Abstract—Intelligent reconfigurable surface (IRS) is becoming an attractive component of cellular networks due to its ability to shape the propagation environment and thereby improve coverage. While IRS nodes incorporate a great number of phase-shifting elements and a controller entity, the phase shifts are typically determined by the cellular base station (BS) due to its computational capability. Since controlling a large number of phase shifts may become prohibitive in practice, it is important to reduce the control overhead between the BS and the IRS controller. To this end, in this paper, we propose a low-rank modeling approach for the IRS phase shifts. The key idea is to represent the IRS phase shift vector using a low-rank tensor approximation model, where each rank-one component is modeled as the Kronecker product of a predefined number of factors of smaller sizes, obtained via tensor decomposition algorithms. We show that the proposed low-rank models drastically reduce the required feedback requirements associated with the BS-IRS control links. Our simulation results indicate that the proposed method is especially attractive in scenarios with a strong line of sight component, in which case nearly the same spectral efficiency is reached as in the cases with near-optimal phase shifts, but with significantly lower feedback overhead.

I. INTRODUCTION

Intelligent reconfigurable surface (IRS) is a candidate technology for beyond fifth-generation and sixth-generation networks due to its ability to control the electromagnetic properties of the radio-frequency waves by performing an intelligent phase shift to the desired direction [2], [3], [4], [5], [6], [7], [8], [9]. Usually, IRS is defined as a planar (2-D) surface with a large number of independent reflective elements, in which they can be fully passive or with some elements active [10], [11], [12], [13]. IRS is connected to a smart controller that sets the desired phase shift for each reflective element, by applying bias voltages at the elements e.g., positive-intrinsic-negative (PIN) diodes. The main advantage of fully passive IRSs is their full-duplex nature, i.e., no noise amplification is observed since no signal processing is possible. However, the passive nature of the IRS makes the channel state information (CSI) acquisition process difficult, since no pilots can be processed at the IRS, and channel estimation should be carried out at the end nodes of the network. Nevertheless, if the IRS is equipped with a few active elements, some channel estimation capability is available using, for example, compressed sensing tools [10]. An advantage of a passive IRS is its reduced power consumption. This makes the IRS a more attractive technology in terms of energy efficiency than alternative technologies, such as amplify-and-forward and decode-and-forward relays [14], [15], [16].

Several works have addressed the CSI acquisition problem in IRS-assisted networks, e.g., [17], [18], [19], [20], [21], and [22]. The work of [17] proposes a tensor-based method where the authors show the benefits of exploiting the multidimensional structure of the received signal by separating the cascade channel. Reference [19] proposes a compressed sensing approach in a multi-user uplink multiple-input multiple-output (MIMO) scenario. In [20], a two-timescale channel estimation framework is proposed to overcome the pilot overhead in a multi-user IRS-aided system. In [21] and [22], the authors consider IRSs operating under hardware impairments, whereas [21] tackle this problem in millimeter-wave MIMO systems, and [22] uses a tensor-based approach to...
jointly estimate the channels and the IRS impairment matrix. The work of [22] proposes a channel estimation framework for millimeter-wave IRS-assisted MIMO systems based on compressed sensing techniques.

Many papers in the literature focus on channel estimation [17], [18], [19], [20], [21], [22], achievable rate maximization [24], [25], energy efficiency (EE) maximization [26], [27], [28], [29], [30], and interference mitigation problems [31], [32], [33]. The works [34], [35], [36], [37], [38] tackle the problem of reducing the channel training or the feedback overhead of IRS phase shifts to the IRS controller. In [34], the authors propose a framework to reduce the channel training overhead and a low-complexity beamforming design based on the discrete Fourier transform (DFT) codebook. The authors of [35] propose a low-complexity scheme for IRS reconstruction based on the single antenna users' mobility in a downlink system. Reference [36] proposes a protocol design to maximize the transmission rate in IRS-assisted MIMO-orthogonal frequency division multiplexing (OFDM) systems. In [37], the authors propose a channel overhead reduction based on an off-grid sparse Bayesian learning method in Terahertz channels, exploiting the angle-domain reciprocity. Recently, [38] proposed a framework for overhead-aware feedback and resource allocation in IRS-assisted MIMO systems. The main idea of [38] is to optimize the network resource such as the bandwidth and the total power used for transmission and feedback.

In this work, we propose an overhead-aware model for designing the IRS phase shifts. Our idea is to represent the optimized IRS phase shift vector using a low-rank tensor modeling approach. This is achieved by factorizing a ten-sorized version of the optimized IRS phase shift vector as Kronecker product combinations of a predefined number of factors of smaller sizes. With that, the number of phase shifts to be conveyed via the feedback control channel can be controlled in a flexible way by choosing the number of rank-one components, the number of factors per component, as well as their sizes and resolution. Two tensor decompositions are used for this purpose, namely the PARAllel FACTors (PARAFAC) [39] and Tucker [40] decompositions. Exploiting the known multi-linear structure of the chosen low-rank tensor model, the IRS controller can easily reconstruct the full phase shift vector from the smaller factors sent via a feedback control link with limited capacity. The main contributions of this paper are the following:

1) The proposed IRS phase shift feedback-aware framework controls the total number of phase shifts to be conveyed to the IRS controller in order to reconfigure the IRS. Such control is achieved by varying the factorization parameters of the proposed tensor-based models, namely PARAFAC-IRS and Tucker-IRS, which are the number of factors and their respective sizes, the number of components, and their respective resolution. The proposed method provides the network controller with degrees of freedom and flexibility to meet the desired trade-offs between feedback overhead and spectral efficiency (SE) performance.

2) Our proposed IRS feedback-aware method enables to save network resources by reducing the IRS phase shift feedback overhead. Specifically, for a fixed feedback load, the proposed method enables more frequent IRS phase shift feedback, which significantly improves the end-to-end latency, especially in fast varying channels, high mobility scenarios, and/or IRS panels with a moderate/large number of elements. Also, thanks to the significant reduction in the feedback overhead, the IRS-assisted network can decide to multiplex phase shifts associated with a higher number of users in the same feedback channel.

3) The proposed method provides a flexible feedback design by controlling the parameters of the low-rank factorization model, such as the number of components, the number and the size of the factors, as well as their respective resolution. This is an important feature of our proposed feedback-aware IRS model, since, for limited feedback control links, the low-rank model and its factorization parameters can be efficiently adjusted to the available capacity of the feedback link, providing more degrees of freedom to system design.

4) Our approach is analytical and provides a systematic way of controlling the feedback overhead by adjusting the parameters of the low-rank IRS model, namely, its rank and the corresponding number of factors of each rank-one component. For limited feedback control links, the proposed low-rank modeling approach provides an efficient “compression” mechanism for the IRS phase shifts to meet the available capacity of the limited feedback control link, providing more degrees of freedom to system design.

Different from [36] and [38], we aim to control the IRS phase shift feedback overhead by conveying only the factors of the chosen low-rank model to the IRS controller instead of conveying the full IRS phase shift vector. We show that the proposed low-rank models drastically reduce the required feedback requirements associated with the BS-IRS control links. In particular, our simulation results indicate that the proposed method is especially attractive in scenarios with a strong line of sight component, in which case nearly the same spectral efficiency is reached as in the cases with near-optimal phase shifts, but with significantly lower feedback overhead.

In [1], we presented the simplest case where only one component in the Kronecker factorization is considered. This paper generalizes [1] to low-rank IRS models while providing a unified view of the problem, by formulating two tensor-based IRS models, namely, PARAFAC-IRS and Tucker-IRS. Moreover, we show that the number of rank-one components controls the trade-off between the SE performance and the feedback overhead. We evaluate our models based on their impacts on the total system SE and EE by taking into account the channel estimation and the IRS phase shifts feedback duration.

The rest of the paper is organized as follows. Section II provides an introduction of the tensor notation and decompositions that are exploited in this paper. The system model is described in Section III. Section IV details our proposed feedback overhead-aware method and provides the details of the PARAFAC-IRS and Tucker-IRS models for IRS phase shift vector factorization. Section V describes the phase shift and weighting factors quantization procedure, the feedback
duration of the proposed models, and the reconstruction of the IRS phase shift vector at the IRS controller. The effects of the factorization parameters and the quantization process are also discussed in this section. Simulation results are provided in Section VI and the final conclusions and perspectives are discussed in Section VII.

A. Notation and Properties

Scalars are represented as non-bold lower-case letters $a$, column vectors as lower-case boldface letters $a$, matrices as upper-case boldface letters $A$, and tensors as calligraphic upper-case letters $\mathcal{A}$. The superscripts $\{\cdot\}^T$, $\{\cdot\}^*$, $\{\cdot\}^H$ and $\{\cdot\}^+$ stand for transpose, conjugate, conjugate transpose, and pseudo-inverse operations, respectively. The operator $\|\cdot\|_F$ denotes the Frobenius norm of a matrix or tensor, and $\mathbb{E}\{\cdot\}$ is the expectation operator. The operator diag$(\mathbf{a})$ converts $\mathbf{a}$ into a diagonal matrix, while diag$(A)$ returns a vector whose elements are the main the diagonal of $A$. Moreover, vec$(A)$ converts $A \in \mathbb{C}^{I \times L}$ to a column vector $\mathbf{a} \in \mathbb{C}^{I \times L \times 1}$ by stacking its columns on top of each other. The symbols $\otimes$, $\otimes$, and $\odot$ denote the outer product, the Kronecker product, and the Khatri-Rao product (also known as the column-wise Kronecker product), respectively. Let us consider a vector $\mathbf{a} = a^{(p)} \otimes \cdots \otimes a^{(1)}$ with $L = I_1 \cdots I_p$, where $a^{(p)} \in \mathbb{C}^{I_p \times 1}$ is the $p$-th factor vector, $p = 1, \ldots, P$. We can define a $P$-th order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_P}$ as

$$\mathcal{A} = \mathcal{T}\{\mathbf{a}\} = a^{(1)} \cdots a^{(P)}, \in \mathbb{C}^{I_1 \times \cdots \times I_P},$$

where $\mathcal{T}\{\cdot\}$ is the tensorization operator that maps the $l$-th element of $\mathbf{a}$ as $a_l = A_{i_1 \cdots i_l}$, with $l = i_1 + (i_2 - 1)I_1 + \cdots + (i_p - 1)I_{p-1} \cdots I_2$. Inversely, we have

$$\text{vec}(\mathcal{A}) = \mathbf{a} = a^{(P)} \otimes \cdots \otimes a^{(1)} \in \mathbb{C}^{I_1 \times \cdots \times I_P}.$$

II. Tensor Prerequisites

In this section, tensor preliminaries are provided by focusing on the main notation, operations, and properties that will be useful in the rest of the paper.

