the next section.

**Proposition 2.** The feasible rate region $\mathcal{X} \subset \mathbb{R}^N_+$ is compact.

**Proof.** We have shown by Proposition 1 that a rate vector $r \in \mathcal{X}$ is feasible if and only if the solution to Problem 1, with the positive concave mapping $\Gamma_r$, satisfies $\{\exists \epsilon > 0 \mid \epsilon \leq \lambda \leq 1\}$. We also obtain the following result from Proposition 1 with positive concave mappings $(\Gamma_r)_{r \in \mathbb{R}^N_+}$ and denote by $(\overline{\rho}, \overline{x})$ its limit which, belongs to the compact set $\mathcal{P} = \{r \in \mathbb{R}^N_+ \mid \|r\|_\infty = 1\}$. Since the set $\mathcal{X}$ is bounded, every sequence of rate demand vectors has a convergent subsequence. Let $(r_n)_{n \in \mathbb{N}} \in \mathcal{X}$, be a convergent subsequence and denote by $r \in \mathbb{R}^N_+$ its limit. Let $(\rho_n^*, \lambda_n^*)_{n \in \mathbb{N}} \in \mathcal{P} \times \mathcal{X}$ be the corresponding convergent sequence of solutions to Problem 1. Since the function $\mathcal{G}$ is continuous, $\lim_{r \to (\overline{\rho}, \overline{x})} \mathcal{G}(\rho_n^*, r_n, \lambda_n^*) = \mathcal{G}(\overline{\rho}, r, \overline{x}) = 0$, so $r \in \mathcal{X}$. Since $(r_n)_{n \in \mathbb{N}}$ is an arbitrary sequence, the above holds for each sequence in $\mathcal{X}$ which implies that $\mathcal{X}$ is closed. Furthermore, in finite dimensional metric spaces every bounded and closed set is compact.

The following theorem shows that the fixed point function is continuous over the compact set $\mathcal{X}$.

**Theorem 1.** Consider the fixed point function $f : \mathcal{X} \to \mathcal{Y} : r \mapsto \rho \in \text{Fix}(\Gamma_r)$. The fixed point function $f$ is uniformly continuous over the compact set $\mathcal{X}$.

**Proof.** Since the set $\mathcal{X}$ is compact, it is sufficient to show that $f$ is continuous on $\mathcal{X}$. Extract a convergent sequence $(r_n)_{n \in \mathbb{N}} \subset \mathcal{X}$, and let the point $r^*$ be its limit which, since $\mathcal{X}$ is compact, exists and belongs to $\mathcal{X}$. It suffices to show that $\lim_{n \to r^*} f(r_n) = f(r^*)$. Extract a convergent sequence $(\rho_n)_{n \in \mathbb{N}} \subset \mathcal{Y}$, of corresponding fixed points in the compact set $\mathcal{Y}$ and denote its limit by $\rho^* \in \mathcal{Y}$. Now consider the function $\mathcal{G}(\rho, r) = \rho - q(\rho; r)$, where $q$ is the load mapping, and note that this function is continuous and $\forall n \in \mathbb{N}$, $\mathcal{G}(\rho_n, r_n) = 0$. It follows from the definition of a continuous function that $\lim_{r \to r^*} \mathcal{G}(\rho_n, r_n) = \mathcal{G}(\rho^*, r^*) = 0$. Therefore, the limit of the sequence $(\rho_n)_{n \in \mathbb{N}}$ is the unique fixed point $f(r^*) = \rho^*$ and $\lim_{r \to r^*} f(r_n) = f(r^*)$. 

\[ \square \]
5. 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.

5.1. Minimax Optimal Approximation

Let the training data set be denoted by \( D = \{ (r^k, \rho^k) \}_{k=1}^K \), \((r^k, \rho^k) \in (\mathcal{X} \times \mathcal{Y}) \), where \( \rho^k := f(r^k) \) are values generated by the fixed point function \( f : \mathcal{X} \rightarrow \mathcal{Y} \). Our objective is to approximate the value \( f(r) \) for any \( r \in \mathcal{X} \), i.e., our objective is to solve an interpolation problem given \( D \).

