Recent Improvements of the AVI Algorithm

Martin Kreuzer

Fakultät für Informatik und Mathematik

Universität Passau

martin.kreuzer@uni-passau.de

Algebraic Oil Project Workshop

Rijswijk

November 26, 2008

Contents

1 – Approximate Border Bases

It is difficult to get a man to understand something when his salary depends upon his not understanding it.

(Upton Sinclair)

$P = \mathbb{R}[x_1, \ldots, x_n]$ polynomial ring over the reals

$I \subseteq P$ zero-dimensional polynomial ideal (i.e. $\dim_{\mathbb{R}}(P/I) < \infty$)

$\mathbb{T}^n = \{x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mid \alpha_i \geq 0\}$ monoid of terms

Definition of Border Bases

Definition 1.1 **(a)** A (finite) set $\mathcal{O} \subset \mathbb{T}^n$ is called an **order ideal** if $t \in \mathcal{O}$ and $t' \mid t$ implies $t' \in \mathcal{O}$.

(b) Let $\mathcal{O}$ be an order ideal. The set $\partial\mathcal{O} = (x_1\mathcal{O} \cup \cdots \cup x_n\mathcal{O}) \setminus \mathcal{O}$ is called the **border** of $\mathcal{O}$.

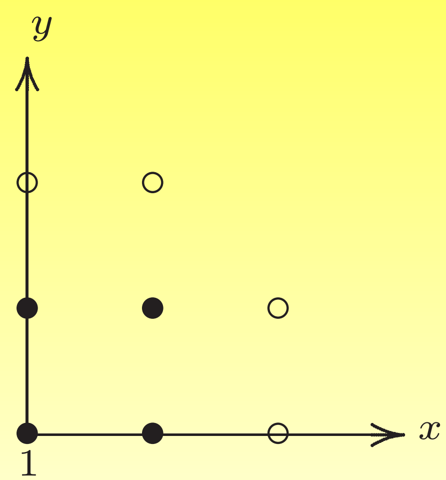
(c) Let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be an order ideal and $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ its border. A set of polynomials $\{g_1, \ldots, g_\nu\} \subset I$ of the form

$$g_j = b_j - \sum_{i=1}^{\mu} c_{ij} t_i$$

with $c_{ij} \in \mathbb{R}$ and $t_i \in \mathcal{O}$ is called an **$\mathcal{O}$-border prebasis** of I.

(d) An $\mathcal{O}$-border prebasis of I is called an **$\mathcal{O}$-border basis** of I if the residue classes of the terms in $\mathcal{O}$ are a $\mathbb{R}$-vector space basis of P/I.

A Picture of an Order Ideal and its Border



Neighbors

Under capitalism, man exploits man.
Under communism, it's just the opposite.
(John Kenneth Galbraith)

Definition 1.2 Let $b_i, b_j \in \partial\mathcal{O}$ be two distinct border terms.

(a) The border terms b_i and b_j are called **next-door neighbors** if $b_i = x_k\, b_j$ for some $k \in \{1, \ldots, n\}$.

(b) The border terms b_i and b_j are called **across-the-street neighbors** if $x_k\, b_i = x_\ell\, b_j$ for some $k, \ell \in \{1, \ldots, n\}$.

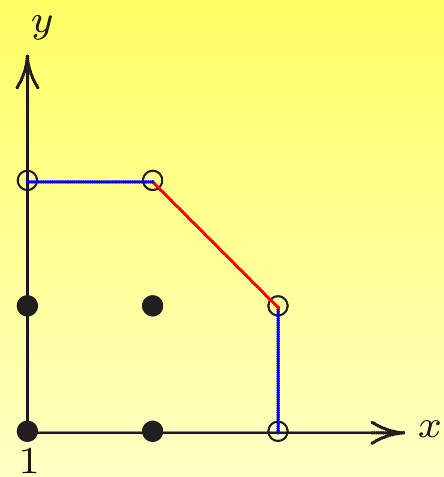
(c) The border terms b_i and b_j are called **neighbors** if they are next-door neighbors or across-the-street neighbors.

(d) The graph whose vertices are the border terms and whose edges are given by the neighbor relation is called the **border web** of $\mathcal{O}$.

