Construction of efficient experimental designs under resource constraints

Radoslav Harman[1,*]
Alena Bachratá[1], Lenka Filová[1], Guillaume Sagnol[2]

[1]Comenius University in Bratislava
[2]Zuse Institut Berlin

Design and Analysis of Experiments in Healthcare
Isaac Newton Institute, Cambridge
8th July 2015

Overview of the talk

- Introduction
- Resource constraints
- Exact optimum designs under resource constraints
- A tabu search heuristic
- Example: Optimal designs for dose escalation trials
- Mixed integer second-order cone programming approach
- Conclusions

The talk is based on papers:

1. Harman, R., Bachratá, A., Filová, L.: Construction of efficient experimental designs under multiple resource constraints, to appear in *Applied Stochastic Models in Business and Industry*
2. Sagnol, G., Harman, R.: Computing exact D-optimal designs by mixed integer second-order cone programming, to appear in *The Annals of Statistics*

Introduction

- An experiment is a set of trials. For each trial, we must select a **design point** from a finite **design space** $\mathfrak{X}$ representing permissible experimental conditions.
- An **exact** experimental design is a mapping $\xi : \mathfrak{X} \to \{0, 1, 2, ...\}$, such that $\xi(x)$ determines the numbers of trials in x.
- An **approximate** (or continuous) experimental design is a mapping $\xi : \mathfrak{X} \to [0, \infty)$, viewed as a relaxation of an exact design*.
- An **optimality criterion** is a mapping $\phi : \Xi \to [0, \infty)$ that measures the "quality" of designs for statistical inference*. (Ξ is a set of designs.)

Assumption of monotonicity:

(M) Augmentation (extension) of an experiment by additional trials cannot decrease its quality for statistical inference, i.e., if designs ξ and ζ satisfy $\xi \leq \zeta$, then $\phi(\xi) \leq \phi(\zeta)$.

Definition of resource constraints

If we consider a general system of linear constraints (such as in Cook and Fedorov 1995, *Statistics*) on ξ, the set of feasible *exact* designs can be very complex.

We suggest to consider the class of **resource constraints** of the form

$$\sum_{x \in \mathfrak{X}} a_{rx} \xi(x) \leq b_r \text{ for all } r \in \{1{:}k\} \quad (\Leftrightarrow \mathbf{A}\xi \leq \mathbf{b}), \tag{1}$$

where a_{rx} represents the "consumption" of the r-th resource by an observation in x, and b_r represents a "limit" on the r-th resource.

(C1) Resource limits are positive and finite, i.e., $b_1, ..., b_k \in (0, \infty)$.

(C2) Augmenting designs cannot decrease the consumption of any resource, i.e., $a_{rx} \geq 0$ for all $r \in \{1{:}k\}^*$ and $x \in \mathfrak{X}$.

(C3) No trial is completely free, and its replication must eventually result in exceeding some resource limit, i.e., for all $x \in \mathfrak{X}$ there is some $r \in \{1{:}k\}$ such that $a_{rx} > 0$.

Special cases of resource constraints

Resource constraints do *not* include all linear constraints, but they do include for instance

- **Size constraints**: $k = 1$, $a_{1x} = 1$ for all x, and $b_1 = N$.
- **Cost constraints**: $k = 1$, a_{1x} are possibly unequal costs of trials in design points x, and $b_1 = B$ is the limit on the budget*.
- **Direct constraints**: $k = n$, $a_{rx} = \delta_{rx}$ (the Kronecker delta).
- **Strata constraints**: $a_{rx} = I[x \in \mathfrak{X}_r]$ (the indicator function), where $\mathfrak{X}_1, ..., \mathfrak{X}_k$ is a partition of $\mathfrak{X}$, and b_r, $r \in \{1{:}k\}$, sum to N.
- "Time separation" constraints: If $\mathfrak{X} = \{1 : n\}$ is a sequence of time moments of trials and consecutive trials must be at least Δ ($\leq n$) time moments apart, set $k = n - \Delta + 1$,$b_r = 1$ for $r \in \{1{:}k\}$, $a_{rx} = I[x \in \{r, r+1, ..., r+\Delta-1\}]$ for all $r \in \{1{:}k\}$, $x \in \mathfrak{X}$.
- "Latin hypercube" constraints for the case that $\mathfrak{X}$ is a cubic grid of points.

