基于深度 ADMM 展开的毫米波 MIMO 系统快速直接定位

Wenzhe Fan，IEEE 研究生会员；Shengheng Liu，IEEE 高级会员；Chunguo Li，IEEE 高级会员；以及 Yongming Huang，IEEE 高级会员

arXiv:2302.02557v1 [eess.SP] 6 Feb 2023

摘要—部署于毫米波系统中的大规模阵列可实现高角分辨率性能，进而促进亚米级定位服务。尽管性能并非最优，迄今最常用的定位方法仍基于一种所谓的两步法：先汇总协作基站的到达角（AoA）测量值，再进行三角测量。这主要是因为现有直接定位方法在大规模系统中的计算成本高得令人难以承受。为解决这一问题，我们提出一种基于深度展开的快速直接定位求解器。首先，将直接定位表述为联合 l_1-$l_{2,1}$范数稀疏恢复问题，并使用交替方向乘子法（ADMM）求解。接下来，我们开发深度 ADMM 展开网络（DAUN），从训练数据中学习 ADMM 参数设置，并为 DAUN 提出一种位置细化算法。最后，仿真结果表明，与基线求解器相比，所提出的 DAUN 在定位精度、收敛速度和计算复杂度方面均表现出显著优势。

索引词——直接定位、毫米波（mmWave）、多输入多输出（MIMO）、交替方向乘子法（ADMM）、深度展开。

I. 引言

MILLIMETER-WAVE（毫米波）多输入多输出（MIMO）已被公认为第五代/第六代蜂窝通信网络中应对数据速率和系统容量需求不断增长的前景技术，这得益于其高频谱效率和波束方向性 [1]。除了前所未有的传输容量外，研究还表明，毫米波 MIMO 系统有望实现亚米级用户定位及相关增值服务 [2], ，这归功于大规模阵列所提供的高角分辨率。

在经典定位系统中，用户定位通常包含两个步骤：首先估计位置相关参数，例如到达角和到达时间（ToA），然后通过三边测量或三角测量 [3]，根据这些参数恢复用户位置。在 [4], 中，作者提出了一种基于到达角的方法，用于大规模 MIMO 系统中的二维定位。针对三维定位，[5]中提出了一种闭式估计器。最近，[2], [6]中提出了几种适用于室内/室外毫米波 MIMO 系统的新型定位方法。然而，上述两步定位方法可能会因在参数估计阶段丢弃有用信息，导致性能不佳，因为估计参数的本质是接收信号的低阶矩 [7]。

作为一种效率更高的替代方案，文献 [8], 提出了直接定位方法，该方法无需估计中间参数，便可直接从接收数据中恢复用户位置。文献 [9], 提出了一种面向大规模 MIMO 系统的基于网格的直接定位方法，通过求解凸稀疏恢复问题来估计与到达角相关的用户位置。文献 [10], 在直接定位系统中考虑了用户移动，并开发了一种高效的变分贝叶斯求解器。尽管直接定位在提升定位性能方面潜力巨大，但当前直接定位方法（例如 [7], [9]–[12], ）面临的最紧迫问题是计算复杂度高；这一问题在毫米波 MIMO 系统中尤为突出，并阻碍了其实用化部署。

为了解决上述问题，我们提出了一种适用于毫米波 MIMO 系统的快速直接定位方法。首先，将直接定位表述为联合 l_1-$l_{2,1}$ 范数稀疏恢复问题。需要指出的是，现有的压缩感知方法均不适用于此类联合 l_1-$l_{2,1}$ 范数稀疏恢复问题，例如迭代收缩阈值算法（ISTA） [13], 、近似消息传递（AMP） [14]或稀疏贝叶斯学习（SBL） [15], 。我们的主要贡献如下：1）针对直接定位开发了一种专用的交替方向乘子法（ADMM）求解器，其计算复杂度极低；2）将 ADMM 求解器展开为基于模型的深度网络，以学习更优的 ADMM 迭代性能；3）设计了一种位置细化算法，进一步提高所提出的深度 ADMM 展开网络（DAUN）的精度。总体而言，所提出的方法首次尝试利用 ADMM 和展开架构解决此类 l_1-$l_{2,1}$ 范数直接定位问题。仿真结果表明，与其他基线方法相比，DAUN 定位精度更高、收敛速度更快，且计算复杂度大幅降低。

记号：小写（大写）粗体字符表示向量（矩阵），默认情况下，向量采用列方向。上标 $(\cdot)^{\mathsf{T}}$ and 和 $(\cdot)^{\mathsf{H}}$分别表示转置和共轭转置运算符。$\mathbb{E}$ 表示离散随机变量的期望值。$\|\cdot\|_1$、$\|\cdot\|_2$ 和 $\|\cdot\|_{2,1}$ 分别表示 l_1 范数、l_2 和 $l_{2,1}$ 范数。运算符 vec(X) 将矩阵 X 向量化，而 blkdiag($\mathbf{X}_1$,$\mathbf{X}_2$, . . . , $\mathbf{X}_M$) 表示块对角矩阵，其中 $\mathbf{X}_1$,$\mathbf{X}_2$, . . . , $\mathbf{X}_M$ 位于其对角线上。

