

# Сравнительный анализ критериев адаптивного планирования экспериментов на основе гауссовских процессов

Евгений Бурнаев, Максим Панов

1. Институт Проблем Передачи Информации,  
127994, г. Москва, ГСП-4, Большой Каретный переулок, 19, стр.1

2. DATADVANCE,

105064, г. Москва, Садово-Черногрязская улица, 13/3

3. PreMoLab,

141700, г. Долгопрудный, Институтский переулок, 9

{evgeny.burnaev,maxim.panov}@datadvance.net

## Аннотация

В работе рассматривается задача адаптивного планирования эксперимента для задачи регрессии. В качестве регрессионной модели используется стохастическая модель гауссовского процесса. В работе рассмотрен как ряд классических критериев адаптивного планирования эксперимента, так и ряд новых критериев. Проведено сравнение рассматриваемых методов на большом количестве тестовых функций.

## 1. Введение

Одной из основных задач, которые приходится решать при построении метамоделей (моделей на основе данных) является задача аппроксимации неизвестной зависимости по данным [1, 2]. Наиболее популярная модель для построения аппроксиматоров, основанная на гауссовских процессах [3, 4, 5], используется в большом количестве разнообразных прикладных задач, включая концептуальное проектирование [6], структурную оптимизацию [7], многокритериальную оптимизацию при проектировании [8], конструирование в аэрокосмической [9] и автомобильной отраслях [10].

Во многих инженерных задачах количество вычислений целевой функции  $f(x)$  существенно ограничено по причине очень большого времени, требуемого для одного вычисления, и/или его высокой стоимости. В связи с этим крайне важной является задача построения методов выбора обучающей выборки  $D$  ограниченного объема  $N$  таким образом, чтобы максимизировать качество аппроксимации. Одним из наиболее популярных в литературе подходов является метод адаптивного планирования экспери-

мента [11]. Цель данной работы состоит в том, чтобы провести сравнительный анализ критериев адаптивного планирования экспериментов на основе гауссовских процессов.

Статья устроена следующим образом:

- в разделе 2 описывается постановка задачи аппроксимации;
- в разделе 3 рассказывается про построение аппроксиматора на основе стационарных гауссовских процессов;
- в разделе 4 описывается постановка задачи адаптивного планирования регрессионного эксперимента ;
- в разделе 5 описаны конкретные критерии адаптивного построения выборки точек;
- в разделах 7 и 8 рассказывается об экспериментальных результатах работы рассмотренных методов и делаются выводы об их особенностях и качестве работы.

## 2. Постановка задачи аппроксимации

В наиболее общем виде задача аппроксимации может быть сформулирована следующим образом. Пусть  $y = f(\mathbf{x})$  некоторая неизвестная функция со входом  $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^n$  и выходом  $y \in \mathbb{R}$ . Пусть  $D_{learn} = (X, \mathbf{y}) = \{(\mathbf{x}_i, y_i = f(\mathbf{x}_i)), i = 1, \dots, N\}$  - обучающая выборка. Задача состоит в построении аппроксимации  $\hat{y} = \hat{f}(\mathbf{x}) = \hat{f}(\mathbf{x}|D_{learn})$  для исходной зависимости  $y = f(\mathbf{x})$  по обучающей выборке  $D_{learn}$ .

Если для всех  $\mathbf{x} \in \mathbb{X}$  (не только для  $\mathbf{x} \in D_{learn}$ ) имеет место примерное равенство  $\hat{f}(\mathbf{x}) \approx f(\mathbf{x})$ , то считается, что аппроксиматор хорошо воспроизводит исходную зависимость. Это факт проверяется

на независимой тестовой выборке  $D_{test} = (X^*, \mathbf{y}^*) = \{(\mathbf{x}_j^*, y_j^* = f(\mathbf{x}_j^*)), j = 1, \dots, N_*\}$ . Мерой качества аппроксимации является среднеквадратичная ошибка на тестовой выборке:

$$Q(\hat{f}|D_{test}) = \sqrt{\frac{1}{N_*} \sum_{j=1}^{N_*} (y_j^* - \hat{f}(\mathbf{x}_j^*))^2}. \quad (1)$$

