УДК 519.233.5:519.242.5

**С.Г. РАДЧЕНКО[*]**

# ТОПОЛОГИЧЕСКИЕ И СТРУКТУРНЫЕ СВОЙСТВА МНОГОФАКТОРНЫХ ПЛАНОВ ЭКСПЕРИМЕНТОВ

[*]Национальный технический университет Украины "Киевский политехнический институт", Киев, Украина

***Анотація.*** *Проаналізовано властивості багатофакторних планів експериментів при топологічному відображенні добре обумовленого факторного простору – прообразу – в погано обумовлений простір – образ. З використанням введеного поняття структури плану експерименту показано збереження інформаційних властивостей плану прообразу і в образі.*
***Ключові слова:*** *планування експерименту, регресійний аналіз, топологічне відображення, структура плану експерименту.*

***Аннотация.*** *Проанализированы свойства многофакторных планов экспериментов при топологическом отображении хорошо обусловленного факторного пространства – прообраза – в плохо обусловленное пространство – образ. С использованием введенного понятия структуры плана эксперимента показано сохранение информационных свойств плана прообраза и в образе.*
***Ключевые слова:*** *планирование эксперимента, регрессионный анализ, топологическое отображение, структура плана эксперимента.*

***Abstract.*** *The properties of multifactorial designs of experiments under topological mapping of well-conditioned factor space – pre-image – into poor-conditioned factor space – image were analyzed. The maintenance of informative properties of experiment for pre-image and image was illustrated with the usage of introduced concept of design of experiment structure.*
***Keywords:*** *design of experiment, regression analysis, topological mapping, design of experiment structure.*

## 1. Введение

*Постановка проблемы*

Статистические регрессионные модели широко используются при моделировании различных сложных систем. Качество моделей определяется выбранным планом экспериментов, системой формирования структуры модели, устойчивостью оцениваемых коэффициентов, формой факторного пространства, числом возможных экспериментов и др. условиями. Для получения моделей с хорошими свойствами необходимо использовать системное оптимальное планирование регрессионного эксперимента [1, с. 71–75].

По мнению многих специалистов, одной из основных проблем множественного регрессионного анализа является мультиколлинеарность (взаимная сопряженность) эффектов – главных и взаимодействий [2, с. 186, 108–114]. Не допустить мультиколлинеарность эффектов можно, если использовать планирование экспериментов и получать планы, в которых эффекты ортогональны друг к другу или близки к ортогональным. К таким планам экспериментов относятся полные факторные эксперименты, многофакторные регулярные планы и планы на основе $\text{ЛП}_\tau$ равномерно распределенных последовательностей [3].

При использовании планов экспериментов предполагается, что форма факторного пространства соответствует прямоугольному параллелепипеду (кубу), сфере или симплексу. В реальных прикладных задачах форма факторного пространства может не соответствовать вышеприведенным, так как между параметрами технических и технологических систем наблюдаются связи, которые близки к линейным, т.е. некоторые факторы $X_i$ и $X_j$ коррелированы между собой [3, с. 134]. Коррелированность факторов приводит к решению некорректно поставленных задач, что весьма усложняет получение статистических моделей с необходимыми хорошими критериями: ортогональность эффектов, адекватность, информативность и т.п. Соответствие плана эксперимента критериям *D*-, *A*-, *E*-, *G*-оптимальности и другим позволяет получить хорошие модели.

Другой причиной, приводящей к возникновению мультиколлинеарности, является обработка результатов неспланированных («пассивных») экспериментов или результатов наблюдений. Этот случай наиболее сложный для получения хороших моделей.

*Цель статьи*

Изложение новых возможных подходов планирования эксперимента для получения регрессионных моделей в нестандартных областях факторного пространства и условий эксперимента, в которых факторы коррелированы.

## 2. Изложение разработанных методов

В традиционной практике планирования экспериментов используется ортогональная декартова система координат. Возможности ортогонального представления эффектов можно существенно расширить, если использовать такие системы координат, в которых эффекты модели будут представлены ортогонально, а не коррелированно.