本研究部分受到国家自然科学基金委员会资助，项目编号为 U1936201 和 62001103。（通讯作者：Yongming Huang 和 Shengheng Liu）作者就职于中国南京 210096 东南大学信息科学与工程学院，同时也就职于中国南京 211111 紫金山实验室（电子邮箱：{wzfan、s.liu、chunguoli、huangym}@seu.edu.cn）。

II. 预备知识

A. 信号模型

我们考虑一个二维目标区域 $\mathcal{R}$，其中有一个单天线用户，以及 M 个基站（BS），每个基站均配备由 N_m 根天线组成的阵列。阵列的重心位于 $\tilde{\mathbf{p}}_m = [\tilde{p}_m^x, \tilde{p}_m^y], m = 1, 2, \ldots, M$，其位置已精确知晓，并假设相对于用户位于远场。用户位置为 $\mathbf{p} = [p^x, p^y] \in \mathcal{R}$，并广播由所有基站发射的导频信号 s_{d}。第 m 个基站接收到的窄带信号表示为

$$\mathbf{y}_m = \sqrt{\omega}\mathbf{h}_m s_{\mathrm{d}} + \mathbf{n}_m, m = 1, 2, \ldots, M, \tag{1}$$

其中，ω 表示信噪比（SNR），$\mathbf{h}_m$ 表示用户与第 m个基站之间的毫米波信道向量，$\mathbf{n}_m \sim \mathcal{CN}(0, \mathrm{I}_{N_m})$表示功率归一化的接收噪声。通常，毫米波信道向量 $\mathbf{h}_m$ 由一条视距（LoS）路径和若干条非视距（NLoS）路径组成，可表示为 $\mathbf{h}_m = \alpha_m \mathbf{a}(\mathbf{p}) + \sum_{i=1}^{P_m} \alpha_m^i \mathbf{a}(\theta_m^i)$，其中$P_m$表示非视距路径的数量，$\mathbf{a}_m(\theta)$为阵列响应，$\alpha_m$ 和 $\theta_m(\mathbf{p})$分别表示视距路径的路径增益和到达角（AoA），$\{\alpha_m^i\}_{i=1}^{P_m}$ 和 $\{\theta_m^i\}_{i=1}^{P_m}$ 分别表示非视距路径的路径增益和到达角。视距路径的到达角与用户位置相关，其关系为 $\theta_m(\mathbf{p}) = \arctan((p^y - \tilde{p}_m^y)/(p^x - \tilde{p}_m^x))$。最后，所有基站的接收信号都会发送至融合中心进行联合处理[1]。

B. 问题表述

由于用户的到达角在角域中具有稀疏性，且位置是与到达角相关的参数，因此我们可以利用压缩感知方法，直接根据接收信号 $\mathbf{y}_m$估计 $\mathbf{p}$。其基本思路是将连续地图采样为网格，并寻找毫米波信道 $\mathbf{h}_m$的稀疏表示。这样，目标区域 $\mathcal{R}$ 可采样为 K 个均匀网格位置，表示为 $\mathcal{K} = \{\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_K\} \in \mathcal{R}$，其中 $\mathbf{\Phi}_k = [\phi_k^x, \phi_k^y]$是第 k个网格位置的坐标。然后，我们进一步定义一个包含 L_m 个角度的均匀网格，表示为 $\mathcal{L}_m = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_{L_m}\}$，其中 $L_m \gg P_m, m = 1, 2, \ldots, M$。如果上述区域网格被密集离散化，即 K 和 L_m 足够大，则接收信号 $\mathbf{y}_m$ 可表示为

$$\mathbf{y}_m = \mathbf{A}_m \mathbf{x}_m + \mathbf{B}_m \mathbf{z}_m + \mathbf{n}_m, m = 1, 2, \ldots, M, \tag{2}$$

其中，$\mathbf{A}_m = [\mathbf{a}_m(\theta_m(\mathbf{\Phi}_1)), \mathbf{a}_m(\theta_m(\mathbf{\Phi}_2)) \ldots, \mathbf{a}_m(\theta_m(\mathbf{\Phi}_K))]$,$\mathbf{B}_m = [\mathbf{a}_m(\vartheta_1), \mathbf{a}_m(\vartheta_2) \ldots, \mathbf{a}_m(\vartheta_{L_m})]$，$\theta_m(\mathbf{\Phi}_k)$ 表示从第 k 个网格到第 m 个基站的到达角，$\mathbf{x}_m = [x_{m,1}, x_{m,2}, \ldots, x_{m,K}]^T$ 是一个稀疏向量，其$x_{m,\,l}$对应于从网格位置 k 到第 m 个基站的视距路径的有效增益，并且 $\mathbf{x}_m$中只有一个非零元素。同时，$\mathbf{z}_m = [z_{m,1}, z_{m,2}, \ldots, z_{m,L_m}]^{\mathsf{T}}$ 也是一个稀疏向量，其$z_{m,\,l}$ 对应于用户到第 m 个基站、网格角度为 ϑ_l的非视距路径的有效增益。接下来定义矩阵 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]$。

[1]所提出的定位方案可扩展至自主位置感知系统。在此类系统中，基站（锚点）发送包含其物理位置和用于信道信息获取的导频序列的信号。用户设备融合所有接收信号，并在无需上行传输的情况下恢复自身位置。

