

ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION
TECHNOLOGIES
AND TELECOMMUNICATIONS

УДК 519.242

DOI: 10.17212/2782-2001-2024-2-55-68

Алгоритмы построения дискретных приближенно Q -оптимальных планов эксперимента при активной идентификации регрессионных моделей многофакторных систем*

А.А. ПОПОВ

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

a.popov@corp.nstu.ru

В теории оптимального эксперимента имеется группа критериев оптимальности (например, D -, A -, E -критерии), отражающих точность оценивания параметров модели. Существует также группа критериев, связанных с точностью прогноза по модели, которая может характеризоваться дисперсией оценок математических ожиданий откликов. Например, использование критерия G -оптимальности позволяет получать планы, на которых построенные модели будут обеспечивать минимизацию максимальной дисперсии прогноза. К числу подобных относится и критерий Q -оптимальности, предполагающий минимизацию средней по области планирования дисперсии прогноза по регрессионной модели.

В прикладных исследованиях, связанных с построением планов экспериментов, часто используется критерий D -оптимальности. Это объясняется еще и тем, что критерии D - и G -оптимальности связаны между собой. В то же время необходимо отметить, что минимизация максимальной дисперсии в общем случае может не приводить к снижению средней по области дисперсии прогноза. Поэтому использование Q -оптимальных планов в практических задачах регрессионного моделирования актуально. Для широкого внедрения в практику активной идентификации регрессионных моделей концепции Q -оптимальности планов эксперимента необходим арсенал эффективных алгоритмов их построения.

В работе предлагаются и описываются алгоритмы построения дискретных приближенно Q -оптимальных планов. Предлагаемые алгоритмы построены на базе развиваемого подхода последовательного наращивания числа точек в плане, а также процедур замены точек в плане. Полученные такими алгоритмами планы рекомендуются к использованию на практике, когда требуется в среднем хорошая точность прогноза по модели на всей области действия входных факторов.

Ключевые слова: регрессионная модель, план эксперимента, критерии оптимальности, дисперсия прогноза по модели, Q -оптимальные планы, дискретные планы, алгоритм добавления, алгоритм добавления-удаления

* Статья получена 30 января 2024 г.

ВВЕДЕНИЕ

При проведении активной идентификации регрессионных моделей важное значение имеет выбор критерия оптимальности планов эксперимента. Их предложено много [1]. Для ряда критериев эллипсоид рассеяния оценок параметров имеет особенности: минимальный объем, минимальные габаритные размеры. В начальный период развития теории оптимального планирования эксперимента основное внимание уделялось критерию D -оптимальности [2–5]. Для D -оптимальных планов часто случается так, что эллипсоид рассеивания оценок параметров оказывается вытянутым по какой-либо главной оси. В теории планирования эксперимента отмечается эквивалентность D - и G -оптимальных непрерывных планов эксперимента. Однако при этом точность предсказания математического ожидания отклика в различных частях области планирования может существенно различаться. Сгладить эти различия можно при использовании Q -оптимальных планов. Публикаций по вопросам построения и использования Q -оптимальных планов немного. Здесь особо нужно отметить работу [6], в которой рассмотрены вопросы аналитического построения таких планов для некоторых типов регрессионных моделей.

1. ПРИБЛИЖЕННО Q -ОПТИМАЛЬНЫЕ ПЛАНЫ ДЛЯ ОДНОФАКТОРНОЙ РЕГРЕССИИ

Предположим, что заданная модель исследуемого объекта имеет вид

$$\eta(x, \theta) = f^T(x)\theta = \sum_{l=1}^m f_l(x)\theta_l, \quad (1)$$

где факторы $x = (x_1, \dots, x_k)^T$ могут изменяться в области $\tilde{X}$. Предполагается, что на исследуемом объекте можно проводить так называемые активные эксперименты по заранее построенному плану. Оговорим здесь, что будем использовать в одних случаях обозначение x_j как j -ю координату вектора переменных x , а в других случаях – как j -ю точку наблюдений. Смысл обозначений будет ясен из контекста. Непрерывный нормированный план ϵ задается совокупностью точек $x_1, x_2, \dots, x_n$ с весами $p_1, p_2, \dots, p_n$. Плану ϵ соответствует информационная матрица

$$M(\epsilon) = \sum_{i=1}^n p_i f(x_i) f^T(x_i).$$

Качество плана эксперимента оценивают по значению некоторого функционала от информационной матрицы. Точность прогноза по модели может характеризоваться дисперсией $\sigma^2(\hat{y}(x))$ оценки $\hat{y}(x) = f^T(x)\hat{\theta}$ математического ожидания отклика, которая вычисляется так:

$$\sigma^2(\hat{y}(x)) = d(x, \epsilon) = f^T(x)M^{-1}(\epsilon)f(x).$$

Оценки $\hat{\theta}$ будем находить по методу наименьших квадратов. В частности, отметим, что план ε^* называется G -оптимальным, если

$$\varepsilon^* = \operatorname{Arg} \min_{\varepsilon} \max_{x \in \tilde{X}} d(x, \varepsilon).$$

И, наконец, план ε^* называется Q -оптимальным, если

$$\varepsilon^* = \operatorname{Arg} \min_{\varepsilon} \int_{\tilde{X}} d(x, \varepsilon) dx. \quad (2)$$

Q -оптимальные планы минимизируют среднюю по области $\tilde{X}$ дисперсию оценки поверхности отклика. Исторически так сложилось, что оптимальные по D -критерию планы часто применяют на практике. В работе [7] нами рассмотрены вопросы сравнения A -оптимальных планов с другими известными. Очень часто робастным критерием оказывался критерий A -оптимальности, поскольку получаемый по нему план был близок к оптимальным планам по другим критериям. В то же время найти публикации по теме построения Q -оптимальных планов крайне сложно. В отмеченной работе [6] Q -оптимальные планы для полиномиальной регрессии на отрезке строились аналитически с привлечением символьных вычислений, что ограничивает применение данного подхода на практике.

