参数化全息MIMO信道建模及其贝叶斯估计

袁正道*① 郭亚博② 高大伟③ 郭庆华④ 黄崇文⑤ 廖桂生③

①(河南开放大学信息工程与人工智能学院 郑州 450000)

②(河南工业大学信息科学与工程学院 郑州 450000)

③(西安电子科技大学杭州研究院 杭州 311200)

④(澳大利亚伍伦贡大学 伍伦贡 2522)

⑤(浙江大学信息与电子工程学院 杭州 310027)

摘 要：全息多输入多输出(HMIMO)技术因其高空间复用效率和信道容量被视为6G通信系统的关键技术之一，但电磁传播模型复杂、用户角度随机给电磁信道建模和估计带来较大困难。现有方法依赖简化假设或统计模型，存在模型失配问题，且难以同时解耦信道、位置与角度。针对上述挑战，该文提出一种融合神经网络、凸优化和因子图的混合信道建模与估计方法，该方法首先学习信道与坐标的非线性映射关系，构建参数化信道模型；其次基于欧拉角旋转理论描述用户角度，并将其嵌入因子图实现信道、坐标及角度的全局建模；最后利用消息传递算法完成参数联合解耦与信道估计。仿真结果表明，所提方法的信道估计误差较现有近似方法降低3 dB以上。该研究突破了现有方法对天线平行假设的依赖，为复杂电磁环境下的高精度信道估计与位置感知提供了新的解决方案。

关键词：全息多输入多输出；神经网络；贝叶斯估计；欧拉角旋转；消息传递算法

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2025)11-4176-12

DOI: 10.11999/JEIT250436 **CSTR**: 32379.14.JEIT250436

1 引言

全息多输入多输出(Holographic MIMO, HMIMO)技术通过连续孔径和超表面天线实现电磁波的全面操控，被视为6G突破5G固有极限的核心技术之一[1]。图1给出了HMIMO与传统大规模MIMO (massive MIMO, mMIMO)的对比示意图。相较于mMIMO的离散天线阵列，HMIMO能够以亚波长连续孔径天线，精确操控电磁场分布，显著提升空间复用效率与信道容量[2]。通过减小天线间距，HMIMO系统能够充分利用耦合[3]实现超方向性增益。以上技术优势使得HMIMO通信已经成为国内外研究热点，例如针对HMIMO系统的波束成形设计[4]、全息定位[5]、全息超表面设计[6]和通信感知一体化(Integrated Sensing And Communication, ISAC)[7]等，但以上研究的前提是准确的信道建模和估计。

针对HMIMO信道建模和估计问题，文献[8]考虑到非各向同性散射和定向天线，研究了包含HMIMO角度信息的信道模型，并利用阵列几何结构导致的低秩性，提出了对应的信道估计方案。通过对信道协方差矩阵进行适当近似，文献[9]设计了一种低复杂度的HMIMO信道估计方案，实现了较高的估计精度。文献[10]提出了基于压缩感知解构的变分贝叶斯推理方法，用于估计方位角、仰角、距离参数和稀疏信道。文献[11]利用连续表面产生的辐射波束的特定结构，提出了一种参数化物理信道模型。上述信道估计方法均基于一些近似模型假设，并未直接从电磁场理论出发，使得这些模型在近场区域存在较大失配[4,12]。在容量研究方面，文献[13]利用并矢格林函数推导了HMIMO系统的信道容量，但其假定用户端配置偶极子单天线，简化了模型复杂度；文献[14]则扩展到了平面天线，得到了信道容量和信道估计精度下界；文献[4]研究了针对超表面HMIMO系统3重极化信道建模和预编码方案，但并未给出信道估计方法。

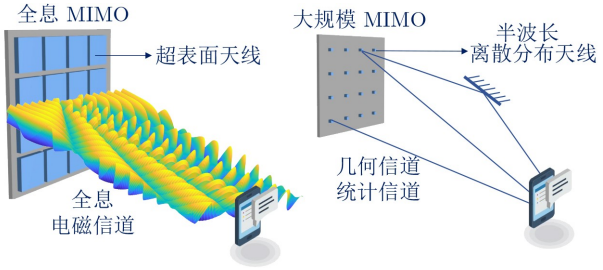


图 1 HMIMO和大规模MIMO区别

收稿日期：2025-05-20；改回日期：2025-08-31；网络出版：2025-09-08

*通信作者： 袁正道 yuanzhengdao@haou.edu.cn

基金项目：国家自然科学基金(62301394, 62331023, 62394292)，河南省科技攻关项目(252102210228)，国家重点研发计划(2021YFA1000500, 2023YFB2904804)

Foundation Items: The National Natural Science Foundation of China(62301394, 62331023, 62394292), The Science and Technology Research Project of Henan Province(252102210228), China National Key R and D Program (2021YFA1000500, 2023YFB2904804)

为准确估计近场HMIMO信道，须利用电磁传播理论或并矢格林函数，但会导致难以处理的积分和非线性问题[15]。将复杂信道由少量参数建模和估计，称之为参数化方法，能够显著降低自由变量个数，提升估计性能[16]。此外，参数化估计方法可以提取用户坐标，同步实现位置感知。基于参数化信道模型，文献[17]利用酉变换近似消息传递(Unitary Approximate Message Passing,UAMP)[18,19]推导了HMIMO近场信道估计算法，得到了良好的效果，但其要求用户端天线与基站端天线平行且投影对齐。由于用户坐标、角度与信道存在紧耦合关系且难以解耦，使得目前针对用户随机角度条件下的信道估计问题尚未得到解决。

为了解耦“信道-坐标-角度”参数，在随机用户角度下实现信道、坐标的估计，本文提出一种将神经网络、凸优化方法嵌入因子图的混合信道估计方法。本方法首先利用反向传播(Back Propagation, BP)神经网络学习全息信道和基站-用户相对坐标之间的非线性关系，然后将神经网络嵌入到因子图中实现全局优化，利用动态泰勒展开将神经网络线性近似，最终实现了神经网络的双向消息穿透。针对用户随机角度问题，本文引入图形学中的欧拉旋转方法，利用凸优化动态估计旋转映射矩阵，最终实现坐标、姿态参数解耦和信道估计。本文的主要贡献如下：第一，利用因子图将欧拉角纳入参数化信道模型中，实现了任意用户天线角度情况下的参数化信道建模；第二，将神经网络和凸优化方法作为因子节点嵌入因子图，实现非线性函数的全局优化；第三，利用泰勒展开等方法，实现了神经网络和凸优化节点的双向消息穿透，得到信道、坐标和角度的联合解耦与估计。仿真结果表明，相对于文献中现有方法，本文所提算法表现出较高的性能增益和鲁棒性。

2　参数化HMIMO信道建模

2.1　并矢格林函数全息信道

考虑如图2所示的上行HMIMO通信系统，其中接收端(基站)和发射端(用户)均配备有全息表面，分别由 $M = M_{\mathrm{row}} \times M_{\mathrm{col}}$ 和 $N = N_{\mathrm{row}} \times N_{\mathrm{col}}$ 个超表面贴片天线(以下简称贴片)组成。每个贴片能够在3个极化方向上独立发送或接收信号，发射和接收贴片的尺寸分别为 $\Delta_x^{\mathrm{t}} \times \Delta_y^{\mathrm{t}}$ 和 $\Delta_x^{\mathrm{r}} \times \Delta_y^{\mathrm{r}}$。假设接收表面位于 xoy 平面，且第1个接收贴片的中心位于坐标系原点。将第 m 个接收贴片和第 n 个发射贴片分别记为 S_m^{r} 和 S_n^{t}，对应中心的坐标向量分别为 $\boldsymbol{r}_m^{\mathrm{r}} = (x_m^{\mathrm{r}}, y_m^{\mathrm{r}}, z_m^{\mathrm{r}})^{\mathrm{T}}$ 和 $\boldsymbol{r}_n^{\mathrm{t}} = (x_n^{\mathrm{t}}, y_n^{\mathrm{t}}, z_n^{\mathrm{t}})^{\mathrm{T}}$。

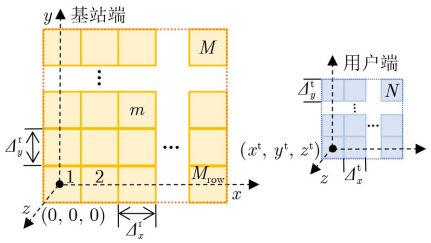


图 2 HMIMO发射和天线结构示意图

当发射表面 S^{t} 上位置 $\boldsymbol{r}^{\mathrm{t}}$ 处存在电流 $\boldsymbol{J}(\boldsymbol{r}^{\mathrm{t}})$，在自由空间中位置 $\boldsymbol{r}^{\mathrm{r}}$ 处的辐射电场 $\boldsymbol{E}(\boldsymbol{r}^{\mathrm{r}})$ 可表示为 $\boldsymbol{E}(\boldsymbol{r}^{\mathrm{r}}) = \mathrm{i}\omega\mu \int_{S^{\mathrm{t}}} \boldsymbol{G}(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}})\boldsymbol{J}(\boldsymbol{r}^{\mathrm{t}})\mathrm{d}s^{\mathrm{t}}$，其中 ω 是介电常数，μ 是磁导率，$\boldsymbol{G}(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}})$ 是并矢格林函数，表示为 $\boldsymbol{G}(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}}) = g(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}})[c_1(r)\boldsymbol{I}_3 + c_2(r)\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}}]$，其中标量函数 $c_1(r)$ 和 $c_2(r)$ 分别表示为

$$c_1(r) = 1 + \mathrm{i}/k_0 r - 1/k_0^2 r^2,\ c_2(r) = 3/k_0^2 r^2 - 3\mathrm{i}/k_0 r - 1 \tag{1}$$

单位方向向量 $\boldsymbol{r}$ 表示为 $\boldsymbol{r} = (\boldsymbol{r}^{\mathrm{t}} - \boldsymbol{r}^{\mathrm{r}})/\left\|\boldsymbol{r}^{\mathrm{t}} - \boldsymbol{r}^{\mathrm{r}}\right\|$，标量 $r = \left\|\boldsymbol{r}^{\mathrm{t}} - \boldsymbol{r}^{\mathrm{r}}\right\|$ 表示发送和接收位置间距离，$\boldsymbol{I}_3$ 为3阶单位矩阵，$k_0 = 2\pi/\lambda$ 是波数，λ 为波长，标量格林函数 $g(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}})$ 为

$$g(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}}) = \exp(\mathrm{i}k_0\left\|\boldsymbol{r}^{\mathrm{t}} - \boldsymbol{r}^{\mathrm{r}}\right\|)/4\pi\left\|\boldsymbol{r}^{\mathrm{t}} - \boldsymbol{r}^{\mathrm{r}}\right\| \tag{2}$$

首先考虑单对发射-接收贴片，假设发射贴片上的电流均匀分布，且不存在非视距传播，则第 n 个发射贴片 S_n^{t} 和第 m 个接收贴片 S_m^{r} 之间具有极化特性的无线信道可表示为

$$\bar{\boldsymbol{H}}_{mn} = \mathrm{i}\omega\mu \int_{x_m^{\mathrm{r}}-\Delta_x^{\mathrm{r}}/2}^{x_m^{\mathrm{r}}+\Delta_x^{\mathrm{r}}/2} \int_{y_m^{\mathrm{r}}-\Delta_y^{\mathrm{r}}/2}^{y_m^{\mathrm{r}}+\Delta_y^{\mathrm{r}}/2} \int_{x_n^{\mathrm{t}}-\Delta_x^{\mathrm{t}}/2}^{x_n^{\mathrm{t}}+\Delta_x^{\mathrm{t}}/2} \int_{y_n^{\mathrm{t}}-\Delta_y^{\mathrm{t}}/2}^{y_n^{\mathrm{t}}+\Delta_y^{\mathrm{t}}/2} \cdot \boldsymbol{G}\left(\boldsymbol{r}^{\mathrm{t}},\boldsymbol{r}^{\mathrm{r}}\right)\mathrm{d}y^{\mathrm{t}}\mathrm{d}x^{\mathrm{t}}\mathrm{d}y^{\mathrm{r}}\mathrm{d}x^{\mathrm{r}} \tag{3}$$

