

Weighted Sum-Rate Maximization With Causal Inference for Latent Interference Estimation

Lei You

Department of Engineering Technology, Technical University of Denmark, Denmark
leiyo@dtu.dk

Abstract—The paper investigates the weighted sum-rate maximization (WSRM) problem with latent interfering sources outside the known network, whose power allocation policy is hidden from and uncontrollable to optimization. The paper extends the famous alternate optimization algorithm weighted minimum mean square error (WMMSE) [1] under a causal inference framework to tackle with WSRM. Specifically, with the possibility of power policy shifting in the hidden network, computing an iterating direction based only on the observed interference inherently implies that counterfactual is ignored in decision making. A method called synthetic control (SC) is used to estimate the counterfactual. For any link in the known network, SC constructs a convex combination of the interference on other links and uses it as an estimate for the counterfactual. Power iteration in the proposed SC-WMMSE is performed taking into account both the observed interference and its counterfactual. SC-WMMSE requires no more information than the original WMMSE in the optimization stage. To our best knowledge, this is the first paper explores the potential of SC in assisting mathematical optimization in addressing classic wireless optimization problems. Numerical results suggest the superiority of the SC-WMMSE over the original in both convergence and objective.

I. INTRODUCTION

The wireless network evolution has been driven by a need for consistently higher rates over the past decades, along with the emergence of the internet of everything (IoE), which aims to serve as a platform for connecting processes, people and data [2]. The requirement for high spectral efficiency in IoE scenarios calls for a need to revisit classic wireless optimization problems under highly dynamic environments. The weighted sum-rate maximization (WSRM) problem [3] is one of these many, solving which plays a key role in determining the effective capacity of a wireless channel. Finding the global maximum of WSRM is generally considered to be an \mathcal{NP} -hard problem [3].

As a result, significant research efforts have been devoted to developing high-quality sub-optimal solutions for the problem. The weighted minimum mean square error (WMMSE) algorithm, first proposed in [1], is one such solution which has been shown to be an efficient algorithmic framework for many cross-layer transmission tasks [3]. Therefore, this algorithmic framework has been significantly extended by many other literature and the research exploring remains still active [4]–[9]. Additionally, the WSRM problem has also been addressed by data-driven methodologies—more specifically—graph neural

networks (GNN) based methods [10]–[13]. The basic idea is to use a GNN to encode the network topology information into a GNN that maps channel state information (CSI) to power control policy. Unsupervised model training is also possible by in-cooperating learning in the processing of a primal-dual algorithm [10].

The single-input-single-output (SISO) case of [4]–[13] and many other research summarized in [14] falls into a special case of the problem investigated in this paper, i.e. when there are no latent interfering links whose power allocation is unknown and uncontrollable. It's important to note that *the latent interfering links cannot be treated simply as noise*, because the power allocation of the latent links may be dependent on those of the known links, or, they may evolve over time. Treating it as noise omits these facts.

The paper examines the WMMSE optimization mechanism from a causal inference's perspective. WMMSE algorithm and other similar methods can suffer performance loss in this scenario because of the influence of these latent factors on the iterating directions towards the ground-truth optimality. To overcome this challenge, the paper proposes the use of the synthetic control (SC) method [15], which uses counterfactual reasoning to infer the causal relationship between optimization variables and the objective. The resulting algorithm requires no additional information beyond that used in the traditional WMMSE algorithm, yet it demonstrates improved convergence and objective value, as well as resilience to emerging and disappearing latent interference sources. The trick in this paper hence applies to most of its previous extensions [3]–[9]. Numerically, the proposed SC-WMMSE shows advantage over its origin on both convergence and objective value as well as demonstrates resilience to emerging and disappearing of latent interference sources.

II. MODEL AND PROBLEM

Consider a wireless network consisting of multiple communications links. The set of all links is denoted by \mathcal{K} , and for any link k ($k \in \mathcal{K}$), the set of links that interfere with k is denoted by \mathcal{I}_k , where $\mathcal{I}_k \subset \mathcal{K}$. The channel gain between the transmitter and receiver of link k is denoted by $h_{k,k}$, and the channel gain from the transmitter of link k to the receiver of link j is denoted by $h_{k,j}$ ($k \neq j$). It is assumed that $|h_{k,j}| = 0$ for any $j \notin \mathcal{I}_k$ ($j \neq k$), and the channel matrix is denoted by

$$R_k = \log \left(1 + \frac{|h_{k,k}|^2 p_k}{\sum_{j \in \mathcal{K}^+} |h_{j,k}|^2 p_j + \sum_{i \in \mathcal{K}^-} |h_{i,k}|^2 q_i + \sigma_k^2} \right) \quad k \in \mathcal{K}^+ \quad (1)$$