Sets of feasible exact and approximate designs

Assume that $\xi^{(0)}$ is a **design to be augmented**, i.e., either an exact design representing trials that have already been performed or a required initial part of the experiment, $\mathbf{A}\xi^{(0)} \leq \mathbf{b}$.

Assumptions (C1)-(C3) guarantee that the set

$$\Xi^{\mathrm{ex}} = \{\xi \text{ exact} : \xi^{(0)} \leq \xi, \mathbf{A}\xi \leq \mathbf{b}\}$$

of all **feasible exact designs** is non-empty, finite, and *simple to explore* using heuristic optimization methods based on transitions between "neighbouring" feasible design (by adding and removing trials).

The assumptions imply that the set of **feasible approximate designs:**

$$\Xi^{\mathrm{ap}} = \{\xi \text{ approximate} : \xi^{(0)} \leq \xi, \mathbf{A}\xi \leq \mathbf{b}\}$$

is a non-empty, compact and convex polyhedron.

Sets of feasible exact and approximate designs

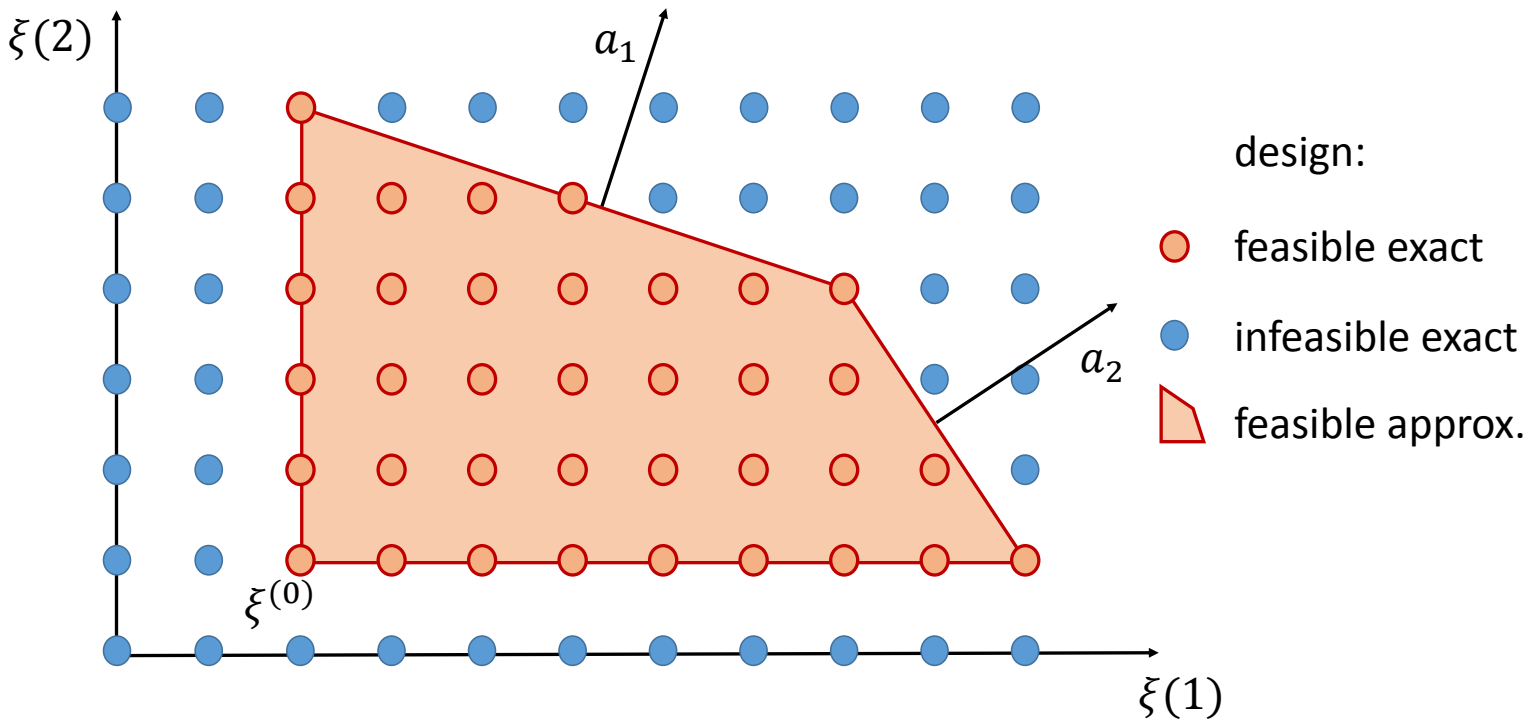


Figure : Illustration of the feasible set defined by resource constraints with $\mathfrak{X} = \{1, 2\}$. The set Ξ^{ex} is formed by the red points and Ξ^{ap} is the red polygon.