Заметим, что во многих практических задачах обучающая выборка не задана априори, причем точки обучающей выборки  $X$  могут быть выбраны произвольным образом из множества  $\mathbb{X}$ , а соответствующие значения функции  $\mathbf{y} = f(X)$  могут быть подсчитаны с помощью алгоритма вычисления  $f(\mathbf{x}), \mathbf{x} \in \mathbb{X}$ . Выбор оптимального в некотором смысле подмножества точек  $X \subset \mathbb{X}$  называется задачей планирования эксперимента. Одним из возможных критериев оптимальности может быть минимизация ошибки на тестовом множестве  $Q(\hat{f}|D_{test})$ . В последнее время большое развитие получило так называемое адаптивное планирование эксперимента [2], при котором обучающая выборка строится итеративно, причем на каждой новой итерации для выбора новых точек используется аппроксимация, построенная на предыдущей итерации (см. раздел 4). В следующем разделе рассматривается построение аппроксиматора на основе гауссовских процессов [12, 13], который в дальнейшем используется для адаптивного планирования эксперимента.

### 3. Аппроксиматор на основе гауссовского процесса

#### 3.1. Гауссовские процессы

Гауссовский процесс является одним из возможных способов задания распределения на пространстве функций. Гауссовский процесс  $f(\mathbf{x})$  полностью определяется своей функцией среднего  $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  и ковариационной функцией  $cov(y, y') = k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$ . Если положить функцию среднего нулевой:  $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] = 0$ , а ковариационную функцию считать известной, то функция апостериорного (для заданной обучающей выборки) среднего значения гауссовского процесса в точках контрольной выборки  $X^*$  выглядит следующим образом [12]:

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

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

В типичных, более реалистичных ситуациях при моделировании мы не имеем доступа непосредственно к значениям функции, а наблюдаем их только в зашумленном виде:

$$\mathbf{y}(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x}), \quad (3)$$

где шум  $\boldsymbol{\varepsilon}(\mathbf{x})$  моделируется независимыми одинаково распределенными нормальными случайными величинами с нулевым средним и дисперсией  $\tilde{\sigma}^2$ . В таком случае наблюдения  $\mathbf{y}(\mathbf{x})$  будут гауссовским процессом с нулевым средним и ковариационной функцией  $cov(y(\mathbf{x}), y(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') + \tilde{\sigma}^2$ . Таким образом, функция апостериорного (для заданной обучающей выборки) среднего значения гауссовского процесса  $f(\mathbf{x})$  в точках контрольной выборки  $X_*$  принимает вид:

$$\hat{f}(X^*) = K_*(K + \tilde{\sigma}^2 I)^{-1} \mathbf{y}, \quad (4)$$

где  $I$  - единичная матрица размера  $N \times N$ .

Заметим, что наличие в формуле (4) дисперсии шума  $\tilde{\sigma}^2$  фактически приводит к регуляризации, что позволяет улучшить обобщающую способность аппроксиматора. При этом апостериорная ковариационная функция гауссовского процесса в точках контрольной выборки имеет вид:

$$\mathbb{V}[\hat{f}(X^*)] = K(X^*, X^*) + \tilde{\sigma}^2 I_* - K_*(K + \tilde{\sigma}^2 I)^{-1} K_*^T \quad (5)$$

где  $K(X^*, X^*) = [k(\mathbf{x}_i^*, \mathbf{x}_j^*), i, j = 1, \dots, N_*]$ ,  $I_*$  - единичная матрица размера  $N_* \times N_*$ .

Дисперсии гауссовского процесса в точках контрольной выборки могут быть использованы как оценки ожидаемой ошибки аппроксимации в этих точках. Заметим, что для этого нет необходимости вычислять по формуле (5) всю матрицу  $\mathbb{V}[\hat{f}(X^*)]$ , а достаточно вычислить только элементы ее главной диагонали, которые и являются искомыми дисперсиями. Обозначим их  $\hat{\sigma}^2(X^*) = \hat{\sigma}^2(X^*|X) = \text{diag}(\mathbb{V}[\hat{f}(X^*)])$ . Заметим, что величины  $\hat{\sigma}^2(X^*)$  не зависят от значений аппроксимируемой функции  $\mathbf{y}$  в точках обучающей выборки  $X$ , а зависят только от самих точек  $X$ .

При работе с реальными данными ковариационная функция породившего их гауссовского процесса, как правило, не известна, поэтому необходимо уметь идентифицировать ее по данным.

