

# クリギング法のための 非凸最適化に基づく観測点選択

数理情報学専攻 48156202 石川 大智

指導教員 長尾 大道 准教授

## 1 はじめに

地球科学分野や環境科学分野では、興味のある空間の広さに対して比較的少数の点的観測しか実施できないような状況があり、観測範囲よりも広い空間や、観測点分布よりも密な解像度に関する知見を得るための手段として、空間補間は重要である。クリギング法 [4] は、空間内の各点に定まる確率変数に対して、観測値から空間補間を行うための手法であり、確率変数の空間共分散の構造を決めているセミバリオグラムと呼ばれる空間相関を記述する関数を観測データから推定したのち、予測値が最良線形不偏予測量となるように、セミバリオグラムと観測データから定まる線形方程式を解くという手続きで行われる。クリギング法をナイーブに実装すると、空間上のある一つの点における確率変数の推定のために  $n$  個の観測点データを用いる場合には、 $O(n^3)$  の計算量を要するという問題点がある。そこで、本研究では、クリギング法における観測点選択を非凸最適化に基づいて行うことを提案する。本研究で取り扱う非凸正則化項とは、Lou et al. [3] によって提案された  $l_1 - l_2$  正則化項や、Gotoh et al. [2] によって  $l_0$  制約付き問題との同値な定式化が提案された  $l_1 - \text{largest-}K$  正則化項である。これらの最適化問題は、制約領域が閉凸集合であり、かつ目的関数が 2 つの凸関数の差で表すことが可能である場合の計画問題である、DC 計画問題というクラスに属している。そして、その解は、DC 計画問題に対する汎用解法の一つとして知られている Tao and Souad [5] によって提案された DC アルゴリズムを用いて、近似的に得ることができる。数値実験では lasso,  $l_1 - l_2$  正則化,  $l_1 - \text{largest-}K$  正則化を比較し、提案する非凸最適化、特に  $l_1 - \text{largest-}K$  正則化が観測点選択を安定的に行えることを示す。

## 2 クリギング法

クリギング法は確率場  $Z = \{Z(\mathbf{s}) : \mathbf{s} \in D\}$  が固有定常である、すなわち任意の 2 地点  $\mathbf{s}_1, \mathbf{s}_2 \in D$  について  $E[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)] = 0$  かつ  $\text{Var}[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]$  が  $\mathbf{d} = \mathbf{s}_1 - \mathbf{s}_2$  のみに依存するときに適用できる。クリギ

ング法は、地点  $\mathbf{s}_1, \dots, \mathbf{s}_n$  での観測値が与えられたとき、セミバリオグラム  $\gamma(\mathbf{d}) = \frac{1}{2} \text{Var}[Z(\mathbf{s} + \mathbf{d}) - Z(\mathbf{s})]$  を推定し、地点  $\mathbf{s}_0$  での推定値  $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n w_i Z(\mathbf{s}_i)$  が線形最小分散推定量となるように最適化問題

$$\begin{aligned} & \text{minimize} && -\mathbf{w}^\top \mathbf{\Gamma} \mathbf{w} + 2\gamma_0^\top \mathbf{w} \\ & \text{subject to} && \mathbf{1}^\top \mathbf{w} = 1 \end{aligned} \quad (1)$$

(ただし  $\mathbf{\Gamma} = [\gamma(\mathbf{s}_i - \mathbf{s}_j)]_{1 \leq i, j \leq n}$ ,  $\gamma_0 = [\gamma(\mathbf{s}_i - \mathbf{s}_0)]_{1 \leq i \leq n}$ ,  $\mathbf{w} = [w_i]_{1 \leq i \leq n}$ .) の解を求めるという手続きで行われる。最適化問題 (1) の目的関数は  $\hat{Z}(\mathbf{s}_0)$  の平均自乗誤差、制約領域は  $\hat{Z}(\mathbf{s}_0)$  の不偏性に対応している。

## 3 提案手法

本研究では、クリギング法で観測点選択を行なうために、推定値  $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n w_i Z(\mathbf{s}_i)$  に対応する  $\mathbf{w}$  を最適化問題

$$\begin{aligned} & \text{minimize} && -\mathbf{w}^\top \mathbf{\Gamma} \mathbf{w} + 2\gamma_0^\top \mathbf{w} + \lambda h(\mathbf{w}) \\ & \text{subject to} && \mathbf{1}^\top \mathbf{w} = 1 \end{aligned} \quad (2)$$