Продолжая исследования работы [7], добавим в число сравниваемых критериев Q -оптимальности. В табл. 1 приведен ряд построенных приближенно Q -оптимальных планов для полиномиальной регрессии на отрезке. Построение приближенно Q -оптимальных планов проводилось с использованием в качестве функционала средней дисперсии оценки отклика, вычисленной на отрезке $[-1, +1]$ по равномерной сетке $X_{\#}$ с числом узлов в ней $N_{\#}$. Число узлов $N_{\#}$ выбиралось равным $N_{\#} = 161$. В данном случае мы при конструировании функционала Q -оптимальности не используем аналитическое выражение для интеграла в (2), а заменяем его на приближенное выражение в виде средней дисперсии на сетке $X_{\#}$. Характеристики приближенно Q -оптимальных планов незначительно отличаются от точных аналитических, представленных, например, в работе [6]. Однако данный подход существенно проще использовать на практике при исследовании многофакторных систем и процессов. Для полноты в табл. 2 приведены характеристики оптимальных планов для полиномиальной степени q регрессии на отрезке.

Построение приближенно Q -оптимальных планов проводилось численными методами до выполнения условий оптимальности. В нашем случае

$$\max_{x \in \tilde{X}} \varphi(x, \varepsilon^*) = \sum_{x' \in X_{\#}} d(x', \varepsilon^*), \quad (3)$$

$$\begin{aligned} \text{где } \varphi(x, \varepsilon^*) &= \sum_{x' \in X_{\#}} d^2(x, x', \varepsilon^*), \quad d(x, x', \varepsilon) = f^T(x) M^{-1}(\varepsilon) f(x'), \quad d(x', \varepsilon) = \\ &= f^T(x') M^{-1}(\varepsilon) f(x'). \end{aligned}$$

Выполнение условий (3) в численных методах построения оптимальных планов осуществляется до достижения заданной точности. В рассматриваемых примерах точность выполнения условий (3) была равна 0,001. Поиск максимума функции $\varphi(x, \varepsilon)$ осуществлялся на сетке с шагом 0,025.

Анализ числовых характеристик, приведенных в табл. 2, позволяет сделать вывод, что приближенно Q -оптимальные планы близки по степени робастности к A -оптимальным планам. Таким образом, список робастных критериев оптимальности можно расширить, добавив в него критерий Q -оптимальности.

Таблица 1

Table 1

Приближенно Q -оптимальные планы для полиномиальной регрессии степени q на отрезке

Approximately Q -optimal designs for polynomial degree q regression on the interval

План	q	n	x_1	x_2	x_3	x_4	x_5
			p_1	p_2	p_3	p_4	p_5
ε_Q^*	2	3	-1	0	+1	—	—
			0,2514	0,4972	0,2514	—	—
ε_Q^*	3	4	-1	-0,4338	0,4338	+1	—
			0,1573	0,3427	0,3427	0,1573	—
ε_Q^*	4	5	-1	-0,642	0	0,642	+1
			0,1106	0,2482	0,2824	0,2482	0,1106

Таблица 2

Table 2

Характеристики оптимальных планов для полиномиальной регрессии степени q на отрезке

Characteristics of optimal designs for polynomial degree q regression on the interval

q	План	$ M(\varepsilon) $	$trM^{-1}(\varepsilon)$	$\lambda_{\max}(M^{-1}(\varepsilon))$	$Q(M^{-1}(\varepsilon))$	$\max_{x \in X} d(x, \varepsilon)$
2	ε_D^*	0,148	9,00	6,842	2,403	3,0
	ε_A^*	0,125	8,00	5,236	2,145	4,0
	ε_E^*	0,096	8,333	5,000	2,239	5,0
	ε_Q^*	0,1257	8,0	5,249	2,145	3,97
3	ε_D^*	$0,512 \cdot 10^{-2}$	44,00	33,58	3,432	4,0
	ε_A^*	$0,363 \cdot 10^{-2}$	37,52	25,77	3,032	6,61
	ε_E^*	$0,285 \cdot 10^{-2}$	38,68	25,00	3,142	7,85
	ε_Q^*	$0,380 \cdot 10^{-2}$	37,94	27,143	3,011	6,35

Окончание табл. 2

End of the Tab. 2

q	План	$ M(\epsilon) $	$trM^{-1}(\epsilon)$	$\lambda_{\max}(M^{-1}(\epsilon))$	$Q(M^{-1}(\epsilon))$	$\max_{x \in X} d(x, \epsilon)$
4	ϵ_D^*	$0,429 \cdot 10^{-4}$	224,87	174,42	4,449	5,0
	ϵ_A^*	$0,265 \cdot 10^{-4}$	188,72	132,75	3,949	9,47
	ϵ_E^*	$0,211 \cdot 10^{-4}$	194,52	129,00	4,084	10,75
	ϵ_Q^*	$0,285 \cdot 10^{-4}$	193,48	144,04	3,901	9,04

