Group Sparse Optimization for Inpainting of Random Fields on the Sphere *

Chao Li ‡ and Xiaojun Chen †

Abstract

We propose a group sparse optimization model for inpainting of a square-integrable isotropic random field on the unit sphere, where the field is represented by spherical harmonics with random complex coefficients. In the proposed optimization model, the variable is an infinite-dimensional complex vector and the objective function is a real-valued function defined by a hybrid of the $\ell_2$ norm and non-Liptchitz $\ell_p (0 < p < 1)$ norm that preserves rotational invariance property and group structure of the random complex coefficients. We show that the infinite-dimensional optimization problem is equivalent to a convexly-constrained finite-dimensional optimization problem. Moreover, we propose a smoothing penalty algorithm to solve the finite-dimensional problem via unconstrained optimization problems. We provide an approximation error bound of the inpainted random field defined by a scaled KKT point of the constrained optimization problem in the square-integrable space on the sphere with probability measure. Finally, we conduct numerical experiments on band-limited random fields on the sphere and images from Earth topography data to show the promising performance of the smoothing penalty algorithm for inpainting of random fields on the sphere.

Keywords: group sparse optimization; exact penalty; smoothing method; random field.

1 Introduction

Let $(\Omega, F, P)$ be a probability space, and let $\mathcal{B}(S^2)$ be the Borel sigma algebra on the unit sphere $S^2 := \{x \in \mathbb{R}^3 : \|x\| = 1\}$, where $\|\cdot\|$ is the $\ell_2$ norm. A real-valued random field on $S^2$ is an $F \otimes \mathcal{B}(S^2)$-measurable function $T : \Omega \times S^2 \to \mathbb{R}$, and it is said to be 2-weakly isotropic if the expected value and covariance of $T$ are rotationally invariant (see [14]). It is known that a 2-weakly isotropic random field on the sphere has the following Karhunen-Loève (K-L) expansion [23][25],

$$T(\omega, x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m}(\omega) Y_{l,m}(x), \quad x \in S^2, \quad \omega \in \Omega, \quad (1.1)$$

---

*This work was partly supported by Hong Kong Research Grant Council PolyU15300120 and the Shenzhen Research Institute of Big Data Grant 2019ORF01002.
‡Department of Applied Mathematics, The Hong Kong Polytechnic University (chaoo.li@connect.polyu.hk).
†Department of Applied Mathematics, The Hong Kong Polytechnic University (xiaojun.chen@polyu.edu.hk).
where $\alpha_{l,m}(\omega) = \int_{S^2} T(\omega, x) Y_{l,m}(x) d\sigma(x) \in \mathbb{C}$ are random spherical harmonic coefficients of $T(\omega, x)$, $Y_{l,m}$, for $m = -l, \ldots, l$, $l = 0, 1, 2, \ldots$ are spherical harmonics with degree $l$ and order $m$, and $\sigma$ is the surface measure on $S^2$ satisfying $\sigma(S^2) = 4\pi$. For convenient, let

$$\alpha(\omega) = (\alpha_{0,0}(\omega), \alpha_{1,0}(\omega), \alpha_{1,1}(\omega), \ldots, \alpha_{l,m}(\omega), \ldots)^T$$

denote the coefficient vector of an isotropic random field $T(\omega, x)$. The infinite-dimensional coefficient vector $\alpha(\omega)$ can be grouped according to degrees $l$ and written in the following group structure

$$\alpha(\omega) = (\alpha_{l,0}^T(\omega), \alpha_{l,1}^T(\omega), \ldots, \alpha_{l,l}^T(\omega), \ldots)^T,$$

where

$$\alpha_l(\omega) = (\alpha_{l,-l}(\omega), \ldots, \alpha_{l,0}(\omega), \ldots, \alpha_{l,l}(\omega))^T \in \mathbb{C}^{2l+1}, \quad l \geq 0.$$  

From [14], we know that for any given degree $l \geq 1$ and any $\omega \in \Omega$, the sum

$$\sum_{m=-l}^{l} |\alpha_{l,m}(\omega)|^2 = \|\alpha_l(\omega)\|^2$$

is rotationally invariant, while $\sum_{m=-l}^{l} |\alpha_{l,m}(\omega)|$ is not invariant under rotation of the coordinate axes. Moreover, in [14] the angular power spectrum is given by

$$C_l = \frac{1}{2l+1} \mathbb{E}[\|\alpha_l(\omega)\|^2]$$

if the expected value $\mathbb{E}[T(\omega, x)] = 0$ for all $x \in S^2$. In the study of isotropic random fields [12, 23, 25], the angular power spectrum plays an important role since it contains full information of the covariance of the field and provides characterization of the field. Hence, considering the group structure (1.2) of the coefficients of an isotropic random field is essential.

Sparse representation of a random field $T$ is an approximation of $T$ with few non-zero elements $\alpha_{l,m}(\omega)$ of $\alpha(\omega)$. In group sparse representation, instead of considering elements individually, we seek an approximation of $T$ with few non-zero groups $\alpha_l(\omega)$ of $\alpha(\omega)$, i.e., few non-zero entries of the following vector

$$(\|\alpha_{0,0}(\omega)\|, \|\alpha_{0,1}(\omega)\|, \ldots, \|\alpha_{l,l}(\omega)\|, \ldots)^T.$$ 

In recent years, sparse representation of isotropic random fields has been extensively studied. In [6] the authors studied the sparse representation of random fields via an $\ell_1$-regularized minimization problem. In [14], the authors considered isotropic group sparse representation of Gaussian and isotropic random fields on the sphere through an unconstrained optimization problem with a weighted $\ell_{2,1}$ norm, which is an example of group Lasso [34]. In [24], the authors proposed a non-Lipschitz regularization problem with a weighted $\ell_{2,p}$ ($0 < p < 1$) norm for the group sparse representation of isotropic random fields. The non-Lipschitz regularizer also preserves the rotational invariance property of $\|\alpha_l(\omega)\|^2$ for any given degree $l \geq 1$ and any $\omega \in \Omega$. 

2
Isotropic random fields on the sphere have many applications [5, 20, 26, 28, 31], especially in the study of the Cosmic Microwave Background (CMB) analysis [1, 30]. However, the true spherical random field usually presents masked regions and missing observations. Many inpainting methods [1, 13, 30, 32] have been proposed for recovering the true field based on sparse representation. However, they did not take the group structure (1.2) of the coefficients into consideration.

Group sparse optimization models have been successfully used in signal processing, imaging sciences and predictive analysis [3, 11, 17, 18, 27]. In this paper, we propose a constrained group sparse optimization model for inpainting of isotropic random fields on the sphere based on group structure (1.2). Moreover, to recover the true random field by using information only on a subset \( \Gamma \subseteq \mathbb{S}^2 \) which contains an open set, we need the unique continuation property [19] of any realization of a random field, that is, for any fixed \( \omega \in \Omega \), if the value of a field equals to zero on \( \Gamma \), then the random field is identically zero on the sphere (see Appendix A for more details).

Let \( L^2(\Omega \times \mathbb{S}^2) \) be the \( L^2 \) space on the product space \( \Omega \times \mathbb{S}^2 \) with product measure \( P \otimes \sigma \) and the induced norm \( \| \cdot \|_{L^2(\Omega \times \mathbb{S}^2)} = \mathbb{E}[\| \cdot \|_{L^2(\mathbb{S}^2)}] \), where \( L^2(\mathbb{S}^2) \) is the space of square-integrable functions over \( \mathbb{S}^2 \) endowed with the inner product

\[
\langle f, g \rangle_{L^2(\mathbb{S}^2)} = \int_{\mathbb{S}^2} f(x)\overline{g(x)}\,d\sigma(x), \quad \forall f, g \in L^2(\mathbb{S}^2),
\]

and the induced \( L^2 \)-norm \( \| f \|_{L^2(\mathbb{S}^2)} = (\langle f, f \rangle_{L^2(\mathbb{S}^2)})^{\frac{1}{2}} \). By Fubini’s theorem, for \( \omega \in \Omega \), \( T(\omega, x) \in L^2(\mathbb{S}^2) \), \( P \)-a.s., where “\( P \)-a.s.” stands for almost surely with probability 1. For brevity, we write \( T(\omega, x) \) as \( T(x) \) or \( T \) if no confusion arises.

Let the observed field be given by

\[
T^\circ := A(T^*) + \Delta, \quad (1.3)
\]

where \( T^* \in L^2(\Omega \times \mathbb{S}^2) \) is the true isotropic random field that we aim to recover, \( \Delta : \Omega \times \mathbb{S}^2 \to \mathbb{R} \) is observational noise and \( A : L^2(\Omega \times \mathbb{S}^2) \to L^2(\Omega \times \mathbb{S}^2) \) is an inpainting operator defined by

\[
A(T(x)) = \begin{cases} 
T(x) & \text{if } x \in \Gamma \\
0 & \text{if } x \in \mathbb{S}^2 \setminus \Gamma,
\end{cases} \quad (1.4)
\]

where \( \mathbb{S}^2 \setminus \Gamma \subset \mathbb{S}^2 \) is a nonempty inpainting area and the set \( \Gamma \subset \mathbb{S}^2 \) has an open subset. In our optimization model, we consider the observed field \( T^\circ \) as one realization of a random field. For notational simplicity, let

\[
\alpha = (\alpha^T_0, \alpha^T_1, \ldots, \alpha^T_l, \ldots)^T, \quad (1.5)
\]

denote the spherical harmonic coefficient vector of an isotropic random field \( T \) for a fixed \( \omega \in \Omega \), where

\[
\alpha_l = (\alpha_{l,-l}, \ldots, \alpha_{l,0}, \ldots, \alpha_{l,l})^T \in \mathbb{C}^{2l+1}, \ l \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}.
\]

By Parseval’s theorem, we have

\[
\|T\|_{L^2(\mathbb{S}^2)}^2 = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |\alpha_{l,m}|^2 = \sum_{l=0}^{\infty} \|\alpha_l\|^2 < \infty.
\]
Since $\rho > \ell$ is widely used in sparse approximation [11,18]. In this paper, we use a weighted $\ell_2$ norm. The number of non-zero groups of $\alpha$ with group structure (1.5) is calculated by
\[
\|\alpha\|_{2,0} = \sum_{l=0}^{\infty} \|\alpha_l\|^0,
\]
where $\|\alpha_l\|^0 = \begin{cases} 1 & \text{if } \|\alpha_l\| > 0 \\ 0 & \text{otherwise,} \end{cases}$ and $\|\alpha\|_{2,0}$ is called $\ell_{2,0}$ norm of $\alpha$. The $\ell_{2,0}$ norm is discontinuous and $\|T\|_{L_2(S^2)}^2 < \infty$ does not ensure $\|\alpha\|_{2,0} < \infty$. In [14], the authors used a weighted $\ell_{2,1}$ norm of $\alpha$ defined by
\[
\|\alpha\|_{2,1} = \sum_{l=0}^{\infty} \beta_l \|\alpha_l\|,
\]
where $\beta_l > 0$. It is easy to see that $\lim_{p \to 0} \|\alpha_l\|^p = \|\alpha_l\|^0$. The $\ell_{2,p}$ norm $(0 < p < 1)$ has been widely used in sparse approximation [11,18]. In this paper, we use a weighted $\ell_{2,p}$ norm of $\alpha$. Let
\[
\ell_{p,\beta}^p := \left\{ \alpha \in \ell^2 : \|\alpha\|_{2,p}^p := \sum_{l=0}^{\infty} \beta_l \|\alpha_l\|^p < \infty \right\}
\]
be a weighted $\ell^p$ $(0 < p < 1)$ space with positive weights $\beta_0 = 1$ and $\beta_l = \eta^l p$ for $l \geq 1$, where $\eta > 1$ is a constant. For a fixed $\omega \in \Omega$, we consider the following constrained optimization problem
\[
\min_{\alpha \in \ell_{p,\beta}^p} \|\alpha\|_{2,p}^p \quad \text{s.t. } \|A(T(x)) - T^\circ(x)\|_{L_2(S^2)}^2 \leq \varrho, \tag{1.6}
\]
where $\varrho > \int_{S^2 \backslash \Gamma} |T^\circ(x)|^2 d\sigma(x)$.