Optimum design under resource constraints

The main optimization problem:

$$\xi^* \in \operatorname{argmax}\{\phi(\xi) : \xi \in \Xi^{\mathrm{ex}}\}. \tag{2}$$

In addition to optimum design, optimization (2) covers knapsack problems, optimal redundancy allocation in reliability theory, optimization problems in graph theory, and other very hard problems.

Classes of computational methods for discrete optimization:

- Enumeration methods (Perfect solutions of small problems)
- Heuristic methods (Solutions of large problems)

However, it is not completely trivial to construct an efficient (i.e., not brute-force) enumeration method, nor is it simple to implement an efficient heuristic for solving Problem (2).

- We will describe a tabu search heuristic method.
- We will (very briefly) describe a theoretical method that allows us to convert the problem to mixed integer second-order cone programming, and use a solver to construct a solution.

A tabu search heuristic method

An exact design is **maximal** if it cannot be augmented without violation of some of the resource constraints. Monotonicity of ϕ implies that the set of maximal designs contains an optimal design.

For each design $\xi \in \Xi^{\mathrm{ex}}$ we choose:

- A **local heuristic evaluation** $val(\xi)$, which is a real number that roughly estimates how promising ξ is as part of an excursion leading to an efficient design. In the actual implementation, we set $val(\xi) = \phi(\xi^+)$, where ξ^+ is an extremal approximate design augmenting ξ, based on the vector of residual amounts of resources.
- A **characteristic attribute** $attr(\xi)$. It assigns different values to substantially different designs and the same values to essentially the same designs (for instance, to algebraically isomorphic designs).

The attributes of all visited designs are stored in a **tabu list** V.