В то же время следует отметить несколько неожиданный момент. Несмотря на то что критерий Q -оптимальности предполагает минимизацию средней дисперсии прогноза на области планирования, можно видеть, что максимальная дисперсия прогноза значительно выше, чем у D -оптимальных планов. Например, если для полиномиальной регрессии 4-й степени для D -оптимального плана она равна пяти, то для приближенно Q -оптимального плана она уже равна 9,04. Справедливости ради необходимо отметить, что для E -оптимальных планов она еще выше и достигает значения 10,75. Для A -оптимальных планов она также велика. На рис. 1 и 2 представлены графики функции $d(x, \epsilon)$ соответственно для приближенно Q -оптимального и D -оптимального плана для полинома 4-й степени. Можно видеть, что для приближенно Q -оптимального плана на практически 80...90 % ширины отрезка действия фактора дисперсия оценки функции отклика близка или меньше 4,0. И наоборот: для D -оптимального плана на большей части отрезка она больше 4,0.

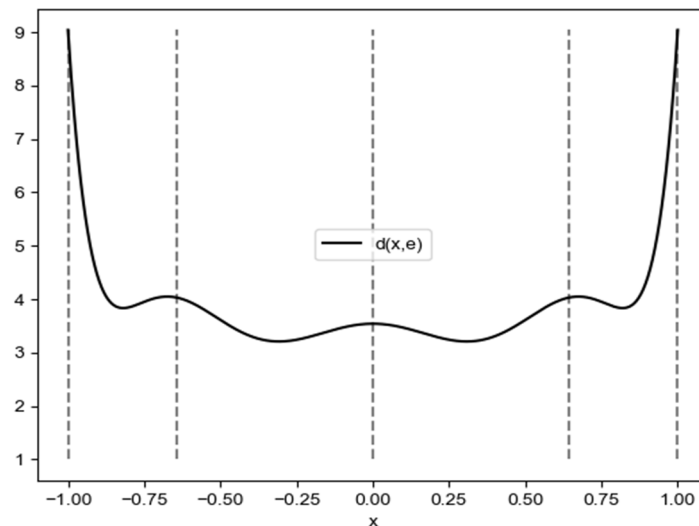


Рис. 1. График функции $d(x, \epsilon)$ приближенно Q -оптимального плана для полинома 4-й степени на отрезке

Fig. 1. Graph of the function $d(x, \epsilon)$ for the approximately Q -optimal design for a polynomial of degree 4 on the interval

Для приближенно Q -оптимального плана дисперсия оценки функции отклика начинает расти ближе к границам отрезка, оставаясь почти постоянной в других частях и в середине отрезка. Такое поведение дисперсии оценки функции отклика характерно для так называемых равномерных планов. На практике это означает, что поскольку на большей части области определения факторов вокруг центра эксперимента точность предсказания по модели примерно одинакова, то мы можем уверенно решать задачу поиска оптимальных режимов протекания исследуемого процесса.

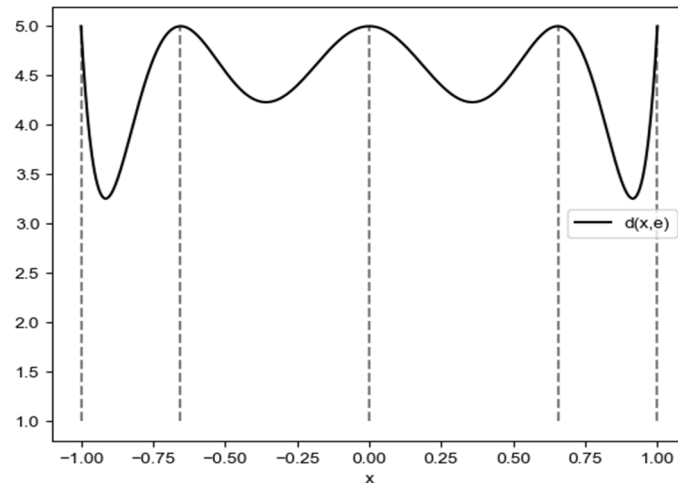


Рис. 2. График функции $d(x, \epsilon)$ D -оптимального плана для полинома 4 степени на отрезке

Fig. 2. Graph of the function $d(x, \epsilon)$ for the D -optimal design for a polynomial of degree 4 on the interval

Непрерывные Q -оптимальные планы на практике не применяются, а используются так называемые дискретные планы.

2. АЛГОРИТМЫ ПОСТРОЕНИЯ ДИСКРЕТНЫХ ПРИБЛИЖЕННО Q -ОПТИМАЛЬНЫХ ПЛАНОВ

Дискретные оптимальные планы можно строить, используя известные алгоритмы из теории нелинейной оптимизации [8]. Однако это будет малоэффективно. Учитывая большую размерность решаемых задач, важно иметь набор алгоритмов специализированного типа [9].