A feasible field of problem (1.6) takes the form
\[
T(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \alpha_{l,m} Y_{l,m}(x), \quad x \in S^2, \quad \alpha \in \ell_{p,\beta}^p.
\]
Since $T^\circ \in L_2(S^2)$, for $\epsilon = \varrho - \int_{S^2 \backslash \Gamma} |T^\circ(x)|^2 d\sigma(x)$, there is a finite number $L$ such that $\|T^\circ(x) - T_L^\circ(x)\|_{L_2(S^2)}^2 < \epsilon$, where $T_L^\circ(x) = \sum_{l=0}^L \sum_{m=-l}^l b_{l,m} Y_{l,m}(x)$ and $b_{l,m} = \int_{S^2} T^\circ(x) Y_{l,m}(x) d\sigma(x)$. Let $b_{l,m} = 0$ for $l = L + 1, \ldots$, $m = -l, \ldots, l$, then $(b_{0,0}, \ldots, b_{L,L}, 0, 0, 0, \ldots) \in \ell_{p,\beta}^p$. From the definition of $A$, we have
\[
\|A(T_L^\circ(x)) - T^\circ(x)\|_{L_2(S^2)}^2 = \int_{S^2 \backslash \Gamma} |T^\circ(x)|^2 d\sigma(x) + \int_{\Gamma} |T^\circ(x) - T_L^\circ(x)|^2 d\sigma(x) < \int_{S^2 \backslash \Gamma} |T^\circ(x)|^2 d\sigma(x) + \epsilon = \varrho.
\]
Thus, the feasible set of problem (1.6) is nonempty. Moreover the objective function of problem (1.6) is continuous and level-bounded. Hence an optimal solution of problem (1.6) exists.

Since $\alpha = 0$ implies that $\|\alpha\|_{2,p}^p = 0$, we assume $\|T^\circ(x)\|_{L_2(S^2)}^2 > \varrho$, which implies that $T(x) = 0$ is not a feasible field of problem (1.6).

Our main contributions are summarized as follows.
Based on the K-L expansion, the constraint of problem (1.6) can be written in a discrete form with variables \( \alpha \in \ell^p \beta \). We derive a lower bound for the \( \ell_2 \) norm of nonzero groups of scaled KKT points of (1.6). Based on the lower bound, we prove that the infinite-dimensional problem (1.6) is equivalent to a finite-dimensional optimization problem.

We propose a penalty method for solving the finite-dimensional optimization problem via unconstrained optimization problems. We establish the exact penalization results regarding local minimizers and \( \varepsilon \)-minimizers. Moreover, we propose a smoothing penalty algorithm and prove that the sequence generated by the algorithm is bounded and any accumulation point of the sequence is a scaled KKT point of the finite-dimensional optimization problem.

We give the approximation error of the random field represented by scaled KKT points of the finite-dimensional optimization problem in \( L_2(\Omega \times S^2) \).

The rest of this paper is organised as follows. In Section 2, we prove that the infinite-dimensional discrete optimization problem is equivalent to a finite-dimensional problem. In Section 3, we present the penalty method and give exact penalization results. In Section 4, we discuss optimality conditions of the finite-dimensional optimization problem and its penalty problem. Moreover, we propose a smoothing penalty algorithm and establish its convergence. In Section 5, we give the approximation error in \( L_2(\Omega \times S^2) \). In Section 6, we conduct numerical experiments on band-limited random fields and images from Earth topography data to compare our approach with some existing methods on the quality of the solutions and inpainted images. Finally, we give conclusion remarks in Section 7.

2 Discrete formulation of problem (1.6)

In this section, we propose the discrete formulation of problem (1.6) and prove that the discrete problem is equivalent to a finite-dimensional problem (2.5).

Based on the definition of \( \mathcal{A} \) and the spherical harmonic expansion of \( T \), we obtain that

\[
\| \mathcal{A}(T(x)) - T^\circ(x) \|_{L_2(S^2)}^2 = \int_{S^2} |\mathcal{A}(T(x)) - T^\circ(x)|^2 d\sigma(x) = \int_{\Gamma} |T(x) - T^\circ(x)|^2 d\sigma(x) + \int_{S^2\setminus \Gamma} |T^\circ(x)|^2 d\sigma(x) \\
= \int_{\Gamma} \left| \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(x) \right|^2 d\sigma(x) - 2Re \left( \int_{\Gamma} T^\circ(x) \left( \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\alpha}_{l,m} \overline{Y_{l,m}(x)} \right) d\sigma(x) \right) \\
+ \int_{S^2} |T^\circ(x)|^2 d\sigma(x) \\
= \alpha^T \bar{\alpha} - 2Re \left( \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\alpha}_{l,m} \int_{\Gamma} T^\circ(x) \overline{Y_{l,m}(x)} d\sigma(x) \right) + \int_{S^2} |T^\circ(x)|^2 d\sigma(x) \\
= \alpha^H Y \alpha - 2Re(\alpha^H \alpha^\circ) + c,
\]
where $Y$ is an infinite-dimensional matrix with $(Y)_{l^2+l+m+1,l^2+l+m+1} = \int_{\Gamma} Y_{l,m}(x)Y_{l,m}(x)d\sigma(x) \in \mathbb{C}$, $\alpha^o = ((\alpha^o_0)^T, \ldots, (\alpha^o_l)^T, \ldots)^T$ with $\alpha^o_l \in \mathbb{C}^{2l+1}$, $\alpha^o_{l,m} = \int_{\Gamma} T^o(x)Y_{l,m}(x)d\sigma(x)$ and $c = \int_{S^2} |T^o(x)|^2d\sigma(x)$.

The minimization problem (1.6) can be written as the following optimization problem

$$
\min_{\alpha \in \ell^p_{\mathbb{R}}} \|\alpha\|_2^p \tag{2.2}
$$

s.t. $\alpha^H Y\alpha - 2\text{Re}(\alpha^H \alpha^o) + c \leq \varrho.$

From the setting of $\varrho$ in problem (1.6) and Theorem 8, the feasible set of (2.2) has an interior point and bounded. The following assumption follows from our assumption on problem (1.6).

**Assumption 1.** The feasible set of problem (2.2) does not contain $\alpha = 0$.

Note that the objective function and constraint function in (2.2) are real-valued functions with complex variables. Thus, following [24], we apply the Wirtinger calculus (See Appendix B for more details) in this paper.

Since the objective function in problem (2.2) is not Lipschitz continuous at points containing zero groups. We shall consider first-order stationary points of (2.2) and extend the definition of scaled KKT points for finite-dimensional optimization problem with real variables in [8,10,29].

**Definition 1.** We call $\alpha^* \in \ell^p_{\mathbb{R}}$ a scaled KKT point of (2.2), if there exists $\nu \in \mathbb{R}$ such that

$$
p\beta_l\|\alpha^*_l\|^p(p\alpha^*_l + 2\nu\|\alpha^*_l\|^2(Y_l\alpha^* - \alpha^o_l)) = 0, \quad \forall l \in \mathbb{N}_0,
\nu((\alpha^*)^H Y\alpha^* - 2\text{Re}((\alpha^*)^H \alpha^o) + c - \varrho) = 0,
(\alpha^*)^H Y\alpha^* - 2\text{Re}((\alpha^*)^H \alpha^o) + c \leq \varrho,
\nu \geq 0. \tag{2.3}
$$

From (2.3), for any nonzero vector $\alpha^*_l \in \mathbb{C}^{2l+1}$, we have

$$p\beta_l\|\alpha^*_l\|^p - 2\nu(Y_l\alpha^* - \alpha^o_l) = 0,$$

and $\nu > 0$. Then,

$$p\beta_l\|\alpha^*_l\|^p = 2\nu\|Y_l\alpha^* - \alpha^o_l\| \leq 2\nu\|Y\alpha^* - \alpha^o\|.
$$

By definition of $Y$ and feasibility of $\alpha^*$, there exists $\hat{c} > 0$ such that $\|Y\alpha^* - \alpha^o\| \leq \hat{c}$. Hence for any nonzero vector $\alpha^*_l \in \mathbb{C}^{2l+1}$, we have

$$\|\alpha^*_l\| \geq \left(\frac{p\beta_l}{2\nu\hat{c}}\right)^{\frac{1}{1-p}}. \tag{2.4}
$$

By definition of $\beta_l$, $l \in \mathbb{N}_0$, we obtain that

$$\infty > \|\alpha^*\|_2^p = \sum_{l=0}^{\infty} \beta_l\|\alpha^*_l\|^p = \sum_{\{l \in \mathbb{N}_0: \|\alpha^*_l\| \neq 0\}} \beta_l\|\alpha^*_l\|^p \geq \left(\frac{p}{2\nu\hat{c}}\right)^{\frac{1}{1-p}} \sum_{\{l \in \mathbb{N}_0: \|\alpha^*_l\| \neq 0\}} (\nu\hat{c})^{\frac{1}{1-p}},$$

which completes the proof.
which implies that \( \{ l \in \mathbb{N}_0 : \| \alpha^*_l \| \neq 0 \} \) is a finite set. Thus, the number of nonzero vectors \( \alpha^*_l \in \mathbb{C}^{2l+1} \) of a scaled KKT point of problem (2.2) is finite and there exists an \( L \in \mathbb{N}_0 \) such that \( L = \max \{ l \in \mathbb{N}_0 : \| \alpha^*_l \| \neq 0 \} \). Therefore, the infinite-dimensional problem (2.2) can be truncated to a finite-dimensional problem and approximated in a finite-dimensional space.

For notational simplicity, we truncate \( \alpha \in \ell^p_\beta \) to \( \alpha = (\alpha_{0}, 0, \alpha^T_1, \ldots, \alpha^T_L)^T \in \mathbb{C}^d \), where \( d := (L+1)^2 \). We use \( \hat{\alpha} \in \mathbb{C}^d \) to denote the truncated vector whose elements are the first \( d \) elements of \( \alpha^o \) and \( \hat{Y} \in \mathbb{C}^{d \times d} \) to denote the leading principal submatrix of order \( d \) of \( Y \). Since \( \Gamma \) has an open subset, we have \( z^H \hat{Y} z > 0 \) for any nonzero \( z \in \mathbb{C}^d \). Thus, the matrix \( \hat{Y} \) is positive definite.

The truncated finite-dimensional problem of problem (2.2) has the following version

\[
\begin{align*}
\min_{\alpha \in \mathbb{C}^d} & \quad \Phi(\alpha) := \sum_{l=0}^{L} \beta_l \| \alpha_l \|^p \\
\text{s.t.} & \quad \alpha^H \hat{Y} \alpha - 2 \text{Re}(\alpha^H \hat{\alpha}^o) + c \leq g.
\end{align*}
\]

By the definition of \( g \), the feasible set of (2.5) is nonempty and has an interior point. The objective function \( \Phi(\alpha) \) is continuous, nonnegative with \( \Phi(0) = 0 \), and differentiable except at points containing zero groups. The penalty formulation for (2.5) is

\[
\begin{align*}
\min_{\alpha \in \mathbb{C}^d} & \quad F_\lambda(\alpha) := \Phi(\alpha) + \lambda \big( \alpha^H \hat{Y} \alpha - 2 \text{Re}(\alpha^H \hat{\alpha}^o) + c - g \big)_+,
\end{align*}
\]

for some \( \lambda > 0 \), where \((\cdot)_+ := \max\{\cdot, 0\}\). For any nontrivial solution \( \alpha^* \) of (2.6), we have

\[
0 < F_\lambda(\alpha^*) = \min_{\alpha \in \mathbb{C}^d} F_\lambda(\alpha) < F_\lambda(0) = \lambda(c - g)_+ = \lambda(c - g) < \lambda c.
\]

In Sections 3-4, we focus on problems (2.5) and (2.6).

### 3 Exact penalization

In this section, we consider the relationship between problems (2.5) and (2.6). We first give some notations. For a closed set \( S \subset \mathbb{C}^n \), \( \text{dist}(z, S) = \inf_{z' \in S} \| z - z' \| \) denotes the distance from a point \( z \in \mathbb{C}^n \) to \( S \) and \( \mathbf{B}(b; r) = \{ z \in \mathbb{C}^n : \| z - b \| \leq r \} \) denotes a closed ball with radius \( r > 0 \) and center \( b \in \mathbb{C}^n \). Let \( g : \mathbb{C}^d \to \mathbb{R} \) be defined as

\[
g(\alpha) := \alpha^H \hat{Y} \alpha - 2 \text{Re}(\alpha^H \hat{\alpha}^o) + c - g,
\]

\( S(\alpha) := \{ \delta \in \mathbb{R} : g(\alpha) \leq \delta \} \) for \( \alpha \in \mathbb{C}^d \), and \( S^{-1}(\delta) := \{ \alpha \in \mathbb{C}^d : g(\alpha) \leq \delta \} \). We denote the feasible set of problem (2.5) by \( F_\varepsilon := \{ \alpha \in \mathbb{C}^d : g(\alpha) \leq 0 \} \). Let \( L := \{ 0, 1, \ldots, L \} \).

Since \( \hat{Y} \) is positive definite, \( g \) is strongly convex and has a unique global minimizer \( \hat{\alpha}^o \neq \hat{\alpha}^o \) which implies that \( g(\hat{\alpha}^o) \in (\inf_{\alpha \in \mathbb{C}^d} g(\alpha), \infty) \). Since \( g \) is strongly convex and quadratic, there is \( C \) such that \( \| \alpha - \bar{\alpha} \| \leq C | g(\alpha) - g(\bar{\alpha}) | \) for \( \alpha, \bar{\alpha} \in \mathbb{C}^d \). Choosing \( \bar{\alpha} \) such that \( g(\bar{\alpha}) = 0 \), from [29] Theorem 9.48, we obtain the following lemma.
Lemma 1. There exists a constant $C > 0$ such that for any $\alpha \in \mathbb{C}^d$,
\[
\text{dist}(\alpha, \mathcal{F}_e) = \text{dist}(\alpha, S^{-1}(0)) \leq C \text{dist}(0, S(\alpha)) = C(g(\alpha))_+.
\]

Theorem 1. There exists a $\lambda^* > 0$ such that a local minimizer $\alpha^* \in \mathbb{C}^d$ of problem (2.5) is a local minimizer of problem (2.6) whenever $\lambda \geq \lambda^*$.

