

НОВЫЙ ПОДХОД К НАСТРОЙКЕ ГАУССОВОЙ СУРРОГАТНОЙ МОДЕЛИ ЦЕЛЕВОЙ ФУНКЦИИ В ЗАДАЧЕ ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ ПРОЕКТНЫХ РЕШЕНИЙ

Т.А. Агасиев¹

agasiev@bmstu.ru

Н.П. Гвоздев²

gvozdevnp@gmail.com

А.П. Карпенко¹

apkarpenko@bmstu.ru

Н.В. Пивоварова¹

nv pivovarova@bmstu.ru

¹ МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

² ООО «ДАТАДВАНС», Москва, Российская Федерация

Аннотация

Рассмотрены методы решения задачи параметрической оптимизации проектных решений на основе построения гауссовой суррогатной модели целевой функции. Приведена задача поиска оптимальных значений свободных параметров (гиперпараметров) суррогатной модели, называемая задачей настройки. Задача настройки надстраивается над задачей синтеза суррогатной модели и имеет высокую вычислительную сложность. Предложен подход к настройке суррогатной модели, который может сделать процедуру настройки приемлемой по вычислительным затратам. Подход включает в себя этапы настройки и эксплуатации. Этап настройки содержит следующие шаги: формирование набора тестовых целевых функций, генерация совокупности обучающих выборок для каждой функции, определение для сгенерированных выборок значений их характерных признаков, определение оптимальных значений гиперпараметров для рассматриваемых тестовых функций и обучающих выборок, формирование совокупности пар вектор значений характерных признаков выборки–оптимальные значения гиперпараметров, построение на этой основе прогнозирующей модели, предсказывающей оптимальные значения гиперпараметров по характерным признакам обучающей выборки. Для исходной задачи на этапе эксплуатации сгенерирована обучающая выборка и опреде-

Ключевые слова

Параметрическая оптимизация, суррогатное моделирование, байесовский подход к оптимизации, гиперпараметры

лены ее характерные признаки, спрогнозированы оптимальные значения гиперпараметров суррогатной модели. На основе указанной обучающей выборки синтезирована суррогатная модель целевой функции. С помощью суррогатной модели решена исходная задача оптимизации, в качестве оптимальных значений гиперпараметров которой применены прогнозные значения. Подход может обеспечить повышение до 30 % эффективности базового алгоритма оптимизации

Поступила 17.03.2023

Принята 08.06.2023

© Автор(ы), 2023

Введение. Одной из задач, решаемых в процессе автоматизированного проектирования изделий, является задача определения наилучших значений его внутренних параметров с использованием некоторой математической модели изделия. Такую задачу называют задачей параметрической оптимизации проектных решений.

В общей постановке задача параметрической оптимизации состоит в поиске экстремума целевой функции с заданными ограничениями на варьируемые параметры [1]. Особенностью целевых функций в задачах автоматизированного проектирования является их высокая вычислительная сложность, поскольку в этих задачах, как правило, используют математические модели проектируемого изделия в форме систем обыкновенных дифференциальных уравнений или уравнений в частных производных. Для решений этих уравнений приходится применять численные методы. В результате даже при использовании современных высокопроизводительных вычислительных систем время вычисления одного значения целевой функции может составлять несколько дней [2, 3].

Для уменьшения вычислительной сложности задачи параметрической оптимизации проектных решений на этапе предварительного проектирования используют такие аппроксимирующие (суррогатные) модели целевой функции [3, 4], как модели линейной регрессии [5], радиальные базисные функции [6], нейронные сети [7], случайные процессы [8]. В последнем случае могут быть использованы суррогатные модели на основе гауссовых случайных, винеровских, пуассоновых процессов и процессов с ортогональными и независимыми приращениями.

В качестве суррогатных моделей целевой функции рассматриваем модели на основе гауссовых случайных процессов, т. е. процессов, в которых любой конечный набор их значений имеет многомерное нормальное распределение. Использование модели обусловлено следующими ее преимуществами [9]:

- гауссов случайный процесс исчерпывающим образом определяют всего две моментные функции: математическое ожидание и симметричная неотрицательно определенная действительная ковариационная матрица;

- для гауссовых случайных процессов понятия стационарности в широком и узком смыслах совпадают;

- на основе гауссовых процессов можно предсказать как приближенное значение целевой функции, так и степень достоверности этого предсказания, что повышает точность выявления подобластей пространства поиска, имеющих наибольшую вероятность локализации экстремума целевой функции;

- гауссовы процессы особенно эффективны при небольших размерах обучающих выборок, которые приходится использовать в случае вычислительно сложных целевых функций.

Задача оптимизации (метаоптимизации) суррогатной модели целевой функции состоит в поиске оптимальных значений свободных параметров (гиперпараметров) этой модели. Назовем эту задачу «Настройка суррогатной модели» или «Настройка гиперпараметров». В результате настройки могут быть определены, очевидно лишь приближенно, оптимальные значения гиперпараметров. Для простоты записи этот нюанс далее опускаем и используем термин «оптимальные значения гиперпараметров».

Задача настройки гиперпараметров надстраивается над задачей синтеза суррогатной модели целевой функции. Используем скалярный критерий оптимальности настройки (хотя в более общей постановке следует рассматривать векторный критерий). Градиент критерия оптимальности настройки, как правило, аналитически вычислить не удастся. Потому задачу настройки гиперпараметров приходится решать прямыми численными методами, основными из которых являются следующие: поиск на сетке; случайный поиск; байесовская оптимизация; применение разного рода эвристик [10, 11].

Поскольку задачу синтеза оптимальной суррогатной модели приходится решать на каждом шаге алгоритма оптимизации целевой функции, вычислительная сложность задачи настройки гиперпараметров оказывается очень высокой и возрастает по мере накопления данных о целевой функции в процессе решения исходной задачи оптимизации. В результате процедура настройки может оказаться нецелесообразной (эффект от настройки не покрывает вычислительные затраты на нее). Предлагается подход к настройке гиперпараметров, который можно отнести к классу однократных методов метаоптимизации [12] и который может сделать процедуру настройки приемлемой по вычислительным затратам.

Предлагаемый подход включает в себя два этапа, идею которых передают их следующие основные шаги. На первом этапе настройки:

- формируем представительный набор тестовых целевых функций;
- для каждой указанной функции генерируем совокупность обучающих выборок, каждая выборка представляет собой набор пар вектор значений варьируемых параметров исходной задачи оптимизации–соответствующее значение целевой функции;
- для каждой сгенерированной выборки определяем значения характерных признаков [13];
- для всех рассматриваемых тестовых целевых функций и всех обучающих выборок реализуем настройку гиперпараметров суррогатных моделей этих функций — определяем оптимальные значения гиперпараметров;
- формируем совокупность пар вектор значений характерных признаков выборки–оптимальные значения гиперпараметров и на этой основе строим прогнозирующую модель, предсказывающую оптимальные значения гиперпараметров по характерным признакам обучающей выборки.