2.1. ПОСЛЕДОВАТЕЛЬНЫЙ АЛГОРИТМ ПОСТРОЕНИЯ ПРИБЛИЖЕННО Q -ОПТИМАЛЬНЫХ ПЛАНОВ (АЛГОРИТМ SEQ-ADD-Q)

Алгоритм предполагает наращивание числа наблюдений в плане до необходимого и хорошо себя зарекомендовал при решении задачи синтеза дискретных D -оптимальных планов произвольного объема [10, 11]. На каждом шаге в план ϵ_s идет добавление новой точки x_{s+1} с весом $\alpha = 1/(s+1)$.

Добавление еще одного наблюдения в план изменяет расчет информационной матрицы, в результате чего получается новое значение, которое вычисляется по формуле

$$\begin{aligned} M(\epsilon_{s+1}) &= (1 - \alpha)M(\epsilon_s) + \alpha f(x_{s+1})f^T(x_{s+1}) = \\ &= (1 - \alpha) \left[M(\epsilon_s) + \frac{\alpha}{1 - \alpha} f(x_{s+1})f^T(x_{s+1}) \right]. \end{aligned}$$

Обратная к ней вычисляется как

$$\begin{aligned} M^{-1}(\epsilon_{s+1}) &= \left[(1 - \alpha)M(\epsilon_s) + \alpha f(x_{s+1})f^T(x_{s+1}) \right]^{-1} = \\ &= (1 - \alpha)^{-1} \left[I_m + \frac{\alpha}{1 - \alpha} M^{-1}(\epsilon_s) f(x_{s+1})f^T(x_{s+1}) \right]^{-1} M^{-1}(\epsilon_s). \end{aligned}$$

Будем использовать свойство обращения матриц из работы В.В. Федорова [5]:

$$(I_p + AB)^{-1} = I_p - A(I_q + BA)^{-1}B, \quad (4)$$

где $A - p \times q$, $B - q \times p$ – матрицы.

Полагая в (4) $A = \frac{\alpha}{1 - \alpha} M^{-1}(\epsilon_s) f(x_{s+1})$ и $B = f^T(x_{s+1})$, получим

$$\begin{aligned} &\left[I_m + \frac{\alpha M^{-1}(\epsilon_s) f(x_{s+1}) f^T(x_{s+1})}{1 - \alpha} \right]^{-1} = \\ &= I_m - \frac{\alpha}{1 - \alpha} M^{-1}(\epsilon_s) f(x_{s+1}) \left[1 + f^T(x_{s+1}) \frac{\alpha}{1 - \alpha} M^{-1}(\epsilon_s) f(x_{s+1}) \right]^{-1} f^T(x_{s+1}). \end{aligned}$$

И окончательно получим

$$M^{-1}(\epsilon_{s+1}) = (1 - \alpha)^{-1} \left[I_m - \frac{M^{-1}(\epsilon_s) f(x_{s+1}) f^T(x_{s+1})}{1 - \alpha + \alpha d(x_{s+1}, \epsilon_s)} \right] M^{-1}(\epsilon_s).$$

Учитывая, что $d(x, x', \epsilon) = f^T(x) M^{-1}(\epsilon) f(x') = \text{tr} M^{-1}(\epsilon) f(x') f^T(x)$, выпишем критерий Q -оптимальности:

$$\begin{aligned} Q(\epsilon_{s+1}) &= (1 - \alpha)^{-1} Q(\epsilon_s) - \sum_{x'_j \in X_{\#}} (1 - \alpha)^{-1} \text{tr} \frac{\alpha D(\epsilon_s) f(x_{s+1}) f^T(x_{s+1}) D(\epsilon_s)}{1 - \alpha + \alpha \cdot d(x_{s+1}, \epsilon_s)} M(x'_j) = \\ &= (1 - \alpha)^{-1} Q(\epsilon_s) - \sum_{x'_j \in X_{\#}} (1 - \alpha)^{-1} \frac{\alpha d^2(x'_j, x_{s+1}, \epsilon_s)}{1 - \alpha + \alpha d(x_{s+1}, \epsilon_s)}, \end{aligned}$$

где $D(\epsilon) = M^{-1}(\epsilon)$.

При добавлении точки в план вес $\alpha = 1/(s+1)$, и изменение критерия Q -оптимальности окончательно можно выразить как

$$Q(\varepsilon_{s+1}) = \frac{s}{s+1} \left[Q(\varepsilon_{s+1}) - \sum_{x'_j \in X_{\#}} \frac{d^2(x'_j, x_{s+1}, \varepsilon_s)}{s + d(x_{s+1}, \varepsilon_s)} \right]. \quad (5)$$

Чтобы достичь последовательного увеличения плана до необходимого объема, мы добавляем дополнительную точку в текущий план на каждом этапе в соответствии с решением

$$x_{s+1} = \operatorname{Arg} \max_{x \in \tilde{X}} \frac{\varphi(x, \varepsilon_s)}{s + d(x, \varepsilon_s)}, \quad (6)$$

где $\varphi(x, \varepsilon_s) = \sum_{x'_j \in X_{\#}} d^2(x, x'_j, \varepsilon_s)$.