Example 1.3 The border of $\mathcal{O} = \{1, x, y, xy\}$ is $\partial\mathcal{O} = \{x^2, x^2y, xy^2, y^2\}$. Here the border web looks as follows:

(x^2, x^2y) and (y^2, xy^2) are next-door neighbor pairs

(x^2y, xy^2) is an across-the-street neighbor pair.



Proposition 1.4 *The border web is connected.*

Neighbor Syzygies

Definition 1.5 **(a)** For $t, t' \in \mathbb{T}^n$, we call the pair $(\mathrm{lcm}(t,t')/t, -\mathrm{lcm}(t,t')/t')$ the **fundamental syzygy** of (t,t').

(b) The fundamental syzygies of neighboring border terms are also called the **neighbor syzygies**.

Definition 1.6 Let $g_i, g_j \in G$ be two distinct border prebasis polynomials. Then the polynomial

$$S_{ij} = (\mathrm{lcm}(b_i, b_j)/b_i) \cdot g_i - (\mathrm{lcm}(b_i, b_j)/b_j) \cdot g_j$$

is called the **S-polynomial** of g_i and g_j.

Remark 1.7 Let $g_i, g_j \in G$.

(a) If (b_i, b_j) are next-door neighbors with $b_j = x_k\, b_i$ then $\mathrm{NR}_G(S_{ij})$ is of the form $\mathrm{NR}_G(S_{ij}) = g_j - x_k\, g_i - \sum_{m=1}^{\mu} a_m e_m$ with $a_m \in \mathbb{R}$.

(b) If (b_i, b_j) are across-the-street neighbors with $x_k\, b_i = x_\ell\, b_j$ then $\mathrm{NR}_G(S_{ij})$ is of the form $\mathrm{NR}_G(S_{ij}) = x_k\, g_i - x_\ell\, g_j - \sum_{m=1}^{\mu} a_m e_m$ with $a_m \in \mathbb{R}$.

(3) If $\mathrm{NR}_{\mathcal{O},G}(S_{ij}) = 0$, we shall say that the syzygy $e_j - x_k\, e_i - \sum_{m=1}^{\mu} a_m e_m$ resp. $x_k\; e_i - x_\ell\, e_j - \sum_{m=1}^{\mu} a_m e_m$ is a **lifting** of the neighbor syzygy $e_j - x_k\, e_i$ resp. $x_k\; e_i - x_\ell\, e_j$.

Theorem 1.8 *(Stetter)*
An $\mathcal{O}$-border prebasis G is an $\mathcal{O}$-border basis if and only if the neighbor syzygies lift.

Definition of Approximate Border Bases

Definition 1.9 Let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be an order ideal, let $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$, and let $\varepsilon > 0$.

A set of polynomials $G = \{g_1, \ldots, g_\nu\}$ is called an **ε-approximate $\mathcal{O}$-border basis** if the following conditions are satisfied:

1. For $j = 1, \ldots, \nu$, we have $\|g_j\| = 1$.
2. If a_j denotes the coefficient of b_j in g_j then $|a_j| > \varepsilon$ and $\{\frac{1}{a_j}\, g_1, \ldots, \frac{1}{a_\nu}\, g_\nu\}$ is an $\mathcal{O}$-border prebasis.
3. For all pairs (i,j) such that (b_i, b_j) are neighbors, we have $\|\operatorname{NR}_{\mathcal{O},G}(S_{ij})\| < \varepsilon$.

Remark 1.10 If $G = \{g_1, \dots, g_\nu\}$ is an ε-approximate border basis then the point $(c_{11}, \dots, c_{\mu\nu})$ in $\mathbb{R}^{\mu\nu}$ given by its coefficients is **close to** the **border basis scheme**.

Example 1.11 Let $\mathcal{O} = \{1,\, x,\, y,\, xy\}$. Then the set

$$g_1 = x^2 + 0.02xy - 0.01y - 1.01 \qquad g_2 = x^2y + 0.03x - 0.98\,y$$

$$g_3 = xy^2 - 1.02\,x \qquad g_4 = y^2 - 0.99$$

is an approximate $\mathcal{O}$-border basis. The ideal $I = \langle g_1, g_2, g_3, g_4 \rangle$ is the unit ideal, since $g_3 - x\,g_4 = 0.03x$ shows $-g_1 \equiv 0.01y + 1.01$ and $g_4 \equiv 101^2 - 0.99 \pmod{I}$.