На втором этапе эксплуатации (решение исходной задачи параметрической оптимизации):

- генерируем для исходной задачи обучающую выборку и определяем ее характерные признаки;
- прогнозируем оптимальные значения гиперпараметров суррогатной модели с помощью указанной прогнозирующей модели;
- на основе указанной обучающей выборки синтезируем суррогатную модель целевой функции;
- решаем исходную задачу оптимизации, используя суррогатную модель, в качестве оптимальных значений гиперпараметров которой применяем их прогнозные значения.

Цель настоящей работы — реализация и исследование эффективности предлагаемого подхода.

Постановка задачи и основные обозначения. *Базовая задача оптимизации.* Рассмотрим базовую (исходную) задачу глобальной параметрической оптимизации в следующей постановке:

$$\min_{X \in D_X \subset \mathfrak{R}^{|X|}} f(X) = f(X^*) = f^*.$$

Здесь $X = (x_1, x_2, \dots, x_{|X|})$ — вектор варьируемых параметров размерности $|X|$; $f(X) \in \mathfrak{R}^1$ — целевая функция; D_X — односвязное множество допустимых значений вектора варьируемых параметров.

Задача суррогатного моделирования. Пусть $L = \{(X_i, f_i = f(X_i))\}; i \in [1 : |L|]\}$ — обучающая выборка, где $X_i \in D_X$; $|L|$ — объем выборки.

Задачу синтеза суррогатной модели $\phi(X)$ целевой функции $f(X)$ ставим следующим образом: на основе выборки L определить функцию $\phi(X)$, аппроксимирующую функцию $f(X)$ на множестве D_X . В качестве ошибки аппроксимации (целевой метафункции) используем нормированную среднеквадратическую ошибку RMSE (дисперсию ошибки аппроксимации на выборке).

Рассмотрим суррогатные модели целевой функции на основе гауссова процесса. При фиксированном типе используемой ковариационной функции данного процесса основными гиперпараметрами такой модели являются свободные параметры этой функции.

Обозначим $d(X_{i_1}, X_{i_2})$ — евклидово расстояние между точками X_{i_1}, X_{i_2} области D_X ; λ — гиперпараметр, отвечающий за масштаб ковариационной функции. Имеем в виду гауссовы процессы со следующими типами ковариационных функций:

- RBF

$$K_1(X_{i_1}, X_{i_2}, \lambda) = e^{-\frac{d(X_{i_1}, X_{i_2})^2}{2\lambda^2}}; \quad (1)$$

- ExpSineSquared

$$K_2(X_{i_1}, X_{i_2}, \lambda, \rho) = e^{-\frac{2 \sin^2(\pi d(X_{i_1}, X_{i_2})/\rho)}{\lambda^2}}, \quad (2)$$

где ρ — параметр, отвечающий за период ковариационной функции;

- Matern

$$K_3(X_{i_1}, X_{i_2}, \lambda, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{\lambda} d(X_{i_1}, X_{i_2}) \right)^\nu B_\nu \left(\frac{\sqrt{2\nu}}{\lambda} d(X_{i_1}, X_{i_2}) \right), \quad (3)$$

где B_ν — функция Бесселя; Γ — гамма-функция; ν — гиперпараметр, определяющий плавность ковариационной функции;

- Rational Quadratic

$$K_4(X_{i_1}, X_{i_2}, \alpha, \lambda) = \left(1 - \frac{d(X_{i_1}, X_{i_2})^2}{2\alpha\lambda^2} \right)^{-\alpha}, \quad (4)$$

где α — масштабный параметр.

Рекомендованные значения указанных гиперпараметров следующие: $\lambda^{\text{rec}} = 1, 0$; $\rho^{\text{rec}} = 1, 0$; $\nu^{\text{rec}} = 1, 5$; $\alpha^{\text{rec}} = 1, 0$.

Задача настройки гиперпараметров. Принимаем, что $T = \{(X_i, f_i); i \in [1 : |T|]\}$ — тестовая выборка мощностью $|T|$, аналогичная выборке L . При фиксированном типе ковариационной функции $K(X_{i_1}, X_{i_2}, P)$ задачу настройки ставим как задачу глобальной оптимизации (метаоптимизации) вида

$$\min_{P \in D_P \subset \mathfrak{R}^{|P|}} \mu(P, T) = \mu(P^*, T), \quad (5)$$

где $\mu(\cdot)$ — целевая метафункция; P — вектор гиперпараметров; P^* — искомый оптимальный вектор гиперпараметров; D_P — множество допустимых значений компонентов вектора P .

В качестве метафункции $\mu(P, T)$ используем нормированную среднеквадратическую ошибку RMSE, т. е. нормированную дисперсию ошибки аппроксимации на выборке T :

$$\mu(P, T) = \frac{\sqrt{\frac{1}{|T|} \sum_{i=1}^{|T|} (f_i - \phi_i(P))^2}}{\sqrt{\frac{1}{|L+T|} \sum_1^{|L+T|} (f_i - \bar{f}_i)^2}}.$$

Здесь $\phi_i = \phi(X_i, P, T)$ — предсказанное с помощью суррогатной модели значение целевой функции $f(X)$ в точке $X_i \in T$; $|L+T|$ — суммарный объем этих выборок; \bar{f}_i — среднее арифметическое значений этой функции в точках обучающей и тестовой выборок.

Характерные признаки обучающей выборки. Рассмотрим следующие характерные признаки обучающей выборки $L = \{(X_i, f_i), i \in [1 : |L|]\}$, вычисление которых реализовано в открытой программной библиотеке FLACCO [14]:

C_1 (pca_expl_var_cov_x) — доля дисперсии значений варьируемых параметров в выборке L , объясняемая главными компонентами, полученными путем анализа ковариационной матрицы;

C_2 (pca_expl_var_cog_x) — величина, аналогичная C_1 , но основанная на корреляционной матрице значений варьируемых параметров в выборке L ;

C_3 (pca_expl_var_cov_init) — величина, аналогичная C_1 , но основанная на анализе ковариационной матрицы всей выборки L , включая значения варьируемых параметров и целевой функции;

C_4 (pca_expl_var_cor_init) — величина, аналогичная C_2 , но основанная на корреляционной матрице всей выборки L ;

C_5 (pca_expl_var_cov_PC1_x) — доля дисперсии значений варьируемых параметров в выборке L , объясняемая первым главным компонентом;

C_6 (pca_expl_var_cor_PC1_x) — величина, аналогичная C_5 , но основанная на корреляционной матрице значений варьируемых параметров в выборке L ;

C_7 (pca_expl_var_cov_PC1_init) — величина, аналогичная C_5 , но вычисленная на основе ковариационной матрицы всей выборки L ;