На начальных шагах при $s < m$ информационная матрица будет неполного ранга. Применим известный прием, связанный с регуляризацией:

$$\tilde{M}(\varepsilon_s) = M(\varepsilon_s) + \gamma I_m,$$

где γ – малый положительный параметр регуляризации. На первом шаге в качестве $\tilde{M}(\varepsilon_s)$ будем рассматривать $\tilde{M}(\varepsilon_s) = \gamma I_m$. Зафиксируем $\gamma = 1$. Необходимо найти выражение, обратное $(I + f(x_{s+1})f^T(x_{s+1}))$. Применим формулу Вудбери [12–14]:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}.$$

Для нашего случая, когда матрица A – единичная размерности $m \times m$, а C – единичная размерности 1×1 , имеем

$$(I_m + f(x_{s+1})f^T(x_{s+1}))^{-1} = I_m - f(x_{s+1}) \left(1 + f^T(x_{s+1})f(x_{s+1}) \right)^{-1} f^T(x_{s+1}).$$

Вычисляем $Q(\varepsilon_1)$:

$$Q(\varepsilon_1) = \sum_{x'_j \in X_{\#}} (f^T(x'_j)f(x'_j)) - \sum_{x'_j \in X_{\#}} \frac{(f^T(x'_j)f(x_1))^2}{(1 + f^T(x_1)f(x_1))}.$$

Следовательно, выбирать первую точку необходимо, решая задачу

$$x_1 = \operatorname{Arg} \max_{x \in \tilde{X}} \sum_{x'_j \in X_{\#}} \frac{(f^T(x'_j)f(x))^2}{1 + f^T(x)f(x)}.$$

Необходимость регуляризации информационной матрицы на первых шагах отпадает, если работать с информационной матрицей в виде матрицы Грамма [11].

2.2. АЛГОРИТМ ДОБАВЛЕНИЯ-УДАЛЕНИЯ ТОЧЕК СИНТЕЗА ДИСКРЕТНЫХ ПРИБЛИЖЕННО Q -ОПТИМАЛЬНЫХ ПЛАНОВ (АЛГОРИТМ ADD-DEL- Q)

Основной принцип алгоритма заключается в выполнении повторного процесса добавления точки к плану и удаления точки из плана. Впервые он был предложен в работе [15]. Нами в работе [7] данная идея была использована для построения дискретных A -оптимальных планов с заданным числом наблюдений N . Адаптируем данный алгоритм для построения дискретных приближенно Q -оптимальных планов N .

Шаг добавления точки в план. Начальный невырожденный план ϵ_N . При добавлении $(N+1)$ -й точки с весом $\alpha = 1/(N+1)$ критерий Q -оптимальности в соответствии с (5) изменится:

$$Q(\epsilon_{N+1}) = \frac{N}{N+1} \left[Q(\epsilon_N) - \sum_{x'_j \in X_{\#}} \frac{d^2(x'_j, x_{N+1}, \epsilon_N)}{N + d(x_{N+1}, \epsilon_N)} \right].$$

Максимальное уменьшение Q -критерия при x_{N+1} будет соответствовать

$$x_{N+1} = \text{Arg max}_{x \in X} \frac{\varphi(x, \epsilon_N)}{N + d(x, \epsilon_N)},$$

где $\varphi(x, \epsilon_N)$ взято из формулы (6). После добавления точки получаем план ϵ_{N+1} .

Шаг исключения точки из плана. Если из плана ϵ_{N+1} исключить некоторую точку x_j^- , то Q -критерий изменится следующим образом:

$$Q(\tilde{\epsilon}_N) = \frac{N}{N+1} \left[Q(\epsilon_{N+1}) + \sum_{x'_j \in X_{\#}} \frac{d^2(x'_j, x_j^-, \epsilon_{N+1})}{N+1 - d(x_j^-, \epsilon_{N+1})} \right],$$

где $\tilde{\epsilon}_N$ – план из N точек, полученный из ϵ_{N+1} .

Минимальное увеличение Q -критерия будет достигаться, если x_j^- будет соответствовать

$$x_j^- = \text{Arg min}_{x \in X} \frac{\varphi(x, \epsilon_{N+1})}{N+1 - d(x, \epsilon_{N+1})}.$$

По результату шагов по добавлению и исключению точки из плана получим, что $Q(\tilde{\epsilon}_N) \leq Q(\epsilon_N)$. Равенство будет, например, тогда, когда исключается та точка, которая была включена на предыдущем шаге.

Этот алгоритм можно эффективно задействовать в ранее рассмотренном алгоритме Seq-Add- Q . Идея состоит в том, чтобы использовать алгоритм Add-Del- Q на первом этапе для построения насыщенного дискретного плана с числом наблюдений $N = m$. Видим, что в этом случае нам не приходится прибегать к процедуре регуляризации информационной матрицы.

3. ИССЛЕДОВАНИЯ

В вычислительном эксперименте в качестве области планирования будем рассматривать квадрат $[-1, +1] \times [-1, +1]$. Полином третьей степени от двух факторов будет регрессионной моделью. Построение приближенно Q -оптимальных планов проводилось с использованием в качестве функционала средней дисперсии оценки отклика, вычисленной на квадрате $[-1, +1] \times [-1, +1]$ по равномерной сетке $X_{\#}$ с числом узлов в ней $N_{\#} = 16 \times 16 = 256$. Сравнивать работу алгоритмов будем со случайно сгенерированным планом. На рис. 3 приведены графики изменения функционала Q при увеличении числа наблюдений. Видим, что случайно выбранный план проигрывает планам, построенным по предложенным алгоритмам, примерно в два раза, т. е. требуется в два раза больше числа наблюдений для достижения необходимой величины средней дисперсии оценки функции отклика. Например, если использовать приближенно Q -оптимальный план из 20 наблюдений, то для достижения такой же эффективности потребуется случайный план из 40 наблюдений.

