Adaptive Design of Experiments Based on Gaussian Processes

Evgeny Burnaev[1,2,3] and Maxim Panov[1,2,3](✉)

[1] Institute for Information Transmission Problems, Bolshoy Karetny per. 19, Moscow 127994, Russia
[2] DATADVANCE, LLC, Pokrovsky blvd. 3, Moscow 109028, Russia
[3] PreMoLab, MIPT, Institutsky per. 9, Dolgoprudny 141700, Russia
evgeny.burnaev@datadvance.net, panov.maxim@gmail.com

Abstract. We consider a problem of adaptive design of experiments for Gaussian process regression. We introduce a Bayesian framework, which provides theoretical justification for some well-know heuristic criteria from the literature and also gives an opportunity to derive some new criteria. We also perform testing of methods in question on a big set of multidimensional functions.

Keywords: Active learning · Computer experiments · Sequential design · Gaussian processes

1 Introduction

Construction of an approximation for an unknown dependency is one of the main steps of surrogate modelling (modelling based on the data) [1,2]. One of the most popular approaches is based on the Gaussian process model [3,4]. It is used in a variety of applied projects including conceptual design, structural optimization, multicriteria optimization, design in aerospace and automobile industries.

In many engineering applications the number of evaluations of a target function $f(\mathbf{x})$ is sufficiently limited due to a long execution time of one evaluation and/or its high cost. That's why the problem of creating methods for construction of a training set D_N of limited size N in a way to get the best possible quality of approximation is of the great importance. The design of experiments (DoE), which optimizes some statistical criterion, is called optimal design of experiments. For parametric models the problem of design of experiments, which optimizes the quality of approximation, is very close to the problem of optimal experimental design for estimation of the model's parameters. The theory of optimal DoE for parametric models is well-developed, see [5,6] and many others. However, in the situation, when approximation model is non-parametric (for example, Gaussian process regression model), we come to the different statement of the DoE problem [7,8], in which the DoE is constructed in a way to obtain the best possible prediction of the target function value. In this paper we will consider the adaptive variant of such a problem which recently gained much

A. Gammerman et al. (Eds.): SLDS 2015, LNAI 9047, pp. 116–125, 2015.
DOI: 10.1007/978-3-319-17091-6_7

attention [1,9]. The adaptive DoE process constructs training set iteratively and on each iteration design points are selected based on the properties of the approximation constructed on the previous iteration.

In this paper we focus on the adaptive design of experiments for Gaussian process (or kriging) regression models. This kind of models along with a prediction of the unknown function $f(\mathbf{x})$ allows to estimate a variance of the prediction $\sigma^2(\mathbf{x})$. Such a point-wise error estimate gives an opportunity to construct different strategies for reducing uncertainty of the approximation and therefore improving its quality. This area of research originates from the field of design and analysis of computer experiments [10–12]. The adaptive sampling strategies for improving the quality of approximation are closely related to the surrogate-based optimization [13], sequential designs for estimating the probability of a failure [14] and adaptive designs for estimating the sensitivity indices [15].

This paper makes a contribution in two main directions. First of all, we provide a Bayesian framework which allows to define optimal adaptive sampling strategies. Then we show, that many of the commonly used adaptive sampling criteria are approximations of optimal one-step look-ahead strategies. Also some new adaptive sampling strategies are obtained as optimal one-step look-ahead solutions. Second, we make an extensive testing of the most of currently used adaptive sampling approaches for the Gaussian process regression and compare them numerically.

2 Gaussian Process Regression

Let $y = f(\mathbf{x})$ be some unknown function with input vector $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^n$ and output $y \in \mathbb{R}$, $D_N = \big(X_N, \mathbf{y}_N\big) = \big\{\big(\mathbf{x}_i, y_i = f(\mathbf{x}_i)\big), i = 1, \ldots, N\big\}$ be a training set. Let's assume that $f(\mathbf{x})$ is a realization of a Gaussian process. Gaussian process, being a one of the possible ways to define a distribution on the functional space, is completely defined by its mean $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and covariance function $k(\mathbf{x}, \tilde{\mathbf{x}}) = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\tilde{\mathbf{x}}) - m(\tilde{\mathbf{x}}))]$. In applications the weighted exponential covariance function $k(\mathbf{x}, \tilde{\mathbf{x}}) = \sigma^2 \exp(-\sum_{i=1}^{n} \theta_i^2 (x^i - \tilde{x}^i)^2)$ is widely used.