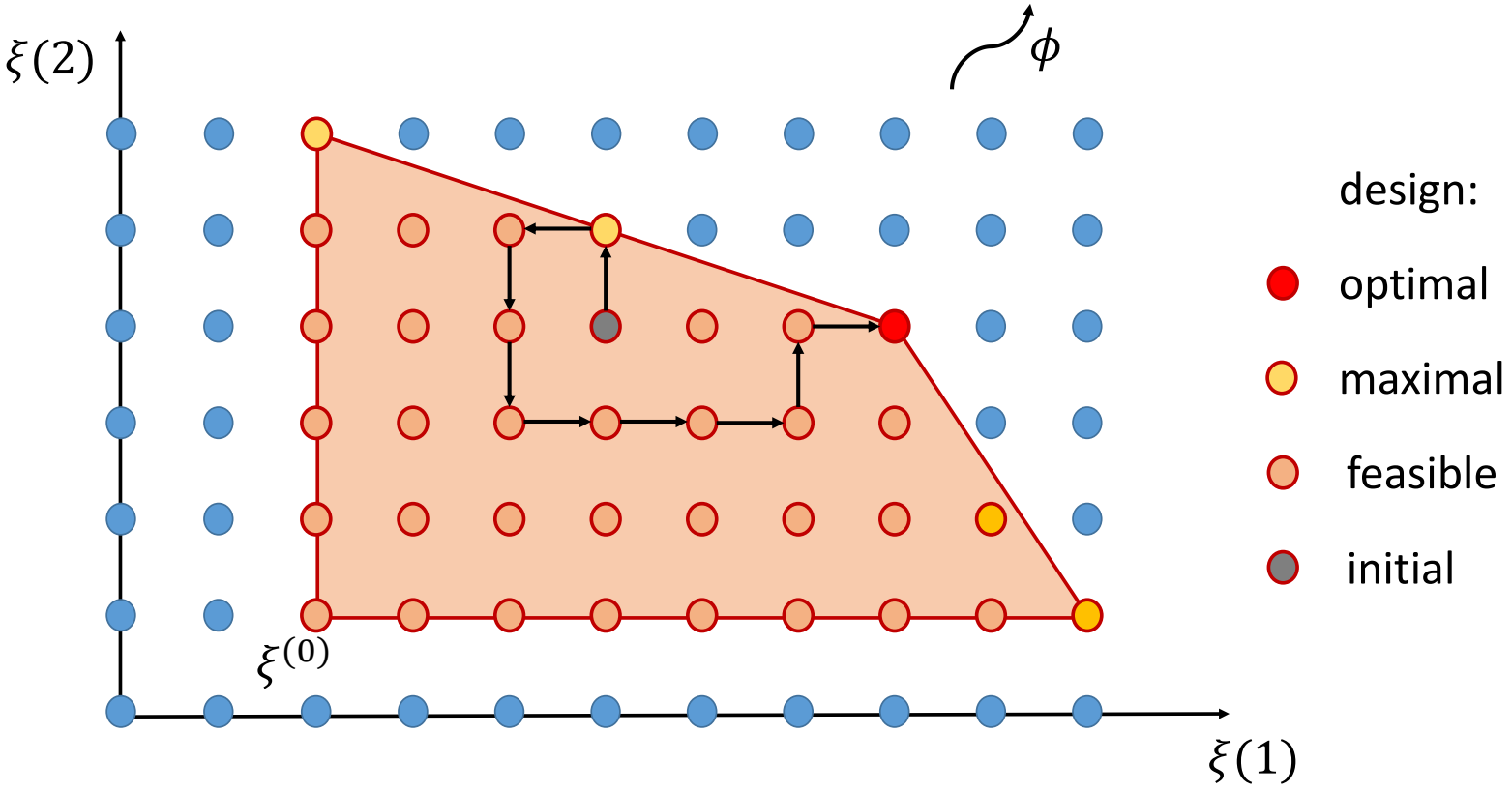
A tabu search heuristic method

Let ξ represent the current design in the excursion.

- The algorithm attempts a **forward step** (if $attr(\xi) \notin V$) or a **backward step** (if $attr(\xi) \in V$), moving to a neighboring feasible exact design ζ. Design ζ is chosen such that it maximizes *val* among all designs satisfying $attr(\zeta) \notin V$.
- If the algorithm attempts a forward step but there is no "upper neighbour" ζ of ξ such that $attr(\zeta) \notin V$, or if it attempts a backward step but there is no "lower neighbour" ζ such that $attr(\zeta) \notin V$, the algorithm tries to reverse the direction.
- If all these attempts fail, that is, if the attributes of all neighboring designs of ξ are contained in the list V, the algorithm randomly selects a neighbour of ξ for the next step.
- The output of the algorithm is the best maximal design.

The algorithm resembles the DETMAX procedure for constructing optimal exact designs under the standard constraint on the size (Mitchell 1974, *Technometrics*).

Illustration of the tabu heuristic

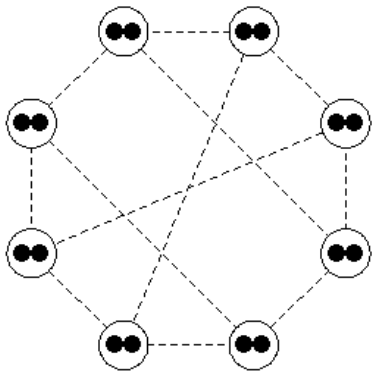


Performance of the tabu search heuristic

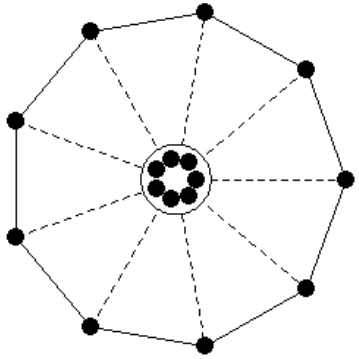
Tested on examples: $\rightarrow$ block models with limits on the numbers of blocks and on the replications of treatments, $\rightarrow$ regression models with simultaneous marginal, cost and direct constraints.

It usually takes less than 300,000 criterion evaluations to obtain a (nearly) optimal design for problems with up to 100 design points.