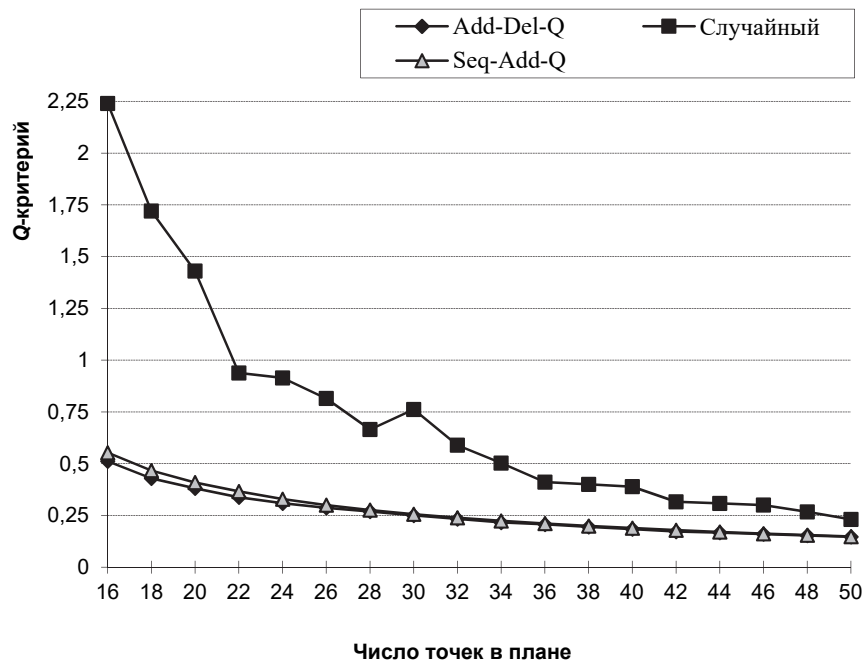


Рис. 3. Значения функционала Q -оптимальности для кубической модели от двух факторов в зависимости от числа наблюдений

Fig. 3. The values of the Q -optimality functional for the cubic model from two factors depending on the number of observations

При рассмотрении полиномиальной регрессии на отрезке мы отмечали тот момент, что применение критериев Q - и A -оптимальности давало близкие результаты с точки зрения достижения различных показателей эффективности планов. На рис. 4 представлены графики изменения величины следа ненорми-

рованной дисперсионной матрицы от числа точек в планах, построенных алгоритмами Add-Del-Q и Add-Del-A.

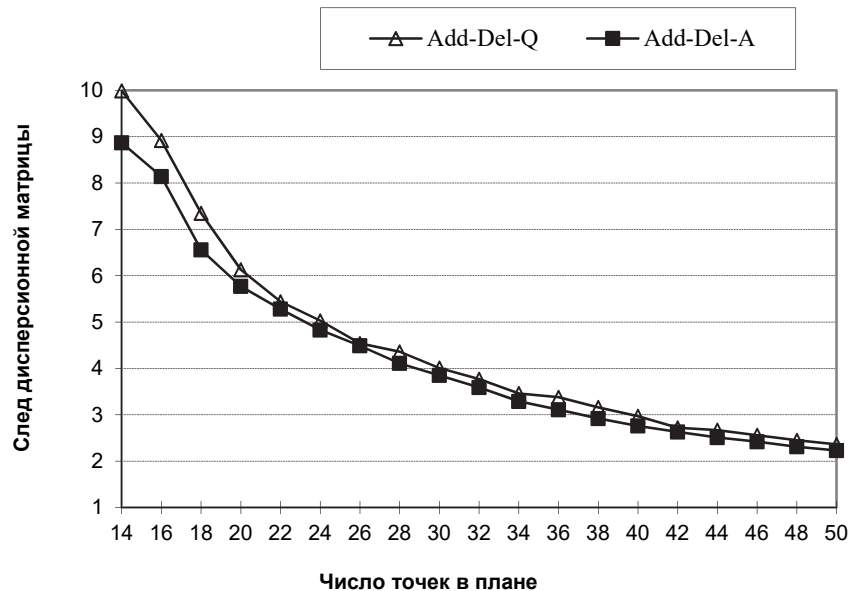


Рис. 4. След дисперсионной матрицы в зависимости от числа наблюдений

Fig. 4. Trace of the dispersion matrix depending on the number of observations

Видим, что использование критерия Q -оптимальности позволяет получать планы, близкие к A -оптимальным. Ранее мы отмечали, что для приближенно Q -оптимальных планов функция $d(x, \epsilon)$ внутри области планирования меняется более плавно, нежели для D -оптимальных, что важно, когда построенная модель используется для поиска оптимальных режимов моделируемого процесса. При этом для конкретной прикладной задачи область $X_{\#}$ может подбираться индивидуально. Отмеченные моменты говорят в пользу того, что при активной идентификации многофакторных регрессионных моделей критерию Q -оптимальности следует отводить, возможно, больше внимания, чем это было ранее.

