



Д.А. РАЧКОВСКИЙ, Е.Г. РЕВУНОВА

УДК 004.942 + 623.454.862

РАНДОМИЗИРОВАННЫЙ МЕТОД РЕШЕНИЯ
ДИСКРЕТНЫХ НЕКОРРЕКТНЫХ ЗАДАЧ

Ключевые слова: *дискретная некорректная задача, случайные проекции, регуляризация.*

ВВЕДЕНИЕ

Для многих практических приложений необходимо решение дискретной задачи $Ax \approx b$, где матрица A и вектор b известны, вектор x требуется оценить. Если сингулярные значения A плавно убывают до нуля и отношение между наибольшим и наименьшим ненулевым сингулярными значениями велико, задачу называют дискретной некорректной задачей [1].

Приближенные решения дискретных некорректных задач как задач наименьших квадратов с использованием численных методов линейной алгебры, таких как разложения LU, Холецкого, QR, являются неустойчивыми. Это означает, что малые возмущения во входных данных приводят к большим возмущениям в решении.

Решение некорректных задач востребовано во многих областях науки и техники. Дискретные некорректные задачи возникают, например, при дискретизации интегральных уравнений Фредгольма и Вольтерры первого рода. Важные проблемы спектроскопии [2], гравиметрии, магнитометрии, электроразведки и другие обладают свойствами дискретных некорректных задач.

Для преодоления нестабильности и повышения точности решения дискретных некорректных задач предложены методы регуляризации [1–5]. Регуляризация накладывает на искомое решение ограничения, обеспечивающие его устойчивость. Например, регуляризация Тихонова [3, 1] накладывает штраф на решения с большими l_2 -нормами.

Недостатки, присущие методам решения дискретных некорректных задач на основе регуляризации Тихонова, связаны с высокой вычислительной сложностью и трудностью выбора надлежащего параметра регуляризации (веса штрафа), влияющего на устойчивость решения. Это определяет практическую потребность в развитии альтернативных подходов к решению дискретных некорректных задач, имеющих точность, сравнимую с регуляризацией Тихонова, но при меньших вычислительных затратах.

В настоящей работе предлагается подход, использующий идеи нейросетевого распределенного представления информации и случайных проекций [6–9]. В последнее время исследователи, работающие в области численных методов линейной алгебры, применяют подобные идеи для получения быстрых рандомизированных алгоритмов решения задач наименьших квадратов, факторизации мат-

© Д.А. Рачковский, Е.Г. Ревунова, 2012

риц, анализа главных компонент и др. В данной статье обобщены результаты предыдущих исследований авторов по разработке рандомизированных методов на основе случайных проекций, которые позволяют ускорить и обеспечить устойчивость решения дискретных некорректных обратных задач. Приведены также новые результаты по выбору параметров рандомизированного преобразования для получения близкой к оптимальной (минимальной) ошибки решения.

1. ЛИНЕЙНЫЕ ЗАДАЧИ НАИМЕНЬШИХ КВАДРАТОВ И ДИСКРЕТНЫЕ НЕКОРРЕКТНЫЕ ЗАДАЧИ

Во многих приложениях математики, физики, анализа данных требуется приближенное решение системы линейных уравнений, которая не имеет точного решения. Например, в задаче

$$\mathbf{Ax} \approx \mathbf{b}, \quad (1)$$

где матрица $\mathbf{A} \in \mathbb{R}^{m \times n}$ и вектор \mathbf{b} известны, вектор \mathbf{x} нужно оценить, для $m > n$, или $m < n$, или $\min(m, n) \neq \text{rank}(\mathbf{A})$ в общем случае не существует такого \mathbf{x} , что $\mathbf{Ax} = \mathbf{b}$. Метод наименьших квадратов выбирает \mathbf{x}' , минимизирующий сумму квадратов компонент вектора невязки, путем решения оптимизационной задачи

$$\mathbf{x}' = \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2. \quad (2)$$

Решение (2) можно получить, взяв производную от $\|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{Ax} - \mathbf{b})^T \times (\mathbf{Ax} - \mathbf{b})$ по \mathbf{x} и положив ее равной нулю. Это дает систему так называемых нормальных уравнений [10]

$$\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}, \quad (3)$$

которую надо решить для получения минимизирующего вектора \mathbf{x}' . При таком решении вектор остатка (невязки) $(\mathbf{Ax}' - \mathbf{b})$ ортогонален пространству столбцов \mathbf{A} , т.е. $(\mathbf{Ax}' - \mathbf{b})^T \mathbf{A} = 0$. Если матрица $\mathbf{A}^T \mathbf{A}$ имеет полный ранг и хорошо обусловлена, систему (3) можно решить с помощью разложения Холецкого $\mathbf{A}^T \mathbf{A} = \mathbf{R}^T \mathbf{R}$, что дает $\mathbf{R}^T \mathbf{Rx} = \mathbf{A}^T \mathbf{b}$, где \mathbf{R} является верхней треугольной матрицей.

В случае, когда \mathbf{A} имеет неполный ранг или плохо обусловлена, может использоваться решение QR-разложением: $\mathbf{A} = \mathbf{QR}$, где \mathbf{Q} — ортогональная матрица, \mathbf{R} — верхняя треугольная матрица. Тогда из (3) получаем систему $\mathbf{Rx} = \mathbf{Q}^T \mathbf{b}$, которая благодаря треугольности \mathbf{R} решается последовательно, путем обратной подстановки.

При очень плохо обусловленной матрице \mathbf{A} для решения используют разложение по сингулярным значениям — SVD (Singular Value Decomposition). Метод SVD представляет \mathbf{A} как

$$\mathbf{A} = \mathbf{USV}^T, \quad (4)$$

где \mathbf{U} — матрица левых сингулярных векторов с ортонормированными столбцами, \mathbf{V} — матрица правых сингулярных векторов с ортонормированными столбцами, \mathbf{S} — диагональная матрица сингулярных значений. С помощью SVD решение получают в виде

$$\mathbf{x}' = \mathbf{VS}^{-1} \mathbf{U}^T \mathbf{b} = \mathbf{A}^+ \mathbf{b}, \quad (5)$$

где \mathbf{A}^+ — обобщенное обращение Мура–Пенроуза (псевдообращение) матрицы \mathbf{A} . Это решение дает вектор минимальной длины среди векторов, удовлетворяющих (2).

Методы, основанные на (5), находят решение за время $O(mn^2)$. Метод SVD наиболее вычислительно нагрузочный среди упомянутых выше (с точностью до постоянных множителей), но его использование предпочтительно, если \mathbf{A} очень плохо обусловлена. Обозначим i -е сингулярное значение \mathbf{A} как $\sigma_i(\mathbf{A})$, $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \dots \geq 0$, а максимальное и минимальное сингулярные значения \mathbf{A} — как $\sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A})$ и $\sigma_{\min}(\mathbf{A})$. Число обусловленности \mathbf{A} есть $\text{cond}(\mathbf{A}) = \sigma_{\max}(\mathbf{A}) / \sigma_{\min}(\mathbf{A})$. Если $\text{cond}(\mathbf{A})$ велико, матрица плохо обусловлена, что подразумевает потенциально неустойчивое и неточное решение. Проявлением неустойчивости является тот факт, что незначительные изменения в векторе \mathbf{b} (например, из-за шума) вызывают большие изменения в решении \mathbf{x}' и ошибка решения, как правило, велика, особенно при повышении уровня шума. Действительно, для больших $\text{cond}(\mathbf{A})$ величины, обратные сингулярным значениям в \mathbf{S}^{-1} , становятся очень большими, поэтому значения компонентов \mathbf{x}' также будут очень большими, как следует из (3) и (5).

Если в спектре сингулярных значений имеется резкий перепад, а сингулярные значения после перепада очень малы, они могут рассматриваться как шумовые и в некоторых методах устраняются применением порога, что обеспечивает повышение устойчивости. Однако этот прием не работает в дискретных некорректных задачах, поскольку, как отмечалось во введении, в ряду сингулярных значений нет разрывов и численный ранг не определен.

Классическим методом решения дискретных некорректных задач является регуляризация Тихонова [3, 1]. Задача регуляризации Тихонова стандартного вида формулируется следующим образом:

$$\mathbf{x}_{\text{reg}} = \arg \min_{\mathbf{x}} (\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda^2 \|\mathbf{x}\|^2), \quad (6)$$

где λ — параметр регуляризации.

Решение (6) можно получить методом фильтрованного SVD [1]:

$$\mathbf{x}_{\text{reg}} = \mathbf{V} \text{diag}(f_i / \sigma_i) \mathbf{U}^T \mathbf{y}, \quad (7)$$

где $f_i = \sigma_i^2 / (\sigma_i^2 + \lambda^2)$ — так называемые фильтрующие множители [1].

Одной из проблем является использование SVD-разложения для \mathbf{A} , так как оно имеет большую вычислительную сложность $O(mn \min\{m, n\})$. Еще одна важная проблема — выбор надлежащего параметра регуляризации λ . Это требует дополнительных вычислений и не всегда позволяет достичь необходимого результата, т.е. обеспечить устойчивое и точное решение \mathbf{x}' .