信道矩阵 $\bar{\boldsymbol{H}}_{mn} \in \mathbb{C}^{3\times 3}$ 可写为

$$\bar{\boldsymbol{H}}_{mn} = \begin{bmatrix} h_{mn}^{xx} & h_{mn}^{xy} & h_{mn}^{xz} \\ h_{mn}^{xy} & h_{mn}^{yy} & h_{mn}^{yz} \\ h_{mn}^{xz} & h_{mn}^{yz} & h_{mn}^{zz} \end{bmatrix} \tag{4}$$

其中 $h_{mn}^{\kappa}, \kappa \in \{xx, xy, xz, yy, yz, zz\}$ 表示对应不同发射和接收极化组合的信道分量。例如 xx 代表发射和接收均为 x 极化，而 xy 则代表交叉极化干扰。从文献[17]中可以看出，单极化扩展到3极化是比较直接的。本文未深入讨论极化维度，在推导中仅考虑 x 极化，即忽略 y, z 极化和对应的交叉极化干扰，使得信道矩阵 $\bar{\boldsymbol{H}}_{mn}$ 退化为标量形式 h_{mn}。从而在本文后续部分使用 h_{mn} 替代 $h_{mn}^{\kappa}, \kappa \in \{xx, xy, xz, yy, yz, zz\}$ 及对应 $\bar{\boldsymbol{H}}_{mn}$。考虑所有发射和接收贴片，得到信道矩阵为式(5)

$$\tilde{\boldsymbol{H}}=\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{bmatrix}\in\mathbb{C}^{M\times N} \tag{5}$$

2.2 信道估计问题数学抽象

假设全数字系统，将L个连续的接收向量排列成矩阵，可得$\tilde{\boldsymbol{Y}}=\tilde{\boldsymbol{H}}\tilde{\boldsymbol{S}}+\tilde{\boldsymbol{W}}$，其中$\tilde{\boldsymbol{Y}}\in\mathbb{C}^{M\times L}$表示接收信号，$\tilde{\boldsymbol{S}}\in\mathbb{C}^{N\times L}$为导频矩阵，$\tilde{\boldsymbol{W}}\in\mathbb{C}^{M\times L}$表示高斯白噪声，其噪声精度为$\gamma$(即方差为$\gamma^{-1}$)。将接收模型整理为

$$\bar{\boldsymbol{Y}}=\boldsymbol{S}\boldsymbol{H}+\bar{\boldsymbol{W}} \tag{6}$$

其中，$\bar{\boldsymbol{Y}}=\tilde{\boldsymbol{Y}}^{\mathrm{T}}\in\mathbb{C}^{L\times M}$，$\boldsymbol{S}=\tilde{\boldsymbol{S}}^{\mathrm{T}}\in\mathbb{C}^{L\times N}$，$\boldsymbol{H}=\tilde{\boldsymbol{H}}^{\mathrm{T}}\in\mathbb{C}^{N\times M}$。信道估计的目标是基于导频信号$\boldsymbol{S}$和接收信号$\boldsymbol{Y}$估计信道矩阵$\boldsymbol{H}$。根据式(3)，信道$\boldsymbol{H}$由$\boldsymbol{r}^{\mathrm{r}}$和$\boldsymbol{r}^{\mathrm{t}}$参数化，可以大幅减少待估计的变量数量，提升信道估计性能。从式(3)还可以看出，$\bar{\boldsymbol{H}}_{mn}$与$\boldsymbol{r}^{\mathrm{t}}$之间存在复杂的积分运算关系，甚至没有闭式解，这给参数化信道估计带来较大挑战。文献[4]通过一些近似得到了HMIMO信道的近似解析表达式，表示为

$$\begin{aligned}\bar{\boldsymbol{H}}_{mn}\approx&\mathrm{i}\omega\mu\Delta^{\mathrm{t}}\Delta^{\mathrm{r}}\frac{\exp\left(\mathrm{i}k_0r_{mn}\right)}{4\pi r_{mn}}\\&\cdot\operatorname{sinc}\frac{k_0\left(x_m^{\mathrm{r}}-x_n^{\mathrm{t}}\right)\Delta_x^{\mathrm{t}}}{2r_{mn}}\\&\cdot\operatorname{sinc}\frac{k_0\left(y_m^{\mathrm{r}}-y_n^{\mathrm{t}}\right)\Delta_y^{\mathrm{t}}}{2r_{mn}}C_{mn}\end{aligned} \tag{7}$$

其中$\Delta^{\mathrm{t}}=\Delta_x^{\mathrm{t}}\Delta_y^{\mathrm{t}}$和$\Delta^{\mathrm{r}}=\Delta_x^{\mathrm{r}}\Delta_y^{\mathrm{t}}$分别表示发射和接收贴片的面积，$r_{mn}=\|\boldsymbol{r}_n^{\mathrm{t}}-\boldsymbol{r}_m^{\mathrm{r}}\|$，$\boldsymbol{C}_{mn}=c_1(r_{mn})\boldsymbol{I}_3+c_2(r_{mn})\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}}$，$c_1(r_{mn})$和$c_2(r_{mn})$通过将式(3)中的$\boldsymbol{r}^{\mathrm{t}}$和$\boldsymbol{r}^{\mathrm{r}}$分别替换为$\boldsymbol{r}_n^{\mathrm{t}}$和$\boldsymbol{r}_m^{\mathrm{r}}$得到。本文采用神经网络模型来描述$\boldsymbol{H}$与$\boldsymbol{r}^{\mathrm{t}}$之间的非线性关系，该模型表达更精确，能够实现高效的参数化信道估计。

2.3 基于神经网络的参数化信道模型

从式(3)中可以看出在HMIMO系统中信道$\boldsymbol{H}$与坐标$\boldsymbol{r}^{\mathrm{t}}$紧密耦合，难以通过常用解耦方式进行参数化信道估计。本节首先提出基于神经网络的信道解耦模型，并对神经网络进行训练。

根据式(7)，第m个接收贴片与第n个发射贴片之间的信道分量h_{mn}由$\boldsymbol{r}_n^{\mathrm{t}}=(x_n^{\mathrm{t}},y_n^{\mathrm{t}},z_n^{\mathrm{t}})^{\mathrm{T}}$参数化。直接的方法是训练简单神经网络模型替代式(3)，但从式(7)中可以看出信道h_{mn}中存在周期项$(\mathrm{e}^{\mathrm{i}k_0r_{mn}})$和尺度项(剩余部分)，这会使得信道变化复杂，需要使用具有高表达能力的神经网络(如卷积网络)，进而导致需要大量的网络参数，将极大提高参数训练和信道估计复杂度。由此本文不直接拟合h_{mn}，转而拟合其中除去了周期项的尺度部分，即$h_{mn}/\mathrm{e}^{\mathrm{i}k_0r_{mn}}$，表示为

$$\bar{h}_{mn}=h_{mn}\left(x_{mn},y_{mn},z_{mn}\right)/\exp\left(\mathrm{i}k_0r_{mn}\right) \tag{8}$$

其中，变量$x_{mn}=x_n^{\mathrm{t}}-x_m^{\mathrm{r}}$，$y_{mn}=y_n^{\mathrm{t}}-y_m^{\mathrm{r}}$，$z_{mn}=z_n^{\mathrm{t}}-z_m^{\mathrm{r}}$，$\boldsymbol{r}_{mn}=(x_{mn},y_{mn},z_{mn})^{\mathrm{T}}$，$r_{mn}=\|\boldsymbol{r}_{mn}\|$。如图3所示，神经网络将$\bar{\boldsymbol{h}}_{mn}$的实部和虚部分开处理，其输入为相对坐标$x_{mn}$，$y_{mn}$和$z_{mn}$，输出为$\{\mathrm{Re}(\bar{\boldsymbol{h}}_{mn}),\mathrm{Im}(\bar{\boldsymbol{h}}_{mn})\}$。根据图3，神经网络的输出可以表示为

$$\begin{aligned}\mathbf{NN}\left(x_{mn},y_{mn},z_{mn}\right)=&\boldsymbol{W}_2^{\mathrm{T}}g_a\left(\boldsymbol{w}_1^xx_{mn}+\boldsymbol{w}_1^yy_{mn}\right.\\&\left.+\boldsymbol{w}_1^zz_{mn}+\boldsymbol{b}_1\right)+\boldsymbol{b}_2\\=&\boldsymbol{\xi}_{mn}\in\mathbb{R}^{2\times1}\end{aligned} \tag{9}$$

其中，$\boldsymbol{W}_1=[\boldsymbol{w}_1^x,\boldsymbol{w}_1^y,\boldsymbol{w}_1^z]\in\mathbb{R}^{N_{\mathrm{h}}\times3}$是输入层权重矩阵，$\boldsymbol{W}_2=[\boldsymbol{w}_{2,1},\boldsymbol{w}_{2,2}]\in\mathbb{R}^{N_{\mathrm{h}}\times2}$是输出层权重矩阵，$N_{\mathrm{h}}$是隐层神经元的数量，$\boldsymbol{b}_1$和$\boldsymbol{b}_2$分别是隐层和输出层的偏置向量。隐层的激活函数为$g_a(\cdot)=\mathrm{tansig}(\cdot)$，输出层使用线性激活函数。将图3中的2个输出合并，记为$\varphi(x_{mn},y_{mn},z_{mn})=\xi_{mn}(1)+\mathrm{i}\xi_{mn}(2)$，最终得到参数化信道模型

$$h_{mn}\approx\varphi\left(x_{mn},y_{mn},z_{mn}\right)\exp\left(\mathrm{i}k_0r_{mn}\right) \tag{10}$$

借助神经网络，本文将复杂的信道模型式(3)转换为更简洁的式(10)。借助精确模型式(3)，通过数值方法可以生成足够的训练样本，使用反向传播算法，通过最小化以下损失函数离线训练神经网络

$$\begin{aligned}\mathcal{L}=&\sum\nolimits_{m,n}\|\boldsymbol{\varphi}_{mn}\left(x_{mn},y_{mn},z_{mn}\right)\\&-\bar{h}_{mn}\left(x_{mn},y_{mn},z_{mn}\right)\|^2/(M\times N)\end{aligned} \tag{11}$$

其中$\bar{h}(x_{mn},y_{mn},z_{mn})$通过式(8)计算得到。训练方法采用结合牛顿法和梯度下降法。后续将训练后的神经网络记为$\varphi(x_{mn},y_{mn},z_{mn})$，并将其作为一个函数节点嵌入因子图，实现全局统一优化。

3 用户天线随机角度建模与因式分解

从式(10)可以看出信道$\boldsymbol{H}$可以参数化为坐标$\boldsymbol{r}_n^{\mathrm{t}}$，$\forall n$，共$3N$个未知参数。此外用户天线在一个平面上，贴片坐标$\boldsymbol{r}_n^{\mathrm{t}},\forall n$之间存在约束关系，可以利用其降低独立参数个数，提升信道估计性能。本节首

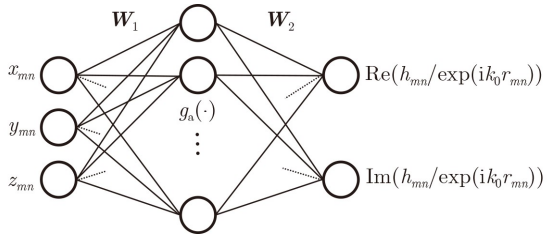


图 3 BP神经网络示意图

先引入图形学中的三维旋转理论，将信道参数降低为一组坐标和旋转角度，然后利用贝叶斯方法建立因子图模型。

3.1 基于欧拉角的随机角度建模

由于基站天线位于xoy平面，且第1个贴片中心坐标为$(0,0,0)$，则基站端第m个贴片坐标具有以下关系$x_m^{\mathrm{r}}=(c_m^{\mathrm{r}}-1)\Delta_x^{\mathrm{r}}$, $y_m^{\mathrm{r}}=(l_m^{\mathrm{r}}-1)\Delta_y^{\mathrm{r}}$, $z_m^{\mathrm{r}}=0$，其中$c_m^{\mathrm{r}}=\mathrm{mod}\,(m,M_{\mathrm{col}})$, $l_m^{\mathrm{r}}=\mathrm{ceil}(m/M_{\mathrm{col}})$，此处$\mathrm{mod}\,()$表示取模运算，$\mathrm{ceil}()$为向下取整运算。对于用户端，当假定用户表面与基站表面平行且投影对齐(如文献[17]中假设条件)时，第n和第1个发射贴片中心坐标之间存在关系为

$$\begin{aligned}&x_n^{\mathrm{t}}=x^{\mathrm{t}}+(c_n^{\mathrm{t}}-1)\Delta_x^{\mathrm{t}}=x^{\mathrm{t}}+\bar{b}_n^x,\\&y_n^{\mathrm{t}}=y^{\mathrm{t}}+(l_n^{\mathrm{t}}-1)\Delta_y^{\mathrm{t}}=y^{\mathrm{t}}+\bar{b}_n^y,\ z_n^{\mathrm{t}}=z^{\mathrm{t}}\end{aligned}\tag{12}$$