2 – The AVI Algorithm

A good algorithm is
10% inspiration, 15% perspiration,
and 75% desperation.
(Anonymous)

Goal: Given a set of (approximate) points $\mathbb{X} = \{p_1, \ldots, p_s\}$ in $\mathbb{R}^n$ and $\varepsilon > 0$, find an order ideal $\mathcal{O}$ and an approximate $\mathcal{O}$-border basis G such that the polynomials in G vanish ε-approximately at the points of $\mathbb{X}$.

Notice that, in general,

- we have $\#\mathcal{O} << \#\mathbb{X}$,
- the ideal $\langle G \rangle$ is the unit ideal.

Theorem 2.1 (The Singular Value Decomposition)
Let $\mathcal{A} \in \mathrm{Mat}_{m,n}(\mathbb{R})$.

1. *There are orthogonal matrices* $\mathcal{U} \in \mathrm{Mat}_{m,m}(\mathbb{R})$ *and* $\mathcal{V} \in \mathrm{Mat}_{n,n}(\mathbb{R})$ *and a matrix* $\mathcal{S} \in \mathrm{Mat}_{m,n}(\mathbb{R})$ *of the form* $\mathcal{S} = \begin{pmatrix} \mathcal{D} & 0 \\ 0 & 0 \end{pmatrix}$ *such that*

$$\mathcal{A} = \mathcal{U} \cdot \mathcal{S} \cdot \mathcal{V}^{\mathrm{tr}} = \mathcal{U} \cdot \begin{pmatrix} \mathcal{D} & 0 \\ 0 & 0 \end{pmatrix} \cdot \mathcal{V}^{\mathrm{tr}}$$

   where $\mathcal{D} = \mathrm{diag}(s_1, \ldots, s_r)$ *is a diagonal matrix.*

2. *In this decomposition, it is possible to achieve* $s_1 \geq s_2 \geq \cdots \geq s_r > 0$. *The numbers* $s_1, \ldots, s_r$ *depend only on* $\mathcal{A}$ *and are called the* **singular values** *of* $\mathcal{A}$.

3. *The number r is the rank of $\mathcal{A}$.*

4. *The matrices $\mathcal{U}$ and $\mathcal{V}$ have the following interpretation:*

$$\begin{aligned}
\text{first } r \text{ columns of } \mathcal{U} &\equiv \text{ONB of the column space of } \mathcal{A} \\
\text{last } m-r \text{ columns of } \mathcal{U} &\equiv \text{ONB of the kernel of } \mathcal{A}^{\text{tr}} \\
\text{first } r \text{ columns of } \mathcal{V} &\equiv \text{ONB of the row space of } \mathcal{A} \\
&\equiv \text{ONB of the column space of } \mathcal{A}^{\text{tr}} \\
\text{last } n-r \text{ columns of } \mathcal{V} &\equiv \text{ONB of the kernel of } \mathcal{A}
\end{aligned}$$

Corollary 2.2 *Let $\mathcal{A} \in \mathrm{Mat}_{m,n}(\mathbb{R})$, and let $\varepsilon > 0$ be given. Let $k \in \{1, \ldots, r\}$ be chosen such that $s_k > \varepsilon \geq s_{k+1}$. Form the matrix $\widetilde{\mathcal{A}} = \mathcal{U}\,\widetilde{\mathcal{S}}\,\mathcal{V}^{\mathrm{tr}}$ by setting $s_{k+1} = \cdots = s_r = 0$ in $\mathcal{S}$. The matrix $\widetilde{\mathcal{A}}$ is called the* **singular value truncation** *of $\mathcal{A}$.*

1. *We have $\min\{\|\mathcal{A} - \mathcal{B}\| : \mathrm{rank}(\mathcal{B}) \leq k\} = \|\mathcal{A} - \widetilde{\mathcal{A}}\| = s_{k+1}$. (Here $\|\cdots\|$ denotes the 2-operator norm of a matrix.)*

