Group sparse optimization for inpainting of random fields on the sphere

CHAO LI†
School of Mathematics and Statistics, Taiyuan Normal University, Taiyuan, Shanxi, China
Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
AND
XIAOJUN CHEN‡
Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

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 ℓ_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 CMB 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 $(\Omega, \mathscr{F}, P)$ be a probability space, and let $\mathfrak{B}(\mathbb{S}^2)$ be the Borel sigma algebra on the unit sphere $\mathbb{S}^2 := \{\mathrm{x} \in \mathbb{R}^3 : \|\mathrm{x}\| = 1\}$, where $\|\cdot\|$ is the ℓ_2 norm. A real-valued random field on $\mathbb{S}^2$ is an $\mathscr{F} \otimes \mathfrak{B}(\mathbb{S}^2)$-measurable function $T : \Omega \times \mathbb{S}^2 \to \mathbb{R}$, and it is said to be 2-weakly isotropic if the expected value and covariance of T are rotationally invariant (see Gia *et al.* (2019)). It is known that a 2-weakly isotropic random field on the sphere has the following Karhunen-Loève (K-L) expansion (see, for example, Lang & Schwab, 2015; Marinucci & Peccati, 2011),

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

where $\alpha_{l,m}(\omega) = \int_{\mathbb{S}^2} T(\omega, \mathrm{x}) \overline{Y_{l,m}(\mathrm{x})} d\sigma(\mathrm{x}) \in \mathbb{C}$ are random spherical harmonic coefficients of $T(\omega, \mathrm{x})$, $Y_{l,m}$, for $m = -l, \ldots, l$, $l = 0, 1, 2, \ldots$ are spherical harmonics with degree l and order m, and σ is the

†Email: chaoo.li@connect.polyu.hk
‡Corresponding author. Email: xiaojun.chen@polyu.edu.hk

surface measure on $\mathbb{S}^2$ satisfying $\sigma(\mathbb{S}^2)=4\pi$. For convenience, let

$$\alpha(\omega)=(\alpha_{0,0}(\omega),\underbrace{\alpha_{1,-1}(\omega),\alpha_{1,0}(\omega),\alpha_{1,1}(\omega)}_{3},\ldots,\underbrace{\alpha_{l,-l}(\omega),\ldots,\alpha_{l,l}(\omega)}_{2l+1},\ldots)^T$$

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

$$\alpha(\omega)=(\alpha_{0\cdot}^T(\omega),\alpha_{1\cdot}^T(\omega),\ldots,\alpha_{l\cdot}^T(\omega),\ldots)^T, \tag{1.2}$$

where

$$\alpha_{l\cdot}(\omega)=(\alpha_{l,-l}(\omega),\ldots,\alpha_{l,0}(\omega),\ldots,\alpha_{l,l}(\omega))^T\in\mathbb{C}^{2l+1}, \qquad l\geqslant 0.$$

From Gia *et al.* (2019), we know that for any given degree $l\geqslant 1$ and any $\omega\in\Omega$, the sum

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

is rotationally invariant, while $\sum_{m=-l}^{l}|\alpha_{l,m}(\omega)|$ is not invariant under rotation of the coordinate axes. Moreover, in Gia *et al.* (2019) the angular power spectrum is given by

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

if the expected value $\mathbb{E}[T(\omega,\mathrm{x})]=0$ for all $\mathrm{x}\in\mathbb{S}^2$. In the study of isotropic random fields (see, for example, Creasey & Lang, 2018; Lang & Schwab, 2015; Marinucci & Peccati, 2011), 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 $\alpha_{l\cdot}(\omega)$ of $\alpha(\omega)$, i.e., few non-zero entries of the following vector

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

In recent years, sparse representation of isotropic random fields has been extensively studied. In Cammarota & Marinucci (2015) the authors studied the sparse representation of random fields via an ℓ_1-regularized minimization problem. In Gia *et al.* (2019), 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 (see Yuan & Lin, 2006). In Li & Chen (2022), 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 $\|\alpha_{l\cdot}(\omega)\|^2$ for any given degree $l\geqslant 1$ and any $\omega\in\Omega$.

Isotropic random fields on the sphere have many applications (see, for example, Cabella & Marinucci, 2009; Jeong *et al.*, 2017; Oh & Li, 2004; Porcu *et al.*, 2018; Stein, 2007), especially in the study of the Cosmic Microwave Background (CMB) analysis (see, for example, Abrial *et al.*, 2007; Starck *et al.*, 2013; Bucher & Louis, 2012; Gruetjen *et al.*, 2017; Kim *et al.*, 2012). However, the true spherical random field usually presents masked regions and missing observations. Many inpainting methods (see, for example, Abrial *et al.*, 2007; Feeney *et al.*, 2014; Starck *et al.*, 2013; Wallis *et al.*, 2017) have been proposed for recovering the true field based on sparse representation. The group structure (1.2) was also previously considered and associated to an isotropy assumption but with a non sparse penalty (a weighted ℓ_2 norm) in Starck *et al.* (2013).

Group sparse optimization models have been successfully used in signal processing, imaging sciences and predictive analysis (see Beck & Hallak, 2019; Chen & Toint, 2021; Huang & Zhang, 2010; Huang *et al.*, 2009; Pan & Chen, 2021, and references therein). 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 (see, Isakov, 2006) 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 Γ, 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

$$\langle f, g\rangle_{L_2(\mathbb{S}^2)} = \int_{\mathbb{S}^2} f(\mathrm{x})\overline{g(\mathrm{x})}d\sigma(\mathrm{x}), \qquad \forall f, g \in L_2(\mathbb{S}^2),$$

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 $\omega \in \Omega$, $T(\omega, \mathrm{x}) \in L_2(\mathbb{S}^2)$, P-a.s., where "P-a.s." stands for almost surely with probability 1. For brevity, we write $T(\omega, \mathrm{x})$ as $T(\mathrm{x})$ or T if no confusion arises. Let the observed field be given by

$$T^\circ := \mathscr{A}(T^*) + \Delta, \tag{1.3}$$

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 $\mathscr{A} : L_2(\Omega \times \mathbb{S}^2) \to L_2(\Omega \times \mathbb{S}^2)$ is an inpainting operator defined by

$$\mathscr{A}(T(\mathrm{x})) = \begin{cases} T(\mathrm{x}) & \text{if } \mathrm{x} \in \Gamma \\ 0 & \text{if } \mathrm{x} \in \mathbb{S}^2 \setminus \Gamma, \end{cases} \tag{1.4}$$

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° as one realization of a random field. For notational simplicity, let

$$\alpha = (\alpha_{0\cdot}^T, \alpha_{1\cdot}^T, \ldots, \alpha_{l\cdot}^T, \ldots)^T \tag{1.5}$$

denote the spherical harmonic coefficient vector of an isotropic random field T for a fixed $\omega \in \Omega$, where $\alpha_{l\cdot} = (\alpha_{l,-l}, \ldots, \alpha_{l,0}, \ldots, \alpha_{l,l})^T \in \mathbb{C}^{2l+1}$, $l \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$.

By Parseval's theorem, we have

$$\|T\|^2_{L_2(\mathbb{S}^2)} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} |\alpha_{l,m}|^2 = \sum_{l=0}^{\infty} \|\alpha_{l\cdot}\|^2 < \infty.$$

Hence the sequence $\{\|\alpha_{l\cdot}\|\}_{l\in\mathbb{N}_0} \in \ell^2(\mathbb{N}_0)$, where $\ell^2(\mathbb{N}_0)$ is the space of square-summable sequences. We write $\alpha \in \ell^2$ to indicate $\{\|\alpha_{l\cdot}\|\}_{l\in\mathbb{N}_0} \in \ell^2(\mathbb{N}_0)$.

The number of non-zero groups of α with group structure (1.5) is calculated by

$$\|\alpha\|_{2,0} = \sum_{l=0}^{\infty} \|\alpha_{l\cdot}\|^0,$$

where $\|\alpha_{l\cdot}\|^0 = \begin{cases} 1 & \text{if } \|\alpha_{l\cdot}\| > 0 \\ 0 & \text{otherwise,} \end{cases}$ and $\|\alpha\|_{2,0}$ is called $\ell_{2,0}$ norm of α. The $\ell_{2,0}$ norm is discontinuous and $\|T\|^2_{L_2(\mathbb{S}^2)} < \infty$ does not ensure $\|\alpha\|_{2,0} < \infty$. In Gia *et al.* (2019), the authors used a weighted $\ell_{2,1}$ norm of α defined by

$$\|\alpha\|_{2,1} = \sum_{l=0}^{\infty} \beta_l \|\alpha_{l\cdot}\|,$$

where $\beta_l > 0$. It is easy to see that $\lim_{p\downarrow 0} \|\alpha_{l\cdot}\|^p = \|\alpha_{l\cdot}\|^0$. The $\ell_{2,p}$ norm $(0 < p < 1)$ has advantages for finding group sparse solutions and has been widely used in group sparse approximation (see Chen & Toint, 2021; Huang & Zhang, 2010, and references therein). In this paper, we use a weighted $\ell_{2,p}$ norm of α. Let

$$\ell^p_\beta := \left\{ \alpha : \ \|\alpha\|^p_{2,p} := \sum_{l=0}^{\infty} \beta_l \|\alpha_{l\cdot}\|^p < \infty \right\}$$

be a weighted ℓ^p $(0 < p < 1)$ space with positive weights $\beta_0 = 1$ and $\beta_l = \eta^l l^p$ for $l \geqslant 1$, where $\eta > 1$ is a constant. In Appendix A we show that with such choice of β_l, any realization of a random field whose coefficient $\alpha \in \ell^p_\beta$ in the K-L expansion has the unique continuation property. Moreover, from $\beta_l \|\alpha_{l\cdot}\|^p \to 0$ as $l \to \infty$, there is L_0 such that for all $l \geqslant L_0$, $\beta_l \|\alpha_{l\cdot}\|^p < 1$, which implies $\|\alpha_{l\cdot}\| < \beta_l^{-\frac{1}{p}} = l^{-1}(\eta^{-\frac{1}{p}})^l < (\eta^{-\frac{1}{p}})^l$ and $\|\alpha_{l\cdot}\|^2 < \beta_l^{-\frac{2}{p}} = l^{-2}\eta^{-\frac{2l}{p}} < l^{-2}$. Hence we have $\alpha \in \ell_{2,1} := \{\alpha : \sum_{l=0}^{\infty} \|\alpha_{l\cdot}\| < \infty\}$ and $\alpha \in \ell^2$ if $\alpha \in \ell^p_\beta$. Some results for isotropic sparse regularization for spherical harmonic representations of random fields on the sphere, with a hybrid of the norms imposed on the coefficients $\alpha \in \ell_{2,1}$ and $\alpha \in \ell^2$ in Cammarota & Marinucci (2015); Gia *et al.* (2019) can be applied to $\alpha \in \ell^p_\beta$. Moreover, for the CMB experiments, the angular power spectrum $C_l = \frac{1}{2l+1}\mathbb{E}[\|\alpha_{l\cdot}(\omega)\|^2] \leqslant O(l^{-3})$ for sufficiently large l.

For a fixed $\omega \in \Omega$, we consider the following constrained optimization problem

$$\begin{array}{ll} \min & \|T(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} \\ \text{s.t.} & \|\mathscr{A}(T(\mathrm{x})) - T^\circ(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} \leqslant \rho, \end{array} \tag{1.6}$$

where $\rho > \int_{\mathbb{S}^2 \setminus \Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x})$ and

$$T(\mathrm{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\mathrm{x}), \quad \mathrm{x} \in \mathbb{S}^2, \quad \alpha \in \ell^p_\beta.$$

Since $T^\circ \in L_2(\mathbb{S}^2)$, for $\varepsilon = \rho - \int_{\mathbb{S}^2\setminus\Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x})$, there is a finite number L such that $\|T^\circ(\mathrm{x}) - T^\circ_L(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} < \varepsilon$, where $T^\circ_L(\mathrm{x}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} b_{l,m} Y_{l,m}(\mathrm{x})$ and $b_{l,m} = \int_{\mathbb{S}^2} T^\circ(\mathrm{x}) \overline{Y_{l,m}(\mathrm{x})} d\sigma(\mathrm{x})$. Let $b_{l,m} = 0$ for $l = L+1, \ldots, m = -l, \ldots, l$, then $(b_{0,0}, \ldots, b_{L,L}, 0, 0, 0, \ldots)^T \in \ell^p_\beta$. From the definition of $\mathscr{A}$, we have

$$\begin{aligned} \|\mathscr{A}(T^\circ_L(\mathrm{x})) - T^\circ(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} &= \int_{\mathbb{S}^2\setminus\Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) + \int_\Gamma |T^\circ(\mathrm{x}) - T^\circ_L(\mathrm{x})|^2 d\sigma(\mathrm{x}) \\ &< \int_{\mathbb{S}^2\setminus\Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) + \varepsilon = \rho. \end{aligned}$$

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

Our main contributions are summarized as follows.

- Based on the K-L expansion, problem (1.6) can be written in a discrete form with variables $\alpha \in \ell^p_\beta$. Based on the discrete form, we propose a sparse optimization problem and derive a lower bound for the ℓ_2 norm of nonzero groups of scaled KKT points of the sparse optimization problem. Furthermore, we prove that the infinite-dimensional sparse optimization problem 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 ε-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 $L_2(\Omega \times \mathbb{S}^2)$.

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 $L_2(\Omega \times \mathbb{S}^2)$. In Section 6, we conduct numerical experiments on band-limited random fields and images from CMB 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 (2.10).