C_8 (pca_expl_var_cor_PC1_init) — величина, аналогичная C_7 , но вычисленная на основе корреляционной матрицы всей выборки L ;

C_9 (nbc_nn_nb_sd_ratio) — отношение стандартных отклонений расстояний от каждой точки выборки L до другой ближайшей точки и до другой ближайшей точки с лучшим значением целевой функции (расстояние между точками оцениваем на основе евклидовой нормы в пространстве $\mathfrak{R}^{|X|}$);

C_{10} (nbc_nn_nb_mean_ratio) — отношение средних арифметических значений расстояний, вычисленных по аналогии с C_9 ;

C_{11} (nbc_dist_ratio_coeff_var) — дисперсия отношений расстояний, вычисленных по аналогии с C_9 ;

C_{12} (nbc_nb_fitness_cor) — корреляция между значением целевой функции в выборке L и числом других точек, для которых данная точка является ближайшим лучшим соседом;

C_{13} (ic_h_max) — максимальное значение информационной энтропии, вычисленной для упорядоченной некоторым образом последовательности значений целевой функции $\{f_i, i \in [1 : |L|]\}$;

C_{14} (ic_eps_s) — значение параметра чувствительности метода информационного содержания, при котором энтропия последовательности значений целевой функции $\{f_i, i \in [1 : |L|]\}$ близка к нулю;

C_{15} (ic_eps_max) — величина, аналогичная C_{14} , но при которой энтропия максимальна;

C_{16} (ic_eps_ratio) — значение чувствительности метода информационного содержания, при котором достигается среднее значение коэффициента частичного информационного содержания;

C_{17} (ic_m0) — значение коэффициента частичного информационного содержания при минимальном значении чувствительности метода информационного содержания.

Схема исследования. В качестве целевых используем 15 функций $f(X)$ из библиотеки тестовых функций [15], имеющих следующие размерности $|X|$ вектора варьируемых параметров:

$$\begin{aligned} f_1(X) \text{ (ALLINIT)} &— |X^1|=4; & f_2(X) \text{ (CAMEL6)} &— |X^2|=2; \\ f_3(X) \text{ (EG1)} &— |X^3|=3; & f_4(X) \text{ (HADAMALS)} &— |X^4|=2; \\ f_5(X) \text{ (HART6)} &— |X^5|=6; & f_6(X) \text{ (HIMMELP1)} &— |X^6|=2; \\ f_7(X) \text{ (HS38)} &— |X^7|=4; & f_8(X) \text{ (HS4)} &— |X^8|=2; \\ f_9(X) \text{ (HS45)} &— |X^9|=5; & f_{10}(X) \text{ (HS5)} &— |X^{10}|=2; \\ f_{11}(X) \text{ (LOGROS)} &— |X^{11}|=2; & f_{12}(X) \text{ (MCCORMCK)} &— |X^{12}|=10; \\ f_{13}(X) \text{ (MDHOLE)} &— |X^{13}|=2; & f_{14}(X) \text{ (S368)} &— |X^{14}|=8; \\ f_{15}(X) \text{ (SINEALI)} &— |X^{15}|=10. \end{aligned}$$

Схема исследования включает в себя 15 итераций для каждой ковариационной функции $K_n(X_{i_1}, X_{i_2}, P)$, $n \in [1 : 4]$. Схема итерации $i \in [1 : 15]$ имеет следующий вид (для простоты записи индекс n опускаем).

Шаг 1. Из списка тестовых функций $S = \{f_1(X), \dots, f_{15}(X)\}$ удаляем без повторений функцию $f_i(X)$, так что в этом списке остается набор функций $S_i = \{f_k(X), k \in [1 : 15], k \neq i\}$.

Шаг 2. Для каждой из 14 функций списка S_i генерируем 10 обучающих выборок $\{L_{k,r}, r \in [1 : 10]\}$ объемом $10 |X^k|^2$ каждая. Точки этих выборок размещаем равномерно случайно в интервале $[-1, 0; 1, 0]$ по каждому измерению пространства поиска, так что в результате эти точки оказываются распределенными равномерно случайно в гиперкубе $\otimes \prod_{k=1}^{|X^j|} [-1, 0; 1, 0]$, где $\otimes \prod$ — символ прямого произведения.

Шаг 3. Для каждой выборки $\{L_{k,r}\}$ вычисляем значения характерных признаков $\{C_m(L_{k,r}), m \in [1 : 17]\}$, указанных в п. 1.

Шаг 4. Аналогично шагу 2 для тестовой функций $f_i(X)$ генерируем тестовую выборку T_i мощностью 5000.

Шаг 5. Для каждой функции списка S_i и каждой обучающей выборки $\{L_{k,r}\}$ находим значения гиперпараметров $P_{k,r}^*$, при которых соответствующее значение метафункции $\mu(P_{k,r}^*, T)$ минимально.

Для данного i и для каждой данной ковариационной функции $K_n(X_{i_1}, X_{i_2}, P)$ результатом выполнения шагов 2–5 является набор значений характерных признаков $\{C_m(L_{k,r}), m \in [1 : 17]\}$, а также набор оптимальных значений гиперпараметров $P_{k,r}^*$, т. е. 140 пар векторов $\{\{C_m(L_{k,r})\}, P_{k,r}^*\}$. Эти пары векторов определяют отображение $C \mapsto P^*$, т. е. отображение пространства характерных признаков обучающей выборки в пространство оптимальных гиперпараметров суррогатной модели с рассматриваемой ковариационной функцией $K_n(X_{i_1}, X_{i_2}, P)$.

Шаг 6. Для тестовой функции $f_i(X)$, не вошедшей в набор S_i , выполняем следующие действия:

– по схеме шагов 2, 3 формируем 10 обучающих выборок $\{L_{i,r}\}$ и определяем значения их характерных признаков $\{C_m(L_{i,r})\}$;

– вычисляем средние значения $\{\bar{C}_{i,m}\}$ каждого характерного признака $C_1 - C_{17}$ на выборках $\{L_{i,r}\}$ (эти величины необходимы только для определения наиболее значимых признаков, на основе которых строим прогнозирующую модель), т. е. полагаем, что

$$\bar{C}_{i,m} = \frac{1}{10} \sum_{r=1}^{10} C_m(L_{i,r}), \quad m \in [1 : 17];$$

– на основе каждой выборки $\{L_{i,r}\}$ строим суррогатную модель $\phi_{i,r}^{\text{rec}}(X)$ функции $f_i(X)$, используя рекомендуемые значения гиперпараметров P^{rec} рассматриваемой ковариационной функции;

– методом случайного леса [16], используя отображение $C \mapsto P^*$ и выбранные значимые характерные признаки, строим прогнозирующую модель;

– с использованием построенной прогнозирующей модели находим приближенно оптимальные значения гиперпараметров $P_{i,r}^*$ для данной выборки, затем строим суррогатную модель $\phi_{i,r}^*(X)$ функции $f_i(X)$ с этими значениями гиперпараметров для рассматриваемой ковариационной функции $K_n(X_{i_1}, X_{i_2}, P_{i,r}^*)$.

