Optimal Decisions in Statistics and Data Analysis

Cardiff, December 16-18, 2013

Abstracts

A conference organized on the occasion of the 60th birthday of Professor Anatoly Zhigljavsky

Conference organizers:

- Henry P. Wynn (LSE, London, UK)
- Werner G. Müller (Linz, Austria)
- Luc Pronzato (Sophia Antipolis, France)

Local organizing committee:

- Jonathan Gillard
- Gareth Davies
- Iskander Aliev
- Andrey Pepelyshev

Optimum dose regimen selection for a target drug concentration

Kabir Soeny
Barbara Bogacka
Queen Mary, University of London, UK

Byron Jones
Thomas Bouillon
Novartis Pharma AG, Basel, Switzerland

In this talk we will present an optimum design for dose regimen selection in clinical trials, where the criterion of optimality meets the requirement that the drug concentration be close to the target drug exposure needed to cure the disease. Furthermore, in cases when the drug is a combination of salts, we find an optimum ratio of the components as well as an optimum dosing regimen.

We define new criteria of optimality and present their properties as well as the new optimization algorithm. We compare various dosing regimens and present a sensitivity analysis for the choice of the model parameter values.

[Barbara Bogacka; School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK]
[b.bogacka@qmul.ac.uk – www.maths.qmul.ac.uk/~bb]

Of copulas, quantiles, ranks, and spectra: an $L1$-approach to spectral analysis

Holger Dette, Tobias Kley, Stanislav Volgushev
Ruhr Universität Bochum

Marc Hallin
ECARES, Université Libre de Bruxelles and ORFE, Princeton University

We present an alternative method for the spectral analysis of a strictly stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$. We define a "new" spectrum as the Fourier transform of the differences between copulas of the pairs (Y_t, Y_{t-k}) and the independence copula. This object is called *copula spectral density kernel* and allows to separate marginal and serial aspects of a time series. We show that it is intrinsically related to the concept of quantile regression. Like in quantile regression, which provides more information about the conditional distribution than the classical location-scale model, the copula spectral density kernel is more informative than the spectral density obtained from the autocovariances. In particular the approach provides a complete description of the distributions of all pairs (Y_t, Y_{t-k}). Moreover, it inherits the robustness properties of classical quantile regression, because it does not require any distributional assumptions such as the existence of finite moments. In order to estimate the copula spectral density kernel we introduce rank-based Laplace periodograms which are calculated as bilinear forms of weighted L_1-projections of the ranks of the observed time series onto a harmonic regression model. We establish the asymptotic distribution of those periodograms, and the consistency of adequately smoothed versions. The finite-sample properties of the new methodology, and its potential for applications are briefly investigated by simulations and a short empirical example.

[Holger Dette; Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany]
[holger.dette@rub.de – www.ruhr-uni-bochum.de/mathematik3/en/dette.html]

Robust multiobjective optimisation

Joerg Fliege
CORMSIS, University of Southampton, UK

Ralf Werner
University of Augsburg, Germany

Motivated by Markowitz portfolio optimization problems under uncertainty in the problem data, we consider general convex parametric multiobjective optimization problems under data uncertainty. For the first time, this uncertainty is treated by a robust multiobjective formulation in the gist of Ben-Tal and Nemirovski. For this novel formulation, we investigate its relationship to the original multiobjective formulation as well as to its scalarizations. Further, we provide a characterization of the location of the robust Pareto frontier with respect to the corresponding original Pareto frontier and show that standard techniques from multiobjective optimization can be employed to characterize this robust efficient frontier. We illustrate our results based on a standard meanvariance problem.

[Joerg Fliege; CORMSIS, University of Southampton, Southampton SO17 1BJ, UK]
[J.Fliege@soton.ac.uk – www.cormsis.soton.com]

Efficient computation of IMSE optimal designs of experiments

Bertrand Gauthier
Luc Pronzato
Laboratoire I3S, University of Nice-Sophia Antipolis/CNRS, France

We address the problem of computing IMSE (Integrated Mean Square Error) optimal designs for random fields interpolation with known covariance structure.

We consider a spectral representation of the IMSE criterion obtained from the (Karhunen-Loève) eigendecomposition of the integral operator defined by the random field covariance kernel and the given integration measure. The IMSE criterion can then be evaluated without resorting to the explicit integration of the MSE, and spectral truncation naturally defines an approximate *truncated-IMSE* criterion, see [1].

This approach is of particular interest when a quadrature rule (*i.e.*, a discrete integration measure) is used to approximate the IMSE. Indeed, in this situation, both IMSE and truncated-IMSE optimal designs are supported by quadrature points [1]. In addition, numerical experiments indicate that retaining a small number of eigenpairs is sufficient to obtain IMSE-optimal designs, or at least good approximations of them. A simulated-annealing based search algorithm, see [2], can then be used for the efficient computation of IMSE optimal designs for discrete integration measures. Several examples are presented.

References:

[1] B. Gauthier and L. Pronzato (2013), *Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models* in review (available at `http://www.i3s.unice.fr/~bgauthie/`).

[2] R. Jin, W. Chen and A. Sudjianto (2005), *An efficient algorithm for constructing optimal design of computer experiments.* Journal of Statistical Planning and Inference, 134(1):268– 287.

[Bertrand Gauthier; Laboratoire I3S, Bât, Euclide, Les Algorithmes, 2000 route des lucioles, 06903 Sophia Antipolis Cedex, France]
[`bgauthie@i3s.unice.fr` – `http://www.i3s.unice.fr/~bgauthie/`]

Matrix optimization problems arising in matrix low-rank approximation

Jonathan Gillard
Anatoly Zhigljavsky
School of Mathematics, Cardiff University

The matrix low-rank approximation can be defined as follows. Let L, K and r be given positive integers such that $1 \leq r < L \leq K$. Denote the set of all real-valued $L \times K$ matrices by $\mathbb{R}^{L\times K}$. Let $\mathcal{M}_r = \mathcal{M}_r^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing all matrices with rank $\leq r$.

Assume we are given a matrix $\mathbf{X}_* \in \mathcal{H}$. The problem of low rank approximation is:

$$f(\mathbf{X}) \to \min_{\mathbf{X}\in\mathcal{M}_r} \tag{1}$$

where $f(\mathbf{X}) = \rho^2(\mathbf{X}, \mathbf{X}_*)$ is a squared distance on $\mathbb{R}^{L\times K} \times \mathbb{R}^{L\times K}$.

If $f(\cdot)$ is the standard Frobenius norm, then (1) has a unique, explicit solution given by truncating the singular value decomposition of $\mathbf{X}_*$.

In this talk we will describe optimization problems that arise when:

1. The norm $f(\cdot)$ is chosen to be a weighted Frobenius norm. Some of the weights may be selected to be 0 or ∞, relating to missing and exact observations, respectively. This is known as the weighted low rank approximation problem.

2. The constraint $\mathbf{X} \in \mathcal{M}_r$ in (1) is replaced by $\mathbf{X} \in \mathcal{M}_r \cap \mathcal{H}$ where $\mathcal{H} = \mathcal{H}^{L\times K} \subset \mathbb{R}^{L\times K}$ is subset of $\mathbb{R}^{L\times K}$ containing matrices of some known structure (for example, Hankel). This is known as the structured low rank approximation problem (SLRA).

For example in this talk, we will demonstrate, that for SLRA the optimization problem arising is typically very difficult: in particular, the objective function is multiextremal even for simple cases and possesses large Lipschitz constants. The number of local minima is a linear function of the number of elements in $\mathbf{X}_*$, yet existing methods to find a solution of the SLRA problem are solely based on methods of local optimization. Recent discussion of this is contained in [1] and [2].

References:

[1] J. W. Gillard and A. A. Zhigljavsky. 2011. Analysis of structured low rank approximation as an optimization problem. Informatica 22(4), pp. 489-505.

[2] J. W. Gillard and A. A. Zhigljavsky. 2013. Optimization challenges in the structured low rank approximation problem. Journal of Global Optimization 57(3), pp. 733-751.

[Jonathan Gillard; School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, CF24 4YH, UK]
[GillardJW@Cardiff.ac.uk – http://www.jonathangillard.co.uk]

Singular spectrum analysis as a nonparametric approach to parametric models

Nina Golyandina
St.Petersburg State University, Russia

Singular Spectrum Analysis (SSA) is known as a nonparametric method of time series analysis, which can decompose a noisy series into a sum of identifiable components with no given model [1-3]. In particular, SSA can extract series trends and periodicities with no parametric model of the trends and unknown frequencies of the periodic components. However, there is a model of time series that suits SSA very well. The signal $\mathrm{S}_N = (s_1, \ldots, s_N)$ that is governed by a linear recurrent relation (LRR) $s_n = \sum_{i=1}^{r} a_i s_{n-i}$ matches SSA in the following sense: only such time series components can be exactly extracted and continued by the SSA methods.

