Robust Gridless Sound Field Decomposition Based on Structured Reciprocity Gap Functional in Spherical Harmonic Domain

Yuhta Takida, Shoichi Koyama, Natsuki Ueno and Hiroshi Saruwatari (The University of Tokyo)

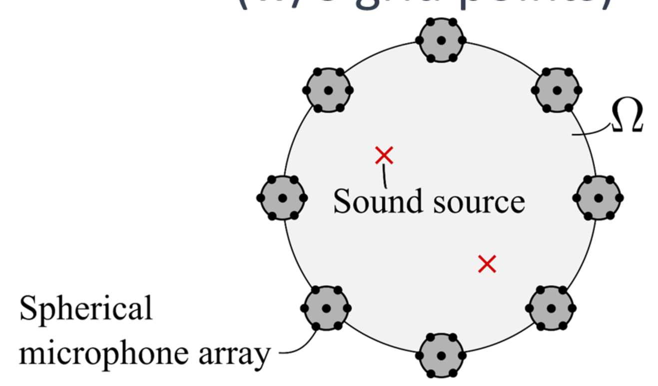
Abstract

Sound field decomposition

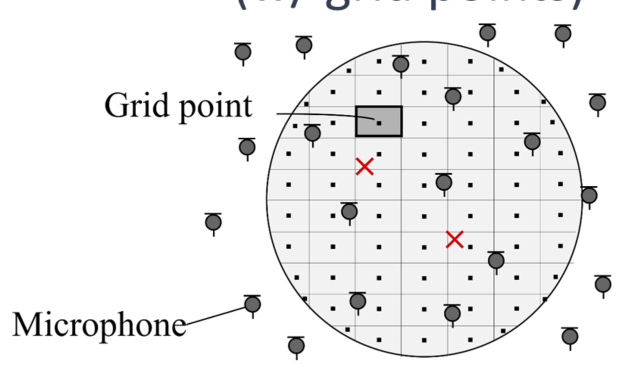
- Goal is to interpolate and reconstruct sound field inside region including sources (ill-posed problem!)
- Sound field should be decomposed into fundamental solutions of Helmholtz eq., i.e., point sources

Proposed method and its relation to prior works

Gridless sound field decomposition[1] (w/o grid points)



Sparse sound field decomposition[2] (w/ grid points)



- Based on spherical-harmonic-domain reciprocity gap functional (SHD-RGF)
- Reconstruction accuracy will be strongly affected by noise

- Based on discretization of possible source region into grid points
- Discrete set of point-source dictionary causes off-grid problem

Grouping time-frequency bins to improve robustness

▶ Proposed Structured RGF
- Formulation of structured SHD-RGF using annihilating filter (AF)
- Decomposition algorithm to find valid solution

▶ Group-sparse representation[2]
- Exploiting group-sparse structure in time-frequency domain
- Off-grid problem is still a major issue

[1] Y. Takida et al, "Gridless sound field decomposition based on reciprocity gap functional in spherical harmonic domain", in *IEEE SAM*, pp. 627—631, 2018.
[2] S. Koyama et al, "Sparse sound field decomposition for super-resolution in recording and reproduction", *JASA*, vol. 143, no.6, pp. 3780—3895, 2018.

Problem Statement

Sound field decomposition

- Source distribution $Q(\mathbf{r},k)$ inside Ω is approximated as a linear combination of J point sources

$$Q(\mathbf{r},k) \approx \sum_{j=1}^{J} c_j \delta(\mathbf{r}-\mathbf{r}_j) \quad \Rightarrow \quad u(\mathbf{r},k) \approx \sum_{j=1}^{J} c_j(k) G(\mathbf{r}|\mathbf{r}_j,k)$$

Spatial convolution of source distribution $Q(\cdot)$ with three-dimensional free-field Green's function $G(\cdot)$

- Pressure on $\partial\Omega$ is approximated in SHD by truncating harmonic order

$$u(\mathbf{r}) \approx \sum_{\nu=0}^{N} \sum_{\mu=-\nu}^{\nu} u_{\nu,\mu} h_\nu(kr) Y_{\nu,\mu}(\theta,\phi)$$

- Measurement model: Q N_mth-order microphone arrays are used

$$\boldsymbol{\alpha} = \mathbf{T}\mathbf{u} + \boldsymbol{\epsilon}$$

- $\boldsymbol{\alpha} \in \mathbb{C}^{Q(N_\mathrm{m}+1)^2}$: Measurements
- $\mathbf{T} \in \mathbb{C}^{Q(N_\mathrm{m}+1)^2 \times (N+1)^2}$: Translation matrix
- $\mathbf{u} \in \mathbb{C}^{(N+1)^2}$: Coefficients in SHD
- $\boldsymbol{\epsilon} \in \mathbb{C}^{Q(N_\mathrm{m}+1)^2}$: Measurements errors