После завершения указанных 15 итераций оцениваем эффективность полученных суррогатных моделей $\phi_{i,r}^{\text{rec}}(X)$, $\phi_{i,r}^*(X)$, $i \in [1 : 15]$, $r \in [1 : 10]$ на тестовых выборках $\{T_i\}$, т. е. находим среднее на этих выборках значение метафункции $\bar{\mu}_i^{\text{rec}}$ при рекомендованных значениях гиперпараметров, а также аналогичные значения метафункции $\bar{\mu}_i^*$ при оптимальных значениях гиперпараметров.

Программная реализация и организация вычислительных экспериментов. В качестве языка программирования использован Python. Достоинствами языка являются его простота, удобство, наличие большого числа готовых решений в виде открытых программных библиотек, высокая скорость разработки.

В качестве основной математической библиотеки применена библиотека NumPy. С помощью этой библиотеки реализованы матричные операции над данными, генерация обучающих и тестовых выборок, выполнены оценки качества аппроксимации (значений метафункции) и вывод результатов в виде графиков с применением библиотеки Matplotlib.

Для вычисления значений тестовых целевых функций использован известный пакет для бенчмаркинга алгоритмов оптимизации PyCUTEst, реализующий Python интерфейс для коллекции тестовых задач оптимизации CUTEst (The Constrained and Unconstrained Testing Environment).

Вычислительные эксперименты выполнены для указанных в п. 1 ковариационных функций со следующими областями допустимых значений гиперпараметров:

- $\lambda \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$ — для ковариационных функций $K_1 - K_4$;
- $\rho \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$ — для функции K_2 ;
- $\nu \in \{0,5; 1,5; 2,5; \infty\}$ — для функции K_3 ;
- $\alpha \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$ — для функции K_4 .

Оптимальные значения гиперпараметров определяем методом поиска по сетке, включающей 30 уникальных значений в указанных пределах для каждого гиперпараметра, исключая ν . Суммарно проверено 30, 900, 120 и 900 комбинаций значений гиперпараметров для ковариационных функций $K_1 - K_4$ соответственно.

Результаты вычислительных экспериментов. Выполнены три серии вычислительных экспериментов:

1) определение значимости характерных признаков обучающей выборки;

2) исследование чувствительности критерия качества аппроксимации (метафункции) к изменению гиперпараметров суррогатной модели;

3) оценка эффективности использования прогнозных значений гиперпараметров.

Значимость характерных признаков обучающей выборки. Для каждой рассматриваемой ковариационной функции реализованы следующие действия:

– по схеме шагов 1–5 для всех 150 выборок $\{L_{k,r}, k \in [1:15], r \in [1:10]\}$ вычислены значения характерных признаков $\{C_m(L_{k,r}), m \in [1:17]\}$, а также соответствующие значения метафункции $\mu_{k,r}^{\text{rec}}$;

– для каждой из $k \in [1:15]$ функций на основе характерных признаков выборок $\{L_{k,r}, r \in [1:10]\}$ определены средние значения каждого характерного признака $\{\bar{C}_{k,m}, m \in [1:17]\}$, а также средние значения метафункции $\bar{\mu}_k^{\text{rec}}$;

– для каждой пары значений $(\{\bar{C}_{k,m}\} / \bar{\mu}_k^{\text{rec}})$ вычислены коэффициенты корреляции Пирсона $r_m, m \in [1:17]$ [17] (коэффициент r_m определяет относительную значимость характерного признака C_m).

Результаты исследования приведены на рис. 1.

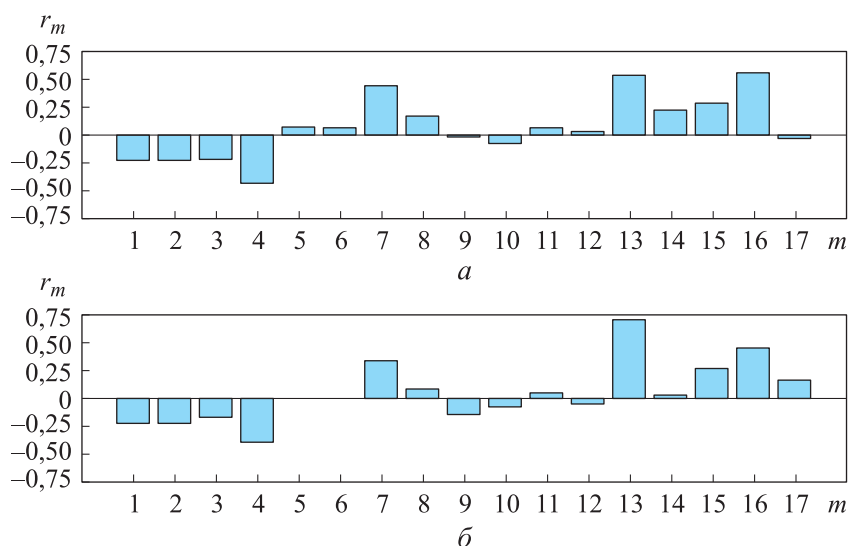


Рис. 1 (начало). Коэффициенты корреляции Пирсона для характерных признаков $\{C_m, m \in [1:17]\}$ ковариационных функций RBF (а), ExpSineSquared (б)

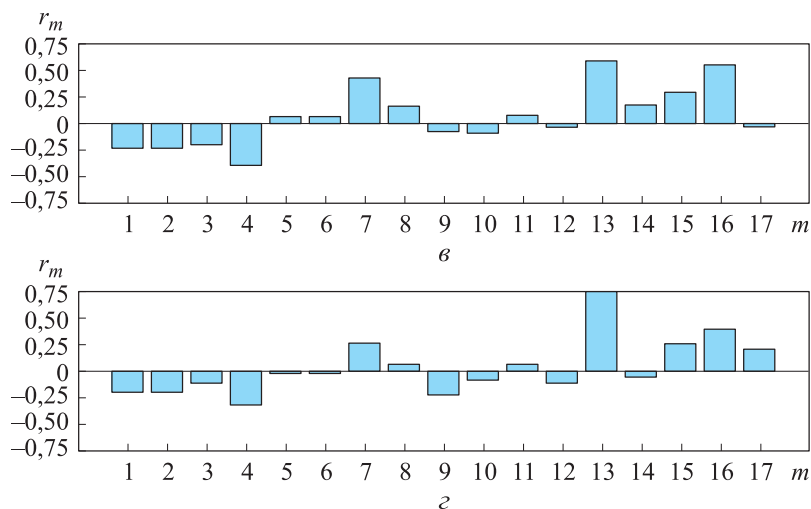


Рис. 1 (окончание). Коэффициенты корреляции Пирсона для характерных признаков $\{C_m, m \in [1 : 17]\}$ ковариационных функций Matern (v), Rational Quadratic (z)

Согласно приведенным результатам, с точки зрения влияния на значения метафункции наибольшую значимость имеют следующие четыре характерных признака:

- C_4 (pca_expl_var_cor_init);
- C_7 (pca_expl_var_cov_PC1_init);
- C_{13} (ic_h_max);
- C_{16} (ic_eps_ratio).

