Report 99-056

Workshop GROSTAT III
September 9-10, 1999
ISSN 1389-2355

Eindhoven, December 1999
The Netherlands

EURANDOM
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Secretariat: Laplace Building 1.09
Telephone: +31 40 247 8100
Telefax: +31 40 247 8190
Email: all@eurandom.tue.nl
Internet: http://www.eurandom.nl

ISSN 1389-2355

Foreword

This report contains the slides of the talks presented at GROSTAT III. This workshop was held at EURANDOM on September 9 and 10, 1999. The workshop was supported by both a grant of NWO, the Dutch Science Foundation, and EURANDOM. GROSTAT III was the third workshop in a series of workshops on applications of Gröbner basis methodologies to statistics. Previous workshops were held in Coventry (1997) and Menton (1998). These workshops provide a forum for statisticians and algebraists to meet and exchange ideas.

December 1999,

A. Di Bucchianico
G. Pistone
H.P. Wynn

GROSTAT III
Workshop on Gröbner bases and statistics

September 9 and 10, 1999

PROGRAMME:

Thursday September 9

8.30 Bus leaves from the Hotel
9.30 - 9.35 Opening by Prof. W. van Zwet, scientific director of EURANDOM

Algebraic theory of DOE

The theme of this section is the state-of-the-art of the application of algebraic theory to DOE. It also shows the relevance for applications and tries to bridge the gaps between statisticians, algebraists and practical users. The speakers are currently working on writing basic material on the subject.

09.35 - 10.10 Di Bucchianico/Cohen: Introduction to Gröbner bases and DOE
10.10 - 10.20 coffee/tea
10.20 - 11.00 Denteneer: Improved PCB estimation: design for surface estimation of DOE
11.00 - 11.20 coffee/tea
11.20 - 12.00 Riccomagno: A glimpse into the future of nonlinear theory and fans
12.00 - 14.00 Lunch (Van Trierzaal)

Algebraic Theory of Special Designs and Methods

This section is devoted to special applications of the algebraic theory to DOE. The emphasis is on current development of specific statistical methodologies.

14.00 - 14.40 Bates/Giglio/Wynn: Regression strategies and Gröbner bases
14.40 - 14.50 coffee/tea
14.50 - 15.30 Rogantin: Estimability and resolution for binary factorial fractions
15.30 - 15.50 coffee/tea
15.50 - 16.30 Ye: Response surface designs and Gröbner bases
16.30- ... coffee/tea

Algorithms and Software

Real application of the algebraic theory requires the development of efficient special algorithms and good interfaces with widely used software. This section is devoted to projects that try to accomplish this.

16.40 - 17.20 Cohen/Capprotti: OpenMath
18.00 Bus leaves for the hotel
20.00 - ... conference dinner in El Teatro (3 minutes walking from the hotel)

EURANDOM

Friday September 10

8.45 Bus leaves from the hotel

Algorithms and Software (continued)

Real application of the algebraic theory requires the development of efficient special algorithms and good interfaces with widely used software. This section is devoted to projects that try to accomplish this.

09.30 - 10.10 Robbiano/Caboara: Algebraic Tools for Statisticians.
10.10 - 10.20 coffee/tea

Probability and Statistics

This section is devoted to algebraic theory for probability computations on a finite sample case and statistical applications.

10.20 - 11.00 Dinwoodie: Conditional Distributions on Integer Data and Gröbner Bases
11.00 - 11.20 coffee/tea
11.20 - 12.00 Pistone/Wynn: Probability computations on finite sample space and applications
12.00 - 14.00 Lunch (Postumuszaal)

Differential Algebra, System Theory and Stochastics

This section is devoted to dynamical aspects of statistical modelling and design. New problems appear in the dynamic case such as controllability, estimability, identifiability and stochastic calculus.

14.00 - 14.40 Karcanias: Nonlinear Problems in Linear Systems: Gröbner Bases and Global Linearisation Methods
14.40 - 14.50 coffee/tea
14.50 - 15.30 Carrà-Ferro: Differential Resultant Systems and Statistics
15.30 - 15.50 coffee/tea
15.50 - 16.30 Margaria: Structural identifiability for polynomial differential systems
16.30 - 16.40 coffee/tea
16.40 - 17.20 Kendall: Some reflections on the algebra of stochastic calculus
17.20 - 17.30 coffee/tea
17.30 - 18.00 Conclusion + discussion of future

EURANDOM

GROSTAT III

Participants:

R.A. Bates, University of Warwick , Department of Statistics, Coventry CV4 7AL, England
Email: r.a.bates@warwick.ac.uk

E.E.M. van Berkum, Eindhoven University of Technology, Department of Mathematics,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Email: E.E.M.v.Berkum@tue.nl,
URL: http://www.win.tue.nl/~eberkum

M. Caboara, Università di Genova, Dipt. di Matematica, Via Dodecaneso 35, 16146 Genova, Italy.
Email: caboara@dima.unige.it

O. Caprotti, Eindhoven University of Technology, RIACA (Research Institute for Applications of Computer Algebra), P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Email: olga@win.tue.nl,
URL: http://www.riaca.win.tue.nl/people/olga/ .

G. Carr à-Ferro, Università di Catania, Department of Mathematics, Piazza Università 2, 95124 Italy.
Email: carra@dipmat.unict.it

A.M. Cohen, Eindhoven University of Technology, Department of Mathematics,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Email: amc@win.tue.nl,
URL: http://www.win.tue.nl/math/dw/personalpages/amc/.

D. Denteneer, Philips Research, Prof. Holstlaan 4 (WY 81), 5656 AA Eindhoven,
The Netherlands.
Email: dee.denteneer@philips.com

A. Di Bucchianico, Eindhoven University of Technology, Faculty Technology
Management, Section Quality of Products &Processes, P.O. Box 513,
5600 MB Eindhoven, The Netherlands.
Email: A.d.Bucchianico@tm.tue.nl,
URL: http://www.tm.tue.nl/vakgr/ppk/bucchianico.htm

I.H. Dinwoodie, Tulane University, Department of Mathematics, New Orleans,
Louisiana 70118.
Email: ihd@math.tulane.edu, URL: http://www.math.tulane.edu/~ihd

R. Fontana, Centro Ricerche FIAT, Strada Torino 50, I-10043 Orbassano (Torino),
Italy.
Email: r.fontana@crf.it

B. Giglio, University of Warwick, Department of Statistics, Coventry CV4 7AL,
England.
Email: b.giglio@warwick.ac.uk, URL: http://www.warwick.ac.uk/staff/B.Giglio/

E. Guardo, Università di Catania, Department of Mathematics, Piazza Università
2, 95124 Italy.
Email: guardo@dipmat.unict.it

F. Huele, CQM, Postbus 414, 5600 AK Eindhoven
Email: huele@cqm.nl

W.S. Kendall, University of Warwick, Department of Statistics, Coventry CV4 7AL,
England.
Email: W.S.Kendall@warwick.ac.uk ,
URL: http://www.warwick.ac.uk/statsdept/Staff/WSK/

P. van der Laan, Eindhoven University of Technology, Department of Mathematics, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Email:P.v.d.Laan@tue.nl,
URL: http://www.win.tue.nl/math/bs/statistics/vdlaan/

G. Margaria, Politecnico di Torino, Dipartimento di matematica, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
Email: margaria@calvino.polito.it

G. Pistone, Politecnico di Torino, Dipartimento di matematica, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
Email: pistone@calvino.polito.it,
URL: http://www.polito.it/~pistone

E.M. Riccomagno, EURANDOM, P.O.Box 513, 5600 MB Eindhoven, The Netherlands.
Email: E.Riccomagno@tue.nl,
URL: http://euridice.tue.nl/~ericcoma

M.P. Rogantin, Università di Genova, Dipt. di Matematica, Via Dodecaneso 35, 16146 Genova, Italy.
Email: rogantin@dima.unige.it,
http://www.dima.unige.it/STAFF/ROGANTIN//index.html

L. Robbiano, Università di Genova, Dipt. di Matematica, Via Dodecaneso 35, 16146 Genova, Italy.
Email: robbiano@dima.unige.it,
URL: http://www.dima.unige.it/STAFF/ROBBIANO/index.html

R. Schwabe, Freie Universität, Institut für Mathematik I, Arnimallee 2-6, D-14 195 Berlin, Germany.
Email: schwabe@math.fu-berlin.de,
URL: http://www.math.fu-berlin.de/~schwabe

J.Q. Smith, University of Warwick, Department of Statistics, Coventry CV4 7AL, England.
Email: J.Q.Smith@warwick.ac.uk,
URL: http://www.warwick.ac.uk/statsdept/Staff/JQS/index.html

M.A. van de Wiel, Eindhoven University of Technology, Department of Mathematics, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Email: markvdw@win.tue.nl,
URL: http://www.win.tue.nl/~markvdw

H.P. Wynn, University of Warwick, Department of Statistics Room 2.81, Coventry CV4 7AL, England.
Email: h.p.wynn@warwick.ac.uk
URL: http://www.warwick.ac.uk/statsdept/risk/hpw.html

N. Karcanias, City University, Department of Electrical, Electronic & Information Engineering, Northampton Square, London EC1V 0HB, England.
Email: n.karcanias@city.ac.uk , URL:
http://www.city.ac.uk/engineering/eeie/Staff/professo.htm

J.J. Seidel, Eindhoven University of Technology, Department of Mathematics, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

K.Q.Ye, University at Stony Brook, Department of Applied Mathematics and Statistics, Stony Brook, NY 11794-3600, U.S.A..
Email: kye@ams.sunysb.edu,
URL: http://www.ams.sunysb.edu/~kye

Introduction to Gröbner bases and DOE

A. Di Bucchianico and A.M. Cohen

EURANDOM / Eindhoven University of Technology

Contents

- Short practical introduction to Design of Experiments (DOE)
- Mathematical background of DOE
- The importance of Gröbner bases for DOE

What is design of experiments?

Design of Experiments is a general term for a collection of statistical techniques for *systematic experimentation*.

The most important benefits of using design of experiments are:

- *less* experiments are needed (time,costs,...)
- *interactions* between factors are taken into account

(but: Taguchi thinks that interactions are due to bad modelling)

Example

- chemical production process of vegetable oil
- relevant questions:
  - which factors influence yield (e.g., temperature, pressure, concentration of added acids, choice of washing liquid, etc.)
  - find the "optimal" values of the important factors (optimal yield or optimal costs or ...)

Main types of experimental designs

- "screening designs"

These designs are used to investigate *which* factors are important ("significant").

- "response surface designs"

These designs are used to determine the *optimal* settings of the significant factors found after performing a screening design (see talk by Ye)

N.B. For a "non-standard" example of DOE, see talk by Denteneer

Screening designs: effects

How do we determine whether an individual factor is of importance?

Basic idea: measure the outcome at 2 *different* settings of that factor (and do this simultaneously for all factors)

Scale the settings such that they become the values +1 and -1 (reason will become clear later)

• A *design* consists of performing measurements at particular level combinations.

• A *run* is a measurement at one level combination (runs may be performed with the same level combination: *replications*)

Suppose that there are 3 factors: A and B. We perform 4 measurements with the following settings:

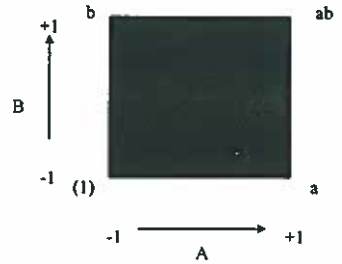
• A -1 and B -1 (short: (1))

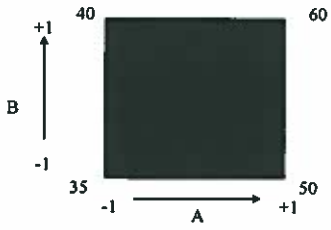
• A +1 and B -1 (short: a)

• A -1 and B +1 (short: b)

• A +1 and B +1 (short: ab)

Graphical display



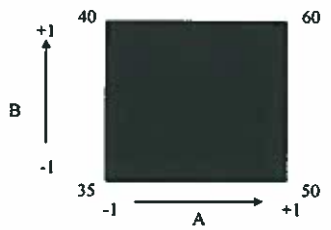


2 estimates for effect A:

50 - 35 = 15 and 60 - 40 = 20

Combine both estimates: ½(50-35) + ½(60-40)

Which estimate is superior?



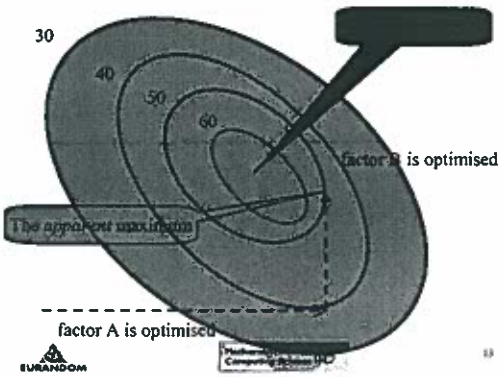
In the same way we estimate the effect B:

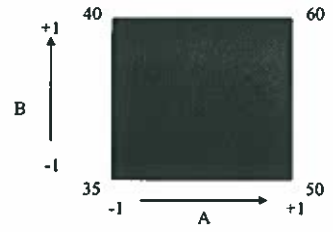
½(40-35) + ½(60-50)

Interactions

Factors may *influence each other*. E.g, the optimal setting of a factor may depend on the settings of the other factors.

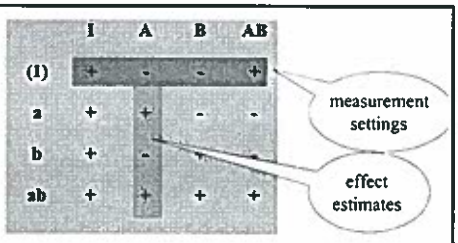
When factors are optimised *separately*, the overall result (as function of all factors) may be *suboptimal* ...





The *interaction effect* AB is the difference between the estimates for the effect A:

½(60-40) - ½(50-35)



from left to right: measurement settings .
from top to bottom: coefficients of effect estimates.

Example: for measurement (1), both A and B are set at -1.
Example: A = - (1) + a - b + ab, where small letters denote the outcomes of the measurements.

General set-up

- calculate effects
- factors whose corresponding effect is "large" are considered to be significant
- being "large" should be a scale-independent notion
- use variance as universal scale

Mathematical model (regression analysis)

- In order to avoid trivialities, assume 3 factors
- We try to fit a linear regression model to the data:

$$Y = \beta_I + \beta_A x_A + \beta_B x_B + \beta_C x_C + \beta_{AB} x_{AB} + \beta_{AC} x_{AC} + \beta_{BC} x_{BC} + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$

N.B. The term $x_A x_B x_C$ is left out, because it has no practical importance (true for any higher-order interaction)

Cross terms and interaction effects

Cross terms in linear regression models are used to model interaction effects. E.g., if

$Y = 3 + 2\,x_A + 4\,x_B + 7\,x_A\,x_B$

$x_A \to x_A + 1 \Rightarrow Y \to Y + 2 + 7\,x_B$,

so increase in Y when x_A changes, depends on x_B. Likewise for x_B.

This explains the notation AB for the interaction effect.

Designs with 3 levels (digression)

- In response surface designs, 3 levels are used in order to include quadratic terms
- A factor now has not a single effect associated to it, but both a linear and a quadratic effect (corresponding to a linear and a quadratic term with some correction to obtain orthogonality).

Regression parameters

- Obtain estimates for the regression parameters through +/-scheme or equivalently, through Least Squares
- Obtain estimate for variance σ^2 through

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{\#\text{measurements} - \#\text{parameters}}$$

where X is the design matrix (= matrix of + /-)

Variances of parameter estimators

- Note that estimates depend on the data and hence, are random variables
- The variance-covariance matrix of the random vector is given by $\sigma^2(X^TX)^{-1}$
- The matrix $(X^TX)^{-1}$ is diagonal, because of choice -1 and +1. Hence, parameter estimates are uncorrelated.

Significance of factors

- variances are used to have universal scale
- significance of individual factors are judged by dividing parameter estimates by (estimated) standard deviation of that parameter estimate.
- Only at this point an explicit distribution for the error term ε is needed.

Fractional experiments

Often the *number of parameters is too large* to allow a complete 2^p design (i.e, all 2^p possible settings -1 and 1 of the p factors).

By performing only *a subset of the 2^p experiments* in a smart way, we can arrange that by performing relatively few, it is possible to estimate the main effects and (possibly) 2nd order interactions.

	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-

- I = -C
- A = -AC
- B = -BC
- AB = -ABC

With this *half fraction* (only 4 = ½ 8 experiments) we see that a number of columns are the same (apart from a minus sign) and hence, the corresponding parameter estimates are the sae.

We say that these factors are *confounded* or *aliased*. In this particular case we have an ill-chosen fraction, because I and C are confounded. One of the two best choices here is the following fraction:

	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
ab	+	+	+	+	-	-	-	-

- I = -ABC
- A = -BC
- B = -AC
- AB = -C

Aliasing and group theory

- there exist *computational rules* for aliases. E.g., it follows from A=C that AB = BC. Note that $I = A^2 = B^2 = C^2$ etc. always holds
- The multiplication rules can be cast in terms of Abelian group theory (going back to Fisher 1942).
- Group theory cannot handle arbitrary fractions in a unified way.

Polynomial interpretation

- Let D be the set of design point and let P the space of polynomials in 3 variables. Define the inner product

$$< f, g >= \sum_{d \in D} f(d)g(d)$$

- Estimation of parameters can be explained in terms of projections in P.
- Choice of -1 and +1 makes monomials orthogonal w.r.t. <,>

EURANDOM

26

Aliasing and polynomials

- Aliasing can now be interpreted as polynomials coinciding when restricted to D.
- f is confounded with g if f-g=0 on D.
- Let I be the set of polynomials that are identically zero on D .

EURANDOM

27

Aliasing structure

- In order to describe the aliasing structure, we need a basis for the vector space P/I.
- A set polynomials $\{g_1,...,g_n\}$ is a Gröbner basis if any polynomial h can be written as

$$h = \sum_{i=1}^{n} q_i g_i + r$$

where r is unique and cannot be reduced further (division with remainder -> term ordering)

EURANDOM

28

Algorithms

- Gröbner bases can be computed. Original algorithm by Buchberger (1965)
- Algorithms for computing Gröbner bases are available in general purpose software as Maple or Mathematica. There also exists dedicated (free!) software: CoCoA (see talk by Robbiano/Caboara).

EURANDOM

29

Conclusion

- aliasing can be interpreted as polynomials coinciding when restricted to the design points
- Gröbner bases can generate the aliasing structure of arbitrary fractions
- there are computational tools for Gröbner bases

EURANDOM

31

Improved PCB-inspection: Design for Surface Estimation

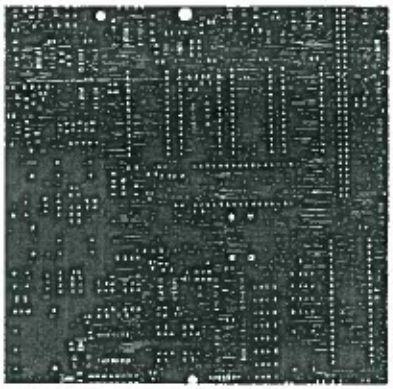
Dee Denteneer
Philips Research-Eindhoven
the Netherlands
Dee.Denteneer@philips.com

September 9, 1999

Summary

- Background and application
- Other issues
- Design issues
- Conclusion

PCB



Background

- Process control for PCBs
- Measure solder paste deposits
- Typically:
  - 1000-5000 on a PCB
  - Area: .1-1. mm^2 , .1-.2 mm high
- Measure optically

Traditionally

- Local polynomials
- Small number of reference points
- Least squares
- Evaluate at target location

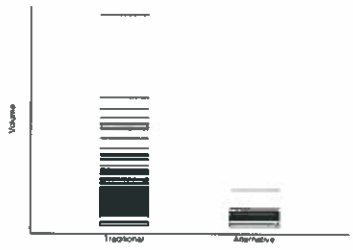
As in SURFACE ESTIMATION

Alternative:

- Local polynomials
- Large number of reference points
- Iterated least squares
- Computational trickery

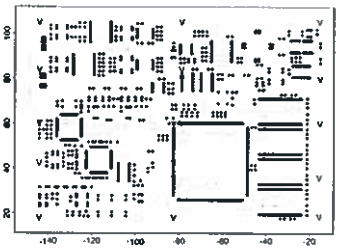
As in LOESS

Result: illustration



- Measure empty board
- Reduce third quartile
- Reduce outliers

Canonical image



- Target or 'PAD' location
- Reference location

V Vertex or fit-location

Speed up by preordering

- For each target location, t, find closest references, S_t,
  - Search all references
  - Determine search interval at start up
- Order (planar) data, π, such that

$$\sum_t \max_{(i,j)\in S_t} \pi(i) - \pi(j), \qquad (1)$$

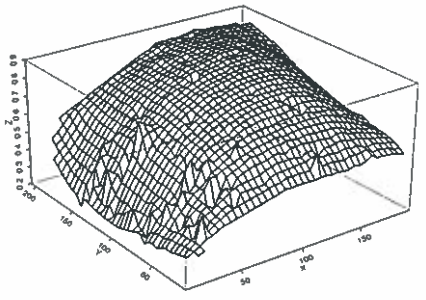
is small.

Speed up Results

- Actual boards
- Search intervals
- With and without preordering

# Targets	# References	Without	With
125	4000	1	.3
175	4500	1	.3
265	13000	1	.15
450	25000	1	.04

Motivation Adaptation



Observed interpolated references

Adaptation

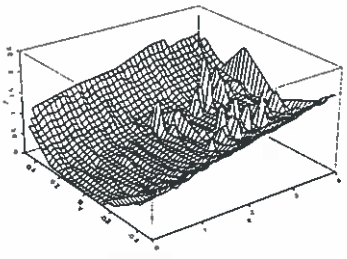
- Smoothing parameter: #Refs/Target
- Adapt:
  High noise: high bandwidth
  Low noise: low bandwidth
- Needed!
- Time restricted, simple approach

Proposal, for each target:
After iteration 1: Compute

$$C = \frac{\max\{W_d - W_d * W_r\}}{\max\{W_d\}} \qquad (2)$$

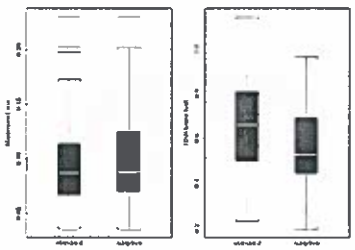
if ($C > .5$) iterate as usual
Else Square distance weights
Fit
Exit

Evaluation Adaptation



- Simulated surface:
- Gaussian noise, contaminated ridges

Results Adaptation



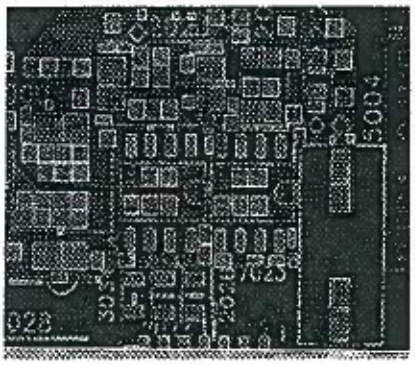
- Reduction in mean RSS: 20%
- Not very sensitive to threshold
- To be developed

Design issue 1: Select vertex locations

- Position
  - Regular grid
  - KD-tree
  - Triangulation methods
- Based on
  - Target locations
  - Reference locations

KD-tree, based on target locations, ...

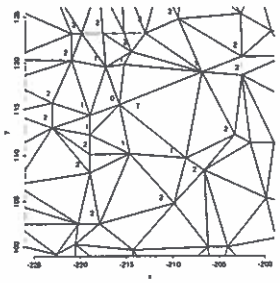
Design issue 2: Select reference locations: Select area in bare cupper



Design issue 3: Select reference locations: A subset of a candidate list

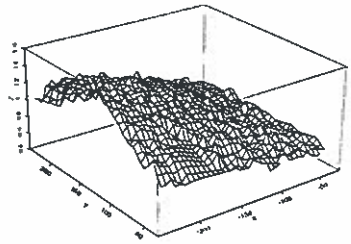
- Select reference closest to target
- Select n-neighbors in Delaunay triangulation
- Borrow rest
- (Extend to better resolution)

Example Design



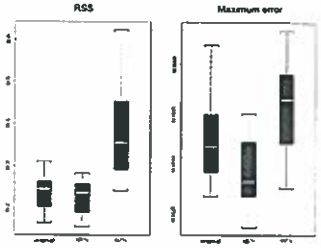
T : Target
0 : Reference nearest to target
1 : 1-neighbors of target
2 : 2-neighbors of target

Evaluation Design:



- Simulation, 16 replications
- Gaussian noise == optimistic
- Vary # targets and # neighbors

Results Design



- Example: 3 and 4 neighbors
  (32 targets, 1600 Refs, 16 replications)
- Poor resolution
- (Neighbours in infinite plane)

Improved resolution:

- Randomisation
- W-opt
  - Improved spread
  - Improved allocation to vertices

Improved resolution: W-opt
Define

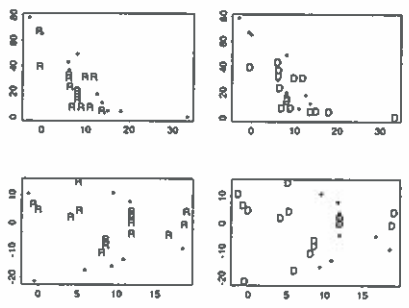
(x_0, y_0) : Vertex location
$(x_i, y_i), i = 1, \ldots, n$: Eglegible locations
$X_{i*} = (1, x_i - x_0, y_i - y_0)$: row design matrix
$d_i = ((x_i - x_0)^2 + (y_i - y_0)^2)^{1/2}/R$: relative distance
$w_i = ((1 - d_i^3)^+)^3$: distance weights
$e \in (0, 1)^n$; $\sum e_i = n^*$: Selected locations

Find e such that

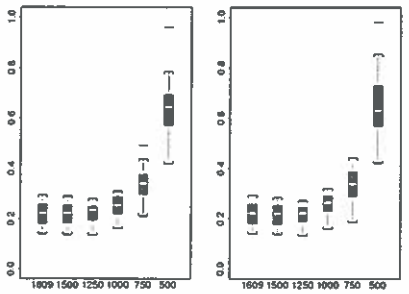
$$\Delta(x_0, y_0) = |X^t D_w D_e X|$$

is maximal.
Also; allocate extra points to vertex (x, y) with minimal $\Delta(x, y)$.

Improved resolution: W-opt
examples



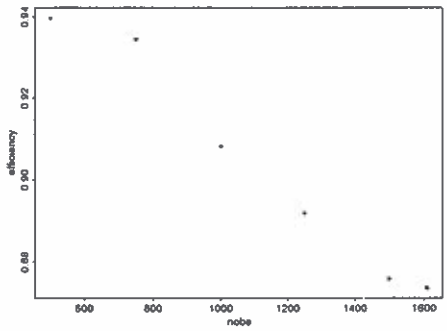
W-opt vs. Random:



Vertex selection-revisited

- Vertex locations 'outside'
- Improve location wrt. reference set
  - Optimize:
    $\Delta(x, y)$ wrt. (x, y) given e.
  - Shrink toward center

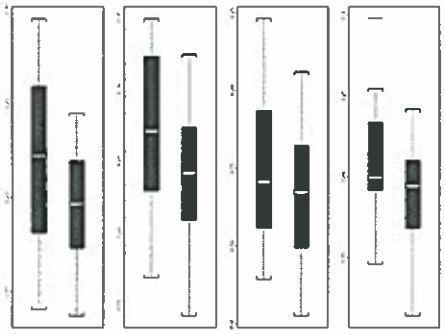
Vertex shrinking Results



Vertex shrinking Results

Various boards

Reduction in mean squared error: 88%-90%



Conclusion

- Succesful application
  - Extends engineering approach
  - Other surface fitters?
- Unresolved new issues
  - Adaptivity, also between PCBs
  - Design issues
- Understand noise process

Gröbner bases in Experimental design

A glimpse into the future of non linear theory and fans

Algebraic Statistics

Eva Riccomagno *et al.*
EURANDOM
Riccomagno@eurandom.nl

Overview

Design:

no replicates
d factors, N points
quantitative/qualitative
complex in number and structure
$D = \{a(i) \in \mathbf{Q}^d : i = 1, \ldots, N\}$

Model:

standard linear regression model
interactions
polynomial but not only
$y(x) = \Sigma_{\alpha \in L} \theta_\alpha x^\alpha + \varepsilon$ where e.g. $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

Identifiability

$$f = \sum_{\alpha \in L} \theta_\alpha x^\alpha \qquad D$$

Design matrix: $X = [a(i)^\alpha]_{i=1,\ldots,N;\alpha \in L}$ $\quad X(f, D)$

Given D, which L?

One L (Est_τ)
A class of Ls (Fan)
Aliasing and confounding structure (NF)

Given L, which D?

stats.pkg (CoCoA)
2^d
Linear algebra, random search, L_f

(D, L_i) $i = 1, \ldots$

Factor screening
Statistical analysis
Orthogonal analysis
Mallow's C_p
Akaike criterion
Model order choice criteria
Smoothness criteria
Stepwise regression

Design Ideal

$D \subset \mathbf{Q}^d$, Variety, $\mathbf{Q}[x_1, \ldots, x_d]$ or $\mathbf{Q}[x]$

$$\text{Ideal}(D) = \{f \in \mathbf{Q}[x_1, \ldots, x_d] : f(a(i)) = 0 \text{ for } i = 1, \ldots, N\}$$

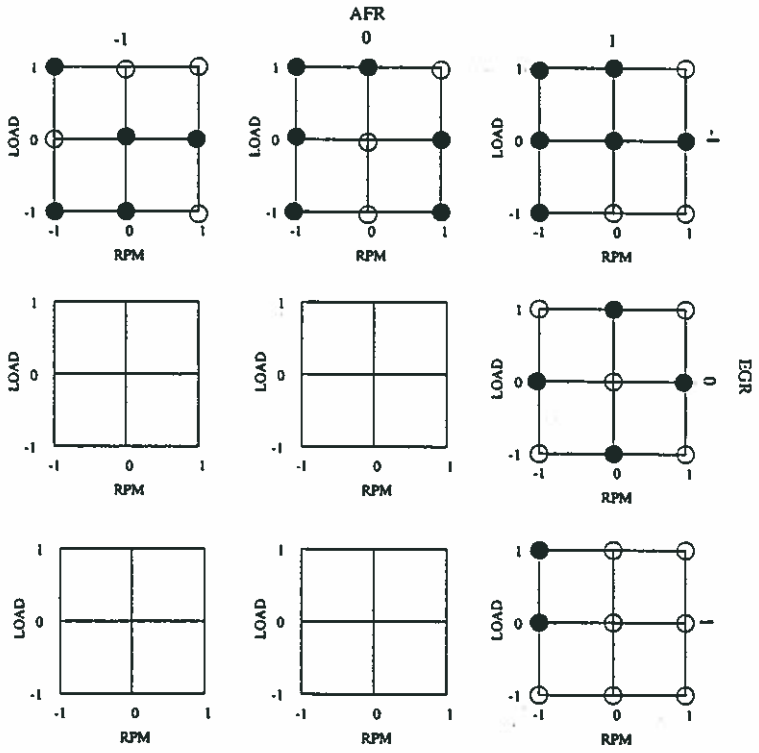
- ideal
- radical (because no replicates)
- finitely generated (Hilbert basis theorem)
- computable

Example

$2^2 = \{(\pm 1, \pm 1)\}$ $\qquad (x_1^2 - 1), (x_2^2 - 1)$

$s_1(x_1^2 - 1) + s_2(x_2^2 - 1)$

The completed design



Bidimensional representation of the 3^4 full factorial. The original designed experiment includes the empty circles; completed runs are given by black dots.

