APPENDIX

Parameter	Value
Number of RF chain/ UPA	$N_{\text{RF}}/\quad S=4$
Number of BS antennas	$N=1024$
Number of antennas in each UPA	$N/S=256$
Carrier frequency	$f_c=300\,\text{GHz}$
Antenna spacing	$d=5.0\times10^{-4}\,\text{m}$
UPA spacing	$d_{\text{UPA}}=5.6\times10^{-2}\,\text{m}$
Pilot length	$P=128$
Azimuth AoA	$\theta_l\sim\mathcal{U}(-\pi/2,\pi/2)$
Elevation AoA	$\phi_l\sim\mathcal{U}(-\pi,\pi)$
Angle of incidence	$\varphi_{\text{in},l}\sim\mathcal{U}(0,\pi/2)$
Number of paths	$L=5$
Rayleigh distance	$D_{\text{Rayleigh}}=20\,\text{m}$
LoS path length	$r_1=30\,\text{m}$
Scatterer distance ($l>1$)	$r_l\sim\mathcal{U}(10,25)\,\text{m}$
Time delay of LoS path	$\tau_1=100\,\text{nsec}$
Time delay of NLoS paths ($l>1$)	$\tau_l\sim\mathcal{U}(100,110)\,\text{nsec}$
Absorption coefficient	$k_{\text{abs}}=0.0033\,\text{m}^{-1}$
Refractive index	$n_t=2.24-j0.025$
Roughness factor	$\sigma_{\text{rough}}=8.8\times10^{-5}\,\text{m}$

TABLE II: Parameter Setting

TRANSFORMATION DETAILS

Let $\boldsymbol{y}=\left[\Re(\bar{\boldsymbol{y}})^T,\Im(\bar{\boldsymbol{y}})^T\right]^T\in\mathbb{R}^{2N_{\text{RF}}P\times1}$, $\boldsymbol{h}=\left[\Re(\bar{\boldsymbol{h}})^T,\Im(\bar{\boldsymbol{h}})^T\right]^T\in\mathbb{R}^{2N\times1}$, $\boldsymbol{n}=\left[\Re(\bar{\boldsymbol{n}})^T,\Im(\bar{\boldsymbol{n}})^T\right]^T\in\mathbb{R}^{2N_{\text{RF}}P\times1}$, and

$$\mathbf{A}=\begin{pmatrix}\Re(\overline{\mathbf{A}}) & -\Im(\overline{\mathbf{A}})\\ \Im(\overline{\mathbf{A}}) & \Re(\overline{\mathbf{A}})\end{pmatrix}\in\mathbb{R}^{2N_{\text{RF}}P\times2N}.$$

Thus, equation (1) can be transformed into (2).

DERIVATION FOR THE DUAL

the problem (3) is equivalently expressed by introducing slackness variable $\boldsymbol{p},\boldsymbol{q}$:

$$\begin{aligned}\min_{\boldsymbol{p},\boldsymbol{q}}\quad & g(\boldsymbol{p})+f(\boldsymbol{q})\\ \text{s.t.}\quad & \boldsymbol{p}-\boldsymbol{q}=0.\end{aligned}$$

The Lagrangian $\mathcal{L}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{x})$ for this problem with the Lagrange multiplier $\boldsymbol{x}$ (also known as the dual variable) is given by:

$$\mathcal{L}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{x})=g(\boldsymbol{p})+f(\boldsymbol{q})+\boldsymbol{x}^\top(\boldsymbol{p}-\boldsymbol{q}).$$

Minimizing the Lagrangian over $\boldsymbol{p}$ and $\boldsymbol{q}$ gives:

$$\inf_{\boldsymbol{p},\boldsymbol{q}}\mathcal{L}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{x})=\inf_{\boldsymbol{p}}(g(\boldsymbol{p})+\boldsymbol{x}^\top\boldsymbol{p})+\inf_{\boldsymbol{q}}(f(\boldsymbol{q})-\boldsymbol{x}^\top\boldsymbol{q}).$$

The Fenchel conjugates f^* and g^* are defined as:

$$f^*(\boldsymbol{x})=\sup_{\boldsymbol{y}}(\boldsymbol{x}^\top\boldsymbol{y}-f(\boldsymbol{y})),\quad g^*(\boldsymbol{x})=\sup_{\boldsymbol{y}}(\boldsymbol{x}^\top\boldsymbol{y}-g(\boldsymbol{y})).$$