Based on the definition of $\mathscr{A}$ and the spherical harmonic expansion of T, we obtain that

$$
\begin{aligned}
&\|\mathscr{A}(T(\mathrm{x})) - T^\circ(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} \\
&= \int_{\mathbb{S}^2} |\mathscr{A}(T(\mathrm{x})) - T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) = \int_{\Gamma} |T(\mathrm{x}) - T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) + \int_{\mathbb{S}^2\setminus\Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) \\
&= \int_{\Gamma} \left| \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\mathrm{x}) \right|^2 d\sigma(\mathrm{x}) - 2\mathrm{Re}\left(\int_{\Gamma} T^\circ(\mathrm{x}) \left(\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\alpha}_{l,m} \overline{Y_{l,m}(\mathrm{x})} \right) d\sigma(\mathrm{x}) \right) \\
&\quad + \int_{\mathbb{S}^2} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) \\
&= \alpha^T Y \bar{\alpha} - 2\mathrm{Re}\left(\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\alpha}_{l,m} \int_{\Gamma} T^\circ(\mathrm{x}) \overline{Y_{l,m}(\mathrm{x})} d\sigma(\mathrm{x}) \right) + \int_{\mathbb{S}^2} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x}) \\
&= \alpha^H Y \alpha - 2\mathrm{Re}(\alpha^H \alpha^\circ) + c,
\end{aligned}
$$

where Y is an infinite-dimensional matrix with $(Y)_{l^2+l+m+1,l'^2+l'+m'+1} = \int_\Gamma Y_{l,m}(\mathrm{x})\overline{Y_{l',m'}(\mathrm{x})}d\sigma(\mathrm{x}) \in \mathbb{C}$, $\alpha^\circ = ((\alpha^\circ_{0\cdot})^T,\ldots,(\alpha^\circ_{l\cdot})^T,\ldots)^T$ with $\alpha^\circ_{l\cdot} \in \mathbb{C}^{2l+1}$, $\alpha^\circ_{l,m} = \int_\Gamma T^\circ(\mathrm{x})\overline{Y_{l,m}(\mathrm{x})}d\sigma(\mathrm{x})$ and $c = \int_{\mathbb{S}^2} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x})$.

By Parseval's theorem, $\|T(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} = \sum_{l=0}^\infty \|\alpha_{l\cdot}\|^2$. Hence the minimization problem (1.6) can be written as the following optimization problem

$$\begin{aligned} \min_{\alpha\in\ell^p_\beta} \quad & \sum_{l=0}^{\infty} \|\alpha_{l\cdot}\|^2 \\ \text{s.t.} \quad & G(\alpha) := \alpha^H Y\alpha - 2\mathrm{Re}(\alpha^H\alpha^\circ) + c - \rho \leqslant 0. \end{aligned} \tag{2.1}$$

Based on this discrete problem, we replace $\|\alpha_{l\cdot}\|^2$ by $\beta_l\|\alpha_{l\cdot}\|^p$ and propose a sparse optimization problem

$$\begin{aligned} \min_{\alpha\in\ell^p_\beta} \quad & \|\alpha\|^p_{2,p} \\ \text{s.t.} \quad & G(\alpha) \leqslant 0. \end{aligned} \tag{2.2}$$

From the setting of ρ in problem (1.6) and Theorem A.1, the feasible set of (2.2) has an interior point and bounded. The following assumption follows from our assumption on problem (1.6).

Assumption 2.1 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 Li & Chen (2022); Sorber *et al.* (2012); Sun *et al.* (2018), 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 extend the definition of scaled KKT points for finite-dimensional optimization problem with real variables in Chen *et al.* (2016, 2010); Rockafellar & Wets (2009).

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

$$p\beta_l\|\alpha^*_{l\cdot}\|^p\alpha^*_{l\cdot} + 2\nu\|\alpha^*_{l\cdot}\|^2(Y_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot}) = 0, \qquad \forall\, l \in \mathbb{N}_0, \tag{2.3}$$

$$\nu G(\alpha^*) = 0, \qquad G(\alpha^*) \leqslant 0. \tag{2.4}$$

Now we show that problem (2.2) has an optimal solution α^* that is a scaled KKT point of (2.2). We introduce the following auxiliary smoothing problem of (2.2),

$$\begin{aligned} \min_{\alpha\in\ell^p_\beta} \quad & \sum_{l=0}^{\infty} \beta_l(\|\alpha_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}} \\ \text{s.t.} \quad & G(\alpha) \leqslant 0, \end{aligned} \tag{2.5}$$

where $0 < \zeta_0 < 1$ and $0 < \zeta_l \leqslant (l^{\frac{1}{1-p}}\beta_l)^{-\frac{2}{p}}$, $l \in \mathbb{N}_0$ are smoothing parameters. By the subadditivity $(t+s)^p \leqslant t^p + s^p$, for $p \in (0,1)$, $t \geqslant 0, s > 0$, we have

$$\sum_{l=0}^{\infty} \beta_l(\|\alpha_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}} \leqslant \sum_{l=0}^{\infty} \beta_l(\|\alpha_{l\cdot}\|^p + \zeta_\ell^{\frac{p}{2}}) < \infty.$$

LEMMA 2.1 Assume Assumption 2.1 holds. Let α^* and α^*_ζ be the optimal solutions of problems (2.2) and (2.5), respectively, then we have $G(\alpha^*) = 0$ and $G(\alpha^*_\zeta) = 0$.

Proof. By Assumption 2.1, $\alpha^* \neq 0$ and $\alpha^*_\zeta \neq 0$. If $G(\alpha^*) < 0$, then there is an $\varepsilon \in (0,1)$ such that $G(\varepsilon\alpha^*) < 0$ and $\|\varepsilon\alpha^*\|^p_{2,p} = \varepsilon^p\|\alpha^*\|^p_{2,p} < \|\alpha\|^p_{2,p}$, which leads to a contradiction. Similarly, if $G(\alpha^*_\zeta) < 0$, then there is an $\varepsilon \in (0,1)$ such that $G(\varepsilon\alpha^*_\zeta) < 0$ and $\sum_{l=0}^\infty \beta_l(\|\varepsilon\alpha^*_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}} = \sum_{l=0}^\infty \beta_l(\varepsilon^2\|\alpha^*_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}} < \sum_{l=0}^\infty \beta_l(\|\alpha^*_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}}$, which also leads to a contradiction. Hence, $G(\alpha^*) = 0$ and $G(\alpha^*_\zeta) = 0$. □

THEOREM 2.3 Assume Assumption 2.1 holds. Then problem (2.2) has an optimal solution that satisfies the scaled KKT conditions (2.3)-(2.4).

Proof. Let $b_{l,m} = \int_{\mathbb{S}^2} T^\circ(\mathrm{x})\overline{Y_{l,m}(\mathrm{x})}d\sigma(\mathrm{x})$, $l \leqslant L$ and $b_{l,m} = 0$, $l > L$ for some $L \in \mathbb{N}_0$ such that $\|T^\circ(\mathrm{x}) - T^\circ_L(\mathrm{x})\|^2_{L_2(\mathbb{S}^2)} < \rho - \int_{\mathbb{S}^2\setminus\Gamma} |T^\circ(\mathrm{x})|^2 d\sigma(\mathrm{x})$, where $T^\circ_L(\mathrm{x}) = \sum_{l=0}^L \sum_{m=-l}^l b_{l,m}Y_{l,m}(\mathrm{x})$. From the discussion on (1.6) in section 1, we have $G(b) < 0$.

Let α^*_ζ be an optimal solution of (2.5). From Lemma 2.1, $G(\alpha^*_\zeta) = 0$ holds. Since $G(b) < G(\alpha^*_\zeta)$, α^*_ζ is not a minimizer of G. Hence, by Wirtinger gradient, $\partial_{\bar{\alpha}_\zeta} G(\alpha^*_\zeta) = Y\alpha^*_\zeta - \alpha^\circ \neq 0$. Choose an $\hat{l} \in \mathbb{N}_0$ such that $Y_{\hat{l}\cdot}\alpha^*_\zeta - \alpha^\circ_{\hat{l}\cdot} \neq 0$. Let $d_{k\cdot} = -Y_{\hat{l}\cdot}\alpha^*_\zeta + \alpha^\circ_{\hat{l}\cdot}$ for $k = \hat{l}$ and $d_{k\cdot} = 0$ for $k \neq \hat{l}$. Then we have that $d \in \ell^p_\beta$ and $(Y\alpha^*_\zeta - \alpha^\circ)^H d < 0$. Let $\mathbb{R}_- = \{x \in \mathbb{R} : x \leqslant 0\}$. Then we have $G(\alpha^*_\zeta) \in \mathbb{R}_-$ and $0 \in \mathrm{int}\{G(\alpha^*_\zeta) + (Y\alpha^*_\zeta - \alpha^\circ)^H d - \mathbb{R}_-\}$. By Maurer & Zowe (1979, (2.3)), α^*_ζ is regular. Moreover, by Maurer & Zowe (1979, Theorem 3.2), there is $\nu_\zeta \geqslant 0$ such that

$$p\beta_l(\|(\alpha^*_\zeta)_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}-1}(\alpha^*_\zeta)_{l\cdot} + 2\nu_\zeta(Y_{l\cdot}\alpha^*_\zeta - \alpha^\circ_{l\cdot}) = 0, \qquad \forall l \in \mathbb{N}_0, \tag{2.6}$$

$$\nu_\zeta G(\alpha^*_\zeta) = 0, \qquad G(\alpha^*_\zeta) \leqslant 0. \tag{2.7}$$

Multiplying both sides of equality (2.6) by $\|(\alpha^*_\zeta)_{l\cdot}\|^2$, we obtain

$$p\beta_l\|(\alpha^*_\zeta)_{l\cdot}\|^2(\|(\alpha^*_\zeta)_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}-1}(\alpha^*_\zeta)_{l\cdot} + 2\nu_\zeta\|(\alpha^*_\zeta)_{l\cdot}\|^2(Y_{l\cdot}\alpha^*_\zeta - \alpha^\circ_{l\cdot}) = 0, \quad \forall l \in \mathbb{N}_0. \tag{2.8}$$

Since b is in the feasible set of (2.5), we have $\sum_{l=0}^\infty \beta_l(\|(\alpha^*_\zeta)_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}} \leqslant \sum_{l=0}^\infty \beta_l(\|b_{l\cdot}\|^2 + \zeta_l)^{\frac{p}{2}}$, which implies that the optimal value of problem (2.5) is uniformly bounded.

From $\alpha^*_\zeta \in \ell^p_\beta$, by Lemma A.1, we have $\sum_{l=0}^\infty \eta^l l\|(\alpha^*_\zeta)_{l\cdot}\| < \infty$, for some $\eta > 1$, which implies that $\{\alpha^*_\zeta\}$ is bounded in ℓ^2, and thus there is a subsequence $\{\alpha^*_{\zeta^k}\}$ of $\{\alpha^*_\zeta\}$, which weakly converges to $\bar{\alpha}$ as $\|\zeta\| \to 0$ in ℓ^2 (see, for example, Kreyszig, 1991). Moreover from $Y_{\hat{l}\cdot}\alpha^*_\zeta - \alpha^\circ_{\hat{l}\cdot} \neq 0$ and (2.8), $\{\nu_\zeta\} \subset \mathbb{R}$ is bounded. Hence there is a subsequence$\{\nu_{\zeta^{k_i}}\}$ of $\{\nu_{\zeta^k}\}$, which converges to $\bar{\nu}$ as $\|\zeta^k\| \to 0$.

Let α^* be an optimal solution of (2.2). From weak convergence of $\{\alpha^*_{\zeta^{k_i}}\}$ to $\bar{\alpha}$, $\|\zeta^{k_i}\| \to 0$, as $k_i \to \infty$, and that $\alpha^*_{\zeta^{k_i}}$ is an optimal solution of (2.5) and $G(\alpha^*_{\zeta^{k_i}}) = 0$ for all $\{\zeta^{k_i}\} \subset \{\zeta^k\}$, we have

$$\|\bar{\alpha}\|^p_{2,p} \leqslant \liminf_{k_i\to\infty} \sum_{l=0}^\infty \beta_l(\|(\alpha^*_{\zeta^{k_i}})_{l\cdot}\|^2 + \zeta^{k_i}_l)^{\frac{p}{2}} \leqslant \lim_{k_i\to\infty} \sum_{l=0}^\infty \beta_l(\|\alpha^*_{l\cdot}\|^2 + \zeta^{k_i}_l)^{\frac{p}{2}} \leqslant \|\alpha^*\|^p_{2,p}$$

and

$$G(\bar{\alpha}) \leqslant \lim_{k_i\to\infty} G(\alpha^*_{\zeta^{k_i}}) = 0.$$

Hence we obtain that $\bar{\alpha}$ is an optimal solution of (2.2). From Lemma 2.1, we have $\bar{\nu}G(\bar{\alpha}) = 0$ and condition (2.4) holds. Since $\alpha^*_{\zeta^{k_i}}$ is weakly convergent to $\bar{\alpha}$ in ℓ^2, as $\|\zeta^{k_i}\| \to 0$, $(\alpha^*_{\zeta^{k_i}})_{l\cdot}$ is convergent to $\bar{\alpha}_{l\cdot}$, for any $l \in \mathbb{N}_0$. Moreover, the product $Y_{l\cdot}\alpha^*_{\zeta^{k_i}}$ is also convergent to $Y_{l\cdot}\bar{\alpha}$ for any $l \in \mathbb{N}_0$ under the weak convergence of $\alpha^*_{\zeta^{k_i}}$ (see, for example, Kreyszig, 1991). Hence $\bar{\alpha}$ satisfies the scaled KKT conditions (2.3). We complete the proof. □

For any nonzero vector $\alpha^*_{l\cdot} \in \mathbb{C}^{2l+1}$ that satisfies (2.3), we have

$$p\beta_l \|\alpha^*_{l\cdot}\|^{p-1} = 2\nu \|Y_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot}\| \leqslant 2\nu \|Y\alpha^* - \alpha^\circ\|.$$

By the definition of Y and feasibility of α^*, there exists $\tilde{c} > 0$ such that $\|Y\alpha^* - \alpha^\circ\| \leqslant \tilde{c}$. Hence for any nonzero vector $\alpha^*_{l\cdot} \in \mathbb{C}^{2l+1}$, we have

$$\|\alpha^*_{l\cdot}\| \geqslant \left(\frac{p\beta_l}{2\nu\tilde{c}}\right)^{\frac{1}{1-p}}. \tag{2.9}$$

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

$$\infty > \|\alpha^*\|^p_{2,p} = \sum_{l=0}^{\infty} \beta_l \|\alpha^*_{l\cdot}\|^p = \sum_{\{l \in \mathbb{N}_0 : \|\alpha^*_{l\cdot}\| \neq 0\}} \beta_l \|\alpha^*_{l\cdot}\|^p \geqslant \left(\frac{p}{2\nu\tilde{c}}\right)^{\frac{p}{1-p}} \sum_{\{l \in \mathbb{N}_0 : \|\alpha^*_{l\cdot}\| \neq 0\}} (\eta^l l^p)^{\frac{1}{1-p}},$$

which implies that $\{l \in \mathbb{N}_0 : \|\alpha^*_{l\cdot}\| \neq 0\}$ is a finite set. Thus, the number of nonzero vectors $\alpha^*_{l\cdot} \in \mathbb{C}^{2l+1}$ of a scaled KKT point of problem (2.2) is finite and there exists an $L \in \mathbb{N}_0$ such that $L = \max\{l \in \mathbb{N}_0 : \|\alpha^*_{l\cdot}\| \neq 0\}$.

