Revisiting Morris method: A polynomial algebra for design definition with increased efficiency and observability

J.-M. Fédou, G. Menez, L. Pronzato, M.-J. Rendas

I3S Laboratory, CNRS/UNS (Sophia Antipolis, France)

Partially supported by ANR DESIRE
(Designs for Spatial Random Fields)

SαMφ, Nice, July 2012

Plan

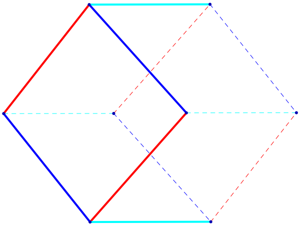
We'll be looking at two related problems

Problem 1

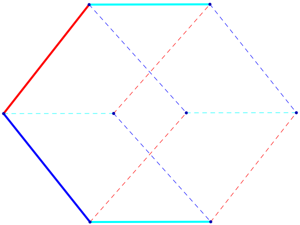
Find subgraphs $G_m^d \subset Q_d$ of the d-dimensional hypercube with the property:

$\forall i \in \{1, \ldots, d\}$, the number of edges joining nodes that **differ only in the i-th coordinate** is equal to m.

We say that graphs with this property are (d, m)-edge equitable.



$(3, 2)$-edge equitable



Not $(3, m)$-edge equitable

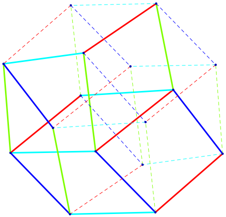
Q_3

Problem 2

Find edge equitable subgraphs $H_c^d \subset Q_d$ of the d-dimensional hypercube with the property:

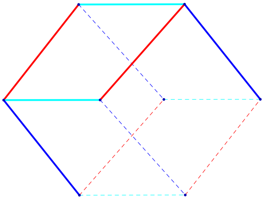
$$\forall i \neq j \in \{1, \ldots, d\}, \text{ the number of cycles in coordinates } i, j \text{ is equal to } c.$$

We say that graphs with this property are (d, c)-cycle equitable.



$(4, 1)$-cycle equitable

(i, j)	2	3	4
1	1	1	1
2		1	1
3			1



not cycle equitable

(i, j)	2	3
1	1	0
2		0

Motivation

Morris elementary effects screening method for sensitivity analysis (Technometrics, 1991)

Commonly used screening method for analysis of $f : \mathbb{R}^d \to \mathbb{R}$

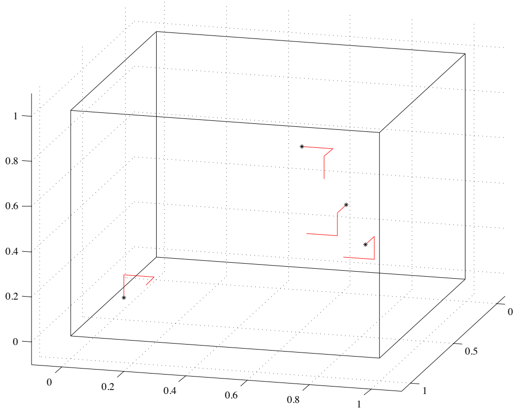
- Partitions input factors into *linear*, *negligible* and *non-linear/mixed*
- Makes no assumptions about f
- Simple (linear in the number of inputs), OAT global method.

Based on statistical analysis of

Elementary effect along direction $i \in \{q, \dots, d\}$

$$d_i(y) \triangleq \frac{1}{\Delta}\left[f(y + \Delta e_i) - f(y)\right], \quad i \in \{1, \dots, d\}$$

Standard Morris method



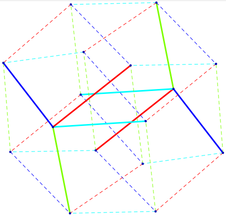
OAT method:
a complete set of d elementary effects is computed along a trajectory contained in a scaled and translated version of Q_d

Our work is concerned with

Morris clustered designs

Design matrices that allow computation of $m > 1$ elementary effects along each direction (i.e., each evaluation of f is used to compute a larger number of d_i's).