Thus, the dual problem is:

$$\max_{\boldsymbol{x}\in\mathbb{R}^{2N}}\left\{-g^*(-\boldsymbol{x})-f^*(\boldsymbol{x})\right\}.$$

DERIVATION OF ALGORITHM 1

We prove this proposition by induction. For $k=0$, we use the definition of $\boldsymbol{q}^1$ in Algorithm 1, yielding the condition $0\in\partial f(\boldsymbol{q}^1)-\boldsymbol{\eta}^0+\sigma\boldsymbol{q}^1$. Thus, $\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1\in\partial f(\boldsymbol{q}^1)$. Invoking Theorem 23.5 in [28], we deduce that

$$\boldsymbol{q}^1\in\partial f^*(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1).$$

Hence, $\sigma\boldsymbol{q}^1=\sigma\partial f^*(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1)$. This means that

$$\boldsymbol{\eta}^0\in\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1+\sigma\partial f^*(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1).$$

Denote that $\mathbf{M}_2=\partial f^*$, it follows that

$$\boldsymbol{\eta}^0\in\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1+\sigma\mathbf{M}_2(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1),$$

Let $\boldsymbol{w}^k\triangleq\mathbf{J}_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}^{k-1})$, which implies

$$\boldsymbol{w}^1=J_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}^0)=\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1.$$

In parallel, by the stipulation of $\boldsymbol{p}^1$, we observe that

$$0\in\partial g(\boldsymbol{p}^1)+(\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1).$$

It follows from Theorem 23.5 in [28] that

$$\boldsymbol{p}^1\in\partial g^*(-(\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1)).$$

Hence, $-\sigma\boldsymbol{p}^1=-\sigma\partial g^*(-(\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1)$, implying that
$2(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1)-\boldsymbol{\eta}^0\in2(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1)-\boldsymbol{\eta}^0+\sigma\boldsymbol{p}^1-\sigma\partial g^*(-(\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1))$.
Since $\mathbf{M}_1=\partial(g^*\circ(-\mathbf{I}))$, we have

$$2(\boldsymbol{\eta}^0-\sigma\boldsymbol{q}^1)-\boldsymbol{\eta}^0\in\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1+\sigma M_1(\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1),$$

which implies $\boldsymbol{x}^1:=\mathbf{J}_{\sigma\mathbf{M}_1}(2\boldsymbol{w}^1-\boldsymbol{\eta}^0)=\boldsymbol{\eta}^0-2\sigma\boldsymbol{q}^1+\sigma\boldsymbol{p}^1$. Hence, we have $\boldsymbol{\eta}^1:=\boldsymbol{\eta}^0+2(\boldsymbol{x}^1-\boldsymbol{w}^1)=\boldsymbol{\eta}^0+2\sigma(\boldsymbol{p}^1-\boldsymbol{q}^1)$.
Then, it follows that

$$\begin{aligned}\boldsymbol{\eta}^1&=\boldsymbol{\eta}^0+2(\boldsymbol{x}^1-\boldsymbol{w}^1)=\boldsymbol{\eta}^0+2(\mathbf{J}_{\sigma\mathbf{M}_1}(2\boldsymbol{w}^1-\boldsymbol{\eta}^0)-\mathbf{J}_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}^0))\\&=\boldsymbol{\eta}^0+2(\mathbf{J}_{\sigma\mathbf{M}_1}(2\mathbf{J}_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}^0))-\mathbf{J}_{\sigma\mathbf{M}_1}(-\boldsymbol{\eta}^0)-\mathbf{J}_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}^0))\\&=(2\mathbf{J}_{\sigma\mathbf{M}_1}-\mathbf{I})(2\mathbf{J}_{\sigma\mathbf{M}_2}-\mathbf{I})(\boldsymbol{\eta}^0)=\mathbf{R}_{\sigma\mathbf{M}_1}\mathbf{R}_{\sigma\mathbf{M}_2}\boldsymbol{\eta}^0.\end{aligned}$$

It follows that the update of $\boldsymbol{\eta}^1$ is the same as our Algorithm. Hence, we prove the statement for $k=0$. Assume that the statement holds for some $k\geq1$. For $k:=k+1$, we can prove that the statement holds similarly to the case $k=0$. Thus, we prove the statement holds for any $k\geq0$ by induction.

