УДК 515.127:519.242

С. Г. Радченко

ИНВАРИАНТНО-ГРУППОВОЙ ПОДХОД В ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Впервые предложен и разработан метод топологического отображения плана эксперимента из хорошо обусловленного факторного пространства в плохо обусловленное, имеющее произвольную форму с определенными условиями ограничения. На основе понятий теории групп и инвариантности показано сохранение статистических (информационных) свойств плана эксперимента в факторном пространстве произвольной формы в собственной кодированной системе координат.

Ключевые слова: регрессионный анализ, некорректно поставленные задачи, устойчивое оценивание, планирование эксперимента, топологическое отображение.

Получение многофакторных статистических моделей с использованием регрессионного анализа в общем случае связано с решением некорректно поставленных задач. Для того чтобы построить статистические модели с наилучшими критериями качества, необходимо применение методов теории планирования эксперимента [1; 2].

В основу классической концепции теории планирования эксперимента положена декартова ортогональная система координат и форма факторного пространства в виде многомерного прямоугольного параллелепипеда. В его вершинах сочетаются минимальные $X_{i\min}$ и максимальные $X_{i\max}$ уровни варьирования факторов X_i при их задании на отрезках $X_{i\min} \le X_i \le X_{i\max}$. Однако в сложных технических и технологических системах такие сочетания не всегда возможны. Реальные формы факторного пространства нередко не соответствуют стандартным, например кубу, сфере или симплексу, при этом различные факторы X_i, X_j $(1 \le i < j \le k)$ будут мультиколлинеарны (сопряжены) в многофакторной статистической модели как близкие к линейной зависимости [3]. В этом случае задача становится некорректно поставленной, а основным методом ее решения будет регуляризация [4–6].

Многие исследователи обращают внимание на проблему мультиколлинеарности факторов (эффектов) в регрессионном анализе.

Так, профессор Дж. Райс приходит к следующему выводу: «Если задача плохо обусловлена, то никакие усилия, потраченные на организацию изощренных вычислений, не могут дать правильных ответов, исключая случайности» [7].

По мнению Э. Фёрстера и Б. Ренца, с возрастанием коррелированности факторов вычисленные коэффициенты уравнения регрессии теряют прикладной смысл в попытке их интерпретации, так как их среднеквадратичные ошибки существенно возрастают, а сами коэффициенты становятся очень чувствительными к выборочным наблюдениям, и тогда даже незначительное изменение исходных данных «может привести к очень сильным сдвигам в значениях оценок» [8].

Если решение прикладных задач отягощено исходной мультиколлинеарностью факторов, то использование известных методов планирования эксперимента, дающих наилучшие критерии качества определяемых моделей, невозможно. Необходима разработка метода и алгоритмов устойчивого оценивания коэффициентов статистических моделей в условиях мультиколлинеарности факторов.

В данной статье впервые предложены и исследованы методы получения устойчивых структур моделей и устойчивых коэффициентов при исходных некорректных условиях, которые используют топологическое отображение (гомеоморфизм) хорошо обусловленного факторного пространства – прообраза – в плохо обусловленное факторное пространство – образ [9].

Разработаны следующие методы отображения прообраза в образ [10]:

– получения функций отображения прообраза в образ с линейными (рис. 1) и криволинейными (рис. 2, 3) ограничениями факторного пространства в образе. Отображение проводится с использованием алгоритмов RASTA4, RASTA5.1 для линейного ограничения образа [9] и алгоритма RASTA4K для криволинейного ограничения [9];

– установления собственной кодированной системы координат в области образа (рис. 4, 5), для чего применяется алгоритм RASTA10 [9].

На рис. 1…3 приведены коэффициенты парной корреляции r_{ij} для натуральных значений факторов $X_{i\text{о}}$ в области образа и для собственной кодированной системы координат $x_{i\text{о}}$ образа, а также для факторов прообраза в натуральной $X_{i\text{пр}}$ и собственной кодированной $x_{i\text{пр}}$ системах координат; на рис. 4 натуральная $\tilde{x}_{i\text{пр}}$ и кодированная $x_{i\text{пр}}$ системы координат эквивалентны друг другу; на рис. 5 $\tilde{x}_{i\text{о}}$ является натуральной системой координат образа, а кодированная система координат показана линиями внутри фигуры образа.