其中，定义$\boldsymbol{r}^{\mathrm{t}}=(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}})^{\mathrm{T}}$表示用户天线第1个贴片的中心坐标，$\bar{\boldsymbol{b}}_n=(\bar{b}_n^x,\bar{b}_n^y,0)^{\mathrm{T}}$为平行偏置向量，表示在平行状态下第$n$个贴片坐标$\boldsymbol{r}_n$与第1个贴片坐标$\boldsymbol{r}^{\mathrm{t}}$之间的偏置，表示为矩阵形式$\bar{\boldsymbol{B}}=(\bar{\boldsymbol{b}}_1,\bar{\boldsymbol{b}}_2,\cdots,\bar{\boldsymbol{b}}_N)\in\mathbb{R}^{3\times N}$，$c_n^{\mathrm{t}}$和$l_n^{\mathrm{t}}$的定义与$c_m^{\mathrm{r}}$和$l_m^{\mathrm{r}}$相同。根据图形学中的三维旋转理论，平面能够通过3次分别绕x, y和z轴的旋转达到任意角度。每次旋转相等效于左乘3×3的映射矩阵，所以当用户平面并不平行时，坐标向量$\boldsymbol{r}_n^{\mathrm{t}}$与$\boldsymbol{r}^{\mathrm{t}}$、欧拉旋转矩阵$\boldsymbol{A}$之间存在映射关系，$\boldsymbol{r}_n^{\mathrm{t}}=\boldsymbol{A}\bar{\boldsymbol{b}}_n+\boldsymbol{r}^{\mathrm{t}}=\boldsymbol{b}_n+\boldsymbol{r}^{\mathrm{t}}$，此处定义$\boldsymbol{b}_n=\boldsymbol{A}\bar{\boldsymbol{b}}_n=(b_n^x,b_n^y,b_n^z)^{\mathrm{T}}$为非平行状态下$\boldsymbol{r}_n^{\mathrm{t}}$相对于$r^{\mathrm{t}}$的偏移向量。旋转矩阵$\boldsymbol{A}$取决于3个坐标轴旋转角度$\boldsymbol{\theta}=(\theta_x,\theta_y,\theta_z)$，其中

$$\begin{aligned}\boldsymbol{A}=\boldsymbol{A}_x\boldsymbol{A}_y\boldsymbol{A}_z=&\begin{pmatrix}1&0&0\\0&\cos\theta_x&-\sin\theta_x\\0&\sin\theta_x&\cos\theta_x\end{pmatrix}\\&\cdot\begin{pmatrix}\cos\theta_y&0&\sin\theta_y\\0&1&0\\-\sin\theta_y&0&\cos\theta_y\end{pmatrix}\\&\cdot\begin{pmatrix}\cos\theta_z&-\sin\theta_z&0\\\sin\theta_z&1&0\\0&0&1\end{pmatrix}\end{aligned}\tag{13}$$

即矩阵$\boldsymbol{A}$可以参数化为角度$\boldsymbol{A}(\boldsymbol{\theta})$。将用户天线的所有贴片中心坐标向量整合为矩阵形式可得

$$\boldsymbol{R}=\boldsymbol{A}(\boldsymbol{\theta})\bar{\boldsymbol{B}}+\boldsymbol{r}^{\mathrm{t}}\otimes\boldsymbol{1}_{1\times N}=\boldsymbol{B}+\boldsymbol{r}^{\mathrm{t}}\otimes\boldsymbol{1}_{1\times N}\tag{14}$$

其中，$\boldsymbol{R}=(\boldsymbol{r}_1^{\mathrm{t}},\boldsymbol{r}_2^{\mathrm{t}},\cdots,\boldsymbol{r}_N^{\mathrm{t}})$, $\boldsymbol{B}=(\boldsymbol{b}_1,\boldsymbol{b}_2,\cdots,\boldsymbol{b}_N)$。根据式(8)，相对坐标$\boldsymbol{r}_{mn}=(x_{mn},y_{mn},z_{mn})^{\mathrm{T}}$可以表示为$\boldsymbol{r}_n^{\mathrm{t}},\forall n$的函数

$$x_{mn}=x^{\mathrm{t}}+b_n^x-x_m^{\mathrm{r}},y_{mn}=y^{\mathrm{t}}+b_n^y-y_m^{\mathrm{r}},z_{mn}=z^{\mathrm{t}}+b_n^z\tag{15}$$

由于接收贴片坐标$\{x_m^{\mathrm{r}},y_m^{\mathrm{r}}\}$已知，本文将式(10)中的$\varphi(x_{mn},y_{mn},z_{mn})$重新定义为函数$\phi_{mn}(x_n^{\mathrm{t}},y_n^{\mathrm{t}},z_n^{\mathrm{t}})$。这意味着任何信道分量$h_{nm}$都可以表示为第1个发射贴片坐标$\boldsymbol{r}^{\mathrm{t}}=(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}})^{\mathrm{T}}$的函数。

3.2 因式分解和因子图表示

针对信道估计问题，本文首先基于奇异值分解(Singular Value Decomposition, SVD) $\boldsymbol{S}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}$对式(6)进行酉变换，得到$\boldsymbol{Y}=\boldsymbol{\Phi}\boldsymbol{H}+\boldsymbol{W}$，其中$\boldsymbol{Y}=\boldsymbol{U}^{\mathrm{H}}\bar{\boldsymbol{Y}}\in\mathbb{C}^{L\times M}$，$\boldsymbol{\Phi}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{S}\in\mathbb{C}^{L\times N}$。由于$\boldsymbol{U}^{\mathrm{H}}$是酉矩阵，可知$\boldsymbol{W}=\boldsymbol{U}^{\mathrm{H}}\bar{\boldsymbol{W}}\in\mathbb{C}^{L\times M}$仍然表示零均值高斯白噪声。在给定观测矩阵$\boldsymbol{Y}$的情况下，未知变量的条件联合概率密度函数可以分解为

$$\begin{aligned}&p\left(\boldsymbol{Y},\boldsymbol{H},\gamma,\boldsymbol{R},\bar{\boldsymbol{R}},\boldsymbol{A},\boldsymbol{\theta},\boldsymbol{r}^{\mathrm{t}}\right)\\&=p\left(\boldsymbol{Y}|\boldsymbol{H},\gamma\right)p(\boldsymbol{H}|\boldsymbol{R})p\left(\boldsymbol{R}|\boldsymbol{B},\boldsymbol{A}\right)\\&\quad\cdot p(\boldsymbol{B}|\boldsymbol{A})p\left(\boldsymbol{A}|\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{r}^{\mathrm{t}}\right)\\&=p\left(\boldsymbol{Y}|\boldsymbol{H},\gamma\right)p(\underbrace{h_{mn},\forall m,n}_{\text{信道}\boldsymbol{H}}|\boldsymbol{R})p(\underbrace{\boldsymbol{r}_n^{\mathrm{t}},\forall n}_{\text{坐标}\boldsymbol{R}}|\underbrace{\boldsymbol{b}_n,\forall n}_{\text{偏置}\boldsymbol{B}},\boldsymbol{r}^{\mathrm{t}})\\&\quad\cdot p(\boldsymbol{B}|\boldsymbol{A})p(\boldsymbol{A}|\underbrace{\theta_x,\theta_y,\theta_z}_{\text{欧拉角}\boldsymbol{\theta}})p\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{r}^{\mathrm{t}}\right)\\&=f_{\boldsymbol{Y}}(\boldsymbol{Y},\boldsymbol{H},\gamma)\prod_{m,n}f_{h_{mn}}\left(h_{mn},\boldsymbol{r}_n^{\mathrm{t}}\right)\\&\quad\cdot\prod_n f_{\boldsymbol{r}_n}(\boldsymbol{r}_n^{\mathrm{t}},\boldsymbol{b}_n,\boldsymbol{r}^{\mathrm{t}})f_{\boldsymbol{b}_n}(\boldsymbol{b}_n,\boldsymbol{A})f_{\boldsymbol{A}}(\boldsymbol{A},\boldsymbol{\theta})\\&\quad\cdot\prod_{\kappa=\{x,y,z\}}f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_\kappa\right)f_{\boldsymbol{r}}\left(\kappa^{\mathrm{t}}\right)\end{aligned}\tag{16}$$

其中，γ是噪声精度，具有先验$p(\gamma)=f_\gamma\propto1/\gamma$；$f_{\boldsymbol{Y}}(\boldsymbol{Y},\boldsymbol{H},\gamma)=\mathrm{CN}(\boldsymbol{Y};\boldsymbol{\Phi}\boldsymbol{H},\gamma^{-1}\boldsymbol{I})$表示似然函数；$f_{h_{mn}}(h_{mn},\boldsymbol{r}_n^{\mathrm{t}})=\delta(h_{mn}-\boldsymbol{\phi}_{mn}\exp(\mathrm{i}k_0r_{mn}))$表示信道$h_{mn}$和坐标$\boldsymbol{r}_n^{\mathrm{t}}$间的参数关系。根据式(14)可知偏置向量$\boldsymbol{b}_n$、平行条件下偏置向量$\bar{\boldsymbol{b}}_n$和旋转映射矩阵$\boldsymbol{A}$之间存在映射关系$f_{\boldsymbol{b}_n}(\boldsymbol{b}_n,\boldsymbol{A})=\delta(\boldsymbol{b}_n-A\bar{\boldsymbol{b}}_n)$。根据式(15)可知$\boldsymbol{r}_n$与第1个贴片的坐标$\boldsymbol{r}^{\mathrm{t}}$之间存在偏置映射关系，可写为$f_{\boldsymbol{r}_n}(\boldsymbol{r}_n,\boldsymbol{b}_n,\boldsymbol{r}^{\mathrm{t}})=\delta(\boldsymbol{r}_n-(\boldsymbol{b}_n+\boldsymbol{r}^{\mathrm{t}}))$，归纳为矩阵形式为$f_{\boldsymbol{R}}(\boldsymbol{R},\boldsymbol{B},\boldsymbol{r}^{\mathrm{t}})=\delta(\boldsymbol{R}-(\boldsymbol{B}+\boldsymbol{r}^{\mathrm{t}}\otimes\boldsymbol{1}_{1\times N}))$。旋转矩阵$\boldsymbol{A}$与欧拉旋转角度$\boldsymbol{\theta}=(\theta_x,\theta_y,\theta_z)^{\mathrm{T}}$之间满足式(13)所示非线性关系。欧拉旋转角度$\theta_x$的先验服$-\pi/2\sim\pi/2$之间的均匀分布$f_{\boldsymbol{\theta}}(\theta_x)=\pi U[-0.5,0.5]$，用户坐标的先验$p(x^{\mathrm{t}})$也服从均匀分布$f_{\mathrm{r}}(x^{\mathrm{t}})=U[-0.5,0.5]$。根据消息传递理论，将式(16)所示因式分解表示为因子图，如图4所示。按照物理含义，本文将图4所示因子图划分为“线性混合”、“坐标-信道解耦”和“姿态解耦”3部分，包含函数$f_{\boldsymbol{B}}$, $f_{\boldsymbol{A}}$和对应的变量节点。表1给出了式(16)所示因式分解、对应的概率分布和函数关系。