На этом основании в дальнейших исследованиях при построении прогнозирующей модели и определении приближенно оптимальных значений гиперпараметров используем только эти характерные признаки.

Приближенно оптимальные интервалы значений гиперпараметров суррогатной модели. Для каждой ковариационной функции $K_n(X_{i_1}, X_{i_2}, P)$, $n \in [1 : 4]$, выполнены шаги 1–4, а шаг 5 реализован в следующем варианте:

- покрываем интервалы допустимых значений каждого гиперпараметра рассматриваемой ковариационной функции логарифмической сеткой из 30 значений (рис. 2–5);
- для значений гиперпараметров суррогатной модели, соответствующих каждому узлу указанной сетки, и для всех 150 обучающих выборок вычисляем значения метафункции $\mu(\cdot)$;
- на основе полученных результатов определяем для каждой обучающей выборки наилучшие значения гиперпараметров, а также число обучающих выборок N , для которых эти значения также являются наилучшими;

– по аналогичной схеме определяем те значения гиперпараметров, а также соответствующие числа обучающих выборок, которые дают близкую к наилучшей оценку метафункции с точностью до 5 %.

Результаты серии вычислительных экспериментов приведены на рис. 2–5.

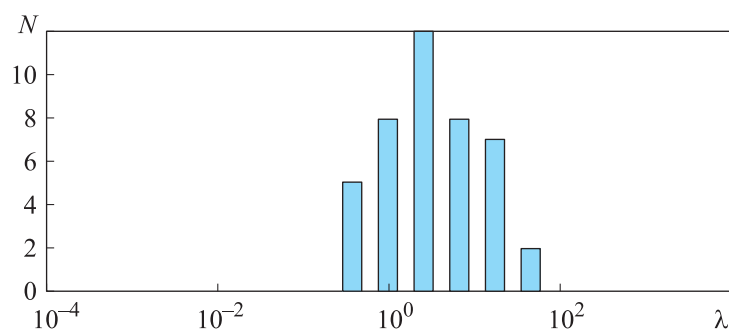


Рис. 2. Число выборок N в функции значений гиперпараметра λ ; ковариационная функция RBF

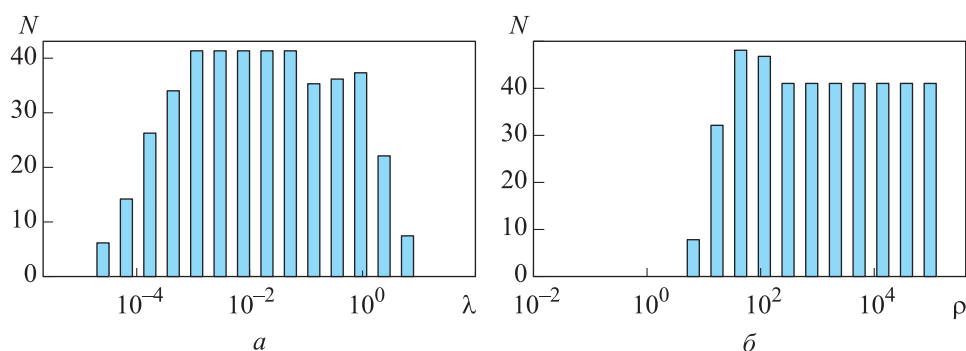


Рис. 3. Число выборок N в функции значений гиперпараметров λ ($\rho \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$) (а) и ρ ($\lambda \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$) (б); ковариационная функция ExpSineSquared

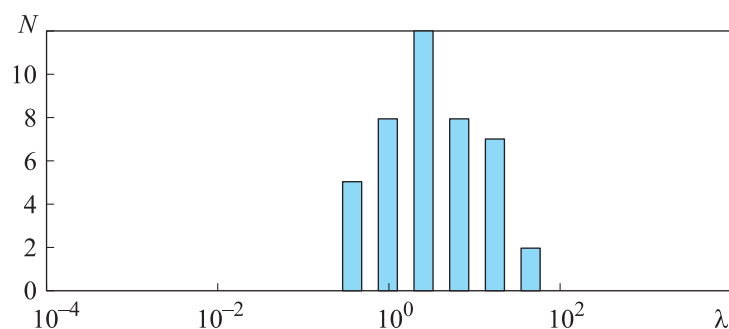


Рис. 4. Число выборок N в функции значений гиперпараметра λ ; ковариационная функция Matern; $\nu \in \{0,5; 1,5; 2,5; \infty\}$