由于 $\mathbf{x}_m$ 中只有一个元素非零，且该元素对应用户位置，因此矩阵 $\mathbf{X}$ 是行稀疏的。基于这一性质，直接定位问题可表述为

$$\begin{aligned} \min_{\hat{\mathbf{X}}, \{\hat{\mathbf{z}}_m\}_{m=1}^M} \quad & \left\|\hat{\mathbf{X}}\right\|_{2,1} + \sum_{m=1}^{M} w_m \|\hat{\mathbf{z}}_m\|_1 \\ \text{s.t.} \quad & \mathbf{y}_m = \mathbf{A}_m \hat{\mathbf{x}}_m + \mathbf{B}_m \hat{\mathbf{z}}_m, \forall m. \end{aligned} \tag{3}$$

其中，$\hat{\mathbf{X}}$ 和 $\hat{\mathbf{z}}_m$ 分别表示 $\mathbf{X}$ 和 $\mathbf{z}_m$的估计值，$\{w_m\}_{m=1}^M$ 是一个未知的加权系数。求解式（3）后，位置估计为 $\mathbf{\Phi}_{\hat{k}}$，其中索引 $\hat{k} = \arg\max_k \left\|[\hat{\mathbf{X}}]_{k,:}\right\|_2$是 $\hat{\mathbf{X}}$ 中的活跃行。

注 1：请注意，尽管问题（3）是凸问题，但高效凸优化求解器的复杂度相对于网格数量 K 和 L_m呈立方增长，因此对于毫米波MIMO系统中的实时定位而言，计算上难以实现。为解决这一问题，下一节将设计一种用于直接定位的深度交替方向乘子法展开框架。

III. 算法描述

A. 直接定位的交替方向乘子法求解器

本质上，直接定位问题（3）是一个联合 l_1-$l_{2,1}$ 范数优化问题，但没有专门的压缩感知方法来处理此类优化问题。根据问题（3）的形式，可以使用交替方向乘子法获得全局最优解。交替方向乘子法的第一步是给出问题（3）的增广拉格朗日函数，其形式为

$$\begin{aligned} J_\rho(\hat{\mathbf{X}}, \hat{\mathbf{z}}, \mathbf{s}) = & \left\|\hat{\mathbf{X}}\right\|_{2,1} + \sum_{m}^{M} w_m \|\hat{\mathbf{z}}_m\|_1 + \mathbf{s}^{\mathsf{H}}(\mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}} - \mathbf{y}) \\ & + \frac{\rho}{2} \|\mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}} - \mathbf{y}\|_2^2, \end{aligned} \tag{4}$$

其中 $\mathbf{A} = \text{块对角矩阵}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_M)$，$\hat{\mathbf{x}} = \text{向量化}(\hat{\mathbf{X}})$，$\mathbf{B} = \text{块对角矩阵}(\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_M)$，$\hat{\mathbf{z}} = [\hat{\mathbf{z}}_1^{\mathsf{T}}, \hat{\mathbf{z}}_2^{\mathsf{T}}, \ldots, \hat{\mathbf{z}}_M^{\mathsf{T}}]^{\mathsf{T}}$，$\mathbf{y} = [\mathbf{y}_1^{\mathsf{T}}, \mathbf{y}_2^{\mathsf{T}}, \ldots, \mathbf{y}_M^{\mathsf{T}}]^{\mathsf{T}}$，$\mathbf{s} \in \mathbb{C}^{\sum_{m=1}^M N_m \times 1}$是拉格朗日乘子，且 ρ 是一个正的惩罚参数。随后，交替方向乘子法重复以下计算

$$\hat{\mathbf{X}}^{(i+1)} = \arg\min_{\hat{\mathbf{X}}} J_\rho(\hat{\mathbf{X}}, \hat{\mathbf{z}}^{(i)}, \mathbf{s}^{(i)}), \tag{5}$$

$$\hat{\mathbf{z}}^{(i+1)} = \arg\min_{\hat{\mathbf{z}}} J_\rho(\hat{\mathbf{X}}^{(i+1)}, \hat{\mathbf{z}}, \mathbf{s}^{(i)}), \tag{6}$$

$$\mathbf{s}^{(i+1)} = \mathbf{s}^{(i)} + \rho(\mathbf{A}\hat{\mathbf{x}}^{(i+1)} + \mathbf{B}\hat{\mathbf{z}}^{(i+1)} - \mathbf{y}), \tag{7}$$

直至收敛。现在我们关注问题 (5)，最小化等价于

$$\min_{\hat{\mathbf{X}}} \left\|\hat{\mathbf{X}}\right\|_{2,1} + \frac{\rho}{2} \left\|\mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}}^{(i)} - \mathbf{y} + \frac{1}{\rho}\mathbf{s}^{(i)}\right\|_2^2. \tag{8}$$

然后，我们在 $\hat{\mathbf{x}}^{(i)}$ 处对目标函数 (8) 中的二阶项进行泰勒展开，并保留至二阶项，

$$\begin{aligned} \left\|\mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}}^{(i)} - \mathbf{y} + \tfrac{1}{\rho}\mathbf{s}^{(i)}\right\|_2^2 \approx & \left\|\mathbf{A}\hat{\mathbf{x}}^{(i)} + \mathbf{B}\hat{\mathbf{z}}^{(i)} - \mathbf{y} + \frac{1}{\rho}\mathbf{s}^{(i)}\right\|_2^2 \\ & + 2(\mathbf{g}_1^{(i)})^{\mathsf{H}}(\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(i)}) + \frac{1}{\tau_1}\left\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(i)}\right\|_2^2, \end{aligned} \tag{9}$$