Рассмотрим два условия задания экспериментов (рис. 1, табл. 1). В первом варианте (рис. 1*а*) факторы $X_{i\text{пр}}, X_{j\text{пр}}$ ортогональны и коэффициент парной корреляции $r_{ij}\left(X_{i\text{пр}}, X_{j\text{пр}}\right)=0$. Во втором – факторы $X_{i\text{о}}, X_{j\text{о}}$ (рис. 1*б*) коррелированы, $r_{ij}\left(X_{i\text{о}}, X_{j\text{о}}\right)=0{,}4472$. Факторное пространство прообраза $R_{\text{пр}}$ ограничено прямоугольником, а образа $R_{\text{о}}$ – параллелограммом.

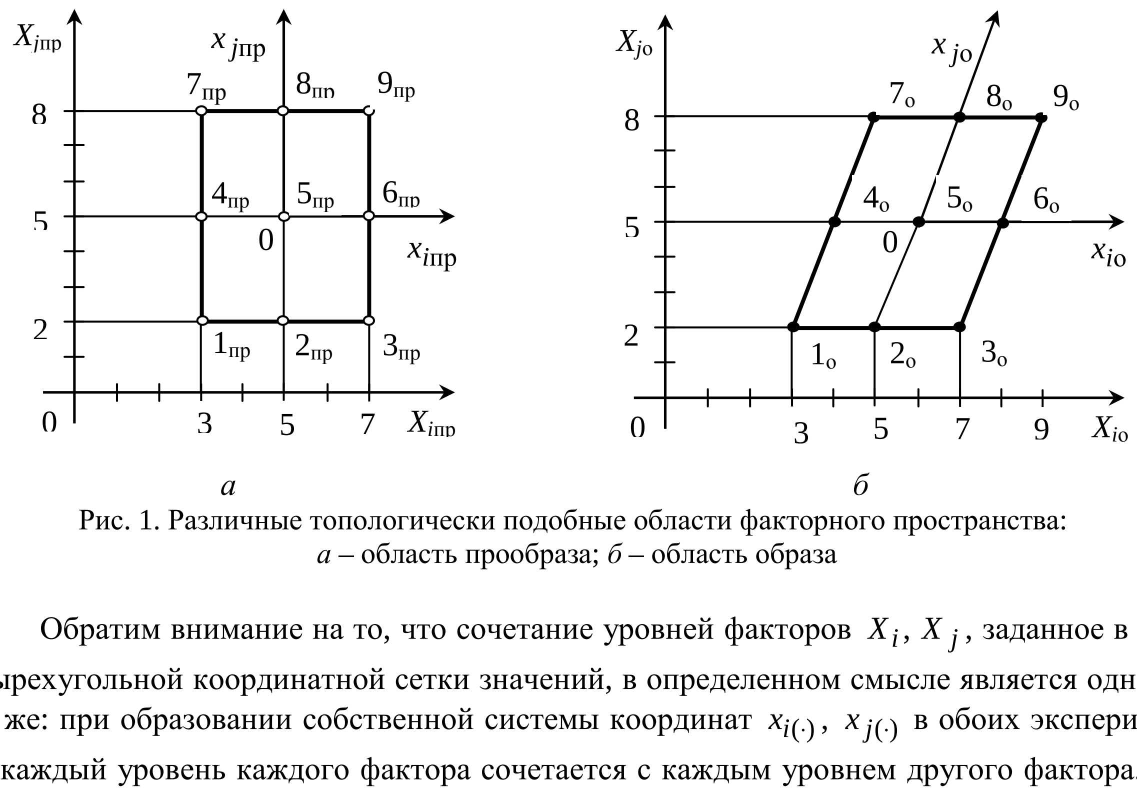


Рис. 1. Различные топологически подобные области факторного пространства:
*а* – область прообраза; *б* – область образа

Обратим внимание на то, что сочетание уровней факторов $X_i, X_j$, заданное в виде четырехугольной координатной сетки значений, в определенном смысле является одним и тем же: при образовании собственной системы координат $x_{i(\cdot)}, x_{j(\cdot)}$ в обоих экспериментах каждый уровень каждого фактора сочетается с каждым уровнем другого фактора. Координатная сетка со значениями уровней факторов $x_{i\text{пр}}, x_{j\text{пр}}$ – область прообраза для прямоугольника – топологически подобна координатной сетке значений уровней факторов $x_{i\text{о}}, x_{j\text{о}}$ – области образа для параллелограмма.