Estimating c_j and $\mathbf{r}_j$ from $\boldsymbol{\alpha}$ makes it possible to reconstruct $u(\cdot)$

Sound Field Decomposition Based on SHD-RGF

Concept of RGF

- Test function $w_n(\cdot)$ and RGF $R(\cdot)$ for $w_n(\cdot)$ is defined as

$$w_n(\mathbf{r}) := p^n e^{ikz}, \quad p = x + iy$$

$$R(w_n) := \int_\Omega w_n(\mathbf{r}) Q(\mathbf{r}) d\mathbf{r}$$

- By applying point source assumption and Green's theorem to $R(w)$, the following equation holds:

$$\sum_{j=1}^{J} c_j w_n(\mathbf{r}_j) = \int_{\partial\Omega} \left(u(\mathbf{r}) \frac{\partial w(\mathbf{r})}{\partial \mathbf{n}} - w(\mathbf{r}) \frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} \right) dS \quad \cdots (*)$$

The parameter of sound sources J, $\mathbf{r}_j$ and c_j can be estimated from pressure and velocity values on $\partial\Omega$.

Localization based on SHD-RGF[1]

- Surface integral (Right-hand side of Eq. (*)) $s_n := R(w_n)$ can be analytically calculated in SHD[1]

$$s_n = \frac{i}{kR^2} \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} (-1)^{\mu+1} w_{\nu,-\mu}^{(n)} u_{\nu,\mu}$$

$w_{\nu,\mu}^{(n)}$: Analytical expansion coefficient of $w_n(\mathbf{r})$ by spherical harmonics[1]

- Compose Hankel matrices with $s_n := R(w_n)$ and estimate the source locations on x-y plane by eigenvalue decomposition.

Closed-form solutions can be greatly affected by measurement errors.

Structured SHD-RGF and Proposed Algorithm

Construction of AF for decomposition

- Construct AF represented by polynomial function

$$H(q) = \prod_{j=1}^{J} (1 - p_j q^{-1}) = \sum_{j=1}^{J} h_j q^{-j}$$

Its roots correspond to source locations $\{p_j\}_{j=1}^{J}$, $p_j = x_j + iy_j$

- Convolution coefficient sequence $\{h_j\}_{j=0}^{J}$ and elements $\{s_n\}_{n\in\mathbb{N}}$ $(n = 1, \ldots, J)$

$$h_{n+J} * s_{n+J} = \sum_{j=0}^{J} h_j s_{n+J-j}$$

$$= \sum_{j'=1}^{J} c_{j'} e^{-ikz_{j'}} p_{j'}^{n+J} \sum_{j=0}^{J} h_j p_{j'}^{-j} = 0 \quad \Rightarrow \quad \textbf{Annihilation: } \mathbf{h} * \mathbf{s} = \mathbf{0}$$

where $H(p_{j'}) = 0$.

- Annihilation holds for multiple time-frequency (T-F) bins when the source locations are assumed to be static for T time frames.

Annihilation using multiple T-F bins: $\mathbf{h} * \mathbf{s}_{f,t} = \mathbf{0}$

$f \in \{1, \ldots, F\}$: index of freq bins
$t \in \{1, \ldots, T\}$: index of time frames

Proposed AF-based Algorithm for SHD-RGF

- Optimization problem for SHD-RGF using AF

$$\underset{\mathbf{u}_{f,t},\mathbf{h}}{\text{minimize}} \quad \sum_{f,t} \|\boldsymbol{\alpha}_{f,t} - \mathbf{T}_f \mathbf{u}_{f,t}\|_2^2$$: Minimization of model error

$$\text{such that} \quad \mathbf{h} * \mathbf{s}_{f,t} = \mathbf{0}, \ \mathbf{h}^\mathsf{H}\mathbf{h} = 1$$: Annihilation and regularization

- Equivalent optimization problem

$$\underset{\mathbf{h}}{\text{minimize}} \quad \mathbf{h}^\mathsf{H}\boldsymbol{\Lambda}(\mathbf{h})\mathbf{h} \qquad \text{such that} \quad \mathbf{h}^\mathsf{H}\mathbf{h} = 1$$

$$\mathbf{h} * \mathbf{s}_{f,t} = \mathbf{W}_f(\mathbf{h})\mathbf{u}_{f,t} = \mathbf{V}_f(\mathbf{u}_{f,t})\mathbf{h}$$
$$\boldsymbol{\Lambda}(\mathbf{h}) := \sum_{f,t} \mathbf{V}_f(\mathbf{v}_{f,t})^\mathsf{H} \boldsymbol{\Sigma}(\mathbf{h})^{-1} \mathbf{V}_f(\mathbf{v}_{f,t})$$
$$\boldsymbol{\Sigma}(\mathbf{h}) := \mathbf{W}_f(\mathbf{h})(\mathbf{T}_f^\mathsf{H}\mathbf{T}_f)^{-1}\mathbf{W}_f(\mathbf{h})^\mathsf{H}$$
$$\mathbf{v}_{f,t} := (\mathbf{T}_f^\mathsf{H}\mathbf{T}_f)^{-1}\mathbf{T}_f^\mathsf{H}\boldsymbol{\alpha}_{f,t}$$