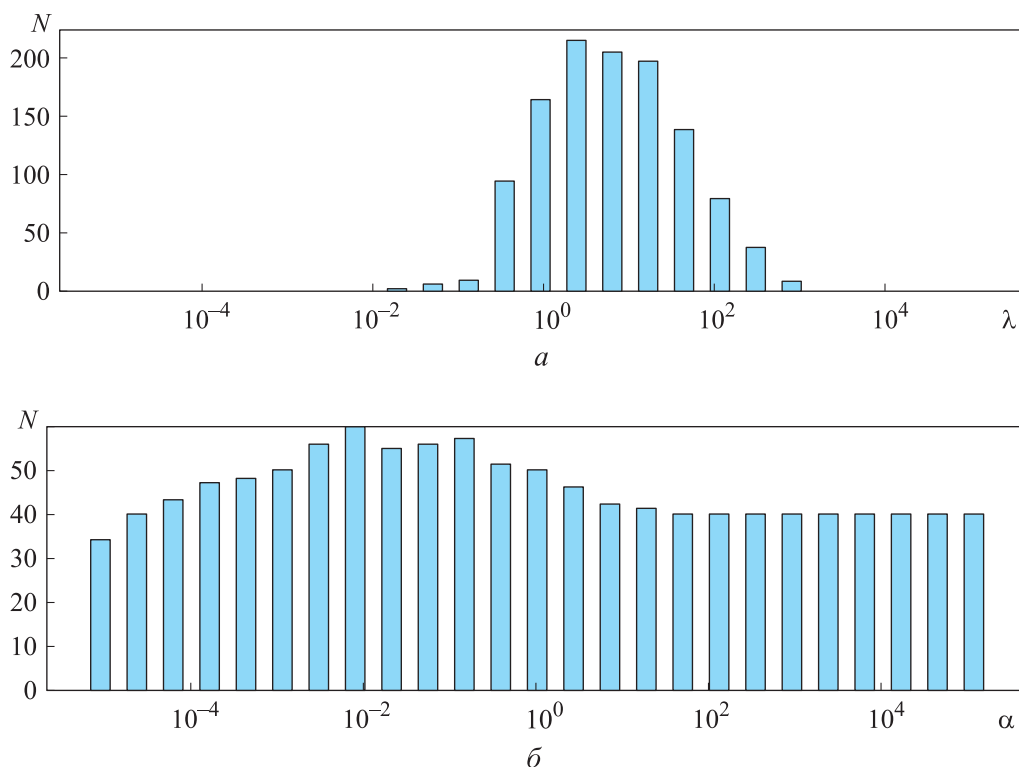


Рис. 5. Число выборок N в функциях значений гиперпараметров λ ($\alpha \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$) (а) и α ($\lambda \in [1 \cdot 10^{-5}; 1 \cdot 10^5]$) (б); ковариационная функция Rational Quadratic

Результаты исследования показывают, что для всех ковариационных функций качество модели существенно зависит от выбранного значения гиперпараметра λ . Этот результат может быть использован в целях сужения областей поиска оптимальных значений гиперпараметров при практическом использовании предлагаемой методики. С этой целью также целесообразно использовать рекомендуемое значение гиперпараметра α ковариационной функции Rational Quadratic ввиду незначительности влияния его на качество суррогатной модели.

Эффективность настройки гиперпараметров. Исследование эффективности предложенного в работе метода настройки гиперпараметров выполнено по схеме, рассмотренной ранее, но с использованием не всех, а наиболее значимых компонентов вектора характерных признаков, т. е. с использованием компонентов C_4 , C_7 , C_{13} , C_{16} . В ходе эксперимента обучено 600 суррогатных моделей с использованием прогнозных значений гиперпараметров P^* и 600 моделей с использованием их рекомен-

дованных значений P^{rec} (по 150 моделей для каждой из четырех рассматриваемых ковариационных функций).

Результаты исследования представлены на рис. 6, где на оси абсцисс приведены значения метафункции μ , на оси ординат — относительные числа $M(P^*) = M^*$, $M(P^{\text{rec}}) = M^{\text{rec}}$ суррогатных моделей со значением этой функции лучше данного. Из рис. 6 следует, что 60 % моделей, обученных с использованием значений гиперпараметров P^{rec} , имеют значение метафункции $\mu < 0,1$. В то же время при использовании прогнозных значений гиперпараметров P^* этому условию удовлетворяет до 80 % моделей.

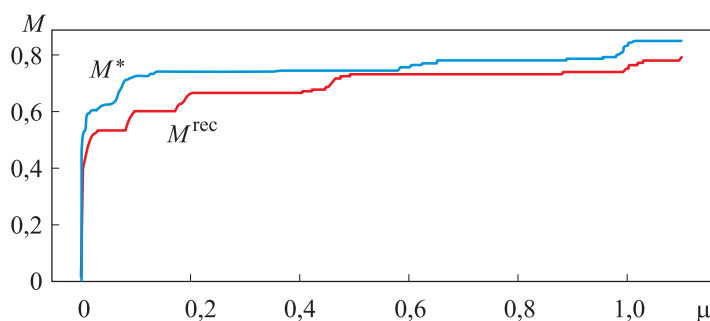


Рис. 6. Эффективность настройки гиперпараметров суррогатной модели

Из рис. 6 также следует, что при использовании прогнозных значений гиперпараметров P^* наблюдается повышение до 30 % качества суррогатных моделей целевой функции по сравнению с использованием рекомендованных значений. Это обстоятельство позволяет надеяться на увеличение эффективности решения исходной задачи параметрической оптимизации на основе синтезированных суррогатных моделей.

Заключение. Предложен новый подход к настройке гиперпараметров гауссовой суррогатной модели целевой функции в задаче параметрической оптимизации проектных решений. Подход предполагает решение исходной задачи оптимизации в два этапа.

Первый этап настройки заключается:

- в формировании совокупности обучающих выборок для представительного набора тестовых целевых функций;
- определении значений характерных признаков каждой указанной выборки, а также приближенно оптимальных значений гиперпараметров используемой суррогатной модели;
- построении на основе полученной совокупности пар вектор значений характерных признаков выборки–оптимальные значения гиперпа-

раметров прогнозирующей модели, которая предлагает приближенно оптимальные значения гиперпараметров по характерным признакам обучающей выборки.

На втором этапе эксплуатации реализовано решение базовой (исходной) задачи параметрической оптимизации по схеме:

- для целевой функции базовой (исходной) задачи сгенерирована обучающая выборка и определены ее характерные признаки;
- спрогнозированы оптимальные значения гиперпараметров используемой суррогатной функции с помощью прогнозирующей модели, полученной на этапе настройки;
- в качестве значений гиперпараметров суррогатной функции использованы их прогнозные значения.

Выполнена программная реализация указанного подхода, а также широкий вычислительный эксперимент по оценке эффективности разработанного математического и программного обеспечения. Показано, что в условиях эксперимента данный подход обеспечивает повышение качества суррогатных моделей целевой функции при приемлемых вычислительных затратах на этапе эксплуатации.

Кроме того, определена значимость рассматриваемых характерных признаков обучающих выборок. Показано, что из числа 17 этих признаков наиболее значимыми являются всего четыре. Именно эти признаки использованы в процессе исследования эффективности подхода.

Кроме того, в экспериментальной части работы определены интервалы наилучших значений гиперпараметров ковариационных функций.

Результаты исследования показывают, что предложенный подход может быть использован в вычислительной практике при решении задач параметрической оптимизации на основе гауссовых суррогатных моделей целевых функций. Если не рассматривать вычислительные затраты на этапе настройки (а они распределяются на все задачи оптимизации, решаемые на этапе эксплуатации), то в результате можно ожидать повышение до 30 % эффективности базового алгоритма оптимизации.

В развитие работы планируется исследовать эффективность подхода на более широком наборе тестовых целевых функций и апробировать его в процессе решения сложных практических задач параметрической оптимизации проектных решений, а также распространить подход на многокритериальную постановку задачи настройки.

Как было отмечено ранее, алгоритмы оптимизации на основе суррогатного моделирования могут использовать не только аппроксимации целевой функции на основе гауссовых процессов, но и другие техники ап-

проксимации — с помощью радиальных базисных функций [18], кригинга [19, 20], поверхностей отклика (Response Surface Methods, RSM) [21, 22]. Представленный подход может быть расширен и на указанные типы аппроксимирующих функций.