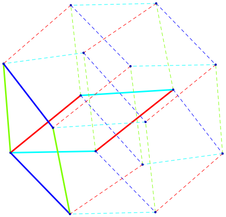
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$



10 points in Q_4

(4, 2)-equitable subgraphs

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



7 points in Q_4

Why coming back to the problem?

Shortcomings of Morris clustered construction

- not guided by m
- cannot yield all possible values of m
- factored version (the most efficient) defined only when d is not prime
- definition in the paper is not always *equitable*
- minimality of the size of the designs (efficiency) is not guaranteed.

Our contribution

Constructive algorithm for generation of the clustered designs of Morris method guided by the target value of m and the dimension d of the input space

- Handles generic values of (d, m).
- Proovably equitable designs.
- For pairs (d, m) for which Morris construction is defined, leads to designs of the same complexity.

Why studying problem 2?

Extend Morris Elementary Effects method to (cross) derivatives of second order

Elementary mixed-effects along directions $i, j \in \{1, \ldots, d\}$

$$d_{ij}^{(2)}(y) = \frac{1}{\Delta}[d_i(y + \Delta e_j) - d_i(y)], \qquad i \in \{1, \ldots, d\}$$

Previous work
The new Morris Method, Campolongo & Braddock (Reliability Engineering and System Safety, 1999) : only defined for $c = 1$, less efficient designs than ours and no complete algorithmic construction.

How do we do it?

Two basic ideas

1. (d, m)-edge and (d, c)-cycle equitable subgraphs are recursively generated, by combining smaller equitable solutions (for smaller values of d, and m or c)
2. use a polynomial representation to manipulate subgraphs and prove their properties

Plan

Polynomial representation of subgraphs of Q_d

Coding points of Q_d by monomials

$$s = \{s_1, s_2, \ldots, s_d\} \longrightarrow \mathcal{P}_s(X_1, X_2, \ldots, X_d) = X_1^{s_1} X_2^{s_2} \ldots X_d^{s_d}$$

Example

$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \in Q_5 \rightarrow X_2 X_3 X_5 \quad \in K(X_1, \ldots, X_5) = K_5$$

Coding subgraphs of Q_d by polynomials

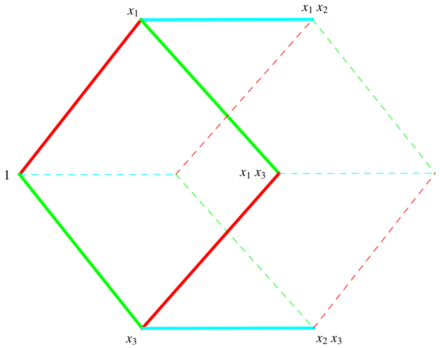
$$G \subset Q_d \rightarrow \mathcal{P}_G = \sum_{s \in G} \mathcal{P}_s$$

$\mathcal{P}_G$: degree at most one in each variable, coefficients in $\{0, 1\}$.

Polynomial representation of subgraphs of Q_d

Example

$$P = 1 + x_1 + x_3 + x_1x_2 + x_1x_3 + x_2x_3 \subset Q_3$$



Edge coloring of Q_3:

- red : x_1
- cyan : x_2
- green : x_3

Polynomial representation of subgraphs of Q_d

Scalar product and structure

Definition of $\langle \cdot, \cdot \rangle$

$\mathcal{P}_s$, $\mathcal{P}_{s'}$ two monomials ($s, s' \in Q_d$)
Define the scalar product

$$\langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle = 1_{s=s'} \quad .$$

Extension to polynomials ($G, G' \subset Q_d$)

$$\langle \mathcal{P}_G, \mathcal{P}_{G'} \rangle = \sum_{s \in G, s \in G'} \langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle \quad .$$