Let us assume without loss of generality that $m(\mathbf{x}) \equiv 0$. If the covariance function $k(\mathbf{x}, \tilde{\mathbf{x}})$ is known, then the posterior process is also Gaussian:

$$\mathbb{P}_N(f) = \mathbb{P}(f|D_N) \sim GP\big(\hat{f}_N(\mathbf{x}), \hat{K}_N(\mathbf{x}, \tilde{\mathbf{x}})\big), \tag{1}$$

where the posterior mean of the Gaussian process $f(\mathbf{x})$ at some set of test points X^* can be written as follows [16]:

$$\hat{f}_N(X^*) = \hat{f}(X^*|D_N) = K_* K^{-1} \mathbf{y}_N, \tag{2}$$

where $K_* = \big[k(\mathbf{x}_i^*, \mathbf{x}_j), i = 1, \ldots, N_*; j = 1, \ldots, N\big]$, $K = \big[k(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, \ldots, N\big]$.

The posterior covariance matrix at test points is given by

$$\hat{K}_N(X^*) = \hat{K}(X^*|X_N) = K_{**} - K_* K^{-1} K_*^T, \tag{3}$$

where $K_{**} = \left[k(\mathbf{x}_i^*, \mathbf{x}_j^*), i, j = 1, \ldots, N_*\right]$. Then the posterior variance $\hat{\sigma}_N^2(X^*) = \hat{\sigma}^2(X^*|X_N) = \mathrm{diag}(\hat{K}_N(X^*))$.

The posterior mean at test points can be used for prediction of function values (approximation) and the posterior variance provides corresponding pointwise errors estimates. Let us note that here we described only basic idea of the Gaussian process regression. For more details about the model and parameters estimation algorithms please see [16–18]. Formulas (2) and (3) allow for simple iterative updates when new points are added to the training set.

Proposition 1. *Let for a Gaussian process $f(\mathbf{x})$ a mean and a covariance functions are equal to $\hat{f}_0(\mathbf{x}) \equiv 0$ and $\hat{K}_0(\mathbf{x}, \tilde{\mathbf{x}}) = k(\mathbf{x}, \tilde{\mathbf{x}})$ correspondingly. Then the following update equations can be used to calculate the mean and the covariance function of the posterior process for $i = 1, \ldots, N$:*

$$\hat{f}_i(\mathbf{x}) = \hat{f}_{i-1}(\mathbf{x}) + \frac{\hat{K}_{i-1}(\mathbf{x}, \mathbf{x}_i)}{\hat{K}_{i-1}(\mathbf{x}_i, \mathbf{x}_i)}\big(y_i - \hat{f}_{i-1}(\mathbf{x})\big),$$

$$\hat{K}_i(\mathbf{x}, \tilde{\mathbf{x}}) = \hat{K}_{i-1}(\mathbf{x}, \tilde{\mathbf{x}}) - \frac{\hat{K}_{i-1}(\mathbf{x}, \mathbf{x}_i)\hat{K}_{i-1}(\mathbf{x}_i, \tilde{\mathbf{x}})}{\hat{K}_{i-1}(\mathbf{x}_i, \mathbf{x}_i)}.$$

Moreover, due to the fact, that a Gaussian random measure is completely defined by the mean and the covariance functions, the corresponding iterative update is available for the posterior Gaussian measure (1).

Corollary 1. *There exists a functional ϕ such that $\mathbb{P}_l = \phi(\mathbf{x}_l, y_l, \mathbb{P}_{l-1}), l = 1, \ldots, N$.*

3 Adaptive Design of Experiments

The area of adaptive design of experiments has been paid much attention in recent decades and numerous criteria have been proposed. To the best of our knowledge, all the criteria are obtained from the common sense arguments, such as uniform filling of the design space, reduction of the error predicted by Gaussian process model, etc. However, it seems reasonable to develop criteria which are optimal in terms of some statistical functional. We are going to introduce Bayesian framework which allows to formally state the adaptive design of experiments problem and start with some definitions.

Definition 1. *Let $\rho = \rho(f, \hat{f})$ be some measurable functional. We will call it error function (see particular examples below).*

Definition 2. *Let $q_\rho(D_l) = \mathbb{E}\left[\rho(f, \hat{f}_l)|D_l\right]$, where expectation is taken as conditional given fixed sample D_l. Let's call $q_\rho(D_l)$ a mean posterior risk.*

Definition 3. *The optimal experimental plan X_N is a solution of the following optimization problem:*

$$J^* = \min_{X_N \in \mathbb{X}^N} \mathbb{E}_{\mathbf{y}_N}\left[\sum_{i=1}^{N} Q_\rho\big(\mathbf{x}_i, y_i, D_{i-1}\big)\right], \tag{4}$$

where $Q_\rho(\mathbf{x}_i, y_i, D_{i-1}) = q_\rho(D_{i-1} \cup (\mathbf{x}_i, y_i)) - q_\rho(D_{i-1})$.

Let us note, that (4) is equivalent to minimizing the error at the final iteration:

$$J^* = \min_{X_N \in \mathbb{X}^N} \mathbb{E}_{\mathbf{y}_N}\left[q_\rho(D_N)\right] - q_\rho(D_0),$$

where we set $q_\rho(D_0) = \mathbb{E}\rho(f, \hat{f}_0)$, $\hat{f}_0 = 0$.

Due to Corollary 1 the problem (4) is a stochastic dynamic programming problem and we can write the Bellman equations for it:

$$J_r(D_{r-1}) = \min_{\mathbf{x}_r \in \mathbb{X}} \mathbb{E}_{y_r}\left[Q_\rho(\mathbf{x}_r, y_r, D_{r-1}) + J_{r+1}(D_{r-1} \cup (\mathbf{x}_r, y_r))\right], r = 1, \ldots, N,$$

where $J_r(D_{r-1})$ is a value function at r-th step and $J_{N+1} \equiv 0$. Unfortunately, this stochastic dynamic programming problem is infinite dimensional and can not be solved explicitly. That is why in the next section we consider various one-step look-ahead solutions.

4 One-Step Look-Ahead Solutions

Value function for the last step of the stochastic dynamic programming problem (4) is given by the following formula:

$$J_N(D_{N-1}) = \min_{\mathbf{x}_N \in \mathbb{X}} \mathbb{E}_{y_N}\left[Q_\rho(\mathbf{x}_N, y_N, D_{N-1})\right].$$

Thus, the criterion for choosing an optimal one-step look-ahead point has the form:

$$\mathbf{x}_N = \arg\min_{\mathbf{x} \in \mathbb{X}} \mathcal{I}_\rho(\mathbf{x}), \tag{5}$$

where $\mathcal{I}_\rho(\mathbf{x}) = \mathbb{E}_{y_N}\left[Q_\rho(\mathbf{x}, y_N, D_{N-1})\right]$.

Depending on the choice of the error function $\rho(f, \hat{f})$ different types of criteria will follow. We consider the following types of the error function:

1. L_2-norm of difference between f and $\hat{f}_N$: $\rho_2(f, \hat{f}_N) = \frac{1}{|\mathbb{X}|}\int_{\mathbb{X}}(f(\mathbf{u}) - \hat{f}_N(\mathbf{u}))^2 \mathrm{d}\mathbf{u}$.
2. L_1-norm of difference between f and $\hat{f}_N$: $\rho_1(f, \hat{f}_N) = \frac{1}{|\mathbb{X}|}\int_{\mathbb{X}}|f(\mathbf{u}) - \hat{f}_N(\mathbf{u})|\mathrm{d}\mathbf{u}$.
3. L_∞-norm of difference between f and $\hat{f}_N$: $\rho_\infty(f, \hat{f}_N) = \max_{\mathbf{u} \in \mathbb{X}}|f(\mathbf{u}) - \hat{f}_N(\mathbf{u})|$.

4.1 L_2 Error Function

Proposition 2. *If the error function is* $\rho(f, \hat{f}_N) = \rho_2(f, \hat{f}_N)$*, then the optimal one-step look-ahead criterion has the form*

$$\mathcal{I}_{\rho_2}(\mathbf{x}) = \int_{\mathbb{X}}\left(\hat{\sigma}^2(\mathbf{u}|X_{N-1} \cup \mathbf{x}) - \hat{\sigma}^2(\mathbf{u}|X_{N-1})\right)\mathrm{d}\mathbf{u}.$$

We will denote this criterion as IntegratedMseGain. Due to the fact that $\hat{\sigma}^2(\mathbf{u}|X_{N-1})$ is independent of $\mathbf{x}_N$ this criterion coincides with the well known criterion IMSE [10]:

$$\mathcal{I}_{IMSE}(\mathbf{x}) = \frac{1}{|\mathbb{X}|}\int_{\mathbb{X}} \hat{\sigma}^2(\mathbf{u}|X_{N-1}\cup\mathbf{x})\mathrm{d}\mathbf{u}.$$

Proposition 3. *Change of the posterior variance at point* $\mathbf{v}$ *after point* $\mathbf{x}$ *is added to the sample, can be written as:*

$$\hat{\sigma}^2(\mathbf{v}|X_{N-1}\cup\mathbf{x}) - \hat{\sigma}^2(\mathbf{v}|X_{N-1}) = -\frac{\hat{K}^2_{N-1}(\mathbf{x},\mathbf{v})}{\hat{\sigma}^2(\mathbf{x}|X_{N-1})}.$$

Corollary 2. *The criterion* $\mathcal{I}_{\rho_2}(\mathbf{x})$ *can be rewritten as:*

$$\mathcal{I}_{\rho_2}(\mathbf{x}) = -\frac{1}{|\mathbb{X}|}\int_{\mathbb{X}} \frac{\hat{K}^2_{N-1}(\mathbf{x},\mathbf{v})}{\hat{\sigma}^2(\mathbf{x}|X_{N-1})}\mathrm{d}\mathbf{v}.$$

This corollary allows to compute $\mathcal{I}_{\rho_2}$ more effectively and increase its computational stability compared to IMSE, as no direct inversion of the covariance matrix for a full training set $(X_{N-1}\cup\mathbf{x})$ is needed.

In some important particular cases the criterion can be computed analytically. For example, for a widely used in applications weighted exponential covariance function we can state the following preposition.

Proposition 4. *Let* $k(\mathbf{x},\tilde{\mathbf{x}}) = \sigma^2\exp(-\sum_{i=1}^n \theta_i^2(x^i-\tilde{x}^i)^2)$, *and* $\mathbb{X} = \prod_{i=1}^n [a_i,b_i]$. *Define* $\mathbf{k}(\mathbf{x}) = \{k(\mathbf{x},\mathbf{x}_i)\}_{i=1}^{N-1}$. *Then*

$$\mathcal{I}_{\rho_2}(\mathbf{x}) = \frac{\mathrm{trace}\big(\mathbf{k}(\mathbf{x})\mathbf{k}^T(\mathbf{x})C\big) - 2\mathbf{k}^T(\mathbf{x})\mathbf{c}(\mathbf{x}) + c_{NN}(\mathbf{x})}{|\mathbb{X}|\hat{\sigma}^2(\mathbf{x}|X_{N-1})},$$

where $\mathbf{c}(\mathbf{x}) = \big[c_{iN}\big]_{i=1}^{N-1}, C = \big[c_{ij}\big]_{i,j=1}^{N-1}$, $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t$,

$$c_{ij} = \sigma^4\prod_{k=1}^n \frac{\sqrt{\pi}\exp(-\frac{1}{2}\theta_k^2(x_i^k - x_j^k)^2)}{\sqrt{8}\theta_k(b_k-a_k)}\Bigg[\mathrm{erf}\bigg(\sqrt{2}\theta_k\big(a_k - \frac{x_i^k+x_j^k}{2}\big)\bigg) -$$

$$-\mathrm{erf}\bigg(\sqrt{2}\theta_k\big(b_k - \frac{x_i^k+x_j^k}{2}\big)\bigg)\Bigg],$$

where for the ease of notations we denoted $\mathbf{x} = \mathbf{x}_N$.

4.2 L_1 Error Function

Proposition 5. *If the error function is* $\rho(f,\hat{f}_N) = \rho_1(f,\hat{f}_N)$, *then the optimal one-step look-ahead criterion has the form*

$$\mathcal{I}_{\rho_1}(\mathbf{x}) = \int_{\mathbb{X}}\sqrt{\frac{2}{\pi}}\big(\hat{\sigma}(\mathbf{u}|X_{N-1}\cup\mathbf{x}) - \hat{\sigma}(\mathbf{u}|X_{N-1})\big)\mathrm{d}\mathbf{u}.$$

To the best of our knowledge this criterion has no analogues in the literature, but its computation is complicated due to the structure of the formula inside the integral.

4.3 L_∞ Error Function

Proposition 6. *If the error function is* $\rho(f, \hat{f}_N) = \rho_\infty(f, \hat{f}_N)$*, then the optimal one-step look-ahead criterion has the form*

$$\mathcal{I}_{\rho_\infty}(\mathbf{x}) = \mathbb{E}\left[\max_{\mathbf{u}\in\mathbb{X}}\left|f(\mathbf{u}) - \hat{f}(\mathbf{u}|D_{N-1}\cup(\mathbf{x}, y_N))\right| - \max_{\mathbf{u}\in\mathbb{X}}\left|f(\mathbf{u}) - \hat{f}(\mathbf{u}|D_{N-1})\right| \middle| D_{N-1}\right].$$

Let us note that this criterion can't be computed analytically. In order to compute it one can use approximation techniques for the distribution of the maximum of the posterior Gaussian random process. In many common situations the maximum is proportional to the maximal value of the posterior variance [19], which justifies usage of the popular criterion Maximum Variance (MaxVar) [10]

$$\mathcal{I}_{MV}(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x}|X_{N-1}).$$

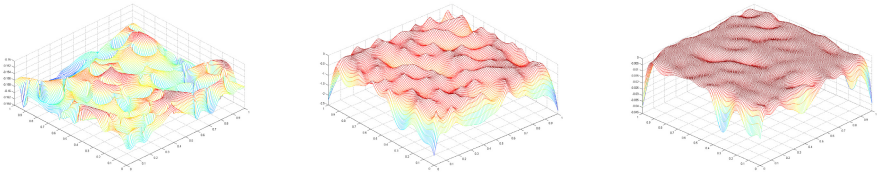
5 Towards More Robust Adaptive DoE Criterion

It is of the great importance that the optimization problem (5) is always multimodal and its complexity highly depends on the complexity of the criterion in question. Some popular criteria as IMSE, have very complicated surface with multiple narrow local optima, which makes their optimization a hard task. In this work we consider a simple multiplicative combination of criteria IntegratedMseGain and Maximum Variance (IntegratedMseGainMaxVar):

$$\mathcal{I}_{IGMV}(\mathbf{x}) = \frac{\hat{\sigma}^2(\mathbf{u}|X_{N-1})}{|\mathbb{X}|}\int_{\mathbb{X}}\left(\hat{\sigma}^2(\mathbf{u}|X_{N-1}\cup\mathbf{x}) - \hat{\sigma}^2(\mathbf{u}|X_{N-1})\right)\mathrm{d}\mathbf{u} =$$

$$= \frac{1}{|\mathbb{X}|}\int_{\mathbb{X}}\left[k(\mathbf{x},\mathbf{u}) - k(\mathbf{x},X)^T K^{-1} k(\mathbf{u},X)\right]^2 d\mathbf{u} = \frac{1}{|\mathbb{X}|}\int_{\mathbb{X}}\hat{K}^2_{N-1}(\mathbf{x},\mathbf{u})d\mathbf{u}.$$

This criterion takes into account global behaviour of the function as IntegratedMseGain does, but is more numerically stable then IntegratedMseGain and MaxVar (see behaviours of these criteria for 2D example in Figure 1).



IMSE **MaxVar** **IntegratedMseGain-MaxVar**

Fig. 1. Surfaces of IMSE, MaxVar and IntegratedMseGainMaxVar criteria

6 Experiments

6.1 Model Experiments

For the experiments we used a big set of artificial functions which are commonly used for testing optimization algorithms [20–22]. The idea behind such a choice is that approximation models are very often used as substitutes for real target functions during surrogate-based optimization process [13].

We performed testing on 10 different functions with 3 and 4 dimensional inputs. For every function we generated 10 independent initial DoEs with the size (10 * dimensionality) points. Adaptive DoE was performed starting from each initial DoE and for each criterion, described in Section 4, except criteria $\mathcal{I}_{\rho_1}$ and $\mathcal{I}_{\rho_\infty}$ as they are hard to compute even numerically. As a benchmark we used the method of random sampling and the non-adaptive method which

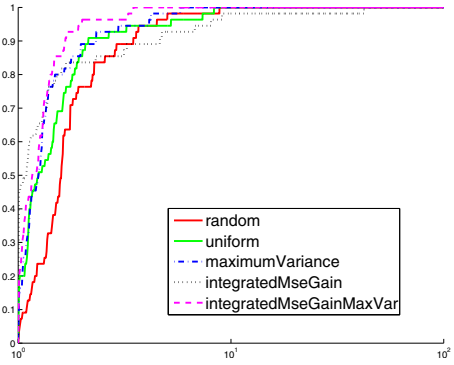


Fig. 2. Dolan-More curves for test functions, 30 added points

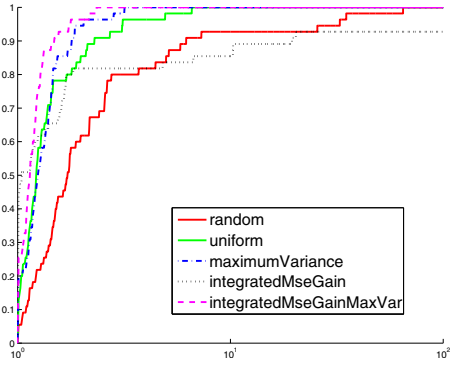


Fig. 3. Dolan-More curves for test functions, 50 added points

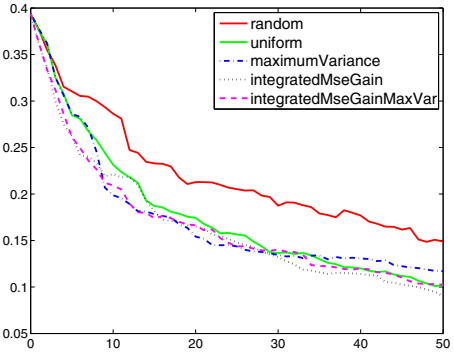


Fig. 4. Error curves for Michalewicz function

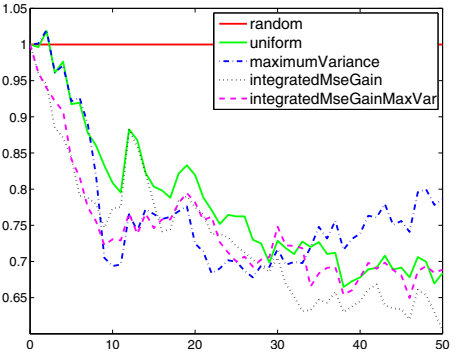


Fig. 5. Error curves for Michalewicz function (normed by the error curve of the random sampling)

Table 1. Input parameters for compression spring design problem

Input parameter	d	D	n	h_f
Lower bound	0.25 mm	1 cm	70	2.5 cm
Upper bound	5 mm	2 cm	100	100 cm

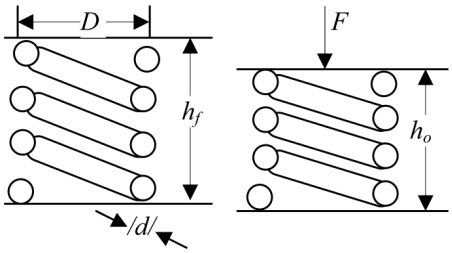


Fig. 6. Compression spring and its parameters

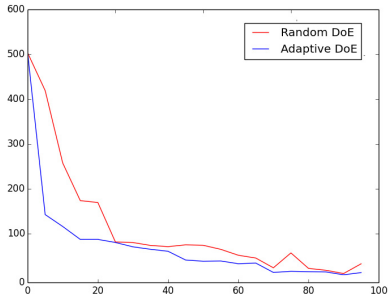


Fig. 7. RMS on the test set depending on the number of points added to DoE

generates new points maximizing minimum inter-point distance, i.e. provides a maximum to the following criterion: $\mathcal{I}_{uniform}(\mathbf{x}) = \min_{\mathbf{v} \in X_{N-1}} d(\mathbf{x}, \mathbf{v})$, where $d(\mathbf{x}, \mathbf{v})$ is a Euclidean distance between points $\mathbf{x}$ and $\mathbf{v}$.

On every iteration we add to the DoE one point optimizing the criterion in question and re-estimate parameters of Gaussian process regression using maximum likelihood approach. Results are compared by means of the root mean squared error on big test set of 10000 points. For convenience results are presented in the form of Dolan-More curves [23]. The higher curve is on the plot, the better is the quality obtained by the corresponding algorithm. Results are presented for 30 adaptively added points (see Figure 2) and for 50 adaptively added points (see Figure 3).

Let us note that IntegratedMseGain criterion provides good results in the most cases (the value of the Dolan-More curve at zero). However for many test functions this criterion also shows bad performance due to numerical instability (the corresponding Dolan-More curve reaches the maximal level with significant delay compared to the curves for other criteria). IntegratedMseGainMaxVar criterion is more robust and provides better results.

Also let's consider an example how the error of approximation changes from iteration to iteration of adaptive DoE process. We consider Michalewicz function in the domain $[0, 1]^2$. On Figures 4 and 5 we plot the dependence of the approximation error on a number of iteration of the adaptive DoE process (the results are averaged over 5 initial train samples). It can be easily seen, that adaptive methods are much better then the random sampling with the error of the best method 1.7 times smaller then the error, obtained in case the random sampling is used.

6.2 Spring Design

We consider a physical model of a compression spring [24]. We wish to construct a DoE which provides an accurate approximation of a force of the spring at its pre-load height, $h_0 = 2.5$ cm. The input design parameters are wire diameter d, coil diameter D, number of coils in the spring n and spring free height h_f, see Figure 6. We provide parameters bounds in Table 1. The force of a spring is computed by means of the physical model [24].

We generated two designs: a random design with 200 points and an adaptive design of the same size, generated using IntegratedMseGainMaxVar criterion (starting from an initial random design with 100 points). We computed a root mean squared error on a big test sample for each iteration of a DoE generation algorithm, see Figure 7. We conclude, that the adaptive design for this test case ensures decreasing trend for the approximation error and better quality than the random design.

7 Conclusions

In this work we propose an approach to adaptive design of experiments for a Gaussian process regression which allows to formulate it as stochastic dynamic programming problem. Optimal one-step look-ahead solutions of this problem allow to justify many criteria of adaptive design of experiments, well known from the literature. Also in this work we introduce new criterion IntegratedMseGainMaxVar, which in simple form combines well-known criteria, but achieves better numerical stability. Experimental results show the benefits of adaptive designs compared to the random sampling. The proposed criterion provides better results then the other criteria. As a direction of future research we plan to construct solutions of dynamic programming problem with horizon of few steps and obtain new, more accurate criteria of adaptive design of experiments.

Acknowledgments. The research was conducted in IITP RAS and supported solely by the Russian Science Foundation grant (project 14-50-00150).

References

1. Forrester A., Sobester A., Keane A.: Engineering Design via Surrogate Modelling. A Practical Guide, pp. 238. Wiley (2008)
2. Bernstein A.V., Burnaev E.V., Kuleshov A.P.: Intellectual data analysis in metamodelling. In: Proceedings of 17th Russian Seminar "Neuroinformatics and its Applications to Data Analysis", pp. 23–28, Krasnoyarsk (2009)
3. Giunta, A., Watson, L.T.: A comparison of approximation modeling technique: polynomial versus interpolating models. In: 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, pp. 392–404. AIAA, Reston (1998)
4. Batill, S.M., Renaud, J.E., Gu, X.: Modeling and simulation uncertainty in multidisciplinary design optimization. In: AIAA Paper, pp. 2000–4803, September 2000

5. Fedorov, V.V.: Theory of Optimal Experiments. Academic Press (1972)
6. Pukelsheim, F.: Optimal Design of Experiments. Wiley, New York (1993)
7. Fedorov, V.V.: Design of spatial experiments: model fitting and prediction. In: Handbook of Statistics, pp. 515–553. Elsevier, Amsterdam (1996)
8. Zimmerman, D.L.: Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics **17**(6), 635–652 (2006)
9. Chen, R.J.W., Sudjianto, A.: On sequential sampling for global metamodeling in engineering design. In: Proceedings of DETC 2002, Montreal, Canada, September 29-October 2 (2002)
10. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and Analysis of Computer Experiments. Statistical Science **4**(4), 409–423 (1989). doi:10.1214/ss/1177012413
11. Currin, C., Mitchell, T., Morris, M., Ylvisaker, D.: A Bayesian approach to the design and analysis of computer experiments. Journal of the American Statistical Association **86**(416), 953–963 (1991)
12. Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D.: Screening, predicting and computer experiments. Technometrics **34**, 15–25 (1992)
13. Jones, D.R., Schonlau, M., William, J.: Efficient global optimization of expensive black-box functions. J. Glob. Optim. **13**(4), 455–492 (1998)
14. Bect, J., Ginsbourger, D., Li, L., Picheny, V., Vazquez, E.: Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing **22**(3), 773–793 (2012)
15. Burnaev, E., Panin, I.: Adaptive design of experiments for sobol indices estimation based on quadratic metamodel. In: Proceedings of the Third International Symposium on Learning and Data Sciences (SLDS 2015), London, England, UK, April 20–22 (to appear, 2015)
16. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press (2006)
17. Burnaev, E., Zaytsev, A., Panov, M., Prikhodko, P., Yanovich, Yu.: Modeling of nonstationary covariance function of Gaussian process on base of expansion in dictionary of nonlinear functions. In ITaS-2011, Gelendzhik, October 2-7 (2011)
18. Belyaev, M., Burnaev, E., Kapushev, Y.: Gaussian process regression for large structured data sets. In: Proceedings of the Third International Symposium on Learning and Data Sciences (SLDS 2015), London, England, UK, April 20–22 (to appear, 2015)
19. Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Springer, 2007
20. Saltelli, A., Sobol, I.M.: About the use of rank transformation in sensitivity analysis of model output. Reliab. Eng. Syst. Safety **50**(3), 225–239 (1995)
21. Ishigami, T., Homma, T.: An importance qualification technique in uncertainty analysis for computer models. In: Proceedings of the Isuma 1990, First International Symposium on Uncertainty Modelling and Analysis. University of Maryland (1990)
22. Rönkkönen, J., Lampinen, J.: An extended mutation concept for the local selection based differential evolution algorithm. In: Proceedings of Genetic and Evolutionary Computation Conference, GECCO 2007, London, England, UK, July 7–11 (2007)
23. Dolan, E.D., More, J.J.: Benchmarking optimization software with performance profiles. In: Mathematical Programmming, Ser. A 91, pp. 201–213 (2002)
24. Fox, R.L.: Optimization methods for engineering design. Addison-Wesley, Massachusetts (1971)