The set of signals governed by LRRs has a natural parametrization. In the complex-valued form, the common term can be written down as follows:

$$s_n = \sum\nolimits_i P_{m_i}(n)\mu_i^n, \tag{2}$$

where P_{m_i} is a polynomial of degree m_i. Let for simplicity $m_i = 1$. The model (2) can be characterized by two ways. First, the coefficients of the minimal governing LRR determine the set of μ_i, and vice versa. Second, the same model is determined by the subspace that is formed by the lagged vectors $S_i = (s_i, \ldots, s_{i+L-1})^\top$, $i = 1, \ldots, N - L + 1$, where L is the so called window length, since the basis of the signal subspace spans the vectors $(1, \mu_i, \ldots, \mu_i^{L-1})^\top$. Therefore, one has the choice: to deal with parametric models (2), with LRRs, or with subspaces.

The approach based on the subspace structure with no explicit estimation of μ_i can be called a nonparametric approach to parametric model. Since the values of μ_i are not necessary for extraction of the corresponding series components by SSA and for their SSA forecasting, Singular Spectrum Analysis as a nonparametric approach can be considered. In the talk, arguments pro and contra are discussed. The main argument pro is that the nonparametric approach is more stable to deviations from the model. If, for example, the trend is approximated by several addends in (2), then the parametric approach consists in the direct estimation of the parameters and then the trend extraction by the obtained dependence on n. However, the procedure can be unstable. The nonparametric SSA approach allows one to apply the methods even if the signal satisfies the model only locally. For example, trends usually do not satisfy LRRs; however, they can be extracted by SSA and its extensions.

Basic SSA is able to extract addends in (2) which are approximately orthogonal to the residual and have contributions different from that of the residual components. Nonparametric extensions of SSA for separation of not necessarily orthogonal series components with possibly equal contributions are suggested. The methods are based on the model (2) but does not estimate the component parameters. Non-parametric approaches to multidimensional objects may be even more promising; thereby, the multidimensional versions of SSA are also considered.

References:

[1] D.Danilov and A.Zhigljavsky (eds). *Principal components of time series.* SPbSU press, St.Petersburg, 1997 (in Russian)

[2] N.Golyandina, V.Nekrutrin, and A.Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*, CRC Press, Boca Raton, 2001.

[3] N.Golyandina and A.Zhigljavsky. *Singular Spectrum Analysis for Time Series*, Springer, Boca Raton, 2013.

[Nina Golyandina; St.Petersburg State University, Faculty of Mathematics and Mechanics, Universitetsky pr. 28, Staryj Petergof, 198504, St. Petersburg, Russia]
[n.golyandina@spbu.ru – gistatgroup.com]

Global optimization algorithms using derivatives and their systematic testing

Dmitri E. Kvasov
Yaroslav D. Sergeyev
Calabria University, Rende (CS), Italy
N. I. Lobachevsky University of Nizhni Novgorod, Russia

A global optimization problem is considered where the objective function $f(x)$ is a multidimensional multiextremal and hard to evaluate function and its gradient $f'(x)$ satisfies the Lipschitz condition over a hyperinterval D with an unknown Lipschitz constant K.

Different methods for solving this problem have been proposed (see, e. g., [4, 9, 11, 15–19]) that can be distinguished either by the mode in which information about the Lipschitz constant K is obtained or by the strategy of exploration of the search hyperinterval D.

There are several ways to specify the Lipschitz constant K: it can be given a priori (see, e. g., [1–3]); its adaptive estimates (local or global) can be obtained during the search (see, e. g., [5, 10, 13, 15, 18]); multiple estimates of the Lipschitz constant can be also used (see, e. g., [7, 8]). All these techniques are considered in this talk, with a particular attention to the local tuning approach (see [12, 13, 15]) and a recently proposed technique (see [7, 8]) for using multiple estimates of K (the existence of such an algorithm was a challenge for more than 15 years since the DIRECT method working in this way with Lipschitz objective functions has been proposed in 1993, see [4]).

In exploring the multidimensional search domain, various adaptive partitioning strategies can be applied. For example, one-point-based algorithms subsequently subdivide the search region in smaller ones and evaluate the objective function at one point within each subregion (see, e. g., [3, 8]). Diagonal partitions are also very interesting for practical applications with expensive black-box functions (see, e. g., [14, 15]). More complex partitions, based on simplices, auxiliary functions of various nature, and so on, can be also used (see, e. g., [4, 11, 17, 19]).

A number of geometric Lipschitz global optimization methods based on constructing auxiliary functions with the usage of different estimates of the Lipschitz constant K are presented in this communication. These methods use either the one-point-based partition strategy or the diagonal one. A special attention in the talk is dedicated to their testing and comparing with some well known Lipschitz global optimization algorithms. A systematic experimental investigation of the methods is performed by using a set of criteria specially developed (see [14, 15]) for comparing different black-box global optimization methods.

The proposed comparison criteria are based on classes of test functions. The GKLS-generator [6] is used for producing such test classes. It constructs three types (non-differentiable, continuously differentiable, and twice continuously differentiable) of classes of multidimensional and multiextremal test functions with known local and global minima. The generation procedure consists of defining a convex quadratic function systematically distorted by polynomials. Each test class provided by the generator consists of 100 functions and is defined by the following parameters: (i) problem dimension, (ii) number of local minima, (iii) global minimum value, (iv) radius of the attraction region of the global minimizer, (v) distance from the global minimizer to the quadratic function vertex. The other necessary parameters are chosen randomly by the generator for each test function of the class. A special notebook with a complete description of all functions is supplied to the user. The GKLS-generator always produces the same test classes for a given set of the user-defined parameters, allowing one to perform repeatable numerical experiments. By changing the user-defined parameters, classes with different properties can be created. More information can be found at: `http://wwwinfo.dimes.unical.it/~yaro`.

Acknowledgments. This work was supported by the Ministry of Education and Science of Russian Federation, project 14.B37.21.0878, as well as by the grants n. 11-01-00682-a of the Russian Foundation for Fundamental Research and n. 1960.2012.9 awarded by the President of the Russian Federation for supporting the leading research groups.

References:

[1] W. Baritompa and A. Cutler, "Accelerations for global optimization covering methods using second derivatives", *Journal of Global Optimization*, **4**(3):329–341, 1994.

[2] L. Breiman and A. Cutler, "A deterministic algorithm for global optimization", *Mathematical Programming*, **58**(1-3):179–199, 1993.

[3] Yu. G. Evtushenko and M. A. Posypkin, "A deterministic approach to global box-constrained optimization", *Optimization Letters*, **7**(4):819–829, 2013.

[4] C. A. Floudas and P. M. Pardalos, (Eds.), *Encyclopedia of Optimization (6 Volumes)*, Springer, New York, 2009.

[5] J. M. Fowkes, N. I. M. Gould, and C. L. Farmer, "A branch and bound algorithm for the global optimization of Hessian Lipschitz continuous functions", *Journal of Global Optimization*, **56**(4):1791–1815, 2013.

[6] M. Gaviano, D. E. Kvasov, D. Lera, and Ya. D. Sergeyev, "Algorithm 829: Software for generation of classes of test functions with known local and global minima for global optimization", *ACM Transactions on Mathematical Software*, **29**(4):469–480, 2003.

[7] D. E. Kvasov and Ya. D. Sergeyev, "A univariate global search working with a set of Lipschitz constants for the first derivative", *Optimization Letters*, **3**(2):303–318, 2009.

[8] D. E. Kvasov and Ya. D. Sergeyev, "Lipschitz gradients for global optimization in a one-point-based partitioning scheme", *Journal of Computational and Applied Mathematics*, **236**(16):4042–4054, 2012.

[9] D. E. Kvasov and Ya. D. Sergeyev, "Univariate geometric Lipschitz global optimization algorithms", *Numerical Algebra, Control and Optimization*, **2**(1):69–90, 2012.

[10] D. Lera and Ya. D. Sergeyev, "Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives", *SIAM Journal on Optimization*, **23**(1):508–529, 2013.

