Algebraic Oil

Martin Kreuzer

Fakultät für Informatik und Mathematik

Universität Passau

martin.kreuzer@uni-passau.de

NOCAS

Passau

December 10, 2008

Contents

1 – Some Problems in Oil Production

Impossible only means
that you haven't found the solution yet.
(Anonymous)

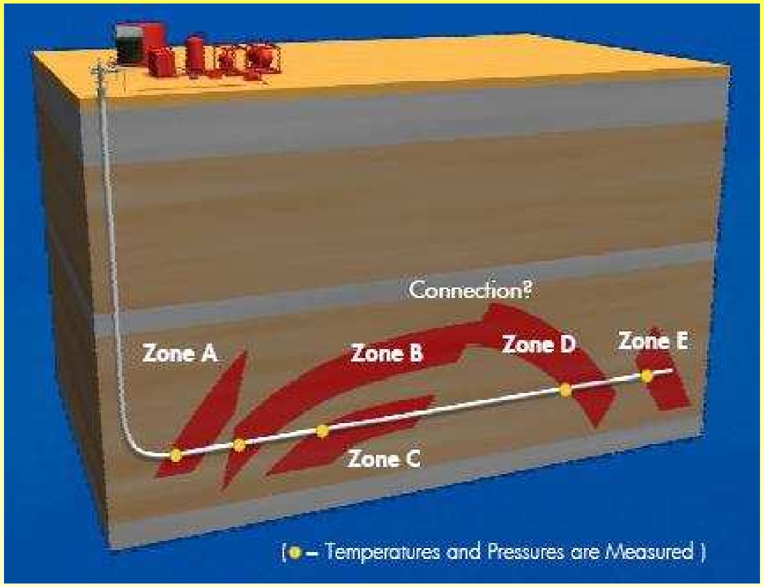
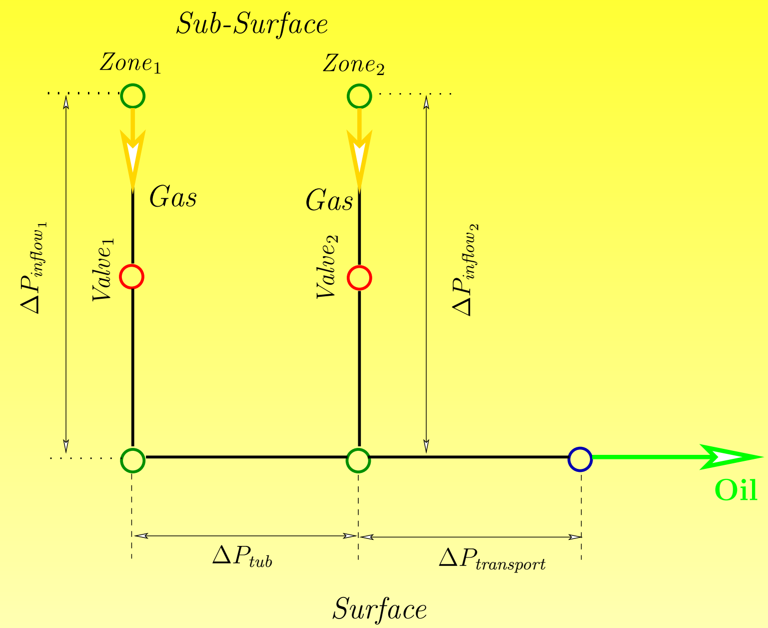


Figure 1: Overview of an Oil Production System

Schematic Representation of the System



Measurement time series are available for individual zone producing separately **(well test)** and for the situation when they all produce simultaneously.

The Production Modelling Problem

Assume that **no a priori model** is available to describe the production of a well in terms of measurable physical quantities.

Find an **algebraic model** of the production in terms of the determining, measurable physical quantities which specifically models the **interactions** occurring in this production unit.

Find such a model which correctly **predicts** the behavior of the production system over **longer time periods** (weeks or even months).