Предложен ряд методов для выбора параметра регуляризации [1]. В методе L-кривой [11] строится график $\|\mathbf{x}_{\text{reg}}\|_2$ от $\|\mathbf{A}\mathbf{x}_{\text{reg}} - \mathbf{b}\|_2$ для допустимых значений параметра регуляризации. При таком представлении L-кривая отражает компромисс между $\|\mathbf{x}_{\text{reg}}\|_2$ и $\|\mathbf{A}\mathbf{x}_{\text{reg}} - \mathbf{b}\|_2$, что и является сущностью регуляризации. Для дискретных некорректных задач L-кривая, построенная в логарифмическом масштабе, часто имеет характерную L-образную форму. Вертикальную и горизонтальную части кривой разделяет выраженный угол. В качестве оптимального выбирается параметр регуляризации, соответствующий точке вблизи от угла L-кривой. Для дискретного параметра регуляризации угол находится как точка с максимальной кривизной на сплайн-аппроксимации.

Согласно принципу невязки [4] параметры регуляризации выбираются так, чтобы норма остатка для регуляризованного решения удовлетворяла $\|\mathbf{A}\mathbf{x}_{\text{reg}} - \mathbf{b}\|_2 = \|\mathbf{d}\|_2$, где \mathbf{d} — возмущение правой части и, следовательно, требу-

ет оценки $\|\mathbf{d}\|_2$. Принцип невязки реализуется как минимизация в смысле наименьших квадратов с квадратичным ограничением: $\min\|\mathbf{x} - \mathbf{x}_1\|$ при условии $\|\mathbf{Ax} - \mathbf{b}\| \leq \delta$, где \mathbf{x}_1 — начальное приближение решения (равно нулю, если не задано), δ — положительная константа. При непрерывном параметре регуляризации для минимизации используется метод Ньютона.

Выбор параметра регуляризации в методе обобщенной перекрестной проверки (кросс-валидации) [12] основан на том, что значение произвольного компонента b_i правой части \mathbf{b} можно спрогнозировать с помощью соответствующего регуляризованного решения. Параметр выбирается так, чтобы минимизировать $\|\mathbf{Ax}_{\text{reg}} - \mathbf{b}\|^2 / D^2$, где D — эффективное число степеней свободы, которое вычисляется как $D = m - \sum_i f_i$ или как $\text{tr}(\mathbf{I} - \mathbf{AA}^T)$, где \mathbf{A}^T — матрица, с по-

мощью которой получают регуляризованное решение $\mathbf{x}_{\text{reg}} = \mathbf{A}^T \mathbf{b}$. Здесь ошибки в \mathbf{b} (вектор аддитивного шума) считаются некоррелированными случайными величинами с нулевым средним и одинаковой дисперсией.

Указанные методы не всегда дают устойчивые результаты, поэтому при неверных значениях параметра регуляризации ошибка решения может быть значительной. В последнее время появились рандомизированные подходы, идеи которых могут быть полезны для решения дискретных некорректных задач. Хотя в этих подходах рандомизация направлена на ускорение вычислений, ее можно также использовать для стабилизации решения дискретных некорректных задач.

2. РАНДОМИЗИРОВАННАЯ АППРОКСИМАЦИЯ НАИМЕНЬШИХ КВАДРАТОВ

При большой размерности матричных данных классические алгоритмы не во всех случаях могут обеспечить эффективную обработку. Традиционные алгоритмы предназначены для обработки матриц, элементы которых заданы с высокой точностью, что не всегда оправдано, так как данные в больших матрицах часто изначально неточные. Кроме того, в некоторых приложениях передача данных может составлять значительную долю общих вычислительных затрат, поэтому требуются алгоритмы с меньшим количеством проходов по данным, позволяющие использовать мощности новых процессорных архитектур, оптимизированных под специфические векторные операции. Для преодоления этих проблем недавно предложены и активно развиваются подходы к матричной обработке с использованием рандомизированных алгоритмов [13].

Следует отметить, что рандомизированные алгоритмы, как показывают оценки их точности, часто являются более точными и надежными, чем детерминированные. Хотя их результаты носят вероятностный характер, вероятность ошибки может регулироваться путем выбора параметров и достигать малых значений (например, менее 10^{-15}), сохраняя при этом преимущества по вычислительным затратам.

Рандомизация реализуется путем случайного сэмплирования (выборки), случайной проекции или их комбинации [14, 15]. Под сэмплированием здесь понимается получение некоторого случайного подмножества матрицы — отдельных элементов (компонентов, ячеек) либо столбцов и строк в соответствии с некоторым распределением вероятностей. Проецирование — случайное линейное отображение (вложение), реализуемое, как правило, с помощью умножения на случайную матрицу, элементы которой генерируются в соответствии с некоторым распределением вероятностей. Как сэмплирование, так и проецирование уменьшают размерность исходной задачи и соответственно вычислительную сложность последующей обработки.

На работы по рандомизированным матричным алгоритмам существенно повлияло направление случайных вложений. В [16] показано, что попарные евклидовы расстояния между точками во входном пространстве приблизительно сохраняются при отображении точек в евклидово пространство гораздо меньшей размерности с использованием случайного гауссова проектора. Указанная и последующие работы зародили идею, что некоторые вычислительные задачи могут быть решены с большей вычислительной эффективностью, если их предварительно перевести в пространство меньшей размерности. Это было использовано в приложениях, требующих быстрой оценки сходства векторов. В [17] такой подход впервые применен к линейной алгебре в контексте «латентного семантического индексирования» на основе SVD.

В ряде работ изучался вопрос уменьшения вычислительных затрат на реализацию самого отображения. В [18] показано, что дискретные случайные матрицы с элементами $\{-1,0,1\}$ дают почти такие же хорошие результаты, как и гауссовы матрицы. В [19, 7] рассмотрены сильно разреженные (и поэтому потенциально «быстрые») дискретные случайные матрицы, однако результаты этих исследований не были применены к численным методам линейной алгебры. Быстрое преобразование Джонсона–Линденштраусса, которое сочетает скорость быстрого преобразования Фурье со свойствами вложений, имеющимися у гауссовой матрицы, вызвало существенный интерес в данной области. В частности, применение этого метода к аппроксимации матриц [20] привело к созданию самых быстрых алгоритмов, известных в настоящее время.

В рандомизированных алгоритмах быстрого решения задачи аппроксимации методом наименьших квадратов рандомизация рассматривается как умножение входной матрицы \mathbf{A} и вектора \mathbf{b} на специально сконструированную, не зависящую от данных, случайную матрицу \mathbf{S} (часто построенную из нескольких других матриц). Таким образом, исходная задача наименьших квадратов (2) заменяется следующей:

$$\mathbf{x}' = \arg \min_{\mathbf{x}} \|\mathbf{Sb} - \mathbf{SAx}\|_2. \quad (8)$$

Решение (8) может быть получено традиционным детерминированным алгоритмом, например с использованием SVD для вычисления обобщенной обратной матрицы:

$$\mathbf{x}' = (\mathbf{SA})^+ \mathbf{Sb}. \quad (9)$$

Могут применяться также стандартные итерационные методы, например метод сопряженных градиентов над нормальным остатком (Conjugate Gradient Normal Residual).

Для рандомизации задачи наименьших квадратов используют как сэмплирование [15], так и проецирование [20]. Границы ошибки и скорости для обоих подходов улучшены в [10] для задач аппроксимации наименьших квадратов в случае избыточных ограничений $m \gg n$. В работе [21] подобные приемы применялись для теоретического и экспериментального анализа алгоритмов случайного проецирования при решении переопределенных ($m > n$) задач наименьших квадратов. Кроме того, метод сэмплирования распространяется на случай недоопределенной ($m < n$) задачи наименьших квадратов [22].

3. РЕШЕНИЕ ДИСКРЕТНЫХ НЕКОРРЕКТНЫХ ЗАДАЧ НА ОСНОВЕ СЛУЧАЙНЫХ ПРОЕКЦИЙ

Рассмотрим дискретную некорректную задачу вида

$$\mathbf{F}\mathbf{x} = \mathbf{y}. \quad (10)$$

Здесь $\mathbf{F} \in \mathfrak{R}^{n \times n}$ — матрица с сингулярными значениями σ_i , плавно убывающими до нуля, и большим числом обусловленности (т.е. матрица \mathbf{F} плохо обусловлена и имеет неопределенный численный ранг). Вектор $\mathbf{y} \in \mathfrak{R}^n$ искажен аддитивным шумом $\boldsymbol{\varepsilon} \in \mathfrak{R}^n$: $\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}$.

Исследования рандомизированных алгоритмов решения задачи наименьших квадратов проведены их авторами для случая, когда вектор \mathbf{b} не искажен шумом, матрица \mathbf{A} хорошо обусловлена и имеет полный ранг. Именно для этого варианта доказывают (например, в [10]), что решение с использованием случайного проецирования (9) с ростом k приближается к решению с помощью обычного псевдообращения (5). Однако для рассматриваемого случая дискретной некорректной задачи решение (5) неустойчиво и имеет недопустимо низкую точность. Поэтому для данной задачи не требуется близости решения, полученного с использованием случайной проекции, к решению на основе псевдоинверсии.