Proof. Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of problem (2.5), that is there exists a neighborhood $\mathcal{N}$ of $\alpha^*$ such that $\Phi(\alpha^*) \leq \Phi(\alpha)$ for $\alpha \in \mathcal{N} \cap \mathcal{F}_e$. We denote the group support set of $\alpha^*$ by $\gamma := \{l \in \mathbb{L} : \|\alpha^*_l\| \neq 0\}$, and the complement set of $\gamma$ in $\mathbb{L}$ by $\tau := \{l \in \mathbb{L} : \|\alpha^*_l\| = 0\}$. Let $\alpha_\gamma$ and $\alpha_\tau$ denote the restrictions of $\alpha$ onto $\gamma$ and $\tau$, respectively. We obtain that $\alpha^*_\gamma$ is a local minimizer of the following problem
\[
\min_{\alpha_\gamma} \sum_{l \in \gamma} \beta_l \|\alpha_l\|^p \\
\text{s.t.} \quad \alpha^H \hat{Y}_\gamma \alpha_\gamma - 2\text{Re}(\alpha^H \hat{A}_\gamma) + c \leq 0.
\]

Let $\bar{\epsilon} = \frac{1}{2} \min\{\|\alpha^*_l\| : l \in \gamma\} > 0$. Then, there exists a small $\delta^*$ such that $\alpha^*_\gamma$ is a local minimizer of \((3.1)\) and $\min\{\|\alpha_l\| : l \in \gamma\} > \bar{\epsilon}$ for all $\alpha_\gamma \in B(\alpha^*_\gamma; \delta^*)$. Let $g_\gamma(\alpha_\gamma) = \alpha^H \hat{Y}_\gamma \alpha_\gamma - 2\text{Re}(\alpha^H \hat{A}_\gamma) + c - 0$ and $\Omega_1 := \{\alpha_\gamma : g_\gamma(\alpha_\gamma) \leq 0\}$. Let $[a_\gamma; 0_\tau]$ be the vector with $([a_\gamma; 0_\tau])_l = a_l, l \in \gamma$ and $([a_\gamma; 0_\tau])_l = 0, l \in \tau$. It is easy to see that $g([a_\gamma; 0_\tau]) = g_\gamma(\alpha_\gamma)$ and $\text{dist}(\alpha_\gamma, \Omega_1) = \text{dist}([a_\gamma; 0_\tau], \mathcal{F}_e)$. By Lemma 1, $\text{dist}(\alpha_\gamma, \Omega_1) \leq C(g_\gamma(\alpha_\gamma))_+$. We denote the group support set of $\alpha_\gamma$ by $\gamma(\alpha_\gamma)$ and the complement set of $\gamma(\alpha_\gamma)$ in $\mathbb{L}$ by $\Omega_1 := \{\alpha_\gamma : g_\gamma(\alpha_\gamma) \leq 0\}$.

The objective function of \((3.1)\) is Lipschitz continuous on $B(\alpha^*_\gamma; \delta^*)$. Then by \cite{8}, Lemma 3.1, there exists a $\lambda^* > 0$ such that, for any $\lambda \geq \lambda^*$, $\alpha^*_\gamma$ is a local minimizer of the following problem
\[
\min_{\alpha_\gamma} F^\gamma_\lambda(\alpha_\gamma) := \sum_{l \in \gamma} \beta_l \|\alpha_l\|^p + \lambda(g_\gamma(\alpha_\gamma))_+,
\]
that is, there exists a neighborhood $U_\gamma$ of $0$ with $U_\gamma \subseteq B(0; \delta^*)$ such that $F^\gamma_\lambda(\alpha_\gamma) \geq F^\gamma_\lambda(\alpha^*_\gamma), \quad \forall \alpha_\gamma \in \alpha^*_\gamma + U_\gamma$.

We now show that $\alpha^*$ is a local minimizer of \((2.6)\) with $\lambda \geq \lambda^*$.

For fixed $\epsilon \in (0, \bar{\epsilon})$ and $\lambda \geq \lambda^*$, we consider problem \((2.6)\) in the neighborhood $U := U_\gamma \times (-\epsilon, \epsilon)^{d_\tau}$, where $d_\tau = \sum_{l \in \tau}(2l + 1)$. Let $\tilde{k}$ be a Lipschitz constant of the function $\lambda(g(\alpha))_+$ over $\alpha^* + U$. There exists an $\epsilon_0 \in (0, \epsilon)$ such that whenever $\|\alpha_0\| < \epsilon_0$, $l \in \tau$,
\[
\Phi(\alpha) = \sum_{l \in \gamma} \beta_l \|\alpha_l\|^p + \sum_{l \in \tau} \beta_l \|\alpha_l\|^p = \sum_{l \in \gamma} \beta_l \|\alpha_l\|^p + \tilde{k} \sum_{l \in \tau} \sum_{l \in \tau} \|\alpha_l\|.
\]

Hence for any $y \in U_\gamma \times (-\epsilon_0, \epsilon_0)^{d_\tau}$, we have
\[
F^\gamma_\lambda(\alpha^* + y) = \lambda(g(\alpha^* + y))_+ + \sum_{l \in \gamma} \beta_l \|\alpha^*_l + y_l\|^p + \sum_{l \in \tau} \beta_l \|y_l\|^p \\
\geq \lambda(g([\alpha^*_\gamma + y_\gamma; 0_\tau]))_+ - \tilde{k} \sum_{l \in \tau} \|y_l\| + \sum_{l \in \gamma} \beta_l \|\alpha^*_l + y_l\|^p + \tilde{k} \sum_{l \in \tau} \|y_l\| \\
\geq F^\gamma_\lambda(\alpha^*_\gamma) = F^\gamma_\lambda(\alpha^*),
\]

8
where the first inequality follows from the Lipschitz continuity of \( g_\lambda \) with Lipschitz constant \( \bar{\kappa} \) and (3.2), and the last inequality follows from the local optimality of \( \alpha^*_\varepsilon \). Thus, \( \alpha^*_\varepsilon \) is a local minimizer of problem (2.6) with \( \lambda \geq \lambda^* \). This completes the proof.

**Theorem 2.** Let \( \tilde{\alpha} = \hat{Y}^{-1}\hat{\alpha}^\circ, \varepsilon > 0 \) and \( \lambda \geq C\bar{\beta}^\frac{1}{L}(\varepsilon(L + 1)\frac{\bar{\kappa}}{2})^{-\frac{1}{L}}\bar{\Phi}(\tilde{\alpha}) \), where \( C \) is defined in Lemma 1 and \( \bar{\beta} = \eta L^p \). Then for any global minimizer \( \alpha^* \) of problem (2.6), the projection \( \alpha_\varepsilon := \mathcal{P}_{F_\varepsilon}(\alpha^*) \) is an \( \varepsilon \)-minimizer of (2.5), that is, \( \Phi(\alpha_\varepsilon) \leq \min\{\Phi(\alpha) : \alpha \in \mathcal{F}_\varepsilon\} + \varepsilon \).

**Proof.** Note that \( g \) is a strongly convex function and \( \tilde{\alpha} \) is a minimizer of \( g \). Since the feasible set \( \mathcal{F}_\varepsilon \) of (2.5) has an interior point, we have \( \tilde{\alpha} \in \text{int}\mathcal{F}_\varepsilon \).

By the global optimality of \( \alpha^* \), we have \( F_\lambda(\alpha^*) \leq F_\lambda(\tilde{\alpha}) \) and

\[
((\alpha^*)^H\hat{Y}\alpha^* - 2\text{Re}((\alpha^*)^H\hat{\alpha}^\circ) + c - g)_+ \leq \frac{1}{\lambda} F_\lambda(\alpha^*) \leq \frac{1}{\lambda} F_\lambda(\tilde{\alpha}) = \frac{1}{\lambda} \bar{\Phi}(\tilde{\alpha}).
\]

Hence for any \( \alpha \in \mathcal{F}_\varepsilon \), we obtain

\[
\Phi(\alpha_\varepsilon) - \Phi(\alpha) \leq \sum_{l=0}^L \beta_l ((\alpha_\varepsilon)_l \|p - \|\alpha_l^*\|p\|) \leq \sum_{l=0}^L \beta_l ((\alpha_\varepsilon)_l - \alpha_l^*\|p \leq \bar{\beta} \sum_{l=0}^L (\|\alpha_\varepsilon)_l - \alpha_l^*\|p) \leq \bar{\beta}(L + 1)^{\frac{1}{L+1}} \sum_{l=0}^L ((\alpha_\varepsilon)_l - \alpha_l^*)^2 \leq \bar{\beta}(L + 1)^{\frac{1}{L+1}} (\text{dist}(\alpha^*, \mathcal{F}_\varepsilon))^p
\]

\[
\leq \bar{\beta}(L + 1)^{\frac{1}{L+1}} (C((\alpha^*)^H\hat{Y}\alpha^* - 2\text{Re}((\alpha^*)^H\hat{\alpha}^\circ) + c - g)_+)^p
\]

\[
\leq \bar{\beta}(L + 1)^{\frac{1}{L+1}} (\frac{C}{\lambda} \Phi(\tilde{\alpha}))^p \leq \varepsilon,
\]

where the first inequality is from the global optimality of \( \alpha^* \), the second inequality is from Lemma 2.4 in [8], the fifth inequality is from the concavity of function \( t \to t^\frac{p}{2} \) for \( t \geq 0 \), the sixth inequality is from Lemma 1, the seventh inequality is from (3.3) and the last inequality follows from the choice of \( \lambda \). Thus, the projection \( \mathcal{P}_{F_\varepsilon}(\alpha^*) \) is an \( \varepsilon \)-minimizer of (2.5). This completes the proof.

**4 Optimality conditions and a smoothing penalty algorithm**

In this section, we first define first-order optimality conditions of (2.5) and (2.6), which are necessary conditions for local optimality. We also derive lower bounds for the \( \ell_2 \) norm of nonzero groups of first-order stationary points of (2.5) and (2.6). Next we propose a smoothing penalty algorithm for solving (2.5) and prove its convergence to a first-order stationary point of (2.5).

**4.1 First-order optimality conditions**

We first present a first-order optimality condition of problem (2.6).

**Definition 2.** We call \( \alpha^* \in \mathbb{C}^d \) a scaled first-order stationary point of problem (2.6) if

\[
p\beta_l\|\alpha_l^*\|^p\alpha_l^* + 2\lambda\xi\|\alpha_l^*\|^2(\hat{Y}_l\alpha^* - \alpha_l^*) = 0, \quad \forall \; l \in \mathbb{L},
\]

(4.1)
for some $\xi$ satisfying $\begin{cases} 0 & \text{if } (\alpha^*)^H \hat{Y} \alpha^* - 2\text{Re}((\alpha^*)^H \hat{\alpha}) + c < \varrho \\ \in [0, 1] & \text{if } (\alpha^*)^H \hat{Y} \alpha^* - 2\text{Re}((\alpha^*)^H \hat{\alpha}) + c = \varrho \\ 1 & \text{otherwise.} \end{cases}$

**Theorem 3.** Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of (2.6). Then $\alpha^*$ is a scaled first-order stationary point of (2.6).

**Proof.** Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of (2.6). We obtain that $\alpha^*_\gamma$ is a local minimizer of

$$\min_{\alpha_\gamma} \sum_{l \in \gamma} \beta_l \|\alpha_l\|^p + \lambda(\alpha^*_\gamma)^H \hat{Y}_\gamma \alpha_\gamma - 2\text{Re}(\alpha^*_\gamma \hat{\alpha}_\gamma) + c - \varrho_+.$$  \hspace{1cm} (4.2)

Note that $\|\alpha^*_l\| \neq 0$, $l \in \gamma$, the first term of the objective function of (4.2) is continuously differentiable at $\alpha^*_l$. The first-order necessary optimality condition for problem (4.2) holds at $\alpha^*_\gamma$, that is,

$$p\beta_l \|\alpha^*_l\|^p - 2\lambda \xi ((\hat{Y}_\gamma \alpha^*_\gamma)_l - \alpha^*_l) = 0, \quad \forall l \in \gamma,$$  \hspace{1cm} (4.3)

for some $\xi$ satisfying $\begin{cases} 0 & \text{if } (\alpha^*_\gamma)^H \hat{Y}_\gamma \alpha^*_\gamma - 2\text{Re}((\alpha^*_\gamma)^H \hat{\alpha}_\gamma) + c < \varrho \\ \in [0, 1] & \text{if } (\alpha^*_\gamma)^H \hat{Y}_\gamma \alpha^*_\gamma - 2\text{Re}((\alpha^*_\gamma)^H \hat{\alpha}_\gamma) + c = \varrho \\ 1 & \text{otherwise.} \end{cases}$

Multiplying $\|\alpha^*_l\|^2$ on both sides of (4.3), we obtain

$$p\beta_l \|\alpha^*_l\|^p \alpha^*_l + 2\lambda \xi \|\alpha^*_l\|^2 ((\hat{Y}_\gamma \alpha^*_\gamma)_l - \alpha^*_l) = 0, \quad \forall l \in \gamma.$$  \hspace{1cm} (4.4)

Since $\|\alpha^*_l\| = 0$ for $l \in \tau$, we have

$$p\beta_l \|\alpha^*_l\|^p \alpha^*_l + 2\lambda \xi \|\alpha^*_l\|^2 (\hat{Y}_l \alpha^* - \alpha^*_l) = 0, \quad \forall l \in \mathbb{L}.$$  \hspace{1cm} (4.5)

Hence (4.1) holds at $\alpha^*$.