の解とすることを方針とする。最適化問題 (2) の正則化項  $h(\mathbf{w})$  について以下の命題が成り立つ：

**命題 1.**  $\mathbf{e}_i = [0 \cdots 0 \ 1 \ 0 \cdots 0]^\top \in \mathbb{R}^n$  とする。凸正則化項  $h$  について  $h(\mathbf{e}_1) = \cdots = h(\mathbf{e}_n) = c_{\text{sparse}}$  が成り立つとき、 $\mathbf{1}^\top \mathbf{w} = 1, \mathbf{w} \geq \mathbf{0}$  を満たす  $\mathbf{w}$  について、 $h(\mathbf{w}) \leq c_{\text{sparse}}$  が成り立つ。

そこで、本研究では、クリギング法において非凸正則化項  $h(\mathbf{w})$  を用いることを提案し、 $l_1 - l_2$  正則化項 [3]  $h(\mathbf{w}) = \|\mathbf{w}\|_1 - \|\mathbf{w}\|_2$  と、 $l_1 - \text{largest-}K$  正則化項 [2]  $h(\mathbf{w}) = \|\mathbf{w}\|_1 - \|\mathbf{w}\|_K$  を取り扱う。ただし、 $\|\mathbf{w}\|_K \stackrel{\text{def}}{=} (\text{絶対値上位 } K \text{ 成分の和})$ 。これらの非凸正則化項  $h$  を最適化問題 (2) に導入した問題は DC 計画問題 [1]、すなわち、凸関数  $f, g$  と閉凸集合  $S$  が存在して、

$$\begin{aligned} & \text{minimize} && f(\mathbf{w}) - g(\mathbf{w}) \\ & \text{subject to} && \mathbf{w} \in S \end{aligned} \quad (3)$$

と定式化することができるクラスに帰着できる。本研究では最適化問題 (2) の目的関数について、凸関数  $f, g$

表 1: 各正則化項に対応させる凸関数  $f, g$  の例.

正則化項	$f$	$g$
$\ell_1$	$-\mathbf{w}^\top \Gamma \mathbf{w} + 2\gamma_0^\top \mathbf{w} + \ \mathbf{w}\ _1$	0
$\ell_1 - \ell_2$	$-\mathbf{w}^\top \Gamma \mathbf{w} + 2\gamma_0^\top \mathbf{w} + \ \mathbf{w}\ _1$	$\ \mathbf{w}\ _2$
$\ell_1 - \text{largest-}K$	$-\mathbf{w}^\top \Gamma \mathbf{w} + 2\gamma_0^\top \mathbf{w} + \ \mathbf{w}\ _1$	$\ \mathbf{w}\ _K$

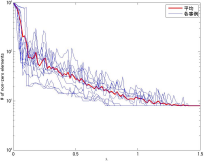
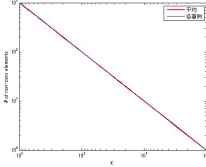
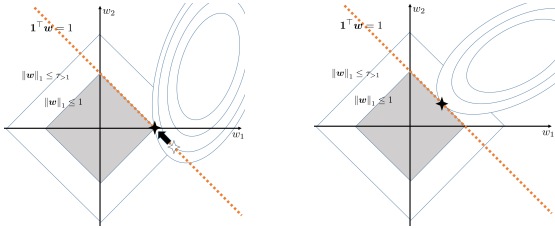

 (a)  $\ell_1$  正則化

 (b)  $\ell_1 - \text{largest-}K$  正則化

 図 1: 各正則化におけるパラメータと  $\mathbf{w}$  の非零要素数の関係. 青線は各観測値集合  $Z_A$  に対するプロット, 赤線はそれらの平均を表す.

 (a)  $\mathbf{w}$  に負の成分あり.

 (b)  $\mathbf{w}$  に負の成分なし.

 図 2:  $\ell_1$  正則化における  $\mathbf{w}$  の観測点選択の振る舞い.

を表 1 のように定めて, DC アルゴリズム [5] によって数値的に解を求めることを提案した.