Рассмотрим подход к устойчивому решению дискретной некорректной задачи на основе рандомизации. Как и при рандомизированном решении обычной задачи наименьших квадратов (9), умножим обе части (10) на матрицу $\mathbf{G} \in \mathfrak{R}^{k \times n}$, $k \leq n$, элементы которой являются реализациями случайной величины с нормальным (гауссовым) распределением с нулевым средним и единичной дисперсией. Число столбцов n матрицы \mathbf{G} определяется размерностью матрицы \mathbf{F} , число строк k априорно неизвестно. Получаем

$$\mathbf{GFx} = \mathbf{Gy}, \quad \mathbf{GF} \in \mathfrak{R}^{k \times n}, \quad \mathbf{Gy} \in \mathfrak{R}^k. \quad (11)$$

Тогда задача наименьших квадратов с учетом проекции записывается как

$$\mathbf{x}_{Pr} = \arg \min_{\mathbf{x}} \|\mathbf{GFx} - \mathbf{Gy}\|_2, \quad (12)$$

а ее решение с применением обычной псевдообратной матрицы — в виде

$$\mathbf{x}_{pinPr} = (\mathbf{GF})^+ \mathbf{Gy}. \quad (13)$$

Точность решения обратной задачи оценим с использованием ошибки e восстановления истинного вектора \mathbf{x}_0 :

$$e = \|\mathbf{x}_0 - \mathbf{x}'\|_2 = \|\mathbf{e}\|_2, \quad (14)$$

где \mathbf{x}' — полученное решение, \mathbf{e} — вектор ошибки решения.

Вектор ошибки \mathbf{e} можно представить как сумму смещения и дисперсии [23]. Вычислим их следующим образом. Обозначим \mathbf{P} оператор, преобразующий \mathbf{y} в $\mathbf{x}' = \mathbf{Py}$. Тогда, принимая во внимание, что $\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}$ и $\mathbf{y}_0 = \mathbf{Fx}_0$, выражение для \mathbf{x}' запишем в виде

$$\mathbf{x}' = \mathbf{P}(\mathbf{y}_0 + \boldsymbol{\varepsilon}) = \mathbf{Py}_0 + \mathbf{P}\boldsymbol{\varepsilon} = \mathbf{PFx}_0 + \mathbf{P}\boldsymbol{\varepsilon}. \quad (15)$$

Используя (15), получаем выражение для \mathbf{e} :

$$\mathbf{e} = \mathbf{x}' - \mathbf{x}_0 = \mathbf{PFx}_0 - \mathbf{x}_0 + \mathbf{P}\boldsymbol{\varepsilon} = (\mathbf{PF} - \mathbf{I})\mathbf{x}_0 + \mathbf{P}\boldsymbol{\varepsilon}. \quad (16)$$

Итак,

$$\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2, \quad \mathbf{e}_1, \mathbf{e}_2 \in \mathfrak{R}^n, \quad \mathbf{e}_1 = (\mathbf{PF} - \mathbf{I})\mathbf{x}_0, \quad \mathbf{e}_2 = \mathbf{P}\boldsymbol{\varepsilon}. \quad (17)$$

Составляющую \mathbf{e}_1 называют смещением, \mathbf{e}_2 — дисперсией.

После проецирования разложение смещение-дисперсия вектора ошибки решения принимает следующие формы. Для ошибки решения методом проецирования можно записать

$$\mathbf{e}_{Pr} = \mathbf{x}_{Pr} - \mathbf{x}_0 = \mathbf{P}_{Pr} \mathbf{G} \mathbf{y} - \mathbf{x}_0 = \mathbf{P}_{Pr} \mathbf{G} (\mathbf{F} \mathbf{x}_0 + \boldsymbol{\varepsilon}) - \mathbf{x}_0 = (\mathbf{P}_{Pr} \mathbf{G} \mathbf{F} - \mathbf{I}) \mathbf{x}_0 + \mathbf{P}_{Pr} \mathbf{G} \boldsymbol{\varepsilon}, \quad (18)$$

где \mathbf{x}_{Pr} — решение после проекции, \mathbf{P}_{Pr} — оператор, преобразующий $\mathbf{G} \mathbf{y}$ в \mathbf{x}_{Pr} : $\mathbf{x}_{Pr} = \mathbf{P}_{Pr} \mathbf{G} \mathbf{y}$.

Для псевдообратного решения после проецирования имеем

$$\mathbf{P}_{pinPr} = \mathbf{C} \text{diag}(\varphi_i / s_i) \mathbf{B}^T. \quad (19)$$

Здесь \mathbf{B} , \mathbf{C} , $\boldsymbol{\Sigma}$ — результат SVD матрицы $\mathbf{G} \mathbf{F} = \mathbf{B} \boldsymbol{\Sigma} \mathbf{C}^T$, $s_i = \text{diag} \boldsymbol{\Sigma}$ — сингулярные значения (элементы диагональной матрицы $\boldsymbol{\Sigma}$), s_i подвергаются пороговому преобразованию:

$$\text{если } \sigma_i > \text{tresh}, \varphi_i = 1, \text{ иначе } \varphi_i = 0; \text{tresh} = \max(K, N) \text{eps}(\max(\sigma_i)), \quad (20)$$

где относительная точность представления числа с плавающей точкой $\text{eps}(z)$ — положительное расстояние от $\text{abs}(z)$ до следующего большего по величине числа с плавающей точкой той же точности, что и z . Здесь смещение и дисперсия ошибки имеют вид

$$\mathbf{e}_{1pinPr} = (\mathbf{P}_{pinPr} \mathbf{G} \mathbf{F} - \mathbf{I}) \mathbf{x}_0, \quad \mathbf{e}_{2pinPr} = \mathbf{P}_{pinPr} \mathbf{G} \boldsymbol{\varepsilon}; \quad \mathbf{e}_{1pinPr} + \mathbf{e}_{2pinPr} = \mathbf{e}_{pinPr}, \quad (21)$$

где \mathbf{e}_{pinPr} — вектор ошибки решения для псевдообращения с использованием случайной проекции.

Псевдообратное решение, основанное на QR-разложении, реализовано следующим образом:

$$\mathbf{x}_{QR} = (\mathbf{Q}^T \mathbf{F})^+ \mathbf{Q}^T \mathbf{y}, \quad \mathbf{P}_{QR} = (\mathbf{Q}^T \mathbf{F})^+ \mathbf{Q}^T, \quad (22)$$

где \mathbf{Q} получается QR-разложением матрицы $(\mathbf{G} \mathbf{F})^T = \mathbf{Q} \mathbf{R}$;

$$\mathbf{e}_{1QR} = (\mathbf{P}_{QR} \mathbf{F} - \mathbf{I}) \mathbf{x}_0, \quad \mathbf{e}_{2QR} = \mathbf{P}_{QR} \boldsymbol{\varepsilon}. \quad (23)$$

Для решений, полученных с помощью случайной проекционной матрицы, можно построить график зависимости от размерности k нормы смещения $e_1 = \|\mathbf{e}_1\|_2$ и дисперсии $e_2 = \|\mathbf{e}_2\|_2$, а также суммарной ошибки $e = \|\mathbf{e}_1 + \mathbf{e}_2\|_2$. Пример зависимостей для дискретной некорректной задачи Phillips [24] и метода проецирования случайной матрицей $\mathbf{G}(k \times n)$ при уровнях шума $1e-3$, $1e-5$, $1e-7$ приведен на рис.1.

При возрастании k $\|\mathbf{e}_1\|$ уменьшается, а $\|\mathbf{e}_2\|$ возрастает, так что $\|\mathbf{e}_1 + \mathbf{e}_2\|$ имеет минимум. Такое поведение этих зависимостей характерно для задач выбора моделей, в которых модель оптимальной сложности обеспечивает минимум ошибки. Подобный характер поведения ошибки имел место для всех исследованных некорректных обратных задач при уровне шума выше некоторого минимального.

Таким образом, для получения решения с минимальной ошибкой необходимо использовать случайную проекционную матрицу с размерностью k , близкой к оптимальной. Однако определение оптимального значения k по графикам зависимости ошибки восстановления истинного сигнала от k представляет

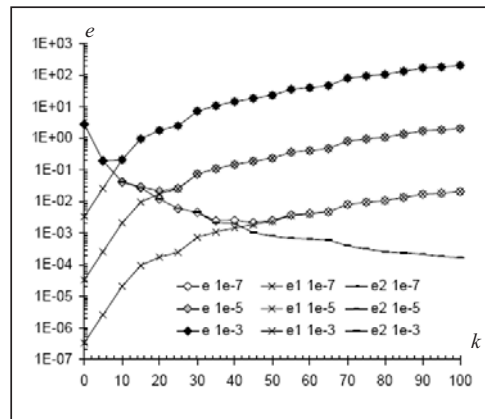


Рис. 1. Графики зависимостей ошибки решения e , ее смещения e_1 и дисперсии e_2 от размерности k случайной проекционной матрицы

только теоретический интерес, так как выражение (14) включает вектор истинного решения, неизвестный при решении практических задач.

