Can clouds of points generate numerical storms?

Claudia Fassino
(joint work with Maria-Laura Torrente)

Dipartimento di Matematica
Università di Genova
www.dima.unige.it/∼fassino/

Genova, 16th-17th February 2012

Introduction

We present a **symbolic-numeric approach** to find multivariate polynomial models describing a set of noisy data, represented as a set $\mathbb{X}$ of limited precision points.

We present a **symbolic-numeric approach** to find multivariate polynomial models describing a set of noisy data, represented as a set $\mathbb{X}$ of limited precision points.

Observations:

- Data without noise: the polynomial model can be described by the elements of the vanishing ideal $\mathcal{I}(\mathbb{X})$ of $\mathbb{X}$

$$\mathcal{I}(\mathbb{X}) = \{f \mid f(p) = 0, \ \forall p \in \mathbb{X}\}$$

Introduction

We present a **symbolic-numeric approach** to find multivariate polynomial models describing a set of noisy data, represented as a set $\mathbb{X}$ of limited precision points.

Observations:

- Data without noise: the polynomial model can be described by the elements of the vanishing ideal $\mathcal{I}(\mathbb{X})$ of $\mathbb{X}$

$$\mathcal{I}(\mathbb{X}) = \{f \mid f(p) = 0, \ \forall p \in \mathbb{X}\}$$

- Data with noise: Each point of $\mathbb{X}$ represents a cloud of points (numerically equivalent to $\mathbb{X}$). Structures based on polynomials involving real data lose many of their rigorous algebraic properties.

Noisy Data

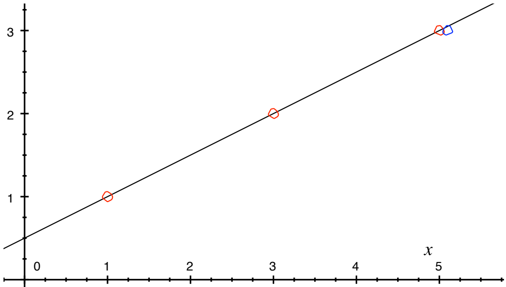
If the points of $\widehat{\mathbb{X}}$ are a "small" perturbation of the points of $\mathbb{X}$ their vanishing ideals $\mathcal{I}(\mathbb{X})$ and $\mathcal{I}(\widehat{\mathbb{X}})$ may be **very different**.

Noisy Data

If the points of $\widehat{\mathbb{X}}$ are a "small" perturbation of the points of $\mathbb{X}$ their vanishing ideals $\mathcal{I}(\mathbb{X})$ and $\mathcal{I}(\widehat{\mathbb{X}})$ may be **very different**.

Example Let $\mathbb{X} = \{(1,1),(3,2),(5,3)\}$ and $\widehat{\mathbb{X}} = \{(1,1),(3,2),(5.1,3)\}$

$$\mathcal{I}(\mathbb{X}) : \begin{cases} x - 2y + 1 \\ y^3 - 6y^2 + 11y - 6 \end{cases} \qquad \mathcal{I}(\widehat{\mathbb{X}}) : \begin{cases} x^2 - 20x + 37y - 18 \\ xy - 43x + 81y - 39 \\ y^2 - 90.1x + 172.2y - 83.1 \end{cases}$$



Data without noise

Given a set $\mathbb{X}$ of points we can compute the τ-Gröbner basis of $\mathcal{I}(\mathbb{X})$

The **Buchberger-Möller algorithm**
Input: A set of points $\mathbb{X}$
Output: A Gröbner basis of $\mathcal{I}(\mathbb{X})$

- **Step Zero:** $\mathcal{O} = \{1\}$
- **Generic Step:** $\mathcal{O} = \{t_1, \ldots, t_k\}$ and $t >_\tau t_i$.
  $t(\mathbb{X}), t_1(\mathbb{X}), \ldots, t_k(\mathbb{X})$ linearly dependent? (*)
  Yes: polynomial of basis formed
  No: t is added to $\mathcal{O}$

Data without noise

Given a set $\mathbb{X}$ of points we can compute the τ-Gröbner basis of $\mathcal{I}(\mathbb{X})$

The **Buchberger-Möller algorithm**
Input: A set of points $\mathbb{X}$
Output: A Gröbner basis of $\mathcal{I}(\mathbb{X})$