The next theorem gives a lower bound for the $\ell_2$ norm of nonzero groups of stationary points of (2.6).

**Theorem 4.** Let $\alpha^* \in \mathbb{C}^d$ be a scaled first-order stationary point of (2.6) and $\|\hat{Y} \alpha^* - \hat{\alpha}\| \leq \tilde{c}$ for some $\tilde{c} > 0$. Then,

$$\|\alpha^*_l\| \geq \left( \frac{p\beta_l}{2\lambda \tilde{c}} \right)^{\frac{1}{p}}, \forall l \in \gamma.$$  \hspace{1cm} (4.6)

The proof is similar with that of (2.4), and thus we omit it here. According to Theorems 1 and 3, Theorem 4 gives a lower bound for the $\ell_2$ norm of nonzero groups of local minimizers of problem (2.5).

Next, we consider the first-order optimality condition of problem (2.5). Similar to Definition 1, we call $\alpha^*$ is a scaled first-order stationary point or a scaled KKT point of (2.5) if the following conditions hold,

$$p\beta_l \|\alpha^*_l\|^p \alpha^*_l + 2\nu \|\alpha^*_l\|^2 (\hat{Y}_l \alpha^* - \alpha^*_l) = 0, \quad \forall l \in \mathbb{L},$$

$$\nu g(\alpha^*) = 0, \quad g(\alpha^*) \leq 0, \quad \nu \geq 0.$$  \hspace{1cm} (4.7)
Let \( \alpha^* \) be a local minimizer of problem (2.5). If there exists \( l \in \mathbb{L} \) such that \( \| \alpha^*_l \|(Y_l \alpha^* - \alpha^*_0) \neq 0 \), then \( \alpha^* \) is a scaled KKT point of (2.5).

**Proof.** Let \( \alpha^* \) be a local minimizer of problem (2.5), then \( \alpha^*_l \) is a local minimizer of problem (3.1). Recall \( g_\gamma(\alpha^*_\gamma) = \alpha^*_\gamma \hat{Y}_\gamma \alpha^*_\gamma - 2 \text{Re}(\alpha^*_\gamma \hat{\alpha}^*_\gamma) + \phi - \frac{1}{2} \), then \( g_\gamma(\alpha^*_\gamma) = g_\gamma(\alpha^*) \).

The condition that there exists \( l \in \mathbb{L} \) such that \( \| \alpha^*_l \|(Y_l \alpha^* - \alpha^*_0) \neq 0 \) implies that \( \hat{Y}_\gamma \alpha^*_\gamma - \hat{\alpha}^*_\gamma \neq 0 \) if \( g_\gamma(\alpha^*_\gamma) = 0 \). Thus the LICQ (linear independence constraint qualification) holds at \( \alpha^*_\gamma \) for problem (3.1).

Note that, the objective function of problem (3.1) and \( g_\gamma \) are continuously differentiable at \( \alpha^*_\gamma \) and \( \partial_{\alpha^*_\gamma} g_\gamma(\alpha^*_\gamma) = \hat{Y}_\gamma \alpha^*_\gamma - \hat{\alpha}^*_\gamma \). Hence \( \alpha^*_\gamma \) is a KKT point of (3.1), that is, there exists \( \nu \) such that
\begin{equation}
\begin{aligned}
\beta \| \alpha^*_l \|^{p-2} \alpha^*_l + 2 \nu \langle \hat{Y}_\gamma \alpha^*_\gamma - \hat{\alpha}^*_\gamma \rangle_l &= 0, \quad \forall l \in \gamma, \\
\nu g_\gamma(\alpha^*_\gamma) &= 0, \quad g_\gamma(\alpha^*_\gamma) \leq 0, \quad \nu \geq 0.
\end{aligned}
\end{equation}

Multiplying \( \| \alpha^*_l \|^2 \) on both sides of the first equality in (4.5), we obtain
\begin{equation}
\begin{aligned}
\beta \| \alpha^*_l \|^{p-2} \| \alpha^*_l \|^2 + 2 \nu \| \alpha^*_l \| \langle \hat{Y}_\gamma \alpha^*_\gamma - \hat{\alpha}^*_\gamma \rangle_l &= 0, \quad \forall l \in \gamma.
\end{aligned}
\end{equation}

Since \( \alpha^*_l = 0 \) for \( l \in \tau \), we obtain
\begin{equation}
\begin{aligned}
\beta \| \alpha^*_l \|^{p-2} \| \alpha^*_l \|^2 + 2 \nu \| \alpha^*_l \| \langle \hat{Y}_\gamma \alpha^*_\gamma - \hat{\alpha}^*_\gamma \rangle_l &= 0, \quad \forall l \in \mathbb{L}.
\end{aligned}
\end{equation}

Combining this with (4.5) and \( g(\alpha^*) = g_\gamma(\alpha^*_\gamma) \), we find that \( \alpha^* \) is a scaled KKT point of (2.5). The proof is completed.

From the definitions we can see that for some \( \lambda > 0 \), any scaled KKT point of problem (2.5) is a scaled first order stationary point of (2.6). Moreover, any scaled first order stationary point of problem (2.6) which belongs to the feasible set \( \mathcal{F}_c \) is a scaled KKT point of problem (2.5).

Following the proof of Theorem 5, we can show the existence of a KKT point of problem (2.2) by replacing \( \mathbb{L} \) by \( \mathbb{N}_0 \).

**Corollary 1.** Let \( \hat{\alpha} \) be a local minimizer of problem (2.2). If there exists \( l \in \mathbb{N}_0 \) such that \( \| \hat{\alpha}_l \|(Y_l \hat{\alpha} - \alpha^*_0) \neq 0 \), then \( \hat{\alpha} \) is a scaled KKT point of (2.2).

### 4.2 A smoothing penalty algorithm for problem (2.5)

We define a smoothing function of the nonsmooth function \( \lambda(g(\alpha))_+ \) as follows
\[
f_{\lambda,\mu}(\alpha) = \psi_{\lambda,\mu}(g(\alpha))
\]
with \( \psi_{\lambda, \mu}(s) := \lambda \max_{0 \leq t \leq 1} \{st - \frac{\mu}{2} t^2\} \) and \( \mu > 0 \) is a smoothing parameter. It is easy to verify that
\[
\psi_{\lambda, \mu}^{\prime}(s) = \lambda \min\{\max\{\frac{s}{\mu}, 0\}, 1\} \geq 0 \quad \text{and} \quad |\psi_{\lambda, \mu}^{\prime}(s_1) - \psi_{\lambda, \mu}^{\prime}(s_2)| \leq \frac{\lambda}{\mu}|s_1 - s_2|, \forall \ s_1, s_2 \in \mathbb{R}.
\]
It is not hard to show that
\[
f_{\lambda, \mu}(\alpha) = \begin{cases} 
0 & \text{if } g(\alpha) \leq 0 \\
\frac{\lambda}{2\mu} (g(\alpha))^2 & \text{if } 0 \leq g(\alpha) \leq \mu \\
\lambda g(\alpha) - \frac{\lambda\mu}{2} & \text{if } g(\alpha) \geq \mu
\end{cases}
\]
and
\[
0 \leq \lambda (g(\alpha))_+ - f_{\lambda, \mu}(\alpha) \leq \frac{\lambda\mu}{2}.
\] (4.6)

By Wirtinger calculus, we obtain \( \nabla f_{\lambda, \mu}(\alpha) = \begin{bmatrix} \partial_\alpha f_{\lambda, \mu}(\alpha) \\ \partial_{\tilde{\alpha}} f_{\lambda, \mu}(\alpha) \end{bmatrix} \), where \( \partial_\alpha f_{\lambda, \mu}(\alpha) = \partial_{\tilde{\alpha}} f_{\lambda, \mu}(\alpha) \) and
\[
\partial_\alpha f_{\lambda, \mu}(\alpha) = \begin{cases} 
0 & \text{if } g(\alpha) \leq 0 \\
\hat{\alpha}\lambda g(\alpha) (\hat{Y}_\alpha - \tilde{\alpha}^o) & \text{if } 0 \leq g(\alpha) \leq \mu \\
\hat{\lambda} (\hat{Y}_\alpha - \tilde{\alpha}^o) & \text{if } g(\alpha) \geq \mu.
\end{cases}
\]

More details about the smoothing function can be found in [7] and references therein.

We consider the following optimization problem
\[
\min_{\alpha \in \mathbb{C}_+^n} F_{\lambda, \mu}(\alpha) := \Phi(\alpha) + f_{\lambda, \mu}(\alpha).
\] (4.7)

For fixed positive parameters \( \lambda \) and \( \mu \), \( F_{\lambda, \mu} \) is continuous and level-bounded since \( \Phi \) is level-bounded and \( f_{\lambda, \mu} \) is nonnegative. Moreover, the gradient of \( f_{\lambda, \mu} \) is Lipschitz continuous.

Now, we propose a smoothing penalty algorithm for solving problem (2.5).

A smoothing penalty algorithm for problem (2.5)

Choose \( \lambda^0 > 0, \mu^0 > 0, \varepsilon^0 > 0, \varsigma_1 > 1, \) and \( 0 < \varsigma_2 < 1 \). Set \( k = 0 \) and \( \alpha^0 = \tilde{\alpha} := \hat{Y}^{-1} \alpha^o \).

1. If \( F_{\lambda^k, \mu^k}(\alpha^k) > F_{\lambda^k, \mu^k}(\tilde{\alpha}) \), set \( \alpha^k = \tilde{\alpha} \); otherwise \( \alpha^k = \alpha^k \).
2. Solve problem (4.7) with initial point \( \alpha^k, \lambda = \lambda^k, \mu = \mu^k \), and find an \( \alpha^{k+1} \) satisfying
\[
||p_{\ell}|| |\alpha_{\ell}^{k+1}|^2 |\alpha_{\ell}^{k+1} + 2|\alpha_{\ell}^{k+1}|^2 (\partial_\alpha f_{\lambda^k, \mu^k}(\alpha^{k+1}))_{\ell}|| \leq \varepsilon^k, \quad \forall \ell \in \mathbb{L}.
\] (4.8)

3. Set \( \lambda^{k+1} = \varsigma_1 \lambda^k, \mu^{k+1} = \varsigma_2 \mu^{k+1}, \varepsilon^{k+1} = \varsigma_2 \varepsilon^k \).
4. Set \( k = k + 1 \) and go to (1).

We give the convergence of Algorithm 4.2 in the following theorem.

**Theorem 6.** Let \( \{\alpha^k\} \) be generated by Algorithm 4.2. Then, the following statements hold.

(i) \( \{\alpha^k\} \) is bounded.

(ii) Any accumulation point \( \alpha^* \) of \( \{\alpha^k\} \) is a scaled KKT point of problem (2.5).
Proof. (i) We can see that
\[
\Phi(\alpha^{k+1}) \leq F_{\lambda^k, \mu^k}(\alpha^{k+1}) \leq F_{\lambda^k, \mu^k}(\hat{\alpha}) = \Phi(\hat{\alpha}),
\]
where the first inequality follows from that \(f_{\lambda^k, \mu^k}(\alpha^{k+1}) \geq 0\), the second inequality follows from step (1) of Algorithm 4.2 and the equality is from \(\hat{\alpha} = \hat{Y}^{-1}\alpha^0 \in \mathcal{K}\) and \(f_{\lambda^k, \mu^k}(\hat{\alpha}) = 0\). Since \(\Phi\) is level-bounded, \(\{\alpha^k\}\) is bounded. Let \(k\) be bounded. Let \(\alpha^*\) be an accumulation point of \(\{\alpha^k\}\) and \(\{\alpha^k\}_{k \in \mathcal{K}}\) be a subsequence of \(\{\alpha^k\}\) such that \(\{\alpha^k\} \rightarrow \alpha^*\) as \(k \rightarrow \infty, k \in \mathcal{K}\). Note that
\[
\lambda^{k-1}(g(\alpha^k)) + \frac{\lambda^{k-1} \mu^{k-1}}{2} \leq f_{\lambda^{k-1}, \mu^{k-1}}(\alpha^k) \leq \lambda^{k-1}(\mu^{k-1}) \leq \Phi(\hat{\alpha}),
\]
where the first inequality follows from (4.6). Then, we have
\[
(g(\alpha^k))_+ \leq \frac{\Phi(\hat{\alpha})}{\lambda^{k-1}} + \frac{\mu^{k-1}}{2}.
\] (4.9)
From step (3) in Algorithm 4.2, \(\lambda^{k-1} \rightarrow \infty\) and \(\mu^{k-1} \rightarrow 0\), as \(k \rightarrow \infty, k \in \mathcal{K}\). Taking limits in (4.9) as \(k \rightarrow \infty, k \in \mathcal{K}\), we obtain that \((g(\alpha^*))_+ \leq 0\). Hence, \(\alpha^* \in \mathcal{F}_e\).