H. The power of the transmitter of link k is denoted by p_k , and the noise power is denoted by σ_k^2 .

The paper considers a scenario where there are latent interference sources whose power distribution and allocation is not known a priori and is neither observable nor controllable. The known network is denoted by \mathcal{K}^+ , and the unknown network is denoted by \mathcal{K}^- , with $\mathcal{K}^+ \cup \mathcal{K}^- = \mathcal{K}$ and $\mathcal{K}^+ \cap \mathcal{K}^- = \emptyset$. Let $K = |\mathcal{K}^+|$. The two networks mutually affect each other through interference. No assumptions are made about the power policies of \mathcal{K}^- . The notation \mathbf{q} is used to represent the power allocation in \mathcal{K}^- , which can proactively or reactively change over time and may be dependent on \mathbf{p} . The problem is defined in (2), with R_k defined in equation (1). The goal is to find a power allocation p_1, p_2, \dots, p_K for all the corresponding transmitters of the K links in the known network such that the weighted sum rate $\sum_k \alpha_k R_k$ of the known network is maximized.

$$\max_{\mathbf{p}} \sum_{k \in \mathcal{K}^+} \alpha_k \mathbb{E}_{\mathbf{H}} \left[\mathbb{E}_{\mathbf{q}|\mathbf{H}, \mathbf{p}} [R_k] \right] \quad (2a)$$

$$\text{s.t.} \quad 0 \leq p_k \leq p_k^{\max}, \quad k \in \mathcal{K}^+ \quad (2b)$$

III. SUM RATE MAXIMIZATION WITH CAUSAL INFERENCE

A. How Latent Interference Affects WMMSE

It is shown that (2) submits to a reformulation as a weighted sum-mean-square-error minimization problem [4], [5] when the power allocation vector of the unknown network, \mathbf{q} , is fixed as constants. When \mathbf{q} is fixed, the mathematical expectation on \mathbf{q} is removed from the problem.

To make this reformulation, the noise plus interference term from the unknown network \mathcal{K}^- is replaced by a new variable η_k . Denote $\eta_k = \sum_{i \in \mathcal{K}^-} |h_{i,k}|^2 q_i + \sigma_k^2$. This replacement can be made in the denominator of the rate expression in (1), and the problem in (2) can be rewritten as (3) without loss of optimality [4], [5].

$$\min_{\mathbf{w}, \mathbf{u}, \mathbf{v}} \mathbb{E}_{\mathbf{H}} \left[\sum_{k \in \mathcal{K}^+} \alpha_k (w_k e_k - \log w_k) \right] \quad (3a)$$

$$\text{s.t.} \quad |v_k|^2 \leq p_k^{\max}, \quad k \in \mathcal{K}^+ \quad (3b)$$

The reformulation of the problem in (2) to a weighted sum-mean-square-error minimization problem utilizes two ad-

ditional variables: w_k and e_k , where w_k is a positive weight variable, and e_k is the mean-square error, defined as:

$$e_k = |u_k h_{k,k} v_k - 1|^2 + \sum_{j \neq k} |u_j h_{j,k} v_k|^2 + \eta_k |u_k|^2 \quad (4)$$

The variable p_k in the original formulation is replaced by v_k , with $p_k = |v_k|^2$ ($k \in \mathcal{K}^+$).

The WMMSE algorithm is designed to solve the reformulation in (3) using the theory of alternate optimization. The algorithm is illustrated in Algorithm 1. It should be noted that there also exists a stochastic version of the WMMSE algorithm that can be used when the channel matrix, \mathbf{H} , is a random variable. The proposed methodology in this paper can also be applied to the stochastic version of the WMMSE algorithm. In practice, the variable η_k can be approximated using techniques such as clear channel assessment (CCA) [16].

Algorithm 1: The WMMSE algorithm for solving the problem in (3).

- 1: Initialize $\mathbf{v}, \mathbf{u}, \mathbf{w}$ randomly
 - 2: **repeat**
 - 3: Observe $\eta_1, \eta_2 \dots \eta_K$
 - 4: **for all** $k = 1, 2 \dots K$ **do**
 - 5: $u_k = |h_{k,k} v_k| / (\sum_{j \in \mathcal{K}^+} |h_{j,k}|^2 |v_j|^2 + \eta_k)$
 - 6: $w_k = 1 / (1 - |u_k h_{k,k} v_k|)$
 - 7: $v_k = \alpha_k h_{k,k} u_k w_k / (\sum_{j \in \mathcal{K}^+} \alpha_j w_j |h_{k,j} u_j|^2 + \lambda_k)$
 - 8: **until** Convergence
 - 9: **return** $\mathbf{v}, \mathbf{u}, \mathbf{w}$
-

The term e_k is a convex quadratic function in \mathbf{u} and \mathbf{v} , which have closed-form solutions. However, when \mathbf{q} is a latent random variable, it can change throughout the optimization process and affect the sampling of η_k , making the optimization process challenging. This is because the collected samples of η_k may not guarantee that u_k will move towards the optimum of the quadratic function e_k when \mathbf{w} is fixed. This problem can be mitigated if η_k is i.i.d., yet such assumption is too strong to be realistic.