[11] R. Paulavičius and J. Žilinskas, *Simplicial Global Optimization*, Springer, New York, 2014 (doi: 10.1007/978-1-4614-9093-7).

[12] Ya. D. Sergeyev, "An information global optimization algorithm with local tuning", *SIAM Journal on Optimization*, **5**(4):858–870, 1995.

[13] Ya. D. Sergeyev, "Global one-dimensional optimization using smooth auxiliary functions", *Mathematical Programming*, **81**(1):127–146, 1998.

[14] Ya. D. Sergeyev and D. E. Kvasov, "Global search based on efficient diagonal partitions and a set of Lipschitz constants", *SIAM Journal on Optimization*, **16**(3):910–937, 2006.

[15] Ya. D. Sergeyev and D. E. Kvasov, *Diagonal Global Optimization Methods*, FizMatLit, Moscow, 2008. In Russian.

[16] Ya. D. Sergeyev and D. E. Kvasov, "Lipschitz global optimization", in J. J. Cochran et al., (Eds.), *Wiley Encyclopedia of Operations Research and Management Science (in 8 volumes)*, John Wiley & Sons, New York, **4**:2812–2828, 2011.

[17] Ya. D. Sergeyev, R. G. Strongin, and D. Lera, *Introduction to Global Optimization Exploiting Space-Filling Curves*, Springer, New York, 2013.

[18] R. G. Strongin and Ya. D. Sergeyev, *Global Optimization with Non-Convex Constraints: Sequential and Parallel Algorithms*, Kluwer Academic Publishers, Dordrecht, 2000.

[19] A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization*, Springer, New York, 2008.

[D.E.Kvasov; DIMES, Calabria University, via P. Bucci, 42C – 87036 Rende (CS), Italy]
[`kvadim@si.dimes.unical.it` – `http://wwwinfo.dimes.unical.it/~kvadim`]

Posterior singular spectrum analysis

Ilkka K. Launonen
University of Oulu, Oulu, Finland

Lasse L. Holmström
University of Oulu, Oulu, Finland

A method is proposed for noise suppression in time series and finding interesting underlying features, such as trends, maxima, minima and oscillations. A combination of Singular Spectrum Analysis and Bayesian modeling is used where the credibility of SSA signal components are analyzed via posterior simulation.

Given a sample from a time series' posterior distribution, the method has two steps. First, the eigentriples are computed, applying basic SSA either on the time series or the posterior mean. Second, the posterior sample is projected in phase space onto the one-dimensional subspaces defined by the eigentriples, and Bayesian inference is made about the eigentriples using their associated projected samples.

The inference is presented in visual form with credibility maps. For each SSA component relating to an eigentriple, a map of tapered pillars of white and black on a gray background is built, with time index and credibility level on the axes. The colors indicate whether the slope of the component is credibly positive or negative, or neither, at the given credibility level. The credibility level relates to the joint probability that the colored features hold and inference is thus simultaneous within a map. Inference on the set of all SSA components is either simultaneous or separate. Components whose maps are mostly gray are considered noise.

The potential of the technique is demonstrated using artificial and real data examples. The method is partly motivated by the analysis of paleoclimate temperature reconstructions, which are based on a hierarchical model with data about the fossil record preserved in glacial lake sediment.

Preprint of the published article [1] is available online at
`http://cc.oulu.fi/~llh/preprints/PSSA.pdf`.

References:

[1] L. Holström and I. Launonen, Posterior Singular Spectrum Analysis, *Statistical Analysis and Data Mining*, (6):387–402, 2013.

[Ilkka Launonen; Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland]
[`ilkka.launonen@oulu.fi` – `http://cc.oulu.fi/~llh/group/ssa.html`]

[Lasse Holmström; Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland]
[`lasse.holmstrom@oulu.fi` — `http://cc.oulu.fi/~llh/`]

Construction of optimum designs to discriminate among several non-normal models

Jesus López–Fidalgo
Raúl Martín–Martín
Department of Mathematics, Institute of Mathematics Applied to Sciences and Engineering, University of Castilla–La Mancha, Ciudad Real, Spain

Chiara Tommasi
Department of Economics, Business and Statistics, University of Milano, Italy

Different optimality criteria have been considered in the recent literature to discriminate between rival models (e.g. [1], [2], [3]). Using the basic idea of discriminating between two models other criteria may be defined for discriminating among a class of several. Most of the criteria studied assume normality while a new criterion based Kullback–Leibler distance considers any possible distribution for each of the rival models ([4], [6]). In this work a new max–min criterion is established for more than two rival models. This criterion seems to be more adequate for discrimination than others previously considered in the literature ([5]). Then some theoretical results are proved in order to produce an operative equivalence theorem. From it some algorithms are provided. Finally an illustrative example is considered and some designs are computed.

The aim is the search of experimental designs to optimally determine which of k rival statistical models given by their pdf's, $f_i(y, x, \theta_i)$, $\quad i = 1, \ldots, k$. Let the Kullback–Leibler distance between two of them be

$$\mathcal{I}\left[f_t(y, x, \theta_t), f_i(y, x, \theta_i)\right] = \int f_t(y, x, \theta_t) \log\left[\frac{f_t(y, x, \theta_t)}{f_i(y, x, \theta_i)}\right] dy.$$

The KL–optimality criterion is defined through the function

$$I_{i,t}(\xi) = \min_{\theta_i \in \Omega_i} \int_{\mathcal{X}} \mathcal{I}\left[f_t(y, x, \theta_1), f_i(y, x, \theta_i)\right] \xi(dx).$$

The "true" model in this situation will be considered as a convex combination of the pdf's of all the rival models, $f_{k+1}(y, x, \theta_{k+1}) = \sum_{i=1}^{k} \beta_i f_i(y, x, \theta_i)$, where $0 \le \beta_i \le 1$, $\sum_{i=1}^{k} \beta_i = 1$. The *Efficiency* of a design ξ for detecting departures from $f_i(y, x, \theta_i)$ is defined as usual,

$$\mathrm{Eff}_{i,k+1}(\xi) = \frac{I_{i,k+1}(\xi)}{I_{i,k+1}(\xi_i^*)},$$

where $\xi_i^* = \arg\max_\xi I_{i,k+1}(\xi)$.

A new minimum efficiency criterion is defined here, which seems more adequate from a discrimination point of view,

$$I_m(\xi) = \min_{i \in \{1,\ldots,k\}} \mathrm{Eff}_{i,k+1}(\xi) = \left[\max_{i \in \{1,\ldots,k\}} \frac{1}{\mathrm{Eff}_{i,k+1}(\xi)}\right]^{-1}.$$

Equivalence theorem A design ξ_m^* is a max-min efficiency optimum design if, and only if there exists a measure $\bar{\eta}$ on $\mathcal{C}(\xi_m^*)$ such that

$$\int_{\mathcal{C}(\xi_m^*)} \psi(x, e_i, \xi_m^*)\, \bar{\eta}(de_i) \ge 0, \qquad \text{fo any} \quad x \in \chi. \tag{3}$$

Iterative algorithm 1 At step l, for any subset of l models $f_{i_1}, \ldots, f_{i_l}$, compute

$$\tilde{\xi} = \arg\max_{\xi} \mathrm{Eff}_{i_1,k+1}(\xi)$$

under the constraint $\text{Eff}_{i_1,k+1}(\xi) = \text{Eff}_{i_2,k+1}(\xi) = \cdots = \text{Eff}_{i_l,k+1}(\xi)$. If

$$\text{Eff}_{r,k+1}(\tilde{\xi}) > \text{Eff}_{i_1,k+1}(\tilde{\xi}) = \cdots = \text{Eff}_{i_l,k+1}(\tilde{\xi})$$

for any $r \neq i_j$ with $j = 1, \ldots, l$, then $\tilde{\xi} = \xi_m^*$ and the procedure stops. Otherwise go back to step 2.

A numerical method for computing the design which maximizes the common efficiency in step 3 is provided in the paper. It is necessary to assume that criterion $I_\alpha(\xi)$ is concave in order to guarantee there is a unique optimum design.