Области образа и прообраза факторного пространства с матрицами натуральных и кодированных значений факторов для отображения области прообраза (фигуры $Ф_{\text{пр}}$) в область образа (фигуру $Ф_{\text{о}}$) (см. рис. 1) приведены в таблице. Функции отображения прообраза $f_{\text{отоб}}$ в образ факторного пространства следующие:

$$\begin{cases} X_{1\text{о}} = 5,2 + 1,9x_{1\text{пр}} - 2,6x_{2\text{пр}} - 1,3x_{1\text{пр}}x_{2\text{пр}}, \\ X_{2\text{о}} = 5,35 + 1,05x_{1\text{пр}} + 2,45x_{2\text{пр}} - 0,85x_{1\text{пр}}x_{2\text{пр}}. \end{cases} \quad (1)$$

Однозначность обратного отображения $f^{-1}_{\text{отоб}}$ для функций $f_{\text{отоб}}$ была установлена с использованием якобиана отображения. В области прообраза якобиан отличен от нуля и, следовательно, обратное отображение является однозначным. Для этого отображения используются алгоритмы RASTA4 и RASTA5.1.

Представив область образа факторного пространства в кодированных значениях факторов (см. таблицу), мож-

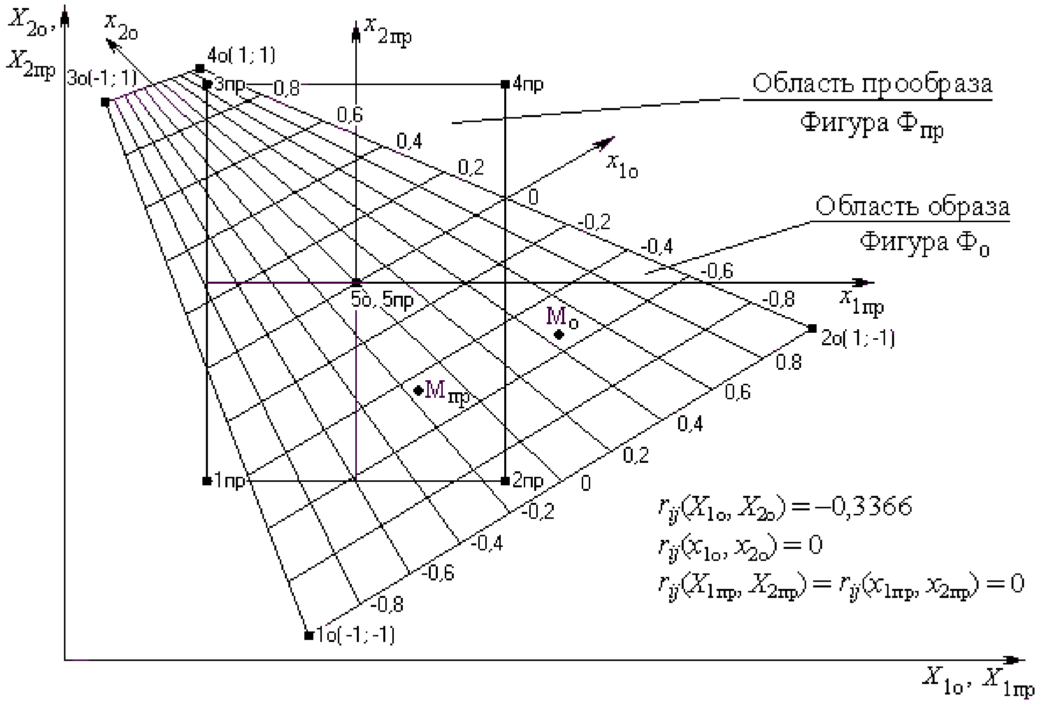

Рис. 1. Системы натуральных и собственных кодированных координат областей образа и прообраза при линейном ограничении образа ($k = 2$)