From (4.8), we have
\[
\|p_{\beta^k} \|s\alpha^k\|^p \alpha^k + 2\|\alpha^k\|^2(\partial_{\alpha} f_{\lambda^k, \mu^k}(\alpha^k))_{t^k} \| \leq \varepsilon^{k-1}, \quad \forall t \in \mathbb{L}.
\] (4.10)
We first assume that \(g(\alpha^*) < 0\). For all sufficiently large \(k\), we obtain \(g(\alpha^k) < 0\) and (4.10) becomes
\[
p_{\beta^k} \|s\alpha^k\|^p \alpha^k + 2\|\alpha^k\|^2(\hat{Y}_\gamma \alpha^k - \hat{\alpha}_\gamma) \| \leq \varepsilon^{k-1}, \quad \forall t \in \mathbb{L}.
\] (4.11)
Now, we prove that \(\{t^k\}_\mathcal{K}\) is bounded. On the contrary, we assume \(\{t^k\}_\mathcal{K}\) is unbounded and \(\{t^k\}_\mathcal{K} \rightarrow \infty\), then,
\[
\left\| \frac{s_{\beta^k}}{t^k} \|s\alpha^k\|^p \alpha^k + 2\|\alpha^k\|^2(\hat{Y}_\gamma \alpha^k - \hat{\alpha}_\gamma) \right\| \leq \frac{\varepsilon^{k-1}}{t^k}, \quad \forall t \in \mathbb{L}.
\]
Passing to the limit in the above relation gives
\[
\|\alpha^k\|^2(\hat{Y}_\gamma \alpha^k - \hat{\alpha}_\gamma) = 0, \quad \forall t \in \mathbb{L}.
\]
Since \(g(\alpha^*) = 0\) implies \(\alpha^*\) is in \(\mathcal{F}_e\), but is not a minimizer of \(g\), we have \(\alpha^* \neq 0\) and \(\hat{Y}_\gamma \alpha^* - \hat{\alpha}_\gamma \neq 0\). Moreover, \(g(\alpha^*) = g_\gamma([\alpha^*_\gamma; 0_{\gamma}]) = 0\) implies that \(\alpha^*_\gamma \neq 0, \forall t \in \gamma\) and \((\hat{Y}_\gamma \alpha^* - \hat{\alpha}_\gamma)_\gamma \neq 0\). Thus, \(\{t^k\}_\mathcal{K}\) is bounded. Let \(\{t^k\}_\mathcal{K} \rightarrow t^*\). Taking limits on both sides of (4.11) gives
\[
\left\| \frac{s_{\beta^k}}{t^*} \|s\alpha^k\|^p \alpha^k + 2t^*\|\alpha^k\|^2(\hat{Y}_\gamma \alpha^k - \hat{\alpha}_\gamma) \right\| = 0, \quad \forall t \in \mathbb{L}.
\]
Therefore, \(\alpha^*\) is a scaled KKT point of (2.5).
5 Approximation error

In [14], the authors gave the approximation error for random field using regularized $\ell_{2,1}$ model based on the observed random field $T^0(\omega, x) \in L_2(\Omega \times S^2)$. In this section, we estimate the approximation error of the inpainted random field in the space $L_2(\Omega \times S^2)$ based on the observed random field $T^0(\omega, x) \in L_2(\Omega \times S^2)$.

For any fixed $\omega \in \Omega$, let $\hat{\alpha}^*(\omega) := (\alpha_{0,0}^*(\omega), \ldots, \alpha_{L,\omega}^*(\omega))^T \in \mathbb{C}^{(L+1)^2}$ with the group support set $\gamma_\omega$ be a scaled KKT point of problem (2.5). By our results in the previous sections, $L_\omega$ is a finite number. Moreover, $\alpha^*(\omega) := ((\hat{\alpha}^*(\omega))^T, 0, \ldots)^T$ is a scaled KKT point of problem (2.2) in the infinite-dimensional space $L_\beta^\rho$.

Let the random field defined by a scaled KKT point $\alpha^* \in L_\beta^\rho$ of problem (2.2) be

$$T^*(\omega, x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m}^*(\omega)Y_{l,m}(x), \quad (5.1)$$

where $\alpha_{l,m}^*(\omega) = 0$, $l = L_\omega + 1, \ldots$, $m = -l, \ldots, l$.

**Lemma 2.** If the random variable $\omega \in \Omega$ has finite second order moment that is $\mathbb{E}[\|\omega\|^2] < \infty$ and there is $\kappa$ such that $\|T^*(\omega_1, x) - T^*(\omega_2, x)\|_{L_2(S^2)} \leq \kappa\|\omega_1 - \omega_2\|$, $\forall \omega_1, \omega_2 \in \Omega$, then $T^*(\omega, x) \in L_2(\Omega \times S^2)$.

**Proof.** Let $\tilde{\omega} \in \Omega$ be fixed. Since for any $\omega \in \Omega$,

$$\|T^*(\omega, x)\|_{L_2(S^2)} - \|T^*(\tilde{\omega}, x)\|_{L_2(S^2)} \leq \|T^*(\omega, x) - T^*(\tilde{\omega}, x)\|_{L_2(S^2)} \leq \kappa\|\omega - \tilde{\omega}\|,$$

we have

$$\|T^*(\omega, x)\|_{L_2(S^2)}^2 \leq \kappa^2\|\omega - \tilde{\omega}\|^2 + 2\|T^*(\tilde{\omega}, x)\|_{L_2(S^2)}^2.$$

Hence, we obtain

$$\mathbb{E}[\|T^*(\omega, x)\|_{L_2(S^2)}^2] \leq 2\|T^*(\tilde{\omega}, x)\|_{L_2(S^2)}^2 + 2\kappa^2\mathbb{E}[\|\omega - \tilde{\omega}\|^2] < \infty,$$

where the last inequality follows from $\mathbb{E}[\|\omega\|^2] < \infty$. Thus, $T^*(\omega, x) \in L_2(\Omega \times S^2)$.

**Theorem 7.** Let $T^0(\omega, x) \in L_2(\Omega \times S^2)$ be the observed random field. Then for any $\epsilon > 0$ there exists $L$ such that

$$0 \leq \|A(T^*_L(\omega, x)) - T^0(\omega, x)\|_{L_2(\Omega \times S^2)}^2 - \varrho < \epsilon,$$

where $T^*_L(\omega, x) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \alpha_{l,m}^*(\omega)Y_{l,m}(x)$.

**Proof.** Since $T^0(\omega, x) \in L_2(\Omega \times S^2)$, by Fubini’s theorem, for $\omega \in \Omega$, $T^0(\omega, x) \in L_2(S^2)$, $P$-a.s., in which case $T^0(\omega, x)$ admits an expansion in terms of spherical harmonics, $P$-a.s., that is, $T^0(\omega, x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha^\text{obs}_{l,m}(\omega)Y_{l,m}(x)$, $P$-a.s., where $\alpha^\text{obs}(\omega) = (\alpha^\text{obs}_{0,0}(\omega), \alpha^\text{obs}_{1,-1}(\omega), \ldots)^T$ is the Fourier coefficient vector of $T^0(\omega, x)$.
By Definition [1] for any $\omega \in \Omega$, $\alpha^*(\omega) \neq 0$, we have $\nu_\omega > 0$. Now, we prove that there exists a positive scalar $\bar{\nu}$ such that $\nu_\omega \geq \bar{\nu}$ for any $\omega \in \Omega$. On the contrary, if there exists a $\omega \in \Omega$ such that $\nu_\omega \to 0$, then from
\[
\|\alpha^*(\omega)\|_{2,p}^2 = \sum_{\{l \in \mathbb{N}_0 : \|\alpha_l(\omega)\|_p \neq 0\}} \beta_l \|\alpha_l^*(\omega)\|^p \geq \left( \frac{p}{2\nu_\omega^2 c} \right)^{\frac{p}{2}} \sum_{\{l \in \mathbb{N}_0 : \|\alpha_l(\omega)\|_p \neq 0\}} (\eta^*)^p \bar{\nu}^{-\frac{p}{2}},
\]
we have $\|\alpha^*(\omega)\|_{2,p}^2 \to \infty$ which is a contradiction with $\alpha^*(\omega) \in \ell_p^0$. Thus, from $\alpha^*(\omega) \in \ell_p^0$ and [2.4], there exists a positive scalar $\bar{\nu}$ such that $\nu_\omega \geq \bar{\nu}$ for any $\omega \in \Omega$. Thus, for any $\epsilon_1 > 0$ there exists $L_1$ such that $\frac{p}{\bar{\nu}} \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_l^*(\omega)\|^p < \frac{\epsilon_1}{2}$ for any $\omega \in \Omega$, which implies that
\[
E \left[ \frac{p}{\nu_\omega} \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_l^*(\omega)\|^p \right] < \frac{p}{\bar{\nu}} E \left[ \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_l^*(\omega)\|^p \right] < \frac{\epsilon_1}{2}.
\]
By Lemma [2] for any $\epsilon_2 > 0$ there exists $L_2$ such that $\sum_{l=L_2+1}^{\infty} E[\|\alpha_l^*(\omega)\|^2] < \frac{\epsilon_2}{2}$. Let $\epsilon = \max\{\epsilon_1, \epsilon_2\}$, $L = \max\{L_1, L_2\}$ and $d = (L + 1)^2$.

For notational simplicity, let $\alpha^*(\omega) = ((\hat{\alpha}^*(\omega))^T, (\hat{\alpha}^*(\omega))^T)^T \in \ell_p^0$, where $\hat{\alpha}^*(\omega) \in \mathbb{C}^d$, $\omega \in \Omega$ and $Y = \begin{bmatrix} \tilde{Y} & \tilde{X} \\ \hat{Y} & \tilde{Y} \end{bmatrix}$, where $\tilde{Y} \in \mathbb{C}^{d \times d}$. Let $\tilde{T}_L^\ast(\omega, x) = \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} \alpha^*_{l,m}(\omega)Y_{l,m}(x)$, we have
\[
\theta = \|A(T^*(\omega, x)) - T^\circ(\omega, x)\|_{L_2(\Omega \times \mathbb{S}^2)}^2 = \|A(T_L^\ast(\omega, x) + \tilde{T}_L^\ast(\omega, x)) - T^\circ(\omega, x)\|_{L_2(\Omega \times \mathbb{S}^2)}^2
\]
\[
= \|A(T_L^\ast(\omega, x)) - T^\circ(\omega, x)\|_{L_2(\Omega \times \mathbb{S}^2)}^2 + E \left[ \int_{\mathbb{S}^2} |A(\tilde{T}_L^\ast(\omega, x))|^2d\sigma(x) + 2 \int_{\mathbb{S}^2} (A(T_L^\ast(\omega, x)) - T^\circ(\omega, x))A(\tilde{T}_L^\ast(\omega, x))d\sigma(x) \right]
\]
\[
= \|A(T_L^\ast(\omega, x)) - T^\circ(\omega, x)\|_{L_2(\Omega \times \mathbb{S}^2)}^2 + \frac{p}{\bar{\nu}} E \left[ \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_l^*(\omega)\|^p \right]
\]
\[
+ E \left[ \int_{\Gamma} |\tilde{T}_L^\ast(\omega, x)|^2d\sigma(x) + 2 \int_{\Gamma} (T_L^\ast(\omega, x) - T^\circ(\omega, x))\tilde{T}_L^\ast(\omega, x)d\sigma(x) \right]
\]
\[
= \|A(T_L^\ast(\omega, x)) - T^\circ(\omega, x)\|_{L_2(\Omega \times \mathbb{S}^2)}^2 + E[\|\hat{\alpha}^*(\omega)\|^2 \tilde{X}^H \tilde{Y} X^H \hat{\alpha}^*(\omega) - 2(\hat{\alpha}^*(\omega))^H \tilde{X}^H \tilde{Y} Y^H \hat{\alpha}^*(\omega)]
\]
where the first equality follows from the second equality in [2.3] with $\nu_\omega > 0$ for any $\omega \in \Omega$. For any fixed $\omega \in \Omega$, by Definition [1] for a scaled KKT point $\alpha^* \in \ell_p^0$ of problem (2.2), there is $\nu_\omega > 0$ such that
\[
p\beta_l \|\alpha_l^*(\omega)\|^{p-2}\alpha_l^*(\omega) + 2\nu_\omega(Y_l, \alpha^*(\omega) - \alpha_l^0(\omega)) = 0, \quad l \in \gamma_\omega.
\]
From (2.1), for $\omega \in \Omega$, $\alpha^\circ(\omega) = Y^H \alpha^\text{obs}(\omega)$. Thus, for any fixed $\omega \in \Omega$,
\[
p\beta_l \|\alpha_l^*(\omega)\|^p + 2\nu_\omega(\alpha_l^*(\omega))^H(Y_l, \alpha^*(\omega) - Y_l \alpha^\text{obs}(\omega)) = 0, \quad l \in \gamma_\omega.
\]
We obtain
\[
\sum_{l=L+1}^{\infty} p\beta_l \|\alpha_l^*(\omega)\|^p + 2\nu_\omega(\tilde{\alpha}^*(\omega))^H(X^H \tilde{Y})(\alpha^*(\omega) - \alpha^\text{obs}(\omega)) = 0, \omega \in \Omega.
\]
Thus,
\[
\mathbb{E}[(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega)] + \mathbb{E}[2(\tilde{\alpha}^*(\omega))^H X^H \tilde{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H X^H \tilde{Y} | \alpha^{\text{obs}}(\omega)]
\]
\[
= \mathbb{E}\left[ (\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega) + 2(\tilde{\alpha}^*(\omega))^H X^H \tilde{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H X^H \tilde{Y} | \alpha^*(\omega) \right] - \mathbb{E}\left[ \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha^*_l(\omega)\|^p \right]
\]
\[
= \mathbb{E}\left[ (\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega) - \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha^*_l(\omega)\|^p \right]
\]
\[
\geq \mathbb{E}\left[ \sum_{l=L+1}^{\infty} \|\alpha^*_l(\omega)\|^2 - \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha^*_l(\omega)\|^p \right] > -\epsilon,
\]
where the first equality follows from (5.4) and the first inequality follows from that \(\|\tilde{Y}\| \leq 1\).
Combining (5.2) and (5.5), we obtain
\[
0 \leq \|A(T^*_L(\omega, x)) - T^0(\omega, x)\|_{L^2(\Omega \times \mathbb{R}^2)}^2 - \varrho < \epsilon.
\]
The proof is completed. \(\square\)

