

Boder Bases による代数曲線回帰について

手島悠人

基幹理工学研究科 数学応用数理専攻

5110A030-8

指導教員名 楢 元

2012/02/07

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

- 既存の方法として最小二乗法が良く用いられる.

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

- 既存の方法として**最小二乗法**が良く用いられる.
 - step1, データの収集

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

- 既存の方法として**最小二乗法**が良く用いられる.
 - step1, データの収集
 - step2, 多項式の次数を決める.

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

- 既存の方法として**最小二乗法**が良く用いられる.
 - step1, データの収集
 - step2, 多項式の次数を決める.
 - step3, 多項式の係数を決める.

Question

誤差を持つ観測値 $\mathbb{X} = \{(x_1, y_1), \dots, (x_s, y_s)\} \subset \mathbb{R}^2$ が与えられているとする. これらはある代数曲線上の点だと考えられる.
その代数曲線を定める多項式 $f(x, y) \in \mathbb{R}[x, y]$ を求めよ.

- 既存の方法として**最小二乗法**が良く用いられる.
 - step1, データの収集
 - step2, 多項式の次数を決める. $\rightarrow$ 次数を決める数学的な評価方法が無い.
 - step3, 多項式の係数を決める.

[2] において, **Border Bases** を用いて近似式を得ている例がある.
Border Bases...近年, 代数幾何学や計算代数等で注目

[2] において, **Border Bases** を用いて近似式を得ている例がある.

Border Bases...近年, 代数幾何学や計算代数等で注目

→ この方法では, 次数を決める必要はない.

他にも有用だと考えられる点はあるが, ほとんど記述はない.

[2] において, **Border Bases** を用いて近似式を得ている例がある.

Border Bases...近年, 代数幾何学や計算代数等で注目

→ この方法では, 次数を決める必要はない.

他にも有用だと考えられる点はあるが, ほとんど記述はない.

→ このアイデアを元に **Border Bases 法**の性質や課題を調べた.

Definition

$P = \mathbb{R}[x_1, \dots, x_n]$, $\mathcal{O} \in \mathbb{T}^n (= \{x_1^{a_1} \cdots x_n^{a_n} \mid a_i \in \mathbb{Z}_{\geq 0}\})$ とする.

(a) $\mathcal{O}$: **order ideal** $\xLeftrightarrow{\text{def}}$ 「 $t \in \mathbb{T}^n$ に対し $\exists t' \in \mathcal{O}$ s.t. $t \mid t' \Rightarrow t \in \mathcal{O}$ 」
を満たす power product の集合.

(b) $\mathcal{I} \subset P$: zero-dimensional ideal (零点集合が有限な ideal) とする.

$\mathcal{O}$: **quotient basis** for $\mathcal{I}$ $\xLeftrightarrow{\text{def}}$ $\mathcal{O}$: order ideal かつ $P/\mathcal{I}$ の basis

(c) $\partial\mathcal{O} = (x_1\mathcal{O} \cup \dots \cup x_n\mathcal{O}) \setminus \mathcal{O}$ を **border** of $\mathcal{O}$ という.

Border Bases とは

Example

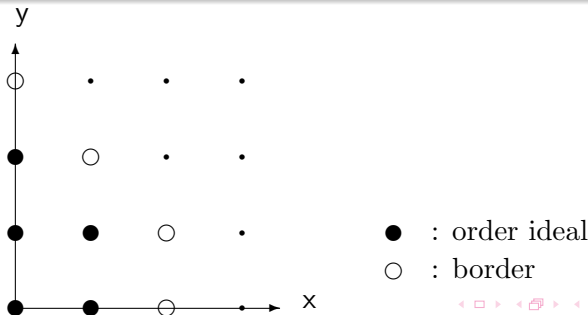
$$\mathcal{I} = \langle x^2 + xy + y^2 - x - y, xy^2 - y, y^3 - y + 1 \rangle$$