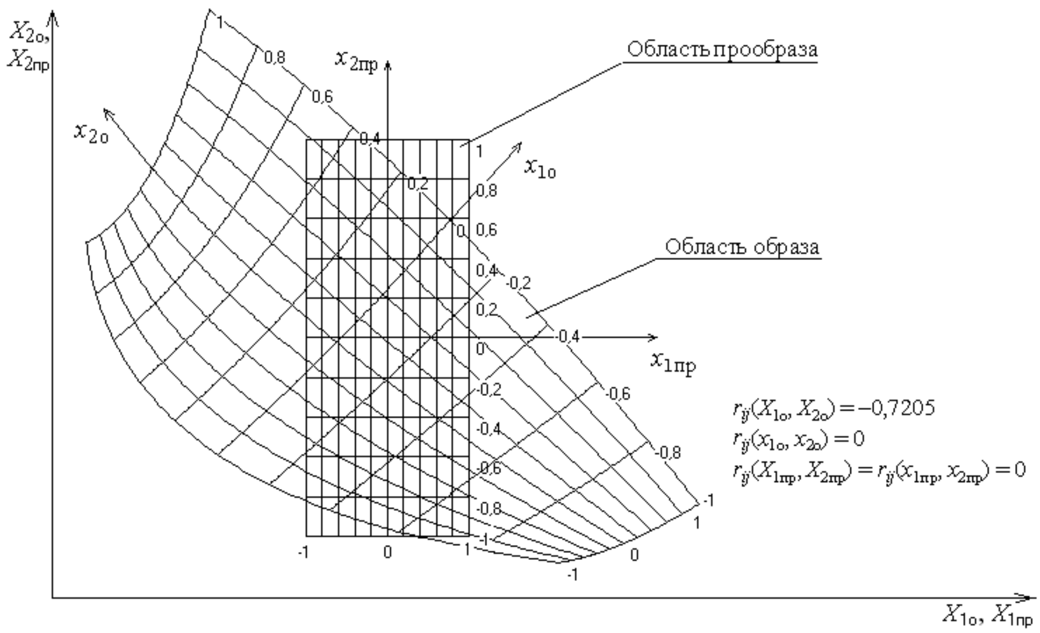

Рис. 2. Системы натуральных и собственных кодированных координат областей образа и прообраза при криволинейном ограничении образа ($k = 2$)

Рабочие матрицы и планы экспериментов областей образа и прообраза факторного пространства

Области факторного пространства									
прообраза					образа				
Номер точки	в натуральных значениях факторов		в кодированных значениях факторов		Номер точки	в натуральных значениях факторов		в кодированных значениях факторов	
	$X_{1пр}$	$X_{2пр}$	$x_{1пр}$	$x_{2пр}$		$X_{1о}$	$X_{2о}$	$x_{1о}$	$x_{2о}$
$1_{пр}$	3,3	2,9	–1	–1	$1_о$	4,6	1,0	–1	–1
$2_{пр}$	7,1	2,9	1	–1	$2_о$	11,0	4,8	1	–1
$3_{пр}$	3,3	7,8	–1	1	$3_о$	2,0	7,6	–1	1
$4_{пр}$	7,1	7,8	1	1	$4_о$	3,2	8,0	1	1
$5_{пр}$	5,2	5,35	0	0	$5_о$	5,2	5,35	0	0

но получить в численном виде функции отображения, связывающие образ с прообразом факторного пространства в натуральных значениях факторов $X_{i\text{пр}}$:

$$\begin{cases} X_{1\text{пр}} = 5,2 + 1,9x_{1\text{о}}, \\ X_{2\text{пр}} = 5,35 + 2,45x_{2\text{о}}. \end{cases} \quad (2)$$

Функции отображения для прообраза факторного пространства в кодированных значениях факторов в численном виде для точки M_u будут следующими:

$$\begin{cases} x_{1u\text{пр}} = x_{1u\text{о}}, \\ x_{2u\text{пр}} = x_{2u\text{о}}. \end{cases} \quad (3)$$

Как в прямом (1), так и в обратном (2), (3) отображениях наблюдается взаимно однозначное соответствие между отображаемыми топологическими пространствами прообраза и образа. Кроме того, в функциях отображения (1)…(3) наглядно представлено взаимно однозначное соответствие между прообразом и образом и наоборот.