4 基于消息传递的信道估计算法

基于前述因子图模型，可以推导对应的消息传递算法。消息传递算法以迭代方式执行，迭代完成

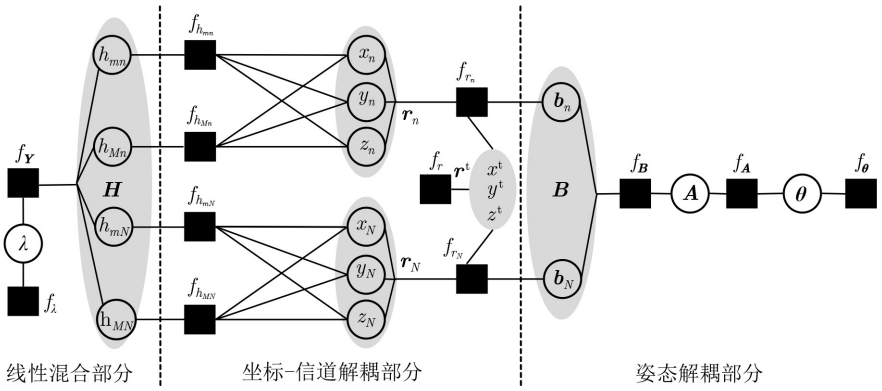


图 4 因式分解式(16)对应因子图

表 1 概率分布和对应函数表达式

概率分布和物理意义	函数表达式	概率分布和物理意义	函数表达式
似然函数$p(\boldsymbol{Y}\|\boldsymbol{H},\gamma)$	$f_{\boldsymbol{Y}}(\boldsymbol{Y},\boldsymbol{H},\gamma)=\mathrm{CN}(\boldsymbol{Y};\boldsymbol{\Phi H},\gamma^{-1}\boldsymbol{I})$	旋转参数化$f_{\boldsymbol{A}}(\boldsymbol{A},\boldsymbol{\theta})$	如式(7)所示
坐标映射$p(\boldsymbol{R}\|\boldsymbol{B},\boldsymbol{r}^{\mathrm{t}})$	$f_{\boldsymbol{R}}=\prod_n f_{\boldsymbol{r}_n}(\boldsymbol{b}_n,\boldsymbol{r}_n^{\mathrm{t}},\boldsymbol{r}^{\mathrm{t}})=\prod_n\delta(\boldsymbol{r}_n-\boldsymbol{b}_n^{\mathrm{t}}-\boldsymbol{r}^{\mathrm{t}})$	参数化信道$p(\boldsymbol{H}\|\boldsymbol{R})$	$f_{h_{mn}}(h_{mn},\boldsymbol{r}_n)=\delta(h_{mn}-\boldsymbol{\phi}_{mn}\exp(\mathrm{i}k_0r_{mn}))$
旋转$p(\boldsymbol{B}\|\boldsymbol{A}(\boldsymbol{\theta}))$	$f_{\boldsymbol{B}}=\prod_n f_{\boldsymbol{b}_n}(\boldsymbol{b}_n,\boldsymbol{A}(\boldsymbol{\theta}))=\prod_n\delta(\boldsymbol{b}_n-\boldsymbol{A}\bar{\boldsymbol{b}}_n)$	噪声先验$p(\gamma)$	$f_\gamma=1/\gamma$

后即可得到图4中每个变量的最大后验估计值，即置信。根据图4上的区域划分，本文也将消息计算划分为“线性混合”、“坐标-信道解耦”和“姿态解耦”3部分。

4.1 基于UAMP的线性混合部分消息计算

从因子图和表1可以看出，似然函数$p(\boldsymbol{Y}|\boldsymbol{H},\gamma)$是一个线性混合模型，可以使用UAMP进行处理[18,19]，分为如下5个步骤：

初始化 初始化噪声精度$\hat{\gamma}=1$、矩阵$\boldsymbol{S_H}=\boldsymbol{0}_{N\times M}$，$\hat{\boldsymbol{H}}=\boldsymbol{0}_{N\times M}$和$\boldsymbol{V_H}=\boldsymbol{1}_{N\times M}$；

步骤1 由$\boldsymbol{V_P}=|\boldsymbol{\Phi}|^2\boldsymbol{V_H}$和$\boldsymbol{P}=\boldsymbol{\Phi}\hat{\boldsymbol{H}}-\boldsymbol{V_P}\cdot\boldsymbol{S_H}$计算矩阵$\boldsymbol{V_P}$和$\boldsymbol{P}$，其中$\hat{\boldsymbol{H}}$和$\boldsymbol{V_H}$分别为$\boldsymbol{H}$的后验均值估计和方差，将在式(29)中给出更新方法；

步骤2 由$\boldsymbol{V_Z}=\boldsymbol{V_P}/(\hat{\gamma}\boldsymbol{V_P}+1)$和$\boldsymbol{Z}=(\hat{\gamma}\boldsymbol{Y}+\boldsymbol{P}/\boldsymbol{V_P})\cdot\boldsymbol{V_Z}$计算无噪声估计均值$\boldsymbol{Z}$和方差$\boldsymbol{V_Z}$；

步骤3 更新噪声精度$\hat{\gamma}=M\times L/\left(\|\boldsymbol{Y}-\boldsymbol{Z}\|^2+\|\boldsymbol{V_Z}\|\right)$；

步骤4 由$\boldsymbol{V_{S_H}}=1/(\boldsymbol{V_P}+\hat{\gamma}^{-1})$和$\boldsymbol{S_H}=\boldsymbol{V_{S_H}}\cdot(\boldsymbol{Y}-\boldsymbol{P})$更新Onsager修正矩阵$S_{\boldsymbol{H}}$和方差$\boldsymbol{V_{S_H}}$；

步骤5 由$\boldsymbol{V_{Q_H}}=1/(|\boldsymbol{\Phi}^{\mathrm{H}}|^2\boldsymbol{V}_{S_H})$和$\boldsymbol{Q_H}=\hat{\boldsymbol{H}}+\boldsymbol{V_{Q_H}}\cdot(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{S_H})$计算外信息矩阵$\boldsymbol{Q_H}$和$\boldsymbol{V_{Q_H}}$。将$\boldsymbol{Q_H}$和$\boldsymbol{V_{Q_H}}$的第$(n,m)$个元素分别记为$q_{mn}$和$\nu_{q_{mn}}$，表示消息$m_{h_{mn}\to f_{h_{mn}}}(h_{mn})$的均值和方差，即$m_{h_{mn}\to f_{h_{mn}}}(h_{mn})=\mathrm{CN}(h_{mn};q_{mn},\nu_{q_{mn}})$。

4.2 坐标-信道解耦部分消息计算

从因子图中可以看出，关于x,y,z 3个坐标的消息更新过程完全对称，以下将仅以x坐标为例并不失一般性。本节将消息计算又划分为神经网络消息穿透和坐标参数估计2部分。

步骤1 神经网络消息穿透计算。从式(9)所示神经网络表达式，可以看出函数节点$f_{h_{mn}}$仍然包含非线性运算，导致消息难以处理。本节利用泰勒展开对节点$f_{h_{mn}}$进行动态线性化，可得

$$\begin{aligned}h_{mn}(x_n^{\mathrm{t}},y_n^{\mathrm{t}},z_n^{\mathrm{t}})\approx&\underbrace{h_{mn}(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})}_{\hat{h}_{mn}}+\underbrace{h_{mn}^{x'}(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})}_{\hat{h}_{mn}^{x'}}\\&\cdot(x_n^{\mathrm{t}}-\hat{x}_n^{\mathrm{t}})+\underbrace{h_{mn}^{y'}(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})}_{\hat{h}_{mn}^{y'}}\\&\cdot(y_n^{\mathrm{t}}-\hat{y}_n^{\mathrm{t}})+\underbrace{h_{mn}^{z'}(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})}_{\hat{h}_{mn}^{z'}}\\&\cdot(z_n^{\mathrm{t}}-\hat{z}_n^{\mathrm{t}})\\=&\hat{h}_{mn}-(\hat{x}_n^{\mathrm{t}}\hat{h}_{mn}^{x'}+\hat{y}_n^{\mathrm{t}}\hat{h}_{mn}^{y'}+\hat{z}_n^{\mathrm{t}}\hat{h}_{mn}^{z'})\\&+x_n^{\mathrm{t}}\hat{h}_{mn}^{x'}+y_n^{\mathrm{t}}\hat{h}_{mn}^{y'}+z_n^{\mathrm{t}}\hat{h}_{mn}^{z'}\\=&\xi_{mn}+x_n^{\mathrm{t}}\hat{h}_{mn}^{x'}+y_n^{\mathrm{t}}\hat{h}_{mn}^{y'}+z_n^{\mathrm{t}}\hat{h}_{mn}^{z'}\end{aligned}\tag{17}$$

其中$h_{mn}^{x'}$表示h_{mn}对x_n^{t}的偏导$\nabla_{x_n}h_{mn}$在坐标$(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})$下的值，即$h_{mn}^{x'}=\nabla_{x_n^{\mathrm{t}}}h_{mn}(\hat{x}_n^{\mathrm{t}},\hat{y}_n^{\mathrm{t}},\hat{z}_n^{\mathrm{t}})$。类似地可得到$h_{mn}^{y'}$和$h_{mn}^{z'}$的定义。偏导$h_{mn}^{x'}$可由$\nabla_{x_n^{\mathrm{t}}}h_{mn}=\left(\nabla_{x_n^{\mathrm{t}}}\boldsymbol{\phi}_{mn}+\boldsymbol{\phi}_{mn}\mathrm{i}k_0\nabla_{x_n^{\mathrm{t}}}r_{mn}\right)\exp(\mathrm{i}k_0r_{mn})$计算，其中$\boldsymbol{\phi}_{mn}$的导数计算为

$$\begin{aligned}\nabla_{x_n^{\mathrm{t}}}\phi_{mn}=&((\boldsymbol{w}_{2,1}+\boldsymbol{w}_{2,2})\cdot\boldsymbol{w}_1^x)^{\mathrm{T}}\\&\times g_{a'}(x_n^{\mathrm{t}}\boldsymbol{w}_1^x+y_n^{\mathrm{t}}\boldsymbol{w}_1^y+z_n^{\mathrm{t}}\boldsymbol{w}_1^z+\boldsymbol{b}_1)\end{aligned}\tag{18}$$

其中$g_a(\cdot)$的导数为$g_{a'}(\cdot)=1-g_a^2(\cdot)$。根据式(17)，可以将$f_{h_{mn}}$表示为$f_{h_{mn}}=\delta\left(h_{mn}-\left(\xi_{mn}+x_n^{\mathrm{t}}\hat{h}_{mn}^{x'}+y_n^{\mathrm{t}}\hat{h}_{mn}^{y'}+z_n^{\mathrm{t}}\hat{h}_{mn}^{z'}\right)\right)$。假设后向消息$m_{y_n^{\mathrm{t}}\to f_{h_{mn}}}(y_n^{\mathrm{t}})$和$m_{z_n^{\mathrm{t}}\to f_{h_{mn}}}(z_n^{\mathrm{t}})$在上一次迭代中得到(类似消息

$m_{x_n^{\mathrm{t}}\to f_{h_{mn}}}(x_n^{\mathrm{t}})$，见式(26))，可以表示为$m_{y_n^{\mathrm{t}}\to f_{h_{mn}}}(y_n^{\mathrm{t}})=\mathcal{CN}(y_n^{\mathrm{t}};\overleftarrow{y}_{mn},\overleftarrow{v}_{y_{mn}})$, $m_{z_n^{\mathrm{t}}\to f_{h_{mn}}}(z_n^{\mathrm{t}})=\mathrm{CN}(z_n^{\mathrm{t}};\overleftarrow{z}_{mn},\overleftarrow{v}_{z_{mn}})$，则从$f_{h_{mn}}$到$x_n^{\mathrm{t}}$的消息可以计算为

$$\begin{aligned}m_{f_{h_{mn}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})=&\int f_{h_{mn}}m_{y_n^{\mathrm{t}}\to f_{h_{mn}}}(y_n^{\mathrm{t}})m_{z_n^{\mathrm{t}}\to f_{h_{mn}}}\\&(z_n^{\mathrm{t}})\times m_{h_{mn}\to f_{h_{mn}}}(h_{mn})\\&\cdot\mathrm{d}y_n^{\mathrm{t}}\mathrm{d}z_n^{\mathrm{t}}\mathrm{d}h_{mn}\\=&\mathrm{CN}(x_n^{\mathrm{t}};\vec{x}_{mn},\vec{\nu}_{x_{mn}})\end{aligned}\tag{19}$$