其中 $\mathbf{g}_1^{(i)} = \mathbf{A}^{\mathsf{H}}(\mathbf{A}\hat{\mathbf{x}}^{(i)} + \mathbf{B}\hat{\mathbf{z}}^{(i)} - \mathbf{y} + \frac{1}{\rho}\mathbf{s}^{(i)})$，且 τ_1 为正的近端参数。将 (9) 代入 (8) 后，(8) 中的最小化问题变为

$$\min_{\hat{\mathbf{X}}} \left\|\hat{\mathbf{X}}\right\|_{2,1} + \rho(\mathbf{g}_1^{(i)})^{\mathsf{H}}(\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(i)}) + \frac{\rho}{2\tau_1}\left\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(i)}\right\|_2^2. \tag{10}$$

对 (10) 进行一些数学运算后，最小化问题可进一步表示为

$$\min_{\hat{\mathbf{X}}} \lambda_1 \left\|\hat{\mathbf{X}}\right\|_{2,1} + \frac{1}{2}\left\|\hat{\mathbf{x}} - \mathbf{c}^{(i)}\right\|_2^2, \tag{11}$$

其中 $\lambda_1 = \tau_1/\rho$ 和 $\mathbf{c}^{(i)} = \hat{\mathbf{x}}^{(i)} - \tau_1 \mathbf{g}_1^{(i)}$。

引理1：式（11）中最小化问题的闭式解可表示为

$$[\hat{\mathbf{X}}^{(i+1)}]_{k,:} = \begin{cases} \frac{\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 - \lambda_1}{\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2}[\mathbf{C}^{(i)}]_{k,:}, & \left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 > \lambda_1 \\ \mathbf{0}^T, & \left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 \le \lambda_1 \end{cases} \tag{12}$$

其中 $\mathbf{C}^{(i)} = [[\mathbf{c}^{(i)}]_{1:K}, [\mathbf{c}^{(i)}]_{K+1:2K}, \ldots, [\mathbf{c}^{(i)}]_{(M-1)K+1:MK}]$和 $k = 1, \ldots, K$。

证明：问题（11）可分解为关于 $\mathbf{X}$ 的行的 M 个相互独立的子问题，其形式为

$$\min_{[\hat{\mathbf{X}}]_{k,:}} \lambda_1 \left\|[\hat{\mathbf{X}}]_{k,:}\right\|_2 + \frac{1}{2}\left\|[\hat{\mathbf{X}}]_{k,:} - [\mathbf{C}^{(i)}]_{k,:}\right\|_2^2, k = 1, \ldots, K. \tag{13}$$

计算关于 $[\hat{\mathbf{X}}]_{k,:}$ 的导数并令其为零，可得 $[\hat{\mathbf{X}}]_{k,:} = 2a_k/(a_k - \lambda_1)[\mathbf{C}^{(i)}]_{k,:}$，其中 $a_k = \left\|[\hat{\mathbf{X}}]_{k,:}\right\|_2$。这意味着 $[\hat{\mathbf{X}}]_{k,:}$ 的解与 $[\mathbf{C}^{(i)}]_{k,:}$ 成线性关系，比例系数为标量 $\eta_k = 2a_k/(a_k - \lambda_1)$，即 $[\hat{\mathbf{X}}]_{k,:} = \eta_k[\mathbf{C}^{(i)}]_{k,:}$。这样，问题（13）变为

$$\min_{[\hat{\mathbf{X}}]_{k,:}} \frac{1}{2}\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2^2 \eta_k^2 + (\lambda_1\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 - \left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2^2)\eta_k + \frac{1}{2}\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2^2, \tag{14}$$

这是关于标量 η_k 的二次函数。通过计算上述二次函数的对称轴，可以立即得到 η_k 的闭式解，即当 $\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 - \lambda_1$ 时为 $\eta_k = (\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 - \lambda_1)/\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2$，而当 $\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 > \lambda_1$ 时为 $\eta_k = 0$；若 $\left\|[\mathbf{C}^{(i)}]_{k,:}\right\|_2 \le \lambda_1$，则为 $\eta_k = 0$。由此得到式（12）所给出的结果。■

接下来，我们聚焦于式（6）。通过进行与式（8）和式（9）类似的推导，式（6）中的最小化等价于

$$\min_{\hat{\mathbf{z}}} \sum_m^M w_m\|\hat{\mathbf{z}}_m\|_1 + \rho(\mathbf{g}_2^{(i)})^{\mathsf{H}}(\hat{\mathbf{z}} - \hat{\mathbf{z}}^{(i)}) + \frac{\rho}{2\tau_2}\left\|\hat{\mathbf{z}} - \hat{\mathbf{z}}^{(i)}\right\|_2^2, \tag{15}$$

其中，$\mathbf{g}_2^{(i)} = \mathbf{B}^{\mathsf{H}}(\mathbf{B}\hat{\mathbf{z}}^{(i)} + \mathbf{A}\hat{\mathbf{x}}^{(i+1)} - \mathbf{y} + \frac{1}{\rho}\mathbf{s}^{(i)})$，且 τ_2 为正的近端参数。经过一些数学变换，式（15）还可以进一步写成更紧凑的单测量向量问题形式，即

$$\min_{\hat{\mathbf{z}}} \lambda_2 \sum_m^M w_m\|\hat{\mathbf{z}}_m\|_1 + \frac{1}{2}\left\|\hat{\mathbf{z}} - \mathbf{d}^{(i)}\right\|_2^2, \tag{16}$$