Математическая модель, полученная по данным, представленным на прямоугольнике, будет соответствовать наилучшим возможным критериям качества. Модель, полученная по данным, представленным на параллелограмме, по мере увеличения коррелированности факторов $X_{i\text{о}}, X_{j\text{о}}$ будет характеризоваться всё более худшими критериями качества, так как факторное пространство в координатах $X_{i\text{о}}, X_{j\text{о}}$ будет плохо обусловлено (рис. 1*б*).

Таблица 1. Матрицы планов экспериментов в областях прообраза и образа

| Номер точки | Значения уровней варьирования факторов в области прообраза | | Номер точки | Значения уровней варьирования факторов в области образа | |
|---|---|---|---|---|---|
| | $X_{i\text{пр}}$ | $X_{j\text{пр}}$ | | $X_{i\text{о}}$ | $X_{j\text{о}}$ |
| $1_{\text{пр}}$ | 3 | 2 | $1_{\text{о}}$ | 3 | 2 |
| $2_{\text{пр}}$ | 5 | 2 | $2_{\text{о}}$ | 5 | 2 |
| $3_{\text{пр}}$ | 7 | 2 | $3_{\text{о}}$ | 7 | 2 |
| $4_{\text{пр}}$ | 3 | 5 | $4_{\text{о}}$ | 4 | 5 |
| $5_{\text{пр}}$ | 5 | 5 | $5_{\text{о}}$ | 6 | 5 |
| $6_{\text{пр}}$ | 7 | 5 | $6_{\text{о}}$ | 8 | 5 |
| $7_{\text{пр}}$ | 3 | 8 | $7_{\text{о}}$ | 5 | 8 |
| $8_{\text{пр}}$ | 5 | 8 | $8_{\text{о}}$ | 7 | 8 |
| $9_{\text{пр}}$ | 7 | 8 | $9_{\text{о}}$ | 9 | 8 |

Для получения универсального метода устойчивого оценивания коэффициентов многофакторных моделей необходимо найти метод перехода от заданного плохо обусловленного факторного пространства образа $R_{\text{о}}$ к наилучше обусловленному факторному пространству прообраза $R_{\text{пр}}$, в котором и необходимо решать поставленную задачу.

Таким методом является метод топологического отображения хорошо обусловленного факторного пространства прообраза $R_{\text{пр}}$, в котором эксперимент можно планировать с наилучшими критериями, в плохо обусловленное факторное пространство образа $R_{\text{о}}$, в котором планировать эксперимент традиционными методами невозможно. Отображение $R_{\text{пр}}$ в $R_{\text{о}}$ при линейных ограничениях образа проводится по алгоритмам RASTA4, RASTA5.1 [3, с. 156–171]. Системы $R_{\text{пр}}$ и $R_{\text{о}}$ с заданными в них отношениями будут изоморфными, т. е. равными по виду, форме. Понятие изоморфизм включает в себя как частный случай понятие гомеоморфизма, играющее основную роль в топологии.

Гомеоморфизм – взаимно однозначное и взаимно непрерывное соответствие между двумя топологическими пространствами. Гомеоморфные пространства топологически эквивалентны.

На рис. 2 показаны топологически эквивалентные формы факторного пространства с линейными и криволинейными ограничениями. Фигуры *а, б, в* представляют систему $R_{\text{о}}$ образа, а фигура *г* представляет систему $R_{\text{пр}}$ прообраза.