其中，变量上方的箭头表示该变量在消息传递计算过程中前向或后向方向，$\vec{x}_{mn}=\left(q_{mn}-\xi_{mn}-\overleftarrow{y}_{mn}\hat{h}_{mn}^{y'}-\overleftarrow{z}_{mn}\hat{h}_{mn}^{z'}\right)/\hat{h}_{mn}^{x'}$, $\vec{\nu}_{x_{mn}}=\left(\nu_{q_{mn}}+\overleftarrow{v}_{y_{mn}}|\hat{h}_{mn}^{y'}|^2+\overleftarrow{v}_{z_{mn}}|\hat{h}_{mn}^{z'}|^2\right)/|\hat{h}_{mn}^{x'}|^2$。

步骤2　坐标和信道联合估计。得到神经网络输出的消息$m_{f_{h_{mn}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})$后即可进行信道和坐标估计。首先，计算从$x_n^{\mathrm{t}}$到$f_{x_n^{\mathrm{t}}}$的前向消息

$$\begin{aligned}m_{x_n^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x_n^{\mathrm{t}})=&\prod\nolimits_m m_{f_{h_{mn}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})\\=&\mathrm{CN}(x_n^{\mathrm{t}},\vec{x}_n,\vec{\nu}_{x_n})\end{aligned}\tag{20}$$

其中，$\vec{\nu}_{x_n}=1/\left(\prod_m 1/\vec{\nu}_{x_{mn}}\right)$, $\vec{x}_n=\vec{\nu}_{x_n}\sum_m\vec{x}_{mn}/\vec{\nu}_{x_{mn}}$。假定第$n$个贴片偏移量为$\hat{b}_n^x$，则可以计算$f_{x_n^{\mathrm{t}}}$到$x^{\mathrm{t}}$的消息为

$$\begin{aligned}m_{f_{x_n^{\mathrm{t}}}\to x^{\mathrm{t}}}(x^{\mathrm{t}})=&\int f_{x_n^{\mathrm{t}}}m_{x_n^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x_n^{\mathrm{t}})\mathrm{d}x_n^{\mathrm{t}}\\=&\mathrm{CN}(x^{\mathrm{t}};\vec{x}_n^{\mathrm{t}},\vec{\nu}_{x_n^{\mathrm{t}}})\end{aligned}\tag{21}$$

其中，$\vec{x}_n^{\mathrm{t}}=\vec{x}_n-\hat{b}_n^x$, $\vec{\nu}_{x_n^{\mathrm{t}}}=\vec{\nu}_{x_n}$。利用消息传递规则可以计算$x^{\mathrm{t}}$坐标的置信(后验概率)为

$$\begin{aligned}b(x^{\mathrm{t}})=&\prod\nolimits_n m_{f_{x_n}\to x^{\mathrm{t}}}(x^{\mathrm{t}})/\int\prod\nolimits_n m_{f_{x_n}\to x^{\mathrm{t}}}(x^{\mathrm{t}})\mathrm{d}x^{\mathrm{t}}\\=&N(x^{\mathrm{t}},\hat{x}^{\mathrm{t}},\nu_{x^{\mathrm{t}}})\end{aligned}\tag{22}$$

其中，$\nu_{x^{\mathrm{t}}}=1/\left(\sum_n 1/\vec{\nu}_{x_n^{\mathrm{t}}}\right)$, $\hat{x}^{\mathrm{t}}=\nu_{x^{\mathrm{t}}}\sum_n\vec{x}_n^{\mathrm{t}}/\vec{\nu}_{x_n^{\mathrm{t}}}$。利用相同的方法可得$y^{\mathrm{t}}$和$z^{\mathrm{t}}$的后验分布$N(y^{\mathrm{t}},\hat{y}^{\mathrm{t}},\nu_{y^{\mathrm{t}}})$和$N(z^{\mathrm{t}},\hat{z}^{\mathrm{t}},\nu_{z^{\mathrm{t}}})$。通过置信传播，后向消息$m_{x^{\mathrm{t}}\to f_{x_n}}(x^{\mathrm{t}})$可计算为

$$m_{x^{\mathrm{t}}\to f_{x_n}}(x^{\mathrm{t}})=b(x^t)/m_{f_{x_n}\to x^{\mathrm{t}}}(x^{\mathrm{t}})=N(x^{\mathrm{t}};\overleftarrow{x}_n^{\mathrm{t}},\overleftarrow{\nu}_{x_n^{\mathrm{t}}})\tag{23}$$

其中均值和方差分别为$\overleftarrow{x}_n^{\mathrm{t}}=\overleftarrow{\nu}_{x_n^{\mathrm{t}}}(\hat{x}^{\mathrm{t}}/\nu_x^{\mathrm{t}}-\vec{x}_n^{\mathrm{t}}/\vec{\nu}_{x_n^{\mathrm{t}}})$和$\overleftarrow{\nu}_{x_n^{\mathrm{t}}}=1/(1/\nu_{x^{\mathrm{t}}}-1/\vec{\nu}_{x_n^{\mathrm{t}}})$。后向消息$m_{f_{x_n^{\mathrm{t}}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})$可更新为

$$\begin{aligned}m_{f_{x_n^{\mathrm{t}}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})=&\int f_{x_n^{\mathrm{t}}}m_{x^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x^{\mathrm{t}})dx^{\mathrm{t}}\\=&\mathcal{N}(x_n;\overleftarrow{x}_n,\overleftarrow{\nu}_{x_n})\end{aligned}\tag{24}$$

其中，$\overleftarrow{x}_n=\overleftarrow{x}_n^{\mathrm{t}}+\hat{b}_n^x$, $\overleftarrow{\nu}_{x_n}=\overleftarrow{\nu}_{x_n^{\mathrm{t}}}$。从而坐标$x_n$的置信可计算为

$$\begin{aligned}b(x_n^{\mathrm{t}})=&m_{f_{x_n^{\mathrm{t}}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})m_{x_n^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x_n^{\mathrm{t}})\\&/\int m_{f_{x_n^{\mathrm{t}}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})m_{x_n^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x_n^{\mathrm{t}})\mathrm{d}x_n^{\mathrm{t}}\\=&N(x_n^{\mathrm{t}};\hat{x}_n^{\mathrm{t}},\nu_{x_n^{\mathrm{t}}})\end{aligned}\tag{25}$$

其均值$\hat{x}_n^{\mathrm{t}}$和方差$\nu_{x_n^{\mathrm{t}}}$分别计算为$\nu_{x_n^{\mathrm{t}}}=1/\left(1/\overleftarrow{\nu}_{x_n}+1/\vec{\nu}_{x_n}\right)$和$\hat{x}_n^{\mathrm{t}}=\nu_{x_n^{\mathrm{t}}}\left(\overleftarrow{x}_n/\overleftarrow{\nu}_{x_n}+\vec{x}_n/\vec{\nu}_{x_n}\right)$。基于$b(x_n^{\mathrm{t}})$，可以计算消息

$$\begin{aligned}m_{x_n^{\mathrm{t}}\to f_{h_{mn}}}(x_n^{\mathrm{t}})=&b(x_n^{\mathrm{t}})/m_{f_{h_{mn}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}})\\=&N(x_n^{\mathrm{t}};\overleftarrow{x}_{mn},\overleftarrow{\nu}_{x_{mn}})\end{aligned}\tag{26}$$

其中均值方差可以计算为$\overleftarrow{\nu}_{x_{mn}}=1/\left(1/\nu_{x_n^{\mathrm{t}}}-1/\vec{\nu}_{x_{mn}}\right)$和$\overleftarrow{x}_{mn}=\overleftarrow{\nu}_{x_{mn}}\left(\hat{x}_n^{\mathrm{t}}/\nu_{x_n^{\mathrm{t}}}-\vec{x}_{mn}/\vec{\nu}_{x_{mn}}\right)$。最后更新函数节点$f_{h_{mn}}$至变量节点$h_{mn}$的消息

$$\begin{aligned}&m_{f_{h_{mn}}\to h_{mn}}(h_{mn})\\&=\int f_{h_{mn}}m_{x_n^{\mathrm{t}}\to f_{h_{mn}}}(x_n^{\mathrm{t}})\times m_{y_n^{\mathrm{t}}\to f_{h_{mn}}}(y_n^{\mathrm{t}})\\&\quad\cdot m_{z_n^{\mathrm{t}}\to f_{h_{mn}}}(z_n^{\mathrm{t}})\mathrm{d}x_n^{\mathrm{t}}\mathrm{d}y_n^{\mathrm{t}}\mathrm{d}z_n^{\mathrm{t}}\\&=N(h_{mn};\overleftarrow{h}_{mn},\overleftarrow{\nu}_{h_{mn}})\end{aligned}\tag{27}$$

其中$\overleftarrow{\nu}_{h_{mn}}=\xi_{mn}+\overleftarrow{x}_{mn}\hat{h}_{mn}^{x'}+\overleftarrow{y}_{mn}\hat{h}_{mn}^{y'}+\overleftarrow{z}_{mn}\hat{h}_{mn}^{z'}$和$\overleftarrow{h}_{mn}=\overleftarrow{\nu}_{x_{mn}}|\hat{h}_{mn}^{x'}|^2+\overleftarrow{\nu}_{y_{mn}}|\hat{h}_{mn}^{y'}|^2+\overleftarrow{\nu}_{z_{mn}}|\hat{h}_{mn}^{z'}|^2$。将前向消息$m_{h_{mn}\to f_{h_{mn}}}(h_{mn})$和后向消息$m_{f_{h_{mn}}\to h_{mn}}(h_{mn})$合并，可以得到信道$h_{mn}$的置信(即后验概率)

$$\begin{aligned}b(h_{mn})=&m_{f_{h_{mn}}\to h_{mn}}(h_{mn})m_{h_{mn}\to f_{h_{mn}}}(h_{mn})\\=&\mathrm{CN}(h_{mn};\hat{h}_{mn},\nu_{h_{mn}})\end{aligned}\tag{28}$$

其中，$\nu_{h_{mn}}=1/\left(1/\overleftarrow{\nu}_{h_{mn}}+1/\nu_{q_{mn}}\right)$, $\hat{h}_{mn}=\nu_{h_{mn}}\left(\overleftarrow{h}_{mn}/\overleftarrow{\nu}_{h_{mn}}+q_{mn}/\nu_{q_{mn}}\right)$。最后将$\nu_{h_{mn}}$和$\hat{h}_{mn},\forall m,n$堆叠为矩阵$\boldsymbol{V_H}$和$\hat{\boldsymbol{H}}$

$$\hat{\boldsymbol{H}}=\begin{bmatrix}\hat{h}_{11}&\hat{h}_{12}&\cdots&\hat{h}_{1N}\\\hat{h}_{21}&\hat{h}_{22}&\cdots&\hat{h}_{2N}\\\vdots&\vdots&\ddots&\vdots\\\hat{h}_{M1}&\hat{h}_{M2}&\cdots&\hat{h}_{MN}\end{bmatrix}\in\mathbb{C}^{M\times N},$$

$$\boldsymbol{V_H}=\begin{bmatrix}v_{h_{11}}&v_{h_{12}}&\cdots&v_{h_{1N}}\\v_{h_{21}}&v_{h_{22}}&\cdots&v_{h_{2N}}\\\vdots&\vdots&\ddots&\vdots\\v_{h_{M1}}&v_{h_{M2}}&\cdots&v_{h_{MN}}\end{bmatrix}\in\mathbb{R}^{M\times N}\tag{29}$$