6 Numerical experiments

In this section, we conduct numerical experiments to compare the \(\ell_p-\ell_2\) optimization model (2.5) with the \(\ell_1\) optimization model (31) in [32] on the inpainting of banded-limited random fields and images from Earth topography data to show the efficiency of problem (2.5) and Algorithm 4.2.

Following [8], we adapt the nonmonotone proximal gradient (NPG) method to solve subproblem (4.7) in Algorithm 4.2. For completeness, we present the NPG method as follows.

**NPG method for problem (4.7)**

Given \(\alpha^0 \in F_\nu\). Choose \(M_{\text{max}} \geq M_{\text{min}} > 0\), \(\bar{\eta} > 1\), \(b > 0\) and an integer \(N \geq 0\). Set \(n = 0\).

1. Choose \(M^0_n \in [M_{\text{min}}, M_{\text{max}}]\). Set \(M_n = M^0_n\).
   a. Solve the subproblem
   \[
   y \in \arg\min_{\alpha \in \mathbb{C}^d} \left\{ \Phi(\alpha) + 2\text{Re}\langle \partial_\alpha f_{\lambda,\mu}(\alpha^n), \alpha - \alpha^n \rangle + M_n \|\alpha - \alpha^n\|^2 \right\}.
   \]
   b. If \(F_{\lambda,\mu}(y) \leq \max_{\lceil n/N \rceil + 1 \leq j \leq n} F_{\lambda,\mu}(\alpha^j) - b \|y - \alpha^n\|^2\) is satisfied, go to (2). Otherwise set \(M_n = \bar{\eta} M_n\), and go to step (a).
2. Set \(\alpha^{n+1} = y\), \(n = n + 1\) and go to (1).

For the NPG method to solve (4.7) in Algorithm 4.2 at \(\lambda = \lambda_k\) \(\mu = \mu_k\), we set \(M_{\text{min}} = 1\), \(M_{\text{max}} = 10^6\), \(\bar{\eta} = 2\), \(b = 10^{-4}\), \(N = 4\), \(M^0_0 = 1\), and for any \(n \geq 1\),
\[
M^0_n = \min \left\{ \max \left\{ \frac{|(\alpha^n - \alpha^{n-1})^H (\partial_\alpha f_{\lambda,\mu}(\alpha^n) - \partial_\alpha f_{\lambda,\mu}(\alpha^{n-1}))|}{\|\alpha^n - \alpha^{n-1}\|^2}, 1 \right\}, 10^6 \right\}.
\]
Algorithm 6 is terminated when
\[ \|\alpha^n - \alpha^{n-1}\|_\infty \leq \sqrt{\varepsilon^k} \quad \text{and} \quad \frac{|F_{\lambda,\mu}(\alpha^n) - F_{\lambda,\mu}(\alpha^{n-1})|}{\max\{1, |F_{\lambda,\mu}(\alpha^n)|\}} \leq \min\{(\varepsilon^k)^{2,2}, 10^{-4}\}. \]

In Algorithm 4.2 we set \( \lambda^0 = 20, \mu^0 = \varepsilon^0 = 1, \varsigma_1 = 2, \varsigma_2 = \frac{1}{2}, \alpha^0 = \hat{Y}^{-1}\hat{\alpha}^\circ \). The smoothing penalty algorithm is terminated when \( \max\{g(\alpha^k) + 0.01\varepsilon^k\} \leq 10^{-6} \), where \( \varepsilon^k \) is updated by \( \varepsilon^{k+1} = \max\{2\varepsilon^k, 10^{-6}\} \) instead of \( 2\varepsilon^k \) in the experiments. All codes were written in MATLAB and the realizations were implemented in Python.

### 6.1 Random data

In the following experiments, we consider randomly generated instances which are generated as follows. First, we choose a subset \( D \subset \{0,1,\ldots,L-1\} \) and generate a group sparse coefficient vector \( \alpha_L^{\text{true}} \in \mathbb{C}^{(L+1)^2 \times (L+1)^2} \) such that \( \alpha_L^{\text{true}} = 0 \) if \( l \in D \) and \( \alpha_L^{\text{true}} = \alpha_L^{\text{cmb}} / \|\alpha_L^{\text{cmb}}\|^{1.5} \) if \( l \in D^c \), where \( \alpha_L^{\text{cmb}} \) is the coefficient vector with maximum degree \( L \) of the CMB 2018 map computed by the HEALPy package \[15\].

We assume that the noise \( \Delta \) also has the K-L expansion and set \( \varrho = \|\Delta\|^{2,2}(\mathbb{S}^2) \). We generate the data for the noise on the HEALPix points with \( N_{\text{side}} = 2048 \) by the MATLAB command: \( \text{randn}(N_{\text{pix}},1) \), where \( N_{\text{pix}} = 12 \times N_{\text{side}}^2 \) and \( \delta > 0 \) is a scaling parameter. Then we use the Python HEALPy package to compute the coefficients of the noise. The masks denoted by \( \Gamma^c = \mathbb{S}^2 \setminus \Gamma \) are shown in Figure 1.

![Masks](Image)

**Figure 1:** Masks (grey part).

In the experiments, we set \( p = 0.5, \beta_l = (1 + 10^{-4})^l \) for \( l \geq 1 \) and \( \beta_0 = 1 \). We compare the \( \ell_p \)-\( \ell_2 \) optimization model \[2.5\] by using Algorithm 4.2 with the \( \ell_1 \) optimization model \[31\] in \[32\] using the YALL1 method \[33\] under the KKM sampling scheme \[21\]. We also compare our method with the group SPG1 method (https://friedlander.io/spgl1/). We present the results in Tables 1 and 2. Following \[9\], \( \text{nnz} := \|\alpha_L^* & \alpha_L^{true}\|_{2,0} \) denotes the number of nonzero groups that \( \alpha_L^* \) and \( \alpha_L^{true} \) have in common and \( \text{false} := \|\alpha_L^*\|_{2,0} - \|\alpha_L^* & \alpha_L^{true}\|_{2,0} \) denotes the number of "false positives" where \( \alpha_L^{true} \) is a zero vector, but \( \alpha_L^* \) is a nonzero vector. The signal-to-noise ratio (SNR) and relative error are defined by

\[
\text{SNR} = 20 \times \log_{10} \frac{\|x^{\text{true}}\|}{\|x^{true} - x^*\|}, \quad \text{RelErr} := \frac{\|x^* - \alpha_L^{true}\|}{\|\alpha_L^{true}\|},
\]

respectively, where \( x^{true} = (T^{true}(x_1), \ldots, T^{true}(x_n))^T \) and \( x^* = (T^*(x_1), \ldots, T^*(x_n))^T \) are estimated on the HEALPix points with \( n = 12 \times 2048^2 \) and \( \alpha_L^* \) is the terminating solution.
From Tables 1 and 2, we can see that our optimization model (2.5) using Algorithm 4.2 achieves smaller relative errors and higher SNR values than the $\ell_1$ optimization model (31) in [32]. Although the group SPG11 method archives small relative errors for some experiments, the SNR values is smaller than ours. Moreover, compared with the $\ell_1$ optimization model and group SPG11, our optimization model (2.5) can exactly recover the number and positions of nonzero groups of $\alpha^*_L$.

We select some numerical results from Table 1 with $L = 50$ and $\|\alpha^*_L\|_{2,0} = 26$ to show the inpainting quality in Figure 2. We observe that our model using Algorithm 4.2 achieves smaller pointwise errors than the $\ell_1$ optimization model (31) in [32] and group SPG11 method.

| Mask $\|\alpha^*_L\|_{2,0}$ | Algorithm 4.2 | YALL1 | Group SPG11 |
|-----------------------------|----------------|--------|-------------|
| $\ell_1$ $\|\alpha^*_L\|_{2,0}$ mnn false | RelErr $\|\alpha^*_L\|_{2,0}$ mnn false | SNR | RelErr $\|\alpha^*_L\|_{2,0}$ mnn false | SNR | RelErr $\|\alpha^*_L\|_{2,0}$ mnn false | SNR |
| $\Gamma_1$ $L = 35$ |
| 7 | 0.0017 | 7 | 7 | 0 | 55.54 | 0.651 | 14 | 6 | 8 | 5.82 | 0.0023 | 12 | 7 | 5 | 52.65 |
| 11 | 0.0012 | 11 | 11 | 0 | 58.28 | 0.575 | 13 | 9 | 4 | 4.80 | 0.0788 | 23 | 11 | 12 | 56.19 |
| 19 | 0.0013 | 19 | 19 | 0 | 57.73 | 0.374 | 25 | 17 | 8 | 8.55 | 0.3965 | 29 | 19 | 10 | 46.25 |
| $\Gamma_2$ $L = 35$ |
| 7 | 0.0023 | 7 | 7 | 0 | 52.47 | 0.668 | 19 | 7 | 12 | 3.08 | 0.0043 | 17 | 7 | 10 | 47.29 |
| 11 | 0.0022 | 11 | 11 | 0 | 53.19 | 0.651 | 25 | 11 | 14 | 2.67 | 0.0788 | 32 | 11 | 21 | 22.07 |
| 19 | 0.0059 | 19 | 19 | 0 | 44.54 | 0.683 | 26 | 19 | 7 | 1.87 | 0.3965 | 33 | 19 | 14 | 8.03 |
| $\Gamma_3$ $L = 35$ |
| 7 | 0.0019 | 7 | 7 | 0 | 54.60 | 0.378 | 21 | 7 | 14 | 8.50 | 0.0035 | 16 | 7 | 9 | 49.01 |
| 11 | 0.0014 | 11 | 11 | 0 | 57.30 | 0.357 | 13 | 11 | 2 | 8.89 | 0.0024 | 26 | 11 | 15 | 52.57 |
| 19 | 0.0015 | 19 | 19 | 0 | 56.28 | 0.321 | 25 | 19 | 6 | 9.70 | 0.1195 | 34 | 19 | 15 | 18.45 |
| $\Gamma_4$ $L = 35$ |
| 7 | 0.0017 | 7 | 7 | 0 | 55.33 | 0.351 | 16 | 6 | 10 | 8.91 | 0.0022 | 12 | 7 | 5 | 52.65 |
| 11 | 0.0013 | 11 | 11 | 0 | 57.72 | 0.321 | 11 | 9 | 2 | 10.78 | 0.0019 | 23 | 11 | 12 | 54.29 |
| 19 | 0.0014 | 19 | 19 | 0 | 56.84 | 0.273 | 22 | 17 | 5 | 11.25 | 0.0073 | 32 | 19 | 13 | 42.76 |
| $\Gamma_1$ $L = 50$ |
| 6 | 0.0027 | 6 | 6 | 0 | 51.43 | 0.651 | 28 | 5 | 23 | 5.18 | 0.0030 | 13 | 6 | 7 | 50.40 |
| 16 | 0.0024 | 16 | 16 | 0 | 52.43 | 0.591 | 33 | 13 | 20 | 4.57 | 0.0024 | 33 | 16 | 17 | 52.29 |
| 26 | 0.0031 | 26 | 26 | 0 | 50.21 | 0.511 | 35 | 23 | 12 | 5.84 | 0.0038 | 37 | 26 | 11 | 48.50 |
| $\Gamma_2$ $L = 50$ |
| 6 | 0.0032 | 6 | 6 | 0 | 49.84 | 0.676 | 39 | 6 | 33 | 2.83 | 0.0044 | 15 | 6 | 9 | 47.04 |
| 16 | 0.0044 | 16 | 16 | 0 | 47.15 | 0.643 | 43 | 16 | 27 | 2.68 | 0.0732 | 45 | 16 | 29 | 22.74 |
| 26 | 0.0090 | 26 | 26 | 0 | 40.93 | 0.664 | 24 | 26 | 16 | 2.73 | 0.3549 | 47 | 26 | 21 | 8.99 |
| $\Gamma_3$ $L = 50$ |
| 6 | 0.0024 | 6 | 6 | 0 | 51.43 | 0.621 | 41 | 6 | 35 | 13.21 | 0.0036 | 14 | 6 | 8 | 48.89 |
| 16 | 0.0024 | 16 | 16 | 0 | 52.43 | 0.191 | 37 | 16 | 21 | 14.61 | 0.0041 | 38 | 16 | 22 | 47.72 |
| 26 | 0.0031 | 26 | 26 | 0 | 50.21 | 0.298 | 38 | 26 | 12 | 10.50 | 0.1393 | 48 | 26 | 22 | 17.11 |
| $\Gamma_4$ $L = 50$ |
| 6 | 0.0027 | 6 | 6 | 0 | 51.25 | 0.197 | 37 | 6 | 31 | 16.42 | 0.0031 | 14 | 6 | 8 | 50.14 |
| 16 | 0.0023 | 16 | 16 | 0 | 52.47 | 0.239 | 37 | 16 | 21 | 14.21 | 0.0030 | 32 | 16 | 16 | 50.51 |
| 26 | 0.0030 | 26 | 26 | 0 | 50.55 | 0.239 | 37 | 25 | 12 | 13.05 | 0.0078 | 44 | 26 | 18 | 42.21 |
Table 2: Numerical results for the inpainting of band-limited random fields with $\delta = 0.1$.