B. Causal Inference Estimator vs. Regression Estimator

The key obstacle in applying the WMMSE algorithm is that the interference at any receiver k ($k \in \mathcal{K}^+$) is difficult to handle from an optimization perspective, due to its dependency on \mathbf{q} . To tackle this obstacle, instead of considering only η_k , the entire denominator of u_k , denoted by I_k , is considered. This allows for better generalization of the proposed inference

methodology, as it can estimate the denominator as a whole for any link k if the CSI of another link j ($j \neq k$, $j \in \mathcal{K}^+$) is unknown or outdated. In the k_{th} loop of Algorithm 1, the power allocation p_k is the variable to be optimized and the other power \mathbf{p}_{-k} are fixed. The mathematical expectation of I_k conditional on the optimization variable p_k is as follows.

$$\begin{aligned} \mathbb{E}[I_k | p_k] &= \mathbb{E}_{\mathbf{H}} \mathbb{E}_{\mathbf{q} | \mathbf{H}, p_k} \left[\sum_{j \in \mathcal{K}^+} |h_{j,k}|^2 p_j + \sum_{i \in \mathcal{K}^-} |h_{i,k}|^2 q_i + \sigma_k^2 \right] \\ &= \underbrace{\mathbb{E}_{\mathbf{H}} \left[\sum_{j \in \mathcal{K}^+} |h_{j,k}|^2 p_j \right]}_{\eta_k} + \mathbb{E}_{\mathbf{H}} \mathbb{E}_{\mathbf{q} | \mathbf{H}, p_k} \left[\sum_{i \in \mathcal{K}^-} |h_{i,k}|^2 q_i + \sigma_k^2 \right] \end{aligned}$$

Estimating the second part of I_k , η_k , is difficult as it depends on the power allocation of the latent interfering sources, which is unknown and may change over time. Supervised machine learning models are limited for this task as the distribution of \mathbf{q} is not i.i.d and may change due to changes in the power allocation of the latent sources or new transmitters/receivers joining or leaving the network. A more suitable approach is to leverage causal inference methods, specifically the structural causal model (SCM) approach, to inexplicitly estimate the counterfactual distribution of \mathbf{q} given the observed interference I_k . This approach allows for estimation of the interference distribution under different power allocation policies, which can then be used to infer the causality relationship between p_k and I_k .

A causal inference estimator targets $\mathbb{E}[I_k | do(p_k)]$, which takes both the actual outcome and potential outcomes into consideration, rather than just the former. Specifically, $\mathbb{E}[I_k | do(p_k)]$ represents the expected value of the interference I_k at receiver k when we intervene and set the power allocation of link k to p_k , and consider all possible power policies of the latent interfering links \mathcal{K}^- . On the other hand, $\mathbb{E}[I_k | p_k]$ represents the expected value of the interference I_k at receiver k when the power allocation of link k is p_k , but the power policies of the latent interfering links \mathcal{K}^- are not intervened, and are instead determined by the current distribution of \mathbf{q} given \mathbf{H} and \mathbf{p} . $\mathbb{E}[I_k | do(p_k)]$ better aligns with the requirements of optimization.

For $\mathbb{E}[I_k | do(p_k)]$ and $\mathbb{E}[I_k | p_k]$ to be equal, the power allocation p_k is independent of all the other interfering links and their power allocations in the network, i.e., there are no confounding variables. In this case, conditioning on p_k would not change the distribution of I_k and the expectation would be the same whether we condition on p_k or intervene on it. This implies that power allocation in \mathcal{K}^+ is being performed randomly, rather than through a specific optimization algorithm.

C. Estimating Interference with SC Methods

The synthetic control method (SC), first proposed in [15], is a powerful technique for estimating the effects of large-

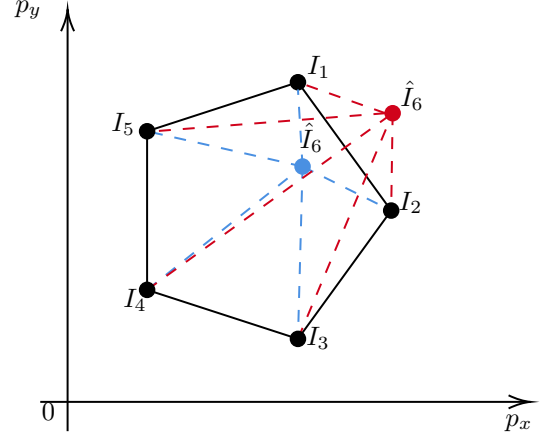


Figure 1. Illustration of interpolation vs. extrapolation. Consider $\mathcal{K}^+ = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{K}^- = \{x, y\}$. Note that q_x and q_y are latent variables to \mathcal{K}^+ , influencing the observed interference at each link of \mathcal{K}^+ . The positions of I_1, I_2, \dots, I_5 are affected by the value of q_x and q_y at the moment of observation. Estimations are performed for the link 6, denoted by \hat{I}_6 with blue and red dots. The blue is from intertropolation, as it falls in-between the known observations (i.e. the convex hull), whereas the red is from extrapolation.

scale interventions [17]. It has been shown to be competitive with other fixed matching estimators [18]. SC approximates the counterfactual outcomes of one unit by constructing a weighted combination of the observed outcomes of other units. It works with panel data, which contains multiple observations for each unit over time. While SC is often used in the context of binary treatment and no interference between units, it can also be applied in more complex settings with units interfering each others, such as in [19], where the potential outcome for unit $k \in \mathcal{K}^+$ is linear to latent factors, with noise in the factor model being additive, zero mean, and independent.