#### 3.2. Нахождение параметров гауссовского процесса

Предположим, что ковариационная функция гауссовского процесса является членом некоторого параметрического семейства  $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}'|a)$ , где  $a \in \mathbb{R}^K$  - вектор параметров ковариационной функции. Семейство  $k(\mathbf{x}, \mathbf{x}'|a)$  обычно берется из класса так называемых стационарных ковариационных функций, т.е. функций значение которых зависит только от разности значений аргументов  $k(\mathbf{x}, \mathbf{x}'|a) = k(\mathbf{x} - \mathbf{x}'|a)$ . Значение параметра  $a$  предлагается восстанавливать по обучающей выборке  $D_{learn}$ , исходя из принципа максимума правдоподобия. Для этого выпишем логарифм правдоподобия гауссовского

процесса в точках обучающей выборки [12]:

$$\log p(\mathbf{y}|X, a, \tilde{\sigma}) = -\frac{1}{2}\mathbf{y}^T(K + \tilde{\sigma}^2 I_N)^{-1}\mathbf{y} - \frac{1}{2}\log|K + \tilde{\sigma}^2 I| - \frac{n}{2}\log 2\pi, \quad (6)$$

где  $|K + \tilde{\sigma}^2 I|$  - детерминант матрицы  $K + \tilde{\sigma}^2 I$ .

Кроме параметров  $a$  ковариационной функции параметром функционала (6) является также значение дисперсии шума наблюдений  $\tilde{\sigma}^2$ , которое также можно настраивать по обучающей выборке. Таким образом, нахождение оптимальных значений параметров сводится к отысканию максимума правдоподобия по параметрам:

$$\log p(\mathbf{y}|X, a, \tilde{\sigma}) \rightarrow \max_{a, \tilde{\sigma}} \quad (7)$$

Выбор конкретного семейства ковариационных функций  $k(\mathbf{x} - \mathbf{x}'|a)$  обычно продиктован соображениями удобства, а также априорными представлениями о свойствах аппроксимируемой зависимости. В данной работе мы используем ковариационные функции вида  $k(\mathbf{x} - \mathbf{x}'|a) = \tilde{\sigma}^2 \exp(-\sum_{i=1}^n \theta_i^2 |x_i - x'_i|^2)$ , где параметры  $a = \{\theta_i, i = 1, \dots, n; \tilde{\sigma}\}$  настраиваются по обучающей выборке при решении задачи (7).

Заметим, что в данном разделе описан только базовый алгоритм регрессии на основе гауссовских процессов, более детальное описание см. в [14].

#### 4. Адаптивное планирование регрессионного эксперимента

Множество различных методов и подходов к адаптивному планированию регрессионного эксперимента достаточно обширно, но при этом разумно выделить два основных направления, а именно [15]:

- Планирование эксперимента с целью наилучшей оценки параметров целевой функции  $f(\mathbf{x})$ ;
- Планирование эксперимента с целью наилучшего предсказания значений целевой функции  $f(\mathbf{x})$ .

Заметим, что по первому подходу существует большой объем литературы, данная теория широко развита в частности в работах Фёдорова [21]. По второму подходу в литературе предложено большое количество различных эвристических методов [11], но работы, в которых проведено их качественное сравнение, практически отсутствуют. В данной работе мы постараемся заполнить эту нишу и сосредоточимся на втором подходе. Рассмотрим алгоритм адаптивного планирования эксперимента, который в общем виде описан ниже:

1. Строим начальную выборку  $D$  с помощью одного из методов планирования эксперимента, которые равномерно заполняют точками пространство дизайна. В данной работе используется метод латинских гиперкубов.
2. Строим аппроксимацию  $\hat{f}(\mathbf{x}|D)$  на основе выборки  $D$ .
3. Если мы добились требуемого качества аппроксимации или исчерпали количество допустимых вычислений целевой функции, то останавливаемся, иначе переходим к шагу 4.
4. Получаем новые точки, используя  $D$ ,  $\hat{f}(\mathbf{x}|D)$  и  $\hat{\sigma}^2(\mathbf{x}|X)$ :

$$X_{new} = \arg \max_{\mathbf{x} \in X} g(D, \hat{f}(\mathbf{x}|D), \hat{\sigma}^2(\mathbf{x}|X)),$$

где  $g$  - некоторый критерий адаптивного планирования эксперимента,  $X_{new}$  - множество точек, которые добавляются в обучающую выборку на текущем шаге.