Therefore, we consider a truncated problem of (2.2) in a finite-dimensional space.

For notational simplicity, we truncate $\alpha \in \ell^p_\beta$ to $\alpha = (\alpha_{0,0}, \alpha^T_{1\cdot}, \ldots, \alpha^T_{L\cdot})^T \in \mathbb{C}^d$, where $d := (L+1)^2$. We use $\hat{\alpha}^\circ \in \mathbb{C}^d$ to denote the truncated vector whose elements are the first d elements of α° and $\hat{Y} \in \mathbb{C}^{d\times d}$ to denote the leading principal submatrix of order d of Y. Since Γ 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

$$\begin{array}{ll} \min\limits_{\alpha \in \mathbb{C}^d} & \Phi(\alpha) := \sum_{l=0}^{L} \beta_l \|\alpha_{l\cdot}\|^p \\ \text{s.t.} & \alpha^H \hat{Y} \alpha - 2\mathrm{Re}(\alpha^H \hat{\alpha}^\circ) + c \leqslant \rho. \end{array} \tag{2.10}$$

By the definition of ρ, the feasible set of (2.10) is nonempty and has an interior point. The objective function $\Phi(\alpha)$ is continuous and level-bounded, nonnegative with $\Phi(0) = 0$, and differentiable except at points containing zero groups. Hence an optimal solution of problem (2.10) exists. The penalty formulation for (2.10) is

$$\min_{\alpha \in \mathbb{C}^d} \quad F_\lambda(\alpha) := \Phi(\alpha) + \lambda(\alpha^H \hat{Y} \alpha - 2\mathrm{Re}(\alpha^H \hat{\alpha}^\circ) + c - \rho)_+, \tag{2.11}$$

for some $\lambda > 0$, where $(\cdot)_+ := \max\{\cdot, 0\}$. For any nontrivial solution α^* of (2.11), we have

$$0 < F_\lambda(\alpha^*) = \min_{\alpha \in \mathbb{C}^d} F_\lambda(\alpha) < F_\lambda(0) = \lambda(c-\rho)_+ = \lambda(c-\rho) < \lambda c. \tag{2.12}$$

In Sections 3-4, we focus on problems (2.10) and (2.11).

3. Exact penalization

In this section, we consider the relationship between problems (2.10) and (2.11). We first give some notations. For a closed set $S \subset \mathbb{C}^n$, $\mathrm{dist}(z,S) = \inf_{z' \in S} \|z - z'\|$ denotes the distance from a point $z \in \mathbb{C}^n$ to S and $\mathbf{B}(b;r) = \{z \in \mathbb{C}^n : \|z-b\| \leqslant r\}$ 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

$$g(\alpha) := \alpha^H \hat{Y} \alpha - 2\mathrm{Re}(\alpha^H \hat{\alpha}^\circ) + c - \rho,$$

$S(\alpha) := \{\delta \in \mathbb{R} : g(\alpha) \leqslant \delta\}$ for $\alpha \in \mathbb{C}^d$, and $S^{-1}(\delta) := \{\alpha \in \mathbb{C}^d : g(\alpha) \leqslant \delta\}$. We denote the feasible set of problem (2.10) by $\mathscr{F}_e := \{\alpha \in \mathbb{C}^d : g(\alpha) \leqslant 0\}$. Let $\mathbb{L} := \{0, 1, \ldots, L\}$.

Since $\hat{Y}$ is positive definite, g is strongly convex and has a unique global minimizer $\hat{Y}^{-1}\hat{\alpha}^\circ \neq \hat{\alpha}^\circ$ which implies that $g(\hat{\alpha}^\circ) \in (\inf_{\alpha \in \mathbb{C}^d} g(\alpha), \infty)$. Since g is strongly convex and quadratic, there is a such that $\|\alpha - \bar{\alpha}\| \leqslant a|g(\alpha) - g(\bar{\alpha})|$ for $\alpha, \bar{\alpha} \in \mathbb{C}^d$. Choosing $\bar{\alpha}$ such that $g(\bar{\alpha}) = 0$, from Rockafellar & Wets (2009, Theorem 9.48), we obtain the following lemma.

LEMMA 3.1 There exists a constant $C > 0$ such that for any $\alpha \in \mathbb{C}^d$,

$$\mathrm{dist}(\alpha, \mathscr{F}_e) = \mathrm{dist}(\alpha, S^{-1}(0)) \leqslant C\mathrm{dist}(0, S(\alpha)) = C(g(\alpha))_+.$$

THEOREM 3.1 There exists a $\lambda^* > 0$ such that a local minimizer $\alpha^* \in \mathbb{C}^d$ of problem (2.10) is a local minimizer of problem (2.11) whenever $\lambda \geqslant \lambda^*$.

Proof. Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of problem (2.10), that is there exists a neighborhood $\mathscr{N}$ of α^* such that $\Phi(\alpha^*) \leqslant \Phi(\alpha)$ for $\alpha \in \mathscr{N} \cap \mathscr{F}_e$. We denote the group support set of α^* by $\gamma := \{l \in \mathbb{L} : \|\alpha^*_{l\cdot}\| \neq 0\}$, and the complement set of γ in $\mathbb{L}$ by $\tau := \{l \in \mathbb{L} : \|\alpha^*_{l\cdot}\| = 0\}$. Let α_γ and α_τ denote the restrictions of α onto γ and τ, respectively. We obtain that α^*_γ is a local minimizer of the following problem

$$\begin{array}{ll} \min\limits_{\alpha_\gamma} & \sum_{l \in \gamma} \beta_l \|\alpha_{l\cdot}\|^p \\ \text{s.t.} & \alpha_\gamma^H \hat{Y}_\gamma \alpha_\gamma - 2\mathrm{Re}(\alpha_\gamma^H \hat{\alpha}_\gamma^\circ) + c \leqslant \rho. \end{array} \tag{3.1}$$

Let $\bar{\varepsilon} = \frac{1}{2}\min\{\|\alpha^*_{l\cdot}\| : l \in \gamma\} > 0$. Then, there exists a small δ^* such that α^*_γ is a local minimizer of (3.1) and $\min\{\|\alpha_{l\cdot}\| : l \in \gamma\} > \bar{\varepsilon}$ for all $\alpha_\gamma \in \mathbf{B}(\alpha^*_\gamma; \delta^*)$. Let

$$g_\gamma(\alpha_\gamma) = \alpha_\gamma^H \hat{Y}_\gamma \alpha_\gamma - 2\mathrm{Re}(\alpha_\gamma^H \alpha_\gamma^\circ) + c - \rho \quad \text{and} \quad \Omega_1 := \{\alpha_\gamma : g_\gamma(\alpha_\gamma) \leqslant 0\}.$$

Let $[\alpha_\gamma; 0_\tau]$ be the vector with $([\alpha_\gamma; 0_\tau])_{l\cdot} = \alpha_{l\cdot}, l \in \gamma$ and $([\alpha_\gamma; 0_\tau])_{l\cdot} = 0, l \in \tau$. It is easy to see that $g([\alpha_\gamma; 0_\tau]) = g_\gamma(\alpha_\gamma)$ and $\mathrm{dist}(\alpha_\gamma, \Omega_1) = \mathrm{dist}([\alpha_\gamma; 0_\tau], \mathscr{F}_e)$. By Lemma 3.1, $\mathrm{dist}(\alpha_\gamma, \Omega_1) \leqslant C(g_\gamma(\alpha_\gamma))_+$ for all $\alpha_\gamma \in \mathbf{B}(\alpha^*_\gamma; \delta^*)$.

The objective function of (3.1) is Lipschitz continuous on $\mathbf{B}(\alpha^*_\gamma; \delta^*)$. Then by Chen *et al.* (2016, Lemma 3.1), there exists a $\lambda^* > 0$ such that, for any $\lambda \geqslant \lambda^*$, α^*_γ is a local minimizer of the following problem

$$\min_{\alpha_\gamma} \quad F^\gamma_\lambda(\alpha_\gamma) := \sum_{l \in \gamma} \beta_l \|\alpha_{l\cdot}\|^p + \lambda(g_\gamma(\alpha_\gamma))_+,$$

that is, there exists a neighborhood U_γ of 0 with $U_\gamma \subseteq \mathbf{B}(0; \delta^*)$ such that

$$F^\gamma_\lambda(\alpha_\gamma) \geqslant F^\gamma_\lambda(\alpha^*_\gamma), \quad \forall \alpha_\gamma \in \alpha^*_\gamma + U_\gamma.$$

We now show that α^* is a local minimizer of (2.11) with $\lambda \geqslant \lambda^*$.

For fixed $\varepsilon \in (0,\bar{\varepsilon})$ and $\lambda \geqslant \lambda^*$, we consider problem (2.11) in the neighborhood $U := U_\gamma \times (-\varepsilon,\varepsilon)^{d_\tau}$, where $d_\tau = \sum_{l\in\tau}(2l+1)$. Let $\tilde{\kappa}$ be a Lipschitz constant of the function $\lambda(g(\alpha))_+$ over α^*+U. For this constant $\tilde{\kappa}$, there exists an $\varepsilon_0 \in (0,\varepsilon)$ such that whenever $\|\alpha_{l\cdot}\| < \varepsilon_0$, $l \in \tau$,

$$\beta_l\|\alpha_{l\cdot}\|^p \geqslant \tilde{\kappa}\|\alpha_{l\cdot}\|, \quad l \in \tau. \tag{3.2}$$

Hence for any $y \in U_\gamma \times (-\varepsilon_0,\varepsilon_0)^{d_\tau}$, we have

$$\begin{aligned} F_\lambda(\alpha^*+y) &= \lambda(g(\alpha^*+y))_+ + \sum_{l\in\gamma}\beta_l\|\alpha^*_{l\cdot}+y_{l\cdot}\|^p + \sum_{l\in\tau}\beta_l\|y_{l\cdot}\|^p \\ &\geqslant \lambda(g([\alpha^*_\gamma+y_\gamma;0_\tau]))_+ - \tilde{\kappa}\|y_\tau\| + \sum_{l\in\gamma}\beta_l\|\alpha^*_{l\cdot}+y_{l\cdot}\|^p + \tilde{\kappa}\|y_\tau\| \\ &\geqslant F^\gamma_\lambda(\alpha^*_\gamma) = F_\lambda(\alpha^*), \end{aligned}$$

where the first inequality follows from the Lipschitz continuity of $\lambda(g(\alpha))_+$ with Lipschitz constant $\tilde{\kappa}$ and (3.2), and the last inequality follows from the local optimality of α^*_γ. Thus, α^* is a local minimizer of problem (2.11) with $\lambda \geqslant \lambda^*$. This completes the proof. □

THEOREM 3.2 Let $\tilde{\alpha} = \hat{Y}^{-1}\hat{\alpha}^\circ$, $\varepsilon > 0$ and $\lambda \geqslant C\bar{\beta}^{\frac{1}{p}}(\varepsilon(L+1)^{\frac{p}{2}-1})^{-\frac{1}{p}}\Phi(\tilde{\alpha})$, where C is defined in Lemma 3.1 and $\bar{\beta} = \eta^L L^p$. Then for any global minimizer α^* of problem (2.11), the projection $\alpha_\varepsilon := \mathscr{P}_{\mathscr{F}_e}(\alpha^*)$ is an ε-minimizer of (2.10), that is, $\Phi(\alpha_\varepsilon) \leqslant \min\{\Phi(\alpha) : \alpha \in \mathscr{F}_e\} + \varepsilon$.

Proof. Note that g is a strongly convex function and $\tilde{\alpha}$ is a minimizer of g. Since the feasible set $\mathscr{F}_e$ of (2.10) has an interior point, we have $\tilde{\alpha} \in \mathrm{int}\mathscr{F}_e$.