ЛИТЕРАТУРА

- [1] Захарова Е.М., Минашина И.К. Обзор методов многомерной оптимизации. *Информационные процессы*, 2014, т. 14, № 3, с. 256–274.
- [2] Liu B., Koziel S., Zhang Q. A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems. *J. Comput. Sc.*, 2016, vol. 12, pp. 28–37.
- [3] Müller J., Shoemaker C.A. Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. *J. Glob. Optim.*, 2014, vol. 60, no. 2, pp. 123–144.
DOI: <https://doi.org/10.1007/s10898-014-0184-0>
- [4] Кулешов А.П. Когнитивные технологии в адаптивных моделях сложных объектов. *Информационные технологии и вычислительные системы*, 2008, № 1, с. 18–29.
- [5] Сунчалин А.М., Сунчалина А.Л. Обзор методов и моделей прогнозирования финансовых временных рядов. *Хроноэкономика*, 2020, № 1.
URL: http://hronoeconomics.ru/01_2020.pdf
- [6] Buhmann M.D. Radial basis functions. Cambridge, Cambridge University Press, 2009.
- [7] Snoek J., Rippel O., Swersky K., et al. Scalable Bayesian optimization using deep neural networks. *PMLR*, 2015, vol. 37, pp. 2171–2180.
- [8] Терехов С.А. Случайные гауссовские процессы в задачах аппроксимации. *X Всерос. науч.-техн. конф. «Нейроинформатика–2008». Лекции по нейроинформатике*. Ч. 1. М., МИФИ, 2008, с. 126–151.
- [9] Binois M., Wuyoff N. A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. *ACM TELO*, 2022, vol. 2, no. 2, art. 8.
DOI: <http://dx.doi.org/10.1145/3545611>
- [10] Luca F., Donini M., Frasconi P., et al. Forward and reverse gradient-based hyperparameter optimization. *Proc. 34th Int. Conf. on Machine Learning*, 2017, vol. 70, pp. 1165–1173.
- [11] Смирнова В.С., Шаламова В.В., Ефимова В.А. и др. Оптимизация гиперпараметров на основе объединения априорных и апостериорных знаний о задаче классификации. *Научно-технический вестник информационных технологий, механики и оптики*, 2020, т. 20, № 6, с. 828–834.
DOI: <https://doi.org/10.17586/2226-1494-2020-20-6-828-834>

- [12] Karpenko A.P., Kuzmina I.A. Structural and parametric synthesis of population algorithms for global optimization. *Procedia Comput. Sc.*, 2021, vol. 186, no. 2, pp. 299–308. DOI: <https://doi.org/10.1016/j.procs.2021.04.207>
- [13] Mersmann O., Bischl B., Trautmann H., et al. Exploratory landscape analysis. *Proc. GECCO'11*, 2011, pp. 829–836. DOI: <https://doi.org/10.1145/2001576.2001690>
- [14] Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems. *IEEE CEC*, 2016, pp. 5262–5269. DOI: <https://doi.org/10.1109/CEC.2016.7748359>
- [15] Fowkes J., Roberts L., Burmen A. PyCUTEst: an open source Python package of optimization test problems. *Open Source Softw.*, 2022, vol. 7, no. 78, art. 4377. DOI: <https://doi.org/10.21105/joss.04377>
- [16] Roy M.H., Larocque D. Robustness of random forests for regression. *J. Nonparametr. Stat.*, 2012, vol. 24, no. 4, pp. 993–1006. DOI: <https://doi.org/10.1080/10485252.2012.715161>
- [17] Наследов А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. СПб., Питер, 2013.
- [18] Gutmann H.M. A radial basis function method for global optimization. *J. Glob. Optim.*, 2001, vol. 19, no. 3, pp. 201–227. DOI: <https://doi.org/10.1023/A:1011255519438>
- [19] De Cock D.R. Kriging as an alternative to polynomial regression in response surface analysis. Ames, Iowa State University, 2003.
- [20] Колодяжный М., Зайцев А. Гетероскедастичные гауссовские процессы и их применение для байесовской оптимизации. *Тр. 42-й Междисциплинарной шк.-конф. ИППИ РАН «ИТиС 2018»*. М., ИППИ РАН, 2018, с. 42–51.
- [21] Barton R.R. Metamodeling: a state of the art review. *Proc. Winter Simulation Conf.*, Philadelphia, Pennsylvania SU, 1994, pp. 237–244. DOI: <https://doi.org/10.1109/WSC.1994.717134>
- [22] Jones D.R. A taxonomy of global optimization methods based on response surfaces. *J. Glob. Optim.*, 2001, no. 21, pp. 345–383. DOI: <https://doi.org/10.1023/A:1012771025575>

Агасиев Талех Азер оглы — канд. техн. наук, доцент кафедры «Системы автоматизированного проектирования» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Гвоздев Никита Павлович — специалист отдела тестирования ООО «ДАТАДВАНС» (Российская Федерация, 117246, Москва, Научный пр-д, д. 17).

Карпенко Анатолий Павлович — д-р физ.-мат. наук, профессор, заведующий кафедрой «Системы автоматизированного проектирования» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Пивоварова Наталья Владимировна — канд. техн. наук, доцент кафедры «Системы автоматизированного проектирования» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Агасиев Т.А., Гвоздев Н.П., Карпенко А.П. и др. Новый подход к настройке гауссовой суррогатной модели целевой функции в задаче параметрической оптимизации проектных решений. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2023, № 3 (144), с. 62–83.

DOI: <https://doi.org/10.18698/0236-3933-2023-3-62-83>

**NEW APPROACH TO ADJUSTING THE OBJECTIVE FUNCTION
GAUSSIAN SURROGATE MODEL IN THE PROBLEM
OF DESIGN SOLUTION PARAMETRIC OPTIMIZATION**

T.A. Agasiev¹

agasiev@bmstu.ru

N.P. Gvozdev²

gvozdevnp@gmail.com

A.P. Karpenko¹

apkarpenko@bmstu.ru

N.V. Pivovarova¹

nvpivovarova@bmstu.ru

¹ **Bauman Moscow State Technical University, Moscow, Russian Federation**

² **LLC “DATADVANCE”, Moscow, Russian Federation**

Abstract

The paper considers methods for solving the problem of design solution parametric optimization based on constructing the Gaussian surrogate model of this problem objective function. The problem is set of finding optimal values of the surrogate model free parameters (hyper-parameters), it is called the problem of its adjustment. The adjustment problem is built over the top of the surrogate model synthesis problem and has a higher computational complexity. The approach to adjusting a surrogate model is proposed, which is able to make the adjustment procedure acceptable in terms of the computational costs. This approach includes the setup and operation stages. The adjustment stage contains the following main steps: formation of a set of test objective functions; generation of a set of learning samples for each of them; determination of their characteristic features values for the generated samples; determination of the hyper-parameters optimal values for all considered