In the classical approximation theory (see, for example, [19, 20, 21]), the data interpolation problem entails computing an approximation \( g \) of the function \( f \) by observing the values in the set \( D \) and then replacing future evaluations of \( f(r) \) with \( g(r) \) for any \( r \in \mathcal{X} \). We have shown in Theorem 1 that the fixed point function based on the true load model of (4) 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 \( D \). Since we are interested in a robust approximation of the unknown \( f^* \in C(\mathcal{X}) \), we aim at minimizing the worst-case error [20],

\[
E_w(g) = \sup_{f \in C(\mathcal{X})} \| f - g \|_{C(\mathcal{X})},
\]

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

Unfortunately, if the only information about \( f^* \) are the observations in \( D \) and the fact that \( f^* \in C(\mathcal{X}) \), the worst-case error can be arbitrarily large for some appropriately chosen \( g \in C(\mathcal{X}) \). However if \( f^* \) belongs to a compact subset of \( C(\mathcal{X}) \), the sup in (5) is attained and we can guarantee a worst-case finite error. A sufficient condition for compactness of a subset in the space \( C(\mathcal{X}) \) is that all functions in the subset are Lipschitz continuous with the same Lipschitz constant. Moreover, Lipschitz continuity imposes a non-linear restriction on the function class. In this case it has been shown in [19], that for any given \( r \in \mathcal{X} \), the values \( f^*(r) \) belong to a closed interval, and the optimal approximation to \( f^*(r) \) 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 the fixed point function we have proved in the previous section, we make the following assumption:

**Assumption 1.** The fixed point function \( f : \mathcal{X} \rightarrow \mathcal{Y} \) is a (component-wise) Lipschitz monotone function (LIMF) on the set \( \mathcal{X} \) (see Definition 2).

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

**Definition 4 (Optimal Approximation).** Let \( X \subset \mathbb{R}_+^N \) and \( Y \subset [0,1]^M \). Let \( D = \{ (x^k, \rho^k) \in (X \times Y) \}_{k=1}^K \), be a data set and assume that \( \rho^k := f(x^k), k = 1, \ldots, K \), are values generated by an unknown function \( f \in \mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y} \), where \( \mathcal{F} \subset C(\mathcal{X}) \) is a set of LIMF functions. The minimax optimal approximation (or optimal recovery) problem is stated as follows:

**Problem 3.** \([20, 11, 22]\) Find \( g : \mathcal{X} \rightarrow \mathcal{Y} \), such that

\[
g \in \arg \min_{g \in S} E_{\max}(g)
\]

where \( S := \{ g \in C(\mathcal{X}) | g(x^k) = f(x^k), \forall k \in \{1, \ldots, K\} \} \), and \( E_{\max}(g) := \max_{f \in \mathcal{F}} \|f - g\|_{C(\mathcal{X})} \) is the worst-case error from (5) 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 \([20, 21]\), that can be used to construct an optimal solution to Problem 3. The following Fact summarizes the important properties of an optimal solution constructed using this framework.

**Fact 2.** Let \( D = \{ (x^k, \rho^k) \in (X \times Y) \}_{k=1}^K \), be a dataset generated by an unknown function \( f \in \mathcal{F} \), where \( \mathcal{F} \) is a set of LIMF functions (see Definition 2). Then, we have the following:

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

\[
(\forall i \in M)(\forall x \in \mathcal{X}) g_i(x) = \frac{\sigma^i_1(x) + \sigma^i_2(x)}{2},
\]

where \( \sigma^i_1(x) = \max_k \{ f_i(x^k) - L_i \| (x - x^k)_+ \| \} \), \( \sigma^i_2(x) = \min_k \{ f_i(x^k) + L_i \| (x - x^k)_+ \| \} \), \( f_i(x^k) = \rho^k_i \), and \( L_i \) is the Lipschitz constant of the \( i \)th component \( f_i \).