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

$$((\alpha^*)^H\hat{Y}\alpha^* - 2\mathrm{Re}((\alpha^*)^H\hat{\alpha}^\circ) + c - \rho)_+ \leqslant \tfrac{1}{\lambda}F_\lambda(\alpha^*) \leqslant \tfrac{1}{\lambda}F_\lambda(\tilde{\alpha}) = \tfrac{1}{\lambda}\Phi(\tilde{\alpha}). \tag{3.3}$$

Hence for any $\alpha \in \mathscr{F}_e$, we obtain

$$\begin{aligned} \Phi(\alpha_\varepsilon) - \Phi(\alpha) &\leqslant \sum_{l=0}^{L}\beta_l(\|(\alpha_\varepsilon)_{l\cdot}\|^p - \|\alpha^*_{l\cdot}\|^p) \leqslant \sum_{l=0}^{L}\beta_l\|(\alpha_\varepsilon)_{l\cdot} - \alpha^*_{l\cdot}\|^p \leqslant \bar{\beta}\sum_{l=0}^{L}\left(\|(\alpha_\varepsilon)_{l\cdot} - \alpha^*_{l\cdot}\|^2\right)^{\frac{p}{2}} \\ &\leqslant \bar{\beta}(L+1)\left(\tfrac{1}{L+1}\sum_{l=0}^{L}\|(\alpha_\varepsilon)_{l\cdot} - \alpha^*_{l\cdot}\|^2\right)^{\frac{p}{2}} \leqslant \bar{\beta}(L+1)^{1-\frac{p}{2}}(\mathrm{dist}(\alpha^*,\mathscr{F}_e))^p \\ &\leqslant \bar{\beta}(L+1)^{1-\frac{p}{2}}(C((\alpha^*)^H\hat{Y}\alpha^* - 2\mathrm{Re}((\alpha^*)^H\hat{\alpha}^\circ) + c - \rho)_+)^p \\ &\leqslant \bar{\beta}(L+1)^{1-\frac{p}{2}}\left(\tfrac{C}{\lambda}\Phi(\tilde{\alpha})\right)^p \leqslant \varepsilon, \end{aligned}$$

where the first inequality is from the global optimality of α^*, the second inequality is from Lemma 2.4 in Chen *et al.* (2016), the fifth inequality is from the concavity of function $t \to t^{\frac{p}{2}}$ for $t \geqslant 0$, the sixth inequality is from Lemma 3.1, the seventh inequality is from (3.3) and the last inequality follows from the choice of λ. Thus, the projection $\mathscr{P}_{\mathscr{F}_e}(\alpha^*)$ is an ε-minimizer of (2.10). This completes the proof. □

4. Optimality conditions and a smoothing penalty algorithm

In this section, we first define first-order optimality conditions of (2.10) and (2.11), which are necessary conditions for local optimality. We also derive lower bounds for the ℓ_2 norm of nonzero groups of first-order stationary points of (2.10) and (2.11). Next we propose a smoothing penalty algorithm for solving (2.10) and prove its convergence to a first-order stationary point of (2.10).

4.1 *First-order optimality conditions*

We first present a first-order optimality condition of problem (2.11).

DEFINITION 4.1 We call $\alpha^* \in \mathbb{C}^d$ a scaled first-order stationary point of problem (2.11) if

$$p\beta_l \|\alpha^*_{l\cdot}\|^p \alpha^*_{l\cdot} + 2\lambda\xi\|\alpha^*_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot}) = 0, \quad \forall\, l \in \mathbb{L}, \tag{4.1}$$

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

THEOREM 4.2 Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of (2.11). Then α^* is a scaled first-order stationary point of (2.11).

Proof. Let $\alpha^* \in \mathbb{C}^d$ be a local minimizer of (2.11). We obtain that α^*_γ is a local minimizer of

$$\min_{\alpha_\gamma} \quad \sum_{l\in\gamma} \beta_l \|\alpha_{l\cdot}\|^p + \lambda(\alpha_\gamma^H \hat{Y}_\gamma \alpha_\gamma - 2\mathrm{Re}(\alpha_\gamma^H \hat{\alpha}^\circ_\gamma) + c - \rho)_+. \tag{4.2}$$

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

$$p\beta_l \|\alpha^*_{l\cdot}\|^{p-2} \alpha^*_{l\cdot} + 2\lambda\xi((\hat{Y}_\gamma\alpha^*_\gamma)_{l\cdot} - \alpha^\circ_{l\cdot}) = 0, \quad \forall\, l \in \gamma, \tag{4.3}$$

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

Multiplying $\|\alpha^*_{l\cdot}\|^2$ on both sides of (4.3), we obtain

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

Since $\|\alpha^*_{l\cdot}\| = 0$ for $l \in \tau$, we have

$$p\beta_l \|\alpha^*_{l\cdot}\|^p \alpha^*_{l\cdot} + 2\lambda\xi\|\alpha^*_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot}) = 0, \quad \forall\, l \in \mathbb{L}.$$

Hence (4.1) holds at α^*. □

The next theorem gives a lower bound for the l_2 norm of nonzero groups of stationary points of (2.11).

THEOREM 4.3 Let $\alpha^* \in \mathbb{C}^d$ be a scaled first-order stationary point of (2.11) and $\|\hat{Y}\alpha^* - \hat{\alpha}^\circ\| \leqslant \tilde{c}$ for some $\tilde{c} > 0$. Then,

$$\|\alpha^*_{l\cdot}\| \geqslant \left(\frac{p\beta_l}{2\lambda\tilde{c}}\right)^{\frac{1}{1-p}}, \forall\, l \in \gamma.$$

The proof is similar with that of (2.9), and thus we omit it here. According to Theorems 3.1 and 4.2, Theorem 4.3 gives a lower bound for the ℓ_2 norm of nonzero groups of local minimizers of problem (2.10).

Next, we consider the first-order optimality condition of problem (2.10). Similar to Definition 2.2, we call α^* is a scaled first-order stationary point or a scaled KKT point of (2.10) if the following conditions hold,

$$\begin{aligned} p\beta_l\|\alpha^*_{l\cdot}\|^p\alpha^*_{l\cdot}+2\nu\|\alpha^*_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^*-\alpha^\circ_{l\cdot})=0,\ \forall\, l\in\mathbb{L},\\ \nu g(\alpha^*)=0,\quad g(\alpha^*)\leqslant 0,\quad \nu\geqslant 0. \end{aligned} \tag{4.4}$$

Since $\alpha=0$ is not a feasible point, thus $\nu\neq 0$. Hence, replacing λ by ν in Theorem 4.3, we can obtain a lower bound for the ℓ_2 norm of nonzero groups of the scaled KKT point of problem (2.10).

THEOREM 4.4 If α^* is a local minimizer of problem (2.10), then α^* is a scaled KKT point of (2.10) with $g(\alpha^*)=0$.

Proof. Let α^* be a local minimizer of problem (2.10). Since $g(\alpha^*)=g_\gamma(\alpha^*_\gamma)$ and $\sum_{l=0}^L\beta_l\|\alpha^*_{l\cdot}\|^p=\sum_{l\in\gamma}\beta_l\|\alpha^*_{l\cdot}\|^p$, following the proof of Theorem 3.1, α^*_γ is a local minimizer of (3.1). Using Wirtinger gradient, we know that $\partial_{\bar{\alpha}_\gamma}g_\gamma(\alpha^*_\gamma)=\hat{Y}_\gamma\alpha^*_\gamma-\hat{\alpha}^\circ_\gamma$.

From Theorem 3.1 and Theorem 4.2, there is $\lambda>0$ such that (4.1) holds, which implies that if $\hat{Y}_{l\cdot}\alpha^*-\hat{\alpha}^\circ_{l\cdot}=0$, then $\alpha^*_{l\cdot}=0$. Hence $\hat{Y}_\gamma\alpha^*_\gamma-\hat{\alpha}^\circ_\gamma\neq 0$, and the LICQ (linear independence constraint qualification) holds at α^*_γ for problem (3.1).

Note that, the objective function of problem (3.1) and g_γ are continuously differentiable at α^*_γ and $\partial_{\bar{\alpha}_\gamma}g_\gamma(\alpha^*_\gamma)=\hat{Y}_\gamma\alpha^*_\gamma-\hat{\alpha}^\circ_\gamma$. Hence α^*_γ is a KKT point of (3.1), that is, there exists ν such that

$$\begin{aligned} p\beta_l\|\alpha^*_{l\cdot}\|^{p-2}\alpha^*_{l\cdot}+2\nu(\hat{Y}_\gamma\alpha^*_\gamma-\alpha^\circ_\gamma)_{l\cdot}=0,\ \forall\, l\in\gamma,\\ \nu g_\gamma(\alpha^*_\gamma)=0,\quad g_\gamma(\alpha^*_\gamma)\leqslant 0,\quad \nu\geqslant 0. \end{aligned} \tag{4.5}$$

Multiplying $\|\alpha^*_{l\cdot}\|^2$ on both sides of the first equality in (4.5), we obtain

$$p\beta_l\|\alpha^*_{l\cdot}\|^p\alpha^*_{l\cdot}+2\nu\|\alpha^*_{l\cdot}\|^2(\hat{Y}_\gamma\alpha^*_\gamma-\alpha^\circ_\gamma)_{l\cdot}=0,\quad \forall\, l\in\gamma.$$

Since $\alpha^*_{l\cdot}=0$ for $l\in\tau$, we obtain

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

Combining this with (4.5) and $g(\alpha^*)=g_\gamma(\alpha^*_\gamma)$, we find that α^* is a scaled KKT point of (2.10).

Now we show $g(\alpha^*)=0$. Assume on contradiction that $g(\alpha^*)<0$. Then for sufficiently small $\delta>0$, we have $g((1-\delta)\alpha^*)<0$ and $\sum_{l=0}^L\beta_l\|(1-\delta)\alpha^*_{l\cdot}\|^p=(1-\delta)^p\sum_{l=0}^L\beta_l\|\alpha^*_{l\cdot}\|^p<\sum_{l=0}^L\beta_l\|\alpha^*_{l\cdot}\|^p$, which implies that α^* cannot be a local minimizer of (2.10). Hence $g(\alpha^*)=0$. The proof is completed. □

From the definitions we can see that for some $\lambda>0$, any scaled KKT point of problem (2.10) is a scaled first order stationary point of (2.11). Moreover, any scaled first order stationary point of problem (2.11) which belongs to the feasible set $\mathscr{F}_e$ is a scaled KKT point of problem (2.10).

4.2 *A smoothing penalty algorithm for problem (2.10)*

We define a smoothing function of the nonsmooth function $\lambda(g(\alpha))_+$ as follows

$$f_{\lambda,\mu}(\alpha)=\psi_{\lambda,\mu}(g(\alpha))$$

with $\psi_{\lambda,\mu}(s) := \lambda \max_{0\leqslant t\leqslant 1}\{st - \frac{\mu}{2}t^2\}$ and $\mu > 0$ is a smoothing parameter. It is easy to verify that $\psi'_{\lambda,\mu}(s) = \lambda \min\{\max\{\frac{s}{\mu},0\},1\} \geqslant 0$ and $|\psi'_{\lambda,\mu}(s_1) - \psi'_{\lambda,\mu}(s_2)| \leqslant \frac{\lambda}{\mu}|s_1 - s_2|, \forall s_1, s_2 \in \mathbb{R}$. It is not hard to show that

$$f_{\lambda,\mu}(\alpha) = \begin{cases} 0 & \text{if } g(\alpha) \leqslant 0 \\ \frac{\lambda}{2\mu}(g(\alpha))^2 & \text{if } 0 \leqslant g(\alpha) \leqslant \mu \\ \lambda g(\alpha) - \frac{\lambda\mu}{2} & \text{if } g(\alpha) \geqslant \mu \end{cases}$$

and

$$0 \leqslant \lambda(g(\alpha))_+ - f_{\lambda,\mu}(\alpha) \leqslant \frac{\lambda\mu}{2}. \tag{4.6}$$

By Wirtinger calculus, we obtain $\nabla f_{\lambda,\mu}(\alpha) = \begin{bmatrix} \partial_\alpha f_{\lambda,\mu}(\alpha) \\ \partial_{\bar\alpha} f_{\lambda,\mu}(\alpha) \end{bmatrix}$, where $\partial_{\bar\alpha} f_{\lambda,\mu}(\alpha) = \overline{\partial_\alpha f_{\lambda,\mu}(\alpha)}$ and

$$\partial_{\bar\alpha} f_{\lambda,\mu}(\alpha) = \begin{cases} 0 & \text{if } g(\alpha) \leqslant 0 \\ \frac{\lambda}{\mu} g(\alpha)(\hat{Y}\alpha - \hat\alpha^\circ) & \text{if } 0 \leqslant g(\alpha) \leqslant \mu \\ \lambda(\hat{Y}\alpha - \hat\alpha^\circ) & \text{if } g(\alpha) \geqslant \mu. \end{cases}$$

More details about the smoothing function can be found in Chen (2012) and references therein.

We consider the following optimization problem

$$\min_{\alpha\in\mathbb{C}^d} \qquad F_{\lambda,\mu}(\alpha) := \Phi(\alpha) + f_{\lambda,\mu}(\alpha). \tag{4.7}$$

For fixed positive parameters λ and μ, $F_{\lambda,\mu}$ is continuous and level-bounded since Φ 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 (2.10).

Algorithm 4.5 A smoothing penalty algorithm for problem (2.10)

Choose $\lambda^0 > 0$, $\mu^0 > 0$, $\varepsilon^0 > 0$, $\varsigma_1 > 1$, and $0 < \varsigma_2 < 1$. Set $k = 0$ and $\alpha^0 = \tilde\alpha := \hat{Y}^{-1}\alpha^\circ$.
(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 α^k, $\lambda = \lambda^k$, $\mu = \mu^k$, and find an α^{k+1} satisfying

$$\|p\beta_l\|\alpha^{k+1}_{l\cdot}\|^p\alpha^{k+1}_{l\cdot} + 2\|\alpha^{k+1}_{l\cdot}\|^2(\partial_{\bar\alpha} f_{\lambda^k,\mu^k}(\alpha^{k+1}))_{l\cdot}\| \leqslant \varepsilon^k, \quad \forall l \in \mathbb{L}. \tag{4.8}$$

(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 4.5 in the following theorem.

THEOREM 4.6 Let $\{\alpha^k\}$ be generated by Algorithm 4.5. Then, the following statements hold.

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

(ii) Any accumulation point α^* of $\{\alpha^k\}$ is a scaled KKT point of problem (2.10).