Iterative algorithm 2

Given a design ξ_s obtained at step s let

$$\begin{aligned} \theta_{i,s} &= \arg\min_{\theta_i \in \Omega_i} \int \mathcal{I}(f_{k+1}, f_i, x, \theta_i)\xi_s(dx) \\ x_{i,s} &= \arg\min_{x \in \chi} \mathcal{I}(f_{k+1}, f_i, x, \theta_{i,s}), \\ & \quad i = 1, \ldots, k. \end{aligned}$$

This means $x_{i,s}$ is the maximum of the partial derivative

$$\psi(x, e_i, \xi) = -I_{i,k+1}(\xi_i^*)\, \frac{\partial I_{i,k+1}(\xi, \xi_x)}{I_{i,k+1}^2(\xi)},$$

Let

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi_{x_s},$$

where $\xi_{x_s} = \left\{ \begin{array}{c} x_{1,s} \cdots x_{k,s} \\ p_{1,s} \cdots p_{k,s} \end{array} \right\}$ for $p_{i,s} = \frac{\mathcal{I}_{i,s}}{\sum_j \mathcal{I}_{j,s}}$. where $\mathcal{I}_{i,s} = \mathcal{I}(f_{k+1}, f_i, x_{i,s}, \theta_{i,s})$.

References:

[1] Atkinson A.C., Fedorov V.V. "The designs of experiments for discriminating between two rival models", *Biometrika*, **62**, 57–70, 1975a.

[2] Atkinson A.C., Fedorov V.V. "Optimal design: experiments for discriminating between several models", *Biometrika*, **62**, 289–303, 1975b.

[3] H. Dette and S. Titoff. "Optimal discrimination designs", *Annals of Statistics*, **37**:2056–2082, 2008.

[4] J. López Fidalgo, Ch. Tommasi, and C. Trandafir. "An optimal experimental design criterion for discriminating between non-normal models", *Journal of the Royal Statistical Society, Series B*, **69**(2):231–242, 2007.

[5] C. Tommasi. "Optimal designs for discriminating among several non-normal models", In López-Fidalgo, Rodríguez-Díaz J., and B. J.M., Torsney, editors, Advances in Model-Oriented Design and Analysis mODa 8, Physica- Verlag: Heidelberg, 213–220, 2007.

[6] C. Tommasi and López-Fidalgo J. "Bayesian optimum designs for discriminating between models with any distribution", *Computational Statistics & Data Analysis*, **54**(1):143–150, 2010.

[Jesus, López–Fidalgo; Edificio Politcnica, University of Castilla–La Mancha, Avda. Camilo José Cela, 3, Ciudad Real, Spain]
[`jesus.lopezfidalgo@uclm.es` – `http://www.uclm.es/profesorado/jesuslopezfidalgo/`]

Equilibrium in bargaining models

Vladimir V. Mazalov
Institute of Applied Mathematical Research, Petrozavodsk, Russia

We consider new trends in game-theoretical modeling of bargaining. Different approaches such as arbitration procedures, multi-staged bargaining, auctions, reputations, stable agreements, etc. are presented. We consider models with incomplete information related with deals between buyers and sellers. A player (buyer or seller) has a private information about his reserved price. Reserved prices are random variables with known probability distributions. Each player declares a price which depends on his reserved price. If the bid price is above the ask price, the good is sold for the average of two prices. Otherwise, there is no deal. We find an equilibrium in a n-threshold form as a solution of a system of algebraic equations. Some results of simulations are presented.

References:

[1] G. Loukas, and G. Oke, " Protection Against Denial of Service Attacks: A Survey", *The Computer Journal*, **53**(7):1020–1037, 2010.

[2]P, Jain, J. Jain, Z. Gupta, Mitigation of Denial of Service (DoS) Attack, IJCEM International Journal of Computational Engineering & Management, **11**:38–44, 2011.

[3] M.J. Hashmi, M. Saxena, R. Saini, Classification of DDoS Attacks and Their Defense Techniques Using Intrusion Prevention System, International Journal of Computer Science & Communication Networks, **2**(5):607–614, 2012.

[Vladimir Mazalov; Pushkinskaya st. 11, Petrozavodsk 185910, Russia]
[vlmazalov@yandex.ru – www.krc.karelia.ru]

Locally D-optimal designs for a nonlinear multiresponse regression model

Viatcheslav B.Melas
St.Petersburg State University, Russia

Let us consider the following nonlinear regression model

$$y_i = \eta(t_i, \theta) + \varepsilon_i, \quad t_i = 1, \ldots, N, \tag{4}$$

where $\eta(t;\theta) = \big(\eta_1(t;\theta), \eta_2(t;\theta), \eta_3(t;\theta)\big)^{\mathsf{T}}$, $t_i \in [0, \infty)$, $i = 1, \ldots, N$, ε_i stands for measurement noise, $\mathrm{E}\{\varepsilon_i\} = 0$, $\mathrm{E}\{\varepsilon_i \varepsilon_j^T\} = R\delta_{ij}$, δ_{ij} is the Kronecker delta, R is assumed to be known. This means that the observations at different time moments are not correlated, but there may be correlations between individual response components for the same time moment. Let us also assume that

$$\eta_1(t;\theta) = \exp(-\theta_1 t), \quad \eta_2(t;\theta) = \frac{\theta_1}{\theta_1 - \theta_2}[\exp(-\theta_2 t) - \exp(-\theta_1 t)], \quad \eta_3(t;\theta) = 1 - \eta_1(t;\theta) - \eta_2(t;\theta). \tag{5}$$

This model is closely connected with a problem in chemical kinetics and was studied in [1]. In that paper it was also proved that the D-optimal design problem for model (1)-(2) is equivalent to that for the model with two components, η_1, η_2 with R replaced by W^{-1}, where $W = W(R)$ is a $2x2$ matrix. Without loss of generality we can assume that W is of the form

$$\begin{pmatrix} a & b \\ b & 1 \end{pmatrix} \tag{6}$$

where a and b are some given numbers.

We would like to construct locally D-optimal designs for the initial nonlinear model that coincided with D-optimal design for the linearized model and depend on unknown proper values of parameters θ_1 and θ_2 as well as of numbers a and b. Note, however, that if we multiply both parameters by a constant, say c, and divide the argument by c our model remains the same. Thus without loss of generality we can assume that $\theta_1 + \theta_2 = c = 1$ in the problem of optimal design. For arbitrary c the optimal design support points should be simply multiplied by c. Let us introduce the new parameter $\Delta = 1 - \theta_1$, then $\theta_2 = 1 + \Delta$. The support points and the corresponding weights of the D-optimal design will be treated as functions of an auxiliary variables Δ, a and b.

In paper [1] it was analytically proved that if a and b satisfy the inequality

$$\frac{a + 1 - 2b}{a - b^2} > \gamma^*, \quad , \gamma^* \approx 2.872,$$

then the optimal design for sufficiently small Δ consists of a single support point and does not depend on a and b. The dependence of this point on Δ was presented by Taylor series using the functional approach. Here we investigate the case when the locally D-optimal design has two support points and expend its dependence on Δ, a and b in Taylor series by the functional approach. To this end we found analytically that the limit of the locally D-optimum design tends to the design having two support points

$$\left\{ \frac{3 - \sqrt{3}}{2}, \frac{3 + \sqrt{3}}{2} \right\}$$

with weights equal to $\frac{1}{2}$ if all these quantities tend to zero.

References:

[1] V.B. Melas, L. A. Krylova and D. Uciński, On the functional approach to locally D-optimum design for multiresponse models. mODa 10 - Advances in Model-Oriented Design and Analysis. Contribution in Statistics 2013, pp. 163-170.

[Viatcheslav Melas; St.Petersburg State University Faculty of Mathematics and Mechanics, University avenue, 28, 198504 St.Petersburg]
[vbmelas@post.ru – http://stmod.math.spbu.ru/staff:melas]

On the limits of the use of OLSE variances for the design of correlated observations

Andrej Pázman
Comenius University, Bratislava, Slovakia

Werner G. Müller
Johannes Kepler University, Linz, Austria

We discuss the circumstances for the use of the design criterion employed in Näther (1985) and recently popularized in Dette, Zhigljavsky and Peplyshev (2013). We will point out its merits as well as its limitations and present a simple example for the latter.

References:

[1] Dette, H., Pepelyshev, A., Zhigljavsky, A., 2013. Optimal design for linear models with correlated observations. *The Annals of Statistics* 41 (1), 143-176.

[2] Näther, W., 1985. *Effective observation of random fields. Teubner-Texte zur Mathematik.* Teubner, Leipzig.