Consider a set of matrices $\{\mathbf{Y}_{i_3}\} \in \mathbb{C}^{I_1 \times I_2}$, for $i_3 = 1, \ldots, I_3$. Concatenating all $I_3$ matrices, we form the third-order tensor $\mathbf{Y} = [\mathbf{Y}_{1}, \mathbf{Y}_{2}, \ldots, \mathbf{Y}_{I_3}] \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $\mathbf{L}_{I_3}$ indicates a concatenation along the third dimension. We can interpret $\mathbf{Y}_{i_3}$ as the $i_3$-th frontal slice of $\mathbf{Y}$, defined as the matrix $\mathbf{Y}_{i_3} = \mathbf{Y}_{i_3} \in \mathbb{C}^{I_1 \times I_2}$. This matrix is built by varying the first and second dimensions for a fixed third dimension index $i_3$. The tensor $\mathbf{Y}$ can be matricized by letting one dimension vary along the rows and the remaining two dimensions along the columns. From $\mathbf{Y}$, we can form three different matrices, referred to as the $n$-mode unfoldings (for $n = \{1, 2, 3\}$ in this case), which are respectively given by

$$[\mathbf{Y}]_{(1)} = \mathbf{Y}_{(1)} \in \mathbb{C}^{I_1 \times I_2 \times I_3},$$

$$[\mathbf{Y}]_{(2)} = \mathbf{Y}_{(2)} \in \mathbb{C}^{I_1 \times I_2 \times I_3},$$

$$[\mathbf{Y}]_{(3)} = \text{vec}(\mathbf{Y}_{(1)}), \ldots, \text{vec}(\mathbf{Y}_{(I_3)}) \in \mathbb{C}^{I_1 \times I_2 \times I_3}.$$

A. PARAFAC Decomposition

The PARAFAC decomposition of a $P$-th order tensor $\mathbf{Y} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_P}$ expresses the tensor as a sum of $R$ rank-one tensor components, and can be defined as [39]

$$\mathbf{Y} = \sum_{r=1}^{R} \lambda_r (a^{(1)}_r \odot a^{(2)}_r \cdots \odot a^{(P)}_r) \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_P},$$

where $a^{(p)}_r \in \mathbb{C}^{I_p \times 1}$ is the $r$-th column of the $p$-th factor matrix $A^{(p)}_r \in \mathbb{C}^{I_p \times R}$, $p = 1, \ldots, P$, herein assumed to have unit norm, while the scalar $\lambda_r$ is the weighting factor associated with the $r$-th rank-one component, and can be interpreted as a factor that reveals the importance of each rank-one component. The tensor $\mathbf{Y}$ can be reshaped into different matrix forms, referred to as “matrix unfoldings”. The $p$-th mode matrix unfolding, defined as $[\mathbf{Y}]_{(p)} \in \mathbb{C}^{I_p \times I_{1-p}I_{p+1}, \ldots, I_P}$, can be written as

$$[\mathbf{Y}]_{(p)} = A^{(p)} \text{diag}(\lambda)(A^{(p+1)} \odot A^{(p+1)} \odot \cdots A^{(1)})^T,$$

where $\lambda = [\lambda_1, \ldots, \lambda_R] \in \mathbb{R}^{1 \times R}$ is the vector collecting the $R$ component weights. Fig. 1 illustrates a PARAFAC decomposition for a third-order tensor $\mathbf{Y}$ ($P = 3$) as the summation of rank-one tensors. In this case, it can be shown that the three matrix unfoldings of $\mathbf{Y}$ admit the following factorizations [41]:

$$[\mathbf{Y}]_{(1)} = A^{(1)} \text{diag}(\lambda)(A^{(2)} \odot A^{(3)})^T \in \mathbb{C}^{I_1 \times I_2 \times I_3},$$

$$[\mathbf{Y}]_{(2)} = A^{(2)} \text{diag}(\lambda)(A^{(3)} \odot A^{(1)})^T \in \mathbb{C}^{I_2 \times I_1 \times I_3},$$

$$[\mathbf{Y}]_{(3)} = A^{(3)} \text{diag}(\lambda)(A^{(2)} \odot A^{(1)})^T \in \mathbb{C}^{I_3 \times I_1 \times I_2}.$$

B. Tucker Decomposition

The Tucker decomposition expresses a tensor as a multi-linear rank combination of factors. For a $P$-th order tensor $\mathbf{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_P}$, it can be written as [41]

$$\mathbf{Y} = G \times_1 B^{(1)} \times_2 \cdots \times_p B^{(P)} \in \mathbb{C}^{I_1 \times \cdots \times I_P},$$

where $G$ and $B^{(p)}$ are the factor matrices and the core tensor, respectively. The Tucker decomposition is a generalization of the PARAFAC decomposition, as each component of $\mathbf{Y}$ can be expressed as a rank-one tensor. This approach can be used to reduce the computational complexity of tensor factorizations, especially when dealing with high-dimensional tensors. The Tucker decomposition is widely used in various applications, such as signal processing, data analysis, and machine learning. It allows for a compact representation of the data, which can be further exploited to perform various operations, such as compression, denoising, and feature extraction.
where $B^{(p)} \in \mathbb{C}^{I_x \times R_p}$ is the $p$-th factor matrix, $p = 1, \ldots, P$, and $\mathcal{G} \in \mathbb{C}^{R_1 \times \cdots \times R_P}$ is the core tensor. The tensor $\mathcal{Y}$ can also be represented in terms of linear combinations of outer product terms as
\[
\mathcal{Y} = \sum_{r_1=1}^{R_1} \cdots \sum_{r_P=1}^{R_P} g_{r_1,\ldots,r_P} \left( b_{r_1}^{(1)} \otimes \cdots \otimes b_{r_P}^{(P)} \right),
\]
where $b_{r_p}^{(p)} \in \mathbb{C}^{I_x \times 1}$ is the $r_p$-th column of the $p$-th factor matrix $B^{(p)} \in \mathbb{C}^{I_x \times R_p}$, $p = 1, \ldots, P$, $r_p = 1, \ldots, R_p$. The Tucker decomposition has a more complex structure than the PARAFAC decomposition since it allows full “interactions” involving all the columns of the associated factor matrices $B^{(1)}, \ldots, B^{(P)}$. In total, the tensor $\mathcal{Y}$ is generated by $R_1R_2 \cdots R_P$ outer product terms as shown in (12), which are combined via the core tensor, the $(r_1, \ldots, r_P)$-th element of which defines the importance of each term. Thus, for each $p$, the $p$-th mode unfolding matrix of $\mathcal{Y}$, defined as $[\mathcal{Y}]_{(p)} \in \mathbb{C}^{I_x \times I_1 \cdots I_{p-1} I_{p+1} \cdots I_P}$, can be written as
\[
[\mathcal{Y}]_{(p)} = B^{(p)} [g]_{(p)} \left( b_{1}^{(p)} \otimes \cdots \otimes b_{r_p}^{(p+1)} \otimes \cdots \otimes b_{1}^{(1)} \right)^T.
\]

For $P = 3$, Figure 2 illustrates the Tucker decomposition. Its matrix unfoldings are given by
\[
[\mathcal{Y}]_{(1)} = B^{(1)} [g]_{(1)} \left( b_{1}^{(3)} \otimes b_{1}^{(2)} \right)^T \in \mathbb{C}^{I_1 \times I_2 I_3},
\]
\[
[\mathcal{Y}]_{(2)} = B^{(2)} [g]_{(2)} \left( b_{1}^{(3)} \otimes b_{1}^{(1)} \right)^T \in \mathbb{C}^{I_2 \times I_1 I_3},
\]
\[
[\mathcal{Y}]_{(3)} = B^{(3)} [g]_{(3)} \left( b_{1}^{(2)} \otimes b_{1}^{(1)} \right)^T \in \mathbb{C}^{I_3 \times I_1 I_2}.
\]

### III. System Model

In this section, we formulate the proposed IRS feedback overhead control framework. Without loss of generality, we assume that the receiver (RX) is the network entity responsible for communicating with the IRS controller (e.g. the base station) and the transmitter (TX) is the mobile user. Note that the proposed framework also applies to the inverse direction. Let us assume the system model of [38], where a MIMO single stream IRS-assisted communication is considered, and the direct TX-RX is assumed to be blocked. The TX and the RX nodes are equipped with $M_T$ and $M_R$ antennas, respectively. The IRS contains $N$ reflective elements, where the $n$-th IRS element is given by $s_n = \beta_n e^{j\theta_n}$, with $\beta_n$ and $\theta_n$ being the reflection amplitude and the phase shift of the $n$-th IRS element, respectively, in which, we assume $\beta_n = 1 \ \forall n \in N$ for simplicity. The received signal is given by
\[
y = w^H G \text{diag} (s) H q + w^H b,
\]
where $b \in \mathbb{C}^{M_R \times 1}$ is the additive noise at the receiver with $\mathbb{E} [bb^H] = \sigma_n^2 I_{M_R}$, $w \in \mathbb{C}^{M_R \times 1}$ and $q \in \mathbb{C}^{M_T \times 1}$ are the receiver and transmitter combiner and precoder (active beamformers), respectively. $H \in \mathbb{C}^{N \times M_T}$ and $G \in \mathbb{C}^{M_R \times N}$ are the TX-IRS and IRS-RX involved channels, and $s = [e^{j\theta_1}, \ldots, e^{j\theta_N}] \in \mathbb{C}^{N \times 1}$ is the IRS phase shift vector (passive beamformer), where $\theta_n$ is the phase shift applied to the $n$-th IRS element.