The intended design

Ideal corresponding to a 3^4 experiment

$$a(a^2-1)=0$$
$$e(e^2-1)=0$$
$$l(l^2-1)=0$$
$$r(r^2-1)=0$$

with the additional equation

$$(a-1)(e+1)=0$$

to give points at level $a=1$ *or* at level $e=-1$.

Note: Universal and total G-bases.

The actual design $\mathrm{tdeg}(r>l>e>a)$

DesId;

$ae+a-e-1$,
re^2+re+e^2+e,
$rle+le^2+rl+le$,
$r^2e+l^2e-le^2+r^2+l^2-le-e-1$,
e^3-e, a^3-a, l^3-l, r^3-r,
$r^2la+rla$, $r^2l^2+r^2l+rl^2+rl$,
$l^2e^2+l^2e-le^2-le$,
$l^2a^2-5/2r^2l-3/2rl^2+1/2r^2a+1/2rla-l^2a+$
$1/2ra^2+3/2l^2e-1/2r^2$,
$-3/2rl-1/2ra+1/2la-1/2a^2-3/2le-1/2l+1/2$,
$rla^2+2r^2l+rl^2-rla+la^2-l^2e+rl-la+le$,
$r^2a^2-r^2l-rl^2+rla+l^2e-r^2-rl+la-a^2-le-l+1$,
$rl^2a+2r^2l+rl^2-rla+l^2a-2l^2e+rl-l^2-la+2le+l$

Design ideal and Gröbner basis

Monomials: $x^\alpha \in \mathbf{Q}[x]$ Ex. $x_1^3 x_2^4 x_4$, x_1

Term ordering, τ: a totally ordering relations on monomials such that $1 < x^\alpha$ for all α

Examples: `tdeg`, `plex`, blocked

$f \in \mathbf{Q}[x]$, $\mathrm{LT}_\tau(f)$

A finite set $G \in \mathbf{Q}[x]$ is a Gröbner basis of a polynomial ideal I with respect to τ iff

$$\langle \mathrm{LT}_\tau(f) : f \in I \rangle = \langle \mathrm{LT}_\tau(g) : g \in G \rangle$$

— $G = \{g_1, \ldots, g_t\}$ is a Gröbner Basis iff

$$< LT(g_i) : i = 1, \ldots, t >=< LT(I) >$$

— is minimal iff

. $LC(g_i) = 1$ for all i

. $LT(g_i) \notin < LT(G \setminus \{g_i\}) >$

— is reduced iff

. $LC(g_i) = 1$ for all i

. for all terms of g_i does NOT belong to

$$< LT(G \setminus \{g_i\}) >$$

for all i.

Properties

- There exist a unique reduced Gröbner bases
- For $f \in \mathbf{Q}[x]$ there exist a unique r such that

$$f = \sum_{g \in G} s_g g + r$$

such that $r <_\tau g$ for all $g \in G$

- ...

Quotient space

$$\begin{aligned}\mathbf{Q}[x]/\text{Ideal}(D) \;=\; &\{[f] : f \in \mathbf{Q}[x] \text{ and} \\ &[f] = \{g : f - g \in \text{Ideal}(D)\}\}\end{aligned}$$

- Vector space
- Finite
- Given τ, reduced the Gröbner basis, G

$$\begin{aligned}\text{Est} \;&=\; \{x^\alpha : x^\alpha \text{ is not divisible by LT}(g)\ g \in G\} \\ &=\; \{x^\alpha : \alpha \in L\}\end{aligned}$$

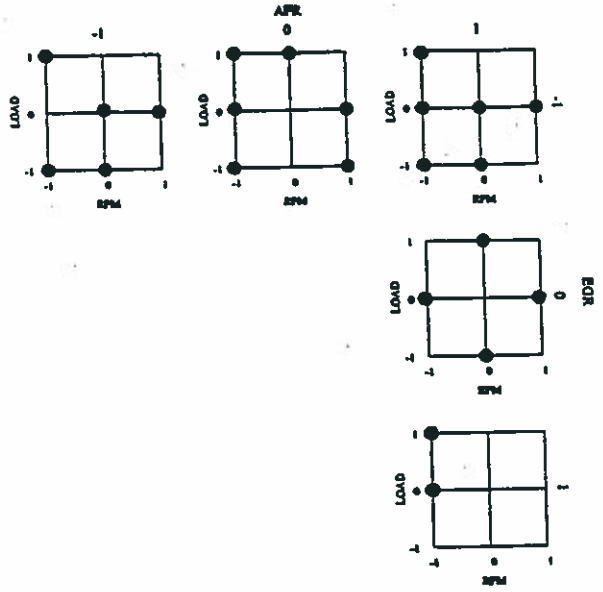
Properties of Est

- #Est =#D
- Marginal funct.ty/order ideal/hierarchical model
- A saturated, linear, identifiable model for D

$$\sum_{\alpha \in L} \theta_\alpha x^\alpha$$

- $D_1 \subseteq D_2$ then $\text{Est}_\tau(D_1) \subseteq \text{Est}_\tau(D_2)$
- Computable
- ...

An engine mapping experiment (FORD)



Full factorial: quadratic model.
`Lm;` a^3, e^3, l^3, r^3;

Complete design:
`Lm;` ea, a^3, e^3, l^3, r^3;
`Est:=OrderId(Lm); Est;`
$r^2l^2e^2, r^2l^2a^2, r^2l^2e, r^2le^2, rl^2e^2, r^2l^2a, r^2la^2, rl^2a^2, r^2l^2,$
$r^2le, rl^2e, r^2e^2, rle^2, l^2e^2, r^2la, rl^2a, r^2a^2, rla^2, l^2a^2,$
$r^2l, rl^2, r^2e, rle, l^2e, re^2, le^2, r^2a, rla, l^2a, ra^2, la^2,$
$r^2, rl, l^2, re, le, e^2, ra, la, a^2,$
$r, l, e, a, 1$

Actual design: tdeg($r > l > a > e$)
`Lm;`
$ae, e^3, re^2, rle, r^2e, a^3, l^3, r^3, r^2la,$
$r^2l^2, l^2e^2, l^2a^2, rla^2, r^2a^2, rl^2a$
`EstBasis:=OrderId(Lm); EstBasis;`
$r^2l, rl^2, r^2a, rla, l^2a, ra^2, la^2, l^2e, le^2,$
$r^2, rl, l^2, ra, la, a^2, re, le, e^2,$
$r, l, a, e, 1$

Division

$f \in \mathbf{Q}[x]$ and $g_1, \ldots, g_t \in \mathbf{Q}[x]$

$$f = \sum_{i=1}^{t} s_i g_t + r(x)$$

Model Confounding

D, τ, G

$$\begin{array}{ccc} \mathbf{Q}[x] & \longrightarrow & \mathbf{Q}[x]/\mathrm{Ideal}(D) \\ h = \sum_{\beta \in L'} \phi_\beta x^\beta & \longmapsto & [h] = NF(h, G) = \sum_{\alpha \in L} \phi'_\alpha x^\alpha = r(x) \end{array}$$

$$h(a(i)) = r(a(i))$$

Parameter Aliasing

$$\phi'_\alpha = \text{linear function}(\phi_\alpha)$$

EXAMPLES

$$
\begin{aligned}
&(p_{2,1,0} + p_{2,3,0})\, {x_1}^2 x_2 + p_{2,0,2}\, {x_1}^2 {x_3}^2 + p_{2,0,0}\, {x_1}^2 + p_{2,2,0}\, {x_1}^2 {x_2}^2 \\
&\quad + p_{2,2,2}\, {x_1}^2 {x_2}^2 {x_3}^2 + (p_{1,2,0} + p_{3,2,0})\, x_1 {x_2}^2 \\
&\quad + (p_{2,0,1} + p_{2,0,3})\, {x_1}^2 x_3 + (p_{0,2,1} + p_{0,2,3})\, {x_2}^2 x_3 \\
&\quad + (p_{1,0,2} + p_{3,0,2})\, x_1 {x_3}^2 \\
&\quad + (p_{1,1,0} + p_{3,1,0} + p_{1,3,0} + p_{3,3,0})\, x_1 x_2 \\
&\quad + (p_{1,0,1} + p_{3,0,1} + p_{1,0,3} + p_{3,0,3})\, x_1 x_3 \\
&\quad + (p_{0,1,1} + p_{0,3,1} + p_{0,1,3} + p_{0,3,3})\, x_2 x_3 \\
&\quad + (p_{0,1,2} + p_{0,3,2})\, x_2 {x_3}^2 + (p_{0,1,0} + p_{0,3,0})\, x_2 \\
&\quad + (p_{1,1,2} + p_{3,1,2} + p_{1,3,2} + p_{3,3,2})\, x_1 x_2 {x_3}^2 + (p_{1,1,1} + p_{3,1,1} \\
&\quad + p_{1,3,1} + p_{3,3,1} + p_{1,1,3} + p_{3,1,3} + p_{1,3,3} + p_{3,3,3})\, x_1 x_2 x_3 \\
&\quad + (p_{2,1,2} + p_{2,3,2})\, {x_1}^2 x_2 {x_3}^2 + (p_{1,2,2} + p_{3,2,2})\, x_1 {x_2}^2 {x_3}^2 \\
&\quad + (p_{2,1,1} + p_{2,3,1} + p_{2,1,3} + p_{2,3,3})\, {x_1}^2 x_2 x_3 \\
&\quad + (p_{1,2,1} + p_{3,2,1} + p_{1,2,3} + p_{3,2,3})\, x_1 {x_2}^2 x_3 \\
&\quad + (p_{2,2,1} + p_{2,2,3})\, {x_1}^2 {x_2}^2 x_3 + (p_{0,0,1} + p_{0,0,3})\, x_3 \\
&\quad + (p_{1,0,0} + p_{3,0,0})\, x_1 + 2\, p_{0,0,0} + p_{0,2,2}\, {x_2}^2 {x_3}^2 + p_{0,2,0}\, {x_2}^2 \\
&\quad + p_{0,0,2}\, {x_3}^2
\end{aligned}
$$

4.5

Algebraic fan

$D \qquad \mathrm{Est}_\tau(D)$

$$\mathcal{F} = \{\ \mathrm{Est}_\tau(D) : \tau \text{ is a term ordering}\}$$

- Fan and leaves
- $\mathcal{F}$ is finite
- Computation

Statistical fan

$N \in \mathcal{Z}_+^*$

$\mathcal{G}(N, d) =$ the set of all models with N terms in d variables and with an order ideal structure

For $d = 2$, partition of identity.

$$\mathcal{E}(D) = \{f \in \mathcal{G}(N, d) : \\ N = \#D \text{ and } X(f, D) \text{ is invertible}\}$$

- $\mathcal{F}(D) \subseteq \mathcal{E}(D)$
- $D = \{(0,0),(0,-1),(1,0),(1,1),(-1,1)\}$

$$\begin{array}{c} \bullet \\ \bullet \\ \bullet\ \bullet\ \bullet \end{array}$$

Caboara and Robbiano (1997)

- If $f \in \mathcal{E}(D)$ and there exists τ such that the term of f are the N smallest terms w.r.t. τ then $f \in \mathcal{F}(D)$
- Example for which $\mathcal{E}(D) = \mathcal{F}(D)$. Ex. Full factorial

Minimal fan designs

$\#\mathcal{F}(D) = 1$

- Full factorial
- Generalised echelon design

Application in reliability

Maximal fan designs

$\mathcal{E}(D) = \mathcal{G}(N, d)$

- Always exist
- A basis of Ideal(D) should be of the type

  $$g_d(x_d), x_{d-1} - g_{d-1}(x_d), \ldots, x_1 - g_1(x_d)$$

  with $\deg g_d = N$ and $g_j(a(i_1)) \neq g_j(a(i_2))$ for all pairs of design points and all $j = 1, \ldots, d-1$ (cfr. Shape Lemma) !!!!

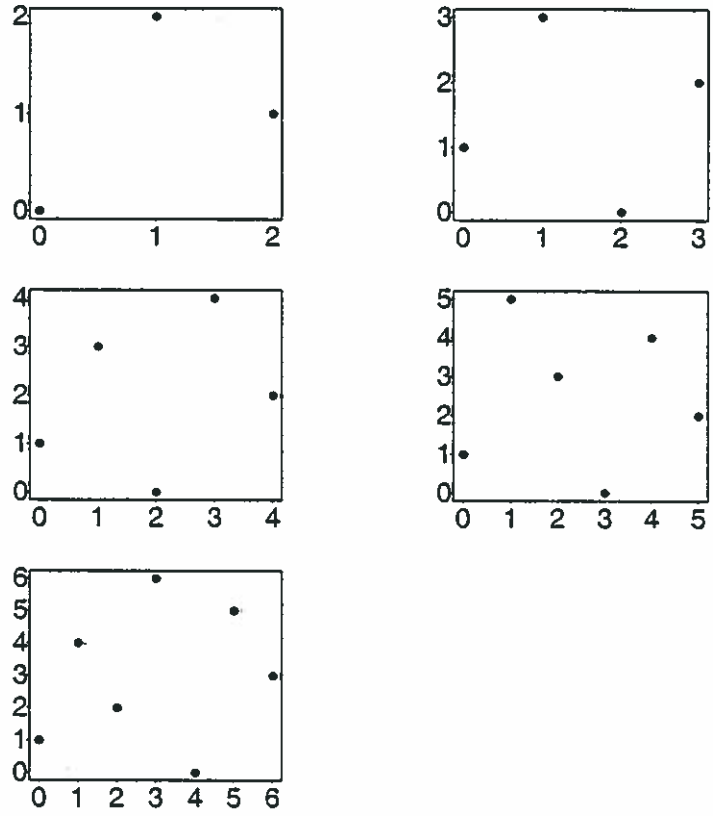
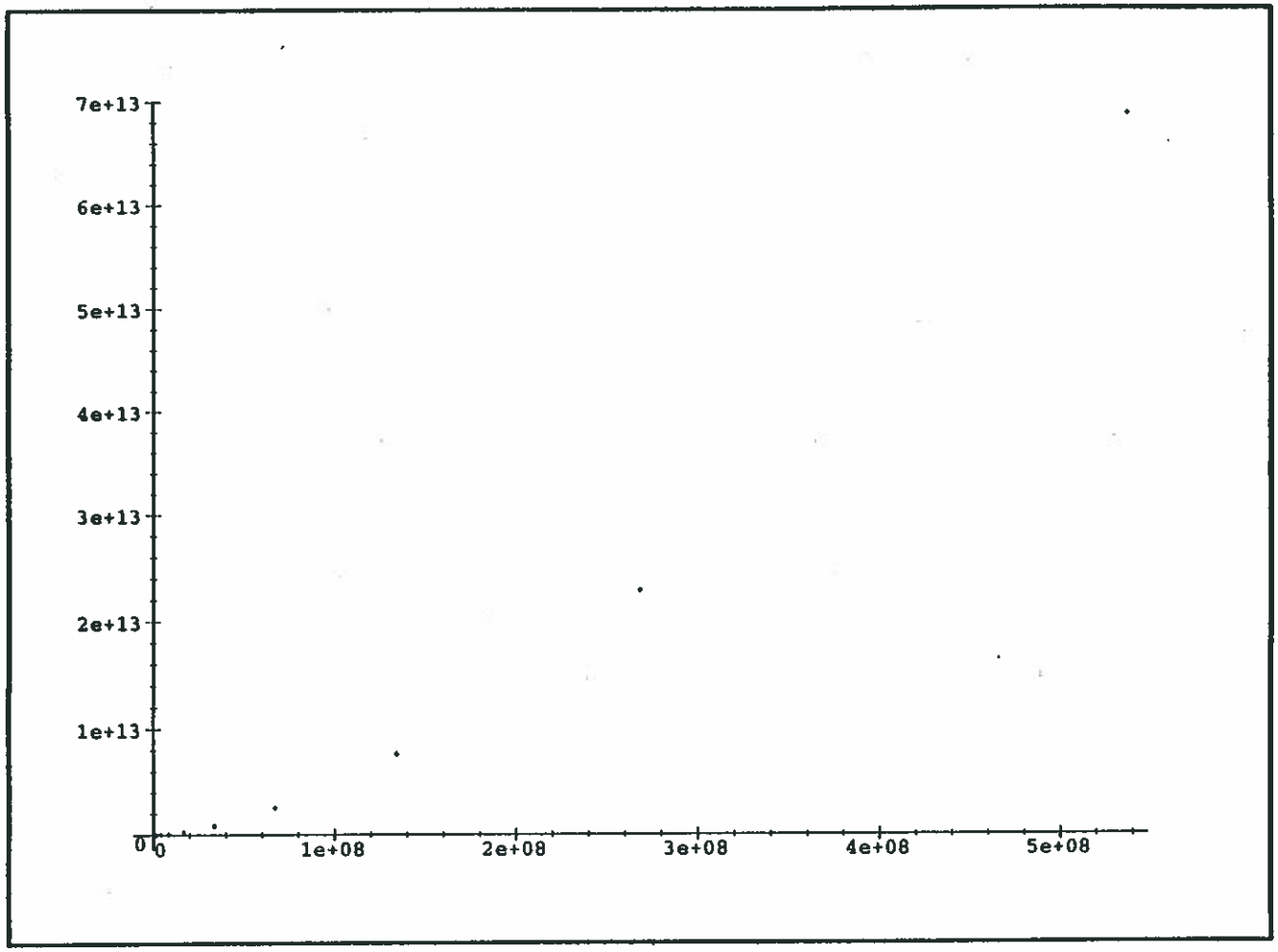


Figure 2: Two dimensional maximal fan designs with n points based on the integer grid $n \times n$ ($n = 3, \ldots, 7$).


7e+13
6e+13
5e+13
4e+13
3e+13
2e+13
1e+13
0
0
1e+08
2e+08
3e+08
4e+08
5e+08

D-optimality

$$\max_{D \subset \Xi} \det\left(X_f(D)^t X_f(D)\right)$$

Fan D-optimality

$$\max_{D \subset \Xi} \left(\prod_{M \in E} \det\left(X_M^t(D) X_M(D)\right) \right)^{1/p}$$

- $M \in E \subset \mathcal{F}$
- $p = nm$ where
  $n = \#D$
  $m = \#E$
- $\Xi = [0,1]^d$.

13.1

Robust design-results

Number of factors	Runs	Number of models	D_r - Determinant using random points	D_l - Determinant using list	D_{av} - Average of 1000 random determinants	The Best LH design	Average LH design	The Average of 1000 Random LH designs	The Best of 1000 Random Latin Hypercube designs
2	3	3	0.5090	0.5093	0.1036	0.5066	0.4053	0.3992	0.5066
2	4	5	0.2873	0.2968	0.0485	0.2809	0.2048	0.2047	0.2809
2	5	7	0.1459	0.1572	0.0241	0.1467	0.0767	0.0745	0.1467
2	6	11	0.0878	0.0882	0.0135	0.0802	0.0525	0.0530	0.0802
2	7	15	0.0540	0.0550	0.0075	0.0452	0.0275	0.0271	0.0452
3	2	3	0.9475	1	0.2638	1		1	1
3	3	6	0.4835	0.5724	0.1163	0.5724	0.1832	0.1402	0.5724
3	4	13	0.2821	0.3488	0.0684	0.3093	0.1917	0.1847	0.3093
3	5	24	0.196	0.2410	0.0454	0.2117	0.0689	0.0683	0.2112

$$D_r, D_l = \sqrt[nm]{\prod_{i=1}^{m} D_i} ,$$

where D_i is the determinant of the i-th model. n - number of the runs.
m - number of the models.

$$D_{av} = \frac{\sum_{j=1}^{1000} \sqrt[nm]{\prod_{i=1}^{m} D_i}}{1000}$$ .where D_{av} is the average determinant

Model ideal

$(a(i), Y_i) \in$ Design Space $\times$ Observation Space

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \frac{\theta_3}{x_1 x_2}$$

$$\begin{cases} t x_1 x_2 - 1 \\ y x_1 x_2 - \theta_0 x_1 x_2 - \theta_1 x_1^2 x_2 - \theta_2 x_1 x_2^2 - \theta_3 \end{cases} \text{!!!}$$

$$D = \{(1,1),(2,1),(1,2)\}\ \texttt{lex}(y > x_2 > x_1)$$

$$\begin{aligned} GB \ = \ & \{\, y + (Y_1 - Y_2)x_1 - 3Y_1 + Y_2 + Y_3 + (Y_1 - Y_3)x_2 \\ & x_1^2 - 3x_1 + 2, x_1 x_2 - x_1 + 1 - x_2, 2 - 3x_2 + x_2^2 \,\} \end{aligned}$$

$$\begin{aligned} NF \ = \ & \{(-\theta_0 + 2Y_2 - 3\theta_1 - \theta_2 - Y_1)x_1 + \\ & (-\theta_1 - Y_1 + 2Y_3 - \theta_0 - 3\theta_2)x_2 \\ & -\theta_3 + \theta_0 - 2Y_2 + 3\theta_1 + 3\theta_2 + 3Y_1 - 2Y_3 \,\} \end{aligned}$$

$$\begin{cases} \theta_0 + 4\theta_2 \ = \ -Y_1 - Y_2 + 3Y_3 \\ \theta_1 - \theta_2 \ = \ -Y_3 + Y_2 \\ \theta_3 - 2\theta_2 \ = \ 2Y_1 - 2Y_3 \end{cases}$$

Gröbner basis methods for structuring and analysing complex industrial experiments

Giovanni Pistone
Dipartimento di Matematica, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino (Italy)
pistone@calvino.polito.it
tel.: +39 +11 5647550 fax.: +39 +11 5647599

Eva Riccomagno (corresponding author)
Department of Statistics, University of Warwick
Coventry, CV4 7AL (UK)
emr@stats.warwick.ac.uk
tel.: +44 +1203 524630 fax.: +44 +1203 524532

Henry P. Wynn (presenting author)
Department of Statistics, University of Warwick
Coventry, CV4 7AL (UK)
emr@stats.warwick.ac.uk
tel.: +44 +1203 523625 fax.: +44 +1203 524532

Gröbner basis in Experimental design

Design: $D = \{a(i) \in Q^d : i = 1, \ldots, N\}$

Interpolator(s): $f \in Q[x_1, \ldots, x_d]$ s.t. $f(a(i)) = 0$, $i = 1, \ldots, N$

Design ideal: $\{f : f(a(i)) = 0, i = 1, \ldots, N\}$

Quotient space: $Q[x_1, \ldots, x_d]/\text{Ideal}(D)$

A vector basis of $Q[x_1, \ldots, x_d]/\text{Ideal}(D)$: $\text{Est} = \{h_\alpha(x) : \alpha \in L\}$

Regression model: $y(x) = \Sigma_{\alpha \in L' \subset L} \theta_\alpha h_\alpha(x)$

Responses Y and X design matrix

Identifiability: $Y = X\theta \longrightarrow \hat{\theta} = (X^t X)^{-1} X^t Y$

A polynomial model: $g(x) \in Q[x_1, \ldots, x_d]$

Model confounding: $\text{NForm}(g, \text{Ideal}(D)) = \Sigma_{\alpha \in L} \phi_\alpha x^\alpha$

ϕ_2

$h_2(x)$

• MODEL AFTER STEPWISE REGRESSION:

Shear =
-29.97 + 282.51*Binder + 9.58*Steel – 559.13*Binder^2

	Value	Std. Error	T value	Pr(>|t|)
Intercept	-29.9731	6.7639	-4.4314	0.0002
Binder	282.5112	58.9834	4.7897	0.0001
Steel	9.5777	1.7206	5.5666	0.0000
Binder^2	-559.1349	126.9370	-4.4048	0.0002

Residual standard error: 0.4316 on 22 dof
Multiple R-Squared: 0.8239
F-statistic: 34.32 on 3 and 22 dof, the p-value is 1.784e-008

Leading terms	a^2	e^2	adx	ea	d^3	ex
RegSS (h)	0.01	0.01	0.02	0.09	0.13	0.14
	ax	ax^3	x^3	x^2	ax^2	ad
	0.14	0.20	0.21	0.44	0.50	0.55
	p	d^2x	dx^3	x^4	a^2x	e
	0.80	1.01	1.07	1.10	1.29	2.56
	dx	dx^2	a	d^2	x	d
	3.76	5.09	7.46	8.79	26.09	27.55
					const	Total
					633.63	722.64

Table 9: Case study: contributions to RegSS

Figure: Observed vs. Fitted Values

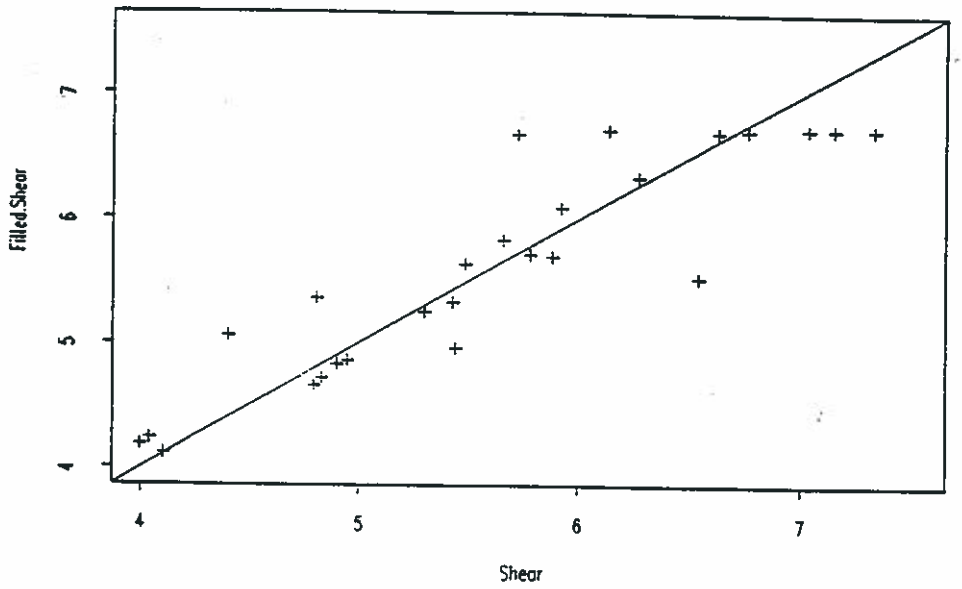
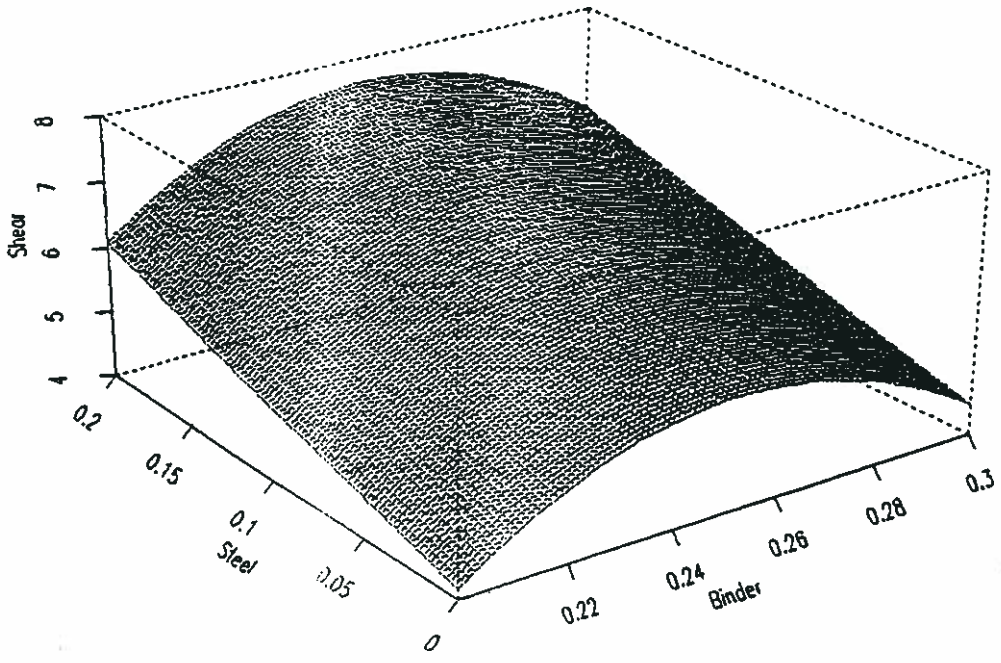


Figure: Relationship between Binder and Steel and the fitted values for Shear



Qualitative Factors

- The given design is

$$Design := \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{array}\right]$$

- $Est = \{1, x_8, x_7, x_6, x_4, x_3, x_2\}$
- Gröbner Basis:
  $Ideal(x_5+x_6+x_7+x_8-1, x_1+x_2+x_3+x_4-1, x_8^2-x_8, x_7x_8, x_6x_8, x_4x_8, x_3x_8, x_2x_8, x_7^2-x_7, x_6x_7, x_4x_7-x_4, x_3x_7+x_4-x_7, x_2x_7, x_6^2-x_6, x_4x_6, x_3x_6-x_3-x_4+x_7, x_2x_6+x_3+x_4-x_6-x_7, x_4^2-x_4, x_3x_4, x_2x_4, x_3^2-x_3, x_2x_3, x_2^2-x_2)$

DOING STATISTICS

1. $Y = X_\tau \theta + \epsilon$

$$X_\tau = \{x^\alpha\}$$

$$x^\alpha \in est_\tau(D), x \in D$$

2. Try to choose τ in order to `favour` terms you want

3. Two-stage procedures:

   (a) Screening

   (b) Favour

BLOCKING

$$\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix}$$

$$A_j \to \{x_i\}_{i \in J_j} \qquad \text{BLOCK}$$

- `plex` between blocks
- `tdeg` within blocks

ORDER

"Graded" ordering

$$x^{\alpha} \succ_{\tau} x^{\beta}$$

$$\Rightarrow \sum \alpha_i \geq \sum \beta_i$$

Theorem:

For all graded orderings

$$\sharp\{x^{\alpha} \in Est \mid deg(x^{\alpha}) = s\}$$

is same

$$deg(x^{\alpha}) = \sum \alpha_i$$

Look at 3 more closely:

We can sequentially build up designs `one point` at a time and we add one term to Est at a time

$$Est_\tau(D_{n+1}) \supseteq Est_\tau(D_n)$$

This is the basis for fast alghoritms.

Estimability and resolution for binary factorial fractions

Maria Piera Rogantin
Department of Mathematics
Università di Genova

Abstract (1)

Each fraction $\mathcal{F}$ of a factorial design is characterized by the ANOVA representation of its **polynomial indicator function** F. Moreover it is known that **lists of estimable functions** $\mathrm{Est}_\sigma(\mathcal{F})$ on the fraction are provided using Gröbner basis theory.

Many statistical properties of binary factorial fractions can be easily detected by inspecting F and $\mathrm{Est}_\sigma(\mathcal{F})$.

Abstract (2)

In particular the **indicator function** provides information about:

- regularity (or, if the fraction is not regular, about the smallest regular fraction containing it);
- projectivity;
- orthogonality between simple or interaction terms;
- a basis of orthogonal terms for $\mathcal{F}$, if there exists.

In the set $\mathrm{Est}_\sigma(\mathcal{F})$ are contained informations about:

- aliasing between simple or interaction terms
- maximal resolution.