Proof. (i) We can see that

$$\Phi(\alpha^{k+1}) \leqslant F_{\lambda^k,\mu^k}(\alpha^{k+1}) \leqslant F_{\lambda^k,\mu^k}(\tilde\alpha) = \Phi(\tilde\alpha),$$

where the first inequality follows from that $f_{\lambda^k,\mu^k}(\alpha^{k+1}) \geqslant 0$, the second inequality follows from step (1) of Algorithm 4.5, and the equality is from $\tilde{\alpha} = \hat{Y}^{-1}\alpha^\circ \in \mathscr{F}_e$ and $f_{\lambda^k,\mu^k}(\tilde{\alpha}) = 0$. Since Φ is level-bounded, $\{\alpha^k\}$ is bounded.

(ii) Let α^* be an accumulation point of $\{\alpha^k\}$ and $\{\alpha^k\}_{k\in\mathscr{K}}$ be a subsequence of $\{\alpha^k\}$ such that $\{\alpha^k\} \to \alpha^*$ as $k \to \infty$, $k \in \mathscr{K}$. Note that

$$\lambda^{k-1}(g(\alpha^k))_+ - \tfrac{\lambda^{k-1}\mu^{k-1}}{2} \leqslant f_{\lambda^{k-1},\mu^{k-1}}(\alpha^k) \leqslant F_{\lambda^{k-1},\mu^{k-1}}(\alpha^k) \leqslant F_{\lambda^{k-1},\mu^{k-1}}(\tilde{\alpha}) \leqslant \Phi(\tilde{\alpha}),$$

where the first inequality follows from (4.6). Then, we have

$$(g(\alpha^k))_+ \leqslant \frac{\Phi(\tilde{\alpha})}{\lambda^{k-1}} + \frac{\mu^{k-1}}{2}. \tag{4.9}$$

From step (3) in Algorithm 4.5, $\lambda^{k-1} \to \infty$ and $\mu^{k-1} \to 0$, as $k \to \infty$, $k \in \mathscr{K}$. Taking limits in (4.9) as $k \to \infty$, $k \in \mathscr{K}$, we obtain that $(g(\alpha^*))_+ \leqslant 0$. Hence, $\alpha^* \in \mathscr{F}_e$.

From (4.8), we have

$$\|p\beta_l\|\alpha^k_{l\cdot}\|^p\alpha^k_{l\cdot} + 2\|\alpha^k_{l\cdot}\|^2(\partial_{\tilde{\alpha}} f_{\lambda^k,\mu^k}(\alpha^k))_{l\cdot}\| \leqslant \varepsilon^{k-1}, \quad \forall l \in \mathbb{L}. \tag{4.10}$$

We first assume that $g(\alpha^*) < 0$. For all sufficiently large k, we obtain $g(\alpha^k) < 0$ and (4.10) becomes

$$p\beta_l\|\alpha^k_{l\cdot}\|^{p+1} \leqslant \varepsilon^{k-1}, \quad \forall l \in \mathbb{L}.$$

Taking limits on both sides of the above relation, we obtain $\alpha^* = 0$ which contradicts to Assumption 2.1. Thus, $g(\alpha^*) = 0$.

Let $t^k := \psi'_{\lambda^k,\varepsilon^k}(g(\alpha^k))$ for notational simplicity, we have $t^k \geqslant 0$. Then (4.10) reduces to

$$\|p\beta_l\|\alpha^k_{l\cdot}\|^p\alpha^k_{l\cdot} + 2t^k\|\alpha^k_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^k - \alpha^\circ_{l\cdot})\| \leqslant \varepsilon^{k-1}, \quad \forall l \in \mathbb{L}. \tag{4.11}$$

Now, we prove that $\{t^k\}_{\mathscr{K}}$ is bounded. On the contrary, we assume $\{t^k\}_{\mathscr{K}}$ is unbounded and $\{t^k\}_{\mathscr{K}} \to \infty$, then,

$$\left\|\tfrac{p\beta_l}{t^k}\|\alpha^k_{l\cdot}\|^p\alpha^k_{l\cdot} + 2\|\alpha^k_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^k - \alpha^\circ_{l\cdot})\right\| \leqslant \tfrac{\varepsilon^{k-1}}{t^k}, \quad \forall l \in \mathbb{L}.$$

Passing to the limit in the above relation gives

$$\|\alpha^*_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot}) = 0, \quad \forall l \in \mathbb{L}.$$

Since $g(\alpha^*) = 0$ implies α^* is in $\mathscr{F}_e$, but is not a minimizer of g, we have $\alpha^* \neq 0$ and $\hat{Y}\alpha^* - \hat{\alpha}^\circ \neq 0$. Moreover, $g(\alpha^*) = g_\gamma([\alpha^*_\gamma; 0_\tau]) = 0$ implies that $\alpha^*_{l\cdot} \neq 0$, $\forall l \in \gamma$ and $(\hat{Y}\alpha^* - \hat{\alpha}^\circ)_\gamma \neq 0$. Thus, $\{t^k\}_{\mathscr{K}}$ is bounded. Let $\{t^k\}_{\mathscr{K}} \to t^*$. Taking limits on both sides of (4.11) gives

$$\|p\beta_l\|\alpha^*_{l\cdot}\|^p\alpha^*_{l\cdot} + 2t^*\|\alpha^*_{l\cdot}\|^2(\hat{Y}_{l\cdot}\alpha^* - \alpha^\circ_{l\cdot})\| = 0, \quad \forall l \in \mathbb{L}.$$

Therefore, α^* is a scaled KKT point of (2.10). □

5. Approximation error

In Gia *et al.* (2019), the authors gave the approximation error for random field using regularized $\ell_{2,1}$ model based on the observed random field $T^{\circ}(\omega,\mathrm{x}) \in L_2(\Omega \times \mathbb{S}^2)$. In this section, we estimate the approximation error of the inpainted random field in the space $L_2(\Omega\times\mathbb{S}^2)$ based on the observed random field $T^{\circ}(\omega,\mathrm{x}) \in L_2(\Omega \times \mathbb{S}^2)$.

For any fixed $\omega \in \Omega$, let $\hat{\alpha}^*(\omega) := (\alpha^*_{0,0}(\omega),\ldots,\alpha^*_{L_\omega,L_\omega}(\omega))^T \in \mathbb{C}^{(L_\omega+1)^2}$ with the group support set γ_ω be a scaled KKT point of problem (2.10). By our results in the previous sections, L_ω is a finite number. Moreover, $\alpha^*(\omega) := ((\hat{\alpha}^*(\omega))^T,0,\ldots)^T$ is a scaled KKT point of problem (2.2) with $\rho(\omega)$ in the infinite-dimensional space ℓ^p_β.

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

$$T^*(\omega,\mathrm{x}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \alpha^*_{l,m}(\omega) Y_{l,m}(\mathrm{x}), \tag{5.1}$$

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

LEMMA 5.1 If the random variable $\omega \in \Omega$ has finite second order moment that is $\mathbb{E}[\|\omega\|^2] < \infty$ and there is κ such that $\|T^*(\omega_1,\mathrm{x}) - T^*(\omega_2,\mathrm{x})\|_{L_2(\mathbb{S}^2)} \leqslant \kappa\|\omega_1-\omega_2\|$, $\forall \omega_1,\omega_2 \in \Omega$, then $T^*(\omega,\mathrm{x}) \in L_2(\Omega\times\mathbb{S}^2)$.

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

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

we have

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

Hence, we obtain

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

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

THEOREM 5.1 Let $T^{\circ}(\omega,\mathrm{x}) \in L_2(\Omega\times\mathbb{S}^2)$ be the observed random field. Then for any $\varepsilon > 0$ there exists L such that

$$-\frac{\varepsilon}{2} \leqslant \|\mathscr{A}(T^*_L(\omega,\mathrm{x})) - T^{\circ}(\omega,\mathrm{x})\|^2_{L_2(\Omega\times\mathbb{S}^2)} - \rho < \varepsilon,$$

where $T^*_L(\omega,\mathrm{x}) = \sum_{l=0}^{L}\sum_{m=-l}^{l}\alpha^*_{l,m}(\omega)Y_{l,m}(\mathrm{x})$ and $\rho = \mathbb{E}[\rho(\omega)]$.

Proof. Since $T^{\circ}(\omega,\mathrm{x}) \in L_2(\Omega\times\mathbb{S}^2)$, by Fubini's theorem, for $\omega \in \Omega$, $T^{\circ}(\omega,\mathrm{x}) \in L_2(\mathbb{S}^2)$, P-a.s., in which case $T^{\circ}(\omega,\mathrm{x})$ admits an expansion in terms of spherical harmonics, P-a.s., that is, $T^{\circ}(\omega,\mathrm{x}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha^{\mathrm{obs}}_{l,m}(\omega)Y_{l,m}(\mathrm{x})$, P-a.s., where $\alpha^{\mathrm{obs}}(\omega) = (\alpha^{\mathrm{obs}}_{0,0}(\omega),\alpha^{\mathrm{obs}}_{1,-1}(\omega),\ldots)^T$ is the Fourier coefficient vector of $T^{\circ}(\omega,\mathrm{x})$.

By Definition 2.2 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 $\nu_\omega \geqslant \bar{\nu}$ for any $\omega \in \Omega$. On the contrary, if there exists a $\omega \in \Omega$ such that $\nu_\omega \to 0$, then from

$$\|\alpha^*(\omega)\|^p_{2,p} = \sum_{\{l\in\mathbb{N}_0:\,\|\alpha^*_{l\cdot}(\omega)\|\neq 0\}} \beta_l\|\alpha^*_{l\cdot}(\omega)\|^p \quad\geqslant\quad \left(\frac{p}{2\nu_\omega\tilde{c}}\right)^{\frac{p}{1-p}} \sum_{\{l\in\mathbb{N}_0:\,\|\alpha^*_{l\cdot}(\omega)\|\neq 0\}} (\eta^l l^p)^{\frac{1}{1-p}},$$

we have $\|\alpha^*(\omega)\|_{2,p}^p \to \infty$ which is a contradiction with $\alpha^*(\omega) \in \ell_\beta^p$. Thus, from $\alpha^*(\omega) \in \ell_\beta^p$ and (2.9), there exists a positive scalar $\bar{\nu}$ such that $\nu_\omega \geqslant \bar{\nu}$ for any $\omega \in \Omega$. Thus, for any $\varepsilon_1 > 0$ there exists L_1 such that $\frac{p}{\bar{\nu}} \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p < \frac{\varepsilon_1}{2}$ for any $\omega \in \Omega$, which implies that

$$\mathbb{E}\left[\frac{p}{\nu_\omega} \sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] < \frac{p}{\bar{\nu}} \mathbb{E}\left[\sum_{l=L_1+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] < \frac{\varepsilon_1}{2}.$$

By Lemma 5.1, for any $\varepsilon_2 > 0$ there exists L_2 such that $\sum_{l=L_2+1}^{\infty} \mathbb{E}[\|\alpha_{l\cdot}^*(\omega)\|^2] < \frac{\varepsilon_2}{2}$. Let $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$, $L = \max\{L_1, L_2\}$ and $d = (L+1)^2$.

For notational simplicity, let $\alpha^*(\omega) = ((\hat{\alpha}^*(\omega))^T, (\tilde{\alpha}^*(\omega))^T)^T \in \ell_\beta^p$, where $\hat{\alpha}^*(\omega) \in \mathbb{C}^d$, $\omega \in \Omega$ and $Y = \begin{bmatrix} \hat{Y} & X \\ X^H & \tilde{Y} \end{bmatrix}$, where $\hat{Y} \in \mathbb{C}^{d \times d}$. Let $\tilde{T}_L^*(\omega, \mathrm{x}) = \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m}^*(\omega) Y_{l,m}(\mathrm{x})$, we have

$$\begin{aligned}
\rho &= \|\mathscr{A}(T^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 = \|\mathscr{A}(T_L^*(\omega,\mathrm{x}) + \tilde{T}_L^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 \\
&= \|\mathscr{A}(T_L^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 \\
&\quad + \mathbb{E}\left[\int_{\mathbb{S}^2} |\mathscr{A}(\tilde{T}_L^*(\omega,\mathrm{x}))|^2 d\sigma(\mathrm{x}) + 2\int_{\mathbb{S}^2} (\mathscr{A}(T_L^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x}))\mathscr{A}(\tilde{T}_L^*(\omega,\mathrm{x}))d\sigma(\mathrm{x})\right] \\
&= \|\mathscr{A}(T_L^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 \\
&\quad + \mathbb{E}\left[\int_{\Gamma} |\tilde{T}_L^*(\omega,\mathrm{x})|^2 d\sigma(\mathrm{x}) + 2\int_{\Gamma} (T_L^*(\omega,\mathrm{x}) - T^\circ(\omega,\mathrm{x}))\tilde{T}_L^*(\omega,\mathrm{x})d\sigma(\mathrm{x})\right] \\
&= \|\mathscr{A}(T_L^*(\omega,\mathrm{x})) - T^\circ(\omega,\mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 + \mathbb{E}[(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega)] \\
&\quad + \mathbb{E}[2(\tilde{\alpha}^*(\omega))^H X^H \hat{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H [\begin{matrix} X^H & \tilde{Y} \end{matrix}] \alpha^{\mathrm{obs}}(\omega)],
\end{aligned} \tag{5.2}$$

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 2.2 for a scaled KKT point $\alpha^* \in \ell_\beta^p$ of problem (2.2), there is $\nu_\omega > 0$ such that

$$p\beta_l \|\alpha_{l\cdot}^*(\omega)\|^{p-2} \alpha_{l\cdot}^*(\omega) + 2\nu_\omega (Y_{l\cdot} \alpha^*(\omega) - \alpha_{l\cdot}^\circ(\omega)) = 0, \quad l \in \gamma_\omega.$$