In this paper, we focus on the IRS phase shifts feedback overhead control problem. Thus, we consider that the steps 1 and 2 illustrated in Fig. 3 are based on state-of-the-art methods. More specifically, as in [38], we assume that the RX has the full knowledge of the channel matrices $G \in \mathbb{C}^{M_R \times N}$ and $H^H \in \mathbb{C}^{N \times M_T}$. Note that the estimates of $G$ and $H$ can be obtained, for instance, from the channel estimation schemes of [17] or [18]. For the beamformer optimization step, we resort to the upper bound algorithm of [38], which provides a joint optimization of the active beamforming vectors (precoder and combiner) $q \in \mathbb{C}^{M_T \times 1}$ and $w \in \mathbb{C}^{M_R \times 1}$, as well as the IRS phase shift vector $s \in \mathbb{C}^{N \times 1}$. Consider the singular value decompositions (SVDs) of $H \in \mathbb{C}^{N \times M_T}$ and $G \in \mathbb{C}^{M_R \times N}$, with rank($H$) = $I$, and rank($G$) = $J$, as $H = \sum_{i=1}^{I} \sigma_i^I u_i^I v_i^H = U^I \Sigma^I V^H$ and $G = \sum_{j=1}^{J} \sigma_j^J u_j^J v_j^H = U^J \Sigma^J V^H$. From the natural ordering of the SVD, we have $\sigma_1^I \geq \sigma_2^I \geq \ldots \geq \sigma_I^I$, and $\sigma_1^J \geq \sigma_2^J \geq \ldots \geq \sigma_J^J$. The upper bound algorithm of [38] finds $w$, $q$, and $s$ as
\[
q = u_1^I \in \mathbb{C}^{M_T \times 1}, \quad w = u_1^J \in \mathbb{C}^{M_R \times 1}, \quad \theta_n = -\angle (u_{n}^J) \in \mathbb{C}^{N \times 1}, \quad n \in N.
\]
\[
s = [e^{j\theta_1}, \ldots, e^{j\theta_N}] \in \mathbb{C}^{N \times 1}.
\]
Note that, (18) and (19) are feasible designs for the active (precoder and combiner) and the passive (IRS phase shifts) beamformers since, in a MIMO single stream scenario, the beamformers must align with the strongest eigenmodes of the involved channels $G$ and $H$. In this work, we adopt the above solution as a starting point for our proposed approach, although different beamforming optimization algorithms can also be used, such as, e.g., [2] and [42]. It is important to mention that the proposed method consists of controlling the...
IRS phase shift feedback overhead by performing a low-rank approximation to a tensorized version of the optimized IRS phase shift vector in (19), consisting of a sum of Kronecker products of smaller factors. Thus, the proposed approach is independent of the chosen channel estimation method and beamforming optimization algorithm.

A. Feedback Overhead Model

After the channel estimation and the beamforming optimization steps, the RX needs to feed the optimized phase shifts (i.e., a quantized version of $s$ in (19)) back to the IRS controller via a control channel, which then tunes the phase shifts of the IRS panel. Considering that this control channel has a limited capacity and that the IRS may contain several shifts of the IRS panel. This control channel shifts (i.e., a quantized version of $s$)

$$E = \frac{T - T_E - T_F}{T} \mu_p + \frac{\mu_p T_F}{T} + P_e,$$

where $N$ is the total number of IRS phase shifts to be fed back, $B_F$, $p_F$ are the feedback bandwidth and power, respectively, $g_F$ is the scalar control channel used, $b_F$ is the resolution of each phase shift, and $N_0$ is the noise power density. The authors of [38] focus on the problem of rate and EE maximization. The rate is given by

$$SE = (1 - \frac{T_E + T_F}{T}) \log \left( 1 + \frac{p_T |s_N|^2 GSH g^2}{N_0} \right),$$

with $T_E$ and $T$ being the duration of the channel estimation phase and the total time interval, and $B$ being the transmission bandwidth. The EE is given by $EE = Rate / P_{tot}$, and the total power consumption $P_{tot}$ can be expressed as

$$P_{tot} = P_e + \frac{T - T_E - T_F}{T} \mu_p + \frac{\mu_p T_F}{T} + P_c,$$

where $P_e$ is the power used for the channel estimation phase, $1/\mu$ is the efficiency of the transmit power amplifier, $p_F$ is the power used during $T_F$ seconds, and $\mu_p$ is the efficiency of the transmit amplifier used for feedback. The values of $p_T$, $p_F$, $B$, $B_F$ are chosen to maximize (21) and (22) [38]. In this paper, we aim to control the feedback duration $T_F$ of (20) by imposing low-rank modeling to a tensorized version of the optimized IRS phase shift vector (19). More specifically, we propose to factorize the phase shift vector as a combination of Kronecker products of smaller phase shift factors. To this end, we develop two different tensor-based models, namely PARAFAC-IRS and Tucker-IRS, which are detailed in the next sections.

IV. LOW-RANK TENSOR-BASED IRS MODELING

In this section, we describe the proposed tensor low-rank models for feedback-aware IRS phase shift representation, allowing us to control the feedback duration $T_F$ given in (20).

For convenience, let us consider the initial idea presented in [1], which consists of factorizing $s$ (optimized in (19)) as the Kronecker product of $P$ factors, i.e.,

$$s \approx s^{(P)} \otimes \cdots \otimes s^{(1)} \in \mathbb{C}^{N_1 \times \ldots \times N_P},$$

where $s^{(p)} \in \mathbb{C}^{N_p \times 1}$ and $N = \prod_{p=1}^{P} N_p$. The equality in (23) only occurs when $(K_H, K_C) \rightarrow \infty$, since, in this case, the channels are characterized by their respective line-of-sight (LOS) components only. To provide an insight into the associated channel decomposition structure, note that a Vandermonde vector of size $N$ can be factorized exactly as the Kronecker product of $P$ smaller vectors, such that $N = \prod_{p=1}^{P} N_p$, with $N_p$ being the size of the $p$-th factor [43]. This result allows to express the channel matrices in the LOS case in terms of Kronecker products of Vandermonde vectors, as shown in (48)-(49) of Appendix. Therein, we demonstrate that in the case of pure LOS channels, the equality in (19) is achieved. On other hand, in the presence of non-line-of-sight (NLOS) components, the Kronecker factorization in (23) is an approximation.

In order to estimate the factors in (23), first we tensorize $s$ as a $P$-th order tensor $\mathbf{T} \in \mathbb{C}^{N_1 \times \ldots \times N_P}$. In other words, applying property (1), we have

$$\mathbf{S} = \mathbf{T} \{s\} \approx s^{(1)} \circ \cdots \circ s^{(P)},$$

**Example:** To shed some light on the impact of the Kronecker-RS model (23) on the feedback overhead, let us consider a simple scenario with $N = 1024$ phase shifts, where $P = 3$ is chosen. Consider, as one example, the following factors $s^{(1)} = [e^{j\theta_1}, \ldots, e^{j\theta_{1024}}]^{\top} \in \mathbb{C}^{1024 \times 1}$, $s^{(2)} = [e^{j\theta_{1025}}, \ldots, e^{j\theta_{1036}}]^{\top} \in \mathbb{C}^{10 \times 1}$ and $s^{(3)} = [e^{j\theta_{1037}}, \ldots, e^{j\theta_{1039}}]^{\top} \in \mathbb{C}^{4 \times 1}$, i.e., $N_1 = 32$, $N_2 = 8$ and $N_3 = 4$. Note that, $N_1$, $N_2$, $N_3$ can assume different values as long $N_1 N_2 N_3 = N = 1024$. In this scenario, instead of conveying to the IRS controller 1024 phase shifts, we only need to convey the phase shifts of the smaller factors, which results in $32 + 8 + 4 = 44$ phase shifts. This drastically reduces the total phase shift feedback overhead. Note that, physically, the Kronecker product in (23) translates into the summation of the phase shifts of the smaller factors. It is clear that in a general case for a large $N$, and based on the choice of $P$, we have

$$\sum_{p=1}^{P} N_p \ll N = \prod_{p=1}^{P} N_p.$$
In the following, we generalize the model (23) to the arbitrary rank-$R$ case, by resorting to low-rank tensor decompositions, which are referred to as the PARAFAC-IRS and Tucker-IRS models.

A. PARAFAC-IRS Model

In (24), the IRS phase shift vector is recast as a rank-one tensor. In this section, we generalize this model to the rank-$R$ case. More specifically, the optimum phase shift vector $s$ of (19) is factorized according to a PARAFAC model, i.e.,

$$ s \approx \sum_{r=1}^{R} \lambda_r \left( s_r^{(1)} \otimes \cdots \otimes s_r^{(P)} \right) \in \mathbb{C}^{N_1 \times \cdots \times N_P}, \quad (25) $$

which, using (1) and (2), implies

$$ S \approx \sum_{r=1}^{R} \lambda_r \left( s_r^{(1)} \odot \cdots \odot s_r^{(P)} \right) \in \mathbb{C}^{N_1 \times \cdots \times N_P}, \quad (26) $$

where $S = \mathcal{F} \{ s \}$, $R$ is the number of components (or tensor rank), $\lambda_r$ is the weighting factor of the $r$-th component which given by, and $s_r^{(p)} \in \mathbb{C}^{N_p \times 1}$ is the $r$-th column of the $p$-th factor matrix $S^{(p)} = \left[ s_1^{(p)}, \ldots, s_R^{(p)} \right] \in \mathbb{C}^{N_p \times R}$, with $\| s_r^{(p)} \|_2 = 1$, $p = 1, \ldots, P$, $r = 1, \ldots, R$.