An example: a fraction, $\mathcal{F}_H$, of a 2^9 design

it is built form a Hadamard matrix of order 16
[H16.5 in Hall (1961)]
cfr. Sun & Wu, 1993

1	1	1	1	1	1	1	1	1
1	1	1	1	-1	-1	-1	-1	-1
1	1	-1	-1	1	1	1	1	-1
1	1	-1	-1	-1	-1	-1	-1	1
1	-1	1	-1	1	1	-1	-1	1
1	-1	1	-1	-1	-1	1	1	-1
1	-1	-1	1	1	-1	1	-1	1
1	-1	-1	1	-1	1	-1	1	-1
-1	1	1	-1	1	1	-1	-1	-1
-1	1	1	-1	-1	-1	1	1	1
-1	1	-1	1	1	-1	-1	1	-1
-1	1	-1	1	-1	1	1	-1	1
-1	-1	1	1	1	-1	1	-1	-1
-1	-1	1	1	-1	1	-1	1	1
-1	-1	-1	-1	1	-1	-1	1	1
-1	-1	-1	-1	-1	1	1	-1	-1

9 factors - 16 runs

Notations

- $\mathcal{D}$ is the set of the points of full factorial design, e.g. $\{-1,1\}^m$
- $\mathcal{F}$ is the set of the points of the fraction, $\mathcal{F} \subseteq \mathcal{D}$ (no restrictions)
- $X_1, X_2, \ldots X_m$ are the simple factors of the design
- $\mathcal{P}$ is the set of parts of $\{1, \ldots, m\}$
- $C_I = \prod_{i \in I} X_i$ is the interaction between the factors with indices in I, $I \in \mathcal{P}$; it order is $\#I$ (C_I could be also a simple factor)
- if I and J are indices in $\mathcal{P}$ then Δ is the "symmetrical difference" (exclusive or):

$$I \Delta J = (I \cup J) - (I \cap J)$$

Note:

$$(\mathcal{P}, \Delta) \sim (\{-1,1\}^m, \times) \quad \text{as a group}$$

1. Indicator function of $\mathcal{F}$

If A is a point in $\mathcal{D}$, the **indicator function** F of $\mathcal{F}$ w.r.t. $\mathcal{D}$ is a $\mathcal{D}$-reduced polynomial such that:

$$F(A) = \begin{cases} 1 & \text{if} \quad A \in \mathcal{F} \\ 0 & \text{if} \quad A \in \mathcal{D} \backslash \mathcal{F} \end{cases}$$

characterizing the ideal of $\mathcal{F}$ as:

$$\begin{cases} I(\mathcal{F}) & = \text{Ideal}\left(X_1^2 - 1, \ldots, X_m^2 - 1, F - 1\right) \\ F^2 - F & \in \text{Ideal}(\mathcal{D}) \end{cases} \tag{2}$$

The indicator function can be represented by

$$F = \sum_{K \in \mathcal{P}} b_K \, C_K \tag{4}$$

- **Relations between the coefficients of the indicator function:**

$$b_I = \sum_{J \in \mathcal{P}} b_J \, b_{I \Delta J}$$

[implied by (2) and (4)]

- **Relations between the coefficients of the indicator function and the points of $\mathcal{F}$:**

$$b_I = \frac{1}{2^m} \sum_{A \in \mathcal{F}} C_I(A)$$

[as $\{C_I\}_{I \in \mathcal{P}}$ is an orthonormal set]

In particular: $b_\emptyset$ is the ratio between the cardinality of the fraction $\mathcal{F}$ and the cardinality of the full design $\mathcal{D}$:

$$b_\emptyset = \frac{\#\mathcal{F}}{2^m}$$

Indicator function and statistical properties:

- A simple or interaction term C_I is a **contrast** on $\mathcal{F}$, that is $\sum_{A \in \mathcal{F}} C_I(A) = 0$, if and only if

$$b_I = 0$$

- Two simple or interaction terms C_I and C_J are **mutually orthogonal** on $\mathcal{F}$, that is $\sum_{A \in \mathcal{F}} C_I(A)\, C_J(A) = 0$, if and only if

$$b_{I \Delta J} = 0$$

- A simple or interaction term C_I is **aliased with the general mean 1** if and only if the corresponding coefficient b_I is, in absolute value, equal to $b_\emptyset$

- A fraction $\mathcal{F}$ is **regular** if and only if **all** the non-null coefficients of its indicator function are equal in absolute value:

$$|b_I| = b_\emptyset$$

- The **resolution** of the fraction is *at least* the smallest length of the indices I, $I \neq \emptyset$, of the non-null b_I's

Coefficients of the indicator function of $\mathcal{F}_H$ multiplied by 2^9 (=512)

16	$b_\emptyset$	16	b_{1234}	-8	b_{12356}	8	b_{125679}
8	b_{1156}	8	b_{1257}	8	b_{12357}	-8	b_{125689}
8	b_{1157}	8	b_{1258}	8	b_{12368}	8	b_{135789}
8	b_{1158}	8	b_{1267}	8	b_{12369}	8	b_{136789}
8	b_{1168}	8	b_{1268}	-8	b_{12378}	8	b_{145679}
8	b_{1178}	-8	b_{1279}	8	b_{12379}	8	b_{145689}
-8	b_{1189}	8	b_{1289}	8	b_{12456}	-8	b_{145789}
8	b_{1256}	-8	b_{1357}	-8	b_{12458}	8	b_{146789}
8	b_{1258}	8	b_{1358}	-8	b_{12467}	8	b_{235679}
-8	b_{1259}	8	b_{1359}	8	b_{12469}	8	b_{235689}
8	b_{1267}	8	b_{1367}	8	b_{12478}	-8	b_{235789}
8	b_{1278}	-8	b_{1368}	8	b_{12489}	8	b_{236789}
8	b_{1279}	8	b_{1369}	8	b_{13456}	8	b_{245789}
8	b_{1356}	8	b_{1459}	8	b_{13458}	8	b_{246789}
-8	b_{1358}	-8	b_{1468}	-8	b_{13459}	8	b_{345679}
-8	b_{1367}	8	b_{1478}	8	b_{13467}	-8	b_{345689}
8	b_{1369}	8	b_{1489}	8	b_{13478}	8	$b_{1235689}$
8	b_{1378}	8	b_{2359}	8	b_{13479}	8	$b_{1235789}$
8	b_{1389}	-8	b_{2369}	-8	b_{13679}	8	$b_{1245679}$
-8	b_{1456}	8	b_{2379}	8	b_{16789}	8	$b_{1245789}$
8	b_{1457}	8	b_{2389}	8	b_{23456}	8	$b_{1345689}$
8	b_{1468}	-8	b_{2457}	8	b_{23457}	-8	$b_{1346789}$
8	b_{1469}	8	b_{2458}	8	b_{23459}	-8	$b_{2345679}$
-8	b_{1478}	8	b_{2459}	8	b_{23468}	8	$b_{2346789}$
8	b_{1479}	8	b_{2467}	8	b_{23478}	16	$b_{12345678}$
		-8	b_{2468}	-8	b_{23489}		
		8	b_{2469}	8	b_{25689}		
		8	b_{2457}	-8	b_{26789}		
		8	b_{3458}	8	b_{35679}		
		8	b_{3467}	8	b_{35789}		
		8	b_{3468}	8	b_{45689}		
		-8	b_{3479}	8	b_{45789}		
		8	b_{3489}				
		16	b_{5678}				

Note that:
- the smallest order of non-null coefficients is 3
- the coefficients b_{5678}, b_{5678} and $b_{12345678}$ equal to $b_\emptyset$
- the "smallest" (lexicographically) index of non-null coefficients is 156

Statistical properties of the previous fraction found by the indicator function:

- the simple factors are contrast and they are mutually orthogonal (there are no coefficients of order 1 and 2)
- the interaction terms C_{12}, C_{13} and C_{14} are orthogonal w.r.t. the simple terms and they are mutually orthogonal (the "smallest" index of non-null coefficients is 156)
- the fraction is not regular
- $1 = C_{1234} = C_{5678} = C_{12345678}$
  that is: the smallest regular fraction containing it is defined by these relations
- the fraction has resolution *at least* 3

2. $\mathrm{Est}_\sigma(\mathcal{F})$: a basis of the vector space $Q[X_1, \ldots, X_m] \,/\, I(\mathcal{F})$

Let $\mathcal{F}$ be a fraction, $I(\mathcal{F})$ be its ideal, σ be a term ordering and $\mathcal{G}$ be a σ-Gröbner basis of $I(\mathcal{F})$. We denote by $\mathrm{Est}_\sigma(\mathcal{F})$ the set of terms which are not divisible by any of the leading terms of the Gröbner basis $\mathcal{G}$.

$\mathrm{Est}_\sigma(\mathcal{F})$ is a basis of the vector space

$$Q[x_1, \ldots, x_m] \,/\, I(\mathcal{F})$$

$\mathrm{Est}_\sigma(\mathcal{F})$ and statistical properties:

Let $\mathcal{F}$ be a fraction with n points, σ a term ordering and $\mathrm{Est}_\sigma = \{C_I \text{ s.t. } I \in B_\sigma\}$ ($\#B_\sigma = n$).

- For any σ and the corresponding B_σ the **saturated model**

$$Y_A = \sum_{I \in B} \theta_I \, C_I(A)$$

is **identifiable**

- **Maximal resolution**
  Let σ be the **DegRevLex** ordering (DRL).
  Let S the maximal order of interaction terms such that all the terms of that order belong to $\mathrm{Est}_{\mathrm{DRL}}$. Then the fraction has resolution at least $2S+1$.
  Moreover, if the DRL-Normal Form of all the interaction terms of order $S+1$ does not contain any term of lower order, the fraction $\mathcal{F}$ has maximal resolution $2S+2$, otherwise its maximal resolution is $2S+1$.

The fraction $\mathcal{F}_H$ has resolution 3 and not 4

- the terms of order lower or equal to 1 are in Est_{DRL}

$$Est_{DRL} = \{1, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{29}, C_{39}, C_{49}, C_{69}, C_{79}, C_{89}\}$$

- there are some terms of order 2 aliased with terms of lower order

C_{12}	$-C_{79} + C_{89} + \frac{1}{2}C_1 + \frac{1}{2}C_2 - \frac{1}{2}C_3 + \frac{1}{2}C_4$
C_{13}	$2C_{89} + C_{79} + C_{89} + \frac{1}{2}C_1 - \frac{1}{2}C_2 - 3/2C_3 - \frac{1}{2}C_4$
C_{14}	$C_{79} + C_{89} + \frac{1}{2}C_1 - \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{15}	$-\frac{1}{2}C_{29} + \frac{1}{2}C_{39} - \frac{1}{2}C_5 + \frac{1}{2}C_6 + \frac{1}{2}C_7 - \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{16}	$-C_{49} + \frac{1}{2}C_5 + \frac{1}{2}C_6 + \frac{1}{2}C_7 + \frac{1}{2}C_8$
C_{17}	$C_{49} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 + \frac{1}{2}C_8$
C_{18}	$\frac{1}{2}C_{29} - \frac{1}{2}C_{39} + \frac{1}{2}C_5 + \frac{1}{2}C_6 + \frac{1}{2}C_7 + \frac{1}{2}C_8 - \frac{1}{2}C_9$
C_{19}	$C_{29} + C_{39} + C_{49} + C_5 - C_6 - C_7 - C_8$
C_{23}	$C_{79} + C_{89} + \frac{1}{2}C_1 - \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{24}	$2C_{89} + C_{79} + C_{89} + \frac{1}{2}C_1 - \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{25}	$\frac{1}{2}C_{29} + \frac{1}{2}C_{39} + C_{49} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 + \frac{1}{2}C_8 - \frac{1}{2}C_9$
C_{26}	$-C_{29} + \frac{1}{2}C_5 + \frac{1}{2}C_6 + \frac{1}{2}C_7 + \frac{1}{2}C_8$
C_{27}	$-\frac{1}{2}C_{29} - \frac{1}{2}C_{39} - C_{49} - \frac{1}{2}C_5 + 5/4C_6 + \frac{1}{2}C_7 + \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{28}	$C_{39} + \frac{1}{2}C_5 - \frac{1}{2}C_6 + \frac{1}{2}C_7 - \frac{1}{2}C_8$
C_{34}	$-C_{79} + C_{89} + \frac{1}{2}C_1 + \frac{1}{2}C_2 - \frac{1}{2}C_3 + \frac{1}{2}C_4$
C_{35}	$C_{29} + \frac{1}{2}C_5 + \frac{1}{2}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8$
C_{36}	$\frac{1}{2}C_{29} + \frac{1}{2}C_{39} + C_{49} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - 5/4C_7 - \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{37}	$C_{29} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 + \frac{1}{2}C_8$
C_{38}	$\frac{1}{2}C_{29} + \frac{1}{2}C_{39} + C_{49} - \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{45}	$C_{29} + C_{39} + C_{49} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8$
C_{46}	$\frac{1}{2}C_{29} - \frac{1}{2}C_{39} - \frac{1}{2}C_5 + \frac{1}{2}C_6 - \frac{1}{2}C_7 + \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{47}	$\frac{1}{2}C_{29} - \frac{1}{2}C_{39} + \frac{1}{2}C_5 + \frac{1}{2}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8 + \frac{1}{2}C_9$
C_{48}	$C_{29} + C_{39} + C_{49} + \frac{1}{2}C_5 - \frac{1}{2}C_6 - \frac{1}{2}C_7 - \frac{1}{2}C_8$
C_{56}	$\frac{1}{2}C_1 + \frac{1}{2}C_2 + \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{57}	$-C_{89} - C_{79} + \frac{1}{2}C_1 + \frac{1}{2}C_2 + \frac{1}{2}C_3 + \frac{1}{2}C_4$
C_{58}	$C_{69} + C_{89} + \frac{1}{2}C_1 + \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{59}	$C_{69} + C_{79} + C_{89} + C_1 - C_2 - C_3 - C_4$
C_{67}	$C_{69} + C_{89} + \frac{1}{2}C_1 + \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_4$
C_{68}	$-C_{69} - C_{79} + \frac{1}{2}C_1 + \frac{1}{2}C_2 + \frac{1}{2}C_3 + \frac{1}{2}C_4$
C_{78}	$\frac{1}{2}C_1 + \frac{1}{2}C_2 + \frac{1}{2}C_3 - \frac{1}{2}C_4$

Notations and definitions

- $\mathcal{O}_{\mathrm{DRL}}(\mathcal{F})$ the set of mutually orthogonal terms in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F})$
- $\mathcal{O}(\mathcal{F})$ a maximal set of mutually orthogonal terms of $\mathcal{F}$
- we say that $\mathcal{O}(\mathcal{F})$ is a "hierarchic set" (or standard set) if $T \in \mathcal{O}$ and T' divides T then $T' \in \mathcal{O}$ (by its very nature $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F})$ is a hierarchic set)

Questions

- what is $\#\mathcal{O}_{\mathrm{DRL}}(\mathcal{F})$?
- what is $\#\mathcal{O}(\mathcal{F})$?
- when $\#\mathcal{O}(\mathcal{F}) = n$?
  i.e. $\mathcal{O}(\mathcal{F})$ is a basis of $Q[X_1, \ldots, X_m] \,/\, I(\mathcal{F})$?
- when $\mathcal{O}(\mathcal{F})$ is a hierarchic set?
- which relations between previous questions and resolution and indicator function?

Next examples show that *the number of identifiable terms with uncorrelated estimators depend neither on the resolution nor on the number of the non null coefficients b_I but only on the "form" of their indices.*

From the point of view of a general methodology of fractional factorial designs, it is remarkable that the framework derived from the regular case cannot be applied to non regular fractions.

Examples: fractions of a 2^5 design

$\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ and $\mathcal{F}_4$ have 16 points ($b_\emptyset = 1/2$), $\mathcal{F}_5$ has 12 points ($b_\emptyset = 3/8$).

Here σ is DevRevLex.

1. Let $\mathcal{F}_1$ be the *regular fraction* whose indicator function is

$$F_1 = \frac{1}{2} + \frac{1}{2} C_{123}$$

$$\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_1) = \{1, C_1, C_2, C_3, C_4, C_5, C_{14}, C_{15}, C_{24}, C_{25}, C_{34}, C_{35}, C_{45}, C_{145}, C_{245}, C_{345}\}$$

Resolution: it is equal to 3.

Orthogonality: the elements of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_1)$ are all mutually orthogonal.

2. Let $\mathcal{F}_2$ be the fraction whose indicator function is

$$F_2 = \frac{1}{2} + \frac{1}{4} C_{123} + \frac{1}{4} C_{1234} + \frac{1}{4} C_{1235} - \frac{1}{4} C_{12345}$$

$$\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_2) = \{1, C_1, C_2, C_3, C_4, C_5, C_{12}, C_{13}, C_{14}, C_{15}, C_{23}, C_{24}, C_{25}, C_{34}, C_{35}, C_{45}\}$$

Resolution: it is at least 5 (all the terms of order 2 are in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_2)$). The fraction has not resolution 6; in fact there exists a confounding relation between interactions of order 3 and lower terms: $1 + C_4 + C_5 - C_{45} - 2\, C_{123} = 0$.

Orthogonality: a maximal subset of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_2)$ with mutually orthogonal terms has 13 elements:

$$\mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_2) = \{1, C_1, C_2, C_3, C_4, C_5, C_{14}, C_{15}, C_{24}, C_{25}, C_{34}, C_{35}, C_{45}\}$$

A set of 16 mutually orthogonal terms (not restricted to $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_2)$) is:

$$\mathcal{O}(\mathcal{F}_2) = \mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_2) \cup \{C_{145}, C_{245}, C_{345}\}$$

3. Let $\mathcal{F}_3$ be the fraction whose indicator function is

$$F_3 = \frac{1}{2} + \frac{1}{4} C_{123} + \frac{1}{4} C_{124} + \frac{1}{4} C_{125} - \frac{1}{4} C_{12345}$$

We have:

$$\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_3) = \{1, C_1, C_2, C_3, C_4, C_5, C_{12}, C_{13}, C_{14}, C_{15}, C_{23}, C_{24}, C_{25}, C_{34}, C_{35}, C_{45}\}$$

Resolution: it is equal to 5: all the terms of order 2 are in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_3)$ and there exist confounding relations between interactions of order 3 and lower terms.

Orthogonality: a maximal subset of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_3)$ with mutually orthogonal terms has 9 elements:

$$\mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_3) = \{1, C_1, C_2, C_3, C_4, C_5, C_{34}, C_{35}, C_{45}\}$$

A set of 16 mutually orthogonal terms is:

$$\mathcal{O}(\mathcal{F}_3) = \mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_3) \cup \{C_{134}, C_{135}, C_{145}, C_{234}, C_{235}, C_{245}, C_{345}\}$$

4. Let $\mathcal{F}_4$ be the fraction whose indicator function is

$$F_4 = \frac{1}{2} + \frac{1}{4} C_{123} + \frac{1}{4} C_{124} + \frac{1}{4} C_{1235} - \frac{1}{4} C_{1245}$$

We have:

$$\begin{aligned} \mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_4) = \{&1, C_1, C_2, C_3, C_4, C_5, \\ &C_{13}, C_{14}, C_{15}, C_{23}, C_{24}, \\ &C_{25}, C_{34}, C_{35}, C_{45}, C_{345}\} \end{aligned}$$

Resolution: it is equal to 3: all the terms of order 1 are in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_4)$ and there exist confounding relations between interactions of order 2 and simple terms.

Orthogonality a maximal subset of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_4)$ with mutually orthogonal terms has 12 elements:

$$\begin{aligned} \mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_4) = \{&1, C_1, C_2, C_3, C_4, C_5, \\ &C_{15}, C_{25}, C_{34}, C_{35}, C_{45}, C_{345}\} \end{aligned}$$

A set of 16 mutually orthogonal terms is:

$$\mathcal{O}(\mathcal{F}_4) = \mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_4) \cup \{C_{134}, C_{234}, C_{1345}, C_{2345}\}$$

5. Let $\mathcal{F}_5$ be the fraction whose indicator function is

$$\begin{aligned} F_5 = \frac{3}{8} + \frac{1}{8}(&C_{123} - C_{124} + C_{125} + C_{134} \\ &- C_{135} - C_{145} - C_{234} - C_{235} \\ &+ C_{245} + C_{345} + C_{1234} + C_{1235} \\ &+ C_{1245} + C_{1345} + C_{2345}) \end{aligned}$$

We have:

$$\begin{aligned} \mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_5) = \{&1, C_1, C_2, C_3, C_4, C_5, \\ &C_{15}, C_{24}, C_{25}, C_{34}, C_{35}, C_{45}\} \end{aligned}$$

Resolution: it is equal to 3: all the terms of order 1 are in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_4)$ and there exist confounding relations between interactions of order 2 and simple terms.

Orthogonality: a maximal subset of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_5)$ with mutually orthogonal terms has 6 elements:

$$\mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_5) = \{1, C_1, C_2, C_3, C_4, C_5\}$$

and this is a maximal set of orthogonal terms.

Summary of these examples:

	$b_\emptyset$	Indices of b_I $b_I \neq 0$ $I \neq \emptyset$	$\mathrm{Est}_{\mathrm{DRL}}$	resolution	$\perp$ terms in Est $\mathcal{O}_*(\mathcal{F})$	$\perp$ terms total $\mathcal{O}(\mathcal{F})$
$\mathcal{F}_1$	1/2	123	all $\# \leq 1$	3	16	16
$\mathcal{F}_2$	1/2	123, 1234, 1235, 12345	all $\# \leq 2$	5	13	16
$\mathcal{F}_3$	1/2	123, 124, 125, 12345	all $\# \leq 2$	5	9	16
$\mathcal{F}_4$	1/2	123, 124, 1235, 1245	all $\# \leq 1$	3	12	16
$\mathcal{F}_5$	3/8	all $\# = 3, 4$	all $\# \leq 1$	3	6	6

In the first four examples the maximal set of orthogonal terms has cardinality equal to the number of points, i.e. it is a basis of the vector space $Q \,/\, I(\mathcal{F}_i)$.

We notice that the sets $\mathcal{O}(\mathcal{F}_3)$ and $\mathcal{O}(\mathcal{F}_4)$ are not "hierarchic sets", that is not all the terms dividing the terms of $\mathcal{O}$ are in $\mathcal{O}$. In other words there exist no term ordering τ whose Est_τ contains either $\mathcal{O}(\mathcal{F}_3)$ or $\mathcal{O}(\mathcal{F}_4)$.

A last example: a Plackett-Burman fraction of a 2^6 design ($b_\emptyset = 3/16$)

-1	-1	1	-1	-1	-1
-1	1	-1	-1	-1	1
1	-1	-1	-1	1	1
-1	-1	-1	1	1	1
-1	-1	1	1	1	-1
-1	1	1	1	-1	1
1	1	1	-1	1	-1
1	1	-1	1	-1	-1
1	-1	1	-1	-1	1
-1	1	-1	-1	1	-1
1	-1	-1	1	-1	-1
1	1	1	1	1	1

$$\begin{aligned} \mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_{PB}) = \{&1, C_1, C_2, C_3, C_4, C_5, C_6, \\ &C_{26}, C_{36}, C_{45}, C_{46}, C_{23}, C_{56}\} \end{aligned}$$

Resolution: it is equal to 3 (all the terms of order 1 are in $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_{PB})$ and there exist confounding relations between interactions of order 2 and simple terms).

Orthogonality: a maximal subset of $\mathrm{Est}_{\mathrm{DRL}}(\mathcal{F}_{PB})$ with mutually orthogonal terms has 7 elements:

$$\mathcal{O}_{\mathrm{DRL}}(\mathcal{F}_{PB}) = \{1, C_1, C_2, C_3, C_4, C_5, C_6\}$$

and this is a maximal set of orthogonal terms.

An open problem for non regular fractions

Let $\mathcal{G}(\mathcal{F})$ be the subgroup of $(\mathcal{P}, \Delta)$ generated by the set of the non-null coefficients of the indicator function.

- what about $\mathcal{G}(\mathcal{F})$ and its cardinality?
- which relation between $\#\mathcal{G}(\mathcal{F})$ and $\#\mathcal{O}(\mathcal{F})$?

In the previous examples, in the fractions:
- $\mathcal{F}_5$ of a 2^5 design
- $\mathcal{F}_{PB}$ of a 2^6 design

the set $\mathcal{G}(\mathcal{F})$ is $\mathcal{P}$.

These are fractions where:
- $b_\emptyset$ is not of the form 2^{-p}
- $\#\mathcal{O}(\mathcal{F})$ is not n

References on factorial design and Gröbner bases

Association pour la Statistique et ses utilisations – SSDF (1996). *Plans d'expérience, applications à l'entreprise.*

Fontana, R., G. Pistone & M.-P. Rogantin (1997a). Algebraic analysis and generation of two-level designs, *Statistica Applicata* 9(1):15–29.

Fontana, R., G. Pistone & M.-P. Rogantin (1997b). Classification of two-level factorial fractions, Preprint no. 349, Dipartimento di Matematica dell'Università di Genova, accepted *JSPI* (april 1999)

Hall, M. Jr. (1961). *Research Summary*, Jet Propulsion Laboratory, Pasadena, California, (36-10):21–26.

Holliday, T., E. M. Riccomagno, G. Pistone & H. P. Wynn (1997). The Application of Algebraic Geometry to the Design and Analysis of Experiments: a Case Study, *Computational Statistics*, 14(2).

Pistone, G. & H. P. Wynn (1996). Generalized confounding with Gröbner bases, *Biometrika*, 83(1): 653–666.

Raktoe, B. L., Hedayat, A. & Federer, W. T. (1981). *Factorial Designs*, J. Wiley & Son Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Robbiano, L. (1998). Gröbner Bases and Statistics, *Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases)*, Buchberger, B. & Winkler, F. ed., Cambridge University Press, 179–204.

Robbiano, L. & Rogantin, M. P. (1998). Full factorial designs and distracted fractions, *Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases)*, Buchberger, B. & Winkler, F. ed., Cambridge University Press, 473–482.

Sun, X. D. & Wu, C. F. J. (1993). Statistical properties of Hadamard matrices of order 16 *Quality through Engineering Design*, W. Kuo ed. Elsevier, 169-179

Geometric aliasing, Generalized defining relations, and Gröbner basis: A New Look at Multi-level Factorial Designs

Kenny Ye
Department of Applied Math and Statistics
State University of New York at Stony Brook

GROSTAT III
Workshop on Grbner bases and statistics
September 9, 1999

Alias structure of regular two-level fractional factorial design can be described by its generating relations.

A 2^{6-2} fractional factorial design

A	B	C	D	E	F
1	-1	-1	1	-1	1
1	1	-1	-1	1	-1
1	1	1	-1	-1	1
-1	1	1	1	-1	-1
1	-1	1	1	1	-1
-1	1	-1	1	1	1
-1	-1	1	-1	1	1
-1	-1	-1	-1	-1	-1

Generating relation: D=-AB, E=BC, F=ABC

Three level fractional factorial design can also be generated by a set of generating relations. However, those generating relations reveal little about alias structure of the design.

3^{4-2} fractional factorial design

A	B	C	D
0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

Generating relation: $C = A + B(mod 3)$, $D = A + 2B(mod 3)$

Popular mixed-level design such as L18 can not be generated by generating relations and its alias structure are complex and was hardly studied.

The L18 orthogonal design

x	y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	1	1	1	1	1	1	1
1	1	2	2	2	2	2	2
1	1	3	3	3	3	3	3
1	2	1	1	2	2	3	3
1	2	2	2	3	3	1	1
1	2	3	3	1	1	2	2
1	3	1	2	1	3	2	3
1	3	2	3	2	1	3	1
1	3	2	1	3	2	1	2
2	1	1	3	3	2	2	1
2	1	2	1	1	3	3	2
2	1	3	2	2	1	1	3
2	2	2	3	1	2	1	3
2	2	3	1	2	3	2	1
2	3	1	3	2	3	1	2
2	3	2	1	3	1	2	3
2	3	3	2	1	2	3	1

Response surface method: Fit the model with a polynomial of contrasts

Contrasts decomposition of three level Designs:

y_0, y_1, y_2: observation a level 0, 1 and 2.

Linear effect: $y_2 - y_0$
Quadratic Effect: $y_0 + y_2 - 2y_1$

Contrasts matrix of the 3^{4-2} fractional factorial design:

A_l	A_q	B_l	B_q	C_l	C_q	D_l	D_q
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	0	2	0	2	1	-1
-1	-1	1	-1	1	-1	0	2
0	2	-1	-1	0	2	0	2
0	2	0	2	1	-1	-1	-1
0	2	1	-1	-1	-1	1	-1
1	-1	-1	-1	1	-1	1	-1
1	-1	0	2	-1	-1	0	2
1	-1	1	-1	0	2	-1	-1

Two goals

- define the aliasing structure of such designs and response surface method
- find a simple representation for the aliasing structure

Solution: Gröbner basis of Contrast matrix

$A_q^2 + A_q - 2,$
$A_l A_q - A_l,$
$A_l^2 - 1/3A_q - 2/3,$
$B_q^2 + B_q - 2,$
$B_l B_q - B_l,$
$B_l^2 - 1/3B_q - 2/3,$
$C_q + 3/4B_l A_l + 3/4B_l A_q + 3/4B_q A_l - 1/4B_q A_q$
$C_l + 3/4B_l A_l - 1/4B_l A_q - 1/4B_q A_l - 1/4B_q A_q,$
$D_q - 3/4B_l A_l + 3/4B_l A_q - 3/4B_q A_l - 1/4B_q A_q$
$D_l + 3/4B_l A_l + 1/4B_l A_q - 1/4B_q A_l + 1/4B_q A_q$

Predictors $X_1, X_2, \cdots, X_p$ are said to be *geometrically aliased* if they are linearly dependent. Particularly, contrasts associated with an experimental design are said to be geometrically aliased if they are linearly dependent.

All the geometric aliasing relations of the contrasts consists an ideal.

Find a base for generalized defining relations

Theorem: Consider a full factorial design $\mathrm{OA}(n, s_1^1 s_2^1 \cdots s_m^1)$, where $n = \prod_{i=1}^n s_i$, s_i is the number of levels of factor X_i.
Let $\{X_{i,c_1}, X_{i,c_2}, \cdots, X_{i,c_{s_i-1}}\}$ be $s_i - 1$ contrasts of factor X_i and $X_{i,c_0} = 1$. The vectors

$$\{X_{1,c_{\alpha_1}} X_{2,c_{\alpha_2}} \cdots X_{m,c_{\alpha m}}, 0 \le \alpha_i \le s_i - 1\}$$

consists an orthogonal basis in R^n. All but $X_{1,c_0} X_{2,c_0} \cdots X_{m,c_0} = 1$ are contrasts.

Corollary: Let D be a design with m factors. Suppose the first l columns is an orthogonal array with strength l. The contrasts of next $m-l$ factors are the linear combinations of the main contrasts of the first l factors and their interactions.

Are generalized defining relations sufficient for describing the ideal of geometric aliasing?

Answer: yes. Under certain monomial ordering, they consists the Gröbner basis of the contrast matrix.

Define $U = A \otimes B$, where $\otimes$ is the Kronecker product.

$$A = \begin{bmatrix} 1 & \cdots & 1 & 1 & 1 \\ 0 & \cdots & 0 & 0 & -1 \\ 0 & \cdots & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

Example: contrasts of 3^{4-2} fractional factorial design

$$U = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Under this ordering, $C_l D_l \succ A_l A_q B_l D_l$, $C_l C_q D_l \succ A_l A_q C_l D_l$.

Let D be the design as in Corollary and V be its contrast matrix. Let monomial ordering as defined by U. Let G_i be the reduced Gröbner basis of the contrast matrix of factor X_i: $V_i = [X_{i,1}, X_{i,2}, \cdots, X_{i,s_i-1}]$ under Graded reverse lexicographic ordering, for $i = 1, 2, \cdots, l$. Let $g_{ij} = X_{ij} - f_{ij}$, where f_{ij} is the defining relation of X_{ij} for $i = l+1, \cdots, m$. Then

$$G = (\cup_{i=1}^{l} G_i) \cup (\cup_{i=l+1}^{m} \cup_{j=1}^{s_i-1} \{g_{ij}\})$$

is the reduced Gröbner basis of $I(V)$ under the defined monomial ordering.