次数付き逆辞書式順序で $\text{LT}(\mathcal{I}) = \{x^2, xy^2, y^3\}$

$\mathcal{O} = \{1, x, y, xy, y^2\}$ は $P/\mathcal{I}$ の basis かつ order ideal

→ $\mathcal{O}$ は quotient basis for $\mathcal{I}$

border は $\partial\mathcal{O} = \{x^2, x^2y, xy^2, y^3\}$



Definition

$\mathcal{O} = \{t_1, \dots, t_\mu\}$: order ideal , $\partial\mathcal{O} = \{b_1, \dots, b_\nu\}$: border of $\mathcal{O}$
 $g_i = b_i - \sum_{j=1}^{\mu} \alpha_{ij} t_j$ とし, $\mathcal{B} = \{g_1, \dots, g_\nu\}$ とおく.
 $\mathcal{I} \subset P$: zero dimensional ideal , $\mathcal{B} \subset \mathcal{I}$ とする.

このとき, $\mathcal{O}$ が quotient basis for $\mathcal{I}$ ならば $\mathcal{B}$ を $\mathcal{O}$ - **border basis** of $\mathcal{I}$ という.

Border Bases とは

Definition

$\mathcal{O} = \{t_1, \dots, t_\mu\}$: order ideal , $\partial\mathcal{O} = \{b_1, \dots, b_\nu\}$: border of $\mathcal{O}$
 $g_i = b_j - \sum_{i=1}^{\mu} \alpha_{ij} t_i$ とし, $\mathcal{B} = \{g_1, \dots, g_\nu\}$ とおく.
 $\mathcal{I} \subset P$: zero dimensional ideal , $\mathcal{B} \subset \mathcal{I}$ とする.

このとき, $\mathcal{O}$ が quotient basis for $\mathcal{I}$ ならば $\mathcal{B}$ を $\mathcal{O}$ - **border basis** of $\mathcal{I}$ という.

Proposition

$\mathcal{B} : \mathcal{O}$ -border basis of $\mathcal{I} \quad \Rightarrow \quad \mathcal{I} = \mathcal{B}$

Example

$$f_1 = \frac{1}{4}x^2 + y^2 - 1, \quad f_2 = x^2 + \frac{1}{4}y^2 - 1, \quad \mathcal{I} = \langle f_1, f_2 \rangle$$

$$\tilde{f}_1 = \frac{1}{4}x^2 + y^2 + \varepsilon xy - 1, \quad \tilde{f}_2 = x^2 + \frac{1}{4}y^2 + \varepsilon xy - 1, \quad \tilde{\mathcal{I}} = \langle \tilde{f}_1, \tilde{f}_2 \rangle$$

	reduced G-Basis	B-Basis
$\mathcal{I}$	$\{x^2 - \frac{4}{5}, y^2 - \frac{4}{5}\}$	$\{x^2 - \frac{4}{5}, x^2y - \frac{4}{5}y, xy^2 - \frac{4}{5}x, y^2 - \frac{4}{5}\}$
$\tilde{\mathcal{I}}$	$\{x^2 - y^2, xy + \frac{5}{4\varepsilon}y^2 - \frac{1}{\varepsilon}, y^3 - \frac{16\varepsilon}{16\varepsilon^2-25}x + \frac{20}{16\varepsilon^2-25}y\}$	$\{x^2 + \frac{4\varepsilon}{5}xy - \frac{4}{5}, x^2y - \frac{16\varepsilon}{16\varepsilon^2-25}x + \frac{20}{16\varepsilon^2-25}y, xy^2 + \frac{20}{16\varepsilon^2-25}y - \frac{16\varepsilon}{16\varepsilon^2-25}y, y^2 + \frac{4\varepsilon}{5}xy - \frac{4}{5}\}$

Definition

$\mathbb{X} = \{p_1, \dots, p_s\} \subset \mathbb{R}^n$, $\mathcal{O} = \{t_1, \dots, t_\mu\}$: order ideal とする.

このとき, $\mathcal{O}$ が **stable** with respect to $\mathbb{X}^\varepsilon$

$\stackrel{\text{def}}{\iff} \forall \tilde{\mathbb{X}} : \text{admissible perturbation of } \mathbb{X}^\varepsilon \text{ において}$

$$M_{\mathcal{O}}(\tilde{\mathbb{X}}) = \begin{bmatrix} t_1(\tilde{p}_1) & \cdots & t_\mu(\tilde{p}_1) \\ \vdots & \ddots & \vdots \\ t_1(\tilde{p}_s) & \cdots & t_\mu(\tilde{p}_s) \end{bmatrix} \in \text{Mat}_{s,\mu}(\mathbb{R})$$

が full rank である.