- **Step Zero:** $\mathcal{O} = \{1\}$
- **Generic Step:** $\mathcal{O} = \{t_1, \ldots, t_k\}$ and $t >_\tau t_i$.
  $t(\mathbb{X}), t_1(\mathbb{X}), \ldots, t_k(\mathbb{X})$ linearly dependent? (*)
  Yes: polynomial of basis formed
  No: t is added to $\mathcal{O}$

(*) $\Leftrightarrow$ - construct $M_{\mathcal{O}}(\mathbb{X}) = (t_1(\mathbb{X}), \ldots, t_k(\mathbb{X}))$
- solve LSP: $M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X}) = t(\mathbb{X})$
- check whether $\rho(\mathbb{X}) = t(\mathbb{X}) - M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X}) = 0$

Example

Example Let $\mathbb{X} = \{(1,1),(3,2),(5,3)\}$ and $\widehat{\mathbb{X}} = \{(1,1),(3,2),(5.1,3)\}$

Consider $\mathcal{O} = \{t_1 = 1, t_2 = y\}$ $t = x$

$$[t_1(\mathbb{X}), t_2(\mathbb{X}), t(\mathbb{X})] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{bmatrix} \qquad [t_1(\widehat{\mathbb{X}}), t_2(\widehat{\mathbb{X}}), t(\widehat{\mathbb{X}})] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5.1 \end{bmatrix}$$

$$t(\mathbb{X}) - 2t_2(\mathbb{X}) + t_1(\mathbb{X}) = 0$$

Independent vectors

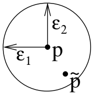
Empirical points

Noisy data set $\Longleftrightarrow$ Empirical points: $\mathbb{X} \subset \mathbb{R}^n$ and a **tolerance** $\varepsilon \in \mathbb{R}^+$

Empirical points

Noisy data set $\Longleftrightarrow$ Empirical points: $\mathbb{X} \subset \mathbb{R}^n$ and a **tolerance** $\varepsilon \in \mathbb{R}^+$

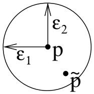
$\widetilde{\mathbb{X}}$ is numerically equivalent to $\mathbb{X}$ iff
$\widetilde{\mathbb{X}}$ consists of points belonging to the "clouds" around the elements of $\mathbb{X}$



Empirical points

Noisy data set $\Longleftrightarrow$ Empirical points: $\mathbb{X} \subset \mathbb{R}^n$ and a **tolerance** $\varepsilon \in \mathbb{R}^+$

$\widetilde{\mathbb{X}}$ is numerically equivalent to $\mathbb{X}$ iff
$\widetilde{\mathbb{X}}$ consists of points belonging to the "clouds" around the elements of $\mathbb{X}$



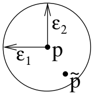
A generic $\widetilde{p}$ can be expressed as $\widetilde{p} = (p_1 + e_1, \ldots, p_n + e_n) \Longrightarrow$
we make use of error variables

$$\mathbf{e} = (e_{11}, \ldots, e_{s1}, e_{12}, \ldots, e_{s2}, \ldots, e_{1n}, \ldots, e_{sn}) \text{ and } \|\mathbf{e}\|_\infty \leq \varepsilon$$

Empirical points

Noisy data set $\Longleftrightarrow$ Empirical points: $\mathbb{X} \subset \mathbb{R}^n$ and a **tolerance** $\varepsilon \in \mathbb{R}^+$

$\widetilde{\mathbb{X}}$ is numerically equivalent to $\mathbb{X}$ iff
$\widetilde{\mathbb{X}}$ consists of points belonging to the "clouds" around the elements of $\mathbb{X}$



A generic $\widetilde{p}$ can be expressed as $\widetilde{p} = (p_1 + e_1, \ldots, p_n + e_n) \Longrightarrow$ we make use of error variables

$$\mathbf{e} = (e_{11}, \ldots, e_{s1}, e_{12}, \ldots, e_{s2}, \ldots, e_{1n}, \ldots, e_{sn}) \text{ and } \|\mathbf{e}\|_\infty \leq \varepsilon$$

A generic admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}$ is expressed as

$$\widetilde{\mathbb{X}} = \mathbb{X}(\mathbf{e}) = \{p_1(\mathbf{e}), \ldots, p_s(\mathbf{e})\}$$