The reduce Gröbner basis contains two kind of elements. One is the order-2 terms of contrasts of factor X_i as linear combination of main contrasts of X_i for $i \leq l$. The other is the defining relations of main contrasts of $X_i, i \geq l+1$ as linear combinations of main contrasts and interaction contrasts of factors $X_1, \cdots, X_l$. The two kinds polynomials generates the ideal that contains all the geometric aliasing.

Applications #1: model isomorphic of 3^{3-1} fractional factorial design

Two designs, $C = A + B(mod 3)$ and $C = -A - B(mod 3)$ are isomorphic in combinatorial sense. But they are model non-isomorphic. This is recently discovered by examining their design efficiency (S.W. Cheng, 1998). But I found it independently using Gröbner basis method.

Gröbner basis of 3^{3-1} with $C = -A - B(mod 3)$:

$$\begin{aligned}
&A_q^2 + A_q - 2,\\
&A_l A_q - A_l,\\
&A_l^2 - 1/3 A_q - 2/3,\\
&B_q^2 + B_q - 2,\\
&B_l B_q - B_l,\\
&B_l^2 - 1/3 B_q - 2/3,\\
&C_q - 3/2 A_l B_l + 1/2 A_q B_q,\\
&C_l - 1/2 A_l B_q - 1/2 A_q B_l
\end{aligned}$$

Gröbner basis of 3^{3-1} with $C = A + B(mod 3)$:

$A_q^2 + A_q - 2,$
$A_l A_q - A_l,$
$A_l^2 - 1/3A_q - 2/3,$
$B_q^2 + B_q - 2,$
$B_l B_q - B_l,$
$B_l^2 - 1/3B_q - 2/3,$
$C_q + 3/4B_l A_l + 3/4B_l A_q + 3/4B_q A_l - 1/4B_q A_q,$
$C_l + 3/4B_l A_l - 1/4B_l A_q - 1/4B_q A_l - 1/4B_q A_q$

Application #2: L18

An already known but not well known result: Interactions between x and y_1 are not in the defining relations of factors y_3 to y_7, this means that x and y_1 interaction is not confounded with any of the main contrasts.

Gröbner basis of the contrast matrix of L18:

$x - 1,$
$y - y - 2,$
$y\ y + y,$
$y + 1/3y - 2/3,$
$y - y - 2,$
$y\ y + y,$
$y + 1/3y - 2/3,$
$y - 3/8xy\ y + 3/8y\ y + 5/8xy\ y + 3/8y\ y - 1/2xy$
$-3/8xy\ y + 3/8y\ y + 1/8xy\ y - 1/8y\ y + 1/2xy,$
$y - 3/8xy\ y + 3/8y\ y + 1/8xy\ y - 1/8y\ y + 1/2xy$
$+1/8xy\ y - 1/8y\ y - 5/24xy\ y - 1/8y\ y + 1/6xy,$
$y + 3/4xy\ y + 3/4y\ y - 1/2xy\ y - 1/2xy - 1/4xy\ y$
$+1/4y\ y + 1/2xy,$
$y - 1/4xy\ y + 1/4y\ y + 1/2xy - 1/4xy\ y - 1/4y\ y$
$+1/6xy\ y + 1/6xy,$
$y - 3/8xy\ y - 9/8y\ y - 3/8xy\ y + 3/8y\ y - 3/8xy\ y$
$+3/8y\ y - 3/8xy\ y - 1/8y\ y,$
$y - 3/8xy\ y + 3/8y\ y - 3/8xy\ y - 1/8y\ y + 1/8xy\ y$
$+3/8y\ y + 1/8xy\ y - 1/8y\ y,$
$y - 3/4xy\ y - 3/4y\ y - 3/4xy\ y + 1/4y\ y,$
$y - 3/4xy\ y + 1/4y\ y + 1/4xy\ y + 1/4y\ y,$
$y + 3/4xy\ y + 1/4xy\ y + xy + 3/4y\ y + 1/4y\ y,$
$y + 3/4y\ y + 1/4y\ y - 1/4xy\ y - 1/12xy\ y - 1/3xy$

Current and future research:

- classify model isomorphic of 3^{4-2} and larger design
- list of specified aliasing relations
- other mixed-level designs

Cohen/Caprotti: GROSTAT III talk on OpenMath

OpenMath is an ideal ingredient for an integrated environment combining systems for computer algebra and for proof checking.

OpenMath consists of

- a formal grammar for mathematical (so-called) OpenMath objects,
- their encodings (e.g., in XML),
- Content Dictionaries (e.g., for fixing the context),
- Phrasebooks (for interfacing with applications)
- additional files (e.g., for signatures, for defining mathematical properties, for display features), and
- other tools (e.g., LaTeX to OpenMath translators).

The OpenMath standard allows to integrate computational systems of different kind.

In the talk, we showed how OpenMath works and went into future applications.

See

http://www.win.tue.nl/~amc/oz/om/index.html

for further information on OpenMath.

J. Symbolic Computation (1999) 11, 1–000

Computing Ideals of Points †‡

J. ABBOTT*, A. BIGATTI*, M. KREUZER**, AND L. ROBBIANO*

* *Department of Mathematics, University of Genova, Italy*
** *Department of Mathematics, University of Regensburg, Germany*

(Received 14 May 1999)

We address the problem of computing ideals of polynomials which vanish at a finite set of points. In particular we develop a modular Buchberger-Möller algorithm, best suited for the computation over $\mathbf{Q}$, and study its complexity; then we describe a variant for the computation of ideals of projective points, which uses a direct approach and a new stopping criterion. The described algorithms are implemented in CoCoA, and we report some experimental timings.

1. Introduction

The easiest geometric object in affine or projective space is a single rational point. It has no secrets, in particular its defining ideal, i.e. the set of all the polynomials which vanish at the point, is straightforward to describe. Namely, for an affine point with coordinates $(a_1, \ldots, a_n)$, the corresponding ideal is $\mathfrak{p} = (x_1 - a_1, \ldots, x_n - a_n)$; while for a projective point with coordinates $(a_0 : \ldots : a_n)$, the corresponding homogeneous ideal is $\mathfrak{P} = (a_i x_1 - a_1 x_i, \ldots, a_i x_n - a_n x_i)$ for any i such that $a_i \neq 0$. But obviously the matter becomes more complicated if we want the algebraic description of the vanishing ideal of a finite set of points. Let K be a field, and let $\mathbf{A}^n_K$ and $\mathbf{P}^n_K$ denote respectively the n-dimensional affine and projective spaces over K. Then let $\mathbf{X} := \{P_1, \ldots, P_s\}$ be a finite set of s rational points, that is points having coordinates in the field K. Why are we interested in the defining ideal $I(\mathbf{X})$? And how do we compute it?

We answer the first question by listing just a few examples we know of where ideals of points have been used in different fields of mathematics.

1) They encode fundamental properties of algebraic varieties of which they are linear sections (see for instance Geramita *et al.* (1993) and the references indicated there).
2) They are the basis for interpolation problems of numerical analysis (see Moeller (1998)).
3) They are used in coding theory (see Sakata (1998)).
4) They have been recently introduced into the realm of statistics, where finite sets of points are called designs and fractions (see Robbiano (1998)).

† Partially supported by the Consiglio Nazionale delle Ricerche (CNR)
‡ abbott,bigatti,robbiano@dima.unige.it martin.kreuzer@mathematik.uni-regensburg.de

0747-7171/90/000000 + 00 $03.00/0 © 1999 Academic Press Limited

As to the second question, we observe that $I(\mathbf{X}) = \bigcap_{i=1}^{s} I(P_i)$, hence there is an obvious answer: it is possible to use a well-known byproduct of Buchberger's Algorithm to compute the intersection. However, this approach is not efficient from the computational point of view and to remedy this an algorithm (called BM-algorithm) of low complexity was presented in Buchberger and Möller (1982); moreover, this algorithm has good performance (see for instance Marinari *et al.* (1993)). Several generalizations have since been proposed for instance to the cases of points with multiplicity, and of points described by differential conditions. A variant for determining a minimal set of generators in the projective case has been given in Cioffi (1998).

So what is left to do? The increasing demand for strong computational tools, in particular in statistics, pushed us to investigate the problem from several points of view.

First of all, previous complexity analyses referred to a model where the cost of operations in the base field is constant, but experience has shown that this is inappropriate, for instance, when the base field is $\mathbf{Q}$. So the first achievement of this paper is a *modular BM-algorithm*, a variant of the classical one, where we use a modular technique to tame the problem of coefficient growth. Our analysis of complexity given in Section 4 takes into account the varying cost of field operations, and agrees with experimental evidence.

Another important matter we address is that of points in projective space. If $\mathbf{X}$, the set of points, resides in projective space then the task of efficiently computing a Gröbner basis of $I(\mathbf{X})$ is more complicated. Let us see where the difficulties arise. Assume that the polynomial ring is $K[x_0, \ldots, x_n]$ and suppose that the hyperplane $x_0 = 0$ does not meet $\mathbf{X}$. Then it is well-known that we may proceed by considering the affine space given by $x_0 = 1$. We compute a Gröbner basis G of the ideal corresponding to the affine image of $\mathbf{X}$ with respect to some degree compatible term ordering, and then homogenize the elements of G to get a Gröbner basis of $I(\mathbf{X})$ with respect to a `DegRevLex`-type term ordering having x_0 as smallest indeterminate. Now one has to use a variant of the FGLM-algorithm to obtain the Gröbner basis with respect to the desired term ordering.

But what happens if $x_0 = 0$ meets $\mathbf{X}$? The usual suggestion is to perform a suitable linear change of coordinates to achieve that property. Then compute a Gröbner basis as described above of the transformed points, and finally reconstruct the Gröbner basis for $I(\mathbf{X})$. This approach was studied in Marinari *et al.* (1993).

Greater care is needed if the projective space is over a finite field. In this case it is clear that a finite set of points can meet every hyperplane, so that no linear change of coordinates is available to do the trick. A way around this obstacle could be to consider a field extension $K(\alpha)$ of the base field K, work in $K(\alpha)[x_0, \ldots, x_n]$, and then come back to K. This approach is clearly more demanding from the computational point of view.

These considerations suggest looking for a direct approach which avoids any change of coordinates. Indeed in Marinari *et al.* (1993) such a variant of the classical BM-algorithm was suggested. The main difficulty is that, unlike for the affine case, in the projective case ideals of points are one dimensional, so that there is a need of an efficient stopping criterion. Unfortunately the key Lemma 12.1 in their paper is flawed and admits counterexamples (e.g. see Example 3.7). Our Section 3 is entirely devoted to the discussion of the projective case. In particular we introduce a well behaved stopping criterion, which relies on combinatorial tools (see Theorem 3.10).

At this point the reader presumably wants to see some practical achievements of our work. All the algorithms described in this paper have been implemented in CoCoA 3.6

(available from `http://cocoa.dima.unige.it`). Some experimental data based on our implementations are reported in section 5.

In the paper we use repeatedly the notion of Gröbner basis. Needless to say this is now one of the main ingredients in Computational Algebra. The layout of the theory of Gröbner bases can be found in the original papers by Buchberger inserted in the bibliography.

2. A Modular Version of the Classical Buchberger-Möller Algorithm

Let us recall the classical version of the Buchberger-Möller Algorithm, as introduced in the paper Buchberger and Möller (1982). We are given a set of points $\mathbf{X} = \{P_1, \ldots, P_s\}$ in an n-dimensional affine space $\mathbf{A}^n$ over a field K via their coordinates, i.e. we have s tuples $P_i = (p_{i1}, \ldots, p_{in}) \in K^n$. Our goal is to compute the ideal I of all polynomials in $K[x_1, \ldots, x_n]$ which vanish at all points of $\mathbf{X}$.

REMARK 2.1. Here we explain precisely what we mean by "row reduction" in a matrix. The most important aspect is that an ordering must be imposed on the columns of the matrix: the most natural is left-to-right, and by permuting the columns we may assume that this is indeed the chosen ordering. In a matrix M we say that a row whose first non-zero element occurs in column c is a **reducer** for column c; a zero row is not a reducer for any column. In Algorithm C below the matrix M is constructed so that every row is a reducer, and no column has more than one associated reducer. A row vector may then be reduced against M by repeatedly subtracting suitable multiples of the reducers in M to kill the first non-zero element in the vector. The process ends when the vector becomes zero or when there is no reducer in M for the column in which the first non-zero entry lies.

Algorithm C: Classical Buchberger-Möller Algorithm

Let K be a field, let $n \geq 1$, let σ be a term ordering on the power products $\mathbb{T}^n$ of $K[x_1, \ldots, x_n]$, and let $P_i = (p_{i1}, \ldots, p_{in}) \in K^n$ for $i = 1, \ldots, s$. Consider the following sequence of instructions.

C1 *Start with empty lists $G = [\,]$, $Q = [\,]$, $S = [\,]$, a list $L = [1]$, and a matrix $M = (m_{ij})$ over K, with s columns and initially zero rows.*

C2 *If $L = [\,]$, return the pair $[G, Q]$ and stop. Otherwise, choose the power product $t = \min_\sigma(L)$, the smallest according to the ordering σ. Delete t from L.*

C3 *Compute the evaluation vector $(t(P_1), \ldots, t(P_s)) \in K^s$, and reduce it against the rows of M to obtain*

$$(v_1, \ldots, v_s) = (t(P_1), \ldots, t(P_s)) - \sum_i a_i(m_{i1}, \ldots, m_{is}) \qquad a_i \in K$$

C4 *If $(v_1, \ldots, v_s) = (0, \ldots, 0)$ then append the polynomial $t - \sum_i a_i s_i$ to the list G, where s_i is the i^{th} element of S. Remove from L all multiples of t. Continue with step* C2.

C5 *If $(v_1, \ldots, v_s) \neq (0, \ldots, 0)$, then add $(v_1, \ldots, v_s)$ as a new row to M, and $t - \sum_i a_i s_i$ as a new element to S. Append the power product t to Q, and add to L those elements of $\{x_1 t, \ldots, x_n t\}$ which are neither multiples of an element of L nor of $\mathrm{LT}_\sigma(G)$. Continue with step* C2.

THEOREM 2.2. Algorithm C *stops after finitely many steps. It returns the reduced σ-Gröbner basis G of the ideal I of all polynomials vanishing on $\mathbf{X} = \{P_1, \ldots, P_s\}$, and the list of power products Q whose residue classes form a K-basis of $K[x_1, \ldots, x_n]/I$.*

PROOF. See Marinari *et al.* (1993). □

With only a slight change, Algorithm C can also yield the separators f_i of the points P_i, i.e. polynomials such that $f_i(P_i) = 1$ and $f_i(P_j) = 0$ whenever $i \neq j$. To do this replace step C2 by the following:

C2 bis *If $L = [\,]$, reduce M to an identity matrix by row operations mimicking these row operations on the elements of S, and then return the triple $[G, Q, S]$ and stop. Otherwise, choose the power product $t = \min_\sigma(L)$ and delete it from L.*

COROLLARY 2.3. Algorithm C *with step* C2 bis *additionally computes a set of separators for the points P_i; these separators are reduced with respect to the Gröbner basis G.*

PROOF. Observe that throughout Algorithm C the matrix M and list S are related as follows: row i of M is just $(s_i(P_1), s_i(P_2), \ldots, s_i(P_s))$ where s_i is the i^{th} element of the list S. Moreover, the last time step C2 bis is entered M is a non-singular $s \times s$ matrix. The linear combinations of the rows of M yielding the identity matrix clearly also convert the elements of S into separators. □

Conceptually, Algorithm C can be split into two principal tasks: determining the quotient basis Q (and implicitly $LT(G)$), and then computing G from $LT(G)$ by linear algebra.

Algorithm C works well when the field K is finite but is less well-suited to the case $K = \mathbb{Q}$ where growth of the entries of M becomes a problem. For this reason we present a modular version of the algorithm: a commonly used paradigm for avoiding intermediate expression swell. Before describing the algorithm we recall some useful facts.

The Chinese Remainder Algorithm (CRA) converts a collection of residue-modulus pairs $\{r_i \bmod p_i : i = 1, 2, \ldots, e\}$ into a single number R such that $R \equiv r_i \pmod{p_i}$ for all i provided the moduli p_i are pairwise coprime. R is determined uniquely modulo $\prod_1^e p_i$. In particular, if the r_i are known to be the residues of some integer bounded in absolute value by B then the integer value can be determined trivially from R provided $\prod_1^e p_i > 2B$.

A rational number can also be recovered from its image modulo a sufficiently large modulus. There are two ways to do this. If we can compute a multiple of its denominator then we can clear the denominator and find the integral numerator as above. Otherwise, if no (reasonable) multiple of the denominator is known then we can use the rational recovery algorithm of Wang *et al.* (1982) which was later refined in Collins and Encarnación (1995). Applying either of these approaches when the modulus is not big enough will produce some "meaningless" rational number (or a warning that no suitable rational exists).

The question of how many primes to use can be resolved in two ways: determine *a priori* bounds on the numerators and denominators and then deduce how many primes to use, or speculatively recover the rationals after including information from each prime

and check whether the result is correct. The second approach is useful when no good bound is available and the result can easily be checked for correctness; this is the method we use.

REMARK 2.4. The main loop of the algorithm has repeatedly to select a new prime p for the modular computation, but the prime has to be suitable. The conditions p is to satisfy are:

1) p must be different from the primes chosen so far;
2) the coordinates of the points $P_1, \ldots, P_s$ should have reductions modulo p, i.e. the prime p must not divide any of their denominators;
3) the reductions $\overline{P}_1, \ldots, \overline{P}_s \in \mathbb{F}_p^n$ should be pairwise distinct points.

Algorithm M: Modular Version of the Buchberger-Möller Algorithm

Let $n \geq 1$, let σ be a term ordering on $\mathbb{T}^n$, and let $P_i = (p_{i1}, \ldots, p_{in}) \in \mathbb{Q}^n$ for $i = 1, \ldots, s$. Consider the following sequence of instructions.

M1 *Let $Q = [\,]$, $G = [\,]$ and $G_m = [\,]$ be empty lists, and let $m = 1$.*
M2 *Pick a new prime number p (see Remark 2.4).*
M3 *Apply Algorithm C (the classical Buchberger-Möller Algorithm) to $K = \mathbb{F}_p$ and the set of points $\{\overline{P}_1, \ldots, \overline{P}_s\}$, where $\overline{P}_i = (\bar{p}_{i1}, \ldots, \bar{p}_{in}) \in \mathbb{F}_p^n$ is the reduction of P_i. Denote the result by $[G_p, Q_p]$.*
M4 *Compare Q_p with Q (see Remark 2.5). If Q is* better *than Q_p then continue with step* M2. *If Q_p is* better *than Q then replace Q by Q_p, set m equal to p, put G_p in G_m, then continue with step* M2.
M5 *At this point we have $Q_p = Q$. Replace m by mp. Update G_m by combining it with G_p using CRA. Perform rational recovery on G_m with modulus m, and call the result G'.*
M6 *If $G' \neq G$ then replace G by G' and continue with step* M2.
M7 *Now $G' = G$. Check whether all polynomials in G vanish at all points of the set $\{P_1, \ldots, P_s\}$. If they do not, go back to step* M2. *Otherwise return $[G, Q]$ and stop.*

REMARK 2.5. The list Q constructed by Algorithm C is generated in increasing order with respect to σ, and always contains exactly s power products at the end of the algorithm. If we have two such lists $Q = [q_1, q_2, \ldots]$ and $Q' = [q'_1, q'_2, \ldots]$, we say that Q is **better** than Q' if for some index j we have $q_i = q'_i$ for $i = 1, 2, \ldots, j-1$ and $q_j <_\sigma q'_j$. The idea behind this definition is that the only way for the computation modulo p to go wrong is to obtain a zero vector after the reduction in step C3 when the same computation over $K = \mathbb{Q}$ would not have found a zero vector at that point, and in this case the corresponding power product t is missing from Q_p. We justify this idea in Lemma 2.6.

LEMMA 2.6. *Let $P_1, \ldots, P_s \in \mathbb{Q}^n$ be distinct points. Let G be the Gröbner basis produced by applying* Algorithm C *over the field $\mathbb{Q}$ to these points, and let D be the determinant of the matrix M just before the algorithm terminated. Let p be a prime which does not divide any denominator appearing in the points, and let G_p be the Gröbner basis produced by applying* Algorithm C *over the field $\mathbb{F}_p$ to the modular reductions of the points. Then $G \equiv G_p \bmod p$ if and only if $D \not\equiv 0 \bmod p$.*

PROOF. First we claim that p cannot divide the (minimal) denominator of D. Consider the computation over $\mathbb{Q}$. The matrix M built up during Algorithm C is just a triangularization of the matrix E whose rows are the evaluation vectors of the quotient basis; moreover the reduction to triangular form is achieved via a unimodular matrix. So $\det(M) = \pm\det(E)$, and by clearing denominators in the rows of E we see that p does not divide the (minimal) denominator of $\det(E)$, thus proving the claim. Note that in the special case where $P_1, \ldots, P_s \in \mathbb{Z}^n$ we have $D \in \mathbb{Z}$.

Furthermore, by Cramer's rule we see that the solution $\underline{x}$ to any linear system $M\underline{x} = \underline{b}$ has least common denominator not divisibile by p whenever the least common denominator of $\underline{b}$ is not divisibile by p. This implies that p does not divide the least common denominator of any element of G (nor of any separator).

First we show that $LT(G) = LT(G_p)$ if and only if $D \not\equiv 0 \bmod p$. Now $LT(G) = LT(G_p)$ if and only if $Q = Q_p$ where Q are Q_p are the respective quotient bases since Q uniquely determines $LT(G)$ and vice versa. Now, Q and Q_p will differ only if the reduction in step C3 produces a zero vector modulo p where a non-zero vector was obtained over $\mathbb{Q}$, and this implies that modulo p there is a linear dependency amongst the rows of M, i.e. $D \equiv 0 \bmod p$.

Now we show that $LT(G) = LT(G_p)$ if and only if $G \equiv G_p \bmod p$. Since the elements of G and G_p are monic, the reverse implication is trivial. So suppose that $LT(G) = LT(G_p)$. Then necessarily $Q = Q_p$, and G is obtained from $LT(G)$ by representing the evaluation vectors of each element of $LT(G)$ as a linear combination of the rows of E. G_p is obtained analogously. Now we have just seen that each element of G can be reduced modulo p. This gives a representation modulo p of the evaluation vector of some element of $LT(G)$ as a linear combination of the rows of E_p (the matrix whose rows are the evaluation vectors modulo p of the quotient basis), but E_p is invertible modulo p, so the representation is unique, and thus coincides with an element of G_p. □

THEOREM 2.7. Algorithm M *stops after finitely many steps. It returns the reduced σ-Gröbner basis G of the vanishing ideal I of $\mathbf{X} = \{P_1, \ldots, P_s\}$ and a list of power products Q whose residue classes form a K-basis of $K[x_1, \ldots, x_n]/I$.*

PROOF. We shall prove termination and correctness together. Let $[G^*, Q^*]$ denote the result Algorithm C would have produced given the same input, and let M^* be the final value of the matrix used there (but not returned).

In step M3 we can have $Q_p \neq Q^*$ only finitely many times since by Lemma 2.6 it implies that p divides $\det(M^*)$. Furthermore, once a prime has been found for which $Q_p = Q^*$ then in Algorithm M we will have $Q = Q^*$ ever after.

Algorithm M cannot return with $Q^* \neq Q$. For suppose we manage to reach step M7 with $Q^* \neq Q$. Then G contains an element, g, with leading term outside $\mathrm{LT}(G^*)$, hence g itself lies outside the ideal generated by G^*, and consequently the verification will fail. So we will eventually pick a prime p with $Q_p = Q^*$.

Now assume that $Q^* = Q$. Any prime for which $Q_p \neq Q^*$ will be discarded by step M4; so assume also that $Q_p = Q^*$. By Lemma 2.6 we must also have $G \equiv G_p \bmod p$. Hence once we have $Q = Q^*$, then G_m in step M5 will be the reduction modulo m of G, and so for sufficiently large m the rational recovery will yield G, at which point the verification in step M7 will succeed. □

COROLLARY 2.8. Algorithm M *can easily be adapted to compute the separators as well (using the variant of* Algorithm C*).*

PROOF. Note that the final verification in step M7 must check the values at each point of both the Gröbner basis elements and the separators.

Recall that the separators are obtained by representing the vectors $(1,0,0,0,\ldots)$, $(0,1,0,0,\ldots)$, etc., as linear combinations of the evaluation vectors of the quotient basis. Now it is easy to extend the proof of Theorem 2.7. □

3. Computing the Ideal of Points in Projective Spaces

Here we describe the direct approach, as promised in the introduction. In the affine case, the ideal of a set of points is 0-dimensional, and hence the algorithm works on a finite set of power products. In the projective case, the ideal of a set of points is 1-dimensional, so we do not have an "obvious" way to deduce when the computation of the Gröbner basis is completed.

The first step towards a stopping criterion comes from the theory of Hilbert functions of sets of points in $\mathbf{P}^n$. We recall some definitions and results (see Mora and Robbiano (1993) and Geramita *et al.* (1993)).

DEFINITION 3.1. Let $A = K[x_0,\ldots,x_n]$, let I be a homogeneous ideal in A. The **Hilbert function** of A/I is defined as $H_{A/I} : \mathbf{N} \longrightarrow \mathbf{N}$ with $H_{A/I}(d) = \dim_K (A/I)_d$.

THEOREM 3.2. (Hilbert Function)
Let $\mathbf{X} = \{P_1,\ldots,P_s\}$ be a set of K-rational points in $\mathbf{P}^n_K$, and let $I \subseteq A = K[x_0,\ldots,x_n]$ be the homogeneous vanishing ideal of $\mathbf{X}$.

1) *There exists $a \geq -1$ such that $H_{A/I}(d) < H_{A/I}(d+1)$ for all $d \leq a$, and $H_{A/I}(d) = s$ for all $d > a$.*
   In particular, since $H_{A/I}(0) = 1$, he have $H_{A/I}(d) = s$ for all $d \geq s-1$.
2) *For any term ordering σ on A we have $H_{A/I} = H_{A/\mathrm{LT}_\sigma(I)}$.*

DEFINITION 3.3. For $\mathbf{X} = \{P_1,\ldots,P_s\}$ a set of K-rational points in $\mathbf{P}^n_K$, we define $a_\mathbf{X}$ to be the integer a of Theorem 3.2.

In the following proposition we use a term ordering σ of x_0-**DegRevLex** type, meaning a degree compatible term ordering σ such that if $t,t' \in \mathbb{T}^{n+1}$, $\deg(t) = \deg(t')$, x_0 does not divide t, and x_0 divides t', then $t >_\sigma t'$. Consequently, for a term ordering of x_0-**DegRevLex** type, if $x_0 \mid \mathrm{LT}_\sigma(f)$, then $x_0 \mid f$.

PROPOSITION 3.4. (Simple Stopping Criterion)
*Let $\mathbf{X} \subseteq \mathbf{P}^n_K$ be a set of K-rational points with homogeneous vanishing ideal $I \subseteq A = K[x_0,\ldots,x_n]$. Suppose that the hyperplane $H = Z(x_0)$ contains no point of $\mathbf{X}$, and that the term ordering σ is of x_0-**DegRevLex** type. Then the leading term ideal $\mathrm{LT}_\sigma(I)$ is generated in degrees $\leq a_\mathbf{X} + 2$.*

PROOF. First we show that there is an isomorphism $(A/\mathrm{LT}_\sigma(I))_{a_\mathbf{X}+1} \to (A/\mathrm{LT}_\sigma(I))_{a_\mathbf{X}+2}$ defined by the multiplication by x_0. This follows from the fact that the two vector spaces have the same dimension and x_0 is not a zero-divisor modulo $\mathrm{LT}_\sigma(I)$, as we have remarked above. Now the proof of the proposition proceeds in the usual way. If $f \in \mathrm{LT}_\sigma(I)_{a_\mathbf{X}+3}$, we can write $f = \sum_{i=0}^n x_i f_i$ with $f_0,\ldots,f_n \in A_{a_\mathbf{X}+2}$. By what we just saw, we can write $f_i = x_0 g_i + h_i$ with $g_i \in A_{a_\mathbf{X}+1}$ and $h_i \in \mathrm{LT}_\sigma(I)_{a_\mathbf{X}+2}$. Then $f = x_0 \sum_{i=0}^n x_i g_i + \sum_{i=0}^n x_i h_i$ implies $x_0 \sum_{i=0}^n x_i g_i \in \mathrm{LT}_\sigma(I)_{a_\mathbf{X}+3}$, and therefore $\sum_{i=0}^n x_i g_i \in \mathrm{LT}_\sigma(I)_{a_\mathbf{X}+2}$. Altogether, $f \in A_1\,\mathrm{LT}_\sigma(I)_{a_\mathbf{X}+2}$ proves the claim. □

Our next examples show that this stopping criterion fails in general.

EXAMPLE 3.5. Consider the set of 6 points $\mathbf{X} = \{(0,2,5),(0,1,2),(1,3,1),(4,3,4),(2,5,4),(1,4,4)\}$ in $\mathbf{P}^2_\mathbf{Q}$. The values of its Hilbert function are 1, 3, 6, 6, ... Hence $a_\mathbf{X} = 1$, there are separators for all 6 points in degree d for each $d \geq 2$, and I is generated in degrees ≤ 3. However, if $\sigma =$ **DegRevLex**, the leading term ideal of I is $(xyz, xz^2, yz^2, z^3, xy^3)$. Thus there is one Gröbner basis elements in degree $4 > a_\mathbf{X} + 2$. This failure depends on the fact that x_0 divides zero modulo I. Indeed two points are on the line $x_0 = 0$.

EXAMPLE 3.6. Consider the set of 5 points $\mathbf{X} = \{(4,1,1),(5,2,-1),(85,4,1),(455,8,-1),(4369,16,1)\}$ in $\mathbf{P}^2_\mathbf{Q}$. The values of its Hilbert functions are $1,3,5,5,\ldots$. Thus $a_\mathbf{X} = 1$ again, but using $\sigma =$ **DegLex** we conclude from $\mathrm{LT}_\sigma(I) = (x^2, xz^2, xy^2, y^5)$ that there is an element in the reduced σ-Gröbner basis of I of degree 5. This failure depends on the fact that σ is not of x_0-**DegRevLex** type.

The example below shows that during the execution of the algorithm the fact that the Hilbert function attains its maximum in some degree does not imply that the Gröbner basis is complete (or nearly complete). In particular it shows that the key Lemma 12.1 of Marinari *et al.* (1993) is incorrect.

EXAMPLE 3.7. Consider the set of 15 points $\mathbf{X} = \{(n,m,1)\}_{n,m\in\{-1,0,1,2\}} \setminus \{(2,2,1)\}$ in $\mathbf{P}^2_\mathbf{Q}$. The values of its Hilbert function are $1,3,6,10,13,15,15,\ldots$. We have that $H(5) = 15$, there is no generator of LT (I) in degree 5, and the generators of degree < 5 generate an ideal of dimension 1 (projective dimension 0). But there is another generator of the ideal in degree 6. In fact, the ideal I is minimally generated by $\{y^4 - 2y^3z - y^2z^2 + 2yz^3,\ x^4 - 2x^3z - x^2z^2 + 2xz^3,\ x^3y^3 - x^3yz^2 - xy^3z^2 + xyz^4\}$.

So we need some better tool in order to construct an efficient stopping criterion.

DEFINITION 3.8. Let $\mathbb{T}^{n+1}_d$ be the set of power products in $K[x_0,\ldots,x_n]$ of degree d. Two power products t,t' are called **connected in** $Q \subseteq \mathbb{T}^{n+1}_d$ if there is a sequence of power products $t_0,t_1,\ldots,t_r \in Q$ with $t_0 = t$ and $t_r = t'$ such that for $i = 1,\ldots,r$ there exist $\alpha,\beta \in \{0,\ldots,n\}$ such that $t_i = \frac{t_{i-1}\cdot x_\alpha}{x_\beta}$. The intuitive meaning of this notion is that one can pass from one power product in the sequence to the next by replacing one indeterminate by another one.