The main idea of assuming $R > 1$ components is to mitigate the approximation error for NLOS channels, where the Vandermonde-based separable structures given in (48)-(49) do not hold. Ideally, when considering continuous phase shifts, the approximation error will decrease as the number of components $R$ of the low-rank model increases. However, in practice, the phase shifts have to be quantized and the quantization errors also affect the representation accuracy. Note that increasing $R$ implies a higher feedback overhead since more phase shift values have to be conveyed. The choice of the number of components $R$, the number of factors $P$, and the size of each factor $N_p$, $p = 1, \ldots, P$, is discussed in details in Sections V and VI.

The factor components are estimated by solving the following problem

$$ \left[ s_r^{(1)}, \ldots, s_r^{(P)} \right]_{r=1,\ldots,R} = \arg \min_{s_r^{(1)}, \ldots, s_r^{(P)}} \left\| S - \sum_{r=1}^{R} \lambda_r \left( s_r^{(1)} \odot \cdots \odot s_r^{(P)} \right) \right\|_F ^{\delta} $$

s.t. $\left\| s_r^{(p)} \right\|_2 = 1$, $p = 1, \ldots, P$, $r = 1, \ldots, R$, \quad (27)

where $s_r^{(p)} \in \mathbb{C}^{N_p \times 1}$ is the unitary norm $p$-th factor component, and $\lambda = [\lambda_1, \ldots, \lambda_R]$ is the weighting vector that contains the weights associated with each rank-one component. Note that Problem (27) is non-convex. To handle it, we resort to classical tensor-based estimation approaches, such as the alternating least squares (ALS) [41] (for the PARAFAC-IRS model) and the high order singular value decomposition (HOSVD) [44] (for the Tucker-IRS model). These algorithms exploit the multi-linearity of tensor models to estimate their associated factors. The ALS algorithm is an iterative solution that has monotonic convergence and always converges to a local or global minimum [45]. Its number of iterations depends on the initialization. On the other hand, the HOSVD provides a closed-form solution to compute a low multilinear rank approximation of the IRS phase shift factors.

Let us define $S^{(p)} = \left[ s_1^{(p)}, \ldots, s_R^{(p)} \right] \in \mathbb{C}^{N_p \times R}$ as the $p$-th factor matrix, $p = 1, \ldots, P$. From (7), the $p$-mode unfolding of $S$, defined as $[S]_{(p)} \in \mathbb{C}^{N_p \times N_1 \times \cdots \times N_{p-1} \times N_{p+1} \times \cdots \times N_P}$, is given as

$$ [S]_{(p)} \approx S^{(p)} \operatorname{diag}(\lambda) \left( S^{(p)} \odot \cdots \odot S^{(p+1)} \odot \cdots \odot S^{(1)} \right)^T. \quad (28) $$

Problem (27) can be solved by means of the ALS algorithm [41], described in Algorithm 1. Basically, each iteration of the ALS algorithm has $P$ LS estimation steps. The $p$-th LS problem is defined as

$$ \hat{S}^{(p)} = \arg \min_{S^{(p)}} \left\| [S]_{(p)} - S^{(p)} \left( S^{(p)} \odot \cdots \odot S^{(p+1)} \odot \cdots \odot S^{(1)} \right)^T \right\|_F ^{2}, \quad (29) $$

the solution of which is given by

$$ \hat{S}^{(p)} = [S]_{(p)} \left( [S]_{(p)} \odot \cdots \odot S^{(p+1)} \odot \cdots \odot S^{(1)} \right)^{+}. \quad (30) $$

At each iteration, an LS estimate of $p$-th factor matrix $\hat{S}^{(p)}$ is found from the estimates of the remaining $P - 1$ factors obtained in the previous estimation steps. Then, we compute the weighting vector $\hat{\lambda} = [\hat{\lambda}_1^{(p)}, \ldots, \hat{\lambda}_R^{(p)}]$ that stores the $2$-norms of the $R$ columns of $\hat{S}^{(p)}$, and a normalized version

---

**Algorithm 1 PARAFAC-IRS ALS**

1. **Inputs:** Phase shift tensor $S \in \mathbb{C}^{N_1 \times \cdots \times N_P}$, number $R$ of components, and maximum number $I$ of iterations.
2. Randomly initialize the factors $\hat{S}_0^{(1)}, \ldots, \hat{S}_0^{(P)}$. Iteration $i = 0$.
3. for $i = 1 : I$ do
   4. for $p = 1 : P$ do
      5. Compute an estimate of the $p$-th factor $\hat{S}_i^{(p)}$ as
         $$ \hat{S}_i^{(p)} = [S]_{(p)} \left( \hat{S}_{i-1}^{(p)} \odot \cdots \odot \hat{S}_{i-1}^{(p+1)} \cdots \odot \hat{S}_{i-1}^{(1)} \right)^{+}. $$
      6. Compute the $p$-th weighting vector $\hat{\lambda}_i^{(p)} = \left[ \| \hat{S}_{i-1}^{(p)(1)} \|_2, \ldots, \| \hat{S}_{i-1}^{(p)(R)} \|_2 \right]$ and update
         $$ \hat{S}_i^{(p)} \leftarrow \hat{S}_i^{(p)} \left[ \text{diag}(\hat{\lambda}_i^{(p)}) \right]^{-1}. $$
3. end for
4. end for
5. Compute the global weighting vector $\hat{\lambda} = [\hat{\lambda}_1^{(1)}, \ldots, \hat{\lambda}_R^{(P)}]$. Return $\hat{S}^{(1)}, \ldots, \hat{S}^{(P)}$ and $\hat{\lambda}$. 

of \( \hat{S}^{(p)} \) is obtained by dividing its \( r \)-th column by \( \hat{\lambda}^{(r)} \), \( r = 1, \ldots, R \). This process is repeated for the subsequent iterations until convergence (or the maximum number of iterations) is achieved. Then, a global weighting vector is computed as \( \hat{\lambda} = \hat{\lambda}^{(1)} \odot \cdots \odot \hat{\lambda}^{(P)} \).

Let us define the normalized mean square error (NMSE) at the \( i \)-th iteration as

\[
e_{(i)} = \frac{\left\| [S]_{(1),i} - [\hat{S}]_{(1),i} \right\|_F^2}{\left\| [S]_{(1),i} \right\|_F^2},
\]

where \([S]_{(1),i}\) is the reconstructed 1-mode unfolding at the \( i \)-th ALS iteration, given by

\[
[S]_{(1),i} = S^{(1)} \text{diag}(\hat{\lambda})(\hat{S}^{(P)} \cdots \hat{S}^{(2)})^\top.
\]

If \( |e_{(i)} - e_{(i-1)}| \leq \epsilon \), where \( \epsilon \) is a pre-defined threshold, convergence is declared and the algorithm stops [41]. In this paper, we consider \( \epsilon = 10^{-4} \). It is worth mentioning that enhanced versions of the ALS algorithm exist in the literature. Efficient and dedicated solutions to estimate the factors of the PARAFAC model can be found in [46], which offers an enhanced convergence speed and involves fast computations.

As one example, Figure 5 illustrates the ratio between the state-of-the-art solution [38], where the IRS phase shifts are fed back, and the proposed PARAFAC-IRS solution. This feedback ratio is given by \( N/(R P \sum_{p=1}^P N_p) \). Let us define a vector \( N_P = [N_1, \ldots, N_P] \in \mathbb{R}^{1 \times P} \) that contains the sizes of the factors for a certain \( P \). We can observe that, for \( R = 1 \), the feedback duration with \( P = 2 \), \( N_1 = 256 \), \( N_2 = 4 \), is almost five times smaller than the state-of-the-art solution. In addition, by increasing the number \( P \) of factors, the size of the factors is reduced, thus decreasing the feedback duration. For instance, by setting \( P = 10 \) and \( N_p = 2 \), \( p = 1, \ldots, P \), the feedback overhead of the PARAFAC-IRS model is approximately 50 times smaller than that of the state-of-the-art [38]. As we discuss in Section VI, for LOS-dominant scenarios, such a feedback reduction comes with minimal or no performance loss, while for NLOS scenarios, a tradeoff involving feedback duration and SE exists. In any case, by properly choosing the parameters of the PARAFAC-IRS model, the desired tradeoff can be achieved. Moreover, for limited feedback links, one can adjust these parameters to meet the capacity of the control channel.

### B. Tucker-IRS Low-Rank Model

In this section, we discuss a second approach to modeling the IRS phase shift vector of (19). It consists of using the Tucker model, i.e.,

\[
s \approx \sum_{r_1=1}^{R_1} \cdots \sum_{r_P=1}^{R_P} \mathcal{G}_{r_1,\ldots,r_P} (s_{r_1}^{(1)} \odot \cdots \odot s_{r_P}^{(1)}) \in \mathbb{C}^{N_1 \times \cdots \times N_P \times 1},
\]

which, using (1) and (2), implies on \( P \)-th order Tucker decomposition

\[
S \approx \sum_{r_1=1}^{R_1} \cdots \sum_{r_P=1}^{R_P} \mathcal{G}_{r_1,\ldots,r_P} (s_{r_1}^{(1)} \circ \cdots \circ s_{r_P}^{(P)}) \in \mathbb{C}^{N_1 \times \cdots \times N_P},
\]

where \( \mathcal{G} \in \mathbb{C}^{R_1 \times \cdots \times R_P} \) is the \( P \)-th order core tensor and \( s_{r_p}^{(p)} \in \mathbb{C}^{N_p \times 1} \) is the \( r_p \)-th column of the \( p \)-th factor matrix \( S^{(p)} \in \mathbb{C}^{N_p \times R_p}, \) \( p = 1, \ldots, P, \) \( r_p = 1, \ldots, R_p \).

We assume that \( s_{r_p}^{(p)} = \sigma_{r_p}^{(p)} u_{r_p}^{(p)} \) where \( u_{r_p}^{(p)} \) has unit norm and \( \sigma_{r_p}^{(p)} \) corresponds to the 2-norm of \( s_{r_p}^{(p)} \). Compared to the PARAFAC-IRS model, the Tucker-IRS model provides additional degrees of freedom to represent the IRS phase shift tensor \( S \) as a linear combination of low multi-linear rank factors. More specifically, comparing (26) with (33), in the \( P \)-th order Tucker-IRS model, each factor can have a different number of components which are combined by means of a core tensor \( \mathcal{G} \). The added degrees of freedom provided by the Tucker-IRS model results may improve the SE performance compared to the PARAFAC-IRS model in NLOS scenarios at the cost of a higher feedback overhead, as it will be discussed in Section VI.