Example

$$\langle X_1 X_2, X_1 X_2 \rangle = 1, \qquad \langle X_1 X_2, X_1 X_2 X_3 \rangle = 0$$

$$\langle 1 + X_1 + X_2 + X_1 X_2, 1 + X_1 X_2 + X_3 \rangle = 2$$

Properties

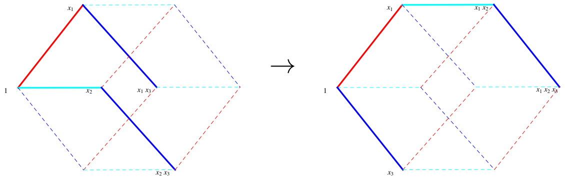
- $\langle P_G, P_{G'} \rangle = |G \cap G'|$
- $\langle P_G, P_G \rangle = |G|$

Algebra over the polynomials

- Addition $+ \Leftrightarrow$ graph sum (nodes multiplicity may be > 1)
- Multiplication is defined modulo $X_i^2 = 1, i \in \{1, \ldots, d\}$
  Multiplication of P_G by monomial $s = X_i \Leftrightarrow$ reflection of G along direction i

Example (X_1 corresponds to red edges)

$$\begin{aligned} X_1(1 + X_1 + X_2 + X_1X_3 + X_2X_3) &= X_1 + X_1^2 + X_1X_2 + X_1^2X_3 + X_1X_2X_3 \\ &= X_1 + 1 + X_1X_2 + X_3 + X_1X_2X_3 \end{aligned}$$



Problem reformulation in terms of polynomials

Facts:

1. edges of color i are preserved by multiplication by X_i. All other edges are moved elsewhere in Q_d
2. (remember that $|G \cap G'| = \langle P_G, P_{G'} \rangle$)
3. $\Rightarrow$ the number of edges of G of color i is exactly $2 \langle P_G, X_i P_G \rangle$
4. $\Rightarrow$ the number of cycles in G in colors i, j is exactly $4 |P_G \cap X_i P_G \cap X_j P_G \cap X_i X_j P_G|$

Problem 1 reformulation

Optimal (d, m)-edge equitable designs are the solutions of

$$P^\star = \underset{P \in K_d}{\arg\min} \langle P, P \rangle$$
$$\text{s.t.} \quad \langle P^\star, X_i P^\star \rangle = 2m, \qquad i \in \{1, 2, \ldots, d\}.$$

We drop minimality, and assess the simpler problem of finding small (d, m)-edge equitable designs (not necessarily minimal).

Problem reformulation in terms of polynomials

Facts:

1. edges of color i are preserved by multiplication by X_i. All other edges are moved elsewhere in Q_d
2. (remember that $|G \cap G'| = \langle P_G, P_{G'} \rangle$)
3. $\Rightarrow$ the number of edges of G of color i is exactly $2\,\langle P_G, X_i P_G \rangle$
4. $\Rightarrow$ the number of cycles in G in colors i, j is exactly $4\,|P_G \cap X_i P_G \cap X_j P_G \cap X_i X_j P_G|$

Problem 2 reformulation

Optimal (d, c)-cycle edge equitable designs are the solutions of

$$P^\star = \arg\min_{P \in K_d} \langle P, P \rangle$$

$$\text{s.t.} \quad |P_G \cap X_i P_G \cap X_j P_G \cap X_i X_j P_G| = 4c, \qquad i \neq j \in \{1, 2, \ldots, d\}.$$

As for Problem 1, we relax the minimality constraint.