We can speak of **connected components** in Q which are clearly disjoint. A connected component of a power product t is empty if $t \notin Q$.

For example, in $\mathbf{Q}[x,y,z]$, let $Q := \{x^3, x^2z, xz^2, yz^2, y^3\}$. In Q we have that x^3 is connected to yz^2, a sequence is x^3, x^2z, xz^2, yz^2. The power product y^3 is connected only to itself.

LEMMA 3.9. (Disconnection Lemma)
Let J be a monomial ideal in $A := K[x_0,\ldots,x_n]$ and $Q_\delta := \mathbb{T}^{n+1}_\delta \setminus J_\delta$ for every $\delta \in \mathbf{N}$. Suppose that there exists $d \in \mathbf{N}$ such that:

1) $H_{A/J}(d) = s$, *i.e.* $|Q_d| = s$;
2) for all $i \in \{0, 1, \ldots, n\}$, *every power product in the connected component of* x_i^d *in* Q_d *is divisible by* x_i;
3) $H_{A/J}(\delta) \geq s$ *for all* $\delta \geq d$.

Then $H_{A/J}(\delta) = s$ *for all* $\delta \geq d$.

PROOF. Let C_i be the connected component of x_i^d in Q_d. It is empty only if $x_i^\delta \in J$. By hypothesis (2) these components are distinct, hence disjoint, therefore $\sum_{i=0}^n |C_i| \leq s$.

First we prove that (1) and (2) imply $Q_\delta \subseteq \bigcup_{i=0}^n x_i^{\delta-d} C_i$ for all $\delta \gg d$:

Let J' be the ideal generated by $J_d \cup C_0 \cup \cdots \cup C_n$. Obviously J' contains all x_i^d, hence $J'_\delta = \mathbb{T}_\delta^{n+1}$ for all $\delta \gg d$ (in particular for all $\delta \geq (n+1)d$).

Now we prove that for each i we have $\{x_0, \ldots, x_n\} \cdot C_i \subseteq J \cup x_i C_i$. Namely, if $t \in C_i$, then $t' = \frac{t \cdot x_j}{x_i}$ is either in J or in C_i, being connected to t in Q_d. From this we have $x_j \cdot t = x_i \cdot t' \in J \cup x_i C_i$.

Therefore $J'_{d+1} = J_{d+1} \cup (\bigcup_i \{x_0, \ldots, x_n\} C_i)$ equals $J_{d+1} \cup (\bigcup_i x_i C_i)$, and by induction it easily follows that $J'_\delta = J_\delta \cup (\bigcup_i x_i^{\delta-d} C_i)$ for all $\delta \geq d$. This means that Q_δ, the complement of J_δ, is all contained in $\bigcup_i x_i^{\delta-d} C_i$ for all $\delta \gg d$.

Now we prove that $|Q_\delta| = s$ for all $\delta \geq d$:

Let $\bar{\delta} \gg d$. From what we have proved and hypothesis (3) it follows $\sum_{i=0}^n |C_i| \geq |Q_{\bar{\delta}}| \geq s$ and therefore $\sum_{i=0}^n |C_i| = s$, which implies $Q_d = \bigcup_{i=0}^n C_i$. Thus $J'_d = \mathbb{T}_d^{n+1}$ and then $J'_\delta = \mathbb{T}_\delta^{n+1}$ for all $\delta \geq d$, hence $Q_\delta \subseteq \bigcup_{i=0}^n x_i^{\delta-d} C_i$ for all $\delta \geq d$.

This and hypothesis (3) imply $|Q_\delta| = s$ for all $\delta \geq d$, which is exacly what we wanted to prove. □

THEOREM 3.10. (Stopping criterion)
Let $\mathbf{X}$ *be a set of* K*-rational points in* $\mathbb{P}_K^n$, *let* I *be the homogeneous vanishing ideal of* $\mathbf{X}$, *and let* σ *be a term ordering on* $A = K[x_0, \ldots, x_n]$. *Assume that there exists* d *such that*

1) $H_{A/\mathrm{LT}_\sigma(I)}(d)$ *equals the number of points in* $\mathbf{X}$,
2) for all i, *every power product in the connected component of* x_i^d *in* $\mathbb{T}_d^{n+1} \setminus \mathrm{LT}_\sigma(I)$ *is divisible by* x_i.

Then the elements of the reduced σ*-Gröbner basis of* I *have degrees* $\leq d$.

PROOF. Let s be the number of points in $\mathbf{X}$ and let J be the ideal generated by the leading power products of the elements of the reduced σ-Gröbner basis of I of degree $\leq d$. We want to prove $J = \mathrm{LT}_\sigma(I)$.

Applying Theorem 3.2 to our hypothesis, we have

$$H_{A/J}(\delta) \geq H_{A/\mathrm{LT}_\sigma(I)}(\delta) = H_{A/I}(\delta) = s \quad \text{for all} \quad \delta \geq d$$

Hence, by Lemma 3.9, we have $H_{A/J}(\delta) = s = H_{A/I}(\delta)$ for all $\delta \geq d$ and then $H_{A/J}(\delta) = H_{A/\mathrm{LT}_\sigma(I)}(\delta)$ for all $\delta \in \mathbb{N}$. This and the fact that $J \subseteq \mathrm{LT}_\sigma(I)$ imply $J = \mathrm{LT}_\sigma(I)$. □

Based on this theorem we define the function `StoppingCriterion(Q)` which returns "TRUE" only if Q contains s elements of maximal degree d and the connected components of every x_i^d in Q satisfy condition (2).

Algorithm CP: Projective Version of the Classical Buchberger-Möller Algorithm

Let K *be a field, let* $n \geq 1$, *let* σ *be a term ordering on the monoid* $\mathbb{T}^{n+1}$ *of power products of* $K[x_0, \ldots, x_n]$, *and let* $P_i = (p_{i0} : \ldots : p_{in}) \in \mathbf{P}^n(K)$ *for* $i = 1, \ldots, s$. *Consider the following sequence of instructions.*

CP1 *Start with empty lists* $G = [\,]$, $Q = [\,]$, $S = [\,]$, *a list* $L = [1]$, *a matrix* $M = (m_{ij})$ *over* K *with* s *columns and initially zero rows, and with* $d = 0$.
CP2 *If the function* `StoppingCriterion(Q)` *defined above yields TRUE then return the pair* $[G, Q]$ *and stop. Otherwise increase* d *by one, reset* M *to be a matrix over* K *with zero rows and* s *columns, reset* $Q = [\,]$, *and let* L *be the list of all power products of degree* d *which are not multiples of a leading power product (w.r.t.* σ*) of an element of* G.
CP3 *If* L *is empty, go to step* CP2; *otherwise choose the power product* $t = \min_\sigma(L)$ *and remove it from* L.
CP4 *Compute the evaluation vector* $(t(P_1), \ldots, t(P_s)) \in K^s$, *and reduce it against the rows of* M, *to obtain*

$$(v_1, \ldots, v_s) = (t(P_1), \ldots, t(P_s)) - \sum_i a_i (m_{i1}, \ldots, m_{is}) \qquad a_i \in K$$

CP5 *If* $(v_1, \ldots, v_s) = (0, \ldots, 0)$ *then append the polynomial* $t - \sum_i a_i s_i$ *to the list* G, *where* s_i *is the* i^{th} *element of the list* S. *Continue with step* CP3.
CP6 *If* $(v_1, \ldots, v_s) \neq (0, \ldots, 0)$ *then add* $(v_1, \ldots, v_s)$ *as a new row to* M *and* $t - \sum_i a_i s_i$ *as a new element to* S. *Append the power product* t *to* Q. *Continue with step* CP3.

REMARK 3.11. The value Q returned by Algorithm CP is normally of little utility; we return it only to simplify the enunciation of Algorithm MP (below).

THEOREM 3.12. Algorithm CP *stops after finitely many steps and returns the reduced* σ*-Gröbner basis* G *of the homogeneous vanishing ideal* I *of* $\mathbf{X} = \{P_1, \ldots, P_s\}$.

PROOF. First we show finiteness of the algorithm: in particular we prove that the algorithm never goes beyond degree s. From the construction it is clear that the number of power products in the list Q at the end of processing some degree d is just $H_{A/I}(d)$. From Theorem 3.2.1 we have that $H_{A/I}(d) = s$ for all degrees $d \geq s - 1$. Let us consider the case $d = s$. We have seen that there are precisely s power products of degree s in Q, that is condition (1) of the stopping criterion. Also condition (2) is satisfied: If a power product t is connected to x_i^s in Q, we have $x_i \mid t$, since otherwise any connection between x_i^s and t would pass through at least one power product exactly divided by x_i^j for every $j = 0, \ldots s$, hence through at least $s + 1$ power products in Q. Hence the Stopping Criterion applies.

To show correctness we prove that $\mathrm{LT}_\sigma(I)$ is generated by $\mathrm{LT}_\sigma(G)$. Consider the end of step CP2. The new list L contains all power products of degree d which are not in the ideal generated by $\mathrm{LT}_\sigma(G)$. Since by construction all elements of G vanish on $\mathbf{X}$, we have $G \subseteq I$, hence $\mathrm{LT}_\sigma(G) \subseteq \mathrm{LT}_\sigma(I)$. We have already remarked that the number of power products in the list Q at the end of processing some degree d is just $H_{A/I}(d)$. Therefore $H_{A/(\mathrm{LT}_\sigma(G))} = H_{A/I}$. This implies that $H_{A/(\mathrm{LT}_\sigma(G))} = H_{A/\mathrm{LT}_\sigma(I)}$, which, together with $\mathrm{LT}_\sigma(G) \subseteq \mathrm{LT}_\sigma(I)$, yields the desired conclusion. □

The following example shows that the stopping criterion is not always optimal.

EXAMPLE 3.13. Consider the set of points $\mathbf{X} = \{(1,1,1),(0,1,1),(0,1,0),(1,1,0)\}$ in $\mathbf{P}^2_{\mathbf{Q}}$. The leading term ideal w.r.t. $\sigma = \texttt{DegRevLex}$ of its homogeneous vanishing ideal I is $\mathrm{LT}_\sigma(I) = (yz, x^2)$. In degree 2, the stopping criterion is not satisfied, because z^2 is connected to xy via $[z^2, xz, xy]$. In degree 3, the stopping criterion is satisfied, but there is no new σ-Gröbner basis element in that degree.

DEFINITION 3.14. Let $A = K[x_0, \ldots, x_n]$, let I be a homogeneous ideal in A, and let $\mathbf{X} = \{P_1, \ldots, P_s\} \subseteq \mathbf{P}^n(K)$ be a set of K-rational points. A homogeneous polynomial $f \in A$ is called a **separator** of P_i, if $f(P_i) \neq 0$ and $f(P_j) = 0$ for $j \neq i$. A separator of minimal degree is also called a **minimal separator**.

REMARK 3.15. As earlier for Algorithm C we may easily adapt Algorithm CP to compute separators. The initial value for d in step CP1 should be -1 to handle correctly the case of just a single input point. In step CP6 we must additionally use the newly added row to bring M into reduced row echelon form. Then in step CP2 once the stopping criterion has been fulfilled we obtain the separators as a subset of the list S: if row i in M contains exactly one non-zero entry, in column j say, then s_i is a separator for the point P_j; if there are several separators for the same point, we take the one of lowest degree. In this way we get minimal separators.

Algorithm MP: Modular Version of the Projective Buchberger-Möller Algorithm

This is identical to Algorithm M *except that it calls* Algorithm CP *in step* MP3 *where* Algorithm C *was called in step* M3.

THEOREM 3.16. Algorithm MP *stops after finitely many steps. It returns the reduced σ-Gröbner basis G of the vanishing ideal I of $\mathbf{X} = \{P_1, \ldots, P_s\}$.*

PROOF. This is entirely analogous to the proof of Theorem 2.7. □

COROLLARY 3.17. Algorithm MP *can be adapted to compute the separators as well (using the variant of* Algorithm CP *described in Remark 3.15).*

PROOF. The adaptation is straightforward, as for Algorithm M, except for two points which do not arise in the affine case. The projective separators are defined only up to multiplication by a non-zero field element. So that the separators can be recovered from their modular images we must eliminate this ambiguity: we do this by scaling the separators so that they evaluate to 1 at their own points.

The other matter is that of recognising "bad" separators. It is possible that the minimal separator modulo p of some point has lower degree than the minimal separator over $\mathbf{Q}$. In such a case the "false modular image" must be discarded as it would otherwise inhibit recovery of the separators over $\mathbf{Q}$. One wasteful approach is to compare the degrees of the separators for the "current prime" with the degrees of the separators built up by Chinese Remaindering: if some separator modulo p has degree lower than the corresponding lifted separator then discard all information from that prime; otherwise if some separator modulo p has degree higher than the corresponding lifted separator then discard all lifted values and start anew from the current prime. Other less wasteful approaches can be used instead. □

4. Complexity Issues

In this section we determine a worst case complexity for Algorithm M to compute both the separators of the points and the Gröbner basis for their ideal. We also determine the complexity for "generic" points with the term ordering DegRevLex; by "generic" we mean random points whose behaviour is as that of true generic points.

As in the paper Marinari *et al.* (1993) the parameters we shall use to express the complexity are:

- s the number of points;
- n the number of variables (i.e. the dimension of the ambient affine space);
- g the number of elements in the Gröbner basis.

Our Algorithm C is the same as algorithm 1 presented in section 5 of that paper. We recall the result of their analysis which showed that the complexity of applying Algorithm C with $K = \mathbf{F}_p$ is $O(s^2(g+s)(\log p)^2 + s^2n^2)$ where the cost of an arithmetic operation in $\mathbf{F}_p$ is $O((\log p)^2)$. Our result is encapsulated in the following theorem.

THEOREM 4.1. *For the case $K = \mathbf{Q}$, where all points have integral coordinates, and assuming that the largest coordinate X satisfies $\log(X) \in O(\log s)$, i.e. X is bounded by some polynomial in s, then* Algorithm M *has worst case bit complexity*

$$O(s^4(g+s)(\log s)^2 + s^4n^2)$$

Furthermore, for "generic" points with the DegRevLex *ordering,* Algorithm M *has bit complexity*

$$O(s^2d(g+s)(\log s)^2 + s^2dn^2)$$

where d is such that $\binom{n+d-1}{d} = s + g$; this is better than the worst case complexity by a factor of s/d.

PROOF. We start by estimating how many bad primes we could encounter. Consider the $s \times s$ matrix, M, whose columns are indexed by the points in $\mathbf{X}$ and whose rows are indexed by the power products in the basis of the quotient R/I; the entries being the value of the power product at the point. By definition this matrix is non-singular; moreover, by Lemma 2.6 a prime is "bad" if and only if it divides $\det(M)$. Now we estimate the greatest possible value for this determinant using Hadamard's bound on the rows. In any row corresponding to a power product of total degree d no entry can exceed X^d in absolute value. Using $d_1, d_2, \ldots, d_s$ to denote the degrees of the power products for the rows in our matrix, we see that Hadamard's bound does not exceed $X^D\sqrt{s^s}$ where $D = d_1 + \cdots + d_s$. Since our quotient basis contains every factor of every element of the basis, we deduce that $D \leq 1 + 2 + \cdots + s = s(s+1)/2$. This worst case is realised with generic points and the Lex ordering; in contrast, generic points with the DegRevLex ordering will typically give a much smaller exponent of $D = sd$ where d is the highest degree of a Gröbner basis element. Here we note that the logarithm of Hadamard's bound in the worst case is: $O(s^2 \log X + s \log s)$.

For any positive integer r the product of the first r primes clearly exceeds $r!$, and this we can estimate using Stirling's asymptotic approximation: $\log(r!) = r \log r - r + o(r)$.

Whence we see that the number of bad primes is $O(s^2)$; in practice it appears that bad primes are typically fairly uncommon.

Next we find out how many good primes we need to be able to reconstruct the answer. To do this we must estimate the sizes of every coefficient of every separator and Gröbner basis element. Again we use the matrix M from above, and shall apply Cramer's rule for obtaining the solutions to a linear system since our coefficients effectively arise from solving linear systems. It is immediate that the determinant of M is a common denominator for both separators and Gröbner basis elements. Hadamard's bound applied to the numerators given by Cramer's rule shows that no numerator exceeds $X^{s+1}\det(M)$. From this we see that the logarithm of the product of the good primes may have to become as large as $O(s^2 \log X + s \log s)$ in the worst case. Thus the total number of primes, both good and bad, which could be necessary is $O(s^2)$ — experiments suggest that random points in $\mathbf{A}^2$ with coordinates in the range $[-2s^2, 2s^2]$ achieve this complexity. The situation is far more favourable for generic points with **DegRevLex** ordering: the same reasoning shows that the total number of primes needed is only $O(sd)$.

So we call Algorithm C up to $O(s^2)$ times at a total cost of $O(s^4(g+s)(\log s)^2 + s^4n^2)$ where the factor of $(\log s)^2$ allows for the fact that an arithmetic operation modulo an s-bit prime costs $O((\log s)^2)$. Following the same line of reasoning we can also deduce that for generic points using **DegRevLex** ordering the total cost of the calls to Algorithm C is $O(s^2d(g+s)(\log s)^2 + s^2dn^2)$.

We comment that the cost of the calls to Algorithm C is at least that of of Chinese Remaindering and verification combined. Combining $O(s^2)$ modular values (each of size $O(\log s)$) into a single value using iterative Chinese Remaindering costs $O(s^3(\log s)^2)$, and we must do this for $s(s+g)$ coefficients, giving a total cost of $O(s^4(g+s)(\log s)^2)$. Verifying the answer by evaluation costs no more than $O(s^4g(\log s)^2)$: the s power products can each be evaluated at the s points once only at a total cost of $O(s^4 \log s)$, and then each polynomial in the Gröbner basis has no more than $s+1$ terms and can be evaluated in time $O(s^4(\log s)^2)$. It can similarly be shown that, in the case of generic points with the **DegRevLex** ordering, the overall complexity is the same as that of the calls to Algorithm C. □

Now we determine the complexity for Algorithm MP (for the field $\mathbb{Q}$). We continue to use the same parameters but observe that now n is one greater than the dimension of the ambient projective space. Following the reasoning in Marinari *et al.* (1993) we find the worst case complexity of Algorithm CP running over the field $K = \mathbf{F}_p$ to be

$$O(s(n+s)(g+s^2)(\log s)^2 + s^2n^2\log(ns))$$

where again we have assumed that each arithmetic operation in $\mathbf{F}_p$ costs $O((\log p)^2)$. Superficially this formula appears different from the one they published. Our result for Algorithm MP is encapsulated in the following theorem.

THEOREM 4.2. *For the case $K = \mathbb{Q}$, where all points have integral coordinates, and assuming that the largest coordinate X satisfies $\log(X) \in O(\log s)$, i.e. X is bounded by some polynomial in s, then* Algorithm MP *has worst case bit complexity*

$$O(s^4(n+s)(g+s^2)(\log s)^2 + s^5n^2\log(ns))$$

Furthermore, for "generic" points with the **DegRevLex** *ordering,* Algorithm MP *has bit*

complexity

$$O(s^2d(n+s)(g+sd)(\log s)^2 + s^3dn^2\log(ns))$$

where d is such that $\binom{n+d-1}{d} = s+g$; this is better than the worst case complexity by a factor of about s^2/d.

PROOF. This proof is largely analogous to that of Theorem 4.1. In place of the single matrix M considered in that proof we must consider several matrices $M_1, M_2, \ldots M_d$ where the rows of M_r are indexed by those power products of degree r..... Since we never need to go beyond degree s we have that $d \leq s$. Each of these matrices is, by definition, of full rank, and a prime is "bad" if and only if at least one M_r is no longer of full rank modulo p; i.e. p divides the determinant of every maximal minor of at least one M_r. Hadamard's bound for the determinant of a maximal minor of M_r gives $X^{r^2}\sqrt{r^r}$. Thus all bad primes must divide a number of size not exceeding $\log\prod_{r=1}^{s}(X^{r^2}\sqrt{r^r}) \in O(s^3\log X + s^2\log s)$. Whence there can be at most s^3 bad primes.

The maximum number of "good" primes we could ever need remains $O(s^2)$ since we never need to go beyond degree s: we have to consider solving linear systems corresponding to the matrices $M_1, M_2, \ldots M_d$, and arguing as previously Cramer's Rule and Hadamard's bound prove that the logarithm of the product of the good primes need never exceed $O(s^2 \log X + s \log s)$ in the worst case.

The argument that the cost of the main loop dominates remains valid for this case. Since we may have to execute the main loop up to $O(s^3)$ times the overall complexity is just s^3 times the complexity of the call to Algorithm CP.

We note that the case of generic points with the **DegRevLex** ordering permits a much more favourable analysis. All elements of the Gröbner basis have degree $d-1$ or d (where $d \in O(\log(s)/\log(n))$). We shall show that a prime p is "bad" if and only if at least one of M_{d-2}, M_{d-1} or M_d is not of full rank modulo p; as in the proof of Theorem 4.1 we can now deduce that the total number of primes to be tried is $O(sd)$, and thus the complexity in this case is just sd times the complexity of a call to Algorithm CP, i.e. $O(s(n+s)(g+s^2)(\log s)^2 + s^2n^2\log(ns))$.

Now we prove that to characterize "bad" primes it suffices to consider only the three matrices M_{d-2}, M_{d-1} and M_d. Let p be a "bad" prime for which the matrix M_r does not have full rank for some $r < d-2$. This implies that there is a polynomial f of degree r which vanishes (modulo p) at every point of $\mathbf{X}$. Now consider the polynomial $x^{d-r-2}f$; this has degree $d-2$ and clearly vanishes (modulo p) at every point of $\mathbf{X}$. But the rows of M_{d-2} are indexed by all power products of degree $d-2$, hence there is a non-trivial linear combination of the rows of M_{d-2}; i.e. the rank of M_{d-2} is not full. □

REMARK 4.3. Our proof suggests that we could waste a large proportion of computation time encountering "bad" primes. In numerous actual computations we have never seen an example where the number of "bad" primes encountered exceeded the number of "good" ones, but with the **DegLex** ordering we have observed as many as 20% of all primes being "bad". The matrices M_i in the proof are far from independent, and it could be that our estimate of the number of "bad" primes is too crude.

5. Implementation Issues and Timings

Algorithm M as described admits a number of implementation refinements. For instance, at step M4 the comparison of the lists Q and Q_p can be effected inside Algorithm C as the list is generated; Algorithm C can also terminate prematurely if a worse Q is being generated.

The rational recovery and stability checks at steps M5 and M6 need only monitor a single coefficient most of the time, though eventually the full answer will be needed. It also is possible to try both rational recovery techniques until one of them yields a stable answer; this is what our implementation does.

The verification in step M7 need not actually evaluate any polynomials: we represent a polynomial with rational coefficients as f/D where D is the least common denominator, and f has integer coefficients; we know the polynomials are correct modulo m, so it suffices to verify that $|f(P_i)| < m$ for each point P_i, and this can be done quickly and easily (e.g. by bounding each term separately and summing these). This cheap verification test may require $O(s)$ further iterations of the main loop, after the correct answer has been found, before it is able to succeed; in practice it works well.

These refinements remain applicable even when the separators are computed.

The following tables give the timings for the computation of the ideal of distinct random points, computed on a Digital Alpha, 192Mb RAM, 433 Mhz, compiled with `gcc -O2`.

DegRevLex	$\mathbf{A}^{10}$	$\mathbf{A}^{20}$	$\mathbf{A}^{40}$	$\mathbf{P}^{9}$	$\mathbf{P}^{19}$	$\mathbf{P}^{39}$	$\mathbf{P}^{39}$
40 small	0.3 s	0.7 s	3.0 s	0.4 s	0.8 s	3.1 s	14 digits
40 medium	1.7 s	2.8 s	8.1 s	2.9 s	5.2 s	12.0 s	45 digits
40 large	3.2 s	5.7 s	14.0 s	7.7 s	12.0 s	23.0 s	123 digits
40 extralarge	7.0 s	12.9 s	26.0 s	11.8 s	22.0 s	47.0 s	264 digits
80 small	3,7 s	5.6 s	14.0 s	2.7 s	5.6 s	17.0 s	39 digits
80 medium	22.0 s	24.0 s	66.0 s	42.0 s	46.0 s	96.0 s	157 digits
80 large	53.0 s	48.0 s	109.0 s	89.0 s	120.0 s	180.0 s	349 digits
80 extralarge	116.0 s	110.0 s	250.0 s	196.0 s	180.0 s	360.0 s	640 digits

Lex	$\mathbf{A}^{10}$	$\mathbf{A}^{20}$	$\mathbf{A}^{40}$
40 small	0.1 s	0.2 s	0.3 s
40 medium	0.4 s	0.6 s	0.9 s
40 large	3.3 s	4.0 s	5.4 s
40 extralarge	29.7 s	38.2 s	59.0 s
80 small	0.5 s	0.6 s	0.9 s
80 medium	2.5 s	2.6 s	4.3 s
80 large	24.0 s	27.0 s	36.0 s
80 extralarge	490.0 s	590.0 s	720.0 s

The first column indicates the number and the size of the points: *small* means that the coordinates are integers between -1 and 1, *medium* are between -9 and 9, *large* between -99 and 99, and *extralarge* between -9999 and 9999. The other columns give the dimension and type of space (affine or projective).

The last column gives the average size of the coefficients in the basis of points in $\mathbf{P}^{39}_{\mathbb{Q}}$. This shows very clearly how the timings over $\mathbb{Q}$ depend on the size of the coordinates.

All these timings are given for computations over $\mathbb{Q}$. The computation of these examples over finite fields is performed by comparison in negligible time; for example the ideal of 80 points in $\mathbf{P}^{39}$ wrt `DegRevLex` over $\mathbf{Z}/(32003)$ is computed in 2.8 seconds.

Note the apparently surprising behaviour when comparing the orderings `DegRevLex` and `Lex` on the smaller examples. This is because we used random points from a relatively small range rather than generic points, so the `Lex` basis comes out "smaller than normal" (smaller coefficients and lower maximal degree of the leading terms of the basis). The "extralarge" rows show the expected behaviour.

References

Buchberger, B. (1965). *An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal.* PhD thesis, University of Innsbruck (Austria). (in German).

Buchberger, B. (1970). An algorithmical criterion for the solvability of algebraic systems of equations. *Aequationes Math.*, 4(3):374–383.

Buchberger, B. (1985). *Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory*, chapter 6, pages 184–232. N.K. Bose Ed., D. Reidel Publishing Co.

Buchberger, B., Möller, H. M. (1982). The construction of multivariate polynomials with preassigned zeros. In Calmet, J., editor, *Proceedings of the European Computer Algebra Conference (EUROCAM '82)*, volume 144 of *Lecture Notes in Comp.Sci.*, pages 24–31, Marseille, France. Springer.

Capani, A., Niesi, G., Robbiano, L. (1998). *CoCoA, a system for doing Computations in Commutative Algebra.* Available via anonymous ftp from `cocoa.dima.unige.it`, 3.6 edition.

Ciolli, F. (1998) Minimally generating ideals of points in polynomial time using linear algebra. Ricerche di Matematica, XLVII fasc. 2 (1998)

Collins, G. E., Encarnación, M. J. (1995). Efficient rational number reconstruction. *J. Symb. Comput.*, 20(3):287–298.

Geramita, A. V., Kreuzer, M., Robbiano, L. (1993). Cayley-Bacharach schemes and their canonical modules. *Trans. Amer. Math. Soc.*, 339:163–189.

Lakshman, Y. (1991). A single exponential bound on the complexity of computing Gröbner bases of zero dimensional ideals. In Mora, T., Traverso, C., editors, *MEGA-90 (Castiglioncello/Italy)*, volume 94 of *Progress in Math.*, pages 227–234, Boston. Birkhäuser.

Marinari, M., Moeller, H., Mora, T. (1993). Groebner bases of ideals defined by functionals with an application to ideals of projective points. *Appl. Alg. in Eng., Comm. and Comp.*, 4(2):103–145.

Mora, T., Robbiano, L. (1993). Points in affine and projective spaces. In Eisenbud, D., Robbiano, L., editors, *Computational Algebraic Geometry and Commutative Algebra, Cortona-91*, volume XXXIV of *Symposia Mathematica*, pages 106–150. Cambridge University Press.

Möller, M. (1998). Gröbner Bases and Numerical Analysis. In B.Buchberger, F.Winkler, editors, *Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases)*, volume 251 of *London Mathematical Society Lecture Notes Series*, pages 159–178. Cambridge University Press.

Robbiano, L. (1998). Gröbner Bases and Statistic. In B.Buchberger, F.Winkler, editors, *Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases)*, volume 251 of *London Mathematical Society Lecture Notes Series*, pages 179–204. Cambridge University Press.

Sakata, S. (1998). Gröbner Bases and Coding Theory. In B.Buchberger, F.Winkler, editors, *Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases)*, volume 251 of *London Mathematical Society Lecture Notes Series*, pages 205–220. Cambridge University Press.

Wang, P. S., Guy, M. J. T., Davenport, J. H. (1982). p-adic reconstruction of rational numbers. *SIGSAM Bulletin (ACM Special Interest Group on Symbolic and Algebraic Manipulation)*, 16(2):2–3.

Conditional Distributions on Integer Data and Gröbner Bases

I. H. Dinwoodie
September 1999, Grostat III

Let S $(= S_{A,a})$ be the set of nonnegative integer lattice points of the form

$$S := \{\mathbf{x} \in Z_+^d : A\mathbf{x} = \mathbf{a}\}$$

for a nonnegative integer vector $\mathbf{a} \in Z_+^r$ and a nonnegative $r \times d$ integer matrix A. We assume that S is finite.

We are interested in computing expectations of functions on S with respect to a probability distribution μ:

$$E_\mu(f) := \sum_{\mathbf{x} \in S} f(\mathbf{x})\, \mu(\mathbf{x}).$$

Brute force enumeration is usually not easy or not possible.

There are at least two approaches that use Gröbner basis methods:

1. **Markov Monte Carlo**
   Construct an irreducible Markov chain on S. A Gröbner basis is used for the irreducibility. The Metropolis algorithm can be used to get an arbitrary stationary distribution μ (Diaconis and Sturmfels 1998).

2. **Multivariate Generating Functions**
   Manipulate the multivariate exponential generating function g (in d indeterminates), using technology of commutative algebra whenever possible:

$$g := \sum_{\mathbf{x} \in S} \frac{\zeta_1^{x_1} \dots \zeta_d^{x_d}}{x_1! \dots x_d!}.$$

   The normal form $\bar{g}^F$ of g is useful in tests of parameter significance, particularly when g is a product.

Markov Monte Carlo

Let $\{\zeta^{f_1^+} - \zeta^{f_1^-}, \ldots, \zeta^{f_m^+} - \zeta^{f_m^-}\}$ be a Gröbner basis for the ideal

$$I_A := < \zeta^{\mathbf{x}} - \zeta^{\mathbf{y}} : A\mathbf{x} = A\mathbf{y}, \mathbf{x}, \mathbf{y} \in Z_+^d > .$$

(The reduced Gröbner basis will be of this form, where $f^{\pm} = \max\{\pm f, 0\}$. Any indeterminate order and any term order is OK, but the resulting Markov chains can have different properties.)

Then run a symmetric and irreducible Markov chain in S as follows: from state $\mathbf{x} \in S$ choose uniformly from the $2m$ vectors $\{\pm f_1, \ldots, \pm f_m\}$ and add the result to $\mathbf{x}$ if the result is nonnegative. Otherwise hold fast. The stationary distribution is uniform, since the transition matrix K is symmetric. (The construction is in Diaconis and Sturmfels 1998.)