从而完成参数和信道估计部分消息迭代闭环。

4.3 用户姿态解耦部分消息计算

从式(13)可知，坐标偏置矩阵$\boldsymbol{B}$与垂直状态下的偏置矩阵$\bar{\boldsymbol{B}}$之间存在非线性关系，利用消息传递难以得到闭式解。此外$\boldsymbol{B}$和$\boldsymbol{A}(\boldsymbol{\theta})$仅依赖于参数$\boldsymbol{\theta}$，由此本文选择将凸优化方法嵌入消息传递算法，实现全局迭代优化，具体步骤如下。首先利用式(22)和式(25)所求贴片坐标$\hat{x}_n^{\mathrm{t}}$和$\hat{x}^{\mathrm{t}}$可得x方向上的偏置

b_n^x的估计值$\hat{\boldsymbol{b}}_n^x=\hat{x}_n^{\mathrm{t}}-\hat{x}^{\mathrm{t}},\forall n$。同样可得$y$和$z$方向上的偏置$\hat{b}_n^y,\hat{b}_n^z,\forall n$。将偏置归纳为矩阵形式$\hat{\boldsymbol{B}}=(\hat{\boldsymbol{b}}_1,\hat{\boldsymbol{b}}_2,\cdots,\hat{\boldsymbol{b}}_N)\in\mathbb{R}^{3\times N}$，其中$\hat{\boldsymbol{b}}_n=(\hat{b}_n^x,\hat{b}_n^y,\hat{b}_n^z)^{\mathrm{T}}$。利用$\hat{\boldsymbol{B}}=\boldsymbol{A}(\boldsymbol{\theta})\bar{\boldsymbol{B}}$和式(13)所示的$\boldsymbol{A}(\boldsymbol{\theta})$表达式，带入优化表达式$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{B}}-\boldsymbol{A}(\boldsymbol{\theta})\bar{\boldsymbol{B}}\right)$，可得$\boldsymbol{A}(\boldsymbol{\theta})$的估计值。但前述优化表达式非凸，难以直接获得全局最优解。本文采用泰勒展开对$\boldsymbol{A}(\boldsymbol{\theta})$进行动态线性化，得到迭代优化表达式为

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{\mathrm{iter}+1}=\arg\min_{\boldsymbol{\theta}}\Big(&\hat{\boldsymbol{B}}^{\mathrm{iter}}-\Big(\boldsymbol{A}(\hat{\boldsymbol{\theta}}^{\mathrm{iter}})+(\theta_x-\hat{\theta}_x^{\mathrm{iter}})\\&\cdot\nabla_{\theta_x}\boldsymbol{A}(\hat{\boldsymbol{\theta}}^{\mathrm{iter}})+(\theta_y-\hat{\theta}_y^{\mathrm{iter}})\nabla_{\theta_y}\boldsymbol{A}(\hat{\boldsymbol{\theta}}^{\mathrm{iter}})\\&+(\theta_z-\hat{\theta}_z^{\mathrm{iter}})\nabla_{\theta_z}A(\hat{\boldsymbol{\theta}}^{\mathrm{iter}})\Big)\bar{\boldsymbol{B}}\Big)\end{aligned}\tag{30}$$

其中$\hat{\boldsymbol{\theta}}^{\mathrm{iter}}$和$\hat{\boldsymbol{B}}^{\mathrm{iter}}$分别为第iter次迭代所得$\boldsymbol{\theta}$和$\boldsymbol{B}$的估计值。将估计得到的$\hat{\boldsymbol{\theta}}$带入$\boldsymbol{A}(\hat{\boldsymbol{\theta}})$，即可得偏置矩阵$\hat{\boldsymbol{B}}=\boldsymbol{A}(\hat{\boldsymbol{\theta}})\bar{\boldsymbol{B}}$，进而得到用户平面每个贴片的偏置向量$\hat{\boldsymbol{b}}_n=\boldsymbol{A}(\hat{\boldsymbol{\theta}})\bar{\boldsymbol{b}}_n,\forall n$。

4.4 算法归纳、复杂度分析和算法实现细节

本节将前述消息更新过程归纳为算法1。

算法初始化时先假定用户天线垂直放置，其他均值设定全0，方差设定为1。算法1中第3～7行对应利用UAMP算法处理线性混合问题，复杂度取决于矩阵乘法，复杂度为$O(MNL)$；第8行对应于神经网络的消息穿透，需要计算如式(18)所示的矩阵计算，涉及$M\times N$个信道元素，每个元素需要执行维度分别为$3\times N_{\mathrm{h}}$和$N_{\mathrm{h}}\times 2$的矩阵乘操作，因此复杂度为$O(MNN_{\mathrm{h}}^2)$；第9～13行表示参数化信道-坐标解耦，每一步均为有限个乘加运算，复杂度可以写为$O(MN)$；第14行表示用户姿态解耦，仅涉及$3\times N$维的矩阵优化求解，且优化参数仅有$\{\theta_x,\theta_y,\theta_z\}$，复杂度较低，所以本文所提算法整体复杂度可以写为$O(MNN_{\mathrm{h}}^2+MNL)$。

5 数值仿真结果

本节从信道估计归一化均方误差(Normalized Mean Square Error, NMSE)和用户位置估计精度2个方面，分析不同算法在信噪比(Signal-to-Noise Ratio, SNR)、导频长度(L)及基站天线数(M)等参数变化下的性能表现，并引入克拉美罗下界(Cramer-Rao Lower Bound, CRLB)作为理论基准。仿真环境设置如下，载波频率设为$f=3$ GHz。在基站端，假设一个具有$M=10\times 10$个天线贴片的表面，用户端表面由$N=5\times 5$个贴片组成。基站和用户的贴片尺寸分别设为$\Delta_x^{\mathrm{r}}=\Delta_y^{\mathrm{r}}=0.5\lambda$和$\Delta_x^{\mathrm{t}}=\Delta_y^{\mathrm{t}}=0.1\lambda$，此时基站的天线孔径为$D=0.5$ m，对应的近场范围(瑞利距离)为$2D^2/\lambda=5$ m。仿真中使用QPSK调制，导频数量设定为$L=50$～600，神经网络隐藏节点数量$N_h=50$。基站第1个贴片的坐标为$(0,0,0)$，用户坐标随机生成，表示为$x^{\mathrm{t}},y^{\mathrm{t}}\sim U[-\lambda,\lambda]$，$z^{\mathrm{t}}\sim U[20\lambda,40\lambda]$。信道和用户坐标估计的$\mathrm{NMSE}_{\boldsymbol{H}}$和$\mathrm{NMSE}_{\mathrm{p}}$分别定义为

$$\begin{aligned}\mathrm{NMSE}_{\boldsymbol{H}}&=\mathbb{E}\left[\|\boldsymbol{H}-\hat{\boldsymbol{H}}\|^2/\|\boldsymbol{H}\|^2\right],\\\mathrm{NMSE}_{\mathrm{p}}&=\mathbb{E}\left[\left\|\left(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}}\right)^{\mathrm{T}}-\left(\hat{x}^{\mathrm{t}},\hat{y}^{\mathrm{t}},\hat{z}^{\mathrm{t}}\right)^{\mathrm{T}}\right\|^2\right.\\&\quad\left./\left\|\left(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}}\right)^{\mathrm{T}}\right\|^2\right]\end{aligned}\tag{31}$$

算法 1 基于消息传递的HMIMO信道估计算法

输入：观测$\boldsymbol{Y}$，导频$\boldsymbol{S}$；输出：用户坐标$\hat{\boldsymbol{r}}^{\mathrm{t}}=(\hat{x}^{\mathrm{t}},\hat{y}^{\mathrm{t}},\hat{z}^{\mathrm{t}})^{\mathrm{T}}$，信道$\hat{\boldsymbol{H}}$。

(1) 初始化噪声精度$\hat{\gamma}$、偏置矩阵$\hat{\boldsymbol{B}}=\bar{\boldsymbol{B}}$，$\boldsymbol{S_H}=\boldsymbol{0}_{N\times M}$，$\hat{\boldsymbol{H}}=\boldsymbol{0}_{N\times M}$和$\boldsymbol{V_H}=\boldsymbol{1}_{N\times M}$。

(2) For iter=1:T

(3) $\boldsymbol{V_P}=|\boldsymbol{\Phi}|^2\boldsymbol{V_H}$和$\boldsymbol{P}=\boldsymbol{\Phi}\hat{\boldsymbol{H}}-\boldsymbol{V_P}\cdot\boldsymbol{S_H}$，

(4) $\boldsymbol{V_Z}=\boldsymbol{V_P}\cdot/(\hat{\gamma}\boldsymbol{V_P}+1)$和$\boldsymbol{Z}=(\hat{\gamma}\boldsymbol{Y}+\boldsymbol{P}./\boldsymbol{V_P})\cdot\boldsymbol{V_Z}$，

(5) $\hat{\gamma}=M\times L/\left(\|\boldsymbol{Y}-\boldsymbol{Z}\|^2+\|\boldsymbol{V_Z}\|\right)$，

(6) $\boldsymbol{V_{S_H}}=1/(\boldsymbol{V_P}+\hat{\gamma}^{-1})$和$\boldsymbol{S_H}=\boldsymbol{V_{S_H}}\cdot(\boldsymbol{Y}-\boldsymbol{P})$，

(7) $\boldsymbol{V_{Q_H}}=1/(|\boldsymbol{\Phi}^{\mathrm{H}}|^2\boldsymbol{V_{S_H}})$和$\boldsymbol{Q_H}=\hat{\boldsymbol{H}}+\boldsymbol{V_{Q_H}}\cdot(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{S_H})$

(8) 利用式(19)计算$f_{h_{mn}}$到x_n^{t}消息$m_{f_{h_{mn}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}}),\forall m,n$，

(9) 利用式(20)和式(21)分别计算消息$m_{x_n^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x_n^{\mathrm{t}}),\forall n$和$m_{f_{x_n^{\mathrm{t}}}\to x^{\mathrm{t}}}(x^{\mathrm{t}}),\forall n$，

(10) 利用式(22)和式(23)分别计算x^{t}的置信$b(x^{\mathrm{t}})$和消息$m_{x^{\mathrm{t}}\to f_{x_n^{\mathrm{t}}}}(x^{\mathrm{t}})$，

(11) 利用式(24)和式(25)分别计算消息$m_{f_{x_n^{\mathrm{t}}}\to x_n^{\mathrm{t}}}(x_n^{\mathrm{t}}),\forall n$和置信$b(x_n^{\mathrm{t}}),\forall n$，

(12) 利用式(26)和式(27)分别计算消息$m_{x_n^{\mathrm{t}}\to f_{h_{mn}}}(x_n^{\mathrm{t}})$和$m_{f_{h_{mn}}\to h_{mn}}(h_{mn}),\forall m,n$，

(13) 利用式(28)计算h_{mn}的置信$b(h_{mn}),\forall m,n$，利用(29)堆叠$b(h_{mn})$，得到$\hat{\boldsymbol{H}}$和$\boldsymbol{V_H}$，

(14) 利用式(30)计算欧拉角$\hat{\boldsymbol{\theta}}^{\mathrm{iter}+1}$，更新$\boldsymbol{A}(\boldsymbol{\theta}),\hat{\boldsymbol{B}}=(\hat{\boldsymbol{b}}_1,\hat{\boldsymbol{b}}_2,\cdots,\hat{\boldsymbol{b}}_N)^{\mathrm{T}}$。