[Werner G. Müller; Johannes Kepler University, Altenberger Straße 69, A-4040 Linz, Austria]
[`werner.mueller@jku.at` – `www.jku.at/ifas`]

Balancing local and global information in simplicial global optimization algorithms

Remigijus Paulavičius
Julius Žilinskas
Vilnius University Institute of Mathematics and Informatics, Lithuania

Yaroslav D. Sergeyev
Dmitri E. Kvasov
Calabria University, Rende (CS), Italy
N. I. Lobachevsky University of Nizhni Novgorod, Russia

In this talk, we consider a global optimization problem for a Lipschitz-continuous functions with an unknown Lipschitz constant. Recently, it has been shown that simplicial-partition-based Disimpl algorithm [4] gives very competitive results to Direct [3] for standard test functions and performs particularly well when the search space and the numbers of local and global optimizers may be reduced according to symmetries [4,5]. However, Disimpl algorithm (similarly to original Direct algorithm) quickly gets close to the optimum but takes longer to achieve a high degree of accuracy.

The present talk is concerned with overcoming this drawback by using the two-phase approach [7]. A globally-biased technique of balancing local and global information during partition is proposed [6], experimentally investigated and compared with the well-known Direct and Direct$l$ [1] methods. Extensive numerical experiments, performed on 800 test functions randomly-generated by the GKLS-generator [2], reveal a significant improvement of the new approach implemented in Gb-Disimpl algorithm. This advantage is more evident for harder global optimization problems.

References:

[1] J. M. Gablonsky and C. T. Kelley, "A Locally-Biased form of the Direct", *Journal of Global Optimization*, **21**(1):27–37, 2001.

[2] M. Gaviano, D.E. Kvasov, D. Lera, Y.D. Sergeyev, "Algorithm 829: Software for generation of classes of test functions with known local and global minima for global optimization", *ACM Transactions on Mathematical Software (TOMS)*, **29**(4):469–480, 2003.

[3] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant", *Journal of Optimization Theory and Application*, **79**(1):157–181, 1993.

[4] R. Paulavičius and J. Žilinskas, "Simplicial Lipschitz optimization without the Lipschitz constant", *Journal of Global Optimization*, in press, 2013.

[5] R. Paulavičius and J. Žilinskas, "Simplicial global optimization", *SpringerBriefs in Optimization*, Springer, New York, 2014.

[6] R. Paulavičius, Ya. D. Sergeyev, D. E. Kvasov, and J. Žilinskas, "Globally-biased Disimpl algorithm for expensive global optimization". *Submitted*, 2013.

[7] Ya. D. Sergeyev and D. E. Kvasov, "Global search based on efficient diagonal partitions and a set of Lipschitz constants", *SIAM Journal on Optimization*,**16**(3):910–937, 2006.

[Remigijus Paulavičius; Vilnius University Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania]
[`remigijus.paulavicius@mii.vu.lt` – `www.mii.vu.lt`]

Optimal design for one-parameter models with correlated observations

Holger Dette
Bochum University, Germany

Andrey Pepelyshev
Anatoly Zhigljavsky
Cardiff University, UK

Suppose that results of experiments satisfy the regression model

$$y_j = \theta f(x_j) + \varepsilon_j,$$

where $x_j \in [a, b]$.

Consider the signed LSE $\hat{\theta}_S$ for the design $\xi = \{x_1, \dots, x_n; p_1, \dots, p_n\}$ and signs $s_1, \dots, s_n \in \{-1, 1\}$.

The variance of $\hat{\theta}_S$ has the form

$$D(\xi) = \sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) f(x_i) f(x_j) s_i s_j p_i p_j \Big/ \Big(\sum_{i=1}^{n} f(x_i) s_i p_i \Big)^2;$$

we consider this expression as the design optimality criterion. Denote $w_i = s_i p_i$ and call it signed weight of a point x_i in the design ξ.

In the talk, new findings on the optimal signed designs in the case when the covariance kernel $K(\cdot, \cdot)$ has the multiplicative structure will be presented. The research is based on our previous results obtained in [1,2].

References:

[1] Dette H., Pepelyshev A., Zhigljavsky A. Optimal design for linear models with correlated observations. *Annals of Statistics*, **41**:143–176, 2013.

[2] Zhigljavsky A., Dette H., Pepelyshev A. A new approach to optimal design for linear models with correlated observations. *J. Amer. Stat. Ass.* **105**:1093–1103, 2010.

[Andrey Pepelyshev; School of Mathematics, Cardiff University, Sennghennydd Road, Cardiff, CF24 4YH, UK]
[pepelyshevan@cardiff.ac.uk – http://ssa.cf.ac.uk/pepelyshev/]

On the perturbative stability of SSA and MSSA forecasts

Karl Michael Schmidt
Maria Vronskaya
Cardiff University, Wales, UK

We discuss the stability of SSA and MSSA forecasts of time series under random perturbations of the input data, observing that some parts of the process, e.g. the SSA reconstruction of the middle part of the time series, are surprisingly stable. For two-series MSSA, we study how the strength of noise propagation for the forecast of the first (main) series depends on scaling of the second (support) series. First results indicate that this may serve as a tool for the discovery of structural relationships between the series. In particular, in view of a general analogy, despite essential differences, of MSSA with a (stationary) VAR stochastic process, this is related to Granger's causality concept.

We also discuss the scaling problem for MSSA forecasting and suggest a way to overcome it, at least partially, by linearisation.

[Karl Michael Schmidt; School of Mathematics, Cardiff University, Sennghennydd Road, Cardiff, CF24 4YH, UK]
[SchmidtKM@cf.ac.uk – www.cf.ac.uk/maths]

Design for mixed models: What is good for all need not be good for everyone

Rainer Schwabe
Maryna Prus
Otto-von-Guericke University, Magdeburg, Germany

In data analysis random effects have become a standard when individual subjects are involved. This is not only valid for medical or biological applications, but this approach receives also increasing interest in economic and social sciences under the name of panel data or in official and regional statistics under the notation of small area estimation.

In opposite to this trend the development of corresponding designs attracts less interest or is even regarded as not substantial. This may be caused by a common believe that everything about designs for linear mixed models is known for decades and that either standard designs should be used, which disregard the random effects, or some true Bayesian designs, which involve the precision of the random effects as prior knowledge.

Our aim is to demonstrate that this believe might be terribly misleading and that the use of the proposed designs may result in an insufficient analysis. This is particularly the case, when the subjects in the study are also the objects of investigation. Instead an Empirical Bayes approach seems to be adequate. We finally also point out that efficient designs for statements on the mean response in the population ("all") may show an unsatisfactory performance for the individual response of the subjects ("everyone").

[Rainer Schwabe; Otto-von-Guericke University, Institute for Mathematical Stochastics, PF 4120, 39016 Magdeburg, Germany]
[rainer.schwabe@ovgu.de – http://www.math.uni-magdeburg.de/institute/imst/ag_schwabe]

Lipschitz global optimization

Yaroslav D. Sergeyev

Calabria University, Rende (CS), Italy
N. I. Lobachevsky University of Nizhni Novgorod, Russia

Global optimization is a thriving branch of applied mathematics and an extensive literature is dedicated to it (see e.g., [1–21]). In this lecture, the global optimization problem of a multidimensional function satisfying the Lipschitz condition over a hyperinterval with an unknown Lipschitz constant is considered. It is supposed that the objective function can be "black box", multiextremal, and non-differentiable. It is also assumed that evaluation of the objective function at a point is a time-consuming operation. Many algorithms for solving this problem have been discussed in literature. They can be distinguished, for example, by the way of obtaining information about the Lipschitz constant and by the strategy of exploration of the search domain. Different exploration techniques based on various adaptive partition strategies are analyzed.

The main attention is dedicated to two types of algorithms. The first of them is based on using space-filling curves in global optimization. A family of derivative-free numerical algorithms applying space-filling curves to reduce the dimensionality of the global optimization problem is discussed. A number of unconventional ideas, such as adaptive strategies for estimating Lipschitz constant, balancing global and local information to accelerate the search, etc. are presented.

Diagonal global optimization algorithms is the second type of methods under consideration. They have a number of attractive theoretical properties and have proved to be efficient in solving applied problems. In these algorithms, the search hyperinterval is adaptively partitioned into smaller hyperintervals and the objective function is evaluated only at two vertices corresponding to the main diagonal of the generated hyperintervals. It is demonstrated that the traditional diagonal partition strategies do not fulfil the requirements of computational efficiency because of executing many redundant evaluations of the objective function.

A new adaptive diagonal partition strategy that allows one to avoid such computational redundancy is described. Some powerful multidimensional global optimization algorithms based on the new strategy are introduced. Results of extensive numerical experiments performed on the GKLS-generator (see [2]) to test the proposed methods demonstrate their advantages with respect to traditional diagonal algorithms in terms of both number of trials of the objective function and qualitative analysis of the search domain, which is characterized by the number of generated hyperintervals.