To get an arbitrary positive distribution μ, one can use the Metropolis algorithm with the symmetric kernel K from above:

$$K_\mu(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) \min\{1, \mu(y)/\mu(x)\}.$$

In words, try to make a move with the symmetric chain. If the possible new state is $\mathbf{y} \in S$, check the value of $\mu(\mathbf{y})/\mu(\mathbf{x})$. If the ratio is at least 1, go ahead and do the move. If the ratio is less than 1, use its value in a coin flip to decide whether or not to make the move.

- Fact: $\mu K_\mu = \mu$, and in fact K_μ is reversible for μ.

Then the expectation is the time average over the resulting chain $x_0 = X_0, X_1, X_2, \ldots$:

$$E_\mu(f) = \lim \frac{1}{n} \sum_{i=0}^{n-1} f(X_i).$$

Example 1: Testing for Independence of Factors

Consider the following multinomial data on hair color and eye color each at four levels:

	Black	Brunette	Red	Blonde
Brown	68 (40.14)	119 (106.28)	26 (26.39)	7 (47.20)
Blue	20 (39.22)	84 (103.87)	17 (25.79)	94 (46.12)
Hazel	15 (16.97)	54 (44.93)	14 (11.15)	10 (19.95)
Green	5 (11.68)	29 (30.92)	14 (7.68)	16 (13.73)

( ) contains the fitted values $\hat{x}_{ij}$

The χ^2-test for independence of factors is a goodness-of-fit test of whether the multinomial distribution is in a 9-dimensional subfamily of distributions. The exact (conditional) p-value is the probability (expectation of an indicator function) given by

$$p - \texttt{value} = \mu_h\{(x_{ij}) \in S : \sum_{i,j=1}^{4} \frac{(x_{ij} - \hat{x}_{ij})^2}{\hat{x}_{ij}} \geq 138.2898\}$$

where S is the set of tables with fixed row and columns sums equal to those of the actual data. Here $d = 16$, and $r = 4 + 4$, and A would be an 8×16 matrix that computes row and column sums. The notation μ_h indicates the hypergeometric distribution on S, the set of tables with the same row and column sums as the one above:

$$\mu_h(\mathbf{x}) := \frac{1}{z} \frac{1}{\mathbf{x}!}.$$

This is the conditional distribution of any distribution in the null family, given the row and column sums. The number of tables with these row and column sums is around 10^{15}.

The p-value is approximately 0.0. The traditional normal approximation uses degrees of freedom of 16-1 - (dimension of parameter space) = 15 - 6 in the χ^2 distribution, and is very accurate for this example. The Markov chain for the computation has basic increments described by matrices $\begin{smallmatrix} + & - \\ - & + \end{smallmatrix}$, corresponding to monomial differences, which clearly preserve row and column sums.

High-dimensional sparse tables make exact methodology necessary, and the Markov chain is only one of several possible methods. Such tables appear in tests of genetic equilibrium such as Hardy-Weinberg equilibrium (see Lange and Lazzeroni 1997).

The general goodness-of-fit problem for exponential families is similar. A rigorous theory for conditional tests is in Witting (1966).

Example 2: Bayesian Computations

Let $\mathcal{X} = \{1, 2, \ldots, d\}$ be a finite set of objects, numbered for convenience. Let π be the uniform prior on the probability distributions Δ on $\mathcal{X}$. Consider the process $X_1, X_2, X_2, \ldots \in \mathcal{X}$ obtained by drawing from $\mathcal{X}$ independently with the multinomial distribution $p \in \Delta$ after getting p with π. We are interested in the succession probability

$$P_\pi\{X_{n+1} = i | A(T_n) = \mathbf{a}\}$$

where A is a nonnegative integer matrix and $T_n \in Z_+^d$ is the total summary data defined by

$$T_n := \sum_{i=1}^{n} e_{X_i}$$

where $\{e_i : i = 1, \ldots, d\}$ is the standard basis for $\mathbf{R}^d$.

It is well-known that the uniform prior π on Δ results in the uniform distribution on summary tables T_n (as opposed to the multinomial distribution without the prior probability), and so

$$P_\pi\{A(T_n) = \mathbf{a}\} = \frac{|\{\mathbf{x} \in Z_+^d : A\mathbf{x} = \mathbf{a}, \sum_i x_i = n\}|}{\binom{n+d-1}{n}}.$$

Furthermore,

$$P_\pi\{T_n = \mathbf{x}, X_{n+1} = i \,|\, p\} = \binom{n}{x_1 \ldots x_d} p_1^{x_1} \ldots p_d^{x_d} p_i.$$

Then with some algebra, the succession probability becomes

$$P_\pi\{X_{n+1} = i | A(T_n) = \mathbf{a}\} = \frac{1}{n+d} \sum_{\mathbf{x} \in S} \frac{x_i + 1}{|S|},$$

an expectation with respect to the uniform distribution on

$$S := \{\mathbf{x} \in Z_+^d : A\mathbf{x} = \mathbf{a}, \sum_i x_i = n\}.$$

The special case with $d = 2, A = (0, 1), \mathbf{a} = (n)$ is the well known rule of succession of Laplace: after n successes in n trials, the probability of a further success is $\frac{n+1}{n+2}$.

Example 3: Network Traffic Estimation

Let $G = (V, E)$ be a directed graph with directed edge set E on numbered vertices V. The probability model for traffic between nodes is independent Poisson processes across all ordered pairs of distinct vertices. The sample space Ω is then the space of nonnegative integer matrices $Z_+^{V \times V}$, with zeros on the diagonal, naturally a subset of a d-dimensional vector space where $d := |V|^2 - |V|$. Let $r = |E|$, the number of edges. The family of probability distributions is also of dimension d, and is indexed by an off-diagonal matrix of Poisson rates $\lambda \in \mathbf{R}_+^{V \times V}$, again of dimension d. Let Λ be the d-dimensional parameter space of matrices. Then for $\lambda \in \Lambda$,

$$P_\lambda(\omega) = \prod_{i \neq j} \frac{\lambda_{ij}^{x_{ij}}}{\omega_{ij}!} e^{-\sum_{i,j} \lambda_{ij}}.$$

The data $\mathbf{y}$ is in the form of counts across directed edges, and is therefore of dimension $r = |E|$. A $0-1$ matrix A of dimension $r \times d$ counts traffic across edges, and can be partitioned into the form $A = (\, A_0 \;|\; I \,)$, where I is an $r \times r$ identity matrix with an appropriate ordering of edges and node pairs.

Star network. In this graph, there is a center node through which all traffic is routed. If the vertices are numbered $1, 2, ..., v-1, v$, we place node v in the center and get edges $E = \{(i, v) : i = 1, 2, \ldots, v-1\} \cup \{(v, j) : j = 1, 2, \ldots, v-1\}$. The outcomes are in dimension $c = v \times (v-1)$, and the data is in dimension $r = 2 \times (v-1)$.

With $v = 4$, order the 12 vertex pairs so the 6 edge pairs come last, say $(1,2), (1,3), (2,1), (2,3), (3,1), (3,2), (1,4), (2,4), (3,4), (4,1), (4,2), (4,3)$, and order the the edges lexicographically $(1,4), (2,4), (3,4), (4,1), (4,2), (4,3)$. Then the $r \times d = 6 \times 12$ matrix A can be written:

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The likelihood of the data $\mathbf{y} \in Z_+^E$ is the probability of the collection of outcomes that result in edge counts represented by $\mathbf{y}$:

$$L_{\mathbf{y}}(\lambda) = \sum_{\omega \in \Omega : A\omega = \mathbf{y}} \prod_{i,j} \frac{\lambda_{ij}^{\omega_{ij}}}{\omega_{ij}!} e^{-\sum_{i,j} \lambda_{ij}}.$$

Any local extremum satisfies the fixed point equation

$$E_{\lambda^*, \mathbf{y}}(\omega) = \lambda^*.$$

With a positive initial guess $\lambda_0 > 0$ close to the true λ, the iteration scheme

$$\lambda_{k+1} := E_{\lambda_k, \mathbf{y}}(\omega)$$

usually converges to a fixed point which maximizes $L_{\mathbf{y}}$. The situation with $K > 1$ repeated, independent network observations can also be treated with this method.

The Markov chain on the set $\{\omega : A\omega = \mathbf{y}\}$ constructed with the Gröbner basis for lex term order and with slack indeterminates ordered last is exactly the ordinary reflecting random walk in the $d - r$ dimensional set

$$S_{\mathbf{y}} := \{(z_{ij})_{ij \in V \times V - E}, z_{ii} = 0, A_0 \mathbf{z} \leq \mathbf{y}\}$$

Computing with this chain is much less efficient than independent uniform sampling from a box $B_{\mathbf{y}}$ which contains $S_{\mathbf{y}}$ (Dinwoodie 1999).

Example 4: Simulated Annealing Algorithm

With arbitary objective function $f : S \to \mathbf{R}$, let μ_t be the probability distribution on S (concentrated on the low points of f) defined by

$$\mu_t(\mathbf{x}) := \frac{e^{-f(\mathbf{x})/t}}{Z_t}.$$

As $t \downarrow 0$, μ_t puts more mass on the low points of f. The Metropolis algorithm with the symmetric Markov chain gives a way to run an ergodic process in S with stationary distribution μ_t. As stationarity is approached, lower t further. It is quite likely that after sufficient time, the chain is in a state where f is near its minimum. Time to equilibrium is studied in Dinwoodie (1998).

Multivariate Generating Functions

This approach can be useful for conditional tests of parameter significance in exponential families, which are more difficult than goodness-of-fit.

Example 1: Odds ratio in logistic model with 2 factors

Consider the exponential generating function

$$g = \sum_{\mathbf{x} \in Z_+^d : A\mathbf{x} = \mathbf{a}} \frac{\zeta^{\mathbf{x}}}{\mathbf{x}!}$$

where we are using the simplified monomial notation $\zeta^{\mathbf{x}} := \zeta_1^{x_1} \ldots \zeta_d^{x_d}$ with indeterminates $\zeta_1, ..., \zeta_d$, and also the factorial notation $\mathbf{x}! := x_1! \ldots x_d!$.

We want to combine terms in g which are equivalent in the sense that the vector exponents have the same value under a further new linear function t_{r+1}. We assume that t_{r+1} can also be represented as a dot product with a vector of nonnegative integers, like a row of A. This combination of terms can be done with another normal form reduction as we describe next.

Let $F_{r+1} := [p_1, \ldots, p_g]$ be a reduced Gröbner basis of monomial differences for the ideal $I_{r+1} := < \zeta^{\mathbf{x}} - \zeta^{\mathbf{y}} : t_{r+1}(\mathbf{x}) = t_{r+1}(\mathbf{y}) >$. Let $\texttt{normalf}(g, F_{r+1})$ be the normal form of the exponential generating function g with respect to F_{r+1} and pure lex order. Define the hypergeometric distribution μ_h on S by

$$\mu_h(\mathbf{x}) := \frac{1}{z} \frac{1}{\mathbf{x}!}$$

where z is a normalizing constant. It is the sum of the coefficients in g. The probability distribution μ_h arises typically as the image distribution on S of the uniform distribution on a subset of a more basic sample space.

Theorem. Let $g_{r+1} := \texttt{normalf}(g, F_{r+1})$ Then g_{r+1} has the form of a finite sum over an integer variable t:

$$g_{r+1} = \sum_{t=0} c_t \zeta^{\mathbf{x}_t}$$

for distinct monomials $\zeta^{\mathbf{x}_t}$ with $t_{r+1}(\mathbf{x}_t) = t$ and coefficients $c_t = z\, \mu_h(\{\mathbf{x} \in S : t_{r+1}(\mathbf{x}) = t\})$.

Proof: Equivalent monomials are reduced to a single equivalent normal form, and their coefficients are combined, since the normal form of a sum is the sum of the

normal forms:

$$\begin{aligned}\texttt{normalf}(g, F_{r+1}) &= \sum_{\mathbf{x}\in Z_+^d : A\mathbf{x}=\mathbf{a}} \frac{1}{\mathbf{x}!}\,\texttt{normalf}(\zeta^{\mathbf{x}}, F_{r+1}) \\ &= \sum_{t=0} \sum_{t_{r+1}(\mathbf{x})=t} \frac{1}{\mathbf{x}!}\,\texttt{normalf}(\zeta^{\mathbf{x}}, F_{r+1}) \\ &= \sum_{t=0} z\,\mu_h(\{t_{r+1}=t\})\,\zeta^{\mathbf{x}_t}\end{aligned}$$

where $\zeta^{\mathbf{x}_t}$ is the normal form of all the monomials whose vector exponents have value $t \in Z_+$ under t_{r+1}. The particular monomials that are present in g_{r+1} may depend on the indeterminate ordering. ∎

The normal has two useful properties:

- the normal form of a sum is the sum of the normal forms
- the normal form of a product is the normal form of the product of the normal forms.

Consider binary data from Collett (1991). The data is in the form of death counts, where a death is a success on a Bernoulli trial. The data comes from an experiment on the lethal effects of a certain harmful agent on the tobacco budworm. We use a logistic model with two factors: dose and sex. The dose factor has 6 levels (corresponding to the logarithm of the actual doses in μg), and the sex factor has two levels "M" and "F". We want to determine the p-value of significance for the sex factor.

Twenty Bernoulli trials were undertaken at each combination of factors, and the data can be summarized in the $d = 2 \times 6 \times 2 = 24$-dimensional table below. The dose factor is labelled "0 1 2 3 4 5" across the top.

Female:

logdose :	0	1	2	3	4	5
alive	20	18	14	10	8	4
dead	0	2	6	10	12	16

Male:

logdose :	0	1	2	3	4	5
alive	19	16	11	7	2	20
dead	1	4	9	13	18	20

In precise terms, the probability model supposes a death probability at level "i" of the dose factor and level "Female" of the sex factor to be

$$p_{i,F} = \frac{e^{\alpha_i}}{1+e^{\alpha_i}}$$

whereas the death probability for level "Male" has an extra parameter $\beta \geq 0$ which reflects the ratio of the odds of death of Male to Female:

$$p_{i,M} = \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}.$$

We will test $\beta = 0$ versus $\beta > 0$.

It is convenient to represent the data with coordinate variables $x_{i,M,1}$ for the number of deaths at dose i and sex M, $x_{i,M,0}$ for the number of failures, etc. Then $r = 18$, since there are 12 design constraints (the total number of trials at each combination of dose and sex is 20, or in precise terms $x_{i,M,0} + x_{i,M,1} = 20$, etc.)

and six linear constraints from sufficient statistics for the submodel:

$$\begin{aligned}
x_{0,F,1} + x_{0,M,1} &= 1 \\
x_{1,F,1} + x_{1,M,1} &= 6 \\
x_{2,F,1} + x_{2,M,1} &= 15 \\
x_{3,F,1} + x_{3,M,1} &= 23 \\
x_{4,F,1} + x_{4,M,1} &= 30 \\
x_{5,F,1} + x_{5,M,1} &= 36.
\end{aligned}$$

Let the actual observed data be denoted by the vector $\mathbf{x}_0$.

The number of terms in the exponential generating function g is $221,760 = 2 \times 7 \times 16 \times 18 \times 11 \times 5$ which one can get using one-variable generating functions for counting integer solutions to linear equations. There is present a simple product structure which implies that the exponential generating function g is a product of 6 homogeneous polynomials each of degree 40, one for each level of the dose factor, and so g can be expressed exactly without difficulty. The actual multiplication should be carried out in combination with computation of the normal form.

The extra linear statistic t_{19} needed to test significance of the odds ratio parameter β is given by

$$t_{19}(\mathbf{x}) = x_{0M1} + x_{1M1} + \ldots + x_{5M1}.$$

the total number of successes at the level "M" of the sex factor.

The normal form g_{19} is a polynomial in x_{5M1} of 54 terms, which greatly simplifies the conditional data of 221,760 tables. The value of the statisic t_{19} can be read from the exponent, and shows a range $t_{19} \in \{29, \ldots, 82\}$.

Then the exact p-value is

$$\mu_h(\hat{\beta} \geq 1.0896510) = \frac{184835984827152341713}{18122964042397296120538 5},$$

which is approximately .0010198993 to ten decimal places.

Other tests of parameter significance in exponential families can be done similarly. In particular, the logistic regression problem with an integer covariate can be treated exactly with generating functions.

References

1. D. Collett (1991). *Modelling Binary Data*. Chapman and Hall, New York.
2. P. Diaconis and A. Gangolli (1995). Rectangular arrays with fixed margins. In *Discrete Probability and Algorithms*, ed. D. Aldous *et al.*, Springer, New York.
3. P. Diaconis and B. Sturmfels (1998). Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, **26**, 363-397.
4. I. H. Dinwoodie (1998). The Diaconis-Sturmfels algorithm and rules of succession. *Bernoulli*, **4**, 401-410.
5. I. H. Dinwoodie (1998). Stochastic simulation on integer constraint sets. *Siam J. Optim.*, **9**, 53-61.
6. I. H. Dinwoodie (1999). Exponential generating functions, normal forms, and exact computation of p-values for integer data. Manuscript.
7. I. H. Dinwoodie (1999). Conditional expectations in network traffic estimation. *Statistics and Probability Letters*, to appear.
8. L. C. Lazzeroni and K. Lange (1997). Markov chains for Monte Carlo tests of genetic equilibrium in multidimensional contingency tables. *Ann. Statist.*, **25**, 138-168.
9. C. Mehta and N. Patel (1983). A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Jour. Amer. Statist. Assoc.*, **78**, 427-434.
10. B. Sturmfels (1996). *Gröbner Bases and Convex Polytopes*. AMS, Providence, Rhode Island.
11. Y. Vardi (1996). Network tomography: estimating source destination traffic intensities from link data. *JASA*, **91**, 365-377.
12. H. Witting (1966). *Mathematische Statistik*. Teubner, Stuttgart.

Probability Computations on finite sample spaces and applications

Giovanni Pistone & Henry P. Wynn

(with Eva Riccomagno)

References:

- Mongraph in progress Ch. 5
- Paper presented at Helsinki 18-8-99

Basic probability

Ω is a finite set, $\mathcal{A}$ is a *σ-algebra* on Ω,

$$\mathbf{P} : \mathcal{A} \mapsto [0,1]$$

is a probability, $\mathcal{L} = \mathcal{L}(\Omega, \mathcal{A})$ is a *Ring* of real random variables, and

$$\mathcal{A} \simeq \left\{ X \in \mathcal{L} \ : \ X^2 = X \right\} \quad \text{with} \quad A \simeq 1_A$$

As the state space is *finite*, the expectation is

$$\mathbf{E}_\mathbf{P} \ : \ \mathcal{L}^1(\Omega, \mathcal{A}) = \mathcal{L} \mapsto \mathbf{R}$$

- The algebraic setting for Ω, $\mathcal{A}$, $\mathcal{L}$ is the same setting as in the "design" case.
- The algebraic setting for $\mathbf{P}$ and $\mathbf{E}_\mathbf{P}$ is "new".

Basic probability theory

Image: Given $(\Omega, \mathcal{A}, \mathbf{P})$ and a random variable X, compute

$$P_X = \mathbf{P} \circ X^{-1}$$

Product: Given $(\Omega_i, \mathcal{A}_i, \mathbf{P}_i)$, $i = 1, 2$, compute

$$(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbf{P}_1 \otimes \mathbf{P}_2)$$

Conditioning: Given $(\Omega, \mathcal{A}, \mathbf{P})$ and an event A and random variables $X, Y_1 \cdots Y_d$, compute

- $\mathbf{E}_\mathbf{P}(X \mid Y_1 \cdots Y_d)$
- $\mathbf{P}(A \mid Y_1 \cdots Y_d)$

Basic Statistical Models

The set of probability measures on a (finite) set has various (parametric) descriptions, each of them pointing to a different *geometry* of the statistical model.

Mixture geometry if

$$\mathbf{P} \leftrightarrow \text{density } P$$

Hilbert geometry if

$$\mathbf{P} \leftrightarrow Q \text{ and } Q = \sqrt{P}, \text{or } P = Q^2$$

Exponential geometry if

$$\mathbf{P} \leftrightarrow U \text{ and } P = \mathrm{e}^{U - K(U)}$$

Steps

Level 1: Ω is described as a 0-dimensional variety af k^d $(k = \mathbf{Q})$

Level 2: $\mathcal{L}$ is descibed as

$$\mathcal{L}(\Omega, \mathcal{A}) \simeq \frac{K[x_1 \cdots x_d]}{\text{Ideal}\,(D)}$$

Level 3: $\mathbf{P}$ is described together with $\mathbf{E_P}$ in the algebraic framework (various definitions are possible).

Level 4: A further algebrisation is possible in the lattice case.

Level ∞: Algebraic geometry of statistical models, non commutative etc, come to the next Grostat to see . . . .

The ring of random variables

Let be given $D = \Omega$, τ, $\text{Est}_\tau\,(D)$ as in design theory. Then

$$\mathcal{L}(D) \simeq \frac{K[x_1 \cdots x_d]}{\text{Ideal}\,((\,)\,D)}$$

The ring $\mathcal{L}(D)$ has a basis as a vector space, i.e.

$$X^\alpha \;:\; D \mapsto K$$
$$\text{Est}_\tau\,(D) = \{x^\alpha \;:\; \alpha \in L\}$$

Note that x^α is a *monomial term*, while X^α is a random variavle (function) on D.

We have $\forall \alpha, \beta \in L$

$$\begin{aligned} X^\alpha X\beta &= X^{\alpha+\beta} \\ &= \sum_{\gamma \in L} r(\alpha+\beta, \gamma) X^\gamma \end{aligned}$$

The array

$$[r(\alpha+\beta, \gamma)]_{\alpha,\beta,\gamma \in L}$$

is the basic computational tool on $\mathcal{L}(D)$.

Sub-σ-algebras(1)

Let

$$Y_1 \cdots Y_k \;:\mapsto K$$

be the *observable* random variables. Any *function* of $Y_1 \cdots Y_k$ is a random variable, and they form a *sub-ring* of $\mathcal{L}(D)$. This sub-ring is *computable* as follows. Let us consider a basis

$$\text{Ideal}\,(g_1 \cdots g_l) = \text{Ideal}\,((\,)\,D)$$

and let us consider the ideal

$$\text{Ideal}\left(g_1 \cdots g_l, y_1 - \sum_{\alpha \in L} c_{1\alpha} x 6 \alpha \cdots\right) \subset K[x_1 \cdots x_d, y_1 \cdots y_l]$$

Elimination of x gives a basis of the ideal in $K[y_1 \cdots y_l]$ whose variety is the *image* of $Y_1 \cdots Y_d$.

Sub-σ-algebras(2)

The usual argument gives a list of monomials

$$\left\{y^\beta \mid \beta \in \tilde{L}\right\}$$

and the sub-ring of functions of $Y_1 \cdots Y_d$ is fully described by the formula

$$\sum_{\beta \in \tilde{L}} d_\beta Y^\beta$$

The σ algebra generated by $Y_1 \cdots Y_l$ is the solution for d_β, $\beta \in \tilde{L}$ of the system

$$\begin{cases} g_1(x) = 0 \\ \vdots \\ g_l(x) = 0 \\ y_1 - \sum_{\alpha \in L} c_{1\alpha} x^\alpha = 0 \\ \vdots \\ y_k - \sum_{\alpha \in L} c_{k\alpha} x^\alpha = 0 \\ \left(\sum_{\beta \in \tilde{L}} d_\beta y^\beta\right)^2 - \sum_{\beta \in \tilde{L}} d_\beta y^\beta = 0 \end{cases}$$

Uniform probability

If $\mathbf{P}_0$ is the uniform probability on D, then for all $X \in \mathcal{L}(D)$

$$\mathbf{E}_0 = \frac{1}{\#D} \sum_{a \in D} X(a)$$

As

$$X = \sum_{\alpha \in L} c_\alpha X^\alpha$$

it is enough to know the uniform moments

$$m(\alpha) = \mathbf{E}_0(X^\alpha)$$

The uniform moments depend on the "algebra" through the following

Proposition
If $R = [r(\alpha + \beta, \gamma)]_{\gamma, \alpha \in L}$, then

$$m(\beta) = \frac{1}{\#L} \text{Trace}\,(R(\beta)$$

General probability, general moments, image probability

It is easy now to introduce general probabilities with densities represented as

$$P = \sum_{\alpha \in L} \theta_\alpha X^\alpha$$

As a first example, we discuss the image under the random variable(s) Y of the probability with density P. We need to compute

$$\mathbf{E}_Y(F) = \mathbf{E}_P(F(Y))$$

We use

$$F(Y) = \sum_{\beta \in \bar{L}} a_\beta Y^\beta$$

so all we need is

$$\mathbf{E}_P(Y^\beta), \quad \beta \in \bar{L}$$

Il $\mathbf{P}_Y(A)$ is needed, use the indicator function.

Conditional expectation

- $\mathcal{L}(D) \simeq \frac{K[x_1 \cdots x_d]}{\text{Ideal}(D)}$ is the ring of random variables.
- $Y_1 \cdots Y_k$ are given observables.
- $\mathcal{L}(Y_1 \cdots Y_k)$ is the sub-ring generated by the observables.
- $\left\{ y^\beta \ : \ \beta \in \bar{L} \right\}$ is the list of estimable terms in the image space.

Then *the regression of X on Y^β, $\beta \in \bar{L}$, is the conditional expectation of X given $Y_1 \cdots Y_k$.*

Proof: $\mathcal{L}(Y_1 \cdots Y_k)$ is a vector space (least squares) and a ring (conditional expectation).

Exponentials and exponential models

The exponential representation of probabilities falls into the algebraic framework because exponential function (as any other function on D have a polynomial representation.

By this polynomial representation, any exponential model has the form

$$\exp\left(\sum_{\alpha \in L} \psi_\alpha X^\alpha \right) = \sum_{alpha \in L} c_\alpha(\psi) X^\alpha$$

where the coefficients of the right-end-side of the equation is in the field

$$K = k(\mathrm{e}^{a^\alpha \psi} \ : \ a \in D, \alpha \in L)$$

Differential Resultant Systems and Statistics

GIUSEPPA CARRA' FERRO

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF CATANIA

Viale Andrea Doria 6, 95125 CATANIA

ITALY

September 7, 1999

Recently many tools from computational differential algebra were used in statistics.

One main problem in such mathematical models is the elimination of some differential variables in a system of ordinary algebraic differential equations.

There are many differential elimination procedures, coming from different theories.

- The characteristic set theory of Wu-Ritt (*Ritt, Kolchin, Wu, Wang*),
- The Gröbner bases theory together the characteristic set theory (*Boulier,Diop, Fliess*),
- The differential resultants theory (*)
- The differential Gröbner bases theory (*,*Ollivier*)

The algebraic multivariate resultant theory was recently studied by many people (*Gelfand-Kapranov-Zelevinski, Jouanolou, Sturmfelds*) and different algorithms were implemented for its calculus in the sparse case (*Canny, Manocha, Sturmfelds*).
Since a differential resultant is a multivariate algebraic resultant, then all results and algorithms can be properly used in the differential case.

The differential resultant theory allows to know some upper bounds on the orders of the differential equations in the output of a differential elimination process.

In fact it is shown that given a system of $m+s$ ordinary algebraic differential equations $\{f_j = 0\colon\ j = 1,\ldots,m+s\}$ in $m+r$ differential variables with $r+1 < s$ and $ord(f_i) = n_i$, it is possible to obtain generalized differential resultant systems of ordinary algebraic differential equations in the coefficients of f_i, when $i = m+1,\ldots,m+s$.

If $n_1 \leq n_2 \leq \ldots \leq n_{m+s}$, then the maximal orders of the differential equations in a generalized differential resultant system is equal to $\sum_{i=0,1,\ldots,m} n_{m+s-i}$. Such results can be applied in the example of n-strain model.

1 Differential Algebra Preliminaries

Every undefined notion is as in [*Ritt (1950)*] and [*Kolchin (1973)*]. Let δ be a derivation and let R be a differential ring, i.e. a commutative ring with unit and a derivation δ acting on it, such that $\mathbf{Q} \subset R$.

C_R=$\{a \in R$: $\delta(a) = 0\}$=ring of constants of R and $\mathbf{N}_0 = \{0, 1, \ldots, n, \ldots\}$.

REMARK 1 $\mathbf{Q} \subseteq C_R$ *because* $\mathbf{Q} \subseteq R$.

EXAMPLE 1 R=$C^\infty(\mathbf{R})$, R=$\mathbf{Q}(t)$, R=$\mathbf{R}(t)$ *and* R=*field of the meromorphic functions on a domain of* $\mathbf{C}$ *with the usual derivation* δ=$\frac{\partial}{\partial t}$ *are differential rings.*

R=$C^m(\mathbf{R})$ *for all* m *in* $\mathbf{N}_0$ *is not a differential ring.*

If $a \in \mathbf{N}_0$, then *order* of δ^a is ord(δ^a)=a.

DEFINITION 1 S=$R\{y_1, \ldots, y_m\}$=$R[\delta^n y_i$: $i = 1, \ldots, m$ *and* $n \in \mathbf{N}_0]$ *is the differential ring of the differential polynomials in the differential indeterminates* $y_1, \ldots, y_m$ *with coefficients in the differential ring* R.

If $f \in S$, then *order* of f is ord(f)=max$\{$ $n \in \mathbf{N}_0$: f contains a power product in $y_1, \ldots, y_m, \delta y_1, \ldots, \delta y_m, \ldots, \delta^n y_1, \ldots, \delta^n y_m$ with nonzero coefficient $\}$.

DEFINITION 2 *An ideal* I *of* S *is a differential ideal iff* $\delta^n(s) \in I$ *for all* $n \in \mathbf{N}_0$ *and all* $s \in I$. *If* $A \subseteq S$, *then* $[A]$=$(\ \delta^n(s)$: $s \in A$, $n \in \mathbf{N}_0\ )$ *denotes the smallest differential ideal containing* A.

2 Algebraic Elimination Theory

2.1 Gröbner bases

The Gröbner bases theory is very useful in the elimination of variables.

Given $k+s$ polynomials $g_1,\ldots,g_{k+s}$ in $S=R[X_1,\ldots,X_{k+r}]$ with $deg(g_j)=d_j$ for all j we want to *eliminate* the variables $X_1,\ldots,X_k$.

In other words given the ideal $I=(g_1,\ldots,g_{k+s})$ we want to find the *k-elimination ideal* $I_k=I\cap R[X_{k+1},\ldots,X_{k+r}]$.

By Gröbner bases theory it is sufficient to a find a Gröbner basis of the ideal I with respect to an elimination term ordering τ_k as in [*Cox, Little and O'Shea (1992)*].

A general elimination term ordering τ_k in this case is given by two arbitrary term orderings σ_1 on the set of power products P_1 in $X_1,\ldots,X_k$ and σ_2 on the set of power products P_2 in $X_{k+1},\ldots,X_{k+s}$.

If $p,p'\in P_1$ and $q,q'\in P_2$, then $pq <_\tau p'q'$ iff either $p <_{\sigma_1} p'$ or $p=p'$ and $q <_{\sigma_2} q'$.

If G is a Gröbner basis of I with respect to τ_k, then $G_k = G\cap R[X_{k+1},\ldots,X_{k+r}]$ is the Gröbner basis of I_k with respect to σ_2.

2.2 Multivariate Algebraic Resultant

Let R be an integral domain and let K be the quotient field of R. Given $k+1$ polynomials $g_1, \ldots, g_{k+1}$ in $S=R[X_1, \ldots, X_k]$ with $deg(g_j)=d_j$ for all j [*Macaulay (1916)*] defines a polynomial Q only in the coefficients of $g_1, \ldots, g_{k+1}$, such that a necessary condition in order that the system $\{g_1 = 0, \ldots, g_{k+1} = 0\}$ has a solution in an algebraic extension of K is $Q=0$ and $Q=Res(g_1, \ldots, g_{k+1})$ is called the *resultant of* $g_1, \ldots, g_{k+1}$.