其中，$\lambda_2 = \tau_2/\rho$，$\mathbf{d}^{(i)} = \hat{\mathbf{z}}^{(i)} - \tau_2\mathbf{g}_2^{(i)}$。式（16）的最小化具有闭式解 [13], [16], ，其形式如下：

$$[\hat{\mathbf{z}}_m^{(i+1)}]_l = \begin{cases} [\mathbf{d}^{(i)}]_{l_s} - \lambda_2 w_m, & [\mathbf{d}^{(i)}]_{l_s} > \lambda_2 w_m, \\ [\mathbf{d}^{(i)}]_{l_s} + \lambda_2 w_m, & [\mathbf{d}^{(i)}]_{l_s} < -\lambda_2 w_m, \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

其中 $l_s = \sum_{j=1}^{m-1} L_j + l, m = 1, 2, \ldots, M$,以及 $l = 1, 2, \ldots, L_m$。

B. 深度 ADMM 展开架构

理论上，给定适当的参数 $\rho, \{w_m\}_{m=1}^M, \tau_1$和 τ_2，ADMM 保证收敛到最优解 [17]。然而，上述参数的选择相当依赖启发式方法，在实际应用中需要进行繁琐的调试。另一方面，通过人工调试获得的参数仅基于非常有限的数据集。因此，对于由任意用户位置、信噪比、基站天线数量等因素产生的后续信号，ADMM 的迭代性能并不稳健。为克服这些困难，我们将式 (5) 至式 (7) 中的 ADMM 迭代展开为模型驱动的深度学习网络，从而能够通过 DAUN 学习未知参数。与黑盒数据驱动神经网络不同，DAUN 将 ADMM 迭代展开为一个可训练框架。因此，ADMM 更新的可解释性可直接延续至所提出的 DAUN。在 DAUN 中，参数 $\rho, \{w_m\}_{m=1}^M, \tau_1, \tau_2$ 在各次 ADMM 迭代之间不共享，因此在第 i 次迭代中，参数 $\rho^{(i)}, \{w_m\}_{m=1}^M \tau_1^{(i)}, \tau_{(i)2}$可训练。请注意，为了提高网络 [18] 的泛化能力和收敛速度，我们在每一层设置不同的参数 $\rho^{(i)}, \tau_{(i)1}, \tau_2^{(i)}$，而不是将它们设为固定值。 I层 DAUN 的信号流图如图 1 所示，其中图中的节点表示 ADMM 更新中的不同运算，有向边表示这些运算之间的数据流，可训练参数总数为 $3I+M$。

由于在低信噪比区域定位精度可能会急剧下降，为了缓解这种退化，我们提出在 DAUN 中引入信噪比加权损失函数，其形式为

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{x}}^{(i)}, \hat{\mathbf{z}}^{(i)}) = (1+\sigma(-\bar{\omega}))\left\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{(i)} + \mathbf{B}\hat{\mathbf{z}}^{(i)}\right\|_2^2, \tag{18}$$

其中，$\bar{\omega}$ 是与接收信号 $\mathbf{y}$ 相关的信噪比（单位：dB），而 $\sigma(x) = 1/(1+e^{-x})$ 表示 Sigmoid 函数。

在模型训练阶段，接收信号 $\mathbf{y}$ 独立生成，其中用户位置从目标区域 $\mathcal{R}$ 中随机抽取，信噪比则在感兴趣区域内服从均匀分布。我们采用小批量训练，并使用随机梯度下降算法训练 DAUN。为解决梯度消失问题，我们采用增量训练机制，按顺序从第一层到最后一层学习所有变量。训练完前 i 层后，将新的第 $(i+1)$ 层添加到 DAUN 中，并再次训练整个网络。变量 $\{\rho^{(t)}, \{w_m^{(i)}\}_{m=1}^M, \tau_1^{(t)}, \tau_{(t)2}\}_{t=1}^i$作为新一轮训练的初始值。正如 [18] 和我们的仿真结果所示，这一操作可有效降低训练复杂度。

C. 位置精化

从物理层面来看，DAUN 的定位精度取决于是否正确识别具有活动视距路径的基站。当某些基站不存在活动视距路径时，简单地选择 $\hat{k} = \arg\max_k \left\|[\hat{\mathbf{X}}]_{k,:}\right\|_2$ 可能会导致次优的位置估计。这是因为，与非活动视距路径对应的 $[\hat{\mathbf{X}}]_{k,:}$ 中的非零元素可能使 $\left\|[\hat{\mathbf{X}}]_{\hat{k},:}\right\|_2$

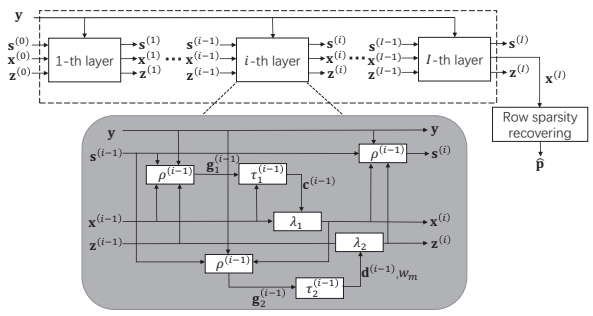

图 1. DAUN 的信号流图，其中包含可学习变量 $\rho^{(i)}$、$\{w_m\}_{m=1}^{M}$ $\tau_1^{(i)}$ $\tau_2^{(i)}$ 、和 。