5.  $D = D \cup D_{new}$ , где  $D_{new} = (X_{new}, Y_{new}) = \{(\mathbf{x} \in X_{new}, y = f(\mathbf{x}))\}$ .
6. Возвращаемся к шагу 2.

Заметим, что в данном алгоритме некоторые пункты требуют дополнительного уточнения. В частности, отдельного внимания заслуживает алгоритм оптимизации критериев. Его описание см. в разделе 6. Возможные критерии адаптивного планирования эксперимента рассматриваются в следующем разделе.

#### 5. Критерии адаптивного добавления точек

В работе рассматривается несколько критериев адаптивного планирования эксперимента. Большинство из этих критериев используют тот факт, что метод построения регрессии на основе гауссовских процессов позволяет оценивать ожидаемую ошибку предсказания  $\hat{\sigma}(\mathbf{x}|X)$ .

##### 5.1. Maximum variance

Одним из наиболее простых методов является выбор точки с максимальной ожидаемой ошибкой предсказания [11]:

$$\max Var(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x}|X).$$

Отличительной особенностью данного метода является склонность к выбору точек на границе области  $X$ . Такое свойство метода может быть нежелательным в случае, если поведение целевой функции  $f(\mathbf{x})$  на границе области является нестабильным, что часто встречается в реальных задачах.

## 5.2. Integrated mean squared error (IMSE)

Данный критерий представляет собой среднюю ошибку по области эксперимента  $\mathbb{X}$  в предположении, что точка  $\mathbf{x}$  добавлена в обучающую выборку [11]:

$$imse(\mathbf{x}) = -\frac{1}{|\mathbb{X}|} \int_{\mathbf{u} \in \mathbb{X}} \hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x}) d\mathbf{u},$$

где  $\hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x})$  - дисперсия предсказания в точке  $\mathbf{u}$  в предположении, что точка  $\mathbf{x}$  добавлена в обучающую выборку, а  $|\mathbb{X}| = \int_{\mathbf{u} \in \mathbb{X}} d\mathbf{u}$  - объем множества  $\mathbb{X}$ .

Заметим, что в силу того, что апостериорная дисперсия  $\hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x})$  не зависит от значений целевой функции в точках выборки  $X \cup \mathbf{x}$ , то она может быть вычислена до того, как наблюдается  $f(\mathbf{x})$ .

## 5.3. Integrated mean squared error gain (imseGain)

В данном критерии в отличие от предыдущего усредняется не ожидаемая ошибка, а ее приращение после добавления точки  $\mathbf{x}$  [16]:

$$imseGain(\mathbf{x}) = \frac{1}{|\mathbb{X}|} \int_{\mathbf{u} \in \mathbb{X}} (\hat{\sigma}^2(\mathbf{u}|X) - \hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x})) d\mathbf{u}.$$

Заметим, что данный критерий фактически отличается от предыдущего критерия только на константное слагаемое. Однако, использование формул обращения блочных матриц позволяет избежать прямого обращения матрицы ковариаций для новой выборки  $X \cup \mathbf{x}$ , поскольку [16]:

$$\hat{\sigma}^2(\mathbf{u}|X) - \hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x}) = \frac{[k(\mathbf{x}, \mathbf{u}) - k(\mathbf{x}, X)^T K^{-1} k(\mathbf{u}, X)]^2}{\hat{\sigma}^2(\mathbf{x}|X)},$$

что значительно упрощает вычисления.

## 5.4. imseGain-MaximumVariance

Несмотря на разумные предположения, которые лежат в основе двух предыдущих критериев, их отличительной чертой является численная нестабильность в том случае, если новая точка помещается близко к другим точкам обучающей выборки. В свою очередь критерий Maximum variance такими проблемами не обладает, а наоборот в точках обучающей выборки имеет локальные минимумы. Таким образом, если совместить критерии Maximum variance и Integrated mean squared error gain в одном критерии с помощью мультипликативной формулы, то можно ожидать, что в полученный критерий будет совмещать наилучшие качества этих критериев, а недостатки скомпенсируются. Получаемый критерий имеет следующий вид:

$$imseGainMV(\mathbf{x}) = \frac{\hat{\sigma}^2(\mathbf{x}|X)}{|\mathbb{X}|} \int_{\mathbf{u} \in \mathbb{X}} (\hat{\sigma}^2(\mathbf{u}|X) - \hat{\sigma}^2(\mathbf{u}|X \cup \mathbf{x})) d\mathbf{u} =$$