- Variable $\mathbf{h}$ is updated by solving above problem iteratively starting with initial value $\mathbf{h}^{(0)}$

$$\mathbf{h}^{(i+1)} = \arg\min_{\mathbf{h}} \mathbf{h}^\mathsf{H}\boldsymbol{\Lambda}(\mathbf{h}^{(i)})\mathbf{h}, \ \text{such that} \ \mathbf{h}^\mathsf{H}\mathbf{h} = 1,$$

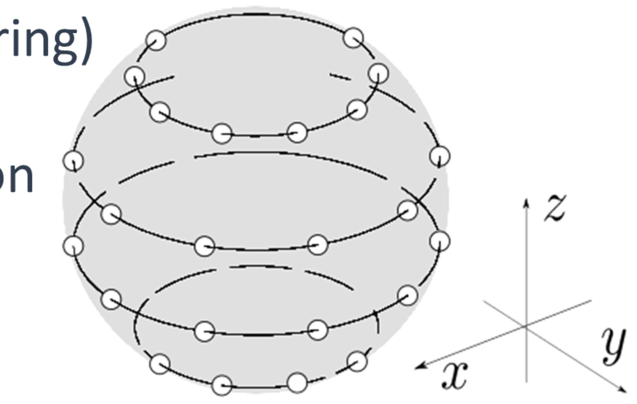
It can be computed in closed-form at each iteration

Numerical Simulations

Simulation conditions

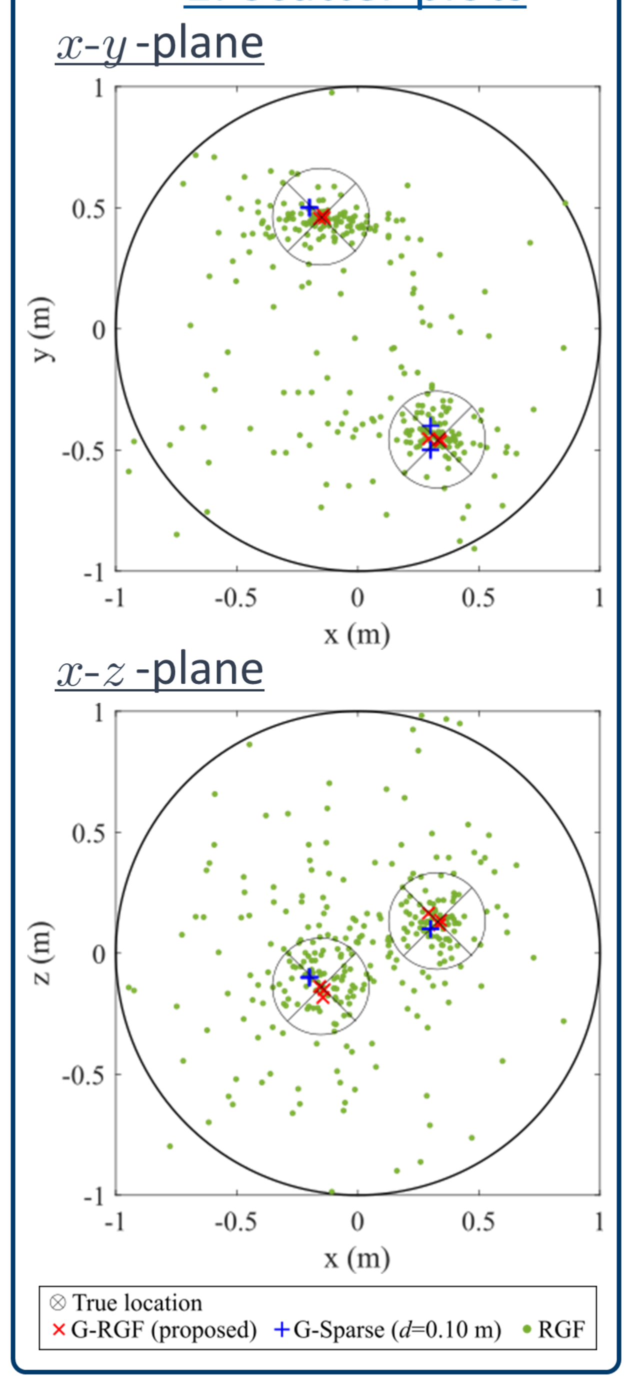
- Comparing proposed method (**G-RGF**) with
  - G-Sparse: group sparse sound field decomposition[2] (interval of grid points d of 0.10, 0.15, and 0.20 m)
  - RGF: RGF in spherical harmonic domain[1] for single time-freq. bin
- Microphone array
  - 24 second-order spherical microphone arrays
  - Four-ring geometry (nine arrays on each ring)
- Evaluation criteria
  - Root-mean-square error of source location

$$\mathrm{RMSE} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} \|\mathbf{r}_{j,\mathrm{true}} - \hat{\mathbf{r}}_j\|_2^2}$$