Ограничительные линии $1_\text{о}2_\text{о}$, $3_\text{о}4_\text{о}$, $1_\text{о}3_\text{о}$, $2_\text{о}4_\text{о}$ в области образа (см. рис. 1) топологически эквивалентны ограничительным линиям $1_\text{пр}2_\text{пр}$, $3_\text{пр}4_\text{пр}$, $1_\text{пр}3_\text{пр}$, $2_\text{пр}4_\text{пр}$ в области прообраза факторного пространства.

В основу изучения общих свойств геометрических преобразований в виде топологического отображения прообраза факторного пространства в образ независимо от вида ограничения факторного пространства и метода отображения следует положить понятия группы преобразований и их инвариантов.

Группа преобразований – это геометрическое преобразование фигур, удовлетворяющее следующим условиям:

– каждая фигура $\Phi_{(\cdot)}$ равна сама себе (свойство рефлексивности). Фигура $\Phi_{(\cdot)}$ представляется как фигура $\Phi_\text{о}$ или $\Phi_\text{пр}$, и каждая из них равна сама себе;

– если фигура $\Phi_\text{пр}$ равна фигуре $\Phi_\text{о}$ (преобразование П), то $\Phi_\text{о}$ равна $\Phi_\text{пр}$ (преобразование П^{-1}) (свойство симметричности). Фигура $\Phi_\text{пр}$ получается из заданной фигуры $\Phi_\text{о}$ по алгоритмам RASTA4 или RASTA4K. Посредством системы функций отображения (1) фигура $\Phi_\text{пр}$ отображается в фигуру $\Phi_\text{о}$. Используя собственную кодированную систему координат для образа (фигуры $\Phi_\text{о}$), которую можно построить путем отображения собственной кодированной системы координат прообраза (фигуры $\Phi_\text{пр}$) в образ (аналогично таблице), по выражению (2) можно отобразить $\Phi_\text{о}$ в $\Phi_\text{пр}$;

– если фигура $\Phi_\text{пр}$ равна фигуре $\Phi'_\text{пр}$ (преобразование П_1), а $\Phi'_\text{пр}$ равна $\Phi_\text{о}$ (преобразование П_2), то и $\Phi_\text{пр}$ равна $\Phi_\text{о}$ (свойство транзитивности). Преобразование $\text{П}_2\text{П}_1$ состоит в том, что сначала производится преобразование П_1, а затем П_2.

Если эти условия выполняются, то геометрические преобразования представляют группу преобразований [11]. Ее геометрические свойства будут заключаться в сохраняющихся при преобразованиях свойствах фигур образа $\Phi_\text{о}$ и прообраза $\Phi_\text{пр}$.

Другой вариант областей образа и прообраза факторного пространства с линейным ограничением образа представлен ниже (рис. 6).

Ортогональность факторов друг к другу в прообразе при указанных преобразованиях сохраняется и в нестандартном факторном пространстве образа в собственной кодированной системе координат (см. рис. 1). Это свой-