Plan

Generation of (d, m)-edge equitable subgraphs of Q_d

Recursive (in m) algorithm

Initialisation

- $m = 1$, generic d

$$G_d^1 = 1 + \sum_{i=1}^{d} X_1 \cdots X_i \quad .$$

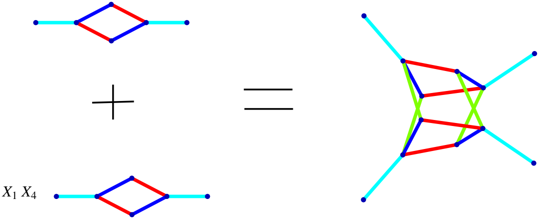
Generation of (d, m)-edge equitable subgraphs of Q_d

Induction

- m even

$$G_d^m = G_{d-1}^{\frac{m}{2}} + X_1 X_d G_{d-1}^{\frac{m}{2}}$$

Example: $G_4^4 = G_3^2 + X_1 X_4 G_3^2$



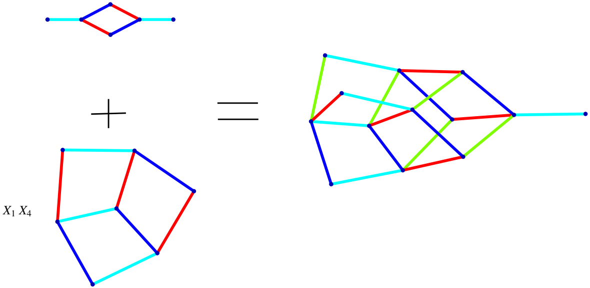
Generation of (d, m)-edge equitable subgraphs of Q_d

Induction

- m odd

$$G_d^m = G_{d-1}^{\frac{m-1}{2}} + X_1 X_d G_{d-1}^{\frac{m+1}{2}}$$

Example: $G_4^5 = G_3^2 + X_1 X_4 G_3^3$



Theorem

G_d^m are (d, m)-edge equitable

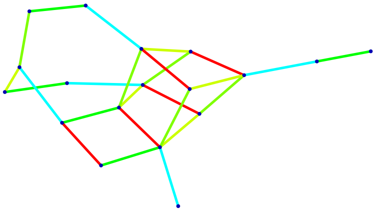
Proof: use properties of scalar product (requires a condition on the solutions for consecutive values of m that is guaranteed by the initialisation of the recursion)

Generation of (d, m)-edge equitable subgraphs of Q_d

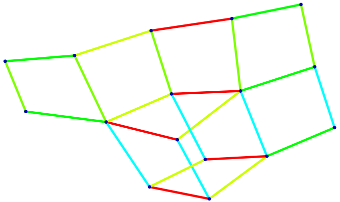
Topology and Initalisation

Other families of solutions can be obtained, by changing the initialization for small values of m

This has an impact on the topology (and on the complexity) of the resulting designs



G_5^5, Init $m = 1$ only



G_5^5, Init $m = 2, 3$

Factored (d, m)-equitable designs

Direct application of our algorithm leads to less efficient designs than Morris when they are defined.

Factored application of our generic solution

$$q_{\min}(m) \triangleq \lceil \log_2(m) \rceil + 1 \ ,$$

$$d = (c-1)q_{\min}(m) + r, \qquad r \in \{q_{\min}(m), \ldots, 2q_{\min}(m) - 1\} \ .$$

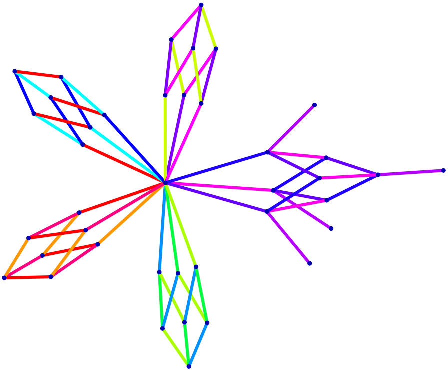
$$G_{Morris}(d, m) = G\left(q_{\min}, m\right) + \sum_{j=1}^{c-2} \left(\mathrm{Shift}_{jq_{\min}} G\left(q_{\min}, m\right) - 1\right) + \mathrm{Shift}_{(c-1)q_{\min}} G(r, m)$$

Fully-defined and provably edge-equitable version of the basic idea of Morris factored designs.