From (13), the \( p \)-th mode unfolding of \( S \) is given by

\[
[S]_{(p)} \approx S^{(p)} \mathcal{G} \left( S^{(p)} \odot \cdots \odot S^{(p+1)} \odot S^{(p-1)} \odot \cdots \odot S^{(1)} \right)^\top
\]

The factor matrix \( S^{(p)} \in \mathbb{C}^{N_p \times R_p}, \) \( p = 1, \ldots, P \), as well as the core tensor \( \mathcal{G} \) can be found by means of the HOSVD [47], which is described in Algorithm 2. The algorithm estimates the factors matrices by computing SVDs of the \( P \) matrix unfoldings of \( S \). From the rank-\( R_p \) SVD of \( [S]_{(p)} = U^{(p)} \text{diag}(\sigma^{(p)}) V^{\mathsf{H}} = S^{(p)} V^{\mathsf{H}} \), we have

\[
\hat{S}^{(p)} = [s_{1}^{(p)}, \ldots, s_{R_p}^{(p)}] = [\sigma_1^{(p)} u_1^{(p)}, \ldots, \sigma_{R_p}^{(p)} u_{R_p}^{(p)}] = U^{(p)} \text{diag}(\sigma^{(p)}) \in \mathbb{C}^{N_p \times R_p}.
\]
Algorithm 2 Tucker-IRS HOSVD

1: Inputs: Tensor $\mathbf{S}$, the number of components $R_p$, $p = 1, \ldots, P$.
2: for $p = 1 : P$ do
3: Compute the SVD of the $p$-mode matrix unfolding of $\mathbf{S}$ as
   $[\mathbf{S}]_p = \mathbf{U}^{(p)} \operatorname{diag}(\sigma^{(p)}) \mathbf{V}^{(p)H} \in \mathbb{C}^{N_p \times N_p \times \cdots \times N_p}$
4: Compute $\tilde{\mathbf{S}}^{(p)} = \mathbf{U}^{(p)} \operatorname{diag}(\tilde{\sigma}^{(p)}) \in \mathbb{C}^{N_p \times R_p}$
5: end for
6: Compute an estimate of the core tensor in vectorized form as
   $\tilde{\mathbf{g}} = \left( \tilde{\mathbf{S}}^{(P)} \otimes \cdots \otimes \tilde{\mathbf{S}}^{(1)} \right)^{\mathsf{H}} \operatorname{vec}(\mathbf{S}) \in \mathbb{C}^{R_1 \times \cdots \times R_P}$
7: Return $\{\tilde{\mathbf{S}}^{(p)}\}, \{\mathbf{u}^{(p)}\}, \{\sigma^{(p)}\}, p = 1, \ldots, P$, and $\mathcal{G} = T\{\tilde{\mathbf{g}}\} \in \mathbb{C}^{R_1 \times \cdots \times R_P}$

### Table I
**Computational Complexities of ALS and HOSVD Algorithms Associated With the PARAFAC-IRS and Tucker-IRS Factorization Models, Respectively**

| Algorithm                        | Computational complexity $\mathcal{O}(\cdot)$ |
|----------------------------------|-----------------------------------------------|
| Alternating Least Squares (ALS)  | $\mathcal{O}(\min\{\mathcal{O}(\text{NR}), \mathcal{O}(\text{VR})\})$ |
| High Order Singular Value Decomposition (HOSVD) | $\mathcal{O}(N \sum_{p=1}^{P} N_p)$ |

where $\mathbf{U}^{(p)} = [\mathbf{u}^{(p)}_1, \ldots, \mathbf{u}^{(p)}_{R_p}]$ is the truncated matrix of left singular vectors to its first $R_p$ columns and $\sigma^{(p)} \in \mathbb{C}^{1 \times R_p}$ holds the corresponding $R_p$ singular values, $p = 1, \ldots, P$. From the $P$ estimated factor matrices, the core tensor $\mathcal{G}$ is then estimated as

$$\tilde{\mathbf{g}} = \left( \tilde{\mathbf{S}}^{(P)} \otimes \cdots \otimes \tilde{\mathbf{S}}^{(1)} \right)^{\mathsf{H}} \operatorname{vec}(\mathbf{S}) \in \mathbb{C}^{R_1 \times \cdots \times R_P},$$

where $\tilde{\mathbf{g}} = \operatorname{vec}(\tilde{\mathbf{G}})$ and $\operatorname{vec}(\mathbf{S})$ are the vectorized versions of the core tensor and the IRS phase shift tensor, respectively.

### C. Computational Complexity

In this section, we discuss the computational complexity of Algorithms 1 and 2, shown in Table I. For the ALS algorithm, the main computational cost is related to the $P$ matrix inverses performed at each iteration. For the ALS algorithm, its main computational cost is represented by the $P$ matrix inversions that are performed at each iteration, which in its turn depends on the number $R$ of components. In our experiments, we have noticed that the total number of iterations for the ALS convergence is less than 10, for $R = 1, 2, 3, 4$ in strong LOS scenarios (see Figure 7). For the HOSVD algorithm, its main computation effort comes from the computation of $P$ independent truncated SVDs. It is important to mention that for a small number of components $R$ or $R_p$, this complexity is negligible compared to the complexity associated with channel estimation (see, e.g., [18] and the references therein).

## V. Quantization, Reconstruction, and Parameter Choices

### A. Phase Shift Quantization

Recall that Algorithms 1 or 2 provide phase shift factors with continuous values. Before conveying these factors via the feedback control channel, quantization is necessary. Let us define $\tilde{\mathbf{a}} = [\exp(i \theta^{(1)}_a), \ldots, \exp(i \theta^{(L)}_a)] \in \mathbb{C}^{L \times 1}$, as the exponential vector that contains the quantized angles of a vector $\mathbf{a} \in \mathbb{C}^{L \times 1}$ with $[\theta^{(1)}_a, \ldots, \theta^{(L)}_a] = Q\{\theta, b\}$ being the $b$-bit quantization operator applied to each entire of the vector $\mathbf{a} \in \mathbb{C}^{L \times 1}$. For the PARAFAC-IRS model, the quantized phase shift factors are given by

$$\tilde{\mathbf{g}}^{(p)}_i = Q\left\{\angle \mathbf{g}^{(p)}_i, \mathbf{b}^{(p)}_i\right\},$$

$p = 1, \ldots, P$, $r = 1, \ldots, R$. For the Tucker-IRS model, the quantized versions of the phase shift factors and the core tensor are respectively given by

$$\tilde{\mathbf{u}}^{(p)}_i = Q\left\{\angle \mathbf{u}^{(p)}_i, \mathbf{b}^{(p)}_i\right\}, \quad \mathcal{G}_{r_1, \ldots, r_P} = Q\left\{\angle \mathcal{G}_{r_1, \ldots, r_P}, \mathbf{b}^{(p)}_i\right\},$$

$r_p = 1, \ldots, R_p$, $p = 1, \ldots, P$. The following codebook is assumed for the $p$-th factor quantization

$$\mathcal{C}^{(p)}_{\bar{\phi}} = \left\{-\pi + \frac{2\pi}{2^{\mathbf{b}^{(p)}_i}}, -\pi + \frac{4\pi}{2^{\mathbf{b}^{(p)}_i}}, \ldots, \pi\right\}.$$

### B. Weighting Factor Quantization

For the PARAFAC-IRS model, the global weighting vector $\lambda$ obtained during the ALS algorithm is quantized as follows. A normalized version of the weighting vector is computed by normalization by its largest element $\lambda_{\text{max}}$, i.e., $\lambda / \lambda_{\text{max}}$. We then obtain a new $(R - 1)$-dimensional vector $\tilde{\lambda}$ holding the elements of the normalized $\lambda$ except the first one (which is equal to one). The quantized version of it is then obtained as $\tilde{\lambda} = Q\{\tilde{\lambda}, b^{(p)}_{\bar{\phi}}\}$. The full $p$-th quantized weighing vector is then given by $\lambda = [1, \tilde{\lambda}]$. In the end, the feedback of this vector costs $(R - 1)b^{(p)}_{\bar{\phi}}$ bits. For the Tucker-IRS model, weighting factor quantization follows in a similar way. Recall that the $P$ weighting vectors $\{\sigma^{(1)}, \ldots, \sigma^{(P)}\}$ hold, respectively, the singular values of the $P$ matrix unfoldings of the phase shift tensor $\mathbf{S}$, where the $p$-th weighting vector is defined as $\tilde{\sigma}^{(p)} = [\sigma^{(p)}_1, \ldots, \sigma^{(p)}_R]$. A normalized version of the weighting vector is computed by normalization by its first (largest) element $\sigma^{(p)}_1$, i.e., $\sigma^{(p)} / \sigma^{(p)}_1$. We then define $\tilde{\sigma}^{(p)}$ as an $(R - 1)$-dimensional vector holding the elements of the normalized $\sigma^{(p)}$ excluding the first one (which is equal to one). The quantized version of it is given by $\tilde{\sigma}^{(p)} = Q\{\tilde{\sigma}^{(p)}, b^{(p)}_{\bar{\phi}}\}$. Finally, the full $p$-th quantized vector is given by $\tilde{\mathbf{s}}^{(p)} = [1, \tilde{\sigma}^{(p)}]$. Hence, for the Tucker-IRS model, the feedback of the quantized versions of the $P$ weighting factors costs $b^{(p)}_{\bar{\phi}} \cdot \sum_{p=1}^{P} (R_p - 1)$ bits. For both PARAFAC-IRS and Tucker-IRS models, uniform quantization of the weighting.
vectors is adopted, and the quantization steps are defined by the following amplitude codebook

\[ C_w = \{0.01, 0.01 + l, 0.01 + 2 l, \ldots, 1\}, \]

where \( l = \frac{4 - 0.01}{2^{2p+1}} \) is the predefined step. For simplicity, the amplitudes in (38) are rounded to the second decimal point.