Каждому объекту $X_{i\text{пр}}$ (фактор), $X_{iu\text{пр}}$ (точка) из системы $R_{\text{пр}}$ однозначно соответствует объект $X_{i\text{о}}$ (фактор), $X_{iu\text{о}}$ (точка) из системы $R_{\text{о}}$, и наоборот. Каждому отношению в системе $R_{\text{пр}}$ однозначно соответствует отношение в системе $R_{\text{о}}$, и наоборот. Если некоторые объекты $\left(X_{1\text{пр}}, X_{2\text{пр}}, ..., X_{k\text{пр}}\right)$ связаны в системе $R_{\text{пр}}$ отношением

$\hat{y}_{w\text{пр}} = f_{w\text{пр}}\left(X_{1\text{пр}}, X_{2\text{пр}}, ..., X_{k\text{пр}}\right)$, то соответствующие объекты $\left(X_{1\text{о}}, X_{2\text{о}}, ..., X_{k\text{о}}\right)$ в системе $R_{\text{о}}$ должны быть связаны отношением $\hat{y}_{w\text{о}} = f_{w\text{о}}\left(X_{1\text{о}}, X_{2\text{о}}, ..., X_{k\text{о}}\right)$, и наоборот.

Рассмотрим отношения для фигуры прообраза $\Phi_{\text{пр}}$ и фигуры образа $\Phi_{\text{о}}$, приведенных на рис. 3. План эксперимента соответствует $2^2//4$. Коррелированность факторов $X_{1\text{о}}, X_{2\text{о}}$ составляет $r_{ij}(X_{1\text{о}}, X_{2\text{о}}) = -0{,}918$. В табл. 2 приведены рабочие матрицы и планы экспериментов областей образа и прообраза факторного пространства, изображенных на рис. 3. Кодированные значения $x_{1\text{пр}}, x_{2\text{пр}}$ и $x_{1\text{о}}, x_{2\text{о}}$ есть ортогональные контрасты факторов $X_{1\text{пр}}, X_{2\text{пр}}$ и $X_{1\text{о}}, X_{2\text{о}}$ [3, с. 54–57, 162].

Функции отображения прообраза в образ $\Phi_{\text{пр}} \to \Phi_{\text{о}}$ (преобразование П), т. е. $X_{1\text{пр}}, X_{2\text{пр}} \to X_{1\text{о}}$; $X_{1\text{пр}}, X_{2\text{пр}} \to X_{2\text{о}}$, имеют вид

$$X_{1\text{о}} = 30 + 5x_{1\text{пр}} - 20x_{2\text{пр}} - 5x_{1\text{пр}}x_{2\text{пр}},$$

$$X_{2\text{о}} = 20 + 2{,}5x_{1\text{пр}} + 15x_{2\text{пр}} + 2{,}5x_{1\text{пр}}x_{2\text{пр}},$$

где $x_{1\text{пр}} = 0{,}2\left(X_{1\text{пр}} - 30\right)$, $x_{2\text{пр}} = 0{,}0666667\left(X_{2\text{пр}} - 20\right)$.