Using $\alpha^\circ(\omega) = Y \alpha^{\mathrm{obs}}(\omega)$, for any fixed $\omega \in \Omega$, we obtain

$$p\beta_l \|\alpha_{l\cdot}^*(\omega)\|^p + 2\nu_\omega (\alpha_{l\cdot}^*(\omega))^H (Y_{l\cdot} \alpha^*(\omega) - Y_{l\cdot} \alpha^{\mathrm{obs}}(\omega)) = 0, \quad l \in \gamma_\omega, \tag{5.3}$$

and

$$\sum_{l=L+1}^{\infty} p\beta_l \|\alpha_{l\cdot}^*(\omega)\|^p + 2\nu_\omega (\tilde{\alpha}^*(\omega))^H [\begin{matrix} X^H & \tilde{Y} \end{matrix}] (\alpha^*(\omega) - \alpha^{\mathrm{obs}}(\omega)) = 0, \omega \in \Omega. \tag{5.4}$$

Thus,

$$\begin{aligned}
&\mathbb{E}[(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega)] + \mathbb{E}[2(\tilde{\alpha}^*(\omega))^H X^H \hat{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H [\ X^H \quad \tilde{Y}\ ]\alpha^{\text{obs}}(\omega)] \\
&= \mathbb{E}\left[(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega) + 2(\tilde{\alpha}^*(\omega))^H X^H \hat{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H [\ X^H \quad \tilde{Y}\ ]\alpha^*(\omega)\right] \\
&\quad - \mathbb{E}\left[\frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] \\
&= \mathbb{E}\left[-(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega) - \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] \\
&\geqslant \mathbb{E}\left[-\sum_{l=L+1}^{\infty} \|\alpha_{l\cdot}^*(\omega)\|^2 - \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] > -\varepsilon,
\end{aligned} \tag{5.5}$$

where the first equality follows from (5.4) and the first inequality follows from that $\|\tilde{Y}\| \leqslant 1$. And

$$\begin{aligned}
&\mathbb{E}[(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega)] + \mathbb{E}[2(\tilde{\alpha}^*(\omega))^H X^H \hat{\alpha}^*(\omega) - 2(\tilde{\alpha}^*(\omega))^H [\ X^H \quad \tilde{Y}\ ]\alpha^{\text{obs}}(\omega)] \\
&= \mathbb{E}\left[-(\tilde{\alpha}^*(\omega))^H \tilde{Y} \tilde{\alpha}^*(\omega) - \frac{p}{\nu_\omega} \sum_{l=L+1}^{\infty} \beta_l \|\alpha_{l\cdot}^*(\omega)\|^p\right] \\
&\leqslant \mathbb{E}\left[\sum_{l=L+1}^{\infty} \|\alpha_{l\cdot}^*(\omega)\|^2\right] \leqslant \frac{\varepsilon}{2}.
\end{aligned} \tag{5.6}$$

Combining (5.2), (5.5) and (5.6), we obtain

$$-\frac{\varepsilon}{2} \leqslant \|\mathscr{A}(T_L^*(\omega, \mathrm{x})) - T^\circ(\omega, \mathrm{x})\|_{L_2(\Omega\times\mathbb{S}^2)}^2 - \rho < \varepsilon.$$

The proof is completed. □

6. Numerical experiments

In this section, we conduct numerical experiments to compare the ℓ_p-ℓ_2 optimization model (2.10) with the ℓ_1 optimization model (31) in Wallis *et al.* (2017) on the inpainting of band-limited random fields and images from CMB data to show the efficiency of problem (2.10) and Algorithm 4.5.

Following Chen *et al.* (2016), we adapt the nonmonotone proximal gradient (NPG) method to solve subproblem (4.7) in Algorithm 4.5. For completeness, we present the NPG method as follows.

Algorithm 6.1 NPG method for problem (4.7)

Given $\alpha^0 \in \mathscr{F}_e$. Choose $M_{\max} \geqslant M_{\min} > 0$, $\tilde{\eta} > 1$, $b > 0$ and an integer $N \geqslant 0$. Set $n = 0$.

(1) Choose $M_n^0 \in [M_{\min}, M_{\max}]$. Set $M_n = M_n^0$.

(a) Solve the subproblem

$$y \in \arg\min_{\alpha\in\mathbb{C}^d} \left\{\Phi(\alpha) + 2\mathrm{Re}\langle \partial_{\bar{\alpha}} f_{\lambda,\mu}(\alpha^n), \alpha - \alpha^n\rangle + M_n\|\alpha - \alpha^n\|^2\right\}.$$

(b) If $F_{\lambda,\mu}(y) \leqslant \max_{[n-N]_+\leqslant j\leqslant n} F_{\lambda,\mu}(\alpha^j) - b\|y - \alpha^n\|^2$ is satisfied, go to (2). Otherwise set $M_n = \tilde{\eta} M_n$, 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 4.5 at $\lambda = \lambda_k$ $\mu = \mu_k$, we set $M_{\min} = 1$, $M_{\max} = 10^6$, $\tilde{\eta} = 2$, $b = 10^{-4}$, $N = 4$, $M_0^0 = 1$, and for any $n \geqslant 1$,

$$M_n^0 = \min\left\{\max\left\{\frac{|(\alpha^n - \alpha^{n-1})^H(\partial_{\bar{\alpha}} F_{\lambda,\mu}(\alpha^n) - \partial_{\bar{\alpha}} F_{\lambda,\mu}(\alpha^{n-1}))|}{\|\alpha^n - \alpha^{n-1}\|^2}, 1\right\}, 10^6\right\}.$$

Algorithm 6.1 is terminated when

$$\|\alpha^n - \alpha^{n-1}\|_\infty \leqslant \sqrt{\varepsilon^k} \quad \text{and} \quad \frac{|F_{\lambda,\mu}(\alpha^n) - F_{\lambda,\mu}(\alpha^{n-1})|}{\max\{1, |F_{\lambda,\mu}(\alpha^n)|\}} \leqslant \min\{(\varepsilon^k)^{2.2}, 10^{-4}\}.$$

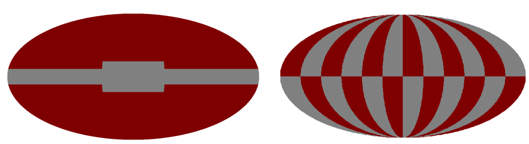
In Algorithm 4.5, we set $\lambda^0 = 20$, $\mu^0 = \varepsilon^0 = 1$, $\varsigma_1 = 2$, $\varsigma_2 = \frac{1}{2}$, $\alpha^0 = \hat{Y}^{-1}\hat{\alpha}^\circ$. The smoothing penalty algorithm is terminated when $\max\{g(\alpha^k)_+, 0.01\varepsilon^k\} \leqslant 10^{-6}$, where ε^k is updated by $\varepsilon^{k+1} = \max\{\varsigma_2\varepsilon^k, 10^{-6}\}$ instead of $\varsigma_2\varepsilon^k$ in the experiments. All codes were written in MATLAB and the realizations were implemented in Python.

6.1 *Random data*

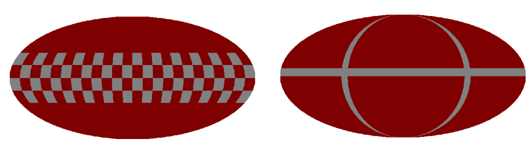
In this subsection, we consider synthetic experiments. We randomly generated instances as follows. First, we randomly choose a subset $D \subset \{0, 1, \ldots, L-1\}$ and generate a group sparse coefficient vector $\alpha_L^{\text{true}} \in \mathbb{C}^{(L+1)^2 \times (L+1)^2}$ such that $\alpha_{l\cdot}^{\text{true}} = 0$ if $l \in D$ and $\alpha_{l\cdot}^{\text{true}} = \alpha_{l\cdot}^{\text{cmb}} / \|\alpha_L^{\text{cmb}}\|^{1.5}$ if $l \in D^c$, where α_L^{cmb} is the coefficient vector with maximum degree L of the CMB 2018 map computed by the HEALPy package. Note that the generated complex coefficients α^{true} are group sparse and the field defined by α^{true} is real-valued.

Next we generate the data for the noise Δ on the HEALPix points with $N_{\text{side}} = 2048$ by the MATLAB command: $\delta\text{randn}(N_{\text{pix}}, 1)$, 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 α^Δ of the noise from $\delta\text{randn}(N_{\text{pix}}, 1)$ and obtain that $\Delta = \sum_{l=0}^{L} \sum_{m=-l}^{l} \alpha_{l,m}^\Delta Y_{l,m}$. We consider the instances in an idealistic scenario and set $\rho = \|\Delta\|_{L_2(\mathbb{S}^2)}^2$. For $\delta = 1$ and 0.1, the values of ρ are around 10^{-3} and 10^{-5}, respectively.

The masks denoted by $\Gamma^c = \mathbb{S}^2 \backslash \Gamma$ are shown in Figure 1.



(a) Γ_1^c (b) Γ_2^c



(c) Γ_3^c (d) Γ_4^c

FIG. 1. Masks (grey part).

In the experiments, we set $p = 0.5$, $\beta_l = (1+10^{-4})^l l^p$ for $l \geqslant 1$ and $\beta_0 = 1$ for Algorithm 4.5. We compare the ℓ_p-ℓ_2 optimization model (2.10) by using Algorithm 4.5 with the following ℓ_1 optimization model (31) in Wallis *et al.* (2017) using the YALL1 method (Yang & Zhang, 2011) under the KKM sampling scheme (Khalid *et al.*, 2014),

$$\min_{\alpha \in \mathbb{C}^d} \quad \|\alpha\|_1 \quad \text{s.t. } \|MA\alpha - b\|^2 \leqslant \rho, \tag{6.1}$$

and the group SPGl1 method (https://friedlander.io/spgl1/) for solving the following problem

$$\min_{\alpha \in \mathbb{C}^d} \quad \textstyle\sum_{l=0}^{L} \beta_l \|\alpha_{l\cdot}\| \quad \text{s.t. } \|MA\alpha - b\|^2 \leqslant \rho, \tag{6.2}$$

where M is a diagonal matrix with elements being 1 or 0, $A \in \mathbb{C}^{n\times d}$ is the measurement matrix, $b = MA\alpha^{\text{true}} + \tilde{\Delta}$ is the observed signal with noise $\tilde{\Delta}$ and $\rho > 0$. Problem (6.2) with $\rho = 0$ is a group version of the weighted ℓ_2 model in Starck *et al.* (2013). Since in the numerical test of Algorithm 4.5 with $p = 0.5$, we set $\beta_l = (1+10^{-4})^l l^p$, for $l \geqslant 1$ and $\beta_0 = 1$, in the experiments of SPGl1 with $p = 1$, we set $\beta_l = 2l+1$ for comparison, which satisfies $\beta_0 = 1$ and $l \leqslant (1+10^{-4})^l l \leqslant 2l+1$ for $1 \leqslant l < 7000$. We set the noise $\tilde{\Delta} = MA\alpha^{\Delta}$, where α^{Δ} is the coefficient vector of noise Δ and $\rho = \|\tilde{\Delta}\|^2$. Following Wallis *et al.* (2017), we set $n = d$. We present the results in Tables 1 and 2. Following Chen & Womersley (2018), nnz $:= \|\alpha_L^* \,\&\, \alpha_L^{\text{true}}\|_{2,0}$ denotes the number of nonzero groups that α_L^* and α_L^{true} have in common and false $:= \|\alpha_L^*\|_{2,0} - \|\alpha_L^* \,\&\, \alpha_L^{\text{true}}\|_{2,0}$ denotes the number of "false positives" where $\alpha_{l\cdot}^{\text{true}}$ is a zero vector, but $\alpha_{l\cdot}^*$ is a nonzero vector. The signal-to-noise ratio (SNR) and relative error are defined by

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

respectively, where $x^{\text{true}} = (T^{\text{true}}(\mathrm{x}_1), \ldots, T^{\text{true}}(\mathrm{x}_n))^T$ and $x^* = (T^*(\mathrm{x}_1), \ldots, T^*(\mathrm{x}_n))^T$ are estimated on the HEALPix points with $n = 12 \times 2048^2$ and α_L^* is the terminating solution,.

From Tables 1 and 2, we can see that our optimization model (2.10) using Algorithm 4.5 achieves smaller relative errors and higher SNR values than the ℓ_1 optimization model (31) in Wallis *et al.* (2017). Although the group SPGl1 method archives small relative errors for some experiments, the SNR values is smaller than ours. Moreover, compared with the ℓ_1 optimization model and group SPGl1, our optimization model (2.10) can exactly recover the number and positions of nonzero groups of α_L^*.

We select some numerical results from Table 1 with $L = 50$ and $\|\alpha_L^{\text{true}}\|_{2,0} = 26$ to show the inpainting quality in Figures 2-8. We observe that our model using Algorithm 4.5 achieves smaller pointwise errors than the ℓ_1 optimization model (31) in Wallis *et al.* (2017) and group SPGl1 method.

Table 1. Numerical results for the inpainting of band-limited random fields with $\delta = 1$.

		Algorithm 4.5					YALL1					Group SPGl1				
Mask	$\\|\alpha_L^{\mathrm{true}}\\|_{2,0}$	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR
$L = 35$																
Γ_1^c	7	0.0032	7	7	0	49.95	0.541	21	7	14	5.28	0.0058	17	7	10	44.73
	11	0.0016	11	11	0	56.18	0.270	13	11	2	11.24	0.0103	28	11	17	39.74
	19	0.0032	19	19	0	49.76	0.238	24	19	5	12.44	0.2033	35	19	16	13.83
Γ_2^c	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
Γ_3^c	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
Γ_4^c	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
$L = 50$																
Γ_1^c	6	0.0035	6	6	0	49.02	0.354	35	6	29	9.00	0.0048	17	6	11	46.41
	16	0.0033	16	16	0	49.52	0.320	41	16	25	9.87	0.0673	43	16	27	27.32
	26	0.0215	26	26	0	33.34	0.304	40	26	14	9.18	0.0975	49	26	23	20.22
Γ_2^c	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	0.0729	45	16	29	22.74
	26	0.0090	26	26	0	40.93	0.664	42	26	16	2.73	0.3549	47	26	21	8.99
Γ_3^c	6	0.0027	6	6	0	51.43	0.212	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
Γ_4^c	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	0.193	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 $\delta = 0.1$.