成为最大值。例如，我们假设真实用户位置与 $[\mathbf{X}]_{k_1,:}$ 相关，即网格位置$\mathbf{\Phi}_{k_1}$应作为位置估计。通过使用 ADMM更新，我们得到估计值 $[\hat{\mathbf{X}}]_{k_1,:} = [1.6, 1.5, 1.7, 0.1]$，其中 $[\hat{\mathbf{X}}]_{k_1,:}$ 中的前三个元素是活动视距路径增益估计，最后一个元素 $[\hat{\mathbf{X}}]_{k_1,:}$ 可视为与非活动视距路径对应的噪声项。同时，对于 k_1 附近的另一个索引 k_2，我们可能得到估计值 $[\hat{\mathbf{X}}]_{k_2,:} = [1.55, 1.45, 1.7, 0.25]$。我们可以发现，通过执行 $\hat{k} = \arg\max_k \left\| [\hat{\mathbf{X}}]_{k,:} \right\|_2$，返回的将是错误索引 k_2，而不是 k_1。为避免此类不匹配，我们提出将 $[\hat{\mathbf{X}}]_{k,:}$ 的活动元素聚类为集合 $\mathcal{S}_{k,1}$，从而粗略地$\hat{}$选出可能具有活动视距路径的基站。随后，我们在与具有活动视距路径的基站相关联的集合 $\mathcal{F}$ 内得到位置估计 k。这样，$\hat{}$位置估计 k 的确定不会受到非活动视距分量的影响。上述位置精化过程的详细描述见算法1。

备注2：直接定位的性能在很大程度上取决于每个基站阵列的角分辨率，因此，为了获得更高的定位精度，网格数量 K 和天线数量 N_m 预计应尽可能大。为求解问题(3)，[9] 中 MOSEK求解器的复杂度为 $\mathcal{O}((KM + \sum_{m=1}^{M} L_m)^{3.5} \sum_{m=1}^{M} N_m)$，这对于实时实现而言计算开销很大，尤其是在采用毫米波大规模阵列时。此外，[8] 中的直接位置确定需要进行高维奇异值分解和穷举网格搜索，复杂度也很高。值得一提的是，所提出的交替方向乘子法是一种一阶原始-对偶算法，其中 (5)、(6) 和 (7) 的计算仅涉及线性运算，例如简单的矩阵乘法/加法。在交替方向乘子法的每次迭代中，复杂度仅为 $\mathcal{O}(((MK)^2 + (\sum_{m=1}^{M} L_m)^2) \sum_{m=1}^{M} N_m)$。此外，DAUN 以离线方式训练，位置精化的复杂度仅为 $\mathcal{O}(KM)$。因此，DAUN 的复杂度主要由 ADMM更新决定，而其复杂度相当低，这一点已在第四节中得到验证。

IV. 数值结果

在本节中，除非另有说明，所有实验均使用以下参数。我们考虑这样一种情况：坐标系原点位于区域中心，$M = 4$ 个基站分别位于 [-50m,-50m],[-50m, 50m], [50 m, 50m], [50m, -50m],。所有基站均配备均匀线性阵列（ULAs），且

算法1 位置精化

初始化：$\chi_{k,1}$ $= \max([\mathbf{X}]_{k,:})\, \chi_{k,2} = 0, \forall k$
当$\mathcal{S}_{k,1}$ 和 $\mathcal{S}_{k,2}$ 发生变化时
  $\mathcal{S}_{k,1} = \emptyset$ 和 $\mathcal{S}_{k,2} = \emptyset$
  当 $i = 1$ 到 M 时
    如果 $|[\hat{\mathbf{X}}]_{k,i} - \chi_{k,1}| > |[\hat{\mathbf{X}}]_{k,i} - \chi_{k,2}|$ 然后$\mathcal{S}_{k,1} = \mathcal{S}_{k,1} \cup \{i\}$
    else $\mathcal{S}_{k,2} = \mathcal{S}_{k,2} \cup \{i\}$
    结束 if
  结束 for 循环
  $\chi_{k,1} = \sum_{m \in \mathcal{S}_{k,1}} \frac{[\hat{\mathbf{X}}]_{k,m}}{|\mathcal{S}_{k,1}|} \chi_{k,2} = \sum_{m \in \mathcal{S}_{k,2}} \frac{[\hat{\mathbf{X}}]_{k,m}}{|\mathcal{S}_{k,2}|}$
结束 while 循环
找出这些索引 $\tilde{k}$ 使得 $|\mathcal{S}_{\tilde{k}}{}_{,1}| \geq 3$ 并收集它们到集合 $\mathcal{F}$ 中
$\hat{k} = \arg\max_{\tilde{k}} \left\| [\hat{\mathbf{X}}]_{\tilde{k},\, \mathcal{S}_{\tilde{k},1}} \right\|_2, \tilde{k} \in \mathcal{F}$