Border Bases とは

Definition

$\mathbb{X} = \{p_1, \dots, p_s\} \subset \mathbb{R}^n$, $\mathcal{O} = \{t_1, \dots, t_\mu\}$: order ideal とする.

このとき, $\mathcal{O}$ が **stable** with respect to $\mathbb{X}^\varepsilon$

$\stackrel{\text{def}}{\iff} \forall \tilde{\mathbb{X}} : \text{admissible perturbation of } \mathbb{X}^\varepsilon \text{ において}$

$$M_{\mathcal{O}}(\tilde{\mathbb{X}}) = \begin{bmatrix} t_1(\tilde{p}_1) & \cdots & t_\mu(\tilde{p}_1) \\ \vdots & \ddots & \vdots \\ t_1(\tilde{p}_s) & \cdots & t_\mu(\tilde{p}_s) \end{bmatrix} \in \text{Mat}_{s,\mu}(\mathbb{R})$$

が full rank である.

stable な order ideal から構成されるような Border Bases を
Stable Border Bases という.

Proposition

$\mathbb{X} \subset \mathbb{R}^n$, $\mathcal{O} = \{t_1, \dots, t_s\}$ quotient basis for $\mathcal{I}(\mathbb{X})$,
 $\mathcal{O}$: stable with respect to $\mathbb{X}^\varepsilon$, $\partial\mathcal{O} = \{b_1, \dots, b_\nu\}$ とする.

- (a) $\forall \tilde{\mathbb{X}}$: admissible perturbation of $\mathbb{X}^\varepsilon$ について,
 $\mathcal{I}(\tilde{\mathbb{X}})$ は $\mathcal{O}$ -Border Basis を持つ.
- (b) $\tilde{\mathcal{B}}$: $\mathcal{O}$ -Border Basis of $\mathcal{I}(\tilde{\mathbb{X}})$ は $\tilde{\mathcal{B}} = \{g_1, \dots, g_\nu\}$
ただし, $g_i = b_i - \sum_{j=1}^s \alpha_{ij} t_j$

Proposition

$\mathbb{X} \subset \mathbb{R}^n$, $\mathcal{O} = \{t_1, \dots, t_s\}$ quotient basis for $\mathcal{I}(\mathbb{X})$,
 $\mathcal{O}$: stable with respect to $\mathbb{X}^\varepsilon$, $\partial\mathcal{O} = \{b_1, \dots, b_\nu\}$ とする.

- (a) $\forall \tilde{\mathbb{X}}$: admissible perturbation of $\mathbb{X}^\varepsilon$ について,
 $\mathcal{I}(\tilde{\mathbb{X}})$ は $\mathcal{O}$ -Border Basis を持つ.
- (b) $\tilde{\mathcal{B}}$: $\mathcal{O}$ -Border Basis of $\mathcal{I}(\tilde{\mathbb{X}})$ は $\tilde{\mathcal{B}} = \{g_1, \dots, g_\nu\}$
ただし, $g_i = b_i - \sum_{j=1}^s \alpha_{ij} t_j$

つまり, $\mathcal{I}(\mathbb{X})$ の Stable Border Basis と $\mathcal{I}(\tilde{\mathbb{X}})$ の Stable Border Basis は
係数部分が少し異なるだけ.

Proposition

$\mathbb{X} \subset \mathbb{R}^n$, $\mathcal{O} = \{t_1, \dots, t_s\}$ quotient basis for $\mathcal{I}(\mathbb{X})$,
 $\mathcal{O}$: stable with respect to $\mathbb{X}^\varepsilon$, $\partial\mathcal{O} = \{b_1, \dots, b_\nu\}$ とする.