Factored (d, m)-edge equitable designs

Example

$\mathbf{G}_{\mathbf{17}}^{\mathbf{4}}$: 4 complete Q_3 ($X_1 \cdots X_3$, $X_4 \cdots X_6$, $X_7 \cdots X_9$, $X_{10} \cdots X_{12}$), together with G_5^4 (over $X_{13} \cdots X_{17}$)



Plan

Some notation

$$Line(X_1, \ldots, X_d) = \sum_{i=1}^{d} \prod_{j \leq i} X_j$$

$$Circle(X_1, \ldots, X_d) = Line(X_1, \ldots, X_d) + \left(\prod_{j=1}^{d} X_j \right) Line(X_1, \ldots, X_d)$$

$$Bubble((X_1, \ldots, X_d) = \textit{Polynomial in the } d \textit{ variables with 3 edges of each colour}$$

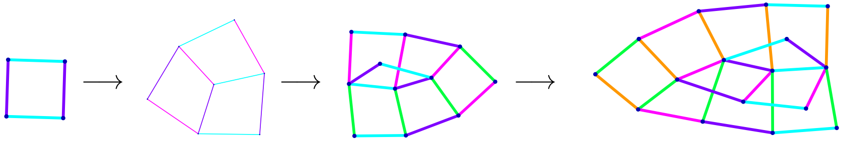
$(d, 1)$-cycle equitable subgraphs

Initialisation

For $d = 2$ and $c = 1$, define $H_2^1 = Q_2$

Induction

For $d > 2$ and $c = 1$, define $H_d^1 = H_{d-1}^1 + X_d\,(1 + Line(X_1, \ldots, X_{d-1}))$



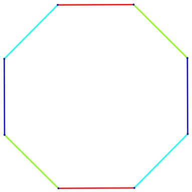
$(d,2)$-cycle and $(d,3)$-cycle equitable subgraphs (H_2^d, H_3^d)

Initialisation

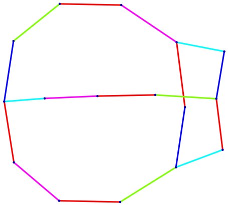
For $d = 3$ and $c = 2$, define $H_2^3 = Q_3$
For $d = 4$ and $c = 3$, define $H_3^4 = Q_4 - X_2 X_4$

Induction

For $d > 3$ and $c = 2$, define $H_2^d = H_2^{d-1} + X_d \mathit{Circle}(X_1, \ldots, X_{d-1})$
For $d > 4$ and $c = 3$, define $H_3^d = H_3^{d-1} + X_d \mathit{Bubble}(X_1, \ldots, X_{d-1})$



$\mathit{Circle}(4)$



$\mathit{Bubble}(6)$

Plan

Size of the design

- If initialization for $m = 1$,

$$\left|G_m^d\right| = m(d - \kappa) + 2^{\kappa+1} - m$$

  where $\kappa = \lfloor \log_2(m) \rfloor$.
- We derived a closed formula $|G_m^d|$ for initialization at $m = 2, 3$

$$|G_m^d| = c(m) + \alpha(m)d$$

  Size of factored solution is also known exactly.
- We also have a closed formula for $|H_c^d|$.

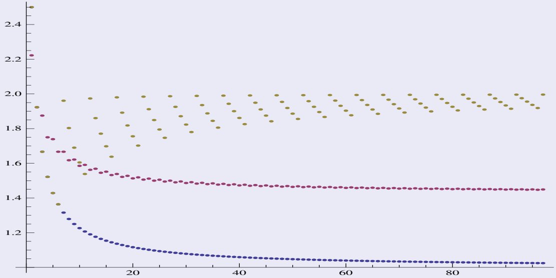
Economy

Definition

Morris index, ($|G_d^m|$ should be small $\Leftrightarrow \chi$ large)

$$\text{Economy: } \chi = \frac{\text{total \# elementary effects}}{|G_m^d|} = \frac{md}{|G_m^d|}$$

Economy of the (d, m)-edge equitable designs



Evolution of χ as d grows, $m = 10$.
Factored designs, designs with init G_1^d, and with init G_2^d, G_3^d.