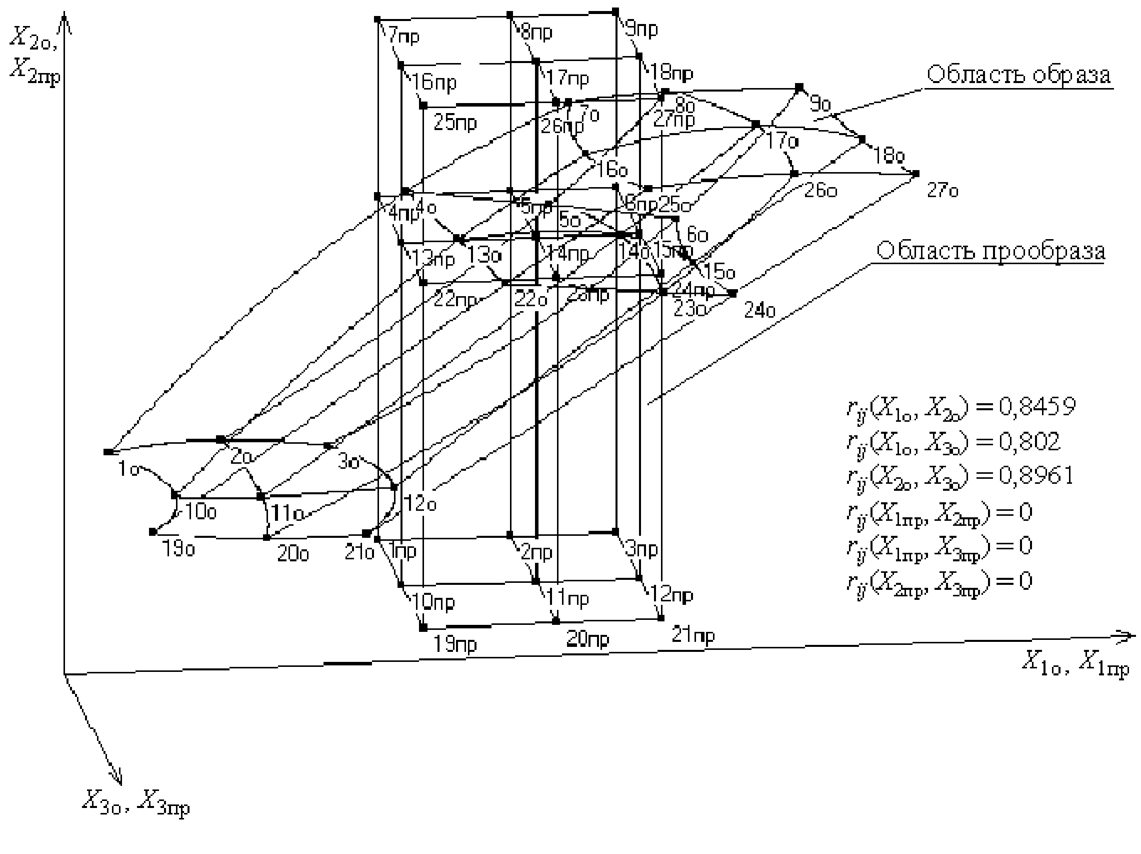

Рис. 3. Области образа и прообраза при криволинейном ограничении образа ($k = 3$)

ство ортогональности является инвариантным свойством по отношению к используемому преобразованию. Рассматриваемые преобразования фигур $\Phi_{пр}$ и $\Phi_{о}$ относятся к теории непрерывных групп преобразований.

Введем понятие инвариантов в исследование отображения прообраза в образ.

Инварианты – это числа или алгебраические выражения, связанные с определенным математическим объектом и остающиеся неизменными при преобразованиях этого объекта или системы отсчета, в которой описывается объект. Характеристика геометрической фигуры и ее свойств осуществляется с помощью чисел и системы координат (или вспомогательной системы отсчета). Величины, описывающие определенную геометрическую фигуру, характеризуют и их отношение к системе координат. Свойства фигуры не должны зависеть от системы координат, т. е. должно выполняться равенство

$$f(x_{1uпр}, \ldots, x_{kuпр}) = f(x_{1uо}, \ldots, x_{kuо}),$$

где $x_{1uпр}, \ldots, x_{kuпр}$ – координаты точек плана эксперимента в собственной кодированной системе координат для фигуры $\Phi_{пр}$ области прообраза; $x_{1uо}, \ldots, x_{kuо}$ – координаты точек плана эксперимента в собственной кодированной системе координат для фигуры $\Phi_{о}$ области образа, которая является отображением фигуры $\Phi_{пр}$ области прообраза; f – определенная функция от значений $x_{1(\cdot)}, ..., x_{k(\cdot)}$ координат точек плана эксперимента в собственной кодированной системе координат для фигур $\Phi_{(\cdot)}$.