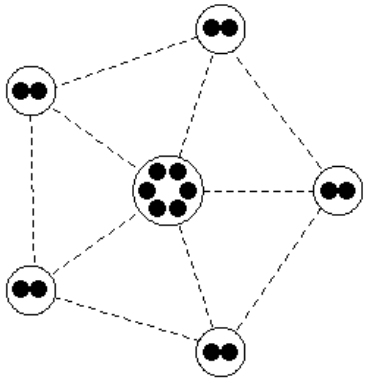
For the block model with blocks of size two and 16 treatments we found a design which is better than the design with a strongly regular concurrence graph (measured by the criterion of D-optimality).



v=16, N=48



v=16, N=72



v=16, N=80

Example: A-optimal designs for dose escalation

Consider an experiment with a placebo and n doses, i.e., $t = n+1$ treatments. The treatments are applied to individuals in t cohorts*.

Block design with cohorts corresponding to blocks: The response for the u-th individual of the k-th cohort receiving the i-th treatment is

$$Y_{ki,u} = \mu + \beta_k + \tau_i + \varepsilon_{ki,u},$$

where μ is the overall mean, β_k is the effect of the k-th cohort, τ_i is the effect of the i-th treatment, and $\varepsilon_{ki,u}$ is a zero-mean random error.

Dose escalation is commonly used in first-in-man studies of a new drug: **for cohort k**, some individuals may be allocated to placebo, some to doses $1, ..., k$, but **no individual is allocated to doses** $\mathbf{k+1}, \mathbf{k+2}, ..., \mathbf{n}$. We also require that in cohort k at least one individual is allocated to dose k.

Bailey 2009, *Statistics in Medicine*,
Haines and Clark 2014, *Statistics and Computing*.

Example: A-optimal designs for dose escalation

The design space = all permissible combinations (cohort,treatment):
$\mathfrak{X} = \{(1,1),(1,2),(2,1),(2,2),(2,3),\ldots,(t-1,t)\} \cup \{(t,1),...,(t,t)\}$.

For an exact design, $\xi(k,i)$ gives the number of individuals of the k-th cohort allocated to the i-th treatment.

Design $\xi^{(0)}$ to be augmented is $\xi^{(0)}(i-1,i) = 1$ for all $i \in \{2,...,t\}$ and $\xi^{(0)}(i-1,i) = 0$ otherwise.

Examples of resource constraints:

- Overall number of individuals: $\sum_{k,i} \xi(k,i) \leq N$.
- Sizes of individual cohorts: $\sum_i \xi(k,i) \leq m_k$ for all k.
- Replication numbers of the treatments: $\sum_k \xi(k,i) \leq r_i$ for all i.
- Cost of the entire experiment: $\sum_k c(k,i)\xi(k,i) \leq B$.
- ...

Example: A-optimal designs for dose escalation

- $\mathbf{S}(\xi)$ is an $t \times t$ matrix with entry ki equal to the number of individuals in cohort k allocated to treatment i.
- $\mathbf{R}(\xi)$ is the diagonal $t \times t$ matrix with entry ii equal to the number of time the treatment i is replicated.
- $\mathbf{K}(\xi)$ is the diagonal $t \times t$ matrix with entry kk equal to the number of individuals in the k-th cohort.

The information matrix for $\tau = (\tau_1, ..., \tau_t)'$ is

$$\mathbf{L}(\xi) = \mathbf{R}(\xi) - (\mathbf{S}(\xi))'(\mathbf{K}(\xi))^{-1}\mathbf{S}(\xi).$$

For the example, we will select the criterion of A-optimality:

$$\phi(\xi) = (t-1)/\sum_{j=2}^{t} \lambda_j^{-1}(\mathbf{L}(\xi)),$$

$0 = \lambda_1(\mathbf{L}(\xi)) < \lambda_2(\mathbf{L}(\xi)) \leq \cdots \leq \lambda_t(\mathbf{L}(\xi))$ are the eigenvalues of $\mathbf{L}(\xi)$.