- (a) $\forall \tilde{\mathbb{X}}$: admissible perturbation of $\mathbb{X}^\varepsilon$ について,
 $\mathcal{I}(\tilde{\mathbb{X}})$ は $\mathcal{O}$ -Border Basis を持つ.
- (b) $\tilde{\mathcal{B}}$: $\mathcal{O}$ -Border Basis of $\mathcal{I}(\tilde{\mathbb{X}})$ は $\tilde{\mathcal{B}} = \{g_1, \dots, g_\nu\}$
ただし, $g_i = b_i - \sum_{j=1}^s \alpha_{ij} t_j$

つまり, $\mathcal{I}(\mathbb{X})$ の Stable Border Basis と $\mathcal{I}(\tilde{\mathbb{X}})$ の Stable Border Basis は
係数部分が少し異なるだけ.

→ $\tilde{\mathbb{X}}$ から $\mathcal{I}(\mathbb{X})$ を推定 (近似) できる.

Example.1 直線回帰

$\tilde{\mathbb{X}} = \{(0, 0), (1, 3), (2, 5.9), (3.1, 9)\}$ が与えられている.

$\xrightarrow{\text{CoCoA}}$ Stable Order Ideal : $\mathcal{O} = \{1, y, y^2, y^3\}$
border of $\mathcal{O}$: $\partial\mathcal{O} = \{x, xy, xy^2, xy^3, y^4\}$

$$\text{Stable Border Basis : } \mathcal{B} = \left\{ \begin{array}{l} x + 0.000031y^3 - 0.00022y^2 - 0.327y, \\ xy - 0.00017y^3 - 0.330y^2 + 0.0049y, \\ xy^2 - 0.360y^3 + 0.168y^2 - 0.267y, \\ y^4 - 17.9y^3 + 97.8y^2 - 159.3y, \\ xy^3 - 6.27y^3 + 34.93y^2 - 57.32y \end{array} \right\}$$

Example.1 直線回帰

$$x + 0.000031y^3 - 0.00022y^2 - 0.327y \approx x - 0.327y$$

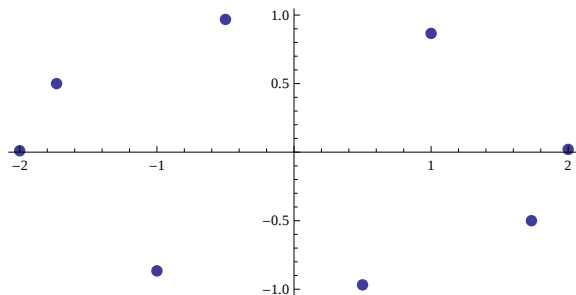
$\therefore \mathcal{I}(\mathbb{X}) = \langle x - 0.327y \rangle$ と推定できる.

最小二乗法では, $y = 2.90x - 0.0494$ と推定できる.

Example.2 二次曲線回帰

$$\tilde{\mathbb{X}} = \{(1, 0.86603), (-1, -0.86603), (1.73205, -0.5), \\ (-1.73205, 0.5), (2, 0.02), (-2.001, 0.01), \\ (0.5, -0.96825), (-0.5, 0.96825)\}$$

が与えられている.



Example.2 二次曲線回帰

$\xrightarrow{\text{CoCoA}}$ Stable Order Ideal : $\mathcal{O} = \{1, x, y, y^2, xy, x^2, y^3, xy^2\}$

$$\mathcal{B} = \left\{ \begin{array}{l} x^3 + 4.0039xy^2 + 0.00759y^3 + 0.001005x^2 - 0.000134xy \\ \quad + 4.021y^2 - 4.00228x - 0.00634y - 4.02, \\ \quad \vdots \end{array} \right\}$$

Example.2 二次曲線回帰

$\xrightarrow{\text{CoCoA}}$ Stable Order Ideal : $\mathcal{O} = \{1, x, y, y^2, xy, x^2, y^3, xy^2\}$

$$\mathcal{B} = \left\{ \begin{array}{l} x^3 + 4.0039xy^2 + 0.00759y^3 + 0.001005x^2 - 0.000134xy \\ \quad + 4.021y^2 - 4.00228x - 0.00634y - 4.02, \\ \quad \vdots \end{array} \right\}$$

10^{-2} 以下を無視すると,

$$\begin{aligned} x^3 + 4.0039xy^2 + 0.00759y^3 + 0.001005x^2 - 0.000134xy \\ + 4.021y^2 - 4.00228x - 0.00634y - 4.02 \end{aligned}$$

$$\approx x^3 + 4xy^2 + x^2 + 4y^2 - 4x - 4 = (x^2 + 4y^2 - 4)(x + 1) \text{ となる.}$$