Let $D=\sum_{j=1,\ldots,k+1} d_j - k$, $L=\binom{k+D}{k}$,

$M_k^D=\{X^a=X_1^{a_1} \cdots X_k^{a_k}: \sum_{i=1,\ldots,k} a_i \leq D, a = (a_1, \ldots, a_k) \in \mathbf{N}_0^k\}$

is the set of all power products in S of degree less than or equal to D. The set M_k^D has cardinality equal to L, while $M_k^{D-d_j}$ has cardinality equal to $L_j=\binom{k+D-d_j}{k}$ for all $j = 1, \ldots, k+1$.

M_k^D and $M_k^{D-d_j}$ are totally ordered using first the degree and then the lexicographic order with $X_1 < \ldots < X_k$.

DEFINITION 3 *$M=M(f_1, \ldots, f_{k+1})$ is the $(\sum_{j=1,\ldots,k+1} L_j \times L)$-matrix defined in the following way:*

for each i such that $\sum_{h=1,\ldots,j} L_h < i \leq \sum_{h=1,\ldots,j+1} L_h$ the coefficients of the polynomial $X^a f_{j+1}$ are the entries of the i-th row for each $X^a \in M_k^{D-d_{j+1}}$ that are written in decreasing order with respect to the power products in M_k^D.

DEFINITION 4 *Let g_j be a polynomial of degree d_j for all $j = 1, \ldots, k+1$ in k variables $X_1, \ldots, X_m$.*

$Res(f_1, \ldots, f_{k+1})$= G.C.D.($det(P)$: P is an $L \times L$-submatrix of M).

In [*Macaulay (1916)*] it is shown that, in the generic case, the resultant of polyomials can be written as $\det(M_0/det(A)$, where M_0

is a square submatrix of M of maximal rank and A is a sumatrix of M_0.

THEOREM 1 *Let g_j be a polynomial of degree d_j for all $j = 1, \ldots, k+1$ in k variables $X_1, \ldots, X_k$. There exists a nonzero $r \in R$ such that $rRes(g_1, \ldots, g_{k+1}) = \sum_{j=1,\ldots,k+1} f_j g_j$, where f_j is a polynomial in $R[X_1, \ldots, X_k, X_{k+1}]$ with $deg(f_j) \leq \sum_{h=1,\ldots,k+1, h \neq j} d_h - k$ for all $j = 1, \ldots, k+1$.*

2.3 Resultants and Elimination Ideals

The resultants can be used for the elimination of a set of variables. In fact the univariate and multivariate resultants can be used for the elimination of the set of k variables $\{X_1, \ldots, X_k\}$ in a set of $k+s$ polynomials $g_1, \ldots, g_{k+s}$ in $R[X_1, \ldots, X_k, X_{k+1}, \ldots, X_{k+r}]$ with $s \geq 1$ as in [*Hodge and Pedoe (1947)*], [*Van der Waerden (1950)*], [*Cox, Little and O'Shea (1992) and (1998)*].

Let $g^*_{k+1} = \sum_{j=k+1,\ldots,k+s} A_j g_j$, where A_j are new indeterminates. Let's consider $\{g_1, \ldots, g_k, g^*_{k+1}\}$ as polynomials in $R[A_{k+1}, \ldots, A_{k+s}, X_{k+1}, \ldots, X_{k+r}][X_1, \ldots, X_k]$. $Q = Res(g_1, \ldots, g_k, g^*_{k+1})$ is a polynomial in $R[A_1, \ldots, A_{k+s}, X_{k+1}, \ldots, X_{k+r}]$. $Q = \sum_{i=1,\ldots,p} Q_i M_i$, where the M_is are distinct power products with the same degree in $A_{k+1}, \ldots, A_{k+s}$, while the Q_is are polynomials in $X_{k+1}, \ldots, X_{k+r}$, because g^*_{k+1} is homogeneous in $A_{k+1}, \ldots, A_{k+s}$.

DEFINITION 5 *Let $s \geq 1$ and let g_j and Q_i be as above for all $j = 1, \ldots, k+s$ and $i = 1, \ldots, p$. Q_i is called a* generalized resultant of $g_1, \ldots, g_k, \ldots, g_{k+s}$ *[Cox, Little and O'Shea (1992)]. E=$(Q_1, \ldots, Q_p)$ is called a* generalized resultant ideal of $g_1, \ldots, g_{k+s}$.

Let $I = (g_1, \ldots, g_{k+s})$.

THEOREM 2 *Let $s \geq 1$ and let g_j and Q_i be as above for all $j = 1, \ldots, k+s$ and $i = 1, \ldots, p$. There exists a nonzero element r in R, such that rQ_i is in the k-elimination ideal I_k of I.*

REMARK 2 *I_k is not always generated by $Q_1, \ldots, Q_p$ as it is remarked in [Cox, Little and O'Shea (1992)].*

The generalized resultants and the generalized resultant ideal depends on which polynomial is the first in the ordered sequence $\{g_1, \ldots, g_k, \ldots, g_{k+s}\}$.

REMARK 3 *Let I as above and let I^*=$(g_1, \ldots, g_k, g^*_{k+1})$ be the corresponding ideal in $R[X_1, \ldots, X_{k+r}, A_{k+1}, \ldots, A_{k+s}]$= $R[X_{k+1}, \ldots, X_{k+r}, A_{k+1}, \ldots, A_{k+s}][X_1, \ldots, X_k]$, as polynomial ring in the variables $X_1, \ldots, X_k$ with coefficients in T=$R[X_{k+1}, \ldots, X_{k+r}, A_{k+1}, \ldots, A_{k+s}]$.*

Let τ be an elimination term ordering on the set of power products in $X_1, \ldots, X_k$. By using Gröbner bases theory let G^* be a Gröbner basis of I^* with respect to τ. Since g^*_{k+1} is homogeneous in $A_{k+1}, \ldots, A_{k+s}$, then every polynomial in $G^* \cap R[X_{k+1}, \ldots, X_{k+r}, A_{k+1}, \ldots, A_{k+s}]$ is homogeneous in the variables $A_{k+1}, \ldots, A_{k+s}$. The generalized resultants are in the ideal generated by the nonzero coefficients of the power products in $A_{k+1}, \ldots, A_{k+s}$, that appear in some polynomial of G^*.

The generalized resultant system and the generalized resultant ideal extend the notions of *resultant system* and *resultant ideal of* a set of $k+s$ polynomials $g_1, \ldots, g_{k+s}$ in $k+r$ variabili.

The resultant system depends on the definite order of elimination of the variables, while the resultant ideal is independent on such order.

The proof is usually given in the case k=1 and then inductively by eliminating one variable at each time

[*Hodge and Pedoe (1947)*].

3 Differential Elimination Theory

3.1 Differential Gröbner bases

Differential Gröbner bases [*, *(1987)*], [*Ollivier (1990)*] can be used for the elimination of the set of m differential variables $\{X_1, \ldots, X_m\}$ in a set of $m+s$ differential polynomials $f_1, \ldots, f_{m+s}$ in $R\{X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+r}\}$ with $0 \leq r$.

In other words given the differential ideal I=$[g_1, \ldots, g_{k+s}]$ we want to find the *differential m-elimination ideal* I_m=$I \cap R[X_{m+1}, \ldots, X_{m+r}]$. By Gröbner bases theory it is sufficient to a find a differential Gröbner basis of the differential ideal I with respect to a differential elimination term ordering.

Unfortunately there are only partial results. Even differential term orderings on the set of differential power products, i.e. power products in the variables and in their derivatives, are not completely classified [*Weispfenning (1993)*].

Let $S=R\{X_1,\ldots,X_{m+r}\}$ and let $S_t=R[X_1,\ldots,X_{m+r},\ldots,\delta^t X_1,\ldots,\delta^t X_{m+r}]$ for all $t\in\mathbf{N}_0$.

The following are useful examples of differential term ordering.

EXAMPLE 2 *Let σ_t be the lexicographic term ordering on the set of power product P_t in $X_1,\ldots,X_{m+r},\ldots,\delta^t X_1,\ldots,\delta^t X_{m+r}$ with $\delta^t X_1 >_{\sigma_t}\ldots>_{\sigma_t}\delta^t X_{m+r}>_{\sigma_t}\ldots>_{\sigma_t} X_1>_{\sigma_t}\ldots>_{\sigma_t} X_{m+r}$ for every $t\in\mathbf{N}_0$. The set $\{\sigma_t\colon\}t\in\mathbf{N}_0$ defines a differential term ordering σ on the set of power products in $X_1,\ldots,X_{m+r}$ and their derivatives by taking its restriction to P_t equal to σ_t. σ is not a differential elimination term ordering.*

EXAMPLE 3 *Let τ_t be the lexicographic term ordering on P_t with $\delta^t X_1>_{\tau_t}\ldots>_{\tau_t} X_1>_{\tau_t}\ldots>_{\tau_t}\delta^t X_{m+r}>_{\tau_t}\ldots>_{\tau_t} X_{m+r}$ for every $t\in\mathbf{N}_0$. The set $\{\tau_t\colon t\in\mathbf{N}_0\}$ defines a differential term ordering τ on the set of power products in $X_1,\ldots,X_{m+r}$ and their derivatives by taking its restriction to P_t equal to τ_t. τ is a differential elimination term ordering.*

EXAMPLE 4 *Let ρ_t be the lexicographic term ordering on P_t with $\delta^t X_1>_{\rho_t}\ldots>_{\rho_t}\delta^t X_m>_{\rho_t}\ldots>_{\rho_t} X_1>_{\rho_t}\ldots>_{\rho_t} X_m>_{\rho_t}\delta^t X_{m+r}>_{\rho_t}\ldots>_{\rho_t} X_{m+r}$ for every $t\in\mathbf{N}_0$. The set $\{\rho_t\colon t\in\mathbf{N}_0\}$ defines a differential term ordering ρ on the set of power products in $X_1,\ldots,X_{m+r}$ and their derivatives by taking its restriction to P_t equal to ρ_t. ρ is a differential elimination term ordering, that allows to eliminate the differentiale variables $X_1,\ldots,X_m$.*

3.2 Multivariate Differential Resultant

Let R be a differential integral domain and let K be the differential quotient field of R.

Let f_j be a differential polynomial in $R\{X_1, \ldots, X_m\}$ of order n_j and degree d_j for each $j = 1, \ldots, s$. $f_j \in R[X_1, \ldots, X_m, \ldots, \delta^{n_j} X_1, \ldots, \delta^{n_j} X_m]$, which is a polynomial ring in $(n_j + 1)m$ variables.

$f_j, \delta f_j, \ldots, \delta^{r_j} f_j$, $j = 1, \ldots, s$ are $\sum_{j=1,\ldots,s}(r_j + 1)$ polynomials in $(\sum_{j=1,\ldots,s} r_j) + s - 1$ variables iff $N = \sum_{j=1,\ldots,m+1} n_j$, $s = m + 1$ and $r_i = (N - n_i$ for all $i = 1, \ldots, s$ by [*, *(1997)*].

The *differential resultant* $\delta Res(f_1, \ldots, f_{m+1})$ of $f_1, \ldots, f_{m+1}$ is the Macaulay's algebraic resultant of the $mN + m + 1$ polynomials $\delta^{N-n_1} f_1, \ldots, \delta f_1, f_1, \ldots, \delta^{N-n_{m+1}} f_{m+1}, \ldots, \delta f_{m+1}, f_{m+1}$ in $S_{mN+m} = R[X_1, \ldots, X_m, \ldots, \delta^N X_1, \ldots, \delta^N X_m]$, which is a polynomial ring in $mN + m$ variables [*, *(1997)*].

EXAMPLE 5 *Let* $f_j = \dot{x}_j - g_j(X_1, \ldots, X_m)$ $j = 1, \ldots, m$ *where* g_j *is in* $R[X_1, \ldots, X_m]$ *and let* f_{m+1} *a differential polynomial of order* 1 *in* $R\{X_1, \ldots, X_m\}$. $ord(f_j)$=1 *for all* j.

$N = m + 1$ *and the differential resultant of* $f_1, \ldots, f_{m+1}$ *is the algebraic resultant of the* $m^2 + 2m + 1$ *polynomials* $\delta^m f_1, \ldots, \delta f_1, f_1, \ldots, \delta^m f_{m+1}, \ldots, \delta f_{m+1}, f_{m+1}$ *in* $S_{m^2+2m} = R[X_1, \ldots, X_m, \ldots, \delta^{m+1} X_1, \ldots, \delta^{m+1} X_m]$, *which is a polynomial ring in* $m^2 + 2m$ *variables.*

EXAMPLE 6 *Let* $f_j = \dot{x}_j - g_j(X_1, \ldots, X_{m+1})$ $j = 1, \ldots, m + 1$, *where* g_j *is in* $R[X_1, \ldots, X_{m+1}]$. *The variables* $\{X_1, \ldots, X_m\}$ *can be eliminated.* $ord(f_j)$=1 *for all* $j = 1, \ldots, m$, *while* $ord(f_{m+1})$=0 *as differential polynomials in* $R\{X_{m+1}\}\{X_1, \ldots, X_m\}$.

$N = m$ *and the differential resultant of* $f_1, \ldots, f_{m+1}$ *is the algebraic resultant of the* $m^2 + m + 1$ *polynomials* $\delta^{m-1} f_1, \ldots, \delta f_1, f_1, \ldots, \delta^{m-1} f_m, \ldots, \delta f_m, f_m, \delta^m f_{m+1}, \ldots, \delta f_{m+1}, f_{m+1}$.

3.3 Differential Resultants and Differential Elimination

The differential resultant can be used for the elimination of the set of m differential variables $\{X_1, \ldots, X_m\}$ in a set of $m+r$ differential polynomials $f_1, \ldots, f_{m+r}$ in $R\{X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+s}\}$ with $s \geq 1$.

Let $f^*_{m+1}=\sum_{j=m+1,\ldots,m+s} A_j f_j$, where A_j are new differential indeterminates. Let's consider $\{f_1, \ldots, f_m, f^*_{m+1})\}$ as differential polynomials in $R\{A_{m+1}, \ldots, A_{m+s}, X_{m+1}, \ldots, X_{m+r}\}\{X_1, \ldots, X_m]$.

$Q=\delta Res(f_1, \ldots, f^*_{m+1})$ is in $R\{A_{m+1}, \ldots, A_{m+s}, X_{m+1}, \ldots, X_{m+r}\}$ by its own definition. $Q=\sum_{i=1,\ldots,p} Q_i M_i q_i$, where the M_is are distinct power products in $A_{m+1}, \ldots, A_{m+s}$ and their derivatives and the Q_is are differential polynomials in $X_{m+1}, \ldots, X_{m+r}$.

The power products M_is have the same degree, because g^*_{m+1} and its derivatives are homogeneous in $A_{k+1}, \ldots, A_{k+s}$ and their derivatives by its own definition.

DEFINITION 6 *Let $s \geq 1$ and let f_j and Q_i be as above for all $j = 1, \ldots, m+s$ and $i = 1, \ldots, p$. Q_i is called a* generalized differential resultant *of $f_1, \ldots, f_m, \ldots, f_{m+s}$. $E=[Q_1, \ldots, Q_p]$ is called a* generalized differential resultant ideal *of $f_1, \ldots, f_{m+s}$.*

THEOREM 3 *Let $s \geq 1$ and let g_j and Q_i be as above for all $j = 1, \ldots, k+s$ and $i = 1, \ldots, p$. There exists a nonzero element r in R, such that rQ_i is in the differential m-elimination ideal I_m of $I=[g_1, \ldots, g_{k+s}]$.*

REMARK 4 *I_m could be not always generated by $Q_1, \ldots, Q_p$ as in the algebraic case. In similar way the generalized differential resultants and the differential generalized resultant ideal depends on which polynomial is the first in the ordered sequence $\{f_1, \ldots, f_m, \ldots, f_{m+s}\}$.*

COROLLARY 1 *Let $s \geq 1$ and let g_j and Q_i be as above for all $j = 1, \ldots, k+s$ and $i = 1, \ldots, p$. Suppose that $n_1 \leq n_2 \leq \ldots \leq n_{m+s}$. $ord(Q_i \leq \sum_{j=1,\ldots,m} n_j + n_{m+s} - n^*_{m+1}$ for all i and for every choice of the first element of the sequence $\{f_1, \ldots, f_m, \ldots, f_{m+s}\}$.*

COROLLARY 2 *Let f_j be a differential polynomial in $R\{X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+r}\}$ of order n_j and degree d_j for each $j = 1, \ldots, s$ and let $s \geq 1$. Let Q_i be a generalized differential resultant of $f_1, \ldots, f_m, \ldots, f_{m+s}$. If $n_1 \leq n_2 \leq \ldots \leq n_{m+s}$, then $ord(Q_i) \leq \sum_{j=0,\ldots,m} n_{m+s-j}$ for all i and for every choice of the first element of the sequence $\{f_1, \ldots, f_m, \ldots, f_{m+s}\}$.*

COROLLARY 3 *Let f_j be a differential polynomial in $R\{X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+r}\}$ of order n_j and degree d_j for each $j = 1, \ldots, s$ and let $s \geq 1$. Let $I=[f_1, \ldots, f_m, \ldots, f_{m+s}]$ and let I_m be the differential m-elimination ideal of I. If $n_1 \leq n_2 \leq \ldots \leq n_{m+s}$, then $I_m = I \cap R\{X_{m+1}, \ldots, X_{m+r}$ has generators polynomials of order either less than or equal to $\sum_{j=0,\ldots,m} n_{m+s-j}$.*

3.4 Examples

EXAMPLE 7 *Let* $f_j = \dot{X}_j - g_j(X_1, \ldots, X_{m+r}$ *for all* $j = 1, \ldots, m+r$, *where* g_j *is a polynomial in* $X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+r}$ *for all* j.

We want to eliminate the variables $X_1, \ldots, X_m$. $n_j = ord(f_j) = 1$ *for all* $j = 1, \ldots, m$, *while* $n_j = ord(f_j) = 0$ *for all* $j = m+1, \ldots, m+r$ *as differential polynomials in the differential variables* $X_1, \ldots, X_m$. *Let* Q_i *be a generalized differential resultant of* $f_1, \ldots, f_m, f_{m+1}, \ldots, f_{m+r}$ *for all* $i = 1, \ldots, p$. *By corollary 1* Q_i *is a differential polynomial in* $X_{m+1}, \ldots, X_{m+r}$ *with* $ord(Q_i) \leq m+1$ *for all* $i = 1, \ldots, p$.

If $r = 1$, *then there is only one generalized differential resultant and its order is either less than or equal to* $m+1$.

If $r = m+1$, *then the generalized resultant* $Q' = Res(f_{m+1}, \ldots,$

$f_{2m+1})$ *of* $f_{m+1}, \ldots, f_{2m+1}$ *is a polynomial of order either less than or equal to* 1 *in* $X_{m+1}, \ldots, X_{2m+1}$, *that lies in the differential* m-*elimination ideal* I_m *of* $I = [f_1, \ldots, f_{2m+1}]$.

If $r > m+1$, *then the generalized resultants* Q'_h *of* $f_{m+1}, \ldots, f_{2m+1}$ *are polynomials of order either less than or equal to* 1 *in* $X_{m+1}, \ldots, X_{2m+1}$ *for all* $h = 1, \ldots, q$. *Moreover each* Q'_h *is in the differential* m-*elimination ideal* I_m *of* $I = [f_1, \ldots, f_{2m+1}]$.

The following example shows the use of algebraic and differential resultants in differential elimination. Furthermore it shows a procedure for the elimination of differential variables by using algebraic and differential Gröbner bases.

EXAMPLE 8 n-strain model *[Riccomagno and White (1997)]*

Let $f_j = \dot{X}_j - g_j(X_1, \ldots, X_{2^n-1}, Y_1, \ldots, Y_{n2^{n-1}}, Z_1, \ldots, Z_n)$

when $j = 1, \ldots, 2^n - 1$,

$f_j = \dot{Y}_j - g_j(X_1, \ldots, X_{2^n-1}, Y_1, \ldots, Y_{n2^n-1}, Z_1, \ldots, Z_n)$

when $j = 2^n, \ldots, n2^{n-1} + 2^n - 1$,

$f_j = \dot{Z}_j - g_j(X_1, \ldots, X_{2^n-1}, Y_1, \ldots, Y_{n2^n-1}, Z_1, \ldots, Z_n)$

when $j = n2^{n-1} + 2^n, \ldots, n2^{n-1} + 2^n + n - 1$,

where g_j *are polynomials. It is possible to eliminate the variables* $X_j s$ *and* $Y_j s$. $ord(f_j) = n_j = 1$ *for all* $j = 1, \ldots, n2^{n-1} + 2^n - 1$, *while* $ord(f_j) = n_j = 0$ *for all* $j = n2^{n-1} + 2^n, \ldots, n2^{n-1} + 2^n - 1 + n$ *as differential polynomial in* $R\{Z_1, \ldots, Z_n\}\{X_1, \ldots, X_{2^n-1}, Y_1, \ldots, Y_{n2^{n-1}}\}$. *So* $N = n2^{n-1} + 2^n - 1$.

If $n = 1$ *it is sufficient to take the algebraic resultant of the polynomials* $\delta f_1, f_1$,

$\delta f_2, f_2, \delta^2 f_3, \delta f_3, f_3$ *in the polynomial ring* $R[Z, \delta Z, \delta^2 Z, \delta^3 Z][X, \delta X, \delta^2 Z, Y, \delta Y, \delta^2 Y]$, *which is in* $R[Z, \delta Z, \delta^2 Z, \delta^3 Z]$.

If $n > 1$, *then let* $f^*_{n2^{n-1}+2^n} = \sum_{j=n2^{n-1}+2^n, \ldots, n2^{n-1}+2^n+n-1} A_j g_j$, *where* A_j *are arbitrary differential variables.*

Now it is possible to eliminate the variables $X_j s$ *and* $Y_j s$ *from the system* $\{f_1 = 0, \ldots, f_{n2^{n-1}+2^n-1} = 0, f^*_{n2^{n-1}+2^n} = 0\}$ *by using the differential resultants and the generalized differential resultants. By above example the order of each generalized differential resultant is either less than or equal to* $n2^{n-1} + 2^n = N + 1$.

$\{ f_j\colon j = 1, \ldots, n2^{n-1} + 2^n + n - 1 \}$ *is a differential Gröbner basis of the differential ideal* $I = [f_j\colon j = 1, \ldots, n2^{n-1} + 2^n + n - 1]$ *with respect to* σ, *but it is not in general a differential Gröbner basis of* I *with respect to* τ.

Let G_{01} *be the Gröbner basis of* $I = I_{01}$ *as a nondifferential ideal in* S_1 *with respect to* τ_1. *If* G_{01} *contains a polynomial only in the* $Z_j s$ *we stop.*

If not we take all derivatives of the g*'s in* G_{01}, *that are in* S_2, *and the corresponding ideal* I_{02}.

Let G_{02} *be the Gröbner basis of* I_{02} *in* S_2 *with respect to* τ_2.

If G_{02} *contains a polynomial only in the* $Z_j s$ *we stop.*

If not we take all derivatives of the g*'s in* G_{02}, *that are in* S_3, *and the corresponding ideal* I_{03} *and so on.*

If N *is the upper bound given by Macaulay's differential resultant theory, then* $N + 1 = n2^{n-1} + 2^n$ *is an upper bound for the order and* $N + 2$ *is an upper bound for the number of steps.*

The same procedure can be used if we take ρ *instead of* τ.

4 The Characteristic Set Method

The characteristic set method can be used as an elimination theory. It was introduced by Ritt in 1932 and it uses pseudodivisions and a ranking of the variables.

Let R be a differential ring. Let f_j be a differential polynomial in $R\{X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+r}\}$ of order n_j and degree d_j for each $j = 1, \ldots, m+s$. f_j is in $R[X_1, \ldots, X_m, \ldots, \delta^{n_j} X_1, \ldots, \delta^{n_j} X_{m+r}]$, which is a polynomial ring in $(n_j + 1)(m + r)$ variables.

In order to eliminate the differential variables $X_1, \ldots, X_m$ first we have to fix a ranking O on the set $\{X_1, \ldots, X_{m+r}\}$, which is a total order on the set

$\{ \delta^n X_j \colon j = 1, \ldots, m+r, n \in \mathbf{N}_0 \}$ compatible with the derivatives.

If $X_1 >_O \ldots >_O X_m$, then the Wu-Ritt algorithm for the extended characteristic set allows to find at most $m + 1$ new differential polynomials $h_j \in R\{X_j, \ldots, X_{m+r}\} \backslash R\{X_{j+1}, \ldots, X_{m+r}\}$, $j = 1, \ldots, m$, $h_{m+1} \in R\{X_{m+1}, \ldots, X_{m+r}\}$, that is an autoreduced set [*Kolchin (1973), Ritt (1950)*].

No bound is known on the number of derivatives of the f_js, that are necessary for finding the h_js.

If m=1, s=1, then an effective upper bound is given by n_2 and respectively n_1, i.e. the extended characteristic set of $\{f_1, f_2\}$ is in $R[X, \ldots, \delta^{n_1+n_2} X]$ [*, (1996)].

Such bound is equal to the bound obtained by using the notion of differential resultant, based on the Macaulay's definition of the resultant of polynomials.

This result can be extended to the general case m=1. It is sufficient to take f_2^*=$\sum_{j=2,\ldots,s+1} A_j f_j$, where $n_1 \geq n_2 \geq \ldots \geq n_{s+1}$ and the A_j's are differential indeterminates such that $\delta^h A_j <_O \delta^k X_1$ for all $j = 2, \ldots, s+1$ and all $h, k \in \mathbf{N}_0$ with respect to a ranking O on $\{A_2, \ldots, A_{s+1}, X_1\}$.

There is still an open question.
If $n_1 \geq n_2 \geq \ldots \geq n_s$, are there effective bounds for the order of derivatives of $f_1, \ldots, f_m, f_{m+1}^*$ given respectively by $\sum_{j=2,\ldots,m} n_j, \ldots$, $\sum_{j=1,\ldots,m-1} n_j$?

References

[*Boulier, Lazard, Ollivier and Petitot (1995)*] Boulier, F., Lazard, D., Ollivier F., Petitot M. (1995). Representation for the radical of a finitely generated differential ideal. *Proc. of ISSAC'95*. New York: ACM Press, 158–166

[*Boulier (1994)*] Boulier, F. (1994). Etude et Implantation de Quelques Algorithmes en algebre differentielle. *Ph. D. Thesis*. Universite de Lille

[*Boulier (1996)*] Boulier, F. (1996). Some improvements of a lemma of Rosenfeld. Preprint

[*Buchberger (1976)*] Buchberger, B. (1976). A theoretical basis for the reduction of polynomials to canonical form. *ACM SIGSAM BULL.* **10** (3), 19–29

[*Buchberger (1976)*] Buchberger, B. (1976). Some properties of Gröbner bases for polynomial ideals. *ACM SIGSAM BULL.* **10**, 19–24

[*Carra' Ferro (1987)*] Carra' Ferro, G. (1987). Gröbner Bases and Differential Algebra. *Lec. Not. in Comp. Sci.* **356**, 129–140

[*Carra' Ferro (1996)*] Carra' Ferro, G. (1996). Ritt's algorithm and Differential Resultants of two Algebraic ODE's", Preprint presented at the 2nd IMACS Conference on Applications of Computer Algebra, Linz luglio 1996

[*Carra' Ferro (1997)*] Carra' Ferro, G. (1997). A Resultant Theory for the Systems of Two Ordinary Algebraic Differential Equations. *Applicable Algebra in Engineering, Communication and Computing*, 539-560

[*Carra' Ferro (1997)*] Carra' Ferro, G. (1997). A Resultant Theory for Ordinary Algebraic Differential Equations. *Lect. Notes in Comp. Sci.* **1255**, 55–65

[*Cox, Little and O'Shea (1992)*] Cox, D. and Little, J. and O'Shea Donal (1992). *Ideals, Varieties and Algorithms*. New York: Springer-Verlag (1992)

[*Cox, Little and O'Shea (1998)*] Cox, D. and Little, J. and O'Shea Donal (1998). *Using Algebraic Geometry*. New York: Springer-Verlag (1998)

[*Diop and Fliess (1991)*] Diop, S., Fliess, M. (1991). On nonlinear observability. *Proc. of the first Europ. Control Conf. 1991*. Paris: Hermes, 152–157

[*Diop (1991)*] Diop, S. (1991). Elimination in control theory. *Math. Control Signal Systems* **4** , 17–32

[*Diop (1991)*] Diop, S. (1991). Finite morphisms of differential algebraic varieties and elimination theory. *Analysis of Controlled Dynamical Systems Proc. of Conference 1990*. Birkhauser , 193–200

[*Diop (1992)*] Diop, S. (1992). Differential algebraic decision methods and some application to system theory. *Theoret. Comp. Sci.* **98**, 137–161

[*Fliess (1989)*] Fliess, M. (1989). Automatique et corps differentiels. *Forum Math.* **1**, 227–238

[*Fliess (1990)*] Fliess, M. (1990). Generalized controller canonical forms for linear and nonlinear dynamics. *IEEE Trans. Autom. Control* **35**, 994–1001

[*Kolchin (1973)*] Kolchin, E. (1973). *Differential Algebra and Algebraic Groups*. New York: Academic Press

[*Hodge and Pedoe (1947)*] Hodge W. V. D., Pedoe D. (1947). *Methods of Algebraic Geometry* vol.1. Cambridge University Press (1947)

[*Macaulay (1916)*] Macaulay, F. S. (1916). *The Algebraic Theory of Modular Systems*, Proc. Cambridge Univ. Press

[*Ollivier (1990)*] Ollivier, F. (1990). Standard Bases of Differential Ideals. *Lec. Not. in Comp. Sci.* **508**, 304–321

[*Riccomagno and White (1997)*] Riccomagno E.,White L.J. (1997) *Multi-Strain Species Modeling via Differential Algebra Reduction* Preprint

[*Ritt (1950)*] Ritt, J.F. (1950). *Differential Algebra*. New York: AMS Coll. Publ. **33**

[*Wang (1993)*] Wang, D.M. (1993). Geometric theorem proving with existing technology. *MEDLAR II Deliverable DII.4.1P*.

[*Weispfenning (1993)*] Weispfenning, V. (1993). Differential-Term Orders. *Proc. of ISAAC'93 Kiev*. New York: ACM Press, 245-253

[*Wu (1987)*] Wu, W.T. (1987). Mechanical Derivation of Newton's Gravitational Laws from Kepler's Laws. *Mathematics Mechanization Research Reprints*. **1**, 53-61

[*Wu (1989)*] Wu, W.T. (1989). On the Foundation of Algebraic Differential Geometry. *Mathematics Mechanization Research Reprints*. **3**, 1-26 *Systems Sciences and Mathematical Sciences*. **2**, 289-312

[*Van der Waerden (1950)*] Van der Waerden B.L., (1950). Modern Algebra II", Third Edition. New York: F.Ungar Publishing Co.

Structural identifiability of polynomial differential systems

Gabriella Margaria
Dipartimento di Matematica, Politecnico di Torino, Italy.
Email: margaria@calvino.polito.it

GROSTAT III, 10 September 1999

We consider differential polynomial model

$$\begin{cases} x'(t,p) &= f(x(t,p),u(t),p) \\ y(t,p) &= g(x(t,p),u(t),p) \\ x(0,p) &= x_0(p) \end{cases} \quad (1)$$

with $x(t,p) \in \mathbf{R}^n$ and $y(t,p) \in \mathbf{R}^m$, each component of f and of g a polynomial in the indicated variables. We assume that the parameter vector p belongs to an open set $\Omega \subset \mathbf{R}^l$.

Taylor series problem The Taylor series approach for the structural identifiability of the model (1) is based on the calculation of successive time derivatives of $y(t,p)$ evaluated at $t = 0$. Successive derivatives are measurable and contain information about the parameters to be identified. The parameter vector p is unidentifiable if the set of solutions is uncountable, is locally identifiable if the set of solutions is finite, and is globally identifiable if there is a unique solution.