Для того чтобы выбрать размерность k проекционной матрицы, при которой ошибка решения близка к минимальной в реальных условиях, т.е. когда точное решение неизвестно, используем различные типы критериев выбора. При этом размерность k выбирается такой, при которой критерий принимает некоторое характерное значение (обычно минимум или максимум).

Используем два типа критериев:

— критерии, предложенные для выбора параметров регуляризации Тихонова (КВПР);

— критерии выбора модели (КВМ), применяемые для выбора оптимальных моделей в области машинного обучения (machine learning), индуктивного обучения (inductive learning) и анализа данных (data mining).

КВПР с использованием L-кривой, обобщенной кросс-валидации и принципа невязки [1, 4, 12] рассмотрены в разд. 1; особенности применения КВМ описаны далее.

3.1. Выбор размерности проекционной матрицы по КВМ. В задачах построения моделей используется ряд подходов, в том числе линейные модели, нейронные сети, классификация и регрессия на основе деревьев, ядерные методы и др. По заданным данным, обычно состоящим из пар вход–выход, строится модель, связывающая вход и выход. Необходимо выбрать модель, близкую к оптимальной для определенного приложения.

Методы выбора модели используют различные критерии выбора [25–27]. КВМ формулируются так, чтобы автоматически уменьшать сложность модели с увеличением уровня шума путем баланса между сложностью модели и ошибкой аппроксимации. На оптимальных моделях критерии достигают минимума. КВМ разработаны на основе различных предположений, например с использованием ошибки прогнозирования (predictive training error) [26], ошибки обобщения (generalization error) [23, 28], информационных статистик (information statistics) [25], длины описания (description length) [29] и др. В отличие от КВПР, которые разрабатывались для градуального (не дискретного) параметра λ регуляризации Тихонова, КВМ создавались для моделей с дискретной сложностью (например, числом базисных функций s).

В настоящей работе исследован выбор размерности случайной проекционной матрицы с использованием наиболее известных критериев выбора модели: C_p (C_p Statistic) Маллоуза [26], AIC (Akaike Information Criterion) Акаике [25], gMDL (Minimum Description Length) Бин Ю [27].

Пусть набор данных D_L представлен L парами $D_L = \{(\mathbf{x}_i, y_i)\}_{i=1, \dots, L}$, $\mathbf{X} \in \mathbb{R}^{L \times N}$, $y_i = y_{0i} + \varepsilon_i$, ε — гауссов аддитивный шум. Модель имеет вид $y = \mathbf{w}^T \mathbf{x} = f(\mathbf{x})$.

Ошибка прогнозирования — это мера ошибки между оцененными и реальными значениями в точках, соответствующих образцам обучающей выборки:

$$I_{\text{PTE}} = E_{\varepsilon} \|\mathbf{B}\mathbf{y} - \mathbf{y}_0\|^2 = \|\mathbf{B}\mathbf{y}_0 - \mathbf{y}_0\|^2 + \text{trace}(\mathbf{B}\mathbf{Q}\mathbf{B}^T), \quad (24)$$

где $\mathbf{B} = \mathbf{X}_s (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T$ — отображение \mathbf{y} в оценку моделью (незашумленными) \mathbf{y}_0 , \mathbf{Q} — ковариационная матрица шума, \mathbf{X}_s — данные, соответствующие $s \leq N$ ненулевым компонентам вектора весов (параметров) \mathbf{w} модели сложности s , E_{ε} — ансамблевое усреднение по шуму.

Критерий C_p [26] дает несмещенную оценку ошибки прогнозирования

$$C_p = \text{RSS} + 2\sigma^2 s, \quad (25)$$

где RSS — квадрат нормы невязки \mathbf{y} , σ^2 — дисперсия шума.

Ошибка обобщения есть разность между значением функционала среднего риска для полученного решения $f_{H,L}$ и значением функционала среднего риска для истинного решения f_0 [23]:

$$I_G = I(f_{H,L}) - I(f_0), \quad f_{H,L} = \arg \min_{f \in H} I_{\text{emp}}(f),$$

$$I_{\text{emp}}(f) = \sum_{i=1}^L V(y_i, f(x_i)) / L, \quad (26)$$

где $f_{H,L}$ — решение, полученное с использованием ограниченной обучающей выборки длиной L для модели H (ограниченной сложности), I_{emp} — эмпирический риск, $V(\cdot)$ — функция потерь.

Информационный критерий AIC [25] оценивает ошибку обобщения с точки зрения информационной статистики. Он вычисляется как

$$\text{AIC} = 2s + L \ln(\text{RSS}), \quad (27)$$

где L — число наблюдений (образцов), s — число параметров.

Критерии минимальной длины описания для регрессии можно записать в виде

$$I_{\text{MDL}} = \text{DL}(\mathbf{y} | \mathbf{X}_s) + \text{DL}(s), \quad (28)$$

где $\text{DL}(\mathbf{y} | \mathbf{X}_s)$ — длина описания \mathbf{y} моделью сложности s , $\text{DL}(s)$ — длина описания модели.

В [27] предложены выражения для критерия gMDL, минимум которого соответствует минимуму длины описания (28):

$$\text{gMDL} = 0,5L \log(\text{RSS} / (L - s)) + 0,5s \log(F) + \log(L) \quad \text{при } R^2 \geq s / L,$$

$$\text{gMDL} = 0,5L \log(\mathbf{y}^T \mathbf{y} / L) + 0,5 \log(L) \quad \text{иначе.} \quad (29)$$

Здесь $F = \frac{\mathbf{y}^T \mathbf{y} - \text{RSS}}{s \text{RSS}} (L - s)$, R — множественный коэффициент корреляции,

$$R^2 = 1 - \frac{\text{RSS}}{\mathbf{y}^T \mathbf{y}}, \quad \text{RSS} = \|\mathbf{A} \mathbf{A}^+ \mathbf{y} - \mathbf{y}\|^2.$$

При проецировании случайной матрицей приведенные выражения для KBM соответствующим образом модифицируются: в (24)–(29) вместо \mathbf{y} , \mathbf{A} используются $\mathbf{S}\mathbf{y}$, $\mathbf{S}\mathbf{A}$.

4. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Экспериментальное исследование проведено на примере известных дискретных некорректных задач Carasso, Phillips, Delves, Vaart [24], а также задачи Spectr, возникающей в аэро-гамма съемке [2].

4.1. Задачи. Задача **Carasso** — реконструкция временного профиля источника тепла на основе измерения температуры с фиксированного расстояния. Обратное уравнение теплового потока является интегральным уравнением Воль-

терры первого рода $\int_a^x K(x, t)f(t)dt = g(x)$ с интервалом интегрирования $[0, 1]$.

Ядро имеет вид

$$K(s, t) = f(s-t), \quad t^{3/2} / (2\pi^{1/2} \exp(-1/4t)). \quad (30)$$

Интегральное уравнение дискретизируется методом квадратур с использованием простой коллокации и правила средней точки с N точками. Конструируется точное решение \mathbf{x}_0 , и затем \mathbf{y}_0 вычисляется как $\mathbf{y}_0 = \mathbf{F}\mathbf{x}_0$, т.е. в этой задаче $\mathbf{F}\mathbf{x}_0$ точно равно \mathbf{y}_0 .

Дискретное представление ядер и правых частей для задач Carasso, Phillips, Delves, Baart получено с использованием программ Regularization Tools [24].

Задача **Phillips** — решение уравнения Фредгольма первого рода $\int_a^b K(s, t)f(t)dt = g(s)$. Решение f , ядро K , правая часть g имеют соответственно вид

$$\begin{aligned} f(t) &= 1 + \cos(\pi t/3) \text{ при } |t| < 3, \quad f(t) = 0 \text{ при } |t| \geq 0, \\ K(s, t) &= f(s-t), \\ g(s) &= (6 - |s|)(1 + 0,5 \cos(\pi s/3)) + (9/2\pi) \sin(\pi |s|/3). \end{aligned} \quad (31)$$

Интервал интегрирования составляет $[-6, 6]$. Дискретизация осуществляется методом Галеркина. Для матрицы \mathbf{F} величины m, n ($m = n$) должны быть кратны четырем. В этой задаче $\mathbf{F}\mathbf{x}_0$ не точно равно \mathbf{y}_0 .

Задача **Delves** — вычисление второй производной. Осуществляется дискретизация аналитически заданного уравнения Фредгольма первого рода с ядром K , которое является функцией Грина второй производной, с правой частью g и решением f :

$$\begin{aligned} K(s, t) &= s(t-1) \text{ при } s < t, \quad K(s, t) = t(s-1) \text{ при } s \geq t, \\ g(s) &= (s^3 - s)/6, \quad f(t) = t. \end{aligned} \quad (32)$$

В данной задаче $\mathbf{F}\mathbf{x}_0$ точно равно \mathbf{y}_0 .