End For

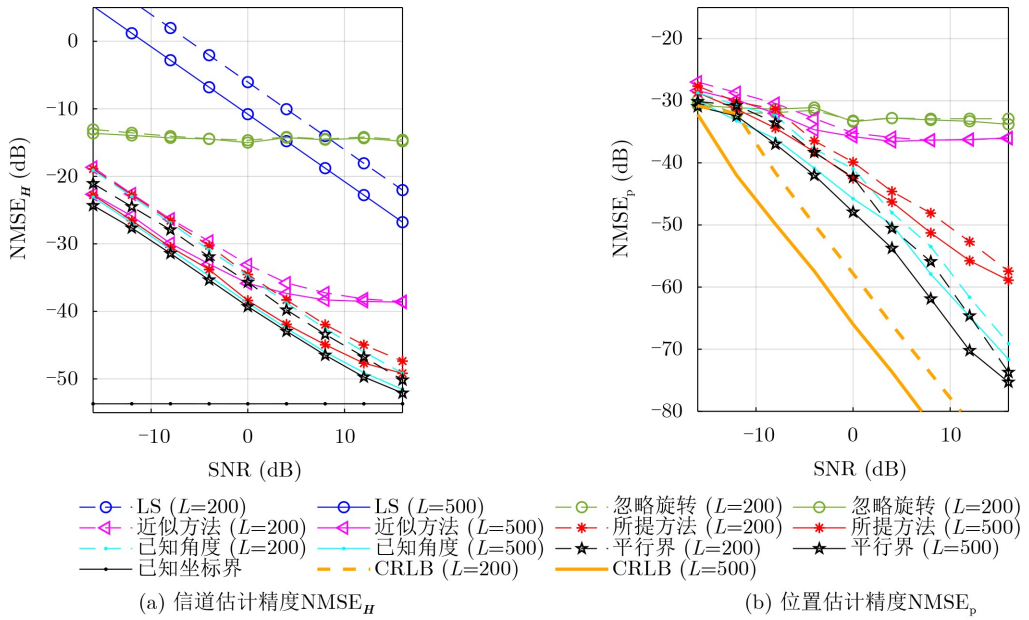
其中 $(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}})$ 和 $\boldsymbol{H}$ 表示用户坐标和信道矩阵，$(\hat{x}^{\mathrm{t}},\hat{y}^{\mathrm{t}},\hat{z}^{\mathrm{t}})$ 和 $\hat{\boldsymbol{H}}$ 表示对应估计值。本文选择的具体对比算法包括：“所提方法”表示本文所提参数化信道估计方法；“LS算法”为基于式(6)的最小二乘估计，表达式为 $\hat{\boldsymbol{H}}_{\mathrm{LS}}=(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{Y}$，但是该方法未利用信道与坐标间的隐含关系，无法提供用户位置的估计；“近似方法”表示利用近似模型(7)，相比式(3)的精确信道，存在一定的模型失配；“平行界”表示基站与用户平行并投影对齐[17]，此时式(13)中映射矩阵 $\boldsymbol{A}=\boldsymbol{I}_3$，作为性能参考界；“已知角度”表示假定用户旋转角度已知，即映射矩阵 $\boldsymbol{A}$ 已知；“忽略旋转”表示用户存在旋转，但信道估计时忽略该旋转，该方法会导致坐标估计误差，用于评估角度估计效果；“CRLB”表示用户坐标估计的理论性能下界[17]；“已知坐标”假设用户坐标 $(x^{\mathrm{t}},y^{\mathrm{t}},z^{\mathrm{t}})$ 已知，直接利用神经网络拟合信道，无需进行迭代计算，代表了信道估计的性能下界。

图5展示了 $\mathrm{NMSE}_{\boldsymbol{H}}$ 和 $\mathrm{NMSE}_{\mathrm{p}}$ 随SNR的变化曲线。从图5(a)可以看出，“平行界”设置发送和接收天线平面平行，避免了天线旋转问题，具有最优的信道估计性能，“已知角度”假设户角度已知，性能略有损失，其 $\mathrm{NMSE}_{\boldsymbol{H}}$ 在高SNR时渐进“已知坐标界”；“忽略旋转”方法未补偿角度引起的坐标偏移，性能损失严重，甚至劣于LS方法，说明天线角度显著影响信道估计精度；“近似方法”采用文献[4]中的近似信道模型，存在性能瓶颈；本文所提方法通过角度估计，补偿了旋转导致的坐标偏移，相比“平行界”有约1 dB性能损失，相比“近似方法”则有超过3 dB以上的性能增益。图5(b)为用户位置估计的 $\mathrm{NMSE}_{\mathrm{p}}$ 误差曲线，可以看出由于本文所提信道模型更好地表征真实信道，其性能显著优于“近似方法”，此外本文所提方法在高SNR下位置误差相较“平行界”差距约3 dB，验证了本文所提算法角度估计的准确性。

图6(a)和图6(b)分别展示了固定SNR下 $\mathrm{NMSE}_{\boldsymbol{H}}$ 和 $\mathrm{NMSE}_{\mathrm{p}}$ 随导频长度 L 的变化曲线。随着 L 的增加，几乎所有估计方法的性能都有所提高，只有“忽略旋转”方法由于严重的模型失配，使得其性能几乎无提升，“近似方法”在–38 dB附近出现性能瓶颈。本文所提方法与“已知角度”“平行界”性能接近，当 L 相对较大时，其性能趋近“已知坐标”。在图6(b)中，还可以观察到所提方法的性能与CRLB之间存在显著差距，这是因为当NMSE非常小(例如小于 -60 dB)时，所提方法在神经网络训练过程中存在的微弱模型失配，主导了误差性能，使得位置估计精度差距变大。

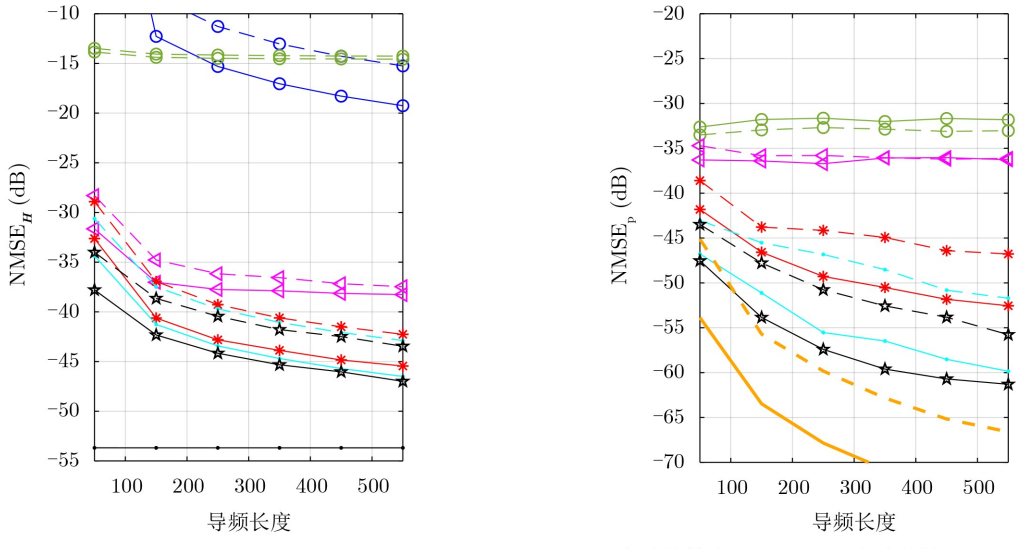
图7给出了固定信噪比为0和4 dB、导频长度 L=500时，$\mathrm{NMSE}_{\boldsymbol{H}}$ 和 $\mathrm{NMSE}_{\mathrm{p}}$ 随基站天线数 M 的变化曲线。从中可以看到，LS方法性能并没有改善，这是因为LS不是参数化估计器，需要估计的信道数量随 M 增加。除LS以外，随 M 的增加其余算法的性能均大幅度提升，直到出现性能瓶颈。此外“近似方法”的 $\mathrm{NMSE}_{\boldsymbol{H}}$ 和 $\mathrm{NMSE}_{\mathrm{p}}$ 在天线数超过100后出现一定的“上翘”，揭示了大天线阵列下模型失配问题的加剧。相比其他方法，本文所提方法参数化建模精度高，性能最接近“平行界”。

图8给出了 $\mathrm{NMSE}_{\boldsymbol{H}}$ 随基站天线数 M 和用户天线数 N 的性能变化矩阵，仿真中固定导频个数 $L=500$,



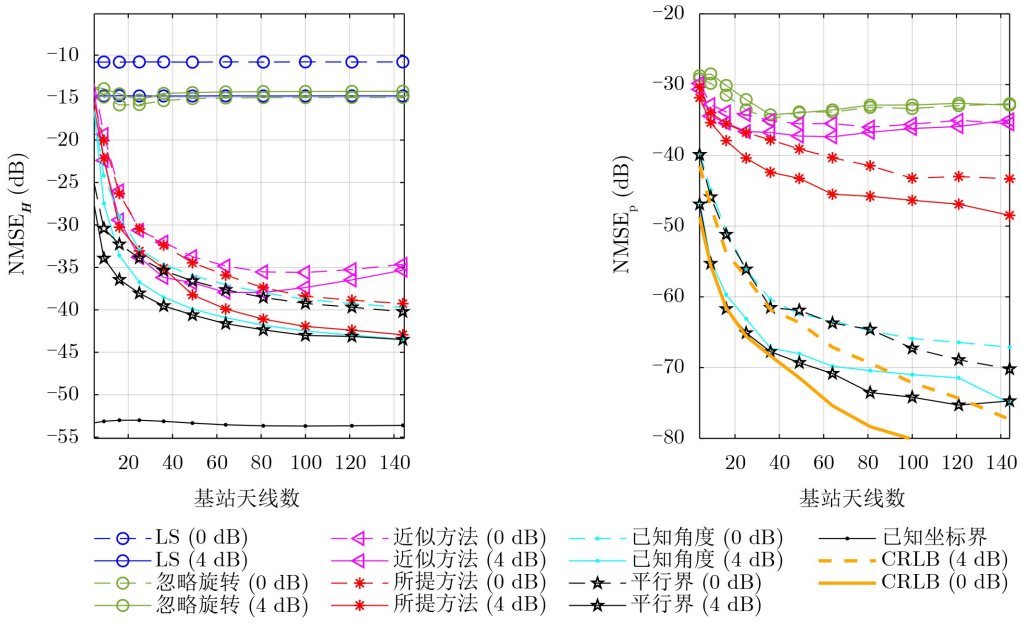
(a) 信道估计精度 $\mathrm{NMSE}_{\boldsymbol{H}}$　　(b) 位置估计精度 $\mathrm{NMSE}_{\mathrm{p}}$

图 5 不同对比算法的信道估计和坐标估计性能随信噪比变化曲线



(a) 信道估计精度NMSE$_H$ (b) 位置估计精度NMSE$_p$

图 6 信道估计和用户位置估计性能随导频长度L的变化曲线



(a) 信道估计精度NMSE$_H$ (b) 位置估计精度NMSE$_p$

图 7 信道估计和用户位置估计性能随基站天线数M变化曲线

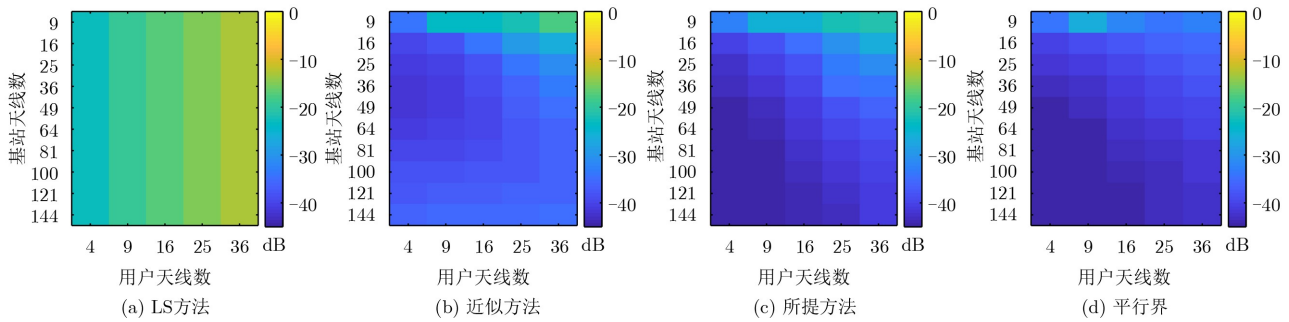


图 8 信道估计性能随用户天线数N和基站天线数M变化对比

SNR = 4 dB。整体上信道估计性能随M的增加和N的减少而提升。但当M >64, N >16后“近似方法”的NMSE$_{\boldsymbol{H}}$性能随M和N的增大均出现一定损失，这与图7结论一致。整体上所提方法显著优于“近似方法”和LS方法，凸显了在不同场景下的鲁棒性。

从以上仿真结果可以得出以下结论：首先，通过参数化信道建模将高维信道矩阵解耦为低维坐标与角度参数，降低了自由变量数量，提升了估计性能；其次，角度估计能够有效补偿用户旋转导致的坐标偏移，使得所提算法在任意角度下的性能与“平行界”假设下渐进，也验证了参数解耦的有效性；第三，为得到闭式解，“近似方法”对并矢格林函数进行了较多近似，存在性能瓶颈，凸显了精确信道建模的必要性。

6　结论

本文针对HMIMO近场信道估计中因复杂电磁环境和用户随机角度导致的参数耦合问题，提出了一种融合神经网络与贝叶斯方法的估计算法。通过神经网络简化信道模型、引入欧拉角旋转理论解耦角度参数，并基于因子图与消息传递算法实现高效联合估计，显著提升了信道与位置估计精度。仿真表明，所提方法在高信噪比下信道估计误差较现有近似方法降低3 dB以上，且在导频长度与天线规模扩展时仍保持鲁棒性。本文成果为HMIMO系统信道估计提供了理论依据，并为6G系统的实际部署提供了技术支撑。

参 考 文 献

[1] GONG Tierui, GAVRIILIDIS P, JI Ran, *et al*. Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions[J]. *IEEE Communications Surveys & Tutorials*, 2024, 26(1): 196–257. doi: 10.1109/COMST.2023.3309529.