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

## 5.5. Uniform

Также в работе рассматривается простой критерий, который основан на том, что разумно располагать точки так, чтобы они равномерно заполняли пространство. Одним из вариантов такого критерия является критерий, максимизирующий минимальное расстояние между точками выборки:

$$uniform(\mathbf{x}) = \min_{\mathbf{u} \neq \mathbf{v} \in X \cup \mathbf{x}} d(\mathbf{u}, \mathbf{v}),$$

где  $d(\mathbf{u}, \mathbf{v})$  - расстояние между точками  $\mathbf{u}$  и  $\mathbf{v}$  в некоторой метрике. В данной работе рассматривается обычная евклидова метрика  $d(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$ .

Заметим, что данный критерий фактически не зависит от значений аппроксимируемой функции  $f(\mathbf{x})$  и от ее аппроксимации  $\hat{f}(\mathbf{x})$ .

## 6. Оптимизация критериев

Критерии адаптивного планирования эксперимента по своей сути являются очень многоэкстремальными. У большинства критериев локальные минимумы расположены в точках текущей выборки, а практически в любой области пространства между точками обучающей выборки имеется локальный максимум. Таким образом качественная оптимизация критериев адаптивного планирования эксперимента является тяжелой задачей, причем трудность задачи возрастает с размерностью пространства, т.к. количество локальных особенностей растет экспоненциально с размерностью пространства входных значений. Для решения этой задачи представляются разумным использовать алгоритмы глобальной оптимизации. В данной работе для оптимизации критериев используется алгоритм имитации отжига [17], который обладает хорошими глобальными свойствами и обеспечивает качественную оптимизацию критериев.

## 7. Экспериментальные результаты

В ходе экспериментов был использован большой набор тестовых функций, которые применяются для тестирования задач оптимизации [18, 19]. Всего тестирование проводилось на 10 различных функциях размерностей 3 и 4, для каждой из которых генерировалось 5 случайных начальных выборок размером (10 \* размерность) точек. Процедура адаптивного планирования эксперимента проводилась для каждого из рассмотренных в разделе 5 критериев и для каждой из 5 начальных выборок. Причем, на каждой

итерации в обучающую выборку добавлялась одна точка, выбранная в соответствии с рассматриваемым критерием. Всего добавлялось 50 точек. Результаты сравнивались по среднеквадратичной ошибке (1) на больших контрольных выборках из 10000 точек. Также для сравнения с критериями приведены результаты для случая, когда в выборку добавляются случайные точки из области дизайна. Для удобства результаты представлены в виде кривых Долан-Мора [20]. Чем выше кривая находится на графике, тем выше качество работы соответствующего алгоритма. Результаты представлены для аппроксимаций, построенных по выборке с 30 адаптивно добавленными точками (см. рисунок 1) и по выборке с 50 адаптивно добавленными точками (см. рисунок 2). Заметим, что критерии IMSE и imseGain на значительной доле функций дают наилучший результат (значение кривой Долан-Мора в нуле), но при этом также на многих функциях очень большой проигрыш, что вызвано численной нестабильностью этих критериев. Предложенный в данной работе критерий imseGain-MaximumVariance позволяет избежать этого эффекта и показывает результат лучше, чем все остальные критерии.

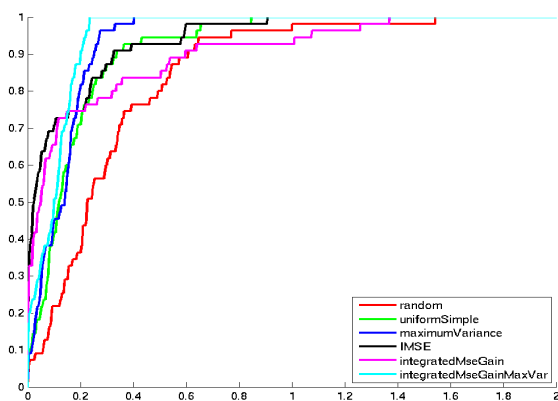


Рис. 1. Кривые Долан-Мора для тестовых функций, 30 добавленных точек.