Изучаемым свойством геометрии факторного пространства в данном исследовании является ортогональность факторов друг к другу. В этом случае функциями $f(\cdot)$ от координат точек фигур $\Phi_{пр}$ и $\Phi_{о}$ являются парные коэффициенты корреляции $r_{ij}(x_{iпр}, x_{jпр})$ и $r_{ij}(x_{iо}, x_{jо})$, 1 J $i < j \leq k$. В различных формах факторного пространства – прообраза и образа – при использовании в прообразе регулярного плана эксперимента коэффициенты парной корреляции факторов $r_{ij}(x_{i(\cdot)}, x_{j(\cdot)})$ равняются нулю и в прообразе, и в образе. Рассматриваемая геометрия факторных пространств тесно связана с теорией инвариантов. По Ф. Клейну, каждая геометрия является теорией инвариантов соответствующей группы преобразований [12].

Разработка теории непрерывных групп преобразований геометрических фигур прообраза $\Phi_{пр}$ и образа $\Phi_{о}$, представляющих собой геометрическое выражение плана эксперимента и сохраняющихся при этом их статистических характеристик – инвариантов – в собственных кодированных системах координат, приводит к созданию инвариантно-группового подхода в теории планирования эксперимента. Этот подход позволяет получать многофакторные статистические модели с наилучшими возможными характеристиками для нестандартных областей факторного пространства в именованной системе координат в условиях исходной мультиколлинеарности факторов в образе.

Таким образом, впервые разработана и исследована концепция инвариантно-группового подхода в теории планирования эксперимента. Данный подход ориентирован на использование для произвольных форм факторного пространства планов экспериментов, разработанных для областей в виде прямоугольных параллелепипедов.

Статистические свойства планов экспериментов для стандартных форм факторного пространства прообраза (фигуры $\Phi_{пр}$): ортогональность факторов и их главных эффектов друг к другу в группе преобразований в собственных кодированных системах координат прообраза и образа – сохраняются и в нестандартном факторном пространстве образа (фигуры $\Phi_{о}$).

Дальнейшее развитие инвариантно-группового подхода в теории планирования эксперимента будет связано с понятием структуры плана эксперимента и сохранением его статистических (информационных) свойств при отображениях прообраза в образ факторных пространств (URL: http://www.n-t.org/sp/lesmi/).

Библиографические ссылки

1. Налимов В. В. Теория эксперимента. М. : Наука, 1971.

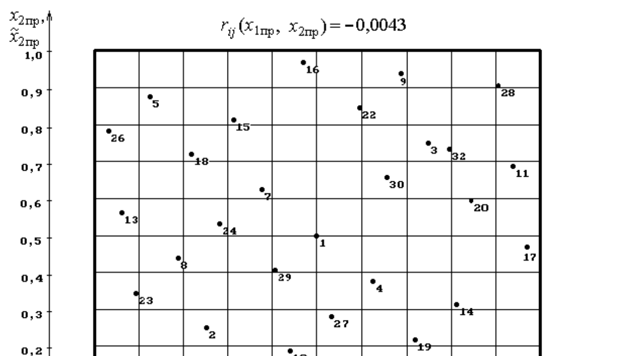

Рис. 4. Расположение в квадрате точек $ЛП_{\tau}$ равномерно распределенных последовательностей $x_{1пр} = \xi_5, x_{2пр} = \xi_7$ прообраза факторного пространства