In the training stage, consider observations of I_1, I_2, \dots, I_k in format of panel data, i.e.

$$\mathbf{X} = \begin{bmatrix} I_1^{(0)} & I_2^{(0)} & \dots & I_K^{(0)} \\ I_1^{(1)} & I_2^{(1)} & \dots & I_K^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ I_1^{(L)} & I_2^{(L)} & \dots & I_K^{(L)} \end{bmatrix} \in \mathbb{R}^{L \times K}$$

where each row ℓ ($1 \leq \ell \leq L$) is an observation across all the K units. For any $k \in \mathcal{K}^+$, denote by \mathbf{x}_k the k_{th} column of \mathbf{X} . Denote by \mathbf{X}_{-k} the matrix without column k . The SC estimator is trained by solving the constrained optimization problem (5) as follows.

$$\boldsymbol{\nu}_k = \arg \min_{\boldsymbol{\beta} \geq 0} \|\mathbf{x}_k - \mathbf{X}_{-k} \boldsymbol{\beta}\| \quad (5a)$$

$$\text{s.t.} \quad \sum_i \beta_i = 1 \quad i = 1, 2, \dots, K-1 \quad (5b)$$

Solving (5) yields a vector $\boldsymbol{\nu}_k$ ($k \in \mathcal{K}^+$), which is essentially a group of coefficients that can be used to construct a linear combination of I_j ($j \in \mathcal{K}^+ \setminus k$). The objective function (5a)

suggests that the computed coefficients lead to an as small as possible mean-squared-error over all the L observations of the unit k and respectively its constructed linear combinations. Note that (5b) imposes a hard constraint on the coefficients such that the obtained linear combination is guaranteed to be a convex combination. In other words, solving the formulation (5) trains a (constrained) linear regression model between the observed interference of unit k and those of the others. In this context, a convex combination has better interpretability. Namely, in terms of interference, it finds a function that accurately estimates the values of the counterfactual interference in the range of the observed interference, as opposed to extrapolation. See Figure 1 for an illustration. Section IV demonstrates the necessity of constraint (5b) by showing numerically that it helps improve optimization significantly.

In the inference stage, the synthetic control method is used to estimate the counterfactual outcome for unit k under treatment p_k . This is done by averaging the outcomes of the synthetic control units ($\mathcal{K}^+ \setminus \{k\}$), weighted by the coefficients obtained in the training stage. The estimation is denoted as $\hat{I}_k = \boldsymbol{\nu}_k^\top \boldsymbol{\mu}_k$ for all $k \in \mathcal{K}^+$, where $\boldsymbol{\mu}_k$ denotes the observations for all units other than k , i.e. any I_j ($j \neq k$).

D. Algorithm Design

The algorithm SC-WMMSE is designed straightforwardly by incorporating the ideas discussed in Sections III-A, III-B, and III-C. It is presented in Algorithm 2.

Algorithm 2: The SC-WMMSE algorithm for solving the problem in (2)

- 1: Initialize $\mathbf{v}, \mathbf{u}, \mathbf{w}$ randomly
 - 2: Train estimators $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \dots \boldsymbol{\nu}_K$ offline by (5)
 - 3: **repeat**
 - 4: Observe I_1, I_2, \dots, I_K under \mathbf{v}
 - 5: **for all** $k = 1, 2 \dots K$ **do**
 - 6: $\boldsymbol{\mu}_k = [I_1, \dots, I_{k-1}, I_{k+1}, \dots, I_K]$
 - 7: **if** $\text{Rand}() < \varepsilon$ **then**
 - 8: $u_k = |h_{k,k} v_k| / \boldsymbol{\nu}_k^\top \boldsymbol{\mu}_k$ // Counterfactual update
 - 9: **else**
 - 10: $u_k = |h_{k,k} v_k| / I_k$ // Factual update
 - 11: $w_k = 1 / (1 - |u_k h_{k,k} v_k|)$
 - 12: $v_k = \alpha_k h_k u_k w_k / (\sum_{j \in \mathcal{K}^+} \alpha_j w_j |h_{k,j} u_j|^2 + \lambda_k)$
 - 13: **until** Convergence
 - 14: **return** $\mathbf{p} = [v_1^2, v_2^2, \dots, v_K^2]$
-

The algorithm SC-WMMSE combines the WMMSE algorithm with SC method, which is designed to estimate the counterfactual outcome of a unit under a certain treatment. The algorithm is trained offline using past observations of

I_1, I_2, \dots, I_K and K SC estimators are trained. During optimization, I_1, I_2, \dots, I_K are observed in every iteration and the SC estimators are used to update u_k with a probability ε , otherwise, the update follows the same rule as WMMSE. The parameter ε is set to decay based on the number of iterations for the convergence of the algorithm¹. The effectiveness of the algorithm is demonstrated in the simulation section of the paper.