Задача **Baart** — дискретизация искусственного интегрального уравнения Фредгольма первого рода с ядром K и правой частью g , которые заданы как

$$K(s, t) = \exp(s \cos(t)), \quad g(s) = 2 \sin(s)/s, \quad s \in [0, \pi/2], \quad t \in [0, \pi]. \quad (33)$$

Решением является $f(t) = \sin(t)$. Дискретизация осуществляется методом Галеркина с ортонормальными базисными функциями. Для матрицы \mathbf{F} величины m, n ($m = n$) должны быть четными. В этой задаче $\mathbf{F}\mathbf{x}_0$ не точно равно \mathbf{y}_0 .

В задаче **Spectr** [30, 31] требуется, используя модельные данные радиоактивного мониторинга посредством аэро-гамма съемки, восстановить поверхностную плотность источников радиоактивности. Значения измерений \mathbf{y} получены умножением сигнала \mathbf{x} , подлежащего восстановлению, на матрицу \mathbf{F} , которая соответствует отклику детектора, находящегося на высоте $h = 100 \text{ m}$ (на борту авианосителя) [2]. В этой задаче $\mathbf{F}\mathbf{x}_0$ точно равно \mathbf{y}_0 .

Во всех задачах матрица \mathbf{F} имела размер 200×200 , большое число обусловленности ($\text{cond}(\mathbf{F}) = \sigma_{\max} / \sigma_{\min} \gg 1$) и сингулярные значения, плавно убывающие до нуля. Правая часть \mathbf{y}_0 искажалась аддитивным шумом с гауссовым распределением и различными амплитудами. Матрица случайных проекций $\mathbf{G} \in \mathcal{R}^{k \times n}$, $n = 200$, $k \leq n$, — гауссова случайная матрица с элементами с нулевым средним и единичной дисперсией.

4.2. Результаты. Проведено экспериментальное сравнение следующих методов решения указанных дискретных некорректных обратных задач.

1. Решение на основе псевдообращения (5),
2. Регуляризация Тихонова (7) с выбором параметра регуляризации λ на основе: обобщенной перекрестной проверки (кросс-валидации) — GCV (Generalized Cross Validation); принципа невязки — DSC (Discrepancy); с использованием L-кривой — LC (L Curve).

Решение для регуляризации Тихонова получено с применением программ Regularization Tools [24].

3. Псевдообращение (PseudoInverse) с проецированием случайной матрицей \mathbf{G} (13) и ортогонализированной матрицей \mathbf{Q} (22) и выбором их размерностей с помощью:
 - КВПР на основе GCV; DSC; с применением LC;
 - КВМ с использованием gMDL; C_p ; AIC.

Как и на рис. 1, для всех исследованных задач и решения псевдообращением с проецированием случайной матрицей существует диапазон уровней шума, в котором при некотором значении k наблюдается минимум ошибки решения. С повышением уровня шума положение минимума смещается к меньшим значениям k и минимальное значение ошибки увеличивается.

Напомним, что в реальных условиях кривые зависимости ошибки от размерности k проекционной матрицы недоступны, так как неизвестно истинное решение, поэтому для выбора k , при котором ошибка решения близка к минимальной, используем КВПР (разд. 1) и КВМ (разд. 3.1). На рис. 2 приведены примеры зависимостей от k значений некоторых из этих критериев в задаче Phillips. Для всех экспериментов уровень шума $1e-5$; проецирование осуществлялось случайными матрицами \mathbf{G} (13) и \mathbf{Q} (22). Можно видеть, что значения k , при которых достигаются минимумы реальной ошибки и критериев, близки.

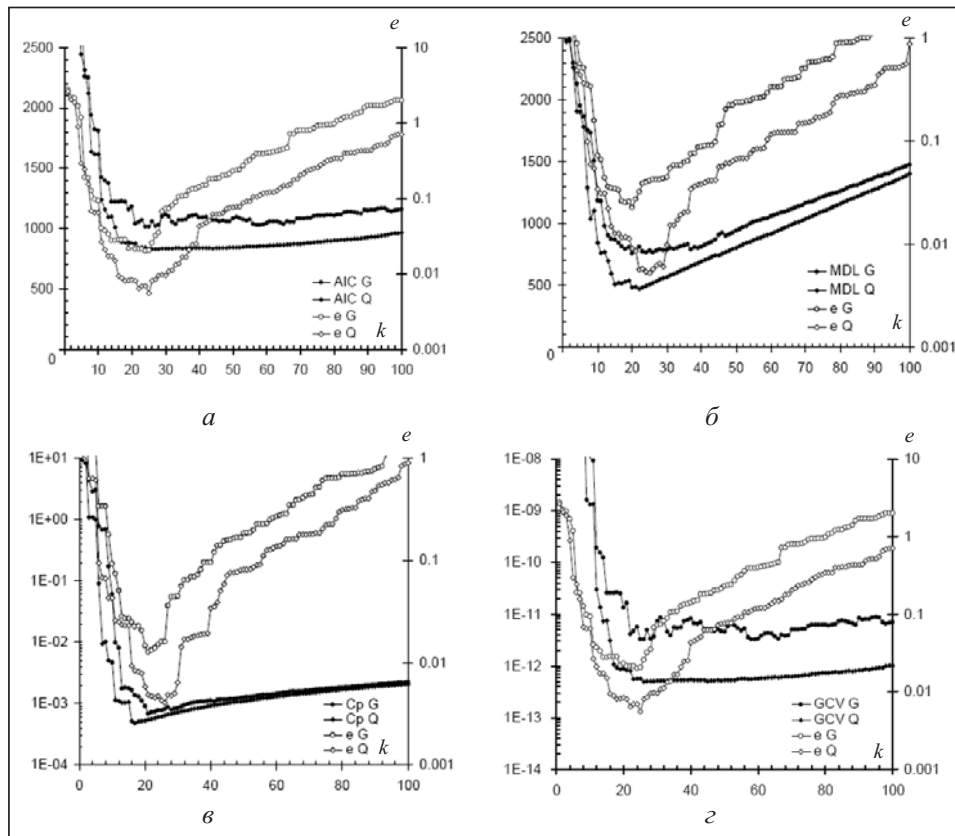


Рис. 2. Графики зависимостей значений критериев выбора модели (левая ось Y) и ошибки решения e (правая ось Y) от k : AIC (а); gMDL (б); C_p (в); GCV (з)

В табл. 1–3 отражены результаты некоторых экспериментов в терминах ошибки решения. В табл. 1 приведено решение дискретных некорректных задач Carasso, Phillips, Delves, Baart, Spectr для трех уровней шума nl в y традиционными методами без проецирования — псевдообращением (PseudoInverse) и регуляризацией Тихонова (Tikhonov) с выбором параметра регуляризации КВПР; даны средние значения $M(terr)$ и среднеквадратичные отклонения (с.к.о.) для десяти реализаций шума в y .

Таблица 1

Задача/метод	$M(terr)$	с.к.о.	$M(terr)$	с.к.о.	$M(terr)$	с.к.о.
Carasso	nl=1e-3		nl=1e-5		nl=1e-7	
PseudoInverse	1.08E+05	1.11E+04	1190.35	166.09	11.79	2.67
Tikhonov GCV	0.736	0.399	0.098	0.016	0.024	0.002
Tikhonov DSC	0.530	0.046	0.069	0.007	0.023	0.001
Tikhonov LC	9852	31153	125.96	396.33	11.79	2.67
Individual LC	H: 98515	L: 0.498	H: 1254	L: 0.529	H: 18.30	L: 8.91
Phillips	nl=1e-3		nl=1e-5		nl=1e-7	
PseudoInverse	8419	2745	78.22	32.51	0.80	0.17
Tikhonov GCV	0.204	0.452	0.009	0.003	0.0013	0.0002
Tikhonov DSC	0.032	0.004	0.005	0.001	0.0009	0.0001
Tikhonov LC	0.350	0.063	0.881	0.143	0.293	0.088
Individual LC	H: 0.448	L: 0.275	H: 1.169	L: 0.703	H: 0.414	L: 0.120
Delves	nl=1e-5		nl=1e-6		nl=1e-7	
PseudoInverse	32.30	2.00	3.36	0.23	0.34	0.04
Tikhonov GCV	0.149	0.084	0.084	0.006	0.057	0.005
Tikhonov DSC	0.120	0.007	0.082	0.002	0.053	0.002
Tikhonov LC	0.128	0.013	0.145	0.010	0.072	0.005
Individual LC	H: 0.152	L: 0.110	H: 0.154	L: 0.120	H: 0.077	L: 0.065
Baart	nl=1e-5		nl=1e-7		nl=1e-9	
PseudoInverse	1.14E+08	7.04E+07	9.40E+05	5.92E+05	11649	3558
Tikhonov GSV	2.92E+07	9.24E+07	40920	116648	2313	7314
Tikhonov DSC	0.107	0.027	0.064	0.002	0.038	0.001
Tikhonov LC	0.075	0.015	0.063	0.007	0.046	0.013
Individual GCV	H: 2.92E+08	L: 0.06	H: 371238	L: 0.061	H: 23131	L: 0.036
Spectrum	nl=1e-1		nl=1e-2		nl=1e-6	
PseudoInverse	4.68E+12	3.76E+12	5.5E+09	2.82E+09	490117	374641
Tikhonov GCV	107.5	49.02	65.8	20.7	2287	6841
Tikhonov DSC	93.6	6.74	62.86	1.146	22.62	0.2
Tikhonov LC	91.87	4.75	57.9	2.2	23.4	2.25
Individual GCV	H: 241	L: 77.1	H: 124	L: 52.9	H: 21741	L: 21.6

Обычное псевдообращение обеспечивает приемлемую ошибку решения только при самом низком уровне шума в некоторых задачах, но не может использоваться во всех задачах при больших уровнях шума из-за очень большой ошибки. Применение регуляризации Тихонова значительно снижает ошибку решения.