Example.2 二次曲線回帰

$\xrightarrow{\text{CoCoA}}$ Stable Order Ideal : $\mathcal{O} = \{1, x, y, y^2, xy, x^2, y^3, xy^2\}$

$$B = \left\{ \begin{array}{l} x^3 + 4.0039xy^2 + 0.00759y^3 + 0.001005x^2 - 0.000134xy \\ \quad + 4.021y^2 - 4.00228x - 0.00634y - 4.02, \\ \quad \vdots \end{array} \right\}$$

10^{-2} 以下を無視すると,

$$\begin{aligned} & x^3 + 4.0039xy^2 + 0.00759y^3 + 0.001005x^2 - 0.000134xy \\ & \quad + 4.021y^2 - 4.00228x - 0.00634y - 4.02 \end{aligned}$$

$$\approx x^3 + 4xy^2 + x^2 + 4y^2 - 4x - 4 = (x^2 + 4y^2 - 4)(x + 1) \text{ となる.}$$

最小二乗法では,

$$x^2 - 0.039xy + 3.984y^2 + 0.011x + 0.034y - 4.015$$

と推定できる.

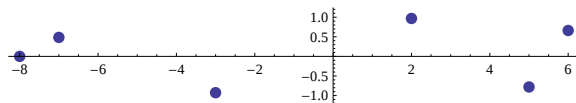
Example.3 楕円 or 直線

$$\tilde{\mathbb{X}} = \{(-8, 0), (-7, 0.484), (-3, -0.927), (2, 0.968), \\ (5, -0.781), (-6, 0.661)\}$$

最小二乗法では,

一次 : $y = 0.00946x + 0.00754$

二次 : $x^2 + 0.00485xy + 64.00y^2 - 0.000645x + 0.00222y - 64.00 = 0$



Example.3 楕円 or 直線

Border Bases 法では,

$$\mathcal{B} = \left\{ \begin{array}{l} x^2 - 0.03494y^3 - 0.00429xy + 64.013y^2 - 0.000557x \\ \quad + 0.0487y - 64.0045, \\ x^2y + 64.024y^3 - 0.001261xy - 0.0009956y^2 + 0.00215x \\ \quad - 64.01y + 0.01721, \\ \vdots \end{array} \right\}$$
$$\approx \left\{ \begin{array}{l} x^2 + 64.01y^2 - 64.00 \\ y(x^2 + 64.02y^2 - 64.01) \\ \vdots \end{array} \right\}$$

となる.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.
- $\mathbb{R}[x_1, \dots, x_n]$ 上でも近似可能 (理論上) .

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.
- $\mathbb{R}[x_1, \dots, x_n]$ 上でも近似可能 (理論上).
- 得られるであろう代数曲線が, 既約で無い多項式の因数として現れる事がある.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.
- $\mathbb{R}[x_1, \dots, x_n]$ 上でも近似可能 (理論上).
- 得られるであろう代数曲線が, 既約で無い多項式の因数として現れる事がある.
- 検定統計量が不明.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.
- $\mathbb{R}[x_1, \dots, x_n]$ 上でも近似可能 (理論上) .
- 得られるであろう代数曲線が, 既約で無い多項式の因数として現れる事がある.
- 検定統計量が不明.
- 「データが少ない, サンプルにばらつきがない, 異常なデータの存在」等により, 推定が困難になる.

考察・課題

- データの量が十分多ければ, 任意の次数の代数曲線で近似できる.
- 何次の代数曲線で近似する事が適当であるか調べる事が出来る.
- $\mathbb{R}[x_1, \dots, x_n]$ 上でも近似可能 (理論上).
- 得られるであろう代数曲線が, 既約で無い多項式の因数として現れる事がある.
- 検定統計量が不明.
- 「データが少ない, サンプルにばらつきがない, 異常なデータの存在」等により, 推定が困難になる.

課題は多いが, Border Bases 法は独自の強みを持った,
代数曲線回帰の新しいアプローチだと言える.