C. Feedback Duration

After the quantization step, the RX conveys to the IRS controller the phase shifts and the weighting factors. Based on the IRS phase shift feedback model in (20), for the proposed PARAFAC-IRS model, the feedback duration is given by

\[
T_F^{\text{(PARAFAC)}} = T_{PR} + R \sum_{p=1}^{P} N_p \cdot b_f^{(p)} + (R - 1) \cdot b_f^{(w)}
\]

where \( T_{PR} \) is the number of bits required for a preamble of the frame, in order to inform the IRS controller of the factorization parameters. The term \( R \sum_{p=1}^{P} N_p b_f^{(p)} \) is the cost, in bits, to convey the phase shifts of the factors, while \( (R - 1) \cdot b_f^{(w)} \) is term related to the cost of the weighting vector \( \hat{\lambda} \). In a similar process, the feedback duration associated with the Tucker-IRS model is given as

\[
T_F^{\text{(Tucker)}} = T_{PR} + \left( \sum_{p=1}^{P} R_p N_p b_f^{(p)} \right) + \sum_{p=1}^{P} R_p + b_f^{(w)} \sum_{p=1}^{P} (R_p - 1)
\]

where \( T_{PR} \) is the preamble duration that informs the IRS controller of the chosen low-rank IRS model, the number of factors \( P \), and the number of components \( R_p \), \( p = 1, \ldots, P \). As in the PARAFAC-IRS model, this preamble cost is negligible compared to the cost related to the feedback of the phase shifts. The term \( \sum_{p=1}^{P} R_p N_p b_f^{(p)} \) represents the cost, in bits, associated with the feedback of the phase shifts, while \( \sum_{p=1}^{P} R_p \) is the feedback cost of the core tensor, and \( b_f^{(w)} \sum_{p=1}^{P} (R_p - 1) \) relates to the feedback cost of the weighting factors.

D. Phase Shift Vector Reconstruction

Upon reception of the quantized parameters of the chosen low-rank IRS model (i.e., factors and weighting vectors), the IRS controller should reconstruct the full IRS phase shift vector. Note that such a reconstruction yields a suboptimal version of the IRS phase shift model, since the quantization steps described in (36) and (37), for the PARAFAC-IRS and Tucker-IRS models, respectively, imply information losses as only the phases of the factors are kept for transmission in the feedback channel. The vector of phase shifts of the IRS is given by \( s^* = e^{j\hat{\mathbf{\lambda}}^T \hat{\mathbf{\Theta}}} \). For the PARAFAC-IRS model, \( \hat{\mathbf{\Theta}} \) is reconstructed as

\[
\hat{\mathbf{\Theta}} = \sum_{r=1}^{R} \hat{\alpha}_r \left( \hat{\mathbf{s}}_r^{(p)} \otimes \cdots \otimes \hat{\mathbf{s}}_r^{(1)} \right)
\]

while \( \hat{\alpha}_r \) is the \( r \)-th phase shift factor quantized based on (36) and adopting the \( \hat{\mathbf{s}}_r^{(p)} \) being the quantized phase shifts according to (37) (38), while \( \hat{\alpha}_r^{(p)} \) is the quantized weighting factor according to (38).

E. On the Effect of the Factorization Parameters

In this section, we discuss the choice of factorization parameters and their impact on the system performance.

- **Number of factors**: Note that its minimum value is \( P = 2 \), while its maximum value is \( \log_2(N) \), where all the factors have the minimum size \( N_p = 2, \ p = \{1, \ldots, P\} \). Note also that \( P = 1 \) implies no factorization. For a fixed \( P \), several choices for \( N_1, \ldots, N_P \) are possible such that \( N = N_1 N_2 \cdots N_P \). Moreover, for a fixed \( N \), higher values of \( P \) imply fewer IRS phase shifts and, hence, a reduction in the feedback overhead, and vice-versa.

- **Number of components**: The number of components defines the rank \( R \) for the PARAFAC-IRS model (26) and the multi-linear ranks \( \{R_1, \ldots, R_P\} \) for the Tucker-IRS model (33). The choice of \( R \) or \( \{R_1, \ldots, R_P\} \) is driven by the channel conditions. For a scenario where the Tx-IRS-RX link is dominated by moderate/strong LOS component, choosing the PARAFAC-IRS model with \( R = 1 \) usually is the most cost-effective option, since it has the maximum feedback overhead with almost no performance loss in terms of SE. On the other hand, for mixed LOS/NLOS channels, one may choose higher performance loss in terms of SE. On the other hand, for mixed LOS/NLOS channels, one may choose higher performance loss in terms of SE.

- **Size of factor components**: The sizes \( \{N_1, N_2, \ldots, N_P\} \) of the factor components indicate the number of independent phase shifts, which may influence the SE. For example, considering \( N = 256 \) and adopting the PARAFAC-IRS model with \( R = 1 \) and \( P = 2 \), there are several possibilities. For example, we may choose \( \{N_1 = 128, N_2 = 2\} \) or \( \{N_1 = N_2 = 16\} \). The first choice implies 110 independent phase shifts to be reported in the feedback channel, while the second choice requires only 32 phase shifts. While the first choice has an increased feedback overhead than the second one, it offers a higher SE. Let us define \( N^{(p)} = \sum_{p=1}^{P} N_p \) as the number of independent phase shifts for a given number of factors \( P \). If the RX (or the network controller) aims to improve the
spectral efficiency based on a fixed number of factors $P$ and of components $R$ (for the PARAFAC-IRS model), the best design is the one that maximizes $N^{(P)}$. On the other hand, if the goal is to minimize the feedback overhead, the best choice for the $N_p$’s translates into minimizing $N^{(P)}$.

F. On the Effect of the Phase Shift and Weighting Factor Quantizations

After the factorization step, the phase shifts of the factor matrices $S^{(1)}, \ldots, S^{(P)}$ are quantized before being conveyed to the IRS controller. From the fact that the proposed method factorizes the IRS phase shift vector into $P$ smaller factors, we can choose different numbers of bits $b_1^{(1)}, \ldots, b_1^{(P)}$ for the $P$ factors, which provides additional flexibility to design the feedback control signaling and meet its capacity requirements. As for the weighting factors, they play a role when $R \geq 2$ (for the PARAFAC-IRS model) and $R_p \geq 2$, $p = 1, \ldots, P$, (for the Tucker-IRS model), since they control the importance of each rank-one component of the chosen model.

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed IRS phase shift overhead-aware feedback model in terms of model fitting, achievable data rate, total SE and EE. We assume the Rician fading model for the involved channels $G \in \mathbb{C}^{M_T \times N}$ and $H^{N \times M_R}$ of (17) [48], [49], [50], [51], i.e.,

$$H = \sqrt{\alpha_H} \frac{K_H}{K_H + 1} H_{\text{LOS}} + \sqrt{\alpha_H} \frac{1}{K_H + 1} H_{\text{NLOS}},$$

$$G = \sqrt{\alpha_G} \frac{K_G}{K_G + 1} G_{\text{LOS}} + \sqrt{\alpha_G} \frac{1}{K_G + 1} G_{\text{NLOS}},$$

where $1/\alpha_H$ and $1/\alpha_G$ are the path-loss components of the TX-IRS and IRS-RX links, respectively. The scalars $K_H$ and $K_G$ are the Rician factors associated with the channel matrices $H$ and $G$, in which, for experimental purposes, we assume $K_H = K_G = K$. The entries of $H_{\text{SLOS}}, G_{\text{NLOS}}$ are modeled as circularly symmetric complex Gaussian random variables, with zero mean and unit variance, i.e., $H_{\text{SLOS}} \sim \mathcal{CN}(0, I_{M_T})$ and $G_{\text{NLOS}} \sim \mathcal{CN}(0, I_{M_R})$. The channel matrices $H_{\text{LOS}}, G_{\text{LOS}}$ follow a geometric-based channel model, i.e., they are given as

$$H_{\text{LOS}} = b_{\text{IRS}} \cdot a_{\text{TX}}^H \in \mathbb{C}^{N \times M_T},$$

$$G_{\text{LOS}} = b_{\text{RX}} \cdot a_{\text{IRS}}^H \in \mathbb{C}^{M_R \times N},$$

where $a_{\text{TX}}, b_{\text{RX}}, b_{\text{IRS}},$ and $a_{\text{IRS}}$ are the arrays steering vectors of the TX, RX, and the IRS, respectively. More details are given in Appendix A.

We assume, as a benchmark, the optimum IRS phase shift vector from the upper bound algorithm of [38], i.e., the output of (19) without factorization. In other words, in this section, we compare the performance of the optimum solution in (19) with our proposed tensor-based low-rank approximation schemes (PARAFAC-IRS and Tucker-IRS).

Fig. 6. NMSE for the PARAFAC-IRS and the Tucker-IRS models. System configuration: $N = 1024$, $P = 3$, $N_1 = 256$, $N_2 = N_3 = 4$.

Fig. 7. Number of iterations to convergence of Algorithm 1. System configuration: $N = 1024$, $P = 3$, $N_1 = 64$, $N_2 = N_3 = 4$.

A. Representation Accuracy

In Figure 6, we illustrate the NMSE between the state-of-the-art optimized IRS phase shift vector, i.e., the output of the upper bound algorithm of [38] in (19), and the factorized versions of these phase shifts obtained with the PARAFAC-IRS and Tucker-IRS models. The NMSE is computed as

$$\text{NMSE}(s) = \frac{1}{L} \sum_{l=1}^{L} \frac{||s(l) - \hat{s}(l)||_F^2}{||s(l)||_F^2},$$

where $L = 10^3$ is the total number of Monte Carlo trials, $\hat{s} \in \mathbb{C}^{N \times 1}$ is the reconstructed phase shift vector obtained according to the PARAFAC-IRS model (using the ALS Algorithm 1) or the Tucker-IRS model (using the HOSVD Algorithm 2). For both cases, it can be observed that increasing the number of components yields a better representation accuracy. For instance, considering the PARAFAC-IRS model with $R = 3$, the NMSE is fairly low for all ranges of Rician factors $K$ (i.e., from NLOS-dominant to LOS-dominant scenarios). Also, when the Rician factor of the channels increases, the NMSE gap between the proposed approach and the state-of-the-art optimized IRS phase shift vector decreases. This is due to the fact that as $K \to \infty$ the channels are dominated by the LOS component which has a separable structure (see Appendix).