E. Intuition Behind

From a causal inference perspective, the optimization process of WMMSE makes treatment decision p_k by its knowledge of how such an intervene would affect the outcome I_k at each loop k . The covariates are \mathbf{q}, \mathbf{H} (which are related to network topology and density). The SC method is used to estimate the impact of this treatment by comparing the treated group ($\{k\}$) to a SC group, which is created from a weighted combination of untreated groups ($\mathcal{K}^+ \setminus \{k\}$).

Ideally, the SC group ought to be selected carefully to closely mimic the characteristics of the treated group before the treatment is applied. Optimizing the selection of the SC group is out of the scope of this paper. The SC method allows us to estimate what would have happened to the link k if p_k had not been applied, making it a counterfactual analysis. In Algorithm 2, line 8 can be viewed as a "counterfactual update" and line 10 as a regular update. The optimization variable u_k is therefore updated using both the observation and the counterfactual, allowing us to address cases where some observed I_k are not statistically caused by the treatment p_k , but rather by other pre-treatment characteristics.

It is worth noting that the power allocation serves both as the intervention of causal inference and the optimization for sum-rate maximization. How to best utilize this interplay remains an open area of study.

IV. SIMULATION

The simulation setups are as follows: multiple transmitters are randomly and uniformly distributed within a circle with a radius of 200 meters. For each transmitter, multiple target receivers are randomly and uniformly distributed within a circle with a radius of 25 meters. The path-loss model follows the *InH-Shopping Malls-NLOS dual slope* model in [20], which takes into account the the probability of line-of-sight shadow fading and blockage. The model applies across 0.5-100 GHz band and 60 GHz is selected for the simulations in this paper. A flat channel is considered and the total bandwidth is 80 MHz. The maximum power on a resource unit (RU) is set to 200 mW uniformly for all links.

A random matrix \mathbf{Z} is generated following a uniform distribution $\mathcal{U}(-1, 1)$ unless specified otherwise. The power

¹In the implementation of this paper, the formula $\varepsilon(t) = [a(1-t/t^{\max})]^b$ is used, where t is the iteration index and a, b are hyper-parameters. As for this paper, the setting $a = 0.2$ and $b = 2$ stays unchanged throughout all simulations in Section IV.