PROOF OF THEOREM 1

Since Fix $\left(\mathbf{T}_\sigma^{\text{PR}}\right)$ is a closed convex set [30, Corollary 4.24], the fixed point is unique. Using the convergence result of [31, Theorem 22], we can directly obtain that the sequence $\{\boldsymbol{\eta}^k\}$ converges to the fixed point $\boldsymbol{\eta}^*$ of the PR iteration. Since the resolvent $\boldsymbol{J}_{\sigma M_2}$ is nonexpansive and $\boldsymbol{w}^{k+1}=\boldsymbol{J}_{\sigma M_2}\left(\eta^k\right)$ as derived in the proof of Proposition 1, we directly obtain

$$\lim_k\boldsymbol{x}^k=\lim_k\boldsymbol{J}_{\sigma M_2}\left(\boldsymbol{\eta}^k\right)=\boldsymbol{J}_{\sigma M_2}\left(\boldsymbol{\eta}^*\right).$$

As show in [18], if $\boldsymbol{\eta}^*$ is the fixed point of the PR iteration, then $\boldsymbol{x}^*=\boldsymbol{J}_{\sigma M_2}(\boldsymbol{\eta}^*)$. Hence, we prove the convergence of sequence $\{\boldsymbol{x}^k\}$ to the optimal solution $\boldsymbol{x}^*$ of (4).

Next, we prove the convergence of the sequence $\{\boldsymbol{q}^k\}$. According to Algorithm 1, we have for all $k\geq0$,

$$\boldsymbol{q}^{k+1}=(\mathbf{A}^\top\mathbf{A}+\sigma\mathbf{I})^{-1}(\mathbf{A}^\top\boldsymbol{y}+\boldsymbol{\eta}^k).$$

Denote $\hat{f}(\boldsymbol{q}) := f(\boldsymbol{q}) + \frac{\sigma}{2}\|\boldsymbol{q}\|^2$, which is a strongly convex function. Thus, $\hat{f}^*$ is essentially smooth [28, Theorem 26.3]. The first-order optimality condition of the above iteration of $\boldsymbol{q}^{k+1}$ implies $0 \in \partial \hat{f}\left(\boldsymbol{q}^{k+1}\right) - \boldsymbol{\eta}^k$. Since $\hat{f}$ is a proper closed convex function, by [28, Theorem 23.5], the first-order optimality condition is equivalent to

$$\boldsymbol{q}^{k+1} = \nabla \hat{f}^*\left(\boldsymbol{\eta}^k\right).$$

It follows from the convergence of $\{\boldsymbol{\eta}^k\}$ and the continuity of $\nabla \hat{f}^*$ [28, Theorem 25.5] that $\{\boldsymbol{q}^k\}$ is convergent. Note that

$$\boldsymbol{\eta}^{k+1} = \boldsymbol{\eta}^k + 2\sigma(\boldsymbol{p}^{k+1} - \boldsymbol{q}^{k+1})$$

from Algorithm 1, and the sequence $\{\boldsymbol{\eta}^k\}$ is convergent, we obtain $\lim_k(\boldsymbol{p}^k - \boldsymbol{q}^k) = 0$. Then, by taking the limit, we obtain $\lim_k \boldsymbol{p}^k = \lim_k \boldsymbol{q}^k = \boldsymbol{q}^*$. Hence, $\{\boldsymbol{p}_k\}$ is convergent.

Assume that $(\boldsymbol{p}^*, \boldsymbol{q}^*, \boldsymbol{x}^*)$ is the limit point of the sequence $\{\boldsymbol{p}^k, \boldsymbol{q}^k, \boldsymbol{x}^k\}$. Since $\boldsymbol{\eta}^{k+1} = \boldsymbol{\eta}^k + 2\sigma(\boldsymbol{p}^{k+1} - \boldsymbol{q}^{k+1})$ from Algorithm 1, we can obtain $\boldsymbol{p}^* - \boldsymbol{q}^* = \boldsymbol{0}$ by taking the limit. It follows that

$$\lim_k \boldsymbol{w}^k = \lim_k [\boldsymbol{w}^k + \sigma(\boldsymbol{p}^k - \boldsymbol{q}^k)] = \lim_k \boldsymbol{x}^k = \boldsymbol{x}^*.$$