2. *The vector subspace $\mathrm{apker}(\mathcal{A}, \varepsilon) = \ker(\widetilde{\mathcal{A}})$ is the largest dimensional kernel of a matrix whose Euclidean distance from $\mathcal{A}$ is at most ε. It is called the* **ε-approximate kernel** *of $\mathcal{A}$.*

3. *The last $n - k$ columns $v_{k+1}, \ldots, v_n$ of $\mathcal{V}$ are an ONB of $\mathrm{apker}(\mathcal{A}, \varepsilon)$. They satisfy $\|\mathcal{A}v_i\| < \varepsilon$.*

Theorem 2.3 **(The AVI Algorithm)**
The following algorithm computes an approximate border basis of an approximate vanishing ideal of a finite set of points $\mathbb{X} \subseteq [\mathbf{-1}, \mathbf{1}]^{\mathbf{n}}$.

A1 *Start with lists* $G = \emptyset$, $\mathcal{O} = [1]$, *a matrix* $\mathcal{M} = (1, \ldots, 1)^{\mathrm{tr}} \in \mathrm{Mat}_{s,1}(\mathbb{R})$, *and* $d = 0$.

A2 *Increase* d *by one and let* L *be the list of all terms of degree* d *in* $\partial\mathcal{O}$, *ordered decreasingly w.r.t.* σ. *If* $L = \emptyset$, *return the pair* $(G, \mathcal{O})$ *and stop. Otherwise, let* $L = (t_1, \ldots, t_\ell)$.

A3 *Let* m *be the number of columns of* $\mathcal{M}$. *Form the matrix*

$$\mathcal{A} = (\mathrm{eval}(t_1), \ldots, \mathrm{eval}(t_\ell), \mathcal{M}) \in \mathrm{Mat}_{s,\ell+m}(\mathbb{R}).$$

Using its **SVD**, *calculate a matrix* $\mathcal{B}$ *whose column vectors are an ONB of the* **approximate kernel** $\mathrm{apker}(\mathcal{A}, \varepsilon)$.

A4 Compute the **stabilized reduced row echelon form** of $\mathcal{B}^{\mathrm{tr}}$ with respect to the given τ. The result is a matrix $\mathcal{C} = (c_{ij}) \in \mathrm{Mat}_{k,\ell+m}(\mathbb{R})$ such that $c_{ij} = 0$ for $j < \nu(i)$. Here $\nu(i)$ denotes the column index of the pivot element in the i^{th} row of $\mathcal{C}$.

A5 For all $j \in \{1, \ldots, \ell\}$ such that there exists a $i \in \{1, \ldots, k\}$ with $\nu(i) = j$ (i.e. **for the column indices of the pivot elements**), append the polynomial

$$c_{ij} t_j + \sum_{j'=j+1}^{\ell} c_{ij'} t_{j'} + \sum_{j'=\ell+1}^{\ell+m} c_{ij'} u_{j'}$$

to the list G, where $u_{j'}$ is the $(j' - \ell)^{\mathrm{th}}$ element of $\mathcal{O}$.

A6 For all $j = \ell, \ell - 1, \ldots, 1$ such that the j^{th} column of $\mathcal{C}$ contains **no pivot element**, append the term t_j as a new first element to $\mathcal{O}$ and append the column $\mathrm{eval}(t_j)$ as a new first column to $\mathcal{M}$.

A7 Using the SVD of $\mathcal{M}$, calculate a matrix $\mathcal{B}$ whose column vectors are an ONB of $\operatorname{apker}(\mathcal{M}, \varepsilon)$.

A8 Repeat steps **A4** – **A7** until $\mathcal{B}$ is empty. Then continue with step **A2**.

This algorithm returns the following results:

(a) The set $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ contains an order ideal of terms which is **strongly linearly independent** on $\mathbb{X}$, i.e. such that there is no unitary polynomial in $\langle \mathcal{O} \rangle_K$ which vanishes ε-approximately on $\mathbb{X}$.

(b) The set G is a δ-approximate $\mathcal{O}$-border basis. (An explicit bound for δ can be given.)