Для задач Carasso, Phillips, Delves среди детерминированных методов наименьшую ошибку решения обеспечивает регуляризация Тихонова с КВПР DSC. Близкие результаты у КВПР GCV. Для задачи Baart хорошие результаты дает регуляризация Тихонова с КВПР LC и DSC, а критерий GCV работает плохо. Для задачи Spectr все детерминированные методы при фиксированном уровне шума показывают сравнимую ошибку.

Таблица 2

Задача/ метод	$M(e)$	с.к.о.	$M(k)$	с.к.о.	$M(e)$	с.к.о.	$M(k)$	с.к.о.	$M(e)$	с.к.о.	$M(k)$	с.к.о.
Carasso	nl=1e-3				nl=1e-5				nl=1e-7			
e G	1.43	0.24	12.1	2.23	0.21	0.05	29.5	2.59	0.03	0.063	61.6	4.14
GCV G	3.59	2.86	17.9	5.47	1.35	1.41	49.4	14.2	0.08	0.016	76.5	10.5
DSC G	1.80	0.84	12.4	2.27	0.28	0.13	33.8	5.03	0.04	0.007	65.4	4.97
MDL G	2.14	1.72	14.4	3.34	0.21	0.05	29.9	3.45	0.04	0.007	53.8	3.22
Ср G	1.48	0.27	12.5	1.84	0.21	0.05	29.9	3.45	0.04	0.028	63.3	4.69
AIC G	2.75	2.27	16.3	4.16	0.92	0.97	45.8	12.6	0.05	0.001	70.5	7.01
e Q	0.48	0.07	16.7	1.70	0.08	0.01	34.1	3.14	0.02	0.001	70.7	3.92
GCV Q	0.55	0.14	17.1	1.10	0.09	0.01	36.2	1.62	0.02	0.006	72.4	1.78
DSC Q	1.91	0.28	8.6	0.84	0.34	0.10	24.4	1.71	0.05	0.007	50.4	2.22
MDL Q	0.65	0.15	12.8	2.10	0.10	0.02	28.8	1.23	0.04	0.005	45.0	7.26
Ср Q	0.92	0.16	10.2	1.23	0.10	0.02	28.2	1.14	0.04	0.001	52.7	3.53
AIC Q	0.56	0.11	16.1	1.10	0.08	0.01	35.0	1.63	0.02	0.063	70.2	2.30
Phillips	nl=1e-3				nl=1e-5				nl=1e-7			
e G	0.13	0.04	9.6	1.17	0.02	0.01	19.7	3.13	0.002	0.000	44.5	3.57
GCV G	0.72	1.34	14.3	6.31	0.25	0.39	44.5	23.1	0.003	0.001	57.2	5.49
DSC G	0.15	0.07	9.5	1.84	0.03	0.02	23.4	5.64	0.002	0.000	45.7	4.30
MDL G	0.17	0.09	11.0	1.94	0.02	0.01	21.3	3.56	0.002	0.000	39.1	3.78
Ср G	0.16	0.08	10.5	1.65	0.03	0.02	23.1	5.95	0.002	0.000	43.9	3.48
AIC G	0.64	1.34	13.3	6.31	0.08	0.10	33.5	11.2	0.003	0.001	54.9	6.24
e Q	0.03	0.00	12.1	1.60	0.005	0.001	24.5	2.37	0.001	0.000	45.6	3.50
GCV Q	0.03	0.01	12.1	0.88	0.006	0.001	24.3	1.89	0.001	0.000	54.7	5.89
DSC Q	0.25	0.11	7.3	0.48	0.030	0.008	14.7	1.25	0.004	0.001	32.1	2.23
MDL Q	0.04	0.02	9.6	1.26	0.007	0.002	18.7	2.21	0.001	0.000	35.8	1.81
Ср Q	0.07	0.02	7.9	0.74	0.009	0.002	17.0	1.76	0.001	0.000	35.4	1.71
AIC Q	0.03	0.00	11.1	0.88	0.005	0.001	23.1	1.91	0.001	0.000	47.9	2.18
Delves	nl=1e-5				nl=1e-6				nl=1e-7			
e G	0.18	0.03	9.4	2.07	0.12	0.01	23.5	3.47	0.07	0.003	58.3	8.18
GCV G	0.26	0.09	13.6	1.65	0.27	0.11	48.5	11.0	0.09	0.037	80.2	23.4
DSC G	0.19	0.03	7.4	2.32	0.12	0.01	22.5	3.89	0.07	0.002	49.1	4.36
MDL G	0.22	0.06	10.3	3.06	0.13	0.01	23.7	5.12	0.08	0.005	39.4	5.97
Ср G	0.21	0.06	9.5	3.21	0.13	0.01	21.1	5.13	0.07	0.004	44.6	5.91
AIC G	0.23	0.06	12.5	1.65	0.25	0.09	46.0	8.74	0.07	0.005	63.1	7.53
e Q	0.12	0.01	17.5	3.57	0.08	0.005	32.7	4.16	0.05	0.002	77.6	8.13
GCV Q	0.13	0.01	14.9	2.02	0.13	0.04	46.6	4.93	0.06	0.003	67.2	6.32
DSC Q	0.22	0.04	5.1	0.74	0.14	0.01	13.4	1.90	0.08	0.004	34.9	1.79
MDL Q	0.13	0.01	10.6	1.90	0.10	0.01	20.8	3.91	0.08	0.006	34.2	5.98
Ср Q	0.16	0.01	7.0	1.25	0.11	0.01	16.1	2.64	0.07	0.004	35.4	4.35
AIC Q	0.13	0.01	13.9	2.02	0.11	0.01	43.2	3.19	0.06	0.002	62.0	6.20

Для некоторых КВПР наблюдаются нестабильные результаты. Это проявляется в том, что при определенных реализациях шума ошибка решения очень велика и может приближаться к ошибке PseudoInverse. Для иллюстрации в строках Individual табл. 1 приведена минимальная (L) и максимальная (H) ошибки решения среди решений, полученных для разных реализаций шума. В частности, нестабильные результаты для КВПР с использованием L-кривой связаны с тем, что процедура нахождения угла L-кривой нетривиальна для тех изменений значений невязки и нормы вектора решения, которые реально встречаются во многих задачах. Классический алгоритм в реализации Хансена [24] часто дает ошибку. Его улучшение открывает перспективы повышения стабильности метода L-кривой,

что может повысить точность решения, так как при правильном срабатывании этот критерий показывает лучшие результаты [8, 9]. Некоторые шаги в данном направлении предприняты в работе [32].

В табл. 2 и 3 отражены результаты для решения с проецированием непосредственно случайной матрицей \mathbf{G} (13) и матрицей \mathbf{Q} после ортогонализации QR разложением (22). Приведены результаты (средние и с.к.о.), полученные при десяти случайных реализациях проекционных матриц и одной реализации шума (три уровня шума в \mathbf{y}) для задач Carasso, Phillips, Delves (табл. 2) и задач Baart, Spectr (табл. 3).