Problem

Given a set $\mathbb{X}$ of empirical points
determine a polynomial of low degree and the corresponding (simple) variety at which the points almost lie

that is

Problem

Given a set $\mathbb{X}$ of empirical points
determine a polynomial of low degree and the corresponding (simple) variety at which the points almost lie

that is

Given $\mathbb{X} = \{p_1, \ldots, p_s\} \subseteq \mathbb{R}^n$ set of points, $\varepsilon \in \mathbb{R}^+$ tolerance
compute

$$\begin{array}{l} f^* \in \mathbb{R}[x_1, \ldots, x_n] \\ \mathbb{X}^* = \{p_1^*, \ldots, p_s^*\} \subseteq \mathbb{R}^n \end{array} \quad \text{s.t.} \quad \left\{ \begin{array}{l} f^*(p_i^*) = 0 \\ \|p_i^* - p_i\|_\infty < \varepsilon \end{array} \right. \quad \forall p_i^* \in \mathbb{X}^*$$

The simplicity of the model judged by the degree of f^*
The goodness of model by the max distance of $\mathbb{X}^*$ from $\mathbb{X}$

The New algorithm (empirical case)
Input: set of empirical points $\mathbb{X}$ and tolerance ε
Out: low degree polynomial f vanishing at $\mathbb{X}(\widehat{\mathbf{e}})$, $\|\widehat{\mathbf{e}}\| < \varepsilon$

- **Generic Step:** $\mathcal{O} = \{t_1, \ldots, t_k\}$ and $t >_\tau t_i$.
  check whether there exists $\widehat{\mathbf{e}}$ s.t. $\mathbb{X}(\widehat{\mathbf{e}})$ is adm.pert. of $\mathbb{X}$ and

$$\rho(\mathbb{X}(\widehat{\mathbf{e}})) = t(\mathbb{X}(\widehat{\mathbf{e}})) - M_{\mathcal{O}}(\mathbb{X}(\widehat{\mathbf{e}}))\alpha(\mathbb{X}(\widehat{\mathbf{e}})) = 0 \qquad (*)$$

(Exact Case $\rho = t(\mathbb{X}) - M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X}) = 0$)

Yes: low degree polynomial is formed and stop
No: t is added to $\mathcal{O}$

Adapting the BM Algorithm to the empirical case

The New algorithm (empirical case)
Input: set of empirical points $\mathbb{X}$ and tolerance ε
Out: low degree polynomial f vanishing at $\mathbb{X}(\widehat{\mathbf{e}})$, $\|\widehat{\mathbf{e}}\| < \varepsilon$

- **Generic Step:** $\mathcal{O} = \{t_1, \ldots, t_k\}$ and $t >_\tau t_i$.
  check whether there exists $\widehat{\mathbf{e}}$ s.t. $\mathbb{X}(\widehat{\mathbf{e}})$ is adm.pert. of $\mathbb{X}$ and

$$\rho(\mathbb{X}(\widehat{\mathbf{e}})) = t(\mathbb{X}(\widehat{\mathbf{e}})) - M_{\mathcal{O}}(\mathbb{X}(\widehat{\mathbf{e}}))\alpha(\mathbb{X}(\widehat{\mathbf{e}})) = 0 \qquad (*)$$

(Exact Case $\rho = t(\mathbb{X}) - M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X}) = 0$)

Yes: low degree polynomial is formed and stop
No: t is added to $\mathcal{O}$

(*) $\rho(\mathbf{e}) = 0 \quad \textbf{in} \quad D = \{\mathbf{e} : \|\mathbf{e}\|_\infty < \varepsilon, M_{\mathcal{O}}(\mathbf{e}) \textbf{ is full rank}\}$

Underdetermined nonlinear systems

Problem: **compute a solution of $\rho(\mathbf{e}) = 0$ s.t. $\mathbf{e} \in D$**

[1]Walker, Watson (1990)

Problem: **compute a solution of $\rho(\mathbf{e}) = 0$ s.t. $\mathbf{e} \in D$**

- **Normal Flow Algorithm.**[1] Let $\bar{x} = 0 \in D$, $\omega \ll 1$, h=(1,. . . ,1)
  - While $\|h\|_2 > \omega$ do

    $$h = -\mathrm{Jac}_\rho(\bar{x})^\dagger \rho(\bar{x})$$
    $$\bar{x} = \bar{x} + h$$

  - Return $\bar{x}$ and stop

[1] Walker, Watson (1990)

Problem: **compute a solution of $\rho(\mathbf{e}) = 0$ s.t. $\mathbf{e} \in D$**

- **Normal Flow Algorithm.**[1] Let $\bar{x} = 0 \in D$, $\omega \ll 1$, h=(1,. . . ,1)
  - While $\|h\|_2 > \omega$ do
    $$h = -\mathrm{Jac}_\rho(\bar{x})^\dagger \rho(\bar{x})$$
    $$\bar{x} = \bar{x} + h$$
  - Return $\bar{x}$ and stop