The Production Allocation Problem

Let $g_1, \ldots, g_s \in \mathbb{R}[x_1, \ldots, x_n]$ be the model polynomials for the individual zones, and let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be the model polynomial for the total production. Since the zones are interacting, we are looking for a relationship of the form

$$f = h_1\, g_1 + \cdots + h_s\, g_s$$

where $h_i \in \mathbb{R}[x_1, \ldots, x_n]$ may also involve indeterminates x_j corresponding to quantities measured at **other zones** $j \neq i$.

This is an **approximate explicit membership problem**.

Applications of Production Allocation

(1) Wells with zones inside and outside the region of tax authority of some country

(2) Wells managed jointly by different companies

The Ultimate Recovery Problem

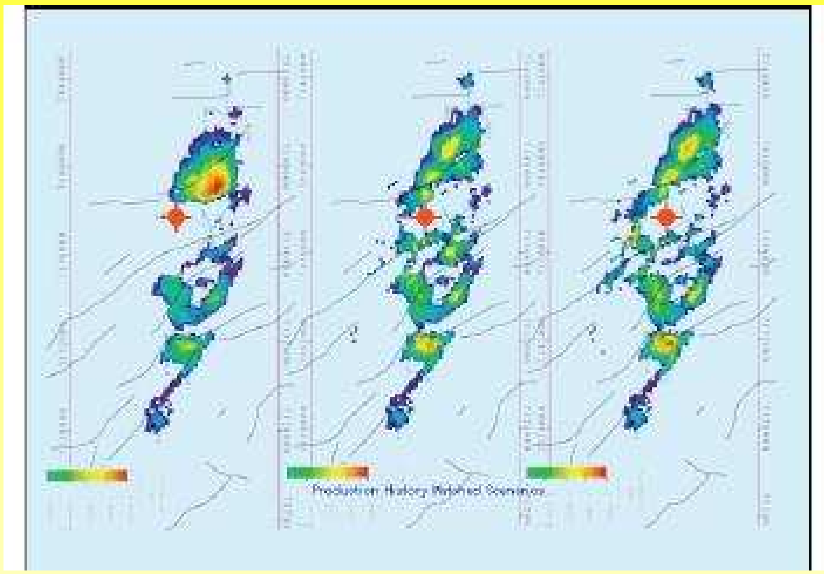


Figure 2: Long Term Changes in an Oil Field

Schematic Representation of the Changes

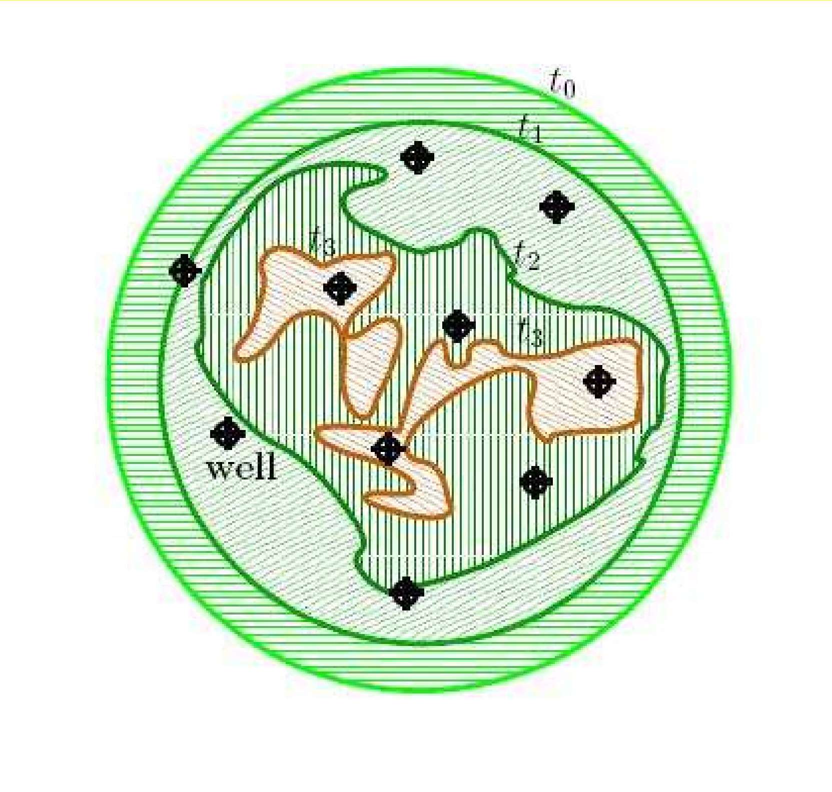


Figure 3: Schematic Changes in an Oil Field

Application to Ultimate Recovery

Based on a long term model, formulate a **production strategy** to increase the current value of **30%** for the **ultimate recovery**.

2 – Border Bases

When they started to prove even the simplest claims, many turned out to be wrong

(Bertrand Russell)

$P = \mathbb{R}[x_1, \ldots, x_n]$ 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 2.1 **(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 $\mathcal{O}$.

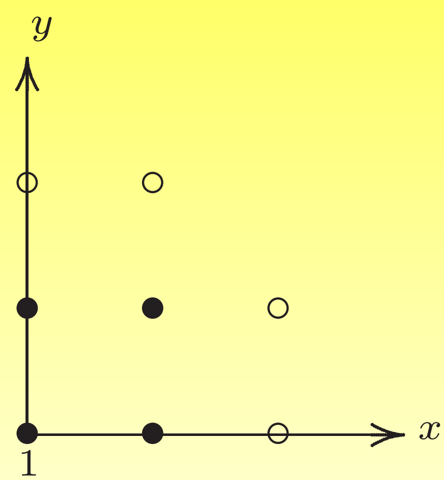
(c) Let $\mathcal{O} = \{t_1, \ldots, t_\mu\}$ be an order ideal and $\partial\mathcal{O} = \{b_1, \ldots, b_\nu\}$ its border. A set of polynomials $\{g_1, \ldots, g_\nu\} \subset I$ of the form