Keywords

Parametric optimization, surrogate simulation, Bayesian approach to optimization, hyper-parameters

test functions and learning samples; formation of a set of pairs, characteristic features of the sample–hyper-parameters optimal values; building on this basis a predictive model forecasting the hyper-parameters optimal values according to the learning sample characteristic features. For the initial problem at the operation stage, a learning sample was generated, its characteristic features were determined, and the hyper-parameters optimal values of the surrogate model were predicted. Based on the specified learning sample, the objective function surrogate model was synthesized. Using the surrogate model, the original optimization problem was solved, where the hyper-parameters predictive values were applied as the optimal values. The approach is able to provide an increase of up to 30 % in efficiency of the basic optimization algorithm

Received 17.03.2023

Accepted 08.06.2023

© Author(s), 2023

REFERENCES

- [1] Zakharova E.M., Minashina I.K. Review of multidimensional optimization techniques. *Informatsionnye protsessy* [Information Processes], 2014, vol. 14, no. 3, pp. 256–274 (in Russ.).
- [2] Liu B., Koziel S., Zhang Q. A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems. *J. Comput. Sc.*, 2016, vol. 12, pp. 28–37.
- [3] Müller J., Shoemaker C.A. Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. *J. Glob. Optim.*, 2014, vol. 60, no. 2, pp. 123–144. DOI: <https://doi.org/10.1007/s10898-014-0184-0>
- [4] Kuleshov A.P. Cognitive technologies in adaptive models of complex objects. *Informatsionnye tekhnologii i vychislitelnye sistemy*, 2008, no. 1, pp. 18–29 (in Russ.).
- [5] Sunchalin A.M., Sunchalina A.L. Overview of methods and models for forecasting financial time series. *Khronoekonomika* [Hronoeconomics], 2020, no. 1 (in Russ.). Available at: http://hronoeconomics.ru/01_2020.pdf
- [6] Buhmann M.D. Radial basis functions. Cambridge, Cambridge University Press, 2009.
- [7] Snoek J., Rippel O., Swrsky K., et al. Scalable Bayesian optimization using deep neural networks. *PMLR*, 2015, vol. 37, pp. 2171–2180.
- [8] Terekhov S.A. [Random Gaussian processes in data approximation problems]. *X Vseros. nauch.-tekh. konf. "Neyroinformatika–2008". Lektsii po neyroinformatike*. Ch. 1 [X. Russ. Sc.-Tech. Conf. Neuroinformatics–2008. Lectures on Neuroinformatics. P. 1]. M., MEPhI, 2008, pp. 126–151 (in Russ.).

- [9] Binois M., Wycoff N. A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. *ACM TELO*, 2022, vol. 2, no. 2, art. 8.
DOI: <http://dx.doi.org/10.1145/3545611>
- [10] Luca F., Donini M., Frasconi P., et al. Forward and reverse gradient-based hyperparameter optimization. *Proc. 34th Int. Conf. on Machine Learning*, 2017, vol. 70, pp. 1165–1173.
- [11] Smirnova V.S., Shalamova V.V., Efimova V.A., et al. Hyperparameter optimization based on a priori and a posteriori knowledge about classification problem. *Nauchno-tekhnicheskiiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Sc. Tech. J. Inf. Technol. Mech. Opt.], 2020, vol. 20, no. 6, pp. 828–834 (in Russ.).
DOI: <https://doi.org/10.17586/2226-1494-2020-20-6-828-834>
- [12] Karpenko A.P., Kuzmina I.A. Structural and parametric synthesis of population algorithms for global optimization. *Procedia Comput. Sc.*, 2021, vol. 186, no. 2, pp. 299–308. DOI: <https://doi.org/10.1016/j.procs.2021.04.207>
- [13] Mersmann O., Bischl B., Trautmann H., et al. Exploratory landscape analysis. *Proc. GECCO'11*, 2011, pp. 829–836. DOI: <https://doi.org/10.1145/2001576.2001690>
- [14] Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems. *IEEE CEC*, 2016, pp. 5262–5269. DOI: <https://doi.org/10.1109/CEC.2016.7748359>
- [15] Fowkes J., Roberts L., Burmen A. PyCUTEst: an open source Python package of optimization test problems. *Open Source Softw.*, 2022, vol. 7, no. 78, art. 4377.
DOI: <https://doi.org/10.21105/joss.04377>
- [16] Roy M.H., Larocque D. Robustness of random forests for regression. *J. Nonparametr. Stat.*, 2012, vol. 24, no. 4, pp. 993–1006.
DOI: <https://doi.org/10.1080/10485252.2012.715161>
- [17] Nasledov A. IBM SPSS Statistics 20 i AMOS: professionalnyy statisticheskiy analiz dannykh [IBM SPSS Statistics 20 and AMOS: professional statistical data analysis]. St. Petersburg, Piter Publ., 2013.
- [18] Gutmann H.M. A radial basis function method for global optimization. *J. Glob. Optim.*, 2001, vol. 19, no. 3, pp. 201–227. DOI: <https://doi.org/10.1023/A:1011255519438>
- [19] De Cock D.R. Kriging as an alternative to polynomial regression in response surface analysis. Ames, Iowa State University, 2003.
- [20] Kolodyazhnyy M., Zaytsev A. [Heteroscedastic Gaussian processes and their application to Bayesian optimisation]. *Tr. 42-y Mezhdisciplinarnoy shk.-konf. IPPI RAN "ITiS 2018"* [Proc. 42nd Interdisciplinary School-Conf. IPPI RAS "ITiS 2018"]. Moscow, IPPI RAS Publ., 2018, pp. 42–51 (in Russ.).
- [21] Barton R.R. Metamodeling: a state of the art review. *Proc. Winter Simulation Conf.*, Philadelphia, Pennsylvania SU, 1994, pp. 237–244.
DOI: <https://doi.org/10.1109/WSC.1994.717134>
- [22] Jones D.R. A taxonomy of global optimization methods based on response surfaces. *J. Glob. Optim.*, 2001, no. 21, pp. 345–383.
DOI: <https://doi.org/10.1023/A:1012771025575>

Agasiev T.A. — Cand. Sc. (Eng.), Assoc. Professor, Department of Computer-Aided Design Systems, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Gvozdev N.P. — Specialist of the Testing Department, LLC “DATADVANCE” (Nauchniy proezd 17, Moscow, 117246 Russian Federation).

Karpenko A.P. — Dr. Sc. (Phys.-Math.), Professor, Head of the Department of Computer-Aided Design Systems, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Pivovarova N.V. — Cand. Sc. (Eng.), Assoc. Professor, Department of Computer-Aided Design Systems, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Agasiev T.A., Gvozdev N.P., Karpenko A.P., et al. New approach to adjusting the objective function Gaussian surrogate model in the problem of design solution parametric optimization. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2023, no. 3 (144), pp. 62–83 (in Russ.).

DOI: <https://doi.org/10.18698/0236-3933-2023-3-62-83>