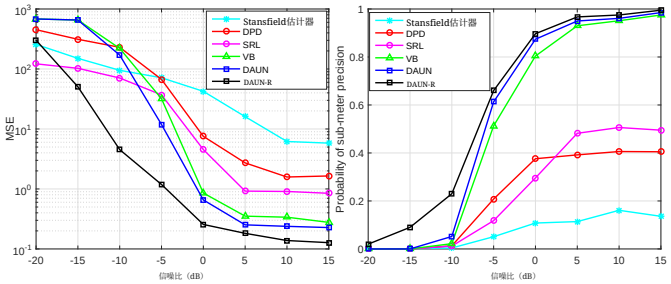

图 2. 不同定位方法的性能比较。(a) 位置估计的均方误差随信噪比变化的情况；(b) 达到亚米级精度的概率随信噪比变化的情况。

天线间距为半个波长。每个基站的天线数量为 $N_m = 50$ $m = 1$、2、3、4。用户位于大小为 40 ×40 m的目标区域 $\mathcal{L}$ 内。为获得最佳定位性能，网格位置和离散角度的数量分别为 $K = 30 \times 30 = 900$ 和 $L_m = 100 \forall m$。毫米波载波频率为30 GHz，信道根据3GPP城市宏蜂窝场景 [19] 生成。DAUN使用Python和PyTorch库 [20]实现。我们每增加一层，就使用700个样本训练网络，批量大小为7。采用Adam优化器，学习率设为0.05；当总层数大于5时，将学习率降至0.01。验证集包含200个样本，并且每次训练过程均在验证损失不再下降时终止。请注意，PyTorch只能处理实值深度学习问题。因此，在模型训练阶段，直接定位问题（3）被转换为其等价的实值形式。所有结果均通过运行1000次蒙特卡洛实验获得。

首先，图 2 展示了性能比较，其中将所提出的 DAUN 及其位置细化变体（DAUN-R）与直接位置确定（DPD） [8], Stansfield估计器 [21], 稀疏重建定位（SRL） [22], 和改进变分贝叶斯（VB） [15]进行了比较。除了采用均方误差（MSE）MSE $= \mathbb{E}[\|\hat{\mathbf{p}} - \mathbf{p}\|_2^2]$之外，还采用了亚米级精度的概率指标。这是因为对于许多应用场景，例如自动驾驶、物联网和救援行动，亚米级定位精度是必需的 [2], [7]。我们发现，与其他技术相比，DAUN-R 在较宽的信噪比范围内实现了更低的均方误差和更高的亚米级

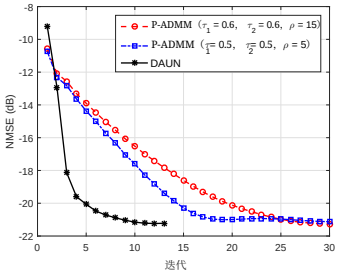

图 3. DAUN 和 P-ADMM 的归一化均方误差随迭代次数的变化。

表 I 不同直接定位求解器（DAUN、P-ADMM ($\tau_1 = 0.5, \tau_2 = 0.5, \rho = 5$)、MOSEK [9], 主化-最小化（MM） [11], 改进变分贝叶斯（VB）[15], 和 DPD [8]）的平均计算时间。

	DAUN	P-ADMM	MOSEK	MM	VB	DPD
时间（秒）	0.017	0.102	4.27	3.81	1.36	2.15

精度。随着信噪比的增加，DAUN 和 DAUN-R 达到亚米级精度的概率都趋近于 1。

此外，与纯交替方向乘子法（P-ADMM）相比，所提出的 DAUN 求解器具有更好的迭代性能。为展示这一优势，图 3 比较了 DAUN 和 P-ADMM 的收敛速度，其中归一化均方误差 NMSE $= \mathbb{E}[\|\mathbf{y} - \hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{B}}\mathbf{z}\|_2^2]/\mathbb{E}[\|\mathbf{y}\|_2^2]$ 随迭代次数变化，且通过多次实验选择 P-ADMM 的参数以保证收敛。图中显示，DAUN 和 P-ADMM 均随迭代单调收敛。我们观察到，DAUN 的收敛速度远快于 P-ADMM。为进一步说明 DAUN 的低复杂度，表 I 列出了不同定位求解器的平均计算时间，所有程序均运行于 Inter Core i7-8700K @4.2 GHz×12 核处理器上。结果表明，得益于低复杂度的交替方向乘子法和展开网络的精细参数调优，DAUN 的计算负担大幅降低。

V. 结论

本文提出了 DAUN，用于毫米波 MIMO 系统中的快速直接定位。首先，将直接定位表述为联合 l_1-$l_{2,1}$ 范数稀疏恢复问题。随后，我们开发了交替方向乘子法求解器来解决该问题。由于交替方向乘子法的收敛性严重依赖精细的参数调优，我们将深度展开融入交替方向乘子法迭代中，以通过网络训练的方式学习参数。最后，我们设计了一种位置细化算法，进一步提高 DAUN 的定位精度。据我们所知，从实现角度来看，所提出的 DAUN 是实现毫米波大规模 MIMO 系统直接定位速度最快的算法，具有重要意义。