		Algorithm 4.5					YALL1					Group SPGl1				
Mask	$\\|\alpha_L^{\mathrm{true}}\\|_{2,0}$	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR	RelErr	$\\|\alpha_L^*\\|_{2,0}$	nnz	false	SNR
$L = 35$																
Γ_1^c	7	3.42e-4	7	7	0	69.29	0.541	16	7	9	5.28	0.0011	12	7	5	39.22
	11	1.51e-4	11	11	0	76.39	0.271	12	11	1	11.24	9.63e-4	27	11	16	60.31
	19	3.39e-4	19	19	0	69.38	0.239	20	19	1	12.44	0.1537	36	19	17	16.26
Γ_2^c	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
Γ_3^c	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
Γ_4^c	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
$L = 50$																
Γ_1^c	6	3.55e-4	6	6	0	68.98	0.353	15	6	9	9.01	5.43e-4	13	6	7	66.47
	16	3.82e-4	16	16	0	68.35	0.330	17	16	1	9.87	0.0479	27	16	11	23.43
	26	0.0034	26	26	0	49.28	0.346	34	26	8	9.18	0.0994	50	26	24	20.23
Γ_2^c	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
Γ_3^c	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
Γ_4^c	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

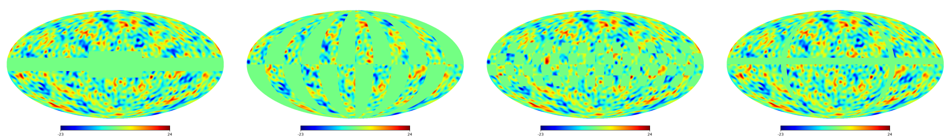


FIG. 2. Observed fields with $L = 50$, $\|\alpha_L^{\text{true}}\|_{2,0} = 6$ and $\delta = 1$ (masked by Γ_1^c, Γ_2^c, Γ_3^c and Γ_4^c from left to right).

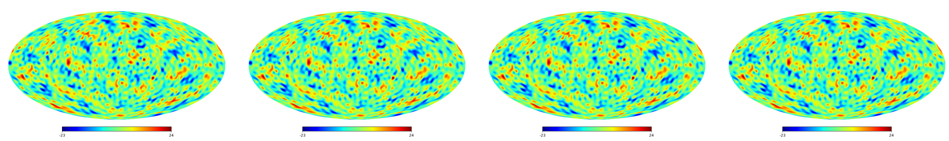


FIG. 3. Inpainted fields by Algorithm 4.5

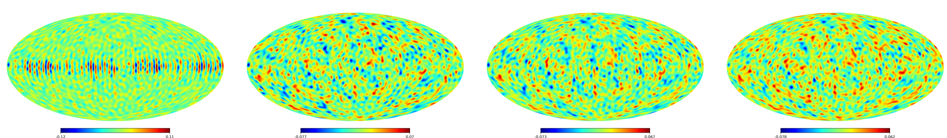


FIG. 4. Pointwise error of inpainted fields in Fig.3 by Algorithm 4.5.

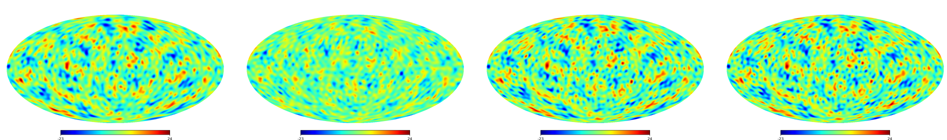


FIG. 5. Inpainted fields by YALL1

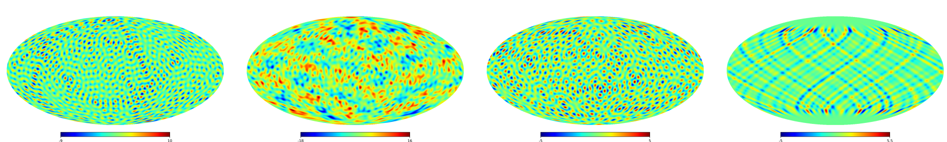


FIG. 6. Pointwise error of inpainted fields in Fig.5 by YALL1.

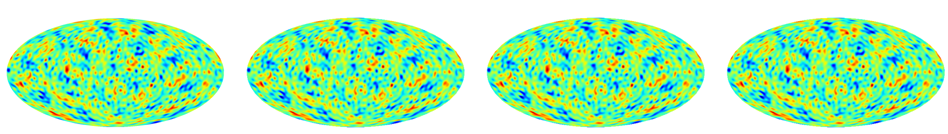


FIG. 7. Inpainted fields by group SPGl1

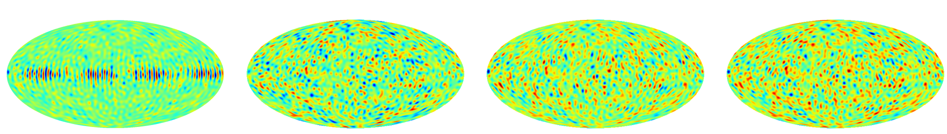
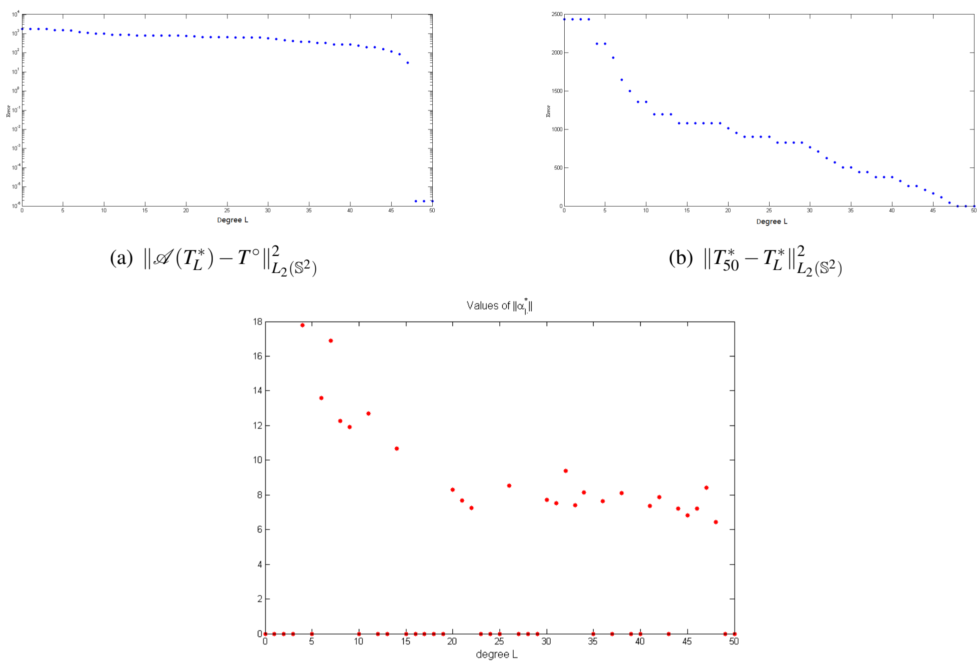


FIG. 8. Pointwise error of inpainted fields in Fig.7 by group SPGl1.

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 Γ_4 and the observed field T° with degree 50. In Figure 9(a) we show the approximation error $\|\mathscr{A}(T_L^*)-T^\circ\|^2_{L_2(\mathbb{S}^2)}$ 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, $\alpha^*_{49\cdot}$ and $\alpha^*_{50\cdot}$ are zero groups of α^*_L for $L = 50$. From our discussion in section 2 and Theorem 5.1, we can guess that the value of L for (a) is 48. We plot the error $\|T^*_{50}-T^*_L\|^2_{L_2(\mathbb{S}^2)}$ for different L in Figure 9(b). We can observe that the error decreases as L increases and the errors are zero for $L = 48, 49, 50$ due to $\alpha^*_{49\cdot} = 0$ and $\alpha^*_{50\cdot} = 0$.



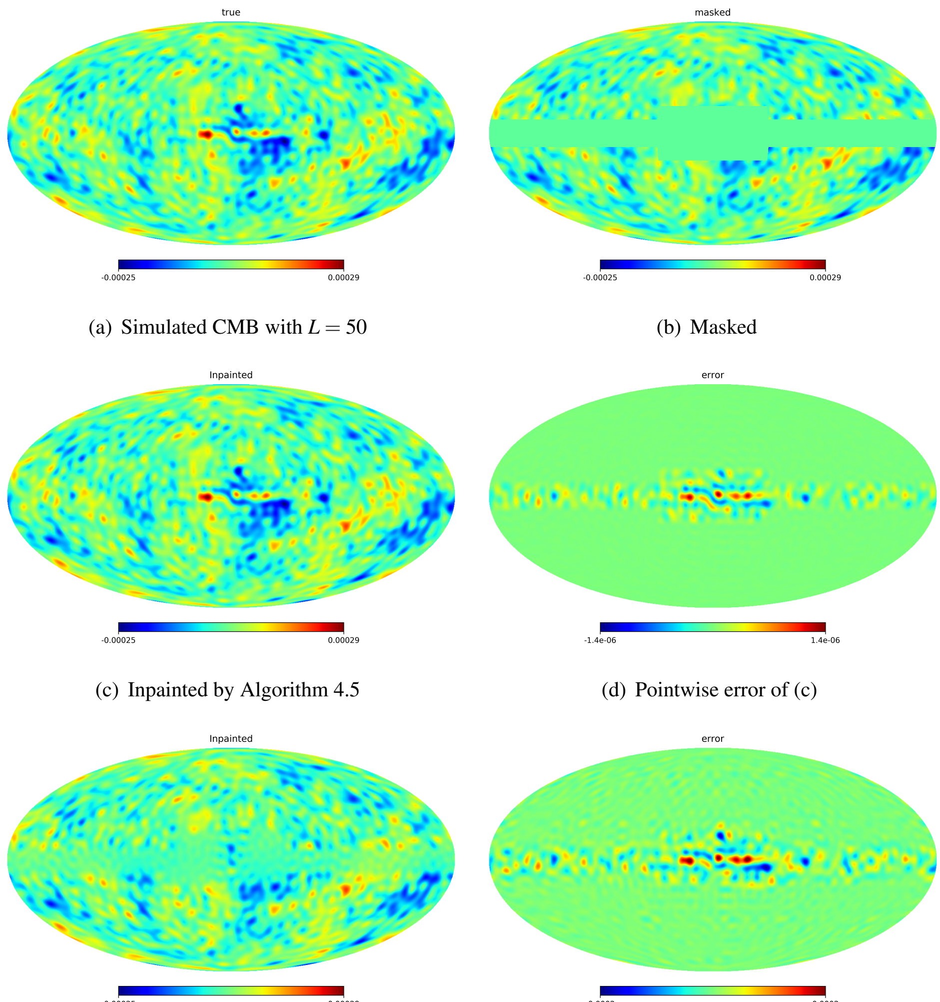
(a) $\|\mathscr{A}(T_L^*)-T^\circ\|^2_{L_2(\mathbb{S}^2)}$ (b) $\|T^*_{50}-T^*_L\|^2_{L_2(\mathbb{S}^2)}$

(c) values of $\|\alpha^*_{l\cdot}\|$

FIG. 9. Approximation errors and values of $\|\alpha^*_{l\cdot}\|$.

6.2 *Real image*

In this section, we conduct experiments on the CMB data which is assumed to be an isotropic random field. We compare our model using Algorithm 4.5 with model (6.2) using the group SPGl1 method. Since the coefficients of the CMB data are not sparse both in elements and groups, we show that our optimization model by using the K-L expansion without discretization of the sphere has good performance. For more detailed study of CMB map, we refer readers to Starck *et al.* (2013); Bucher & Louis (2012); Gruetjen *et al.* (2017); Kim *et al.* (2012) and references therein. We consider the noiseless case (set $\rho = 0$ for both models) and choose the maximum degree of the simulated CMB data to be $L = 50$. The parameters are the same as that in section 6.1. The simulation results are shown in Figure 10. We can see our model achieves small pointwise errors.



(a) Simulated CMB with $L = 50$

(b) Masked

(c) Inpainted by Algorithm 4.5

(d) Pointwise error of (c)

(e) Inpainted by group SPGl1

(f) Pointwise error of (e)

FIG. 10. Inpainting results of the simulated CMB data with maximum degree $L = 50$

7. Conclusion

In this paper, we propose a constrained group 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 (1.6) in a discrete form and derive an equivalence formulation (2.1) of problem (1.6). Based on problem (2.1), we propose a group sparse optimization model (2.2), and derive a lower bound (2.9) for the ℓ_2 norm of nonzero groups of its scaled KKT points. Using this lower bound, we show that problem (2.2) can be solved via the finite-dimensional problem (2.10). For this finite-dimensional problem (2.10) and its penalty problem (2.11), we prove the exact penalization in terms of local minimizers and ε-minimizers. Moreover, we propose a smoothing penalty algorithm for solving problem (2.10) and prove its convergence. We also present the approximation error of the inpainted random field represented by the scaled KKT point in the space $L_2(\Omega \times \mathbb{S}^2)$. Finally, we present numerical results on band-limited random fields on the sphere and the images from CMB data to show the promising performance of our model by using the smoothing penalty algorithm.

8. Funding

In this section, we would like to acknowledge support for this project from the Postdoctoral Fellowship Scheme of CAS AMSS-PolyU Joint Laboratory in Applied Mathematics and Hong Kong Research Grant Council PolyU15300021.

9. Acknowledgement

We would like to thank the Associate Editor and two reviewers for their helpful comments and suggestions on modification of the paper. We are grateful to Prof. Wei Bian, Prof. Ian Sloan and Prof. Buyang Li for their helpful discussion on group sparse optimization problem (2.2) and the space ℓ^p_β for inpainting of random fields on the sphere.