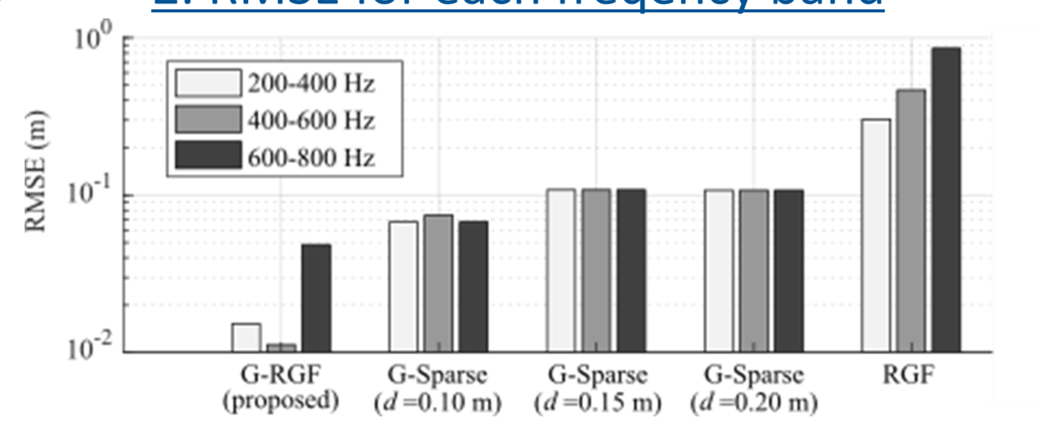


Results

1. Scatter plots

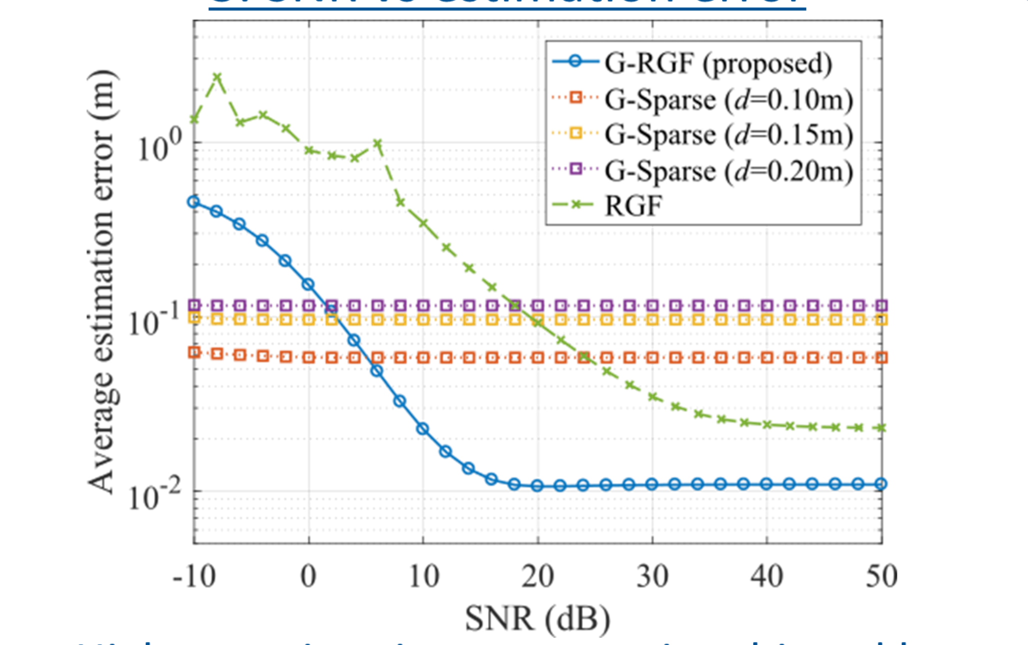


2. RMSE for each freqency band



Highest source localization accuracy is achieved by **G-RGF** for all group of time-frequency bands

3. SNR vs estimation error



Highest estimation accuracy is achieved by proposed **G-RGF** above 6 dB

▶ Off-grid problem is avoided on the basis of RGF
▶ Robustness against noise is improved by grouping T-F bins