ЗАКЛЮЧЕНИЕ

В работе предложены построенные численно приближенно Q -оптимальные планы для регрессии на отрезке. Отмечается, что использование Q -критерия позволяет получать планы, близкие по характеристикам к A -оптимальным. Для построения дискретных Q -оптимальных планов предлагаются два алгоритма. Эффективность предлагаемых алгоритмов продемонстрирована на примере построения приближенно Q -оптимальных планов для двухфакторного полинома третьей степени. Для достижения необходимой величины средней дисперсии оценки функции отклика оптимальные планы требуют существенно меньшего числа экспериментов по сравнению со случайными планами.

СПИСОК ЛИТЕРАТУРЫ

1. *Нахимов В.В., Голикова Т.И.* Логические основания планирования эксперимента. – М.: Металлургия, 1981. – 151 с.
2. *Голикова Т.И., Панченко Л.А., Фридман М.З.* Каталог планов второго порядка. Ч. 2. – М.: Изд-во МГУ, 1974. – 384 с.
3. *Голикова Т.И., Панченко Л.А.* Систематизация планов для оценки полиномиальных моделей второго порядка // Планирование оптимальных экспериментов. – М.: Изд-во МГУ, 1975. – С. 106–149.
4. *Дубова И.С., Федоров В.В.* Таблицы оптимальных планов II (Насыщенные D -оптимальные планы на кубе). – М.: Изд-во МГУ, 1972. – 39 с. – (Межфакультетская лаборатория статистических методов; вып. 40).
5. *Федоров В.В.* Теория оптимального планирования эксперимента. – М.: Наука, 1971. – 312 с.
6. *Григорьев Ю.Д.* Q -оптимальные и близкие к ним планы эксперимента для полиномиальной регрессии на отрезке // Заводская лаборатория. Диагностика материалов. – 2020. – Т. 86, № 5. – С. 65–72. – DOI: 10.26896/1028-6861-2020-86-5-65-72.
7. *Попов А.А.* Алгоритмы построения дискретных A -оптимальных планов эксперимента при активной идентификации регрессионных моделей многофакторных систем // Системы анализа и обработки данных. – 2022. – № 2 (86). – С. 39–54. – DOI: 10.17212/2782-2001-2022-2-39-54.
8. *Денисов В.И., Попов А.А.* A -, E -оптимальные и ортогональные планы регрессионных экспериментов для полиномиальных моделей / Научный совет по комплексной проблеме «Кибернетика». – Препринт. – М., 1976. – 44 с.
9. *Попов А.А.* Оптимальное планирование эксперимента в задачах структурной и параметрической идентификации моделей многофакторных систем: монография. – Новосибирск: Изд-во НГТУ, 2013. – 296 с.
10. *Попов А.А.* Последовательные схемы построения оптимальных планов эксперимента // Сборник научных трудов НГТУ. – 1995. – Вып. 1. – С. 39–44.
11. *Попов А.А.* Последовательные схемы синтеза оптимальных планов эксперимента // Доклады Академии наук высшей школы Российской Федерации. – 2008. – № 1 (10). – С. 45–55.
12. *Беллман Р.* Введение в теорию матриц. – М.: Наука, 1969. – 367 с.
13. *Pao C.P.* Линейные статистические методы и их применения. – М.: Наука, 1968. – 548 с.
14. *Searle S.R.* Matrix algebra useful for statistics. – 2nd ed. – Wiley, 2017. – 479 p.
15. *Mitchell T.J.* An algorithm for the construction of D -optimal experimental designs // Technometrics. – 1974. – Vol. 16 (2). – P. 203–210.

Попов Александр Александрович, доктор технических наук, профессор кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление научных исследований – статистические методы анализа данных, оптимальное планирование экспериментов, методы машинного обучения. Имеет более 250 публикаций, в том числе 3 монографии. E-mail: a.popov@corp.nstu.ru

Popov Alexander Alexandrovich, D.Sc. (Eng.), Professor, Department of Theoretical and Applied Computer Science, Novosibirsk State Technical University. The main area of his research is statistical methods of data analysis, optimal design of experiments, methods of machine learning. He has more than 250 publications, including 3 monographs. E-mail: a.popov@corp.nstu.ru

Algorithms for constructing discrete approximate Q -optimal experimental designs with active identification of regression models of multifactor systems*

A.A. POPOV

Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

a.popov@corp.nstu.ru

Abstract

In the theory of optimal experiment, there is a group of optimality criteria, for example, such as D -, A -, E -criteria, reflecting the accuracy of estimating model parameters. There is also a group of criteria related to the accuracy of the forecast based on the model, which can be characterized by the variance of estimates of mathematical expectations of responses. For example, using the G -optimality criterion allows you to obtain designs on which the constructed models will minimize the maximum variance of the forecast. Among these is the Q -optimality criterion, which assumes minimizing the average variance of the forecast for the regression model over the planning area.

Most of the theoretical and applied research is related to the use of the D -optimality criterion. This is also explained by the fact that the criteria of D - and G -optimality are interconnected. At the same time, it should be noted that minimizing the maximum variance in the general case may not lead to a decrease in the average variance of the forecast area. In this regard, the use of Q -optimal designs in practical regression modeling tasks is relevant. For the widespread introduction into practice of the active identification of regression models of the concept of Q -optimality of experimental designs, an arsenal of effective algorithms for their construction is needed.