Таблица 3

Задача/ метод	$M(e)$	с.к.о.	$M(k)$	с.к.о.	$M(e)$	с.к.о.	$M(k)$	с.к.о.	$M(e)$	с.к.о.	$M(k)$	с.к.о.
Baart	nl=1e-5				nl=1e-7				nl=1e-9			
$e \mathbf{G}$	0.16	0.04	4.0	0.67	0.07	0.01	5.10	0.32	0.05	0.01	6.60	0.52
MDL \mathbf{G}	76.8	170	5.8	1.03	0.11	0.06	5.60	0.52	0.06	0.01	6.40	0.52
Ср \mathbf{G}	76.8	170	5.8	1.03	0.11	0.06	5.60	0.52	0.06	0.01	6.40	0.52
AIC \mathbf{G}	104	183	5.9	1.10	0.11	0.06	5.70	0.48	0.06	0.01	6.50	0.53
$e \mathbf{Q}$	0.07	0.00	5.0	0.00	0.06	0.00	6.00	0.00	0.04	0.00	7.00	0.00
GCV \mathbf{Q}	0.43	0.03	6.0	0.00	0.06	0.00	6.00	0.00	0.04	0.00	7.00	0.00
DSC \mathbf{Q}	0.20	0.11	4.0	0.00	0.07	0.02	5.00	0.00	–	–	–	–
MDL \mathbf{Q}	0.14	0.02	4.1	0.32	0.07	0.00	5.00	0.00	0.04	0.00	7.00	0.00
Ср \mathbf{Q}	0.14	0.00	4.0	0.00	0.07	0.00	5.00	0.00	0.04	0.00	7.00	0.00
AIC \mathbf{Q}	0.07	0.00	5.0	0.00	0.06	0.00	6.00	0.00	0.04	0.00	7.00	0.00
Spectr	nl=1e-1				nl=1e-2				nl=1e-6			
$e \mathbf{G}$	122	12.2	12.3	2.71	64.7	1.27	22.4	1.43	25.1	2.24	33.8	1.14
MDL \mathbf{G}	237	238	14.3	2.16	74.2	14.1	24.0	1.89	35.7	24.9	34.2	1.23
Ср \mathbf{G}	135	21.3	11.5	2.92	72.4	13.7	23.9	1.96	205	324	35.9	1.97
AIC \mathbf{G}	237	238	14.4	2.07	130	115	25.6	2.51	205	324	36.0	1.89
$e \mathbf{Q}$	86.9	5.92	16.3	1.06	57.2	1.34	27.9	0.88	22.2	0.62	35.7	0.67
GCV \mathbf{Q}	116	18.2	18.2	0.42	59.7	1.41	26.2	0.63	24.4	3.40	36.6	0.70
DSC \mathbf{Q}	133	8.40	7.70	2.06	67.6	2.64	20.1	0.74	27.7	2.08	32.5	0.71
MDL \mathbf{Q}	93.1	7.53	14.7	1.49	64.2	1.58	21.7	1.16	23.0	0.87	34.4	0.70
Ср \mathbf{Q}	108	11.5	12.1	1.73	64.9	0.07	20.9	0.74	23.2	0.92	34.3	0.67
AIC \mathbf{Q}	108	17.1	17.2	0.42	61.0	1.11	25.2	0.63	22.9	2.66	35.4	0.84

4.3. Анализ результатов для методов с проецированием. Анализ результатов показал, что в исследованных задачах качество решения с помощью псевдообратной матрицы после проецирования и выбора значения k по минимальному значению соответствующих критериев достигает качества решения регуляризации Тихонова. Рассмотрим детальнее поведение решения для конкретных задач.

Задачи Carasso, Phillips, Delves. Для предложенных проекционных методов минимальная ошибка, достигаемая при получении решения с помощью \mathbf{Q} -проекции, находится на уровне лучших результатов детерминированного метода (отличие 5%). Ошибка методов с \mathbf{G} -проекцией в среднем в 1,5–3 раза выше, чем с \mathbf{Q} -проекцией, однако эти ошибки на несколько порядков меньше, чем ошибка решения, полученная с помощью псевдообратной матрицы без проецирования.

Для \mathbf{G} -проекции среди КВПР лучше работает критерий DSC, который использует априорную информацию об уровне шума в \mathbf{y} . Неплохие результаты дает КВПР LC при «ручном» определении вершины угла L-кривой. Однако из-за значительного числа сбоев при автоматическом определении угла результаты по L-кривой не включены в таблицы. Среди КВМ лучшие результаты демонстриру-

ет S_p — они близки к минимальной достижимой ошибке e для G -проекции (на 10–50% хуже).

Для Q -проекции и КВПР эффективнее работает $GCV Q$. Среди КВМ лучшие результаты демонстрирует $AIC Q$ — они близки к минимальной достижимой ошибке для Q -проекции (ошибка AIC — от 0 до 20% хуже $e Q$) и немного лучше результатов GCV .

Задачи Vaart, Spectr. Спецификой задачи Vaart являются малые оптимальные значения k даже при малом уровне шума. Для задачи Spectr использован очень высокий уровень шума, и абсолютная ошибка решения велика. При таких особенностях КВПР для G -проекции нуждаются в дополнительной настройке, поэтому они не включены в табл. 3.

В задаче Vaart при $n_l = 1e-5$ кривые зависимости КВМ от k очень «крутые», и ошибка в определении оптимального k даже на 1–2 единицы приводит к большой ошибке решения для G -проекции. При уровнях шума $n_l = 1e-7$, $n_l = 1e-9$ все три критерия работают хорошо и показывают близкие значения. Q -проекция улучшает результаты для всех критериев. Для КВПР лучший результат дает DSC ; для КВМ AIC — результаты на уровне детерминированных методов.

В задаче Spectr для G -проекции и КВМ эффективные результаты получены для критерия S_p . Для Q -проекции и КВПР лучшие результаты — для GCV (от 5 до 30% хуже минимальной ошибки), а для КВМ — примерно одинаковые результаты для всех критериев, при меньших уровнях шума — порядка минимальной ошибки, при больших — ошибка до 10–20% хуже минимальной.

В целом, для G -проекции значения k и ошибки решения, близкие к оптимальным, дает кривая для S_p , а для Q -проекции — AIC . Результаты для Q -проекций лучше, чем для G -проекций. Результаты для метода решения с проецированием значительно эффективнее, чем для традиционного метода псевдообращения.

В случае, когда велика ошибка решения с проецированием, достигаемая при минимуме критерия, анализ показывает, что это часто происходит из-за неверного «срабатывания» критерия при некоторой реализации случайной матрицы, когда выбранное значение k значительно отличается от оптимального (выбросы). Обычно это наблюдается для критериев, зависимость которых от k пологая, и G -проекции, при которой график имеет «изрезанный» вид. На рис. 3 это проиллюстрировано примерами зависимостей КВМ и КВПР от k для G -проекции, при которых получена наибольшая ошибка решения для разных задач и уровней шума.

Повышения точности определения оптимального k можно ожидать при сглаживании графиков и выборе в качестве оптимального k значения в первом локальном минимуме вместо глобального. Такие модификации необходимы также для инкрементных методов определения псевдооптимального k . При Q -проекции правая ветвь графика становится более крутой, а график — более плавным, что приводит к более точному определению k и достижению минимальной ошибки.

На рис. 3 приведены зависимости для Q -проекций, иллюстрирующие улучшение результатов. Так, для задачи Phillips и уровня шума $1e-5$ для КВПР GCV при G -проекции $k = 61$ и $e = 0,28$ против истинного значения $k = 21$ и $e = 0,017$, а при Q -проекции $k = 23$ и $e = 0,0058$ в отличие от истинного значения $k = 24$ и $e = 0,0053$ (рис. 3, а); для шума $1e-2$ в задаче Spectr и КВМ AIC для G -проекции $k = 68$ и $e = 3,27E+09$ при истинном значении $k = 23$ и $e = 63,4$, а для Q -проекции $k = 26$ и $e = 61,17$ против истинного $k = 29$ и $e = 55,45$ (рис. 3, б); для шума $1e-3$ в задаче Carasso и КВМ MDL для G -проекции $k = 21$ и $e = 6,87$ при истинном $k = 9$ и $e = 1,69$, а для Q -проекции $k = 14$ и $e = 0,55$ при истинном $k = 16$ и $e = 0,53$ (рис. 3, в); для шума $1e-5$ в задаче Vaart и КВМ S_p при G -проекции $k = 7$ и $e = 60,31$ в отличие от истинного $k = 5$ и $e = 0,086$, а при Q -проекции $k = 4$ и $e = 0,14$ против истинного значения $k = 5$ и $e = 0,0665$ (рис. 3, г).

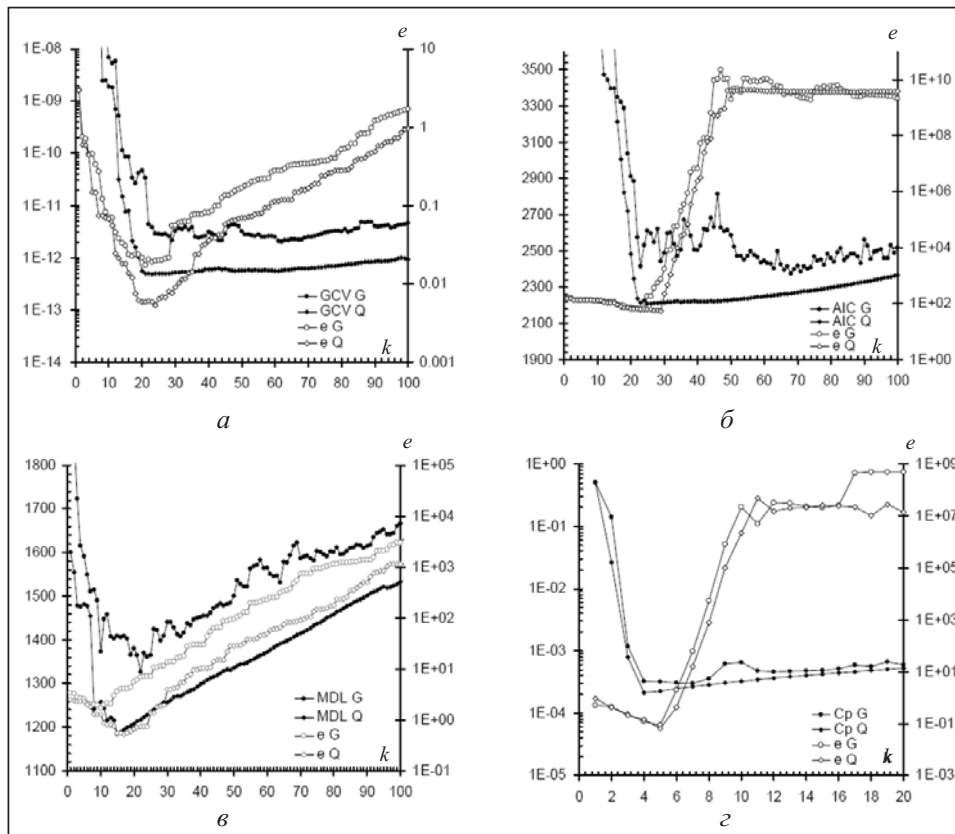


Рис. 3. Графики зависимостей значений ошибки e (правая ось Y) и критериев выбора модели (левая ось Y) от k для G -проекции и Q -проекции: Phillips, GCV, шум $1e-5$ (а); Spectr, AIC, $1e-2$ (б); Carasso, gMDL, $1e-3$ (в); Vaart, Cp, $1e-5$ (г)