Example: A-optimal designs for dose escalation

At most 8 individuals per cohort (Haines and Clark 2014).
$\Rightarrow$ 5 constraints

cohort \ dose	p	1	2	3	4	cohort sizes
1	4	4 ≥ 1				8
2	2	3	3 ≥ 1			8
3	2	1	2	3 ≥ 1		8
4	1	1	1	2	3 ≥ 1	8
5	1	1	1	2	3	8
replications	10	10	7	7	6	40

Example: A-optimal designs for dose escalation

At most 40 individuals $\Rightarrow$ 1 constraint
(Relative efficiency 123%)

cohort \ dose	p	1	2	3	4	cohort sizes
1	1	1 ≥ 1				2
2	1	1	1 ≥ 1			3
3	1	1	1	1 ≥ 1		4
4	2	2	2	3	3 ≥ 1	12
5	3	3	4	4	5	19
replications	8	8	8	8	8	40

Example: A-optimal designs for dose escalation

At most 40 inds. + at most 11 inds. per cohort $\Rightarrow$ 6 res. constraints
(Relative efficiency 116%)

cohort \ dose	p	1	2	3	4	cohort sizes
1	1	1 ≥ 1				2
2	1	2	2 ≥ 1			5
3	2	3	3	3 ≥ 1		11
4	2	2	2	2	3 ≥ 1	11
5	2	1	2	2	4	11
replications	8	9	9	7	7	40

Example: A-optimal designs for dose escalation

At most 40 inds. + at most 11 inds. per cohort + at least 2 inds. on placebo per cohort + 3 inds. on the “escalation dose” per cohort (Relative efficiency 110%)

cohort \ dose	p	1	2	3	4	cohort sizes
1	2 (≥ 2)	3 (≥ 3)				5
2	2 (≥ 2)	1	3 (≥ 3)			6
3	2 (≥ 2)	1	1	3 (≥ 3)		7
4	2 (≥ 2)	2	2	2	3 (≥ 3)	11
5	2 (≥ 2)	2	2	2	3	11
replications	8	9	8	7	6	40

Example: A-optimal designs for dose escalation

At most 40 inds. + at most 11 inds. per cohort + at most j inds. use the dose applied in $j-1$ previous cohorts $\Rightarrow$ 20 constraints
(Relative efficiency 94%)

cohort \ dose	p	1	2	3	4	cohort sizes
1	2	1 $\leq 1 \geq 1$				3
2	2	2 ≤ 2	1 $\leq 1 \geq 1$			5
3	4	3 ≤ 3	2 ≤ 2	1 $\leq 1 \geq 1$		10
4	2	3 ≤ 4	3 ≤ 3	2 ≤ 2	1 $\leq 1 \geq 1$	11
5	2	2 ≤ 5	2 ≤ 4	3 ≤ 3	2 ≤ 2	11
replications	12	11	8	6	3	40

MISOCP approach

What if we need more complex constraints than resource constraints or more control over / information about the quality of the result?

If the constraints are second-order cone representable (and we have a standard model + criterion is D-, A-, G- or c-optimality), it is possible to transform the problem to the form of a **mixed integer second-order cone program** and solve it using specialized software.

For instance, for A-optimality we need to solve:

$$\begin{aligned} \max_{\xi(x), Y_x, \mu_x} \quad & \sum_x \mu_x \\ \text{s.t.} \quad & \sum_x A_x Y_x = \sum_x \mu_x, \ \| Y_x \|_F^2 \leq \mu_x \xi(x) \text{ (for all } x \in \mathfrak{X}), \\ & \mu_x \geq 0 \text{ (for all } x \in \mathfrak{X}), \ \xi \in \Xi^{\mathrm{ex}} \end{aligned}$$

Table : SOCP formulation of the A optimal design problem. The matrices A_x are such that the criterion of A-optimality $\phi_A(\xi)$ is the harmonic mean of the eigenvalues of the "moment" matrix $\sum_x \xi(x) A_x A_x'$.

Comparison of heuristics and MP methods

Heuristic method	MISOCP method
Medium to large-size design space	Small to medium-size design space
Resource constraints	SOC constraints including general linear constraints as a special case
Any monotonic criterion.	Standard optimality criteria (D,A,G,c), regression models with uncorrelated observations.
No lower bound on efficiency, no proof of optimality	Strong lower bound on efficiency, sometimes also proof of prefect optimality

Summary

- Optimization under multiple constraints can be used to construct statistically efficient experimental designs that conform to (a combination of) financial, material, logistic, operational or safety requirements.
- The problems of optimal exact designs under multiple constraints are computationally very challenging, but today we already have efficient heuristic and mathematical programming algorithms that can provide optimal or near-optimal solutions for (at least some) realistic situations.

Thank you for attention!