参考文献

[1] M. Xiao、S. Mumtaz、Y. Huang、L. Dai、Y. Li、M. Matthaiou、G. K. Karagiannidis、E. Bj¨ornson、K. Yang、C.-L. I 和 A. Ghosh，“面向未来移动网络的毫米波通信”，IEEE J. Sel. Areas Commun.，第 35 卷，第 9 期，第 1909–1935 页，2017 年 9 月。[2] A. Shahmansoori、G. E. Garcia、G. Destino、G. Seco-Granados 和 H. Wymeersch，“5G 系统中基于毫米波 MIMO 的位置与姿态估计”，IEEE Trans. Wireless Co mmun.，第 17 卷，第 3 期，第 1822–1835 页，2017 年 3 月。[3] M. Pan、P. Liu、S. Liu、W. Qi、Y. Huang、X. You、X. Jia 和 X. Li，“利用 5G 微微蜂窝基站进行室内定位的高效 DOA 与 TOA 联合估计”，IEEE Trans. Instrum. Meas.，第 71 卷，文章编号 8005219，2022 年 8 月。[4] A. Hu、T. Lv、H. Gao、Z. Zhang 和 S. Yang，“一种基于 ESPRIT 的大规模 MIMO 系统中非相干分布源二维定位方法”，IEEE J. Sel. Topics Signal Process.，第 8 卷，第 5 期，第 996–1011 页，2014 年 10 月。[5] Y. Wang 和 K. Ho，“一种用于传感器网络三维 AOA 定位的渐近高效闭式估计器”，IEEE Trans. Wireless Commun.，第 14 卷，第 12 期，第 6524–6535 页，2015 年 12 月。[6] J. Yang、S. Jin、C.-K. Wen、J. Guo、M. Matthaiou 和 B. Gao，“用于毫米波通信中三维定位的基于模型的学习网络”，IEEE Trans. Wireless Commun.，第 20 卷，第 8 期，第 5449–5466 页，2021 年 8 月。[7] H. Zhao、N. Zhang 和 Y. Shen，“面向大规模天线阵列系统的波束空间直接定位”，IEEE Trans. Signal Process.，第 68 卷，第 3529–3544 页，2020 年 5 月。[8] A. J. Weiss，“窄带射频发射机的直接位置确定”，IEEE Signal Process. Lett.，第 11 卷，第 5 期，第 513–516 页，2004 年 5 月。[9] N. Garcia、H. Wymeersch、E. G. Larsson、A. M. Haimovich 和 M. Coulon，“大规模 MIMO 直接定位”，IEEE Trans. Signal Process.，第 65 卷，第 10 期，第 2475–2487 页，2017 年 5 月。[10] L. Lian、A. Liu 和 V. K. Lau，“基于动态变分贝叶斯推断的大规模 MIMO 系统用户位置跟踪”，IEEE Trans. Signal Process.，第 67 卷，第 21 期，第 5628–5642 页，2019 年 11 月。[11] X. Zheng、A. Liu 和 V. Lau，“存在相位噪声时大规模 MIMO 系统的联合信道与位置估计”，IEEE Trans. Signal Process.，第 68 卷，第 2598–2612 页，2020 年 4 月。[12] J. Li、P. Li、P. Li、L. Tang、X. Zhang 和 Q. Wu，“基于多源数据级联直接定位的自位置感知”，IEEE Trans. Intell. Transp. Syst.，提前出版，2022 年 5 月。[13] A. Chambolle、R. A. DeVore、N. Lee 和 B. J. Lucier，“非线性小波图像处理：变分问题、压缩及通过小波收缩去噪”，IEEE Trans. Image Process.，第 7 卷，第 3 期，第 319–335 页，1998 年 3 月。[14] S. Rangan 、 P. Schniter 、 A. K. Fletcher 和 S. Sarkar，“ 关于任意矩阵下近似消息传递的收敛性”，IEEE Trans. Inf. Theory.，第 65 卷，第 9 期，第 5339–5351 页，2019 年 9 月。[15] D. G. Tzikas、A. C. Likas 和 N. P. Galatsanos，“贝叶斯推断的变分近似”，IEEE Signal Process. Mag.，第 25 卷，第 6 期，第 131–146 页，2008 年 11 月。[16] D. L. Donoho，“通过软阈值处理去噪”，IEEE Trans. Inf. Theory，第 41 卷，第 3 期，第 613–627 页，1995 年 5 月。[17] H. Lu 、 X. Long 和 J. Lv，“ 一种基于交替方向方法的联合稀疏向量快速恢复算法”，载于第 14 届国际人工智能与统计会议论文集，2011 年，第 461–469 页。[18] A. B. Stimming 和 C. Studer，“ 通信系统的深度展开：综述及若干新方向 ”，载于 IEEE 信号处理系统国际研讨会（SiPS），IEEE，2019 年 10 月，第 266–271 页。[19] K. Haneda 等，“用于室外城市微蜂窝和宏蜂窝环境的 5G 3GPP 类信道模型”，载于 IEEE 第 83 届车辆技术会议（VTC Spring）论文集，2016 年，第 1–7 页。[20] A. Paszke 等，“PyTorch 中的自动微分”，载于第 31 届神经信息处理系统会议论文集，2017 年，第 1–4 页。[21] M. Gavish 和 A. J. Weiss，“仅方位目标定位算法的性能分析”，IEEE Trans. Aerosp. Electron. Syst.，第 28 卷，第 3 期，第 817–828 页，1992 年 7 月。[22] D. Malioutov、M. Cetin 和 A. S. Willsky，“从稀疏信号重构角度看利用传感器阵列进行源定位”，IEEE Trans. Signal Proces，第 53 卷，第 8 期，第 3010–3022 页，2005 年 8 月。