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

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

Definition 2.2 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 $k, \ell \in \{1, \ldots, n\}$.

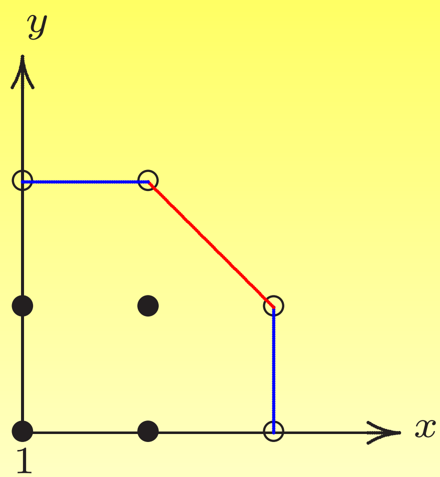
(c) The border terms b_i and b_j are called **neighbors** if they are next-door neighbors or across-the-street neighbors.

(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 2.3 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 2.4 *The border web is connected.*

Neighbor Syzygies

Definition 2.5 **(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 2.6 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 2.7 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 $\mathrm{NR}_G(S_{ij}) = g_j - x_k\, g_i - \sum_{m=1}^{\mu} a_m g_m$ 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 $\mathrm{NR}_G(S_{ij}) = x_k\, g_i - x_\ell\, g_j - \sum_{m=1}^{\mu} a_m g_m$ 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 2.8 *(Stetter)*
An $\mathcal{O}$-border prebasis G is an $\mathcal{O}$-border basis if and only if the neighbor syzygies lift.

Border Bases and Multiplication Matrices

For $r \in \{1, \ldots, n\}$, we define the r-th **formal multiplication matrix** $\mathcal{A}_r$ as follows:

Multiply $t_i \in \mathcal{O}$ by x_r. If $x_r t_i = b_j$ is in the border of $\mathcal{O}$, rewrite it using the prebasis polynomial $g_j = b_j - \sum_{k=1}^{\mu} c_{kj} t_k$ and put $(c_1, \ldots, c_\mu)$ into the i-th column of $\mathcal{A}_r$. But if $x_r t_i = t_j$ then put the j-th unit vector into the i-th column of $\mathcal{A}_r$.