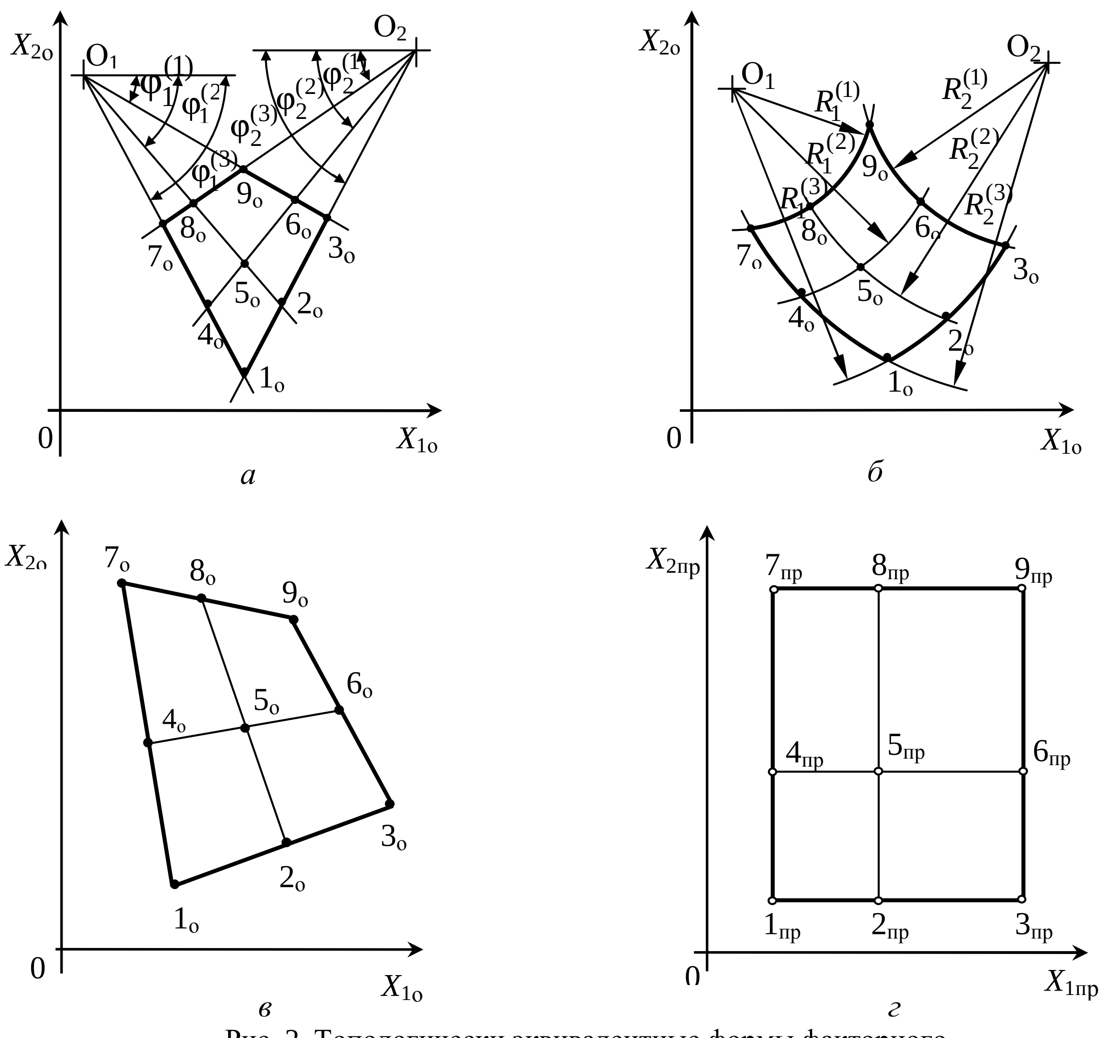


Рис. 2. Топологически эквивалентные формы факторного пространства с линейными и криволинейными ограничениями