| Mask | $\|\alpha_2^{true}\|_2$ | Algorithm 4.2 | YALL1 | Group SPGII |
|------|----------------|----------------|--------|-------------|
|      | $\|\alpha_2^{true}\|_2$, nnz false | RelErr $\|\alpha_2^{true}\|_2$, nnz false | RelErr $\|\alpha_2^{true}\|_2$, nnz false | RelErr $\|\alpha_2^{true}\|_2$, nnz false | SNR | SNR |
| $\Gamma_1^+$ | 7 | 1.90e-4 | 7 | 7 | 0 | 74.42 | 0.551 | 6 | 6 | 0 | 5.83 | 2.88e-4 | 11 | 7 | 4 | 70.78 |
| | 11 | 1.27e-4 | 11 | 11 | 0 | 77.89 | 0.591 | 12 | 9 | 3 | 4.80 | 1.61e-4 | 21 | 11 | 10 | 75.82 |
| | 19 | 1.39e-4 | 19 | 19 | 0 | 77.13 | 0.511 | 25 | 17 | 8 | 8.55 | 0.0035 | 33 | 19 | 14 | 49.18 |
| $\Gamma_1^0$ | 7 | 2.58e-4 | 7 | 7 | 0 | 71.73 | 0.676 | 19 | 19 | 0 | 3.08 | 5.24e-4 | 17 | 7 | 10 | 65.60 |
| | 11 | 2.35e-4 | 11 | 11 | 0 | 72.57 | 0.643 | 21 | 11 | 10 | 2.67 | 0.0859 | 34 | 11 | 23 | 21.32 |
| | 19 | 6.89e-4 | 19 | 19 | 0 | 63.22 | 0.664 | 10 | 7 | 3 | 1.87 | 0.3949 | 33 | 19 | 14 | 8.06 |
| $\Gamma_1^-$ | 7 | 2.08e-4 | 7 | 7 | 0 | 73.60 | 0.212 | 16 | 7 | 9 | 8.50 | 4.05e-4 | 15 | 7 | 8 | 67.84 |
| | 11 | 1.40e-4 | 11 | 11 | 0 | 77.05 | 0.191 | 14 | 11 | 3 | 8.89 | 2.54e-4 | 25 | 11 | 14 | 71.89 |
| | 19 | 1.70e-4 | 19 | 19 | 0 | 75.37 | 0.298 | 21 | 19 | 2 | 9.70 | 0.1172 | 35 | 19 | 16 | 18.62 |
| $\Gamma_2^+$ | 7 | 1.97e-4 | 7 | 7 | 0 | 74.09 | 0.197 | 11 | 6 | 5 | 8.91 | 2.65e-4 | 11 | 7 | 4 | 71.50 |
| | 11 | 1.28e-4 | 11 | 11 | 0 | 77.83 | 0.193 | 9 | 9 | 0 | 10.78 | 1.93e-4 | 22 | 11 | 11 | 74.25 |
| | 19 | 1.56e-4 | 19 | 19 | 0 | 76.13 | 0.239 | 20 | 17 | 3 | 11.26 | 0.0046 | 35 | 19 | 16 | 46.66 |
| $\Gamma_2^0$ | 6 | 2.56e-4 | 6 | 6 | 0 | 71.81 | 0.551 | 13 | 5 | 8 | 5.18 | 3.10e-4 | 11 | 6 | 5 | 70.17 |
| | 16 | 2.15e-4 | 16 | 16 | 0 | 73.32 | 0.591 | 18 | 13 | 5 | 4.57 | 2.67e-4 | 27 | 16 | 11 | 71.44 |
| | 26 | 2.46e-4 | 26 | 26 | 0 | 72.17 | 0.511 | 35 | 22 | 12 | 5.84 | 3.94e-4 | 35 | 26 | 9 | 68.08 |
| $\Gamma_2^-$ | 6 | 3.18e-4 | 6 | 6 | 0 | 69.92 | 0.676 | 16 | 6 | 10 | 2.83 | 4.48e-4 | 16 | 6 | 10 | 66.96 |
| | 16 | 5.09e-4 | 16 | 16 | 0 | 65.85 | 0.643 | 36 | 16 | 20 | 2.68 | 0.0829 | 45 | 16 | 29 | 21.63 |
| | 26 | 9.97e-4 | 26 | 26 | 0 | 60.01 | 0.664 | 32 | 26 | 6 | 2.73 | 0.3562 | 47 | 26 | 21 | 8.98 |
| $\Gamma_3^+$ | 6 | 2.97e-4 | 6 | 6 | 0 | 70.52 | 0.212 | 30 | 6 | 24 | 13.21 | 3.84e-4 | 13 | 6 | 7 | 68.31 |
| | 16 | 2.70e-4 | 16 | 16 | 0 | 71.34 | 0.191 | 24 | 16 | 8 | 14.61 | 4.59e-4 | 34 | 16 | 18 | 66.75 |
| | 26 | 3.28e-4 | 26 | 26 | 0 | 69.65 | 0.298 | 31 | 26 | 5 | 10.50 | 0.1387 | 50 | 26 | 24 | 17.15 |
| $\Gamma_3^0$ | 6 | 2.86e-4 | 6 | 6 | 0 | 70.85 | 0.197 | 23 | 6 | 17 | 16.42 | 3.43e-4 | 11 | 6 | 5 | 69.29 |
| | 16 | 2.56e-4 | 16 | 16 | 0 | 71.82 | 0.193 | 24 | 16 | 8 | 14.21 | 3.27e-4 | 28 | 16 | 12 | 69.69 |
| | 26 | 3.09e-4 | 26 | 26 | 0 | 70.19 | 0.239 | 29 | 25 | 4 | 13.05 | 8.06e-4 | 43 | 26 | 17 | 61.86 |
(a) Observed fields with $L = 50$, $\|\alpha^\text{true}\|_{2,0} = 6$ and $\delta = 1$ (masked by $\Gamma_1^c$, $\Gamma_2^c$, $\Gamma_3^c$ and $\Gamma_4^c$ from left to right)

(b) Inpainted fields by Algorithm 4.2

(c) Pointwise error of inpainted fields in (b) by Algorithm 4.2

(d) Inpainted fields by YALL1

(e) Pointwise error of inpainted fields in (d) by YALL1

(f) Inpainted fields by group SPGl1

(g) Pointwise error of inpainted fields in (f) by group SPGl1

Figure 2: Inpainting results.
To give some insight for the choice of $L$ of the truncated space, we conduct experiments on a random field with different $L$ under the noiseless case. We choose the mask $\Gamma$ and the observed field $T^o$ with degree 50. In Figure 3(a) we show the approximation error $\|A(T^*_L) - T^o\|^2_{L_2(S^2)}$ on a logarithmic scale. We can see that the error decreases as $L$ increases and the errors are same for $L = 48, 49, 50$. Moreover, $\alpha^*_49$, and $\alpha^*_50$ are zero groups of $\alpha^*_L$ for $L = 50$. From our discussion in Section 2 and Theorem 5.2, we can guess that the value of $L$ for (a) is 48. We plot the error $\|T^*_50 - T^*_L\|^2_{L_2(S^2)}$ for different $L$ in Figure 3(b). We can observe that the error decreases as $L$ increases and the errors are zero for $L = 48, 49, 50$ due to $\alpha^*_49 = 0$ and $\alpha^*_50 = 0$. In Figure 3(c) and (d), we plot the values of $\|\alpha^*_L\|$ of $T^*_L$ with different $L$ to give some insight of the sparsity pattern of $\alpha^*$.

![Figure 3](image1.png)

(a) $\|A(T^*_L) - T^o\|^2_{L_2(S^2)}$

(b) $\|T^*_50 - T^*_L\|^2_{L_2(S^2)}$

(c) $\|\alpha^*_L\|$  

(d) $\|\alpha^*_L\|$  

Figure 3: Approximation errors and sparsity pattern.

### 6.2 Real image

We consider the inpainting of images from Earth topographic data, taken from Earth Gravitational Model (EGM2008) publicly released by the U.S. National Geospatial-Intelligence Agency (NGA) EGM Development Team. We generate four resolution test images from the Earth topography data which are shown in Figures 5(a), 6(a), 7(a) and 8(a). For each resolution test image we consider a mask which is shown in Figures 5(d), 6(d), 7(d), and 8(d) respectively. For all the images we set $\delta = 1$. The recovered images and pointwise errors are presented in Figures 5, 6, 7 and 8. We can see that model (2.5) using Algorithm 4.2 yields higher quality images with smaller pointwise errors and higher SNR values than the $\ell_1$ optimization model (31) in [32] and the group SPGL1 method under different resolutions.

\[^{1}\text{http://geoweb.princeton.edu/people/simons/DOTM/Earth.mat}\]
Figure 4: Test images of Earth topographic data constructed to be band-limited at $L = 35$. The SNR of (b) is 97.30 and the SNR of (c) is 1.81 and the SNR of (d) is 13.04.

Figure 5: Test images of Earth topographic data constructed to be band-limited at $L = 35$. The SNR of (b) is 97.30 and the SNR of (c) is 1.81 and the SNR of (d) is 13.04.

7 Conclusion

In this paper, we propose a constrained group sparse optimization model (1.6) for the inpainting of random fields on the unit sphere with unique continuation property. Based on the K-L expansion, we rewrite the constraint in (1.6) in a discrete form and derive an equivalence formulation (2.2) of problem (1.6). For problem (2.2), we derive a lower bound (2.4) for the $\ell_2$ norm of nonzero groups of its scaled KKT points. We show that problem (2.2) is equivalent to the finite-dimensional problem (2.5). For this finite-dimensional problem (2.5) and its penalty problem (2.6), we prove the exact penalization in terms of local minimizers and $\varepsilon$-minimizers. Moreover, we propose a smoothing penalty algorithm for solving problem (2.5) and prove its convergence. We also present the approximation error of the inpainted random field represented by the scaled KKT point in the
Figure 6: Test images of Earth topographic data constructed to be band-limited at $L = 50$. The SNR of (b) is 90.47, the SNR of (c) is 6.44 and the SNR of (d) is 8.07.

Figure 7: Test images of Earth topographic data constructed to be band-limited at $L = 80$. The SNR of (b) is 82.48, the SNR of (c) is 14.70 and the SNR of (d) is 18.04.

space $L_2(\Omega \times S^2)$. Finally, we present numerical results on band-limited random fields on the sphere and the images from Earth topographic data to show the promising performance of our model by using the smoothing penalty algorithm.

References

[1] P. Abrial, Y. Moudden, J.L. Starck, B. Afeyan, J. Bobin, J. Fadili, and M.K. Nguyen. Morphological component analysis and inpainting on the sphere: application in physics and astrophysics. J. Fourier Anal. Appl., 13(6):729–748, 2007.