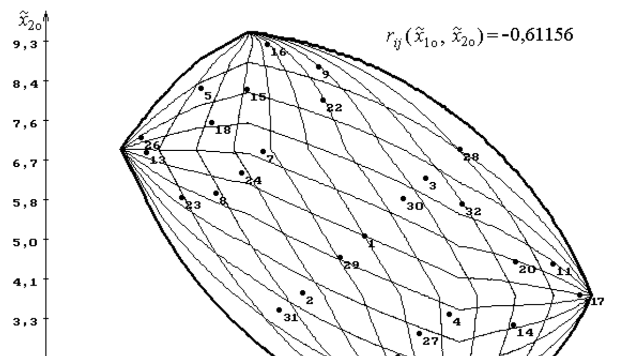

Рис. 5. Произвольная область образа факторного пространства с криволинейными ограничительными линиями и натуральной и собственной системами координат

2. Налимов В. В., Голикова Т. И. Логические основания планирования эксперимента. 2-е изд., перераб. и доп. М. : Металлургия, 1981.

3. Половинкин А. И. Основы инженерного творчества : учеб. пособие для студентов втузов. М. : Машиностроение, 1988.

4. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. 3-е изд. М. : Наука, 1986.

5. Лаврентьев М. М., Романов В. Г., Шишатский С. П. Некорректные задачи математической физики и анализа. М. : Наука, 1980.

6. Жуковский Е. Л. Статистическая регуляризация решений обратных некорректно поставленных задач обработки и интерпретации результатов эксперимента // Методы математического моделирования, автоматизация обработки наблюдений и их применения : сборник / под ред. А. Н. Тихонова, А. А. Самарского. М., 1986. С. 47–72.

7. Райс Дж. Матричные вычисления и математическое обеспечение / пер. с англ. О. Б. Арушаняна ; под ред. В. В. Воеводина. М. : Мир, 1984.

8. Фёрстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. Руководство для экономистов / пер. с нем. и предисл. В. М. Ивановой. М. : Финансы и статистика, 1983.

9. Радченко С. Г. Устойчивые методы оценивания статистических моделей : монография. Киев : Санспарель, 2005.

10. Радченко С. Г. Устойчивые методы оценивания статистических моделей // Матеріали IX-ої Міжнар. наук. конф. ім. акад. М. Кравчука. Київ, 2002. С. 451–452.

11. Радченко С. Г. Теория групп преобразований в устойчивом статистическом моделировании // Системний аналіз та інформаційні технології : матеріали VII Міжнар. наук.-техн. конф. / НТУУ «КПІ». Київ, 2005. С. 150.

12. Кованцов М. І. Проективна геометрія. 2-е вид., перероб. і доп. Київ : Вища шк, 1985.

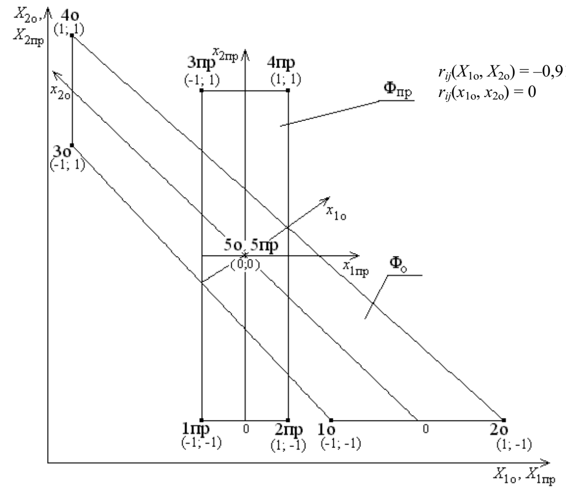

Рис. 6. Области образа и прообраза факторного пространства при линейных ограничениях образа

S. G. Radchenko

INVARIANT GROUP APPROACH IN THE EXPERIMENT DESIGN THEORY

The article for the first time presents method, developed for topological mapping of the experiment design from a well-conditioned factor space into the ill-conditioned one being of arbitrary shape with certain conditions of its restriction. Basing on such notions as the group theory and invariance the author has shown the preservation of statistical properties of the experiment design in the factor space of arbitrary form in its own proper coded coordinate system.

Keywords: regression analysis, ill-posed problems, stable estimation, experiment design, topologic mapping.