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

6. ALGORITHM AND SIMULATION

We consider a neighborhood with \( M = 10 \) BS sites and \( N = 50 \) TPs placed randomly. Each TP is connected to a single 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 (7) is computed independently at each BS.

6.1. Noisy Training Data

Practical systems are subject to noise during measurement, so that instead of a data set \( D = \{ (r^k, f(r^k)) \}_{k=1}^K \), a noisy
Algorithm 1: Load Estimation At Each BS

1: **Training:** (a) Collect the training data set $D^\text{noise} = \{(r^k, y^k = f(r^k) + \epsilon(r^k))\}_{k=1}^K$. (b) Perform the estimation of $L$ and data smoothing to construct the compatible $D = \{(r^k, \rho^k)\}_{k=1}^K$ as described in Section 6.1.
2: **Online Prediction:** Given a new rate demand vector $r \in \mathcal{X}$, perform the direct computation of Fact 2.

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

training data set $D^\text{noise} = \{(r^k, y^k = f(r^k) + \epsilon(r^k))\}$ is available, where $\epsilon(r^k)$ 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 $L := \max_{k \neq j} \frac{|y^k - y^j| - n r^k|}{\| r^k - r^j \|}$, where $\epsilon := \sup_k \epsilon(r^k)$ [23]. The monotone-smoothing problem is given by a linear program (LP) [11]

$$\min_{q^+_k, q^-_k \geq 0} \sum_{k=1}^K |q^k|$$

s.t. $q^k - q^j \leq y^j - y^k + L \| (r^k - r^j)_+ \|$, 

$\forall k, j \in \{1, 2, \ldots, K\}$

where $q^k = q^+_k - q^-_k$, $|q^k| = q^+_k + q^-_k$, 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 or LP solver.

### 6.2. Implementation and Complexity

Note that, for a given $r \in \mathcal{X}$ each BS $i \in \mathcal{M}$ can calculate the component $g_i(r)$ independently of other BSs using (7) by Algorithm 1. Therefore, Algorithm 1 is scalable to a larger network. The training step can be performed by standard convex solvers whereas the complexity of the prediction step, which is a direct computation, is linear in sample size $K$, i.e, $O(K)$. Therefore for small sample sizes considered in this study, Algorithm 1 is fast.

### 6.3. Results

We train the network over the set $\mathcal{X}' = [r^\text{min}, r^\text{max}] \subset \mathbb{R}^{50}$, where $r^\text{min} \geq (10^6) \mathbf{1}$ and $r^\text{max} \leq (10^7) \mathbf{1}$ is the pre-configured feasible range (in bits/s) of rate vectors, where $\mathbf{1} \in \mathbb{R}_{++}^{50}$ is the ones vector. We calculate the cell-load values using the fixed point iterative method [14] with the cell-load mapping (4). Other important parameters are: $RB = 20$ MHz, $p_i = 1$ W, $\sigma^2 = 1.38 \times 10^{-23} \times 300/20 \times 10^5$.

Normally distributed random 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 $\mathcal{X}' = [r^\text{min}, r^\text{max}]$ 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 $f$. Note that even though the cost function (5) 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.

![Fig. 1. RMSE and (Pearson’s) Correlation coefficient](image-url)
7. 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, 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] G. H. Golub and L. B. Smith, “Algorithm 414: Chebyshev approximation of continuous functions by a chebyshev system of functions,” Commun. ACM, vol. 14, no. 11, pp. 737–746, Nov. 1971.

[16] U. Krause, “Concave perronfrobenius theory and applications,” Nonlinear Analysis: Theory, Methods & Applications, vol. 47, no. 3, pp. 1457 – 1466, 2001.

[17] R. L. G. Cavalcante and S. Stanczak, “The role of asymptotic functions in network optimization and feasibility studies,” ArXiv e-prints, June 2017.

[18] C. J. Nuzman, “Contraction approach to power control, with non-monotonic applications,” in IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference, Nov 2007, pp. 5283–5287.

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

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

[21] J.F. Traub and H. Wo ´zniakowski, A general theory of optimal algorithms, ACM monograph series. Academic Press, 1980.

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

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