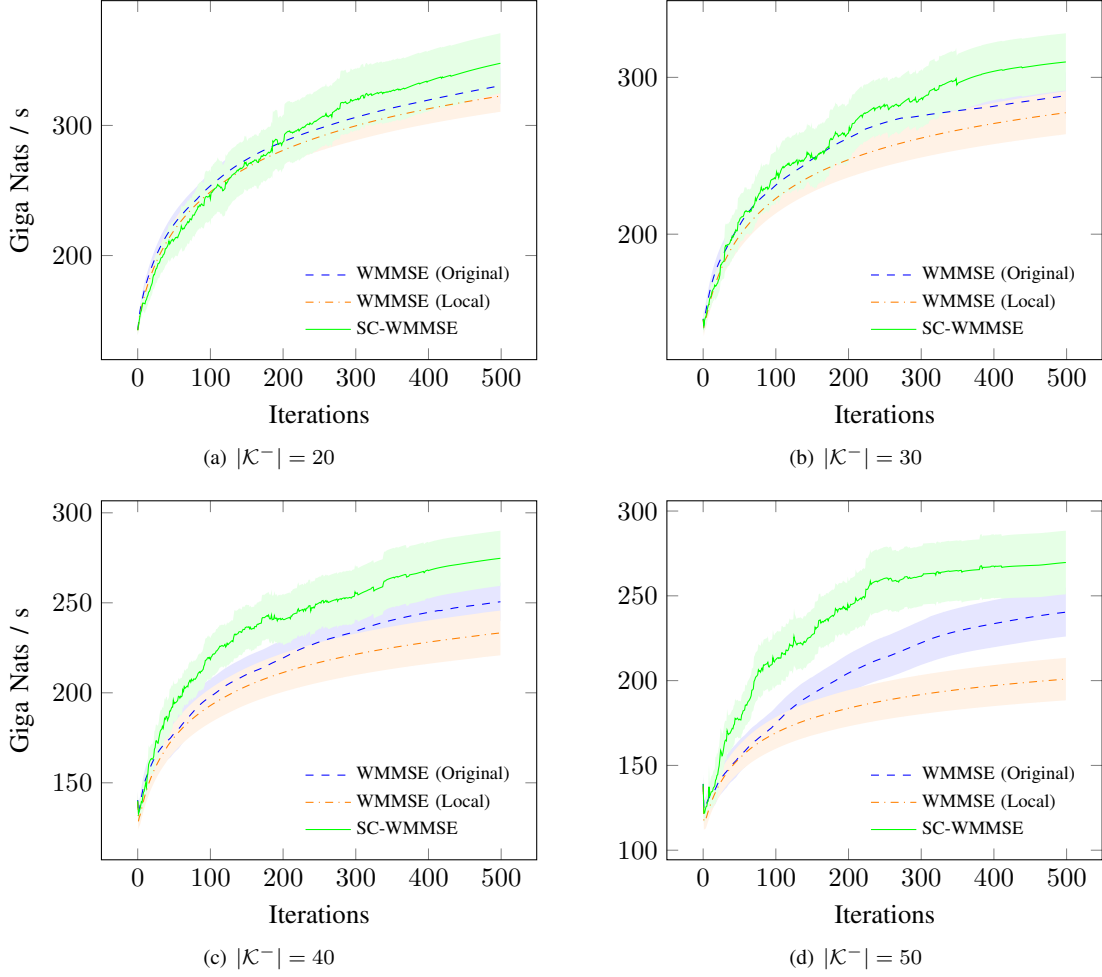


Figure 2. The performance in objective maximization and convergence is evaluated over 50 independent simulations. The 90% confidence interval is plotted for each curve. The setup $|\mathcal{K}^+| = 50$ is used throughout 2(a)–2(d). The y-axis is the sum-rate over all links in \mathcal{K}^+ .

\mathbf{q} is obtained by the linear transformation $\mathbf{q} = \mathbf{Z}\mathbf{p}$ with randomness and is always capped by the maximum power limit after the transformation². Each figure in this section is based on 50 independent simulations to ensure that the results are statistically significant. In every simulation, the network topology is regenerated by following the rules stated above. Additionally, the random matrix \mathbf{Z} , if used, is refreshed in each simulation, for both the training and inference stages.

The code of the simulation is available on [GitHub](#) [21].

A. Objective Performance and Convergence

This subsection aims to evaluate the effectiveness of the proposed algorithm, SC-WMMSE, in optimizing the objective function in (2) and its performance in terms of convergence. The original WMMSE algorithm is used as a baseline for comparison, as well as a version of WMMSE that only uses local network information for optimization, referred to as WMMSE

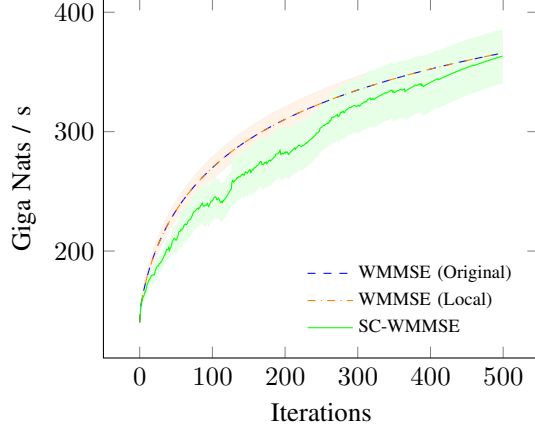
(Local). In the local version, the term η_k in Algorithm 1 is set to σ_k^2 , effectively ignoring all changes in \mathcal{K}^- . This is used to gauge the impact of \mathcal{K}^- on \mathcal{K}^+ and to determine if such impact is too small to expect SC-WMMSE having an effect.

The results are shown in Figure 2 with $|\mathcal{K}^+| = 50$ and $|\mathcal{K}^-|$ ranging from 20 to 50. SC-WMMSE consistently outperforms the baseline in terms of objective function maximization. Particularly, when the transmissions in the latent network \mathcal{K}^- are dense, the algorithm demonstrates significantly better performance in both objective maximization and convergence. On the other hand, when the transmissions in \mathcal{K}^- are sparse, WMMSE (Local) performs similarly to the original WMMSE. This suggests that the impact of \mathcal{K}^- on \mathcal{K}^+ is low, and the term η_k can be well approximated by σ_k^2 . As a result, the update in line 8 becomes less significant in the optimization process.