By Algorithm 1 and the derivation in Proposition 1, we have

$$0 \in \partial f(\boldsymbol{q}^{k+1}) - \boldsymbol{\eta}^k + \sigma \boldsymbol{q}^{k+1}.$$

Since $\boldsymbol{w}^{k+1} = \boldsymbol{\eta}^k - \sigma \boldsymbol{q}^{k+1}$, we have

$$\boldsymbol{w}^{k+1} \in \partial f\left(\boldsymbol{q}^{k+1}\right).$$

Similarly, by Algorithm 1 and the derivation details in Proposition 1, we have

$$0 \in \partial g(\boldsymbol{p}^{k+1}) + (\boldsymbol{\eta}^k - 2\sigma \boldsymbol{q}^{k+1} + \sigma \boldsymbol{p}^{k+1}).$$

Since $\boldsymbol{x}^{k+1} = \boldsymbol{\eta}^k - 2\sigma \boldsymbol{q}^{k+1} + \sigma \boldsymbol{p}^{k+1}$, we have

$$-\boldsymbol{x}^{k+1} \in \partial g(\boldsymbol{p}^{k+1}).$$

Hence, we have

$$\boldsymbol{w}^{k+1} \in \partial f\left(\boldsymbol{q}^{k+1}\right), \quad -\boldsymbol{x}^{k+1} \in \partial g\left(\boldsymbol{p}^{k+1}\right),$$

Together with $\boldsymbol{p}^* - \boldsymbol{q}^* = \boldsymbol{0}$ and taking limit, we have

$$\boldsymbol{x}^* \in \partial f\left(\boldsymbol{q}^*\right), \quad -\boldsymbol{x}^* \in \partial g\left(\boldsymbol{p}^*\right), \quad \boldsymbol{p}^* - \boldsymbol{q}^* = 0.$$

This completes the proof [28, Corollary 28.3.1].

Proof of Theorem 2

As indicated above, the iteration formula for $\boldsymbol{\eta}^k$ in the proposed PR-DEN can be succinctly represented as:

$$\boldsymbol{\eta}_R^{k+1} = f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ R_\theta \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k).$$

That is,

$$\begin{aligned}\boldsymbol{\eta}_R^{k+1} =& f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ (R_\theta - \mathrm{Prox}_{\sigma^{-1}g} + \mathrm{Prox}_{\sigma^{-1}g}) \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k)\\ =& f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ (R_\theta - \mathrm{Prox}_{\sigma^{-1}g}) \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k)\\ &+ f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ \mathrm{Prox}_{\sigma^{-1}g} \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k).\end{aligned}$$

From the proof of proposition 1, we obtain

$$\mathbf{T}_\sigma^{\mathbf{PR}}(\boldsymbol{\eta}_R^k) = f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ \mathrm{Prox}_{\sigma^{-1}g} \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k).$$

Thus, we derive

$$\boldsymbol{\eta}_R^{k+1} = f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ (R_\theta - \mathrm{Prox}_{\sigma^{-1}g}) \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R^k) + \mathbf{T}_\sigma^{\mathbf{PR}}(\boldsymbol{\eta}_R^k). \tag{8}$$

By the Universal approximation theorem [33], there exists a network $R_\theta(\cdot)$ approximating the proximal operator $\mathrm{Prox}_{\sigma^{-1}g}$ with arbitrary precision. Suppose $\max\{\|f_{\mathrm{LT}_1}\|, \|f_{\mathrm{LT}_2}\|, \|f_{\mathrm{LT}_3}\|\} = \alpha$, $\mathrm{Lip}(\mathbf{T}_\sigma^{\mathbf{PR}}) = \beta$. There exists R_θ, such that

$$\|(R_\theta - \mathrm{Prox}_{\sigma^{-1}g})(\boldsymbol{\eta}_R^k)\| < \frac{1-\beta}{\alpha^3}\|\boldsymbol{\eta}_R^k\|. \tag{9}$$

Thus, combined (9) with (8), we derive

$$\|\boldsymbol{\eta}_R^{k+1}\| < (1-\beta)\|\boldsymbol{\eta}_R^k\| + \beta\|\boldsymbol{\eta}_R^k\| = \|\boldsymbol{\eta}_R^k\|.$$

From Banach fix point theorem, $\{\boldsymbol{\eta}_R^k\}$ is convergent.