A number of possible generalizations to problems with multiextremal partially generated constraints is mentioned. The usage of parallel computations is discussed briefly and theoretical results on the possible speed-up are presented.

Acknowledgments. This work was supported by the Ministry of Education and Science of Russian Federation, project 14.B37.21.0878, as well as by the grants n. 11-01-00682-a of the Russian Foundation for Fundamental Research and n. 1960.2012.9 awarded by the President of the Russian Federation for supporting the leading research groups.

References:

[1] C. A. Floudas and P. M. Pardalos, *State of the Art in Global Optimization*. Kluwer, Dordrecht, 1996.

[2] M. Gaviano, D. E. Kvasov, D. Lera, and Ya. D. Sergeyev, "Algorithm 829: Software for generation of classes of test functions with known local and global minima for global optimization", *ACM Transactions on Mathematical Software*, **29**(4):469–480, 2003.

[3] R. Horst and P. M. Pardalos, *Handbook of Global Optimization*, Kluwer, Dordrecht, 1995.

[4] D. E. Kvasov and Ya. D. Sergeyev, "A univariate global search working with a set of Lipschitz

constants for the first derivative", *Optimization Letters*, **3**(2):303–318, 2009.

[5] D. E. Kvasov and Ya. D. Sergeyev, "Lipschitz gradients for global optimization in a one-point-based partitioning scheme", *Journal of Computational and Applied Mathematics*, **236**(16):4042–4054, 2012.

[6] D. E. Kvasov and Ya. D. Sergeyev, "Univariate geometric Lipschitz global optimization algorithms", *Numerical Algebra, Control and Optimization*, **2**(1):69–90, 2012.

[7] D. Lera and Ya. D. Sergeyev, "Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives", *SIAM Journal on Optimization*, **23**(1):508–529, 2013.

[8] R. Paulavičius and J. Žilinskas, *Simplicial Global Optimization*, Springer, New York, 2014 (doi: 10.1007/978-1-4614-9093-7).

[9] J. Pintér, *Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications*, Kluwer, Dordrecht, 1996.

[10] Ya. D. Sergeyev, "An information global optimization algorithm with local tuning", *SIAM Journal on Optimization*, **5**(4):858–870, 1995.

[11] Ya. D. Sergeyev, "Global one-dimensional optimization using smooth auxiliary functions", *Mathematical Programming*, **81**(1):127–146, 1998.

[12] Ya. D. Sergeyev and D. E. Kvasov, "Global search based on efficient diagonal partitions and a set of Lipschitz constants", *SIAM Journal on Optimization*, **16**(3):910–937, 2006.

[13] Ya. D. Sergeyev and D. E. Kvasov, *Diagonal Global Optimization Methods*, FizMatLit, Moscow, 2008. In Russian.

[14] Ya. D. Sergeyev and D. E. Kvasov, "Lipschitz global optimization", in J. J. Cochran et al., (Eds.), *Wiley Encyclopedia of Operations Research and Management Science (in 8 volumes)*, John Wiley & Sons, New York, **4**:2812–2828, 2011.

[15] Ya. D. Sergeyev, R. G. Strongin, and D. Lera, *Introduction to Global Optimization Exploiting Space-Filling Curves*, Springer, New York, 2013.

[16] R. G. Strongin, *Numerical Methods in Multi-Extremal Problems (Information-Statistical Algorithms)*, Nauka, Moscow, 1978. In Russian.

[17] R. G. Strongin and Ya. D. Sergeyev, *Global Optimization with Non-Convex Constraints: Sequential and Parallel Algorithms*, Kluwer, Dordrecht, 2000.

[18] A. A. Zhigljavsky, *Mathematical Theory of the Global Random Search*, St.Petersburg University Press, 1985. In Russian.

[19] A. A. Zhigljavsky, *Theory of Global Random Search*, Kluwer, Dordrecht, 1991.

[20] A. A. Zhigljavsky and A. Žilinskas, *Search for Global Extremum*, Nauka, Moscow, 1991. In Russian.

[21] A. A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization*, Springer, New York, 2008.

[Yaroslav D. Sergeyev; DIMES, Università della Calabria, via P. Bucci, Cubo 42-C, 87036 Rende (CS), Italy]
[`yaro@si.deis.unical.it` – `http://wwwinfo.deis.unical.it/~yaro/`]

The Infinity Computer and numerical computations with infinities and infinitesimals

Yaroslav D. Sergeyev

Calabria University, Rende (CS), Italy
N. I. Lobachevsky University of Nizhni Novgorod, Russia

The lecture introduces a new methodology allowing one to execute numerical computations with finite, infinite, and infinitesimal numbers (see [1–19]) on a new type of a computer – the Infinity Computer (see EU, USA, and Russian patents [8]). The new approach is based on the principle 'The whole is greater than the part' (Euclid's Common Notion 5) that is applied to all numbers (finite, infinite, and infinitesimal) and to all sets and processes (finite and infinite). It is shown that it becomes possible to write down finite, infinite, and infinitesimal numbers by a finite number of symbols as particular cases of a unique framework different from that of the non-standard analysis. The new methodology (see survey [9]) evolves ideas of Cantor and Levi-Civita in a more applied way and, among other things, introduces new infinite integers that possess both cardinal and ordinal properties as usual finite numbers (its relations with traditional approaches are discussed in [4,5]).

It is emphasized that the philosophical triad – researcher, object of investigation, and tools used to observe the object – existing in such natural sciences as Physics and Chemistry, exists in Mathematics, too. In natural sciences, the instrument used to observe the object influences the results of observations. The same happens in Mathematics where numeral systems used to express numbers are among the instruments of observations used by mathematicians. The usage of powerful numeral systems gives the possibility to obtain more precise results in Mathematics, in the same way as the usage of a good microscope gives the possibility to obtain more precise results in Physics. A new numeral system allowing one to express easily infinities and infinitesimals offers exciting capabilities for describing mathematical objects, mathematical modeling, and practical computations. The concept of the accuracy of numeral systems is introduced. The accuracy of the new numeral system is compared with traditional numeral systems used to work with infinity.

The new computational methodology gives the possibility to execute computations of a new type and simplifies fields of Mathematics where the usage of the infinity and/or infinitesimals is necessary (e.g., divergent series, limits, derivatives, integrals, measure theory, probability theory, optimization, fractals, etc., see [1–7,9–19]). Numerous examples and applications are given. In particular, a number of results related to the First Hilbert Problem are established.

The main attention in the lecture is dedicated to the explanation of how practical numerical computations with infinities and infinitesimals can be executed (e.g., see [2,13,16] for applications in optimization, numerical differentiation, and ODEs). The Infinity Calculator using the Infinity Computer technology is presented during the talk.

Numerous research articles of several authors and a lot of an additional information can be downloaded from the page http://www.theinfinitycomputer.com

Acknowledgments. This work was supported by the Ministry of Education and Science of Russian Federation, project 14.B37.21.0878.

References:

[1] L. D'Alotto, "Cellular automata using infinite computations", *Applied Mathematics and Computation*, **218**(16), 8077-8082, 2012.

[2] S. De Cosmis and R. De Leone, "The use of Grossone in Mathematical Programming and Operations Research", *Applied Mathematics and Computation*, **218**(16), 8029-8038, 2012.

[3] D. I. Iudin, Ya. D. Sergeyev, and M. Hayakawa, "Interpretation of percolation in terms of infinity computations", *Applied Mathematics and Computation*, **218**(16), 8099-8111, 2012.

[4] G. Lolli, "Infinitesimals and infinites in the history of Mathematics: A brief survey", *Applied*

Mathematics and Computation, **218**(16), 7979-7988, 2012.

[5] M. Margenstern, "Using Grossone to count the number of elements of infinite sets and the connection with bijections", *p-Adic Numbers, Ultrametric Analysis and Applications*, **3**(3), 196-204, 2011.

[6] M. Margenstern, "An application of Grossone to the study of a family of tilings of the hyperbolic plane", *Applied Mathematics and Computation*, **218**(16), 8005-8018, 2012.

[7] Ya. D. Sergeyev, *Arithmetic of Infinity*, Edizioni Orizzonti Meridionali, CS, (2003 paperback edition, 2013 revised electronic edition).