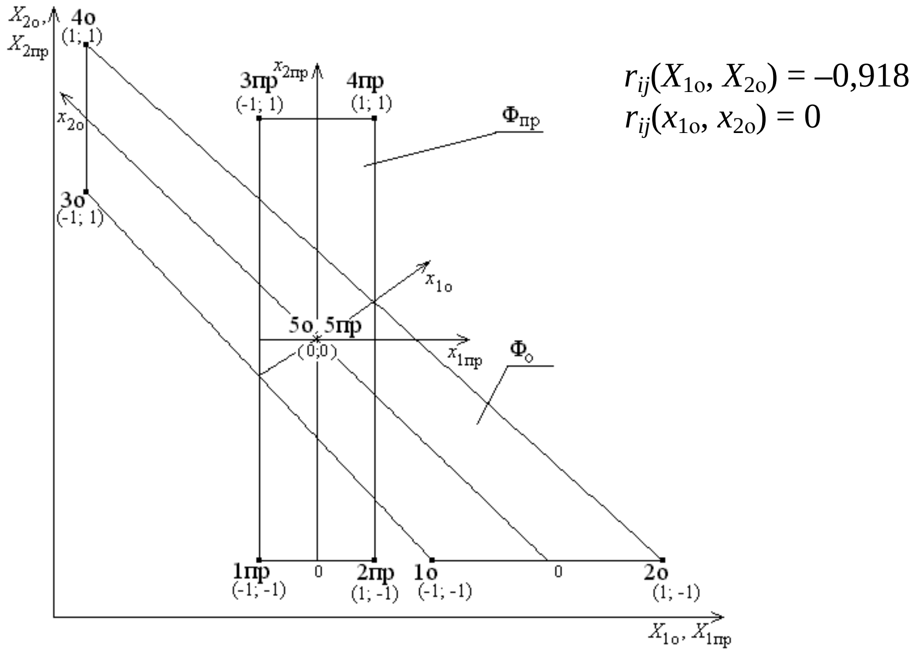


Рис. 3. Области образа и прообраза факторного пространства при линейных ограничениях образа

Таблица 2. Рабочие матрицы и планы экспериментов областей образа и прообраза факторного пространства для линейного ограничения

| Область прообраза | | | | | Область образа | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Номер точки | в натуральных значениях факторов | | в кодированных значениях факторов | | Номер точки | в натуральных значениях факторов | | в кодированных значениях факторов | |
| | $X_{1пр}$ | $X_{2пр}$ | $x_{1пр}$ | $x_{2пр}$ | | $X_{1о}$ | $X_{2о}$ | $x_{1о}$ | $x_{2о}$ |
| $1_{пр}$ | 25 | 5 | –1 | –1 | $1_о$ | 40 | 5 | –1 | –1 |
| $2_{пр}$ | 35 | 5 | 1 | –1 | $2_о$ | 60 | 5 | 1 | –1 |
| $3_{пр}$ | 25 | 35 | –1 | 1 | $3_о$ | 10 | 30 | –1 | 1 |
| $4_{пр}$ | 35 | 35 | 1 | 1 | $4_о$ | 10 | 40 | 1 | 1 |
| $5_{пр}$ | 30 | 20 | 0 | 0 | $5_о$ | 30 | 20 | 0 | 0 |

Функции для отображения образа, кодированного в прообраз $Ф_{о.код} \to Ф_{пр}$ (преобразование $П^{-1}$), т.е. $x_{1о}, x_{2о} \to X_{1пр}$; $x_{1о}, x_{2о} \to X_{2пр}$, имеют вид:

$$X_{1пр} = 30 + 5x_{1о},$$

$$X_{2пр} = 20 + 15x_{2о},$$

где $x_{1о} = x_{1о}$;

$x_{2о} = x_{2о}$.

В кодированных значениях факторов $X_{1о}$, $X_{2о}$ и $X_{1пр}$, $X_{2пр}$, т.е. в ортогональных контрастах $x_{(\cdot)пр}$, $x_{(\cdot)о}$, планы экспериментов эквивалентны.

Покажем, что информационные свойства многофакторных уравнений регрессии при отображении прообраза в образ сохраняются.

Основными классами используемых планов экспериментов являются многофакторные регулярные планы [4] и планы на основе $ЛП_\tau$ равномерно распределенных последовательностей [5].

На основе понятия «структура математическая» введем понятие «структура плана эксперимента».

Структура математическая – задание дополнительных условий (операций, отношений, топологии и т. д.) на множестве, природа элементов которого не определена.

Структура плана эксперимента – отношение между уровнями варьирования факторов, подчиняющееся определенным условиям и/или зависимостям.

В качестве дополнительных условий в структуре плана эксперимента будем рассматривать отношение между уровнями варьирования факторов. Отношение – произвольное подмножество $R_{эк}$ множества $N$ всех кортежей – упорядоченных наборов вида $s_i$ – уровней варьирования факторов, $1 \le i \le k$; $k$ – число факторов. $R_{эк}$ есть $k$-местное отношение на $N$. Понятие «отношение» в математике служит для выражения на теоретико-множественном языке связей между объектами.