In Figure 7, we illustrate the total number of iterations required for the ALS Algorithm 1 to converge based on the...
convergence threshold $\epsilon = 10^{-4}$ from the difference of two consecutive iterations, given in (31). As expected, as the number $R$ of components increases, more iterations are required. However, for a Rician factor of 15 dB, all configurations require less than 10 iterations to achieve convergence.

In terms of achievable data rate, we compare the two proposed low-rank IRS models with the state-of-the-art solution [38] where the phase shift vector is not factorized given by (19). The achievable data rate is calculated as

$$SE = \log_2 \left( 1 + \frac{|w^H G \text{diag}(s) H q|^2}{\sigma_b^2} \right),$$

in bits/s/Hz, (47)

where $s \in \mathbb{C}^{N \times 1}$ is the optimum IRS phase shifts, which are given in (41) for the PARAFAC-IRS model and in (42) for the Tucker-IRS model.

In Figure 8, we assume $P = 3$ for the proposed low-rank models, with $N_1 = 64$, and $N_2 = N_3 = 4$. As expected, in this scenario the state-of-the-art solution [38] provides the performance upper bound, since no factorization is applied. Figure 8 (a) compares the models in an ideal scenario with continuous phase shifts and continuous values for the weighting factors. For notation convenience, for the Tucker-IRS model we define the vector $R_F = [R_1, R_2, \ldots, R_F]$ that contains the number of components for each factor for fixed $P$ (with $P = 3$ in this case). As expected, when the number of components increases, the achievable data rate also increases. We can also observe that the PARAFAC-IRS model with $R = 16$ as well as the Tucker-IRS model with $R_3 = [16, 4, 4]$ achieve the performance of the benchmark solution [38].

In practice, both the phase shift and the weighting factors have to be quantized. As illustrated in the results of Figure 8 (b), there is an optimal choice of $R$ for the PARAFAC-IRS model ($R = 4$). For higher values of $R > 4$ the performance degrades due to model overfitting. For the Tucker-IRS model, when the number of components in $R_3$ increases, the performance in the NLOS region ($K < -5$ dB) also improves at the cost of a higher feedback overhead. Note that, in the moderate/strong LOS scenario ($K > 5$ dB), the number of components for both models does not give a noticeable performance enhancement. The performance gap between the models is explained by the fact that the Tucker-IRS model has more representation degrees of freedom, involving linear combinations of a higher number of factor components. This implies an increased feedback cost compared to the PARAFAC-IRS model. Consequently, there is a trade-off between performance and feedback overhead offered by both low-rank models. From our experience, a proper model for a NLOS scenario would be the Tucker-IRS model, while PARAFAC-IRS one is preferable in moderate/strong LOS cases since it leads to the best performance with the lowest feedback cost.

In the following experiments, we consider the PARAFAC-IRS model, due to its simplicity and lower phase shift and weight feedback cost. We have evaluated extensively the results of the Tucker-IRS model and observed the same qualitative conclusions as those presented.

In Figure 9, we compare the achievable data rate of the PARAFAC-IRS model with $R = 1$, by varying the number of factors. This experiment assumes $N = 1024$, $M_R = M_T = 2$, and $b_F^{(p)} = b_F = 3$, $p = 1, \ldots, P$. We can observe that, in the NLOS-dominant region ($K < -5$ dB), increasing $P$ leads to performance degradation. This is due to the fact that larger values of $P$ imply fewer independent phase shifts, the number of which is given by $\sum_{p=1}^P N_p$. However, in the LOS-dominant region ($K > 5$ dB), the performance gap between the proposed model and the state-of-the-art [38] is reduced and becomes negligible for higher values of $K$. Indeed, as $K \rightarrow \infty$ the involved channels admit a separable Vandermonde structure that can be factorized as the Kronecker product of an arbitrary number $P$ of factors. Since the optimization of the IRS phase shift vector directly depends on these channels (see the Appendix), in a strong LOS scenario, all the factorization

\[Note that in strong LOS conditions (K \rightarrow \infty) one possibility to further reduce the feedback overhead would be to convey to the IRS controller the pair of angle of arrival (AOA) and angle of departure (AOD). In this particular case, the IRS controller would have to reconstruct the steering vectors from these angle pairs, and then find the optimum IRS phase shifts by solving an optimization problem.\]
feedback capacity. For instance, when per factor can be increased accordingly to meet the limited resolution (having 512 elements) can only be quantized with 1 bit. Note that the size of the factors is reduced, and the resolution per factor can be increased accordingly to meet the limited feedback capacity. For instance, when \( N_2 = [256, 4] \) and \( b^{(p)}_T = [3, 16] \), the total number of bits is 256·3 + 4·16 = 832.

We can note from these results that by increasing the resolution of the factors, the performance gets closer to that of the state-of-the-art phase shift with infinity resolution (solid curve). In particular, for \( K > -5 \text{ dB} \), our approach provides the best results.

### B. SE and EE Performances

In this section, we evaluate the performance in terms of SE and EE by taking into account the channel estimation duration \( T_E \) and the IRS phase shift feedback duration \( T_F \), given in (20)-(22). The channel estimation duration \( T_E \) is assumed to be the same for both state-of-the-art [38] and the proposed approach. For the state-of-the-art method, the feedback duration is given by (20) is used, while for the proposed PARAFAC-IRS model it is given by (39).

The channel estimation period, in (21), is given as \( T_E = (M_F N + 1) T_0 \), where \( T_0 = 0.8 \mu \text{ seconds} \) denotes the duration of the pilot sequence [38]. The frame duration is given by \( T = T_{PD} + T_F \), where \( T_{PD} = T_E + T_D \), is divided into 30% for pilot transmissions \( (T_E) \) and 70% for data transmission \( T_D \). Regarding the power parameters in (22), we have \( P_E = P_0 (1 + N M_F) T_0 \), where \( P_0 = 0.8 \text{ mW} \) is the pilot signal power. Other parameters of (21) and (22) are defined as \( P_{max}/P_{c,n} = 45/45/10 \text{ dBm}, B_{max} = 100 \text{ MHz}, N_0 = -174 \text{ dBm/Hz}, \alpha_H = \alpha_G = 110 \text{ dB}, \) and \( \mu_{HF} = 1/1 \).

The feedback channel \( g_F \) is generated from a circularly symmetric complex Gaussian distribution, normalized by \( \sqrt{B_F} = \sqrt{B_T} = \sqrt{\alpha_T} \) to account for the effects of path loss and shadowing, as given in Table I. We assume \( K = 10 \text{ dB}, N = 1024 \) and consider the PARAFAC-IRS model with \( R = 1 \). As for the number of factors, we study three configurations, with \( P = 2, (N_2 = [512, 2]), P = 3 \) \( (N_2 = [256, 2, 2]) \) and \( P = 10 \) \( (N_{10} = [2, \ldots , 2]) \in \mathbb{R}^{1 \times 10} \).

In Figures 11 and 12, we study the SE and EE of the proposed approach using the PARAFAC-IRS model with \( R = 1 \) and we compare its performance with the state-of-the-art [38] by varying the feedback bandwidth \( B_F = B_{max} - B \), where \( B_{max} = 100 \text{ MHz} \) is the total available bandwidth. As shown in these figures, as the number of factors increases the reduction on the feedback duration pays off in the total system SE and EE. The proposed method achieves a gain in the SE of 32% for \( P = 10 \), 20% for \( P = 3 \), and 14% for \( P = 2 \), over the state-of-the-art, considering the \( B_T = 200 \text{ kHz} \). A similar gain in the EE is observed.

In Figures 12(a) and 12(b), we compare the proposed PARAFAC-IRS model with \( R = 1 \) and \( P = 10 \) with the state-of-the-art solution by varying the number of antennas. We can see from Figure 12(a) that for a feedback bandwidth \( B_T \leq 200 \text{ kHz} \), the proposed approach operating with a \( 2 \times 2 \) MIMO setup provides better SE results than the state-of-the-art one operating with the \( 16 \times 16 \) MIMO setup. In Figure 12(b), the proposed approach achieves a higher EE gain than the state-of-the-art in both setups \( (4 \times 4, 64 \times 64) \). This gain can be explained by the fact that for \( K = 10 \text{ dB} \) the proposed approach has a similar data rate than the reference solution in (19) but a significantly reduced (50 times smaller) feedback duration \( T_F \) (see Fig. 9).
resulting in higher SE and EE performances. Finally, Figures 13(a) and 13(b) show the SE and EE performances as a function of the feedback power \( p_F \), with \( p_F = P_{\text{max}} - p_{\text{TX}} \). We notice that the proposed configurations provide remarkable results in all scenarios, outperforming the reference solution. To summarize the results illustrated in Figures 11-13, we conclude that the proposed tensor-based IRS factorization models, by reducing the IRS phase shift feedback overhead, offer SE and EE performance enhancements, when systems aspects, such as the channel estimation duration \( T_E \), and the IRS phase shift feedback duration \( T_F \), are taken into account. In addition, our approach reaches similar performance to the non-factorized IRS solution, especially in moderate/strong LOS scenarios, as can be seen in Figures 8-10.
VII. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed low-rank tensor-based models to represent the IRS phase shifts, which allows for controlling the feedback overhead. We showed that the proposed models offer flexibility to meet the desired tradeoff between performance enhancement and feedback overhead reduction. The PARAFAC-IRS method is preferable in the case of moderate/strong LOS scenarios, achieving a spectral efficiency that is close to that of the state-of-the-art solution while providing a significant feedback overhead reduction. For NLOS scenarios, the Tucker-IRS model achieves higher data rates than the PARAFAC-IRS model at the expense of higher feedback overhead. From a system-level viewpoint, the network can adapt the IRS factorization parameters to meet a determined quality of service. One can also resort to the proposed low-rank IRS models to increase the feedback periodicity, i.e., by providing more frequent feedback, which is crucial in fast time-varying channels, where the IRS should be reconfigured more frequently to follow the environment changes. Moreover, the proposed low-rank IRS models allow the network to multiplex more IRS phase shifts in the same feedback channel, which is useful to accommodate multi-user IRS-assisted communications.