[2] 陈晓明, 魏建川, 黄崇文. 面向6G无线网络的全息多输入多输出技术综述[J]. 电子与信息学报, 2024, 46(5): 1703–1715. doi: 10.11999/JEIT231140.
CHEN Xiaoming, WEI Jianchuan, and HUANG Chongwen. Overview of holographic multiple-input multiple-output technology for 6G wireless networks[J]. *Journal of Electronics & Information Technology*, 2024, 46(5): 1703–1715. doi: 10.11999/JEIT231140.

[3] ZHU Xiaolong, ZHANG Yanming, TANG Zhenzhou, *et al*. Coupling-induced capacity analysis of holographic MIMO communication system[C]. 2024 3rd International Conference on Electronics and Information Technology (EIT), Chengdu, China, 2024: 384–388. doi: 10.1109/EIT63098.2024.10762024.

[4] WEI Li, HUANG Chongwen, ALEXANDROPOULOS G C, *et al*. Tri-polarized holographic MIMO surfaces for near-field communications: Channel modeling and precoding design[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(12): 8828–8842. doi: 10.1109/TWC.2023.3266298.

[5] D'AMICO A A, TORRES A D J, SANGUINETTI L, *et al*. Cramér–Rao bounds for holographic positioning[J]. *IEEE Transactions on Signal Processing*, 2022, 70: 5518–5532. doi: 10.1109/TSP.2022.3222102.

[6] 黄崇文, 季然, 魏丽, 等. 面向全息MIMO 6G通信的电磁信道建模理论与方法[J]. 电子与信息学报, 2024, 46(5): 1940–1950. doi: 10.11999/JEIT231219.
HUANG Chongwen, JI Ran, WEI Li, *et al*. Electromagnetic channel modeling theory and approaches for holographic MIMO wireless communications[J]. *Journal of Electronics & Information Technology*, 2024, 46(5): 1940–1950. doi: 10.11999/JEIT231219.

[7] ZHANG Haobo, ZHANG Hongliang, DI Boya, *et al*. Holographic integrated sensing and communication[J]. *IEEE Journal on Selected Areas in Communications*, 2022, 40(7): 2114–2130. doi: 10.1109/JSAC.2022.3155548.

[8] DEMIR Ö T, BJÖRNSON E, and SANGUINETTI L. Channel modeling and channel estimation for holographic massive MIMO with planar arrays[J]. *IEEE Wireless Communications Letters*, 2022, 11(5): 997–1001. doi: 10.1109/LWC.2022.3152600.

[9] D'AMICO A A, BACCI G, and SANGUINETTI L. DFT-based channel estimation for holographic MIMO[C]. 2023 57th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, USA, 2023: 353–357. doi: 10.1109/IEEECONF59524.2023.10477065.

[10] CHEN Yuanbin, WANG Ying, WANG Zhaocheng, *et al*. Angular-distance based channel estimation for holographic MIMO[J]. *IEEE Journal on Selected Areas in Communications*, 2024, 42(6): 1684–1702. doi: 10.1109/JSAC.2024.3389116.

[11] GHERMEZCHESHMEH M and ZLATANOV N. Parametric channel estimation for LoS dominated holographic massive MIMO systems[J]. *IEEE Access*, 2023, 11: 44711–44724. doi: 10.1109/ACCESS.2023.3273602.

[12] YUAN S S A, HE Zi, CHEN Xiaoming, *et al*. Electromagnetic effective degree of freedom of an MIMO system in free space[J]. *IEEE Antennas and Wireless Propagation Letters*, 2022, 21(3): 446–450. doi: 10.1109/LAWP.2021.3135018.

[13] GONG Tierui, WEI Li, HUANG Chongwen, *et al*. Holographic MIMO communications with arbitrary surface placements: Near-field LoS channel model and capacity limit[J]. *IEEE Journal on Selected Areas in*

Communications, 2024, 42(6): 1549–1566. doi: 10.1109/JSAC.2024.3389126.

[14] CHEN Ang, CHEN Li, CHEN Yunfei, *et al*. Near-field positioning and attitude sensing based on electromagnetic propagation modeling[J]. *IEEE Journal on Selected Areas in Communications*, 2024, 42(9): 2179–2195. doi: 10.1109/JSAC.2024.3413981.

[15] GUO Shuaishuai, YE Jia, QU Kaiqian, *et al*. Green holographic MIMO communications with a few transmit radio frequency chains[J]. *IEEE Transactions on Green Communications and Networking*, 2024, 8(1): 90–102. doi: 10.1109/TGCN.2023.3318986.

[16] AN Jiancheng, YUEN C, HUANG Chongwen, *et al*. A tutorial on holographic MIMO communications-Part I: Channel modeling and channel estimation[J]. *IEEE Communications Letters*, 2023, 27(7): 1664–1668. doi: 10.1109/LCOMM.2023.3278683.

[17] YUAN Zhengdao, GUO Yabo, GAO Dawei, *et al*. Neural network-assisted hybrid model based message passing for parametric holographic MIMO near field channel estimation[J]. *IEEE Transactions on Wireless Communications*, 2025, 24(7): 6211–6224. doi: 10.1109/TWC.2025.3552492.

[18] GUO Qinghua and XI Jiangtao. Approximate message passing with unitary transformation[EB/OL]. https://arxiv.org/abs/1504.04799, 2015.

[19] YUAN Zhengdao, GUO Qinghua, and LUO Man. Approximate message passing with unitary transformation for robust bilinear recovery[J]. *IEEE Transactions on Signal Processing*, 2021, 69: 617–630. doi: 10.1109/TSP.2020.3044847.

袁正道：男，博士，副教授，研究方向为大规模MIMO、稀疏信道估计、消息传递算法和优化理论等.

郭亚博：男，博士，讲师，研究方向为智能超表面信道建模、信道估计、消息传递估计算法等.

高大伟：男，博士，讲师，研究方向为机器学习、阵列信号处理、通信感知一体化等.

郭庆华：男，博士，副教授，研究方向为雷达、通信等领域的机器学习和信号处理算法.

黄崇文：男，博士，研究员，研究方向为HMIMO、智能超表面、太赫兹通信、深度学习方法.

廖桂生：男，博士，教授，研究方向为自适应信号处理、阵列信号处理、信号检测与估计、智能天线等.

责任编辑：余 蓉

Parametric Holographic MIMO Channel Modeling and Its Bayesian Estimation

YUAN Zhengdao[①] GUO Yabo[②] GAO Dawei[③] GUO Qinghua[④]
HUANG Chongwen[⑤] LIAO Guisheng[③]

[①](*School of Artificial Intelligence , Henan Open University, Zhengzhou* 450000, *China*)
[②](*School of Information Science, Henan University of Technology, Zhengzhou* 450000, *China*)
[③](*Hangzhou Institute of Technology, Xidian University, Hangzhou* 311200, *China*)
[④](*University of Wollongong, Wollongong, NSW* 2522, *Australia*)
[⑤](*College of Information science Electronic Engineering, Zhejiang University, Hangzhou* 310027, *China*)

Abstract:

Objective Holographic Multiple-Input Multiple-Output (HMIMO), based on continuous-aperture antennas and programmable metasurfaces, is regarded as a cornerstone of 6G wireless communication. Its potential to overcome the limitations of conventional massive MIMO is critically dependent on accurate channel modeling and estimation. Three major challenges remain: (1) oversimplified electromagnetic propagation models, such as far-field approximations, cause severe mismatches in near-field scenarios; (2) statistical models fail to characterize the coupling between channel coefficients, user positions, and random orientations; and (3) the high dimensionality of parameter spaces results in prohibitive computational complexity. To address these challenges, a hybrid parametric-Bayesian framework is proposed in which neural networks, factor graphs, and convex optimization are integrated. Precise channel estimation, user position sensing, and angle decoupling in near-field HMIMO systems are thereby achieved. The methodology provides a pathway toward high-capacity 6G applications, including Integrated Sensing And Communication (ISAC).

Methods A hybrid channel estimation method is proposed to decouple the “channel-coordinate-angle”

parameters and to enable joint estimation of channel coefficients, coordinates, and angles under random user orientations. A neural network is first employed to capture the nonlinear relationship between holographic channel characteristics and the relative coordinates of the base station and user. The trained network is then embedded into a factor graph, where global optimization is performed. The neural network is dynamically approximated through Taylor expansion, allowing bidirectional message propagation and iterative refinement of parameter estimates. To address random user orientations, Euler angle rotation theory is introduced. Finally, convex optimization is applied to estimate the rotation mapping matrix, resulting in the decoupling of coordinate and angle parameters and accurate channel estimation.

Results and Discussions The simulations evaluate the performance of different algorithms under varying key parameters, including Signal-to-Noise Ratio (SNR), pilot length L, and base station antenna number M. Two performance metrics are considered: Normalized Mean Square Error (NMSE) of channel estimation and user positioning accuracy, with the Cramér-Rao Lower Bound (CRLB) serving as the theoretical benchmark. At an SNR of 10 dB, the proposed method achieves a channel NMSE below –40 dB, outperforming Least Squares (LS) estimation and approximate model-based approaches. Under high SNR conditions, the NMSE converges toward the CRLB, confirming near-optimal performance (Fig. 5a). The proposed channel model demonstrates superior performance over "approximate methods" due to its enhanced characterization of real-world channels. Moreover, the positioning error gap between the proposed method and the "parallel bound" narrows to nearly 3 dB at high SNR, confirming the accuracy of angle estimation and the effectiveness of parameter decoupling (Fig. 5b). Moreover, the proposed method maintains performance close to the theoretical bounds when system parameters, such as user antenna number N, base station antenna number M, and pilot length L, are varied, demonstrating strong robustness (Figs. 6～8). These results also show that the Euler angle rotation-based estimation effectively compensates for coordinate offsets induced by random user orientations.

Conclusions This study proposes a framework for HMIMO channel estimation by integrating neural networks, factor graphs, and convex optimization. The main contributions are threefold. First, Euler angles and coordinate mapping are incorporated into the parameterized channel model through factorization and factor graphs, enabling channel modeling under arbitrary user antenna orientations. Second, neural networks and convex optimization are embedded as factor nodes in the graph, allowing nonlinear function approximation and global optimization. Third, bidirectional message passing between neural network and convex optimization nodes is realized through Taylor expansion, thereby achieving joint decoupling and estimation of channel parameters, coordinates, and angles. Simulation results confirm that the proposed framework achieves higher accuracy—exceeding benchmarks by more than 3 dB, and demonstrates strong robustness across a range of scenarios. Future work will extend the method to multi-user environments, incorporate polarization diversity, and address hardware impairments such as phase noise, with the aim of supporting practical deployment in 6G systems.

Key words: Holographic Multiple-Input Multiple-Output (HMIMO); Neural networks; Bayesian estimation; Euler angle rotation; Message-passing algorithm.