[8] Ya. D. Sergeyev, *Computer system for storing infinite, infinitesimal, and finite quantities and executing arithmetical operations with them*, EU patent 1728149, issued 03.06.2009; RF patent 2395111, issued 20.07.2010; USA patent 7,860,914 issued 28.12.2010.

[9] Ya. D. Sergeyev, "A new applied approach for executing computations with infinite and infinitesimal quantities", *Informatica*, **19**(4), 567-596, 2008.

[10] Ya. D. Sergeyev, "Evaluating the exact infinitesimal values of area of Sierpinski's carpet and volume of Menger's sponge", *Chaos, Solitons & Fractals*, **42**, 3042-3046, 2009.

[11] Ya. D. Sergeyev, "Numerical point of view on Calculus for functions assuming finite, infinite, and infinitesimal values over finite, infinite, and infinitesimal domains", *Nonlinear Analysis Series A: Theory, Methods & Applications*, **71**(12), e1688-e1707, 2009.

[12] Ya. D. Sergeyev, "Counting systems and the First Hilbert problem", *Nonlinear Analysis Series A: Theory, Methods & Applications*, **72**(3-4), 1701-1708, 2010.

[13] Ya. D. Sergeyev, "Higher order numerical differentiation on the Infinity Computer", *Optimization Letters*, **5**, 575-585, 2011.

[14] Ya. D. Sergeyev, "On accuracy of mathematical languages used to deal with the Riemann zeta function and the Dirichlet eta function", *p-Adic Numbers, Ultrametric Analysis and Applications*, **3**(2), 129-148, 2011.

[15] Ya. D. Sergeyev, "Using blinking fractals for mathematical modeling of processes of growth in biological systems", *Informatica*, **22**(4), 559-576, 2011.

[16] Ya. D. Sergeyev, "Solving ordinary differential equations by working with infinitesimals numerically on the Infinity Computer", *Applied Mathematics and Computation*, **219**(22), 10668-10681, 2013.

[17] Ya. D. Sergeyev and A. Garro, "Observability of Turing Machines: a refinement of the theory of computation", *Informatica*, **21**(3), 425-454, 2010.

[18] Ya. D. Sergeyev and A. Garro, "Single-tape and multi-tape Turing machines through the lens of the Grossone methodology", *Journal of Supercomputing*, **65**(2), 645-663, 2013.

[19] A. A. Zhigljavsky, "Computing sums of conditionally convergent and divergent series using the concept of grossone", *Applied Mathematics and Computation*, **218**(16), 8064-8076, 2012.

[Yaroslav D. Sergeyev; DIMES, Università della Calabria, via P. Bucci, Cubo 42-C, 87036 Rende (CS), Italy]
[`yaro@si.deis.unical.it` – `http://wwwinfo.deis.unical.it/~yaro/`]

Optimal design, Lagrangian and linear model theories: a fusion

Ben Torsney
School of Mathematics & Statistics, University of Glasgow

The general approximate optimal design problem (P1) aims to maximise a criterion of several variables, subject to them being nonnegative and summing to 1. If all variables must be positive, a necessary condition of optimality is that all vertex directional derivatives be zero; equivalently all partial derivatives should share a common value, a Lagrange Multiplier value.

We consider the problem of optimizing a criterion of several variables, subject to them satisfying several (non-linear) equality constraints. Lagrangian Theory requires that at an optimum all partial derivatives be exactly linear in a set of Lagrange Multipliers. It seems we can argue that the partial derivatives, viewed as response variables, must exactly satisfy a Linear Model with the Lagrange Multipliers as parameters. This then is a model without errors implying a fitted model with zero residuals. The residuals appear to play the role of directional derivatives.

Further, if we must have all variables nonnegative, we might exploit the multiplicative algorithm for (P1). Strictly speaking this has two steps:

1. A multiplicative step, under which we multiply each variable by a positive function, say g(.), of its vertex directional derivative or of its partial derivative;. and
2. A scaling step to ensure the variables sum to 1.

The multiplicative step naturally extends to our more general problem, but some deeper consideration is needed to devise a "scaling" step to "scale" the resultant products to (approximately) satisfy the required set of equality constraints.

We will explore this idea in the case when the variables form a matrix which must satisfy fixed row and column sum constraints. It can be seen that partial derivatives must exactly satisfy a linear model, additive in two sets of "main effect" parameters..

[Ben Torsney; School of Mathematics & Statistics, University of Glasgow, Mathematics Building, 15 University Gardens, Glasgow G12 8QW]
[Bernard.Torsney@glasgow.ac.uk –]

Decomposing multivariate polynomials with structured low-rank approximation

Konstantin Usevich
Vrije Universiteit Brussel, Brussels, Belgium

We are focused on numerical methods for decomposing a multivariate polynomial as a sum of univariate polynomials in linear forms. The main tool is the recent result on equivalence between the Waring rank of a homogeneous polynomial and the rank of a partially known quasi-Hankel matrix constructed from the coefficients of the polynomial. Based on this equivalence, we show that the original decomposition problem can be reduced to structured low-rank matrix completion (or to structured low-rank approximation in the case of approximate decomposition). Several methods for structured low-rank approximation/completion are compared.

Notation Let $\mathbb{R}[\mathbf{z}]$ and $\mathbb{C}[\mathbf{z}]$ denote the vector spaces of n-variate polynomials with real and complex coefficients, where $\mathbf{z} = \begin{bmatrix} z_1 & \cdots & z_n \end{bmatrix}^\top$. We also denote by $\mathbb{R}_d[\mathbf{z}]$ and $\mathbb{C}_d[\mathbf{z}]$ the spaces of homogeneous polynomials of degree d, and by $\mathbb{R}_{\leq d}[\mathbf{z}]$ and $\mathbb{C}_{\leq d}[\mathbf{z}]$ the spaces of multivariate polynomials with degree at most d.

Problem 1 *Given a polynomial $p \in \mathbb{K}_{\leq d}[\mathbf{z}]$, where $\mathbb{K} = \mathbb{R}, \mathbb{C}$, decompose it into a sum of r univariate polynomials in linear forms, i.e. find a decomposition (possibly minimal)*

$$p(\mathbf{z}) = \widetilde{p}_1(\mathbf{s}_1^\top \mathbf{z}) + \cdots + \widetilde{p}_r(\mathbf{s}_r^\top \mathbf{z}),$$

where $\widetilde{p}_1, \ldots, \widetilde{p}_r \in \mathbb{K}_{\leq d}[t]$ and $\mathbf{s}_1, \ldots, \mathbf{s}_r \in \mathbb{K}^n$.

We are also interested in approximation of multivariate polynomials by sums of univariate polynomials of linear forms. This problem is motivated by approximation of a multivariate function $f(\mathbf{z})$ by sums of *ridge functions* $\widetilde{f}(\mathbf{s}^\top \mathbf{z})$, where $\mathbf{s}^\top \mathbf{z}$ is a linear form and $\widetilde{f}$ is a univariate function. Approximation by ridge functions appears in many problems in statistics and data analysis, see e.g. [1,2]. Recently, Problem 1 was considered in the context of identification of block-structured nonlinear dynamical systems [3].

Problem 1 is closely related to a classical problem in algebra [4].

Problem 2 (Waring decomposition) *Given a homogeneous polynomial $a \in \mathbb{C}_d[\mathbf{z}]$, find the minimal r and vectors $\mathbf{s}_1, \ldots, \mathbf{s}_r \in \mathbb{C}^n$ such that*

$$a(\mathbf{z}) = (\mathbf{s}_1^\top \mathbf{z})^d + \cdots + (\mathbf{s}_r^\top \mathbf{z})^d. \tag{7}$$

The minimal possible r is called the Waring rank of a (denoted by $\operatorname{wrank}(a)$*).*

Problem 2 is equivalent to symmetric tensor decomposition of a symmetric tensor, see e.g. [5]. A recent result [6] states that generically (for a generic polynomial a)

$$\operatorname{wrank}(a) \leq r \iff \operatorname{rank} \mathscr{H}(a, e) \leq r,$$

where $\mathscr{H}(a, e)$ is a linearly structured matrix, which is composed of the coefficients of the polynomial $a(\mathbf{z})$ and some unknown (latent) variables e. The matrix $\mathscr{H}(a, e)$ is *quasi-Hankel* (a multidimensional generalization of Hankel matrices). Therefore, Problem 2 with known r is equivalent to low-rank completion of the quasi-Hankel $\mathscr{H}(a, e)$ (finding the latent variables e given the coefficients of the polynomial a, such that $\operatorname{rank} \mathscr{H}(a, e) \leq r$).