3 – Update Strategies

In Step **A2** we updated the list of terms to be processed as follows:

A2 Increase d by one and let L be the list of all terms of degree d in $\partial\mathcal{O}$, ordered decreasingly w.r.t. σ.

This simple strategy makes it easy to analyze the algorithm (in the paper), but has disadvantages in a practical implementation:

(1) The list L may be very large. In that case the SVD computation requires a lot of resources.

(2) The Stable RREF computation is less stable if the approximate kernel becomes very large.

To counter these problems, the following strategies can be used:

First Variation

A2 Increase d by one. Let L be the list of all terms in $(\partial\mathcal{O})_d$ **which are not in** $\langle \mathrm{LT}_\sigma(G) \rangle$, ordered decreasingly w.r.t. σ.

Advantages:

(1) Smaller lists L yield faster SVD computations.

(2) If a candidate term is a multiple of a known border term, we know already how to rewrite it.

Disadvantages:

(1) This rewriting (i.e. the corresponding border basis polynomial) may not satisfy the quality requirements given by the threshold number.

(2) We have to complete the computed border basis at the end.

Further Variations

(1) Do several iterations in degree d, for instance depending on the number of terms to be processed. Each time, take only a part of the terms of degree d and put them into L.

(2) Using one of the fast approaches, find a suitable order ideal $\mathcal{O}$ first. Then recompute the approximate border basis by projecting the evaluation vectors of the border terms to the space spanned by the evaluation vectors of the terms in $\mathcal{O}$.

(3) Same as (2), but this time project only to evaluation vectors of terms smaller w.r.t. σ than the given border term (to make sure this is the leading term).

4 – Final Cleaning

Question to Radio Eriwan:

Is it true that there are polynomials in the Approximate Vanishing Ideal which do not vanish approximately?

Radio Eriwan answers:

Yes, approximately.

The *approximate kernel* we compute using the SVD is a vector space, but the "true" set of "unitary" relations among the given evaluation vectors is a neighborhood of the vector space, i.e. a topological object.

Consequently, after we perform the stable RREF, we may sometimes find relations among the evaluation vectors of the terms in $\mathcal{O}$ which are not smaller than the threshold.

Methods for Final Cleaning

(1) After each degree, compute an additional SVD of eval($\mathcal{O}$) and remove redundant elements of $\mathcal{O}$. **(expensive)**

(2) At the very end, compute the SVD of eval($\mathcal{O}$) and remove all redundant elements of $\mathcal{O}$. **(quick, not completely stable)**

(3) At the very end, compute the SVD of eval($\mathcal{O}$) and remove redundant elements of $\mathcal{O}$ one by one. **(good compromise)**

In this case, updating the approximate border basis is particularly easy.

5 – Quality Control

Don't trust the computer any farther than you can throw it.
(Approximate English Proverb)

To check the quality of the output of the AVI algorithm, several methods have been implemented.

Smooth Evaluations

Compute the evaluation vectors of the polynomials in the approximate border basis and check that they are smaller than a given threshold w.r.t. a suitable norm.

Approximate Border Basis Properties

Check that G is an approximate border basis: first check the border prebasis property and the size of the border coefficients, then compute the normal remainders of the S-polynomials of the neighbor pairs and check their size.

Predictive Power of the Results

Use only part of the data for the modelling experiment. Use the remaining part to verify the predictions derived from the models.

6 – Automatic Threshold Determination

Automatic simply means
that you can't repair it yourself.
(Anonymous)

The quality of the predictions derived from various chosen threshold numbers ε can be used to find the **"correct"** ε which fits the data. We can use the following predictive powers:

(1) approximate vanishing of the approximate border basis polynomials at further data points;

(2) approximate correctness of the predicted values (for instance, of oil and gas productions) given by model polynomials derived from the order ideal.

7 – Future Developments

Prediction is very difficult,
especially if it's about the future.
(Niels Bohr)

(1) Implement and compare different update strategies in the AVI algorithm with respect to overall efficiency, quality of the computed approximate border basis, and predictive and interpretive power of the results.

(2) Apply the algorithm to the setting of production-exploration.

(3) Use the output of AVI and **rational recovery** as input for further **symbolic** computations to find structural information hidden in the input data.