Отношение в структуре плана эксперимента – подмножество $R_{эк}$ множества $N$ всех кортежей – упорядоченных $k$-местных наборов $(d_{1mi}, d_{2mi}, ..., d_{imi}, ..., d_{kmi}) \in N$, где $d_{imi}$ – уровень варьирования $i$-го фактора $mi$-го уровня, т. е. $m$-й уровень $i$-го фактора, $1 \le i \le k$, $1 \le mi \le s_i$, $k$ – число факторов, $s_i$ – число уровней $i$-го фактора, $N$ число опытов в плане эксперимента.

В каталогах планов экспериментов уровни варьирования факторов $d_{imi}$ для многофакторных регулярных планов принято обозначать как 0, 1, 2, …, $s_i - 1$, а для планов на основе $ЛП_\tau$ равномерно распределенных последовательностей – цифрами 0,5; 0,25; 0,75; 0,125; 0,625; 0,375; 0,875; … на интервале (0; 1).

Структура регулярного плана эксперимента – отношение между уровнями варьирования факторов, подчиняющееся условию пропорциональности частот для факторов: $n_{ij...l}^{pq...r} = n_i^p \, n_j^q \, ... n_l^r \, / N^{t-1}$, где $n_i^p$ – число появлений $p$-го уровня $i$-го фактора, $n_j^q$ – число появлений $q$-го уровня $j$-го фактора и $n_l^r$ – число появлений $r$-го уровня l-го фактора, где $p = 0, 1, ..., s_i - 1$, $q = 0, 1, ..., s_j - 1$, $r = 0, 1, ..., s_l - 1$, $i, j, l = 1, 2, ..., k$, $i \ne j \ne l$, $N$ – число опытов в плане эксперимента, $t$ – мощность плана эксперимента, $2 \le t \le k$, $n_{ij...l}^{pq...r}$ – число одновременных появлений в плане эксперимента указанных уровней для $t$ факторов.

Структура $ЛП_\tau$ плана эксперимента – отношение между уровнями варьирования факторов, подчиняющееся условию содержания не менее чем $2^{\tau+1}$ точек, представляющих собой $П_\tau$-сетку, в любом двоичном участке $ЛП_\tau$ последовательности. Число уровней каждого фактора $s_i$ равно числу опытов $N$ в плане эксперимента, т.е. $s_i = N$. Понятие полного факторного эксперимента для $ЛП_\tau$ планов отсутствует.

С использованием понятия «структура плана эксперимента» рассмотрим сохранение информационных свойств многофакторных уравнений регрессии при отображении прообраза в образ [6].

**Утверждение.** При отображении многофакторного плана эксперимента из прообраза в образ структура плана и его статистические (информационные) свойства сохраняются.

*Доказательство.* Пусть отображение точек декартовой системы координат из прообраза в образ взаимно однозначно и взаимно непрерывно; функции отображения прямые

$f_{\text{iотоб}}$ и обратные $f_{\text{iотоб}}^{-1}$ дифференцируемые (гладкие); якобиан J при обратном отображении в области прообраза не равен нулю. Точки плана эксперимента в прообразе принадлежат декартовой системе координат. Тогда структура плана эксперимента в образе соответствует структуре плана эксперимента в прообразе, так как собственные кодированные системы координат прообраза и образа топологически эквивалентны. План эксперимента в виде точек, координаты которых в собственной кодированной системе координат в прообразе есть уровни варьирования факторов, образует сетку уровней варьирования факторов, которая при отображении топологически эквивалентна сетке уровней варьирования факторов в образе. Такие свойства при переходе к гомеоморфным фигурам будут топологическими.

Статистические (информационные) свойства плана эксперимента (критерии *D*-, *A*-, *E*-, *G*-, *Q*-оптимальности, ортогональности) полностью определяются его структурой и, выраженные как характеристики матрицы дисперсий-ковариаций $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\sigma^2(\varepsilon)$ в образе, будут соответствовать характеристикам плана эксперимента в прообразе, так как матрицы планов экспериментов в прообразе и образе будут эквивалентны друг другу в собственных кодированных системах координат.