REFERENCES

ABRIAL, P., MOUDDEN, Y., STARCK, J., AFEYAN, B., BOBIN, J., FADILI, J. & NGUYEN, M. (2007) Morphological component analysis and inpainting on the sphere: application in physics and astrophysics. *J. Fourier Anal. Appl.*, **13**, 729–748.

AXLER, S., BOURDON, P. & WADE, R. (2013) *Harmonic Function Theory*, vol. 137. Springer Science & Business Media.

BECK, A. & HALLAK, N. (2019) Optimization problems involving group sparsity terms. *Math. Program.*, **178**, 39–67.

BOUBOULIS, P. & THEODORIDIS, S. (2010) Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS. *IEEE Trans. Signal Process.*, **59**, 964–978.

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

BUCHER, M. & LOUIS, T. (2012) Filling in cosmic microwave background map missing data using constrained Gaussian realizations. *Mon. Notices Royal Astron. Soc.*, **424**, 1694–1713.

CABELLA, P. & MARINUCCI, D. (2009) Statistical challenges in the analysis of cosmic microwave background radiation. *Ann. Appl. Stat.*, **3**, 61–95.

CAMMAROTA, V. & MARINUCCI, D. (2015) The stochastic properties of l_1-regularized spherical Gaussian fields. *Appl. Comput. Harmon. Anal.*, **38**, 262–283.

CHEN, X., XU, F. & YE, Y. (2010) Lower bound theory of nonzero entries in solutions of l_2-l_p minimization. *SIAM J. Sci. Comput.*, **32**, 2832–2852.

CHEN, X. (2012) Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.*, **134**, 71–99.

CHEN, X., LU, Z. & PONG, T. (2016) Penalty methods for a class of non-Lipschitz optimization problems. *SIAM J. Optim.*, **26**, 1465–1492.

CHEN, X. & TOINT, P. L. (2021) High-order evaluation complexity for convexly-constrained optimization with non-Lipschitzian group sparsity terms. *Math. Program.*, **187**, 47–78.

CHEN, X. & WOMERSLEY, R. (2018) Spherical designs and nonconvex minimization for recovery of sparse signals on the sphere. *SIAM J. Imaging Sci.*, **11**, 1390–1415.

CREASEY, P. & LANG, A. (2018) Fast generation of isotropic Gaussian random fields on the sphere. *Monte Carlo Methods Appl.*, **24**, 1–11.

FEENEY, S., MARINUCCI, D., MCEWEN, J., PEIRIS, H., WANDELT, B. & CAMMAROTA, V. (2014) Sparse inpainting and isotropy. *J. Cosmol. Astropart. Phys.*, **2014**, 050.

GIA, Q. L., SLOAN, I., WOMERSLEY, R. & WANG, Y. (2019) Sparse isotropic regularization for spherical harmonic representations of random fields on the sphere. *Appl. Comput. Harmon. Anal.*, **49**, 257–278.

GRUETJEN, H., FERGUSSON, J., LIGUORI, M. & SHELLARD, E. (2017) Using inpainting to construct accurate cut-sky cmb estimators. *Physical Review D*, **95**, 043532.

HILLE, E. (2005) *Analytic Function Theory*, vol. 2. American Mathematical Soc.

HUANG, J., MA, S., XIE, H. & ZHANG, C.-H. (2009) A group bridge approach for variable selection. *Biometrika*, **96**, 339–355.

HUANG, J. & ZHANG, T. (2010) The benefit of group sparsity. *Ann. Statist.*, **38**, 1978–2004.

ISAKOV, V. (2006) *Inverse Problems for Partial Differential Equations*, vol. 127. Springer.

JEONG, J., JUN, M. & GENTON, M. (2017) Spherical process models for global spatial statistics. *Stat. Sci.*, **32**, 501–513.

KHALID, Z., KENNEDY, R. & MCEWEN, J. (2014) An optimal-dimensionality sampling scheme on the sphere with fast spherical harmonic transforms. *IEEE Trans. Signal Process.*, **62**, 4597–4610.

KIM, J., NASELSKY, P. & MANDOLESI, N. (2012) Harmonic in-painting of cosmic microwave background sky by constrained Gaussian realization. *Astrophys. J. Lett.*, **750**, L9.

KREUTZ-DELGADO, K. (2009) The complex gradient operator and the CR-calculus. *arXiv:0906.4835*.

KREYSZIG, E. (1991) *Introductory Functional Analysis with Applications*, vol. 17. John Wiley & Sons.

LANG, A. & SCHWAB, C. (2015) Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.*, **25**, 3047–3094.

LI, C. & CHEN, X. (2022) Isotropic non-Lipschitz regularization for sparse representations of random fields on the sphere. *Math. Comp.*, **91**, 219–243.

MARINUCCI, D. & PECCATI, G. (2011) *Random Fields on the Sphere: Representations, Limit Theorems and Cosmological Applications*. Cambridge: Cambridge University Press.

MAURER, H. & ZOWE, J. (1979) First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. program.*, **16**, 98–110.

OH, H. & LI, T. (2004) Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method. *J. R. Statist. Soc. B*, **66**, 221–238.

PAN, L. & CHEN, X. (2021) Group sparse optimization for images recovery using capped folded concave functions. *SIAM J. Imaging Sci.*, **14**, 1–25.

PORCU, E., ALEGRIA, A. & FURRER, R. (2018) Modeling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.*, **86**, 344–377.

ROCKAFELLAR, R. & WETS, R.-B. (2009) *Variational Analysis*. Springer Science & Business Media.

SORBER, L., BAREL, M. V. & LATHAUWER, L. D. (2012) Unconstrained optimization of real functions in complex variables. *SIAM J. Optim.*, **22**, 879–898.

STARCK, J., FADILI, M. & RASSAT, A. (2013) Low-l CMB analysis and inpainting. *Astron. Astrophys.*, **550**, A15.

STEIN, M. (2007) Spatial variation of total column ozone on a global scale. *Ann. Appl. Stat.*, **1**, 191–210.

SUN, J., QU, Q. & WRIGHT, J. (2018) A geometric analysis of phase retrieval. *Found. Comut. Math.*, **18**, 1131–1198.

WALLIS, C., WIAUX, Y. & MCEWEN, J. (2017) Sparse image reconstruction on the sphere: analysis and synthesis. *IEEE Trans. Image Process.*, **26**, 5176–5187.

YANG, J. & ZHANG, Y. (2011) Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM J. Sci. Comput.*, **33**, 250–278.

YUAN, M. & LIN, Y. (2006) Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. B*, **68**, 49–67.

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

$$\mathscr{T} = \left\{ T(\theta,\phi) \in L_2(\mathbb{S}^2) : T(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \alpha_{l,m} Y_{l,m}(\theta,\phi) \text{ and } \alpha \in \ell_\beta^p \right\}^1 \tag{A.1}$$

be a class of fields with spherical harmonic representations and coefficients belong to ℓ_β^p. Then we show that any $T \in \mathscr{T}$ 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 A.1 If $\alpha \in \ell_\beta^p$, then $\sum_{l=0}^{\infty} \eta^l l \|\alpha_{l\cdot}\| < \infty$.

Proof. Since $\alpha \in \ell_\beta^p$ and $\beta_l = \eta^l l^p$, for any $\varepsilon > 0$ there is a positive integer N such that for any $l > N$, $\eta^l l^p \|\alpha_{l\cdot}\|^p < \varepsilon$. Thus, $l\|\alpha_{l\cdot}\| < (\frac{\varepsilon}{\eta^l})^{\frac{1}{p}}$, $\forall l > N$, which implies that $\eta^l l \|\alpha_{l\cdot}\| < \varepsilon^{\frac{1}{p}} \eta^{\frac{l(p-1)}{p}}$, $\forall l > N$. Thus, for any integers $N_1 \geqslant N_2 \geqslant N$, we have $\sum_{l=N_1}^{N_2} \eta^l l \|\alpha_{l\cdot}\| < \varepsilon^{\frac{1}{p}} \sum_{l=N_1}^{N_2} \eta^{\frac{l(p-1)}{p}} \leqslant \varepsilon^{\frac{1}{p}} \sum_{l=0}^{\infty} \eta^{\frac{l(p-1)}{p}} \leqslant C\varepsilon^{\frac{1}{p}}$, where $C = \sum_{l=0}^{\infty} \eta^{\frac{l(p-1)}{p}} < \infty$. Since ε is arbitrary, we obtain $\sum_{l=0}^{\infty} \eta^l l \|\alpha_{l\cdot}\| < \infty$. □

LEMMA A.2 H is real analytic in $B(0;\hat{r})$.

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 (Axler *et al.*, 2013), $r^l Y_{l,m}(\theta,\phi)$, $l \in \mathbb{N}_0$, $m = -l,\ldots,l$ are harmonic functions on $B(0;\hat{r})$. Then we obtain that H_L, $L \in \mathbb{N}_0$ are harmonics functions on $B(0;\hat{r})$. Moreover, for any $(r,\theta,\phi) \in B(0;\hat{r})$ and $l > 0$,

$$\begin{aligned} |H_l(r,\theta,\phi)| &= \left| \sum_{m=-l}^{l} r^l \alpha_{l,m} Y_{l,m}(\theta,\phi) \right| \leqslant \left| \left(\sum_{m=-l}^{l} (r^l \alpha_{l,m})^2 \right)^{\frac{1}{2}} \left(\sum_{m=-l}^{l} (Y_{l,m}(\theta,\phi))^2 \right)^{\frac{1}{2}} \right| \\ &= r^l \|\alpha_{l\cdot}\| \sqrt{\tfrac{2l+1}{4\pi}} < \eta^l l \|\alpha_{l\cdot}\|, \end{aligned}$$

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 $r \in [0,\hat{r})$ and $\hat{r} < \eta$.

By Lemma A.1, $\sum_{l=0}^{\infty} \eta^l l \|\alpha_{l\cdot}\| < \infty$, then for any $\varepsilon > 0$, there exists N such that for any $L' > L > N$, we have that for any $(r,\theta,\phi) \in B(0;\hat{r})$,

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

[1]With a slight abuse of notation, in this subsection, $T(\theta,\phi) \equiv T(\mathrm{x})$ and $Y_{l,m}(\theta,\phi) \equiv Y_{l,m}(\mathrm{x})$, where $\mathrm{x} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T \in \mathbb{S}^2$.

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

By Axler *et al.* (2013, Theorem 1.23), H is harmonic on $B(0;\hat{r})$. Moreover, by Axler *et al.* (2013, Theorem 1.28), H is real analytic in $B(0;\hat{r})$. This completes the proof. □

LEMMA A.3 For any $T \in \mathscr{T}$, 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 A.2, H is real analytic in $B(0;\hat{r})$. Then by Hille (2005, Theorem 8.1.3), we obtain $H \equiv 0$ in $B(0;\hat{r})$ which implies that $T \equiv 0$ on $\mathbb{S}^2$. This completes the proof. □

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

THEOREM A.1 For any $T \in \mathscr{T}$, if $T = 0$ on Γ, then $T \equiv 0$ on $\mathbb{S}^2$.

Proof. Since the coefficients $\alpha \in \ell^p_\beta$ and Γ has an open subset, we know from Lemma A.3 that if $T = 0$ on Γ then $T \equiv 0$ on $\mathbb{S}^2$. □

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 $T \in \mathscr{T}$, $\mathscr{A}(T) \equiv 0$ if and only if $\alpha = 0$.

Appendix B. Wirtinger gradient

In this appendix, we briefly introduce the Wirtinger gradient of real-valued functions with complex variables over the finite-dimensional space $\mathbb{C}^n$ and the infinite-dimensional space ℓ^2. For more details about Wirtinger's calculus, we refer to Brandwood (1983); Kreutz-Delgado (2009); Bouboulis & Theodoridis (2010); Li & Chen (2022); Sorber *et al.* (2012); Sun *et al.* (2018).

We first introduce the Wirtinger gradient in ℓ^2. Let $f : \ell^2 \to \mathbb{R}$ be a real-valued function of a vector of complex variables $z \in \ell^2$. Let $z = x + iy$, where x, y are the real and image parts of z, respectively. We write $f(z) = f_1(x+iy) = f_1^r(x,y) + i f_1^i(x,y)$, where f_1^r and f_1^i are the real and image parts of f_1, respectively. Since f_1 is real-valued, we have $f_1^i \equiv 0$. We say f is Wirtinger differentiable, if f_1^r is Fréchet differentiable with respect to x and y, respectively. Following Bouboulis & Theodoridis (2010), under the conjugate coordinates $(z^T, \bar{z}^T)^T \in \ell^2 \times \ell^2$, the Wirtinger gradient of a Wirtinger differentiable function f at $z \in \ell^2$ is

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

where $\partial_z f(z) := \frac{1}{2}\left(\frac{\partial f_1^r(x,y)}{\partial x} - i\frac{\partial f_1^r(x,y)}{\partial y}\right)$, $\partial_{\bar{z}} f(z) := \frac{1}{2}\left(\frac{\partial f_1^r(x,y)}{\partial x} + i\frac{\partial f_1^r(x,y)}{\partial y}\right)$. Since f is real-valued, $\overline{\partial_z f} = \partial_{\bar{z}} f$. Thus,

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

We say $z^* \in \ell^2$ is a stationary point of f if it satisfies $\partial_{\bar{z}} f(z^*) = 0$.

Following Brandwood (1983); Kreutz-Delgado (2009), the Wirtinger gradient of a function $g : \mathbb{C}^n \to \mathbb{R}$ can be defined in the same way as above. In particular, we say g is Fréchet differentiable, if its real part g_1^r is Fréchet differentiable, and the Wirtinger gradient of g is

$$\nabla g(z) = \begin{pmatrix} \partial_z g(z) \\ \partial_{\bar{z}} g(z) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \frac{\partial g_1^r(x,y)}{\partial x} - i\frac{\partial g_1^r(x,y)}{\partial y} \\ \frac{\partial g_1^r(x,y)}{\partial x} + i\frac{\partial g_1^r(x,y)}{\partial y} \end{pmatrix}.$$