The NF Algorithm depends on the **conditioning** of $\mathrm{Jac}_\rho(\bar{x})$
SO we adopt the following strategy:

- If $\mathrm{Jac}_\rho(x)$ well-conditioned $\Rightarrow$ we apply the NF Algorithm to ρ
- If $\mathrm{Jac}_\rho(x)$ ill-conditioned $\Rightarrow$ we construct $\widehat{\rho} : D \to \mathbb{R}^m$ such that $\mathrm{Jac}_{\widehat{\rho}}(x)$ **well-conditioned** in D and $\|\widehat{\rho}(x) - \rho(x)\|$ **small** in D

[1] Walker, Watson (1990)

Our strategy is based on

- **Numerical Rank**. Let $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, $\delta > 0$ and $k > 1$. If

$$\sigma_1(A) \geq \ldots \geq \sigma_r(A) > k\delta > \delta > \sigma_{r+1}(A) \geq \ldots \geq \sigma_n(A)$$

then r is called the numerical (δ, k)-rank of A.

[2] Hong, Pan (1992)

Our strategy is based on

- **Numerical Rank**. Let $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, $\delta > 0$ and $k > 1$. If

$$\sigma_1(A) \geq \ldots \geq \sigma_r(A) > k\delta > \delta > \sigma_{r+1}(A) \geq \ldots \geq \sigma_n(A)$$

  then r is called the numerical (δ, k)-rank of A.

- **Rank Revealing Decomposition** [2]. Let $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and r=numerical (δ, k)-rank of $A \Rightarrow \exists$ a permutation matrix Π s.t.

$$A\Pi = Q \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

$$\sigma_{\min}(R_{11}) \geq \frac{k}{q(n,r)}\delta \quad \text{and} \quad \|R_{22}\|_2 \leq q(n,r)\delta$$

The first r columns of $A\Pi$ are strongly independent

[2] Hong, Pan (1992)

Back to the problem

We consider, at the first step of the Normal Flow Algorithm, the ill conditioned system

$$\boxed{\mathbf{Jac}_\rho(0)h = -\rho(0)}$$

We select the maximum number of "strongly independent equations"

- computing $r = \text{rank}_{\delta,k}(Jac_\rho(0) \mid \rho(0))$, with $\delta \geq \varepsilon$;
- computing the Rank Revealing Decomposition of

$$\begin{pmatrix} \text{Jac}_\rho(0)^t \\ \rho(0)^t \end{pmatrix} \Pi = Q \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

- partitioning $\Pi = (\Pi_1 \mid \Pi_2)$ with $\Pi_1 \in \mathrm{Mat}_{m \times r}(\mathbb{R})$.

Back to the problem

We consider, at the first step of the Normal Flow Algorithm, the ill conditioned system

$$\boxed{\mathbf{Jac}_{\rho}(0)h = -\rho(0)}$$

We select the maximum number of "strongly independent equations"

- computing $r = \mathrm{rank}_{\delta,k}(Jac_{\rho}(0) \mid \rho(0))$, with $\delta \geq \varepsilon$;
- computing the Rank Revealing Decomposition of

$$\begin{pmatrix} \mathrm{Jac}_{\rho}(0)^t \\ \rho(0)^t \end{pmatrix} \Pi = Q \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

- partitioning $\Pi = (\Pi_1 \mid \Pi_2)$ with $\Pi_1 \in \mathrm{Mat}_{m\times r}(\mathbb{R})$.

$$\widehat{\rho}(\mathbf{e}) = \begin{pmatrix} \Pi_1^t \rho(\mathbf{e}) \\ G\Pi_1^t \rho(\mathbf{e}) \end{pmatrix} \quad \text{where} \quad G = R_{12}^t R_{11}^{-t}$$

Main results

$$\widehat{\rho}(\mathbf{e}) = \begin{cases} \Pi_1^t \rho(\mathbf{e}) & = & 0 \quad \} \, r \\ G\Pi_1^t \rho(\mathbf{e}) & = & 0 \quad \} \, m-r \end{cases}$$