## 4 数値実験

まず, 数値実験の状況設定について述べる. 確率場  $Z = \{Z(\mathbf{s}) : \mathbf{s} \in D\}$  を  $D = [0, 1]^2$  として, 確率変数  $Z$  が従うセミバリオグラムを  $\gamma(d) = 50(1 - e^{-d/0.25})$  とした. 領域  $D$  上の観測点の集合  $A$  は, 推定点  $\mathbf{s}_0 = (1/2, 1/2)$  について  $\mathbf{s}_0 \in A, |A| = 1000$  を満たす一様な配置からランダムに一つ固定し, 観測値集合  $Z_A = \{Z(\mathbf{s}) : \mathbf{s} \in A\}$  を 10 組生成した. 推定点  $\mathbf{s}_0 = (1/2, 1/2)$  における実際の観測値  $Z(\mathbf{s}_0)$  と  $Z_{A \setminus \{\mathbf{s}_0\}}$  から提案手法による推定値  $\hat{Z}(\mathbf{s}_0)$  の平均自乗誤差を比較することによって,  $\ell_1, \ell_1 - \ell_2, \ell_1 - \text{largest-}K$  の各正則化の性能について比較した. 本要旨では  $\ell_1, \ell_1 - \text{largest-}K$  正則化の結果を載せる. 図 1 に,  $\ell_1, \ell_1 - \text{largest-}K$  正則化における正則化パラメータと  $\mathbf{w}$  の非零要素数の関係を示す. 図 1a から,  $\ell_1$  正則化では  $\lambda$  の増加に対し

て  $\mathbf{w}$  の非零要素数が増加する現象を見られる. 一方, 図 1b から,  $\ell_1 - \text{largest-}K$  正則化ではパラメータ  $K$  の値は, 得られた  $\mathbf{w}$  の非零要素数であり,  $\ell_1 - \text{largest-}K$  正則化が安定的に観測点選択を行なえることが見て取れる. 図 2 に, 最適化問題

$$\begin{aligned} & \text{minimize} && -\mathbf{w}^\top \Gamma \mathbf{w} + 2\gamma_0^\top \mathbf{w} \\ & \text{subject to} && \mathbf{1}^\top \mathbf{w} = 1, \|\mathbf{w}\|_1 \leq \tau \end{aligned} \quad (4)$$

において  $\tau$  が  $\infty$  から減少するときの  $\mathbf{w}$  の振る舞いを示す. 図 2a のような状況では,  $\tau (> 1)$  が十分小さくなると目的関数の最小点は実効領域の端点と一致することになり,  $\tau$  の減少にともなって最小点に変化して疎な点に近づく. 一方, 図 2b のような状況では,  $\tau$  の減少に伴って実効領域が変化しても, 実効領域の中で目的関数の最小値を与える点は単体  $\{\mathbf{w} \mid \mathbf{1}^\top \mathbf{w} = 1, \mathbf{w} \geq \mathbf{0}\}$  上から動かないので, 正則化が機能しない.

## 5 結論

本研究では, 空間補間の標準的な手法であるクリギング法のための観測点選択を行なう手法として, 最適化問題 (2) の解  $\mathbf{w}$  を用いて, 推定値  $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n w_i Z(\mathbf{s}_i)$  を与えることを提案した. その際,  $\ell_1$  正則化など, 凸な正則化項では変数選択の目的が達成できず, 導入する正則化項が非凸であることが必要であることを示した. そして, 最適化問題 (2) を DC 計画問題に帰着させて, DC アルゴリズムによって数値的に得られた解を用いることを提案した.

## 参考文献

- [1] L. T. H. An and P. D. Tao: The DC(Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems, *Annals of Operations Research*, vol. 133 (2005), pp.23–46.
- [2] J. Gotoh, A. Takeda and K. Tono: DC Formulations and Algorithms for Sparse Optimization Problems, *Mathematical Programming* (2017), pp. 141–176.
- [3] Y. Lou, P. Yin, Q. He and J. Xin: Computing Sparse Representation in a Highly Coherent Dictionary based on Difference of L1 and L2, *Journal of Scientific Computing*, vol. 64 (2015), pp 178–196.
- [4] G. Matheron: Principles of geostatistics, *Economic Geology*, vol. 58 (1963), pp. 1246–1266.
- [5] P. D. Tao and E. B. Souad: Algorithms for Solving a Class of Nonconvex Optimization Problems. Methods of Subgradients, *North-Holland Mathematics Studies*, vol. 129 (1986), pp. 249–271.
- [6] R. Tibshirani: Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B*, vol. 58 (1996), pp. 267–288.