APPENDIX

SEPARABILITY OF IRS PHASE SHIFTS

FOR LOS CHANNELS

In (45) and (46), we model the LOS components of the channels based on the array steering vectors of a geometric channel. In general, an array steering vector is given as

\[ e^{j \theta} = [1, e^{j \theta_1}, e^{j \theta_2}, \ldots, e^{j \theta_{N}}} \in \mathbb{C}^{N \times 1} \]

where \( \theta \) is the spatial frequency that depends on the array geometry, e.g., uniform linear array (ULA) or uniform rectangular array (URA). An array steering vector is a Vandermonde vector, which has a separable structure. Thus, it can be written in terms of Kronecker products of an arbitrary number \( P \) of smaller Vandermonde vectors, such that \( P = \prod_{p=1}^{P} N_p \) [43], i.e.,

\[ e^{j \theta} = [e^{j \theta_1}, e^{j \theta_2}, \ldots, e^{j \theta_{N}}] \in \mathbb{C}^{N \times 1} \]

is the spatial frequency of the \( p \)-th factor. From this result, we can rewrite the LOS channel matrices \( H_{\text{LOS}} \) and \( G_{\text{LOS}} \) defined in earlier in (45) and (46), as follows

\[ H_{\text{LOS}} = (b_{\text{IRS}}^{(p)} \otimes \cdots \otimes b_{\text{IRS}}^{(1)}) \cdot e^{j \Theta} \in \mathbb{C}^{N \times M_T} \]  

\[ G_{\text{LOS}} = b_{\text{RX}} \cdot (a_{\text{IRS}}^{(p)} \otimes \cdots \otimes a_{\text{IRS}}^{(1)})^H \in \mathbb{C}^{M_T \times N} \]  

To show that, for LOS channels, the Kronecker factorization of the IRS phase shift vector into \( P \) vectors is optimal and matches to the optimized upper bound solution of [38] in (19), first, note that the optimum IRS phase shift vector solves the following problem

\[ \max_{\theta_1, \ldots, \theta_{N}} |w^H G_{\text{LOS}} \text{diag}(s) H q|^2. \]  

Replacing the channels with their LOS and NLOS components yields

\[ \max_{\theta_1, \ldots, \theta_{N}} |w^H (\gamma_{\text{LOS}} G_{\text{LOS}} + \gamma_{\text{NLOS}} G_{\text{NLOS}}) \text{diag}(s) (H_{\text{LOS}} + H_{\text{NLOS}} q')|^2 \]  

where \( \gamma_{\text{LOS}} = \sqrt{\alpha_{\text{LOS}} K_{\text{LOS}}}, \gamma_{\text{NLOS}} = \sqrt{\alpha_{\text{NLOS}} K_{\text{NLOS}}}, H_{\text{LOS}} = \sqrt{\alpha_{\text{HLOS}} K_{\text{HLOS}}}, H_{\text{NLOS}} = \sqrt{\alpha_{\text{HNLOS}} K_{\text{HNLOS}}}. \]  

Assuming the LOS-dominant scenario for both channels, or similarly, assuming \( K_{\text{LOS}} \to \infty \), the above problem simplifies (or can be well approximated) as

\[ \max_{s} |w^H G_{\text{LOS}} \text{diag}(s) H_{\text{LOS}} q|^2. \]  

Let \( s = e^{j \theta} \), where \( \theta = [\theta_1, \ldots, \theta_{N}] \) is the vector of phases. From the rank-1 approximations \( G_{\text{LOS}} = \sigma_0 u_{\text{LOS}} v_{\text{LOS}}^H \) and \( H_{\text{LOS}} = \sigma_0 u_{\text{HLOS}} v_{\text{HLOS}}^H \), obtained, e.g., from truncated SVDs, the upper bound algorithm of [38] sets \( \theta = -\angle (u_{\text{LOS}} \otimes v_{\text{LOS}}) \). Due to the separable structure of LOS channels in (48) and (49), their rank-1 SVDS give

\[ G_{\text{LOS}} = \sigma_0 u_{\text{LOS}} (v_{\text{LOS}}^{(p)} \otimes \cdots \otimes v_{\text{LOS}}^{(1)})^H \]

\[ H_{\text{LOS}} = \sigma_0 u_{\text{HLOS}} (u_{\text{LOS}}^{(p)} \otimes \cdots \otimes u_{\text{LOS}}^{(1)}) v_{\text{HLOS}}^H \]

i.e., the dominant right singular vector of \( G_{\text{LOS}} \) and the dominant left singular vector of \( H_{\text{LOS}} \) are separable into the Kronecker product of \( P \) terms. Hence, the IRS phase shift vector \( s \) that maximizes problem (52), for fixed \( w \) and \( q \), can be factorized exactly as a Kronecker product of smaller subvectors, i.e.,

\[ s = e^{j \theta} = e^{j \angle (u_{\text{LOS}} \otimes v_{\text{LOS}})} \]

(53)

This equivalence implies the optimum IRS phase shift vector admits an exact factorization as a Kronecker product of \( P \) factors, i.e.,

\[ s = e^{j \theta} = e^{j \angle (u_{\text{LOS}}^{(p)} \otimes v_{\text{LOS}}^{(1)})}, \]

where the \( p \)-th factor is given by \( s^{(p)} = e^{j \angle (u_{\text{LOS}}^{(p)} \otimes v_{\text{LOS}}^{(1)})} \in \mathbb{C}^{N_p \times 1}, p = 1, \ldots, P \). In the presence of NLOS components in \( H \) and/or \( G \), the Kronecker factorization of the IRS phase shift vector in (53) is not exact and the accuracy of such an approximation will depend on the importance of the NLOS component with respect to the LOS one. As demonstrated in our numerical experiments, for Rician factors \( K_{\text{LOS}} = K_{\text{HLOS}} = K > 5 \) dB, the achievable data rate loss associated with this Kronecker factorization is negligible, while benefiting from a significantly reduced feedback overhead. On the other hand, for scenarios with moderate/strong NLOS components, the Kronecker factorization is no more a good option, which justifies using generalized low-rank PARAFAC-IRS or Tucker-IRS models that involve a higher number Kronecker product components to capture the more complex channel structure and thus compensating for such a model mismatch.

REFERENCES

[1] B. Sokal, P. R. B. Gomes, A. L. F. de Almeida, B. Makki, and G. Fedor, “IRS phase-shift feedback overhead-aware model based on rank-one tensor approximation,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2022, pp. 3911–3916.
Bruno Sokal (Student Member, IEEE) received the B.Sc. degree in telecommunications engineering from the Federal Institute of Science, Education and Technology of Ceará, Fortaleza, Brazil, in 2015, the M.Sc. degree in teleinformatics engineering from the Federal University of Ceará (UFC), Fortaleza, in 2017, and the dual Ph.D. degrees in electrical and teleinformatics engineering from the Ilmenau University of Technology, Ilmenau, Germany, and UFC, in 2022. Since 2017, he has been a Researcher with the Wireless Telecom Research Group (GTEL), UFC.

Paulo R. B. Gomes received the Diploma degree (magna cum laude) in physics from the Federal University of Ceará (UFC), Ceará, Brazil, in 2011, the Diploma degree in telecommunications engineering from the Federal Institute of Education, Science and Technology of Ceará (IFCE), Ceará, in 2014, and the M.Sc. and Ph.D. degrees in teleinformatics engineering from the UFC in 2014 and 2018, respectively. His research interests include tensor-based signal processing with applications to sensor arrays, radar, and wireless communication systems.

Gabor Fodor (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Budapest University of Technology and Economics in 1998, the Docent degree from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2019, and the D.Sc. degree from the Hungarian Academy of Sciences in 2019. He is currently a Master Researcher with Ericsson Research and an Adjunct Professor with the KTH Royal Institute of Technology. He was a co-recipient of the IEEE Communications Society Stephen O. Rice Prize in 2018. He is also serving as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE WIRELESS COMMUNICATIONS and as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Behrooz Makki (Senior Member, IEEE) received the Ph.D. degree in communication engineering from the Chalmers University of Technology, Gothenburg, Sweden. He was a Post-Doctoral Researcher at the Chalmers University of Technology from 2013 to 2017. He is currently working as a Senior Researcher with Ericsson Research, Gothenburg. He has coauthored more than 70 journal articles, more than 50 conference papers, and more than 90 patent applications. His current research interests include channel estimation, blind and semi-blind signal processing, and tensor-based signal processing. He served as an Associate Editor for several journals, such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2012–2014 and 2014–2016) and the IEEE SIGNAL PROCESSING LETTERS (2016–2018 and 2018–2020).

André L. F. de Almeida (Senior Member, IEEE) received the dual Ph.D. degrees in sciences and teleinformatics engineering from the University of Nice Sophia Antipolis, Sophia Antipolis, France, and the Federal University of Ceará, Fortaleza, Brazil, in 2007. He was awarded multiple times visiting professor positions at the University of Nice Sophia Antipolis (2012–2013, 2015, and 2018–2019). He is currently a Professor with the Department of Teleinformatics Engineering, Federal University of Ceará. His research interests include channel estimation, blind and semi-blind signal processing, and tensor-based signal processing. He served as an Associate Editor for several journals, such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2012–2014 and 2014–2016) and the IEEE SIGNAL PROCESSING LETTERS (2016–2018 and 2018–2020).

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.