- the $m - r$ last equations depend on the first r equations;
- the first r rows of $\mathrm{Jac}_{\widehat{\rho}}(0)$ (or of $\mathrm{Jac}_{\widehat{\rho}}(\mathbf{e})$ if $\mathbf{e} \approx \mathbf{0}$) is well conditioned;
- $\|\widehat{\rho}(\mathbf{e}) - \Pi^t \rho(\mathbf{e})\|_2 \leq q(m,r)\delta + O(\delta^2) \;\; \forall\, \mathbf{e} \in D$
  If $\rho(\mathbf{e}^*) = 0$, $\mathbf{e}^* \in D$, then $\|\widehat{\rho}(\mathbf{e}) - \Pi^t \rho(\mathbf{e})\|_2 = O(\delta^2)$

Conclusion: we apply the NF algorithm to $\Pi_1^t \rho(\mathbf{e})$ which is

- equivalent to $\widehat{\rho}(\mathbf{e})$
- "close" to $\rho(\mathbf{e})$
- with well-conditioned Jacobian matrices on D.

Example

We consider the set of points created by perturbing less than 0.1 the coords of 6 points on $y^2 - x - 2y + 2 = 0$:

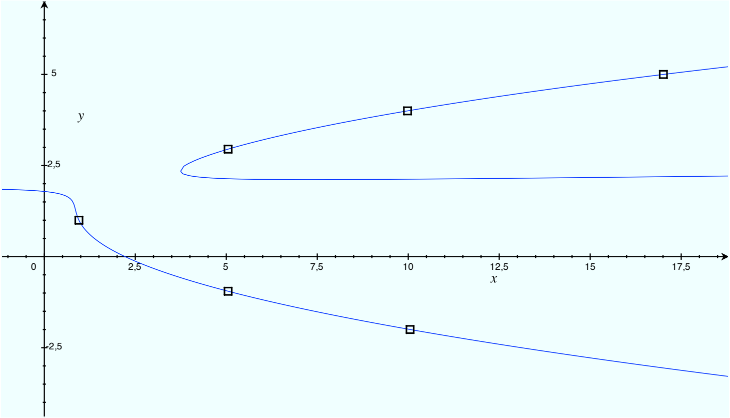
$\mathbb{X} = \{(0.95, 1), (5.05, 2.95), (5.05, -0.95), (9.98, 4), (10.05, -2), (17.01, 5)\}$

Example

We consider the set of points created by perturbing less than 0.1 the coords of 6 points on $y^2 - x - 2y + 2 = 0$:

$\mathbb{X} = \{(0.95, 1), (5.05, 2.95), (5.05, -0.95), (9.98, 4), (10.05, -2), (17.01, 5)\}$

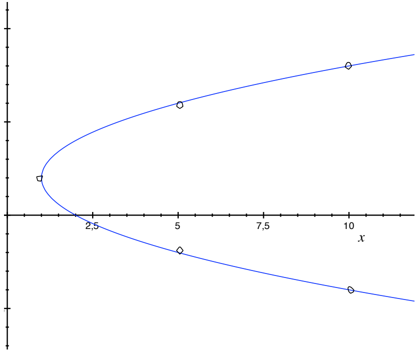
A straightforward computation using BM algorithm gives



Example

Our Algorithm applied to $\mathbb{X}$ with $\varepsilon = 0.1$, $\delta = 2\varepsilon$, $k = 2$ computes

$$\bar{f} = y^2 - 0.9816x - 2.0176y + 1.9660 \qquad \left(\approx\ y^2 - x - 2y + 2\right)$$

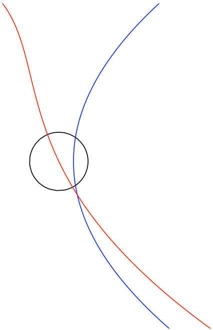


Singular values and condition number of $[\mathrm{Jac}_\rho(0) \mid \rho(0)]$

$$\{6.39, 4.67, 3.33,\ \ 0.035, 0.02, 0.00\} \quad K_2\left(\mathrm{Jac}_\rho(0) \mid \rho(0)\right) \approx 310$$

Example

Zooming around the first point we get...



Open problems

Problem 1: Parameters

Empirical points $\Longleftrightarrow$ tolerance ε on the coordinates of the input points.
Numerical rank of $[\mathrm{Jac}_\rho(0)|\rho(0)]$ $\Longleftrightarrow$ parameters δ and k s.t.

$$\sigma_1 \geq \dots \sigma_r > k\delta > \delta \geq \sigma_{r+1} \geq \sigma_m$$

What is the best choice of δ and k as function of ε?

Problem 2: Strategy

A different strategy for solving the non linear system $\rho(\mathbf{e}) = 0$ avoiding the "almost dependent" equations (using the information about the tolerance...)