4.4. Быстродействие. На рис. 4 приведено время вычислений для различных методов в зависимости от k . Как и ожидалось, решение на основе псевдообращения без случайной проекции занимает постоянное и довольно большое время по сравнению с решениями с использованием случайной проекции при малых значениях k . Например, при $k=30$ время вычисления со случайным проецированием ($\sim 0,003$ s для G - и Q -проекции) в 20 раз меньше, чем для псевдообращения без случайного проецирования (0,06 s).

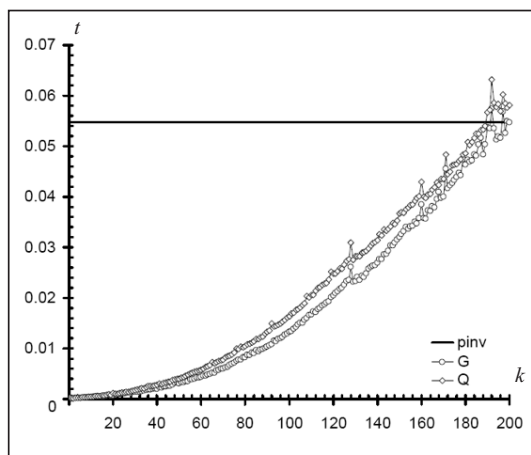


Рис. 4. График зависимости времени вычислений от размерности k случайной проекционной матрицы: без проецирования (pinv), с G -проецированием и Q -проецированием

Уменьшение времени вычисления объясняется тем, что после проецирования разложение по сингулярным значениям осуществляется для результирующей матрицы ($n \times k$), где k составляет малую долю n исходной матрицы F ($n \times n$). С увеличением уровня шума положение минимума ошибки смещается в область меньших значений k . Это отражает тот факт, что зависимость со-

ставляющей смещения ошибки $\|e_1\|$ от k практически одинакова для всех уровней шума, в то время как составляющая дисперсии ошибки $\|e_2\|$ возрастает с повышением уровня шума. Возможность выигрыша по времени увеличивается при более высоких уровнях шума, когда оптимальное k мало.

ЗАКЛЮЧЕНИЕ

Предложен подход к решению дискретных некорректных задач с использованием случайных проекций. В отличие от современных проекционных подходов к повышению вычислительной эффективности решения задачи наименьших квадратов с хорошо обусловленной матрицей A и незашумленным вектором правой части, в которых точность решения возрастает при увеличении размерности случайной проекционной матрицы, в предлагаемом подходе имеется оптимальная промежуточная размерность, при которой достигается минимум ошибки решения.

При оптимальном k (меньшей размерности проекционной матрицы) ошибка решения методом псевдообращения на несколько порядков меньше, чем ошибка решения методом псевдообращения без проецирования, и находится на уровне ошибки решения методами регуляризации Тихонова.

Для выбора k , близкого к оптимальному, на основе анализа поведения составляющих ошибки решения — смещения и дисперсии — предложено использовать два типа критериев: критерии выбора параметра регуляризации Тихонова и критерии выбора моделей.

С увеличением уровня шума величина ошибки решения увеличивается для всех методов. Однако для предложенного метода значение размерности k , при котором достигается минимум ошибки, уменьшается. Это создает возможность сокращения вычислительных затрат на получение решения. Разложение по сингулярным значениям SVD для получения псевдообратной матрицы, с помощью которой вычисляется решение, выполняется для результирующей $(n \times k)$ -матрицы после проецирования. Поэтому, когда k составляет малую долю n , вычислительные затраты на SVD уменьшаются по сравнению с затратами для исходной матрицы $(n \times n)$. Выигрыш увеличивается при большом уровне шума, когда оптимальное k мало.

Таким образом, изучение и применение методов решения дискретных некорректных задач, использующих регуляризацию на основе псевдообращения со случайным проецированием, является перспективным подходом благодаря уменьшению вычислительных затрат, а также устойчивости и точности решения. Направления дальнейших исследований в этой области:

- подбор параметра регуляризации для регуляризации Тихонова по критериям выбора модели;
- создание методов вычислительно эффективного выбора оптимальной размерности проекционной матрицы на основе априорной информации о решении, например гладкости либо неотрицательности решения;
- разработка высокопроизводительной аппаратной реализации для приложений, требующих реального времени обработки, с использованием эффективных систолических архитектур.

СПИСОК ЛИТЕРАТУРЫ

1. Hansen P. C. Rank-deficient and discrete ill-posed problems. Numerical aspects of linear inversion. — Philadelphia: SIAM, 1998. — 247 p.

2. Забулонов Ю.Л., Коростиль Ю.М., Ревунова Е.Г. Оптимизация решения обратной задачи по восстановлению функции плотности распределения поверхностных загрязнений // Сб. науч. тр. ИПМЭ НАН Украины «Моделирование и информационные технологии». — 2006. — С. 77–83.
3. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. — М.: Наука, 1979. — 285 с.
4. Морозов В.А. Регулярные методы решения некорректно поставленных задач. — М.: Наука, 1987. — 239 с.
5. Engl H.W., Hanke M., Neubauer A. Regularization of inverse problems. — Dordrecht: Kluwer Acad. Publ., 2000. — 321 p.
6. Мисуно И.С., Рачковский Д.А., Слипченко С.В. Векторные и распределенные представления, отражающие меру семантической связи слов // *Мат. машины и системы*. — 2005. — № 3. — С. 50–67.
7. Рачковский Д.А., Мисуно И.С., Слипченко С.В. Рандомизированные проекционные методы формирования бинарных разреженных векторных представлений // *Кибернетика и системный анализ*. — 2012. — № 1. — С. 176–188.
8. Ревунова Е.Г., Рачковский Д.А. Повышение точности решения обратной задачи с использованием случайных проекций // 16-th Intern. Conf. Knowledge-Dialogue-Solution. — 2009. — **10**. — P. 95–102.
9. Ревунова Е.Г. Исследование составляющих ошибки для решения обратной задачи с использованием случайных проекций // *Мат. машины и системы*. — 2010. — № 4. — С. 33–42.
10. Drineas P., Mahoney M.W., Muthukrishnan S., Sarlos T. Faster least squares approximation: Tech. Rep. // arXiv: 0710.1435. — 2007. — 21 p.
11. Hansen P.C., O’Leary D.P. The use of L-curve in the regularization of discrete ill-posed problems // *SIAM J. Sci. Comput.* — 1993. — **14**. — P. 1487–1503.
12. Golub G.H., Heath M., Wahba G. Generalised cross-validation as a method for choosing a good ridge parameter. *Technometrics*. — 1979. — **21**(2). — P. 215–223.
13. Halko N., Martinsson P.G., Tropp J.A. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions // *ACM Rep. Caltech*. — 2009. — N 5. — P. 1–82.
14. Boutsidis C., Drineas P., Mahoney M.W. An improved approximation algorithm for the column subset selection problem // STOC’09: Proc. 41st Ann. ACM Symp. Theory of Computing. — 2009. — P. 968–977.
15. Drineas P., Mahoney M.W., Muthukrishnan S. Sampling algorithms for l2 regression and applications // Proc. of the 17th ACM-SIAM Symp. on Discrete Algorithms (SODA). — 2006. — P. 1127–1136.
16. Johnson W.B., Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space // *Contemporary Mathematics*. — 1984. — **26**. — P. 189–206.
17. Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S. Latent semantic indexing: A probabilistic analysis // *J. Comput. System Sci.* — 2000. — **61**. — P. 217–235.
18. Achlioptas D. Database-friendly random projections: Johnson–Lindenstrauss with binary coins // *Ibid.* — 2003. — **66**(4). — P. 671–687.
19. Li P., Hastie T.J., Church K.W. Very sparse random projections // 12th ACM SIGKDD Intern. Conf. on Knowledge discovery and data mining. — Philadelphia: ACM Press, 2006. — P. 287–