In this talk, we extend the approach of [6] for Problem 1 (in the case $\mathbb{K} = \mathbb{C}$) by stacking together quasi-Hankel matrices corresponding to different homogeneous components of the original polynomial $p(\mathbf{z})$. We compare several methods of low-rank completion of structured matrices, including the recently proposed penalty-based method [7]. The advantage of the approach in [7] is its ability

to solve both low-rank approximation and low-rank completion problems. In addition, the method in [7] does not require solving systems of polynomial equations, in contrast to the method of [6].

References:

[1] D. Donoho and I. M. Johnstone, "Projection-based approximation and a duality with kernel methods", *The Annals of Statistics*, **17**(1):58–106, 1989.

[2] V.Y. Lin and A. Pinkus, "Fundamentality of ridge functions", *Journal of Approximation Theory*, **75**(3):295–311, 1993.

[3] M. Schoukens and Y. Rolain, "Cross-term elimination in parallel Wiener systems using a linear input transformation", *IEEE Transactions on Instrumentation and Measurement*, **61**(3):845–847, 2012.

[4] A. Iarobbino and V. Kanev, *Power sums, Gorenstein Algebras and Determinantal Loci*, volume 1721 of *Lecture Notes in Mathematics*, Springer, 1999.

[5] P. Comon, G. Golub, L.-H. Lim and B. Mourrain, "Symmetric tensors and symmetric tensor rank", *SIAM Journal on Matrix Analysis and Applications*, **30**(3):1254–1279, 2008.

[6] J. Brachat, P. Comon, B. Mourrain and E. Tsigaridas, "Symmetric tensor decomposition", *Linear Algebra and its Applications*, **433** (11–12):1851–1872, 2010.

[7] M. Ishteva, K. Usevich and I. Markovsky, "Regularized structured low-rank approximation", arXiv preprint, `http://arxiv.org/abs/1308.1827`.

[Konstantin Usevich; Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Elsene, Belgium]
[`kusevich@vub.ac.be` – `http://homepages.vub.ac.be/~kusevich/`]

Experimental design and learning with ABCD: approximate Bayes computation design

Markus Hainy
Werner G. Müller
Johannes Kepler University of Linz, Austria

Henry P. Wynn
London School of Economics, UK

Bayesian optimal design theory is well-established. It is shown how this can be combined with Approximate Bayes Computation (ABC) to give a very general framework for solving optimum experimental design problems. The present paper extends the work in [1]. A condensed version of ABCD first sets up the general design problem for choice of a sampling distribution $f(y, \theta, D)$ for observations Y conditional on a parameter θ, obtainable with a design D. When θ has the prior distribution $\pi(\theta)$ one seeks to minimise over the choice of design the preposterior risk:

$$\psi(D) = \min_D \mathrm{E}_Y \phi\{\pi(\theta|Y, D)\},$$

where $\pi(\theta|Y, D)$ is the posterior distribution, ϕ is some prespecified loss functional and E_Y is marginal expectation with respect to Y. In ABCD we are essentially using an approximation to $\pi(\theta|Y, D)$, based on the well-known ABC methods but in such a way that the often problematical "outside" integral embodied in E_Y does not need additional simulation. This is because the marginal distribution of Y is constructed as part of the ABC process. This was exploited by P. Muller and coworkers [2], who make extensive use of MCMC. The main expense of ABCD arises from having to vary the design D as part of the optimization process.

The method is highly suited to classical statistical learning theory, using information criteria, which has a long history: Renyi, Lindley, de Groot and others. The key theorem (attributed to de Groot) is that a functional ϕ, above, is *learnable*, in that its value on the prior distribution, $\phi\{\pi(\theta|Y, D)\}$, is less than or equal to $\phi(D)$, for all $f(y, \theta)$ and all $\pi(\theta)$, if and only if it is a convex functional on the space of distributions. The condition includes Shannon information (as negative risk) and all Tsallis informations. There is also a strong connection with a majorization-type peakness condition, in these cases. Another large class of metric-based functionals (generalised variances) are also included with a remarkable link to metric embedding theory.

Give this strong theoretical and computational basis one can expand the notation of an experimental design to cover areas such as selective sampling and other "directed attention" problems which arise in cognitive science and computer vision. It is hoped, in this way, to contribute to modern learning theory, often associated with machine learning.

References:

[1] M. Hainy, W.G. Müller and H.P. Wynn. Approximate Bayesian Computation Design, In Proceedings of the 10th International Workshop on Model-Oriented Design and Analysis, Poland: mODa 10. 135–143, 2013.

[2] P. Müller, B. Sansó and M. Iorio. Optimal Bayesian Design by inhomogeneous Markov chain simulation. J. Amer. Statist. Assoc. 99, 788-798, 2004.

[Henry P Wynn; London School of Economics, Houghton Street, London WC2A 2AE, UK]
[h.wynn@lse.ac.uk – http://www.lse.ac.uk/CATS/Home.aspx]

Multi-objective optimization algorithms based on statistical models of objectives

Antanas Žilinskas
Vilnius University, Lithuania

Nonlinear multi-objective optimization is very active research area. Depending on the properties of a multi-objective optimization problem, different approaches to its solution can be applied. The best direction developed is optimization of convex problems; for the latter problems the methods that generalize the ideas of classical mathematical programming suit well [1]. For the problems with not so nice objectives, metaheuristic methods are frequently favorable [2]. However, there remains a class of important problems without sufficient attention of researchers: namely, the problems with "black-box", multimodal, and "expensive" objectives. Such problems frequently occur in applications where objectives are available either as a computational model or as software. A well established approach to single objective optimization of "black-box", multimodal, and "expensive" objectives is based on statical models of objectives. In the present paper we discuss the possibilities to generalize that approach to the multi-objective case. We focus on the problems where objective functions are "expensive" because of the complexity of the computational model; "expensiveness" here means long lasting computation of a value of the objective function. The complexity of the computational model normally implies not only the "expensiveness" of the objective function but also the uncertainty in its properties. "Black box" optimization of "expensive" functions with many respects is quite opposite to the optimization of objective functions defined by analytical formulae. The limitation in collecting general information about the function, and particulary about its minima, strongly requires the rationality in distribution the points where to compute the objective function values. Therefore the algorithms justified by the principles of rational decision theory here are of especial interest. To construct such algorithms in the single objective optimization case the statistical models of multimodal functions have been proven very helpful [3],[4]. Recently several papers have been published which propose multi-objective optimization algorithms generalizing single-objective optimization algorithms based on statistical models of objective functions. These algorithms are straightforward generalizations of the single-objective prototypes, and their theoretical analysis is absent. In the present paper we discuss theoretical results substantiating the use of statistical models of objective functions in multi-objective optimization algorithms, as well as properties of the algorithms based on those models.

References:

[1] K. Miettinen, *Nonlinear Multi-Objective Optimization*, Kluwer Academic Publishers, Boston, 1999.

[2] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, J.Wiley, Chichester, 2009.

[3] A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization*, Springer, New York, 2008.

[4] R. Strongin and Y. Sergeyev, *Global Optimization with Non-Convex Constraints*, Kluwer Academic Publishers, Dodrecht, 2000.

[Antanas Žilinskas; Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, Vilnius, LT-08663, Lithuania]
[antanas.zilinskas@mii.vu.lt – http://www.mii.lt/index.php?siteaction=personnel.view&id=151]

On bounding vectors and bounding fronts in multi-objective branch and bound algorithms

Julius Žilinskas
Antanas Žilinskas
Vilnius University, Lithuania

The main concept of branch and bound is to detect subsets of feasible solutions which cannot contain optimal solutions. In multi-objective optimization bounding front is used – a set of bounding vectors in objective space dominating all possible objective vectors corresponding to the subset of feasible solutions. The subset cannot contain Pareto optimal solutions if each bounding vector in bounding front corresponding to this subset is dominated by at least one already known decision vector in the current solution set. The simplest bounding front corresponds to a single ideal vector composed of lower bounds for each objective function. However, the bounding fronts with multiple bounding vectors may be tighter and therefore their use may discard more subsets of feasible solutions. In this talk we investigate the use of bounding vectors and bounding fronts for some multi-objective optimization problems.

[Julius Žilinskas; Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, Vilnius, LT-08663, Lithuania]
[julius.zilinskas@mii.vu.lt – http://www.mii.vu.lt/index.php?siteaction=personnel.view&id=442&lang=en]