Theorem 2.9 *(Mourrain)*
The set G is the $\mathcal{O}$-border basis of I if and only if the formal multiplication matrices commute, i.e. iff

$$\mathcal{A}_i \, \mathcal{A}_j = \mathcal{A}_j \, \mathcal{A}_i \qquad \text{for } 1 \leq i < j \leq n.$$

The Border Basis Scheme

Idea: Use the Commuting Multiplication Matrices / Buchberger Criterion to paramatrize $\mathcal{O}$-border bases. Given an $\mathcal{O}$-border prebasis $G = \{g_1, \ldots, g_\nu\}$ with

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

consider the c_{ij} as indeterminates. Let $I_{\mathcal{O}}$ be the ideal in $K[c_{ij}]$ generated by the entries of all $\mathcal{A}_i \mathcal{A}_j - \mathcal{A}_j \mathcal{A}_i$.

The affine subscheme $\mathbb{B}_{\mathcal{O}}$ of $K^{\mu\nu}$ defined by $I_{\mathcal{O}}$ is called the **$\mathcal{O}$-border basis scheme**. Its points correspond 1-1 to the ideals having an $\mathcal{O}$-border basis.

Properties of the Border Basis Scheme

(1) The scheme $\mathbb{B}_{\mathcal{O}}$ corresponds to an open subset of the **Hilbert scheme** $\mathrm{Hilb}^s(\mathbb{A}^n)$.

(2) The scheme $\mathbb{B}_{\mathcal{O}}$ comes equipped with a **universal flat family** whose fibers correspond to the rings $P/\langle G \rangle$ where G is an $\mathcal{O}$-border basis.

(3) A rational curve $C \subseteq \mathbb{B}_{\mathcal{O}}$ which connects the point corresponding to a given ideal $I = \langle G \rangle$ to the point corresponding to the border term ideal $\langle b_1, \ldots, b_\nu \rangle$ is nothing but a **flat deformation** from P/I to $P/\langle b_1, \ldots, b_\nu \rangle$.

3 – Approximate Border Bases

I had a fortune cookie the other day and it said:
"Outlook not so good!"
"Sure, but Microsoft ships it anyway!"

Motivation: Suppose we are given some points $\mathbb{X} = \{p_1, \dots, p_s\}$ in $\mathbb{R}^n$. When does a polynomial **vanish approximately** at $\mathbb{X}$?

Let $\varepsilon > 0$ be a given **threshold number**. We say that $f \in P = \mathbb{R}[x_1, \dots, x_n]$ **vanishes ε-approximately** at $\mathbb{X}$ if $|f(p_i)| < \varepsilon$ for $i = 1, \dots, s$.

Problem 1: The polynomials which vanish ε-approximately at $\mathbb{X}$ do not form an ideal!

Problem 2: All polynomials with very small coefficients vanish ε-approximately at $\mathbb{X}$!

Therefore we need to measure the **size** of a polynomial. In other words, we need a topology on $\mathbb{R}[x_1, \ldots, x_n]$.

Definition 3.1 Let $f = a_1 t_1 + \cdots + a_s t_s \in P$, where $a_1, \ldots, a_s \in \mathbb{R} \setminus \{0\}$ and $t_1, \ldots, t_s \in \mathbb{T}^n$. Then the number $\|f\| = \|(a_1, \ldots, a_s)\|$ is called the **(Euclidean) norm** of f.

Clearly, this definition turns P into a normed vector space. Now it is reasonable to consider the condition that polynomials $f \in P$ with $\|f\| = 1$ vanish ε-approximately at $\mathbb{X}$.