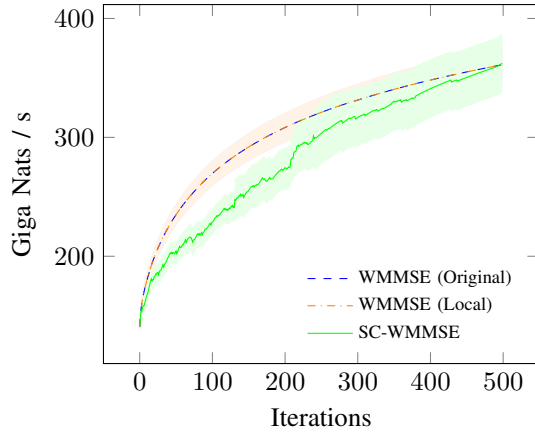
B. Robustness

The robustness of SC-WMMSE in dynamic network changes is evaluated in this section. The algorithm's ability to adapt to

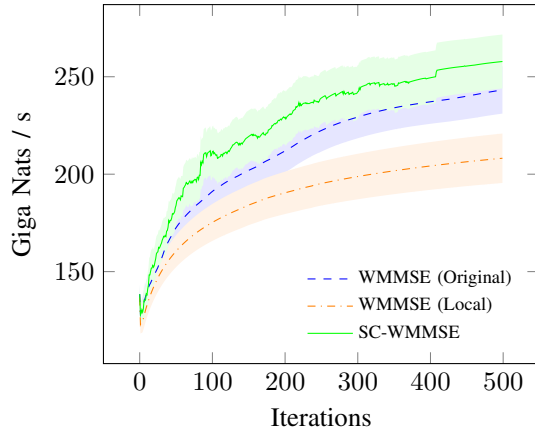
²Note that \mathbf{p} in the optimization process depends on the iteration step (and hence time), making \mathbf{q} is actually evolving with time too.



(a) $|\mathcal{K}^-| = 0$ at both the training and the inference stages



(b) $|\mathcal{K}^-| = 50$ at training. $|\mathcal{K}^-| = 0$ at inference.



(c) $|\mathcal{K}^-| = 0$ at training. $|\mathcal{K}^-| = 50$ at inference.

Figure 3. In this evaluation, the algorithm's ability to adapt to changes in the network is tested by introducing or removing latent interference sources, while keeping the number of links in \mathcal{K}^+ constant at 50. The scenario in which no latent interference exists in either the training or testing phase is used as a baseline, as seen in Figure 3(a). The results are based on 50 independent simulations, with a 90% confidence interval plotted.

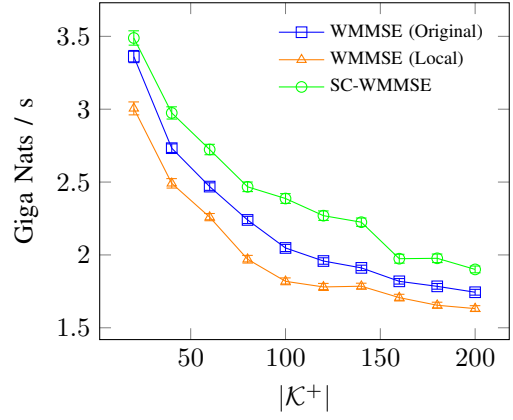


Figure 4. Scalability is evaluated for the proposed algorithm over 50 independent simulations, with $|\mathcal{K}^-| = 50$. The results are presented with a 99 percent confidence interval, which are visually narrow. The y-axis represents the average throughput per link, which is calculated by dividing the sum-rate achieved on average over 500 algorithm iterations by the number of links.

emergence or disappearance of interference sources in \mathcal{K}^- is in question. Figure 3(a) is used as a baseline scenario with no latent interference links. SC-WMMSE is not expected to perform better than the others since there is no latent interference. The algorithm's performance is on par with the others in terms of objective value at convergence, but the convergence is slower due to added noise via the convex hull approximation \hat{I}_k ($k \in \mathcal{K}^+$). Figure 3(c) illustrates SC-WMMSE's robustness in dealing with changes in the network. Even though there is no latent interference present during training, the causal estimator can still make impact in the optimization stage, resulting in SC-WMMSE outperforming the other two in terms of both objective value and convergence. However, the performance is not as good as seen in Figure 2(d) due to the more significant data distribution change between training and testing.

C. Scalability

The scalability of the algorithm is evaluated by analyzing its performance on per-link throughput as the density of transmissions in the optimized network, \mathcal{K}^+ , increases. Results show that when the transmissions in \mathcal{K}^+ are sparse, both WMMSE and SC-WMMSE perform similarly. However, as the number of links in \mathcal{K}^+ increases, the advantage of SC-WMMSE becomes more pronounced, until reaching a threshold around 150 links. Beyond this point, the advantage of SC-WMMSE starts to decrease, possibly due to the inherent limitations of the optimization mechanism used by all three algorithms when dealing with large scale problems. It is also worth noting that the gap between WMMSE and its local version also shrinks with an increase in the number of links in \mathcal{K}^+ . Overall, SC-WMMSE demonstrates better performance than its baseline WMMSE in all scenarios in Figure 4.