Size of the (d, c)-cycle equitable designs

We obtain :

c	Nb Edges	Nb Points
1	d	$\frac{d^2+d+2}{2}$
2	$2d-4$	d^2-d+2
3	$3d-5$	$\frac{3d^2-7d+10}{2}$

For random designs and *New Morris* designs

c	Nb Edges	Nb Points
1	$2\binom{d}{2}$	$4\ \binom{d}{2}$
2	$4\binom{d}{2}$	$8\binom{d}{2}$
3	$6\binom{d}{2}$	$12\binom{d}{2}$

c	Nb Edges	Nb Points
1	not edge equitable	$4\ d^2-d+2$
2	$\star$	$\star$
3	$\star$	$\star$

Plan

Morris example function

$$f(x) = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{5} \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^{4} \beta_{ijls} w_i w_j w_l w_s$$

$w_i = 2X_i - 1, i \in \{1, 2, 4, 6, 8, \ldots, 20\}$, $w_i = 2.2X_i/(X_i + 0.1) - 1, i \in \{3, 5, 7\}$.

$$\begin{array}{llll} \beta_i = 20, & i \in \{1, \ldots, 10\}, & \beta_{ij} = -15, & i, j \in \{1, \ldots, 6\} \\ \beta_{ijl} = -10, & i, j, l \in \{1, \ldots, 5\}, & \beta_{ijls} = 5, & i, j, l, s \in \{1, \ldots, 4\}. \end{array}$$

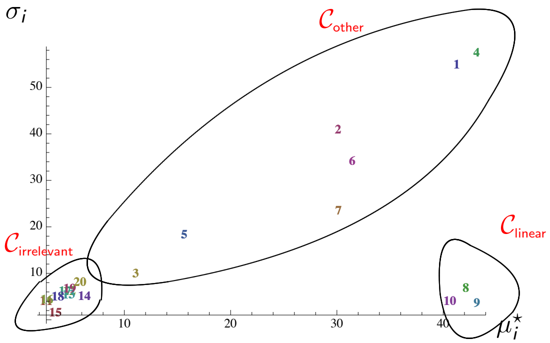
Remaining 1^{st} and 2^{nd} order coefficients are independent realisations of a standard normal distribution, $\beta_i \sim \mathcal{N}(0, 1)$, $i \notin \{1, \ldots, 10\}, \beta_{ij} \sim \mathcal{N}(0, 1)$, $i, j \notin \{1, \ldots, 6\}$. For this function the relevant classes of input factors are

$$\mathcal{C}_{\text{irrelevant}} = \{11, \ldots, 20\}, \qquad \mathcal{C}_{\text{linear}} = \{8, 9, 10\}, \qquad \mathcal{C}_{\text{other}} = \{1, \ldots, 7\}\ .$$

Note: X_7 is a purely non-linear term, while X_6 is an interaction factor.

Screening of Morris example function ($m = 4, r = 3$)

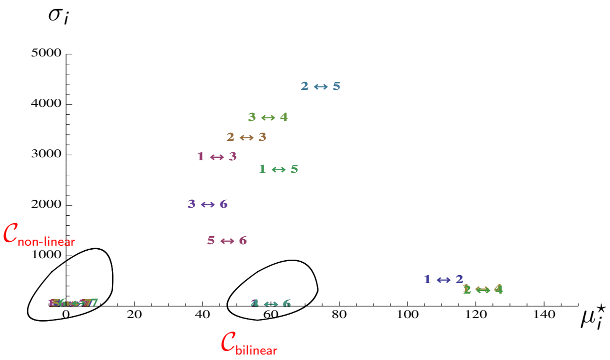
Total number of derivatives per direction: 12



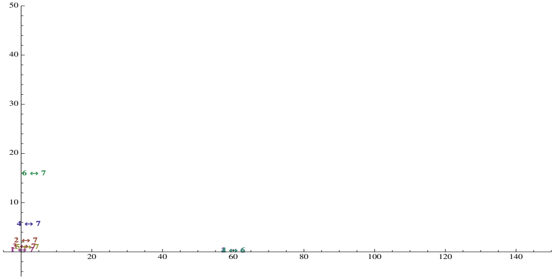
About **half the number of function evaluations** compared to $m = 1$.