Proof of Theorem 3

From Theorem 2, we derive the convergence of $\{\boldsymbol{\eta}_R^k\}$, which denotes the sequence produced by PR-DEN. That is

$$\lim_{k\to\infty} \boldsymbol{\eta}_R^k = \boldsymbol{\eta}_R,$$

where $\boldsymbol{\eta}_R$ is the solution of the following fix point equation:

$$\boldsymbol{\eta}_R = f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ R_\theta \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R). \tag{10}$$

Additionally, since the dual problem (4) is convex, the optimal solution $\mathbf{x}_*$ is the equivalent to the solution of the following problem [18]:

$$\boldsymbol{x}_* = \mathbf{J}_{\sigma\mathbf{M_2}}(\boldsymbol{\eta}_*) \quad \text{s.t.} \quad \boldsymbol{\eta}_* = \mathbf{R}_{\sigma\mathbf{M}_1}\mathbf{R}_{\sigma\mathbf{M}_2}(\boldsymbol{\eta}_*). \tag{11}$$

Thus, subtracting (10) from (11), we obtain

$$\boldsymbol{\eta}_R - \boldsymbol{\eta}_* = \mathbf{T}_\sigma^{\mathbf{PR}}(\boldsymbol{\eta}_R - \boldsymbol{\eta}_*) + f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ (R_\theta - \mathrm{Prox}_{\sigma^{-1}g}) \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R).$$

Suppose $\max\{\|f_{\mathrm{LT}_1}\|, \|f_{\mathrm{LT}_2}\|, \|f_{\mathrm{LT}_3}\|\} = \alpha$, $\mathrm{Lip}(\mathbf{T}_\sigma^{\mathbf{PR}}) = \beta$, then

$$\|\boldsymbol{\eta}_R - \boldsymbol{\eta}_*\| \le \beta\|\boldsymbol{\eta}_R - \boldsymbol{\eta}_*\| + \alpha^3\|(R_\theta - \mathrm{Prox}_{\sigma^{-1}g}\|\|\boldsymbol{\eta}_R)\|.$$

Hence, we have

$$\|\boldsymbol{\eta}_R - \boldsymbol{\eta}_*\| \le \frac{\alpha^3}{1-\beta}\|(R_\theta - \mathrm{Prox}_{\sigma^{-1}g})\|\|\boldsymbol{\eta}_R\|.$$

We claim there exists a consistent upper bound M on $\boldsymbol{\eta}_R$.

$$\|\boldsymbol{\eta}_R\| \le \|f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ R_\theta \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta}_R)\| + 1. \tag{12}$$

By the Universal approximation theorem [33], there exists a network $R_\theta(\cdot)$ approximating the proximal operator $\mathrm{Prox}_{\sigma^{-1}g}$ with arbitrary precision. This indicates that for any $\varepsilon > 0, \boldsymbol{\eta} \in \mathbb{X}$, we have

$$\|f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ (R_\theta - \mathrm{Prox}_{\sigma^{-1}g}) \circ f_{\mathrm{LT}_1}(\boldsymbol{\eta})\| < \epsilon$$

when R_θ sufficiently approximates $\mathrm{Prox}_{\sigma^{-1}g}$.

It demonstrates that

$$\mathrm{Lip}(f_{\mathrm{LT}_3} \circ f_{\mathrm{LT}_2} \circ R_\theta \circ f_{\mathrm{LT}_1}) < 1.$$

Combined with (12), it convinces the fact that $\|\boldsymbol{\eta}_R\|$ is bounded. Here, we denote the upper bound as M. Thus,

$$\|\boldsymbol{\eta}_R - \boldsymbol{\eta}_*\| \le \frac{\alpha^3 M}{1-\beta}\|(R_\theta - \mathrm{Prox}_{\sigma^{-1}g})\|.$$

Since $R_\theta \to \mathrm{Prox}_{\sigma^{-1}g}$, it can be indicated that

$$\boldsymbol{\eta}_R \to \boldsymbol{\eta}_*.$$

Therefore,

$$\boldsymbol{x}_R = \mathbf{J}_{\sigma\mathbf{M}_1}(\boldsymbol{\eta}_R) \to \mathbf{J}_{\sigma\mathbf{M}_1}(\boldsymbol{\eta}_*) = \boldsymbol{x}_*.$$