[2] S. Axler, P. Bourdon, and R. Wade. Harmonic Function Theory, volume 137. Springer Science & Business Media, 2013.
Figure 8: Test images of Earth topographic data constructed to be band-limited at $L = 128$. The SNR of (b) is 73.18, the SNR of (c) is 21.60 and the SNR of (d) is 1.65.

[3] Amir Beck and Nadav Hallak. Optimization problems involving group sparsity terms. *Math. Program.*, 178(1):39–67, 2019.

[4] D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings H - Microwaves, Optics and Antennas*, 130(1):11–16, 1983.

[5] P. Cabella and D. Marinucci. Statistical challenges in the analysis of cosmic microwave background radiation. *Ann. Appl. Stat.*, 3:61–95, 2009.

[6] V. Cammarota and D. Marinucci. The stochastic properties of $l_1$-regularized spherical Gaussian fields. *Appl. Comput. Harmon. Anal.*, 38(2):262–283, 2015.

[7] X. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.*, 134(1):71–99, 2012.

[8] X. Chen, Z. Lu, and T.K. Pong. Penalty methods for a class of non-Lipschitz optimization problems. *SIAM J. Optim.*, 26(3):1465–1492, 2016.

[9] X. Chen and R.S. Womersley. Spherical designs and nonconvex minimization for recovery of sparse signals on the sphere. *SIAM J. Imaging Sci.*, 11(2):1390–1415, 2018.

[10] X. Chen, F. Xu, and Y. Ye. Lower bound theory of nonzero entries in solutions of $l_2$-$l_p$ minimization. *SIAM J. Sci. Comput.*, 32(5):2832–2852, 2010.

[11] Xiaojun Chen and Ph L Toint. High-order evaluation complexity for convexly-constrained optimization with non-lipschitzian group sparsity terms. *Math. Program.*, 187(1):47–78, 2021.

[12] P.E. Creasey and A. Lang. Fast generation of isotropic Gaussian random fields on the sphere. *Monte Carlo Methods Appl.*, 24(1):1–11, 2018.
[13] S.M. Feeney, D. Marinucci, J.D. McEwen, H.V. Peiris, B.D. Wandelt, and V. Cammarota. Sparse inpainting and isotropy. *J. Cosmol. Astropart. Phys.*, 2014(1):050, 2014.

[14] Q.T. Le Gia, I.H. Sloan, R.S. Womersley, and Y.G. Wang. Sparse isotropic regularization for spherical harmonic representations of random fields on the sphere. *Appl. Comput. Harmon. Anal.*, 49:257–278, 2019.

[15] K.M. Górski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, and M. Bartelmann. Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.*, 622(2):759–771, 2005.

[16] E. Hille. *Analytic Function Theory*, volume 2. American Mathematical Soc., 2005.

[17] Jian Huang, Shuange Ma, Huiliang Xie, and Cun-Hui Zhang. A group bridge approach for variable selection. *Biometrika*, 96(2):339–355, 2009.

[18] Junzhou Huang and Tong Zhang. The benefit of group sparsity. *Ann. Statist.*, 38(4):1978–2004, 2010.

[19] V. Isakov. *Inverse Problems for Partial Differential Equations*, volume 127. Springer, 2006.

[20] J. Jeong, M. Jun, and M. Genton. Spherical process models for global spatial statistics. *Stat. Sci.*, 32(4):501–513, 2017.

[21] Z. Khalid, R.A. Kennedy, and J.D. McEwen. An optimal-dimensionality sampling scheme on the sphere with fast spherical harmonic transforms. *IEEE Trans. Signal Process.*, 62(17):4597–4610, 2014.

[22] K. Kreutz-Delgado. The complex gradient operator and the CR-calculus. *arXiv:0906.4835*, 2009.

[23] A. Lang and C. Schwab. Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.*, 25(6):3047–3094, 2015.

[24] C. Li and X. Chen. Isotropic non-Lipschitz regularization for sparse representations of random fields on the sphere. *Math. Comp.*, 91:219–243, 2022.

[25] D. Marinucci and G. Peccati. *Random Fields on the Sphere: Representations, Limit Theorems and Cosmological Applications*. Cambridge University Press, Cambridge, 2011.

[26] H.S. Oh and T.H. Li. Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method. *J. R. Statist. Soc. B*, 66(1):221–238, 2004.

[27] Lili Pan and Xiaojun Chen. Group sparse optimization for images recovery using capped folded concave functions. *SIAM J. Imaging Sci.*, 14(1):1–25, 2021.
Appendix A. Unique continuation property

In this Appendix, we give details about the unique continuation property of a class of fields with spherical harmonic representations.

Let \( \hat{r} \in (1, \eta) \) be a positive constant for some \( \eta > 1 \), \( B(0; \hat{r}) := \{(r, \theta, \phi) : r \in [0, \hat{r}), \theta \in [0, \pi], \phi \in [0, 2\pi]\} \subset \mathbb{R}^3 \) be an open ball centered at 0 of radius \( \hat{r} \). Let

\[
\mathcal{T} = \left\{ T(\theta, \phi) \in L_2(S^2) : T(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\theta, \phi) \text{ and } \alpha \in \ell_\beta^p \right\}^2
\]  

be a class of fields with spherical harmonic representations and coefficients belong to \( \ell_\beta^p \). Then we show that any \( T \in \mathcal{T} \) has the unique continuation property. Let \( H : B(0; \hat{r}) \to \mathbb{R} \) be given by

\[
H(r, \theta, \phi) = \sum_{l=0}^{\infty} r^l \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\theta, \phi),
\]

where \( \alpha \in \ell_\beta^p \).

**Lemma 3.** If \( \alpha \in \ell_\beta^p \), then \( \sum_{l=0}^{\infty} \eta^l \| \alpha_l \| < \infty \).

\footnote{With a slight abuse of notation, in this subsection, \( T(\theta, \phi) \equiv T(x) \) and \( Y_{l,m}(\theta, \phi) \equiv Y_{l,m}(x) \), where \( x = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T \in S^2 \).}
Proof. Since $\alpha \in \ell^p_\beta$ and $\sum_{l=0}^{\infty} \eta^l = \infty$, there exists an integer $N$ such that for any $l > N$, $\eta^l \|\alpha_l\|^p < \eta^l$. Then, we obtain, for any $l > N$, $l\|\alpha_l\| \leq \eta^l \|\alpha_l\|^p < 1$ which implies that $\eta^l \|\alpha_l\| \leq \eta^l \|\alpha_l\|^p$. Thus, $\sum_{l=0}^{\infty} \eta^l \|\alpha_l\| < \infty$. This completes the proof. \hfill \qed

**Lemma 4.** $H$ is real analytic in $B(0; \bar{r})$.

Proof. Let $H_L(r, \theta, \phi) = \sum_{l=0}^{L} H_l(r, \theta, \phi)$, where $H_l(r, \theta, \phi) = r^l \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\theta, \phi)$. By definition of harmonic functions \cite{2}, $r^l Y_{l,m}(\theta, \phi)$, $l \in \mathbb{N}_0$, $m = -l, \ldots, l$ are harmonic functions on $B(0; \bar{r})$. Then we obtain that $H_L$, $L \in \mathbb{N}_0$ are harmonic functions on $B(0; \bar{r})$. Moreover, for any $(r, \theta, \phi) \in B(0; \bar{r})$ and $l > 0$,

$$|H_l(r, \theta, \phi)| = \left| \sum_{m=-l}^{l} r^l \alpha_{l,m} Y_{l,m}(\theta, \phi) \right| \leq \left( \sum_{m=-l}^{l} (r^l \alpha_{l,m})^2 \right)^{\frac{1}{2}} \left( \sum_{m=-l}^{l} (Y_{l,m}(\theta, \phi))^2 \right)^{\frac{1}{2}} = r^l \|\alpha_l\| \sqrt{\frac{2l+1}{4\pi}} < \eta^l \|\alpha_l\|,$$

where the first inequality follows from Cauchy-Schwarz inequality, the second equality follows from addition theorem of spherical harmonics and the last inequality follows from $r \in [0, \bar{r})$ and $\bar{r} < \eta$.

By Lemma 3, $\sum_{l=0}^{\infty} \eta^l \|\alpha_l\| < \infty$, then for any $\epsilon > 0$, there exists $N$ such that for any $L > N$, we have that for any $(r, \theta, \phi) \in B(0; \bar{r})$,

$$|H_{L'}(r, \theta, \phi) - H_L(r, \theta, \phi)| = \left| \sum_{l=L+1}^{L'} H_l(r, \theta, \phi) \right| \leq \sum_{l=L+1}^{L'} |H_l(r, \theta, \phi)| < \sum_{l=L+1}^{L'} \eta^l \|\alpha_l\| < \epsilon.$$

Thus, the sequence of harmonic functions $\{H_L\}$ converges uniformly to $H$ on $B(0; \bar{r})$.

By \cite{2} Theorem 1.23, $H$ is harmonic on $B(0; \bar{r})$. Moreover, by \cite{2} Theorem 1.28, $H$ is real analytic in $B(0; \bar{r})$. This completes the proof. \hfill \qed

**Lemma 5.** For any $T \in \mathcal{T}$, if there exists a sequence of distinct points $\{ (\theta_n, \phi_n) \} \subseteq \mathbb{S}^2$, with at least one limit point in $\mathbb{S}^2$, and if $T(\theta_n, \phi_n) = 0$, $n = 1, 2, \ldots$, then $T \equiv 0$ on $\mathbb{S}^2$.

Proof. It is obvious that $H(1, \theta, \phi) = T(\theta, \phi)$, then $T(\theta_n, \phi_n) = 0$, $n = 1, 2, \ldots$ implies that $H(1, \theta_n, \phi_n) = 0$, $n = 1, 2, \ldots$. By Lemma 4, $H$ is real analytic in $B(0; \bar{r})$. Then by \cite{16} Theorem 8.1.3, we obtain $H \equiv 0$ in $B(0; \bar{r})$ which implies that $T \equiv 0$ on $\mathbb{S}^2$. This completes the proof. \hfill \qed

Now we present the unique continuation property of any $T \in \mathcal{T}$.

**Theorem 8.** For any $T \in \mathcal{T}$, if $T = 0$ on $\Gamma$, then $T \equiv 0$ on $\mathbb{S}^2$.

Proof. Since the coefficients $\alpha \in \ell^p_\beta$ and $\Gamma$ has an open subset, we know from Lemma 5 that if $T = 0$ on $\Gamma$ then $T \equiv 0$ on $\mathbb{S}^2$. \hfill \qed

By Parseval’s theorem, for any $T \in L_2(\mathbb{S}^2)$, we have $\|T\|^2_{L_2(\mathbb{S}^2)} = \|\alpha\|^2$, which implies that $T \equiv 0$ if and only if $\alpha = 0$. Hence, we can claim that for any $T \in \mathcal{T}$, $\mathcal{A}(T) \equiv 0$ if and only if $\alpha = 0$. 27
Appendix B. Wirtinger gradient

In this appendix, we briefly introduce the Wirtinger gradient of real-valued functions with complex variables. For more details about Wirtinger calculus, we refer to \[4,22,24\].

Let $f : \mathbb{C}^n \to \mathbb{R}$ be a real-valued function of a vector of complex variables $z \in \mathbb{C}^n$, which can be written as $f_1(z, \bar{z})$, that is $f(z) = f_1(z, \bar{z})$. Let $z = x + iy$, where $x, y \in \mathbb{R}^n$. We say $f(z) = f_2(x + iy)$ is Wirtinger differentiable, if $f_2$ is differentiable with respect to $x$ and $y$, respectively. Following $[4,22]$, under the conjugate coordinates $(z^T, \bar{z}^T)^T \in \mathbb{C}^n \times \mathbb{C}^n$, the Wirtinger gradient of $f$ at $z \in \mathbb{C}^n$ is

$$
\nabla f(z) = \begin{pmatrix}
\partial_z f(z) \\
\partial_{\bar{z}} f(z)
\end{pmatrix},
$$

where $\partial_z f(z) := \frac{\partial f_1(z, \bar{z})}{\partial z} \bigg|_{\bar{z}=\text{const}}$, $\partial_{\bar{z}} f(z) := \frac{\partial f_1(z, \bar{z})}{\partial \bar{z}} \bigg|_{z=\text{const}}$. Since $f$ is real-valued, $\overline{\partial_z f} = \partial_{\bar{z}} f$. Thus,

$$
\nabla f(z) = 0 \iff \partial_{\bar{z}} f(z) = 0.
$$

We say $z^* \in \mathbb{C}^n$ is a stationary point of $f$ if it satisfies

$$
\partial_{\bar{z}} f(z^*) = 0.
$$

It is easy to verify that the Wirtinger gradient of $f(z) = \|z\|^2$, $z \in \mathbb{C}^n$, is $\nabla f(z) = (\bar{z}^T, z^T)^T$, and $z^* \in \mathbb{C}^n$ is a stationary point of $f$ if $z^* = 0$. 