Definition 3.2 An ideal $I \subseteq P$ is called an **ε-approximate vanishing ideal** of $\mathbb{X}$ if there exists a system of generators $\{f_1, \ldots, f_r\}$ of I such that $\|f_i\| = 1$ and f_i vanishes ε-approximately at $\mathbb{X}$ for $i = 1, \ldots, r$.

Definition of Approximate Border Bases

Definition 3.3 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 3.4 If $G = \{g_1, \ldots, g_\nu\}$ is an ε-approximate border basis then the point $(c_{11}, \ldots, c_{\mu\nu})$ in $\mathbb{R}^{\mu\nu}$ given by its coefficients is **close to** the **border basis scheme**.

Example 3.5 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 $g_4 \equiv 101^2 - 0.99 \pmod{I}$.

4 – The AVI Algorithm

We all know Linux is great...
it does infinite loops in 5 seconds.
(Linus Torvalds)

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 4.1 (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:*

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

Corollary 4.2 *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 4.3 (The AVI Algorithm)
The following algorithm computes an approximate border basis of an approximate vanishing ideal of a finite set of points $\mathbb{X} \subseteq \mathbf{[-1,1]^n}$.

A1 *Start with lists* $G = \emptyset$, $\mathcal{O} = [1]$, *a matrix* $\mathcal{M} = (1,\dots,1)^{\mathrm{tr}} \in \mathrm{Mat}_{s,1}(\mathbb{R})$, *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* $L = (t_1, \dots, t_\ell)$.

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

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

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.)

Optimizations and Improvements of AVI

(1) better update strategies in Step A2 to keep the sizes of the SVD computations manageable and to improve the **strong linear independence** of the evaluation vectors of the computed order ideal

(2) **final cleaning** of the result to remove remaining almost linear dependencies and to get as close to the border basis scheme as possible

(3) recomputation of the approximate border basis using different **term orderings** (dependencies of physical quantities) and **threshold numbers** (inherent data uncertainty)

5 – Modelling Oil Production

If an experiment works,
something has gone wrong.
(Finagle's First Law)

Example 5.1 A certain two-zone oil well in Brunei yields 7200 data points in $\mathbb{R}^8$. We use 80% of the points for modelling the total production in the following way:

(a) Using the AVI algorithm with $\varepsilon = 0.05$, compute an order ideal $\mathcal{O}$ and its evaluation matrix $\mathrm{eval}(\mathcal{O})$.

(b) Find the vector in the linear span of the rows of $\mathrm{eval}(\mathcal{O})$ which is closest to the total production.

(c) The corresponding linear combination of terms in $\mathcal{O}$ is the **model polynomial** for the total production.

(d) Plot all values of the model polynomial at the given points. Compare them at the points which were not used for modelling with the actual measured values.

Example 5.2 Using the same data points, $\varepsilon = 0.1$, and the same procedure, model the total gas production.

Example 5.3 Using the same procedure, model a second oil well. Find the optimal threshold number (i.e. the **inherent variability** in the data).

6 – Further Applications

A few months in the laboratory
can save a few hours in the library.
(Westheimer's Law)

(1) Dynamic Modelling: Using the differential AVI algorithm, construct model polynomials which also capture the dynamic behavior of the system.

(2) Production-Exploration: Based on tomographic data, find **simple** algebraic surfaces which approximate the gas/oil body; deduce new production strategies.

(3) Geometric Exploration: Based on a novel interpretation of seismic data, find oil/gas bodies of non-standard geometric shapes.

References

1. (with D. Heldt, S. Pokutta, H. Poulisse) *Algebraische Modellierung mit Methoden der approximativen Computeralgebra und Anwendungen in der Ölindustrie*, OR News **28** (2006), 15–18

2. (with D. Heldt, S. Pokutta, H. Poulisse) *Approximate computation of zero-dimensional polynomial ideals*, to appear in J. Symb. Computation

3. (with H. Poulisse, L. Robbiano) *From oil fields to Hilbert schemes*, Proc. ApCoA 2008

If you want to know **THE END**,

look at the beginning.

(African Proverb)

Thank you for your attention!