The paper proposes and describes algorithms for constructing discrete approximate Q -optimal designs. The proposed algorithms are based on the developed approach of consistently increasing the number of points in the designs, as well as procedures for replacing points in the design. The designs obtained by such algorithms are recommended for use in practice when, on average, good prediction accuracy is required for the model over the entire range of input factors.

Keywords: regression model, experimental design, optimality criteria, model forecast variance, Q -optimal designs, discrete designs, addition algorithm, addition-removal algorithm

REFERENCES

1. Nalimov V.V., Golikova T.I. *Logicheskie osnovaniya planirovaniya eksperimenta* [Logical foundations for design an experiment]. Moscow, Metallurgiya Publ., 1981. 151 p.
2. Golikova T.I., Panchenko L.A., Fridman M.Z. *Katalog planov vtorogo poryadka*. Ch. 2 [Catalog of plans of the second order. Pt. 2]. Moscow, MSU Publ., 1974. 384 p.
3. Golikova T.I., Panchenko L.A. Sistematizatsiya planov dlya otsenki polinomial'nykh modelei vtorogo poryadka [Systematization of designs for evaluating second order polynomial models]. *Planirovanie optimal'nykh eksperimentov* [Design of optimal experiments]. Moscow, MSU Publ., 1975, pp. 106–149.
4. Dubova I.S., Fedorov V.V. *Tablitsy optimal'nykh planov II (Nasyshchennyye D-optimal'nye plany na kube)* [Tables of optimal designs II. (Saturated D-optimal designs on a cube)]. Moscow, MSU Publ., 1972. 39 p.
5. Fedorov V.V. *Teoriya optimal'nogo planirovaniya eksperimenta* [Theory of optimal experiment design]. Moscow, Nauka Publ., 1971. 312 p.

* Received 30 January 2024.

6. Grigoriev Yu.D. Q-optimal'nye i blizkie k nim plany eksperimenta dlya polinomial'noi regressii na otrezke [Q-optimal experimental designs and close to them experimental designs for polynomial regression on the interval]. *Zavodskaya laboratoriya. Diagnostika materialov = Industrial laboratory. Diagnostics of materials*, 2020, vol. 86 (5), pp. 65–72. DOI: 10.26896/1028-6861-2020-86-5-65-72.
7. Popov A.A. Algoritmy postroeniya diskretnykh A-optimal'nykh planov eksperimentov pri aktivnoi identifikatsii regressionnykh modelei mnogofaktornykh sistem [Algorithms for constructing discrete A-optimal experiment designs in active identification of regression models of multifactor systems]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2022, no. 2 (86), pp. 39–54. DOI: 10.17212/2782-2001-2022-2-39-54.
8. Denisov V.I., Popov A.A. A-, E-optimal'nye i ortogonal'nye plany regressionnykh eksperimentov dlya polinomial'nykh modelei [A-, E-optimal and orthogonal regression experiment designs for polynomial models]. Scientific Council for the Complex Problem "Cybernetics". Moscow, 1976. 44 p.
9. Popov A.A. Optimal'noe planirovanie eksperimenta v zadachakh strukturnoi i parametricheskoi identifikatsii modelei mnogofaktornykh sistem [The optimal experiment planning in problems of structural and parametric identification of multifactor systems models]. Novosibirsk, NSTU Publ., 2013. 296 p.
10. Popov A.A. Posledovatel'nye skhemy postroeniya optimal'nykh planov eksperimenta [The sequential schemes constructing of the optimal experiment designs]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Transaction of scientific papers of the Novosibirsk state technical university*, 1995, iss. 1, pp. 39–44.
11. Popov A.A. Posledovatel'nye skhemy sinteza optimal'nykh planov eksperimenta [Sequential schemes of synthesis of optimum plans of experiment]. *Doklady Akademii nauk vysshei shkoly Rossiiskoi Federatsii = Proceedings of the Russian higher school Academy of sciences*, 2008, no. 1 (10), pp. 45–55.
12. Bellman R. *Vvedenie v teoriyu matrits* [Introduction to matrix analysis]. Moscow, Nauka Publ., 1969. 367 p. (In Russian).
13. Rao C.R. Lineinye statisticheskie metody i ikh primeneniya [Linear statistical inference and its applications]. Moscow, Nauka Publ., 1968. 548 p. (In Russian).
14. Searle S.R. *Matrix algebra useful for statistics*. 2nd ed. Wiley, 2017. 479 p.
15. Mitchell T.J. An algorithm for the construction of D-optimal experimental designs. *Technometrics*, 1974, vol. 16 (2), pp. 203–210.

Для цитирования:

Попов А.А. Алгоритмы построения дискретных приближенно Q-оптимальных планов эксперимента при активной идентификации регрессионных моделей многофакторных систем // Системы анализа и обработки данных. – 2024. – № 2 (94). – С. 55–68. – DOI: 10.17212/2782-2001-2024-2-55-68.

For citation:

Popov A.A. Algoritmy postroeniya diskretnykh priblizhenno Q-optimal'nykh planov eksperimenta pri aktivnoi identifikatsii regressionnykh modelei mnogofaktornykh sistem [Algorithms for constructing discrete approximate Q-optimal experimental designs with active identification of regression models of multifactor systems]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2024, no. 2 (94), pp. 55–68. DOI: 10.17212/2782-2001-2024-2-55-68.