Example:

$$\begin{cases} x_1' &= p_1 x_1^2 + p_2 x_1 x_2 \\ x_2' &= p_3 x_1^2 + p_4 x_1 x_2 \\ y &= x_1 \\ x_1(0) &= 1 \\ x_2(0) &= 0 \end{cases}$$

The set of successive time derivatives:

$$y - 1,$$
$$y' - p_1,$$
$$y'' - 2p_1^2 - p_2 p_3,$$
$$y^{(3)} - 6p_1^3 - 6p_1 p_2 p_3 - p_3 p_4 p_2,$$
$$y^{(4)} - 5p_3^2 p_2^2 - 24p_1^4 - 36p_1^2 p_2 p_3 - 9p_1 p_3 p_4 p_2 - p_3 p_4^2 p_2,$$
$$y^{(5)} - 120p_1^5 - 72p_1^2 p_3 p_4 p_2 - 13p_1 p_3 p_4^2 p_2 - 78p_1 p_3^2 p_2^2 - 240p_1^3 p_2 p_3 - p_3 p_4^3 p_2 - 17p_3^2 p_2^2 p_4,$$
$$\vdots$$

A few upper bounds have been published:

$2n - 1$ for linear systems,

$2^{2n} - 1$ for bilinear systems,

$\frac{q^{2n} - 1}{q - 1}$ for homogeneous polynomial systems.

Problem

Find a upper bound on the number of successive derivatives of $y(t,p)$ needed for the identification analysis.

Solution

Use differential algebra.

A *ranking* of $(x_1, \ldots, x_n)$ is a total ordering of the set of all derivatives $x_i^{(j)}$ such that

$$x_i^{(j)} < x_i^{(j+k)}$$

$$x_i^{(m)} < x_j^{(l)} \Longrightarrow x_i^{(m+k)} < x_j^{(l+k)}.$$

The *leader* v of a differential polynomial f is the largest derivative with respect to the ranking which appears in f.

Let f and g be two differential polynomials and let v^d be the rank of f.

$$f = I_d v^d + \ldots$$

If no proper derivative of v appears in g then g is *partially reduced* with respect to f .

If g is partially reduced with respect to f and its degree in v is less than d then g is *reduced* with respect to f

A set of differential polynomials $\mathcal{A}$ is called *autoreduced* if each element of $\mathcal{A}$ is reduced with respect to all the others.

$\mathcal{A}$ is a *characteristic set* of E if E does not contain any non-zero element reduced with respect to $\mathcal{A}$.

An example:

$$\begin{cases} x_1' &= p_1 x_1^2 + p_2 x_1 x_2 \\ x_2' &= p_3 x_1^2 + p_4 x_1 x_2 \\ y &= x_1 \end{cases}$$

The set

$$x_1' - p_1 x_1^2 - p_2 x_1 x_2, x_2' - p_3 x_1^2 - p_4 x_1 x_2$$
$$p_1', p_2', p_3', p_4', y - x_1$$

is a characteristic set with respect to the ranking

$$y < \ldots < x_1 < x_2 < p_1 < p_2 < p_3 < p_4 < \ldots$$

Let I be the differential ideal generated by the following differential polynomials:

$x'(t,p) - f(t,x,p,u)$, p', $y(t,p) - g(t,x,p,u)$
in $\mathbf{R}\{u,x,y,p\}$.

Let L_f be the *generalized Lie-derivative operator* with respect to f:

$$L_f = \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i} + \sum_{i=1}^{\infty} u^{(i)} \frac{\partial}{\partial u^{(i-1)}}$$

Given $e_1, \ldots, e_m$, non-negative integers, let K be the ring

$$\mathbf{R}[u, u', \ldots, y_1, \ldots, y_1^{(e_1)}, \ldots, y_m, \ldots, y_m^{(e_1)}, x, p]$$

and let K_0 be the ring

$$K_0 = \mathbf{R}[u, u', \ldots, y_1, \ldots, y_1^{(e_1)}, \ldots, y_m, \ldots, y_m^{(e_m)}, p].$$

Let ν be the morphism from K to K_0 which maps x to $x(0)$ and leaves the other variables fixed.

Let $\overline{J}(e_1, \ldots, e_m)$ be the ideal generated by

$$\nu(y_i - g_i), \ldots, \nu(y_i^{(e_i)} - L_f^{e_i} g_i), \qquad (2)$$

for $i = 1, \ldots, m$. Thus one obtains the set of equations in the parameters whose solution set is to be found out.

Theorem 1 *Let*

$$\begin{array}{l} \{A_1(u,y), \ldots, A_m(u,y), \\ B_1(u,y,p), \ldots, B_l(u,y,p), \\ C_1(u,y,p,x), \ldots, C_n(u,y,p,x)\} \end{array} \qquad (3)$$

be a characteristic set of I with respect to the ranking $u << y << p < x$ and $y_i^{(e_i)}$ the leader of A_i.

Let T be a characteristic set of $\overline{J}(e_1, \ldots, e_m)$ with respect to the ranking $u << y << p$, X the set of ranks and Y the set of leaders of T.

Then three situations can arise:

i. If there exists an i such that the parameter p_i is not a leader of T, i.e. $p_i \notin Y$, then the model is not structurally identifiable for any p.

ii. If all parameters p_i are ranks of T, i.e. $p_i \in X$, $i = 1, \ldots, l$, then the model is globally structurally identifiable at p.

iii. If all parameters p_j are in Y, but p_i is not in X, for some i, then the model is locally structurally identifiable at p_i.

Theorem 2 *For $m = 1$ the number of equations that need to be taken into consideration for the Taylor series approach is $n + l + 1$, i.e. it is sufficient to consider $\overline{J}(n + l)$.*

Example 1

$$\begin{cases} x_1' &= p_1 x_1^2 + p_2 x_1 x_2 \\ x_2' &= p_3 x_1^2 + p_4 x_1 x_2 \\ y &= x_1 \\ x_1(0) &= 1 \\ x_2(0) &= 0 \end{cases}$$

We know that $e_1 = 4$. We compute a characteristic set with respect to the ranking

$$y < y' < y'' < y^{(3)} < y^{(4)} < p_1 < p_2 < p_3 < p_4$$

of the ideal $\overline{J}(4)$ generated by:

$$\begin{aligned} &y - 1, y' - p_1, \\ &y'' - 2p_1^2 - p_2 p_3, \\ &y^{(3)} - 6p_1^3 - 6p_1 p_2 p_3 - p_3 p_4 p_2, \\ &y^{(4)} - 5p_3^2 p_2^2 - 24p_1^4 - 36p_1^2 p_2 p_3 \\ &\quad -9p_1 p_3 p_4 p_2 - p_3 p_4^2 p_2 \end{aligned}$$

A characteristic set is

$$y - 1,$$

$$(-y'' + 2y'^2)y^{(4)} + 30y''y'^4 - 12y''^2y'^2 + 5y''^3 - 16y'^6 - 3y^{(3)}y''y' - 6y^{(3)}y'^3 + y^{(3)2},$$

$$y' - p_1,$$

$$-y'' + p_3 p_2 + 2y'^2$$

and

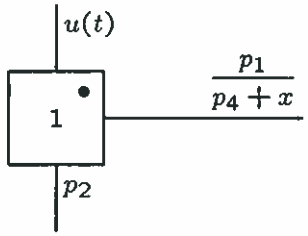
$$(-y'' + 2y'^2)p_4 + y^{(3)} + 6y'^3 - 6y''y'$$

As we do not find a polynomial with leader p_2, thus we conclude, by Theorem 1 that the model is unidentifiable.

The polynomial models don't cover the whole range of models of real interest, e.g., models where the coordinate function f are rational functions have been discarded.

Example 2 Now we consider the model of Figure 1

Figure 1:



$$x' = -\frac{p_1 x}{p_4 + x} - p_2 x + u(t)$$

The experiment considered is $u = 0$, $x(0) = a$.

$$\begin{cases} x'(t,p) &= \dfrac{r(x,p,u)}{q(x,p,u)} \\ y(t,p) &= g(x(t,p), u(t), p) \\ x(0,p) &= x_0(p) \end{cases}$$

We introduce the new state variable

$$v(t,p) = \frac{1}{q(x,p,u)}$$

The state vector $\overline{x} = (x, v)$ satisfies a differential equation $\dot{\overline{x}} = \overline{r}(x,p,u)$ where each component $\overline{r}_i$ is a polynomial in the variables x, v, p, u.

Let $x_1 = x$ and $x_2 = \dfrac{1}{p_4 + x_1}$

$$\begin{cases} x_1' = -p_1 x_1 x_2 - p_2 x_1 \\ x_2' = p_1 x_1 x_2^3 + p_2 x_1 x_2^2 \\ x_1(0) = a \\ x_2(0) = p_3 \end{cases}$$

A characteristic set of

$$\nu(y^{(i)} - L_f^i x_1), \quad i = 0, \ldots, 3$$

is composed by

$$y^{(0)} - a$$

$$\begin{aligned}(&-2(y^{(1)})^4 a + 2a^2 y^{(2)}(y^{(1)})^2)p_3 - a^2(y^{(2)})^2 \\ &+ (y^{(1)})a^2(y^{(3)}) - 2a(y^{(2)})(y^{(1)})^2 + 2(y^{(1)})^4\end{aligned}$$

$$\begin{aligned}(&-y^{(1)}a^2 y^{(3)} - 2(y^{(1)})^4 + 2a(y^{(2)})(y^{(1)})^2 \\ &+ a^2(y^{(2)})^2)ap_2 + a(-2y^{(2)}(y^{(1)})^3 \\ &+ 3a(y^{(2)})^2 y^{(1)} - (y^{(1)})^2 a y^{(3)})\end{aligned}$$

$$\begin{aligned}((&y^{(1)})^2 a^4 (y^{(3)})^2 - 2y^{(1)}a^4 y^{(3)}(y^{(2)})^2 \\ &- 4(y^{(1)})^3 a^3 y^{(3)} y^{(2)} + 4(y^{(1)})^5 a^2 y^{(3)} \\ &- 8ay^{(2)}(y^{(1)})^6 + 4(y^{(1)})^8 + 4a^3(y^{(2)})^3(y^{(1)})^2 \\ &+ a^4(y^{(2)})^4)p_1 + 12a^2(y^{(2)})^2(y^{(1)})^5 \\ &- 12ay^{(2)}(y^{(1)})^7 - 4a^3(y^{(2)})^3(y^{(1)})^3 + 4(y^{(1)})^{\text{[illegible]}}\end{aligned}$$

The model is globally identifiable.

Example 3 The non linear differential equations

$$\begin{cases} x_1' = \dfrac{p_1 x_2}{p_5 + x_2} x_1 - p_3 x_1 \\ x_2' = -p_4 \dfrac{p_1 x_2}{p_5 + x_2} x_1 \\ x_1(0) = a \\ x_2(0) = b \end{cases}$$

describe microbial growth in a batch reactor where x_1 is the concentration of micro-organisms, x_2 is the concentration of growth-limiting substrate, $\dfrac{p_1 x_2}{p_5 + x_2}$ is overall velocity of the reaction, p_1 is the maximum velocity, p_3 is the decay rate coefficient, p_4 is the inverse of the yield coefficient and p_5 is the Michaelis-Menten constant. Both x_1 and x_2 are observed and the identifiability analysis of the parameters $\{p_1, p_5, p_3, p_4\}$ is required. We introduce the new state variable $x_3 = \dfrac{1}{p_5 + x_2}$. The system becomes:

$$\begin{cases} x_1' = p_1 x_1 x_2 x_3 - p_3 x_1 \\ x_2' = -p_1 p_4 x_1 x_2 x_3 \\ x_3' = p_1 p_4 x_1 x_2 x_3^3 \\ x_1(0) = a \\ x_2(0) = b \\ x_3(0) = p_2 \end{cases}$$

where $p_2 = \dfrac{1}{p_5 + b}$. Only x_1 and x_2 are observed and the parameter set is $\{p_1, p_2, p_3, p_4\}$.

We compute a characteristic set of the set

$$\begin{aligned}&\nu(y_1 - x_1), \nu(y_1^{(1)} - L_f x_1), \nu(y_1^{(2)} - L_f^2 x_1), \\ &\nu(y_2 - x_2), \nu(y_2^{(1)} - L_f x_2), \nu(y_2^{(2)} - L_f^2 x_2)\end{aligned}$$

with respect to the ranking

$$\begin{aligned}y_1 < y_2 < y_1^{(1)} < y_2^{(1)} < y_1^{(2)} < y_2^{(2)} < \\ p_1 < p_2 < p_3 < p_4.\end{aligned}$$

We find the following 6 differential polynomials

$$-y_1 + a$$

$$-y_2 + b$$

$$\begin{aligned}(&a^2 b(y_2^{(2)})^2 - 2aby_2^{(2)}y_2^{(1)}y_1^{(1)} - a^2(y_2^{(1)})^2 y_2^{(2)} \\ &+ a(y_2^{(1)})^3 y_1^{(1)} + (y_2^{(1)})^2 b(y_1^{(1)})^2)p_1 \\ &+ (y_2^{(1)})^3 a y_1^{(2)} - (y_2^{(1)})^3 (y_1^{(1)})^2\end{aligned}$$

$$aby_2^{(2)} + ap_2(y_2^{(1)})^2 b - a(y_2^{(1)})^2 - y_2^{(1)} b y_1^{(1)}$$

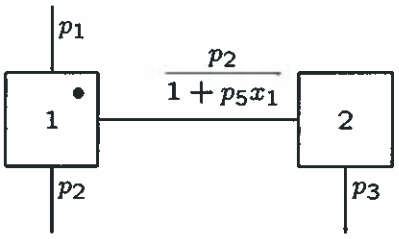
$$\begin{aligned}&a(y_2^{(1)})^2(ay_2^{(2)} - y_2^{(1)}y_1^{(1)})p_3 \\ &+ a(y_2^{(1)})^2(y_2^{(2)}y_1^{(1)} - y_2^{(1)}y_1^{(2)})\end{aligned}$$

$$(y_2^{(1)})^3(ay_1^{(2)} - (y_1^{(1)})^2)p_4$$
$$+ (y_2^{(1)})^3(ay_2^{(2)} - y_2^{(1)}y_1^{(1)})$$

Then the model is globally identifiable.

Example 4 The following equations correspond to a two-compartmental system with the flow from compartment 1 into compartment 2 Michaelis-Menten saturative

Model 2



$$\begin{cases} x_1' = -p_1x_1 - \dfrac{p_2x_1}{1+p_5x_1} \\ x_2' = \dfrac{p_2x_1}{1+p_5x_1} - p_3x_2 \\ x_1(0) = 1 \\ x_2(0) = 0 \end{cases}$$

By introducing the new state variable $x_3 = \dfrac{1}{1+p_5x_1}$ the system becomes:

$$\begin{cases} x_1' = -p_1x_1 - p_2x_1x_3 \\ x_2' = p_2x_1x_3 - p_3x_2 \\ x_3' = p_1x_1x_3^2 - p_2x_1x_3^3 \\ x_1(0) = 1 \\ x_2(0) = 0 \\ x_3(0) = p_4 \end{cases}$$

where $p_2 = \dfrac{1}{1+p_5}$. Only compartment 1 is observed and the set of parameters is $\{p_1, p_2, p_3, p_4\}$. A characteristic set of

$$\nu(y^{(i)} - L_f^i x_1), \quad i = 0, \ldots, 7$$

gives a differential polynomial of degree 3 in the parameter p_1, one linear in the parameter p_2 and one linear in p_4

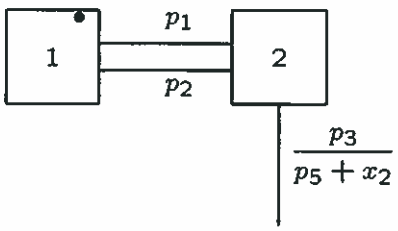
$$(-y^{(2)}+(y^{(1)})^2)p_2+(y^{(1)})^3+2(y^{(1)})^2p_1+y^{(1)}p_1^2$$

$$((y^{(1)})^3 + 2(y^{(1)})^2p_1 + y^{(1)}p_1^2)p_4$$
$$+ y^{(2)}y^{(1)} + y^{(2)}p_1 - (y^{(1)})^3 - (y^{(1)})^2p_1.$$

None of the differential polynomials contains the parameters p_3. This completes the identification. The model is not identifiable.

Example 5 Let us consider

Model 3:



$$\begin{cases} x_1' = -p_1x_1 + p_2x_2 \\ x_2' = p_1x_1 - \dfrac{p_3x_2}{p_5+x_2} - p_2x_2 \\ x_1(0) = a \\ x_2(0) = 0 \end{cases}$$

We rewrite the system as:

$$\begin{cases} x_1' = -p_1x_1 + p_2x_2 \\ x_2' = p_1x_1 - p_2x_2 - p_3x_2x_3 \\ x_3' = -p_1x_1x_3^2 + p_2x_2x_3^2 + p_3x_2x_3^3 \\ x_1(0) = a \\ x_2(0) = 0 \\ x_3(0) = p_4 \end{cases}$$

where $p_2 = \dfrac{1}{p_5}$ and $x_3 = \dfrac{1}{p_5+x_2}$. Only compartment 1 is observed and the set of parameters is $\{p_1, p_2, p_3, p_4\}$. A characteristic set of

$$\nu(y^{(i)} - L_f^i x_1), \quad i = 0, \ldots, 5$$

contains

$$y^{(1)} + p_1a$$

$$ay^{(2)} + y^{(1)}ap_2 - (y^{(1)})^2$$

$$y^{(1)}a^3(-(y^{(1)})^3y^{(4)} - (y^{(2)})^3y^{(1)} + 2(y^{(1)})^2y^{(3)}y^{(2}$$
$$- y^{(1)}a(y^{(3)})^2 + ay^{(2)}y^{(1)}y^{(4)})p_3$$
$$+ y^{(1)}a^3(-4y^{(1)}ay^{(3)}(y^{(2)})^2 + 2(y^{(1)})^2a(y^{(3)})^2$$
$$+ 2a(y^{(2)})^4)$$

$$y^{(1)}a^4(-2(y^{(1)})^4(y^{(3)})^2 + 2a(y^{(1)})^2(y^{(3)})^2y^{(2)} + 4(y^{(1)})^3y^{(3)}(y^{(2)})^2 - 4ay^{(1)}y^{(3)}(y^{(2)})^3 - 2(y^{(1)})^2(y^{(2)})^4 + 2a(y^{(2)})^5)p_4 + y^{(1)}a^4(y^{(4)}(y^{(1)})^3y^{(3)} - y^{(4)}(y^{(1)})^2(y^{(2)})^2 - 2(y^{(3)})^2y^{(2)}(y^{(1)})^2 + 3y^{(3)}(y^{(2)})^3y^{(1)} + y^{(1)}a(y^{(3)})^3 - a(y^{(3)})^2(y^{(2)})^2 - y^{(1)}ay^{(2)}y^{(4)}y^{(3)} + a(y^{(2)})^3y^{(4)} - (y^{(2)})^5)$$

The model is globally identifiable.

Now we suppose that compartment 2 is observed.

It was possible to decide only local identifiability with this method.

Some reflections on the algebra of stochastic calculus

W. S. Kendall
Statistics, Warwick

www.warwick.ac.uk/statsdept/Staff/WSK/

* Work in progress

... SD Jacka

1. Symbolic Itô calculus

* 2. Playing with algebra for diffusions

* 3. Branched semimartingales

4. Zeros of Brownian polynomials

www.warwick.ac.uk/statsdept/Staff/WSK/

Semimartingale

Infinitesimal level:

$$\Delta X = \underbrace{m\,\Delta t}_{\text{drift / signal}} \pm \underbrace{\sqrt{v\,\Delta t}}_{\text{volatility}}$$

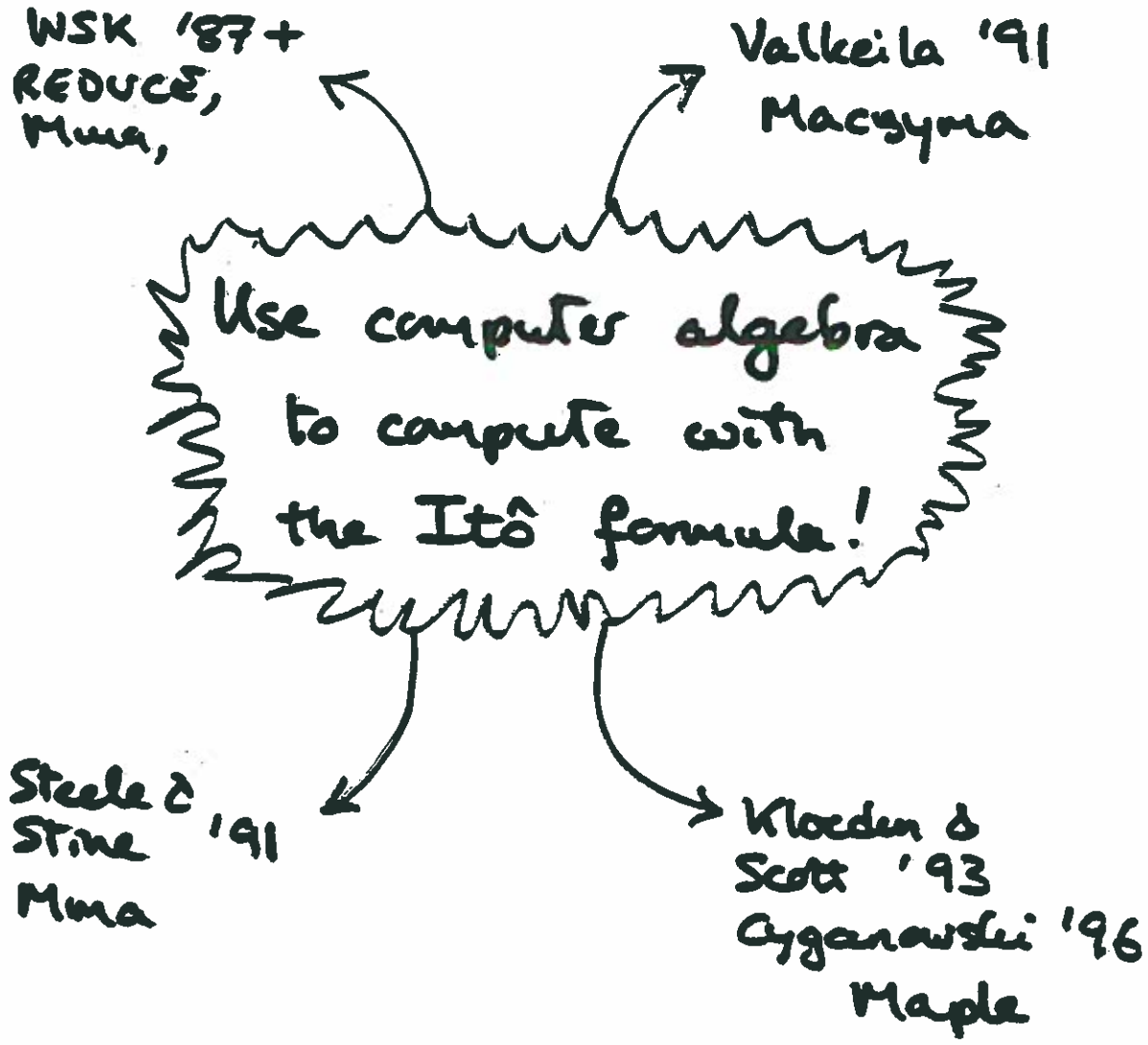
random

Continuous-time level:

$$X = \underbrace{V}_{\substack{\text{locally} \\ \text{bdd} \\ \text{variation}}} + \underbrace{M}_{\substack{\text{local} \\ \text{martingale}}}$$

1. The heart of stochastic calculus

$$df(X) = f'(X)\,dX + \tfrac{1}{2} f''(X)\,dX^2$$



Algebraic formalism (Itô, '75)

X, $\int f dX$ — semimartingales $\mathcal{A}$

$\xrightarrow{d}$

$\xleftarrow{\text{integration}}$

dX, $f dX$ — stochastic differentials $\mathcal{D}_2$

$\mathcal{D}_2$ is <u>module</u> (over $\mathcal{A}$) and <u>ring</u>

multiplication: $\mathcal{D}_2 \times \mathcal{D}_2 \to \mathcal{D}_1$

$dX, dY \mapsto dXdY$

drift: $\mathcal{D}_2 \to \mathcal{D}_1$

$dX \mapsto \text{drift}(dX)$

direct implementation of algebraic structures in AXIOM.

... WSK '99

http://www.warwick.ac.uk/statsdept/Staff/WSK/ca.html

- REDUCE
- Mma
- AXIOM

Computer algebra & stochastic calculus:-

- finding good representations;
- identifying diffusion geometry;
- searching for supermartingales;
- solving (some) s.d.e.;
- preparation of simulation code.

Fragment from AXIOM Itovsn3.

time semimartingale

```
dt := introduce! (t, dt);
classical SD! (dt);
```

Brownian motion

```
dB := introduce! (B, dB);
brownian SD! (dB, dt);
```

elicit details of $\int_0^T \exp\left(B(T)-B(u)-\frac{\nu}{2}(u-T)\right) du$

```
Y := exp(B + v*t/2) *
        itoIntegral (-dB - v*dt/2)
```

hence, using Itô formula, identify it as X where X solves an S.d.e:

```
dX := introduce!(X, dX);
itoSDE!(
dX = X*(dB - v*dt/2 + dt/2) + dt);
```

2. Searching for invariant varieties

Polynomial Itô diffusion (X, Y):

- First-order or drift structure:

BSD	dX	dY	dt
$Drift$	$-\frac{X}{2}\,dt$	$-\frac{Y}{2}\,dt$	dt

- Second-order structure:

$$dX^2 = \left(Y^4 - 2\,X\,Y^3 + \left(3\,X^2 - 2\right)Y^2 + \left(-2\,X^3 + 2\,X\right)Y + 2\,X^4 - 4\,X^2 + 2\right)dt$$
$$dX\,dY = \left(-Y^4 + X\,Y^3 + \left(-X^2 + 2\right)Y^2 + \left(X^3 - 2\,X\right)Y + X^2 - 1\right)dt$$
$$dY^2 = \left(Y^4 + \left(X^2 - 2\right)Y^2 + 1\right)dt$$

Simple AXIOM computations, leading to vanishing normal forms.

```
v := [[coefficient(du*dv,dt) for du in [dX,dY]]
                             for dv in [dX,dY]]
V0 := (determinant v)::POLY INT
V1 := coefficient(drift d V0,dt)::POLY INT
V2 := coefficient((d V0)^2,dt)::POLY INT

H:= gcd(V0,gcd(V1,V2))
normalForm(coefficient(drift d H,dt)::POLY INT,[H])
normalForm(coefficient((d H)^2,dt)::POLY INT,[H])
```

Hence H defines an invariant variety:

$$H \quad = \quad (X^2 - Y^2 - 1)^2$$

TOWARDS FINDING INVARIANT SUBSETS:

X diffusion solving martingale problem:

$$f(X) - \int \left\{ \sum_i m^i \frac{\partial f}{\partial x^i} + \frac{1}{2} \sum_{ij} v^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j} \right\} dt$$

is martingale for smooth f of compact support.

Suppose polynomial drift m
volatility v

Compute invariant varieties

- cf pde theory
- work in progress
- specific probabilistic questions
- strong implications of being defined everywhere

Target output:

Either

- no invariant manifolds;
- fibration given by algebraic data;
- union of invariant algebraic varieties on each of which algorithm may be repeated.

Work in progress ...

- Trivial answer if $\det(v) > 0$ everywhere;

- Reduce to sub-variety if $\det(v)$ nontrivial and $\det(v) = 0$ somewhere;

  (Can also use $\det(v + s \cdot \text{identity})$)

* ISSUE: deal with situations such as hypoellipticity

(eg X, Y, $\int X dY - Y dX$)

— Lie algebra ✓

— Function-Theoretic ?

* also issue of non-polynomial fibrations ...

Identifiability for s.d.e?

$$dx_1 = P_1(x_1, x_2, \ldots)\,dt + \varepsilon\, dW_1$$

$$dx_2 = P_2(x_1, x_2, \ldots)\,dt + \varepsilon\, x_2\, dW_2$$

$$\ldots$$

added noise

$\dot{x}_1, \ddot{x}_1, \ldots$ doesn't make sense anymore

observe paths not derivatives

CONJECTURE: for "nice" sde as above, identifiability properties inherited from a deterministic system

because of Cameron-Martin-Girsanov formula?

3. Formal computations:

Useful if $X \cdot Y \equiv 0$

implies $X \equiv 0$ or $Y \equiv 0$.

(integral domain)

Let $P = P(B_0, B_1, \ldots, B_{n+1})$ be generated by $B_0 = t$ and $n+1$ independent BM's

$$B_1, \ldots, B_{n+1}$$

under sums, multiplication, stochastic integration.

Then P is an integral domain

Reason:

- Alternative A: "Almost sure eternal law"

  $P[X(t) = X(T) \text{ for } T \le t \le T+\varepsilon] > 0$
  forces $X \equiv 0$ if $X \in \mathcal{P}$.

  use Doob-Meyer decomposition, induction on "chaotic complexity".

- Alternative B:

  $P[X(T+\varepsilon) = X(T)] > 0$
  forces $X \equiv 0$ if $X \in \mathcal{P}$.

  use potential theory and (again) induction

Similar result for differentials:

$X\,dY \equiv 0$ on an open stochastic interval
forces $X \equiv 0$ or $dY \equiv 0$

Algebraic perspective on semimartingales:

Field-of-fractions:

$R = \{X/Y : X, Y \in P,\ Y \neq 0\}$

"projective semimartingale" uniqueness

Algebraic closure:
given a polynomial, eg

$P(z) = B_1 z^n + B_2 z^{n-1} + \dots + B_n z + B_{n+1}$,

extend R to $A = A(z)$ using solutions z of $P(z) \equiv 0$.

"branched semimartingale"

typically not semimartingales.

partial uniqueness

Algebra / Galois theory ...

Differential closure:

example

closure under integration

$$Z = \int X \, dY$$

for $X, Y \in \mathcal{X}$, which is to say branched semimartingales.

We can no longer expect global definition: consider

$$Z = \int \frac{dt}{1-t}$$

So use sheaf theory, localized by stochastic time intervals (stopping times ...)

Prospect: stochastic differential Calcis Theory ...

4. Zeros of Brownian polynomials.

Example of "branched semimartingale":

for $P(z) = B_1 z^n + \ldots + B_{n+1}$ as above, away from a repeated root

$$dZ = -\sum_{r=1}^{n+1}{}' \frac{z^{n+1-r}}{P'(z)}\, dB_r + \sum_{r=1}^{n+1}{}' \frac{z^{2(n+1-r)}}{P'(z)^2}\left(\frac{n+1-r}{z} - \frac{P''(z)}{2P'(z)}\right) dt,$$

while different roots only collide in real / complex-conjugate pairs ("dance of the roots").

WSK + CJ Price, 1999.

Follow "preprint" links from
http://www.warwick.ac.uk/staffdept/Staff/WSK

Fragments of Theory:

- $B_1 z^n + B_2 z^{n-1} + \dots + B_n z + B_{n+1}$ is irreducible over P (Eisenstein criterion)
- So is $z^n + B z^{n-1} + B^2 z^{n-2} + \dots + B^n$ if $n+1$ is prime. Notice that all roots collapse to 0 when B hits 0.
- Almost sure eternal law delivers similar consequence for factorizations of semimartingale polynomials.
- Options for continuation through coincidence of roots:
  - ad hoc deterministic;
  - inject extra randomness.

Perspectives

- a framework for computational algebra applied to stochastic calculus

- stochastic algebraic geometry?

http://www.warwick.ac.uk/statsdept/Staff/WSK