Study of cross derivatives

Analysis concentrated on smaller class $\mathcal{C}_{\text{other}}$



(Zoom)



.

We detect that X_7 as a **non-linear factor with no interaction** with the other factors as well as the **bilinear term** $X_2 X_6$.

Plan

Up to now

1. Recursive algorithm for (d, m)-edge equitable graphs that completes the definition of clustered Morris designs
2. Recursive algorithm for (d, c)-cycle equitable graphs for $c = 1, 2, 3$ (can be exploited to build the skeleton of the FANOVA graph)
3. Explicit formulas for the size of the designs
4. Uses polynomial representation of subgraphs of Q_d and an appropriate definition of inner product as formal tools.
5. Polynomial representation enables direct identification of pairs of design points involved in the derivatives along each direction (or pairs of directions, for mixed effects).

Further work

Open issues ...

- minimality (of factored designs) ?
- effect of initialization
- relation to other classes of subgraphs of the hypercube (median graphs, mesh graphs,...)
- Generalize to subgraphs of $\{0, 1, \ldots, k\}^d$ for detection of higher order effects in each input factor

Generation of (d, m)-equitable subgraphs of Q_d

m even. Assume $G_{d-1}^{m/2}$ is $(d-1, m)$-equitable.

$$\left\langle G_m^d, X_i G_m^d \right\rangle = \begin{cases} \left\langle G_{d-1}^{\frac{m}{2}}, X_i G_{d-1}^{\frac{m}{2}} \right\rangle + \\ \qquad \left\langle X_1 X_d G_{d-1}^{\frac{m}{2}}, X_i X_1 X_d G_{d-1}^{\frac{m}{2}} \right\rangle = 2m, & \text{if } i < d \\ \left\langle G_{d-1}^{\frac{m}{2}}, X_1 G_{d-1}^{\frac{m}{2}} \right\rangle + \\ \qquad \left\langle X_1 X_d G_{d-1}^{\frac{m}{2}}, X_1 G_{d-1}^{\frac{m}{2}} \right\rangle = 2m, & \text{if } i = d \end{cases}.$$

Generation of (d,m)-equitable subgraphs of Q_d

Proof (equitable designs)

m odd. Assume $G_{d-1}^{\frac{m-1}{2}}$ and $G_{d-1}^{\frac{m+1}{2}}$ equitable

$$\langle G_d^m, X_i G_d^m \rangle = \begin{cases} \left\langle G_{d-1}^{\frac{m-1}{2}}, X_i G_{d-1}^{\frac{m-1}{2}} \right\rangle + \\ \quad + \left\langle G_{d-1}^{\frac{m+1}{2}}, X_i G_{d-1}^{\frac{m+1}{2}} \right\rangle, & \text{if } i < d \\ 2\left\langle G_{d-1}^{\frac{m-1}{2}}, X_1 G_{d-1}^{\frac{m+1}{2}} \right\rangle, & \text{if } i = d \end{cases}$$

$$= \begin{cases} (m-1) + (m+1) = 2m, & \text{if } i < d \\ 2\left\langle G_{d-1}^{\frac{m-1}{2}}, X_1 G_{d-1}^{\frac{m+1}{2}} \right\rangle, & \text{if } i = d \end{cases}$$

Thus

$$G_d^m \text{ is } (d,m)\text{-equitable} \Leftrightarrow \left\langle G_{d-1}^{\frac{m-1}{2}}, X_1 G_{d-1}^{\frac{m+1}{2}} \right\rangle = m$$