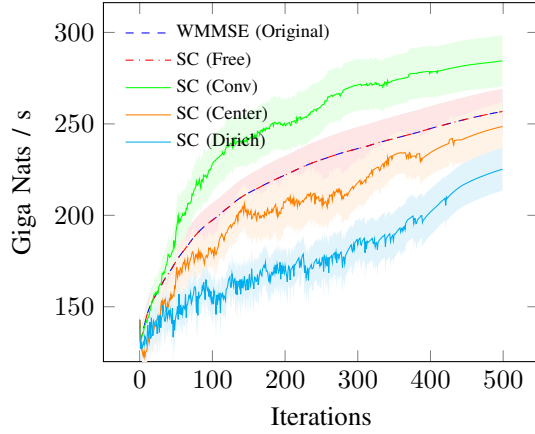


Figure 5. SC (Free) is the unconstrained version of (5), i.e., SC (Conv). Additionally, the evaluation includes two other methods: SC Center, which uses the center point of the convex hull for estimation of all units; and SC (Dirich), which generates a random Dirichlet distribution as the coefficients β and selects a random point inside the convex hull in each iteration step.

D. Necessity of the Convexity Constraint

The effectiveness of the convex combination constraint in (5b) for optimization is demonstrated in Figure 5. On top of the synthetic control method used in this paper (the method of solving (5), referred to as SC (Conv)), several other methods are compared, including its unconstrained version (SC (Free)). Additionally, the center point of the convex hull can be used as the estimate (SC (Center)). One could also generate a Dirichlet distribution for the coefficients (SC (Dirich)). These methods are used in place of the convex combination constraint in Algorithm 2 to compare with the baseline WMMSE.

SC (Free) performs similarly to the baseline WMMSE, which suggests that relaxing the constraint leads to an overfitting to observation, neglecting the counterfactual. Both SC (Center) and SC (Dirich) perform worse than the baseline, indicating that the inference task is not trivial and that SC (Conv) is by-far an effective method for obtaining the coefficients for optimization.

V. CONCLUSION AND DISCUSSION

This paper has showcased the potential of using causal inference to assist in solving optimization problems of wireless communications. The proposed SC-WMMSE demonstrates how a causal inference framework can address the challenges posed by confounding variables in optimization. As future work, further analysis and theoretical proofs of the algorithm's convergence are suggested. Additionally, the research in causal inference has typically been based on the assumption of no inter-unit interference, which is not applicable in wireless communications. Therefore, it would be beneficial to develop causal factor models that are better suited for this field.

REFERENCES

- [1] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.
- [2] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.
- [3] H. Baligh, M. Hong, W.-C. Liao, Z.-Q. Luo, M. Razaviyayn, M. Sanjabi, and R. Sun, "Cross layer provision of future cellular networks," *arXiv preprint arXiv:1407.1424*, 2014.
- [4] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 1106–1110.
- [5] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [6] J. Shin and J. Moon, "Weighted-sum-rate-maximizing linear transceiver filters for the k-user mimo interference channel," *IEEE transactions on communications*, vol. 60, no. 10, pp. 2776–2783, 2012.
- [7] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex mimo interference channels," *IEEE Transactions on communications*, vol. 63, no. 3, pp. 801–815, 2015.
- [8] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell mimo networks," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1590–1608, 2017.
- [9] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.
- [10] M. Eisen and A. Ribeiro, "Optimal wireless resource allocation with random edge graph neural networks," *IEEE transactions on signal processing*, vol. 68, pp. 2977–2991, 2020.
- [11] N. Naderializadeh, M. Eisen, and A. Ribeiro, "Wireless power control via counterfactual optimization of graph neural networks," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020, pp. 1–5.
- [12] A. Chowdhury, G. Verma, C. Rao, A. Swami, and S. Segarra, "Unfolding wmmse using graph neural networks for efficient power allocation," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 6004–6017, 2021.
- [13] I. Nikoloska and O. Simeone, "Fast power control adaptation via meta-learning for random edge graph neural networks," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2021, pp. 146–150.
- [14] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, C. Fischione *et al.*, "Weighted sum-rate maximization in wireless networks: A review," *Foundations and Trends® in Networking*, vol. 6, no. 1–2, pp. 1–163, 2012.
- [15] A. Abadie and J. Gardeazabal, "The economic costs of conflict: A case study of the basque country," *American economic review*, vol. 93, no. 1, pp. 113–132, 2003.
- [16] D. Yang, Y. Xu, and M. Gidlund, "Wireless coexistence between IEEE 802.11 and IEEE 802.15. 4-based networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 7, no. 1, p. 912152, 2011.
- [17] C. Shi, D. Sridhar, V. Misra, and D. Blei, "On the assumptions of synthetic control methods," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 7163–7175.
- [18] J. Chen, "Synthetic control as online linear regression," *arXiv preprint arXiv:2202.08426*, 2022.
- [19] A. Agarwal, S. Cen, D. Shah, and C. L. Yu, "Network synthetic interventions: A framework for panel data with network interference," *arXiv preprint arXiv:2210.11355*, 2022.
- [20] K. Haneda, L. Tian, H. Asplund, J. Li *et al.*, "Indoor 5g 3gpp-like channel models for office and shopping mall environments," in *2016 IEEE ICC*. IEEE, 2016, pp. 694–699.
- [21] [Online]. Available: <https://github.com/youlei202/research-weighted-sum-rate>