Также рассмотрим пример, на котором видно, как изменяется ошибка аппроксимации по мере добавления точек в выборку. Для этого рассмотрим аппроксимацию функции Michalewicz в области  $[0, 1]^2$ . На рисунках 3 и 4 изображена зависимость ошибки аппроксимации от итерации алгоритма, причем результаты усреднены по 5 запускам для 5 различных начальных выборок. Хорошо видно, что адаптивные методы заметно выигрывают у случайного добавления точек, причем ошибка лучшего из адаптивных методов почти в 2 раза меньше, чем у метода случайного добавления точек.

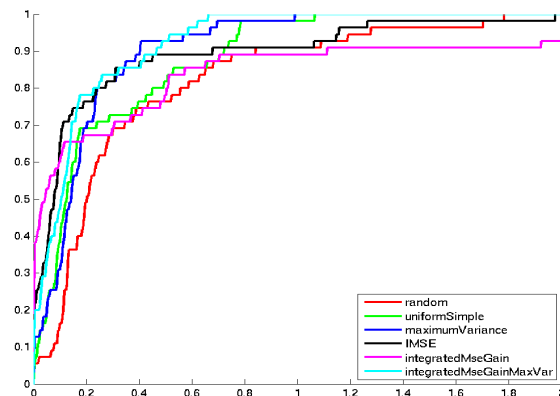


Рис. 2. Кривые Долан-Мора для тестовых функций, 50 добавленных точек.

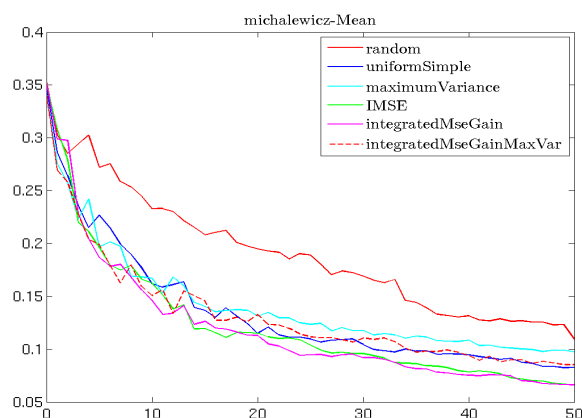
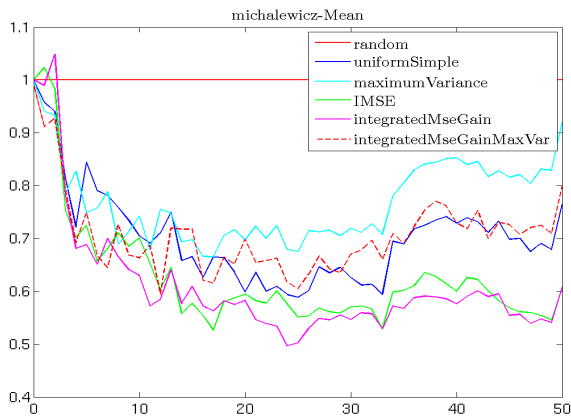


Рис. 3. Кривые ошибок для функции Michalewicz.

## 8. Выводы

В работе рассмотрен ряд критериев адаптивного планирования эксперимента с целью наилучшего предсказания значений функции. Также предложен новый критерий imseGain-MaximumVariance, который в простой форме комбинирует уже известные критерии, что позволяет добиться более хороших свойств и численного поведения критерия. Экспериментальные результаты показали значительное преимущество адаптивных методов над случайным добавлением точек в обучающую выборку. При этом предложенный критерий показал более хорошие результаты по сравнению с известными критериями, а также оказался более численно стабильным.

Заметим, что рассмотренный в данной работе подход является полностью эвристическим, хотя и исходит из разумных предположений о свойствах рассматриваемой задачи. Поэтому дальнейшие исследования предполагается сосредоточить на



**Рис. 4. Кривые ошибок для функции Michalewicz, нормированные на ошибку метода случайного до- бавления точек.**

построении хорошей теоретической постановки задачи адаптивного планирования эксперимента с целью наилучшего предсказания значений целевой функции. Такая постановка позволит получить новые критерии адаптивного планирования эксперимента и исследовать их теоретические свойства.