It can be shown that

$$\left\langle G_{d-1}^{k-1}, X_1 G_{d-1}^{k} \right\rangle = 2k-1 \Rightarrow \left\langle G_d^{2k-1}, X_1 G_d^{2k} \right\rangle = 4k-1$$

$$\left\langle G_{d-1}^{k}, X_1 G_{d-1}^{k+1} \right\rangle = 2k+1 \Rightarrow \left\langle G_d^{2k}, X_1 G_d^{2k+1} \right\rangle = 4k+1$$

Generation of (d, m)-equitable subgraphs of Q_d

Demonstration

$$\left\langle G_{d-1}^{k}, X_1 G_{d-1}^{k+1} \right\rangle = 2k + 1$$

Check that is true for $k = 1$, using the construction G_d^2.

$$\begin{aligned}\left\langle G_d^1, X_1 G_d^2 \right\rangle &= \left\langle (1 + \sum_{i=1}^{d} X_1 \cdots X_i), (X_1 + X_d)(1 + \sum_{j=1}^{d-1} X_1 \cdots X_j) \right\rangle \\ &= \langle 1, 1 \rangle + \langle X_1, X_1 \rangle + \langle X_1 \cdots X_d, X_1 \cdots X_d \rangle \\ &= 3\end{aligned}$$

The identity is thus valid for all k, completing the proof that our algorithm generates (d, m)-equitable subgraphs of Q_d.

Morris designs

$$\mathbb{R}^d = \prod_{j=1}^{t} \mathbb{R}^q, \qquad d = tq \qquad Y = \bigcup_{j=1}^{t} Y^j,$$

where

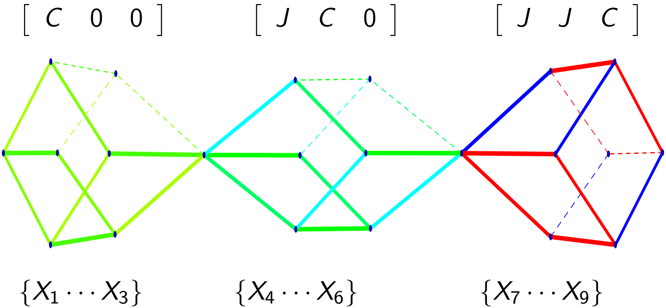
$$Y^j = v_j + C\,[\underbrace{O_q \;\cdots\; O_q}_{j-1 \text{ blocks}} \; I_q \; \underbrace{O_q \;\cdots O_q}_{t-j \text{ blocks}}], \qquad j = 1, \ldots, t \;,$$

$$B_M = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ C & O & O & \cdots & O \\ J & C & O & \cdots & O \\ J & J & C & \cdots & O \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ J & J & J & \cdots & C \end{bmatrix}$$

0: q-element (row) vector of zeros, J: $n_C \times q$ matrix of ones.

Morris designs

$$d = 9 = 3 \times 3$$



Morris designs

Choice of C

Chose $\mathcal{I} \subset \{1, \ldots, q\}$. Let the rows of C(of dimension $n_C \times q$) be the set of all binary vectors with ℓ entries equal to one, $\forall \ell \in \mathcal{I}$.

$$n_C = \sum_{\ell \in \mathcal{I}} C_\ell^q$$

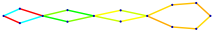
$$m(\mathcal{I}) = I(1)I(q) + \sum_{j=2}^{q} I(j-1)I(j)C_{j-1}^{q-1}$$

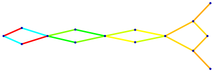
Size of Morris designs

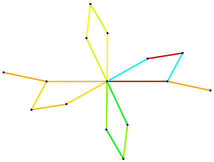
$$n_M = t n_C + 1 = \frac{d}{q} \sum_{\ell \in \mathcal{I}} C_\ell^q + 1$$

Initialisation

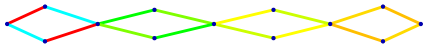
$m = 2$ d odd

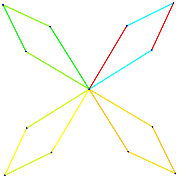






$m = 2,$ d even





Initialisation

$m = 3$