Если статистические характеристики плана эксперимента (совокупность критериев качества плана) в прообразе «хорошие», то в образе они будут также «хорошие» в собственных кодированных системах координат. □

Статистическое моделирование с использованием отображения прообраза факторного пространства в образ факторного пространства предполагает, что полученные с применением метода наибольшего правдоподобия (метода максимального правдоподобия) оптимальные оценки в одном пространстве будут характеризоваться такими же свойствами и для другого пространства. Единственным оценкам коэффициентов регрессии, полученным в одном пространстве, должны соответствовать единственные оценки в другом пространстве.

При выполнении приведенных условий получение статистических моделей будет корректным: решение существует, оно единственно и устойчиво.

С разработанными методами решения регрессионных задач и полученными результатами можно ознакомиться в [7, 8].

## 3. Выводы и перспективы дальнейших исследований

1. При соответствии нестандартных областей факторного пространства (образ факторного пространства) требованиям выпуклости, связности и компактности для них методом топологического отображения можно найти хорошо обусловленные факторные пространства – прообраз факторного пространства, в которых и следует решать поставленные задачи.
2. Прообраз и образ факторного пространства топологически эквивалентны (гомеоморфны), т.е. в математическом понимании одинаковы как топологические пространства. В метрическом понимании указанные пространства отличаются своими статистическими свойствами в натуральных значениях координат по критериям коррелированности факторов между собой.
3. Устойчивое оценивание коэффициентов многофакторных моделей $(1 \le \operatorname{cond}(\mathbf{X}^{\mathrm{T}}\mathbf{X}) < 10)$ достигается не использованием метода регуляризации или изменением заданных начальных условий, а путем решения задачи в наилучших из возможных условий, найденных путем топологического отображения реальных условий или использования других подходов.
4. Хорошие свойства статистических оценок коэффициентов уравнения регрессии, полученных в прообразе с использованием метода максимального правдоподобия, сохраняются и для условий, полученных путем топологического отображения прообраза в образ; единственность оценок также выполняется.

Дальнейшие исследования планов экспериментов для регрессионных моделей целесообразно проводить в разработке концепции ортогональности для условий неспланированного «пассивного» эксперимента и нестандартных областей факторного пространства.

## СПИСОК ЛИТЕРАТУРЫ

1. Радченко С.Г. Системное оптимальное планирование регрессионного эксперимента / С.Г. Радченко // Заводская лаборатория. Диагностика материалов. – 2012. – Т. 78, № 7. – С. 71 – 75.
2. Демиденко Е.З. Линейная и нелинейная регрессии / Демиденко Е.З. – М.: Финансы и статистика, 1981. – 302 с.
3. Радченко С.Г. Методология регрессионного анализа / Радченко С.Г. – К.: Корнійчук, 2011. – 376 с.
4. Бродский В.З. Введение в факторное планирование эксперимента / Бродский В.З. – М.: Наука, 1976. – 224 с.
5. Соболь И.М. Выбор оптимальных параметров в задачах со многими критериями: учеб. пособие для вузов / И.М. Соболь, Р.Б. Статников. – [2-е изд., перераб. и доп.]. – М.: Дрофа, 2006. – 175 с.
6. Радченко С.Г. Сохранение информационных свойств многофакторных планов экспериментов при отображении прообраза в образ / Радченко С.Г. // Тринадцята міжнарод. наук. конф. ім. академіка М. Кравчука: матер. конф. в т. 3, (Київ, 13–15 травня 2010 р.). – Київ: НТУУ «КПІ», 2010. – С. 97.
7. Лаборатория экспериментально-статистических методов исследований (ЛЭСМИ) [Электронный ресурс]. – Режим доступа: http://www.n-t.org/sp/lesmi.
8. Сайт кафедры «Технология машиностроения» Механико-машиностроительного института Национального технического университета Украины «Киевский политехнический институт» [Электронный ресурс]. – Режим доступа: http://tm-mmi.kpi.ua/index.php/ru/1/publications.

*Стаття надійшла до редакції 08.11.2012*