Работа выполнена при поддержке Лаборатории структурных методов анализа данных в предсказательном моделировании, МФТИ, грант правительства РФ дог. 11.G34.31.0073.

## Список литературы

[1] Бернштейн А.В., Бурнаев Е.В., Кулешов А.П. Интеллектуальный анализ данных в метаоптимизации. // Труды 17 Всероссийского Семинара "Нейроинформатика и ее приложения к Анализу Данных Красноярск, 2009 – с. 23-28.

[2] Forrester A., Sobester A., Keane A. Engineering Design via Surrogate Modelling. A Practical Guide. – Wiley, 2008. – 238 p.

[3] Giunta A., Watson L.T. A Comparison of Approximation Modeling Technique: Polynomial Versus Interpolating Models. // 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Vol. 1, AIAA, Reston, VA, 1998 – pp. 392–404.

[4] Simpson T.W., Booker A.J., Ghosh S., Giunta A., Koch P.N., Yang, R.J. Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion. // Structural and Multidisciplinary Optimization, Vol. 27, No. 5, 2004 – pp. 302–313.

[5] Batill S.M., Renaud J.E., Gu X., Modeling and Simulation Uncertainty in Multidisciplinary Design Optimization // AIAA Paper 2000-4803, Sept. 2000.

[6] Pacheco J. E., Amon C. H., Finger S. Bayesian Surrogates Applied to Conceptual Stages of the Engineering Design Process // Journal of Mechanical

Design, Vol. 125, No. 4, 2003 – pp. 664–672.

[7] Booker A.J., Dennis J.E., Frank P.D., Serafini D.B., Torczon, V., Trosset, M. A Rigorous Framework for Optimization of Expensive Functions by Surrogates // Structural Optimization, Vol. 17, No. 1, 1999, – pp. 1–13.

[8] Koch P.N., Wujek B.A., Golovidov O., Simpson, T.W. Facilitating Probabilistic Multidisciplinary Design Optimization Using Kriging Approximation Models // AIAA Paper 2002-5415, Sept. 2002.

[9] Simpson T.W., Maurey T.M., Korte J.J., and Mistree F. Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization // AIAA Journal, Vol. 39, No. 12, 2001 – pp. 2233–2241.

[10] Yang R. J., Wang N., Tho C.H., Bobineau J.P., and Wang B.P. Metamodeling Development for Vehicle Frontal Impact Simulation // American Society of Mechanical Engineers, ASME Design Engineering Technical Conf.—Design Automation Conf., DETC2001/DAC-21012, Sept. 2001.

[11] Chen R.J.W., Sudjianto A. On sequential sampling for global metamodeling in engineering design // Proceedings of DETC'02: ASME 2002 Design Engineering Technical Conferences And Computers and Information in Engineering Conference Montreal, Canada, September 29-October 2, 2002.

[12] Rasmussen C.E., Williams C.K.I. Gaussian Processes for Machine Learning. – the MIT Press, 2006.

[13] Cressie N.A.C. Statistics for Spatial Data. – Wiley, 1993.

[14] Burnaev E., Zaytsev A., Panov M., Prikhodko P. and Yanovich Yu. Modeling of nonstationary covariance function of Gaussian process on base of expansion in dictionary of nonlinear functions // In Information Technologies and Systems–2011, Gelendzhik, October 2–7 2011.

[15] Zimmerman D.L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction // Environmetrics, Volume 17, Issue 6, 2006 – pages 635–652.

[16] Gramacy R.B., Lee H.K.H. Adaptive design and analysis of supercomputer experiments // arXiv:0805.4359v4

[17] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P. Optimization by Simulated Annealing // Science, New Series, Vol. 220, No. 4598, 1983 – pp. 671–680.

[18] GDR MASCOT-NUM Toy Functions benchmark. <http://gdr-mascotnum.math.cnrs.fr/data2/benchmarks/jm.pdf>

[19] Lappeenranta University of Technology: evolutionary computation pages - the function testbed. <http://www.it.lut.fi/ip/evo/functions/functions.html>

[20] Dolan E.D., Moré J.J. Benchmarking optimization software with performance profiles // Mathematical Programming, Ser. A 91, 2002 – pp. 201–213

[21] Fedorov V.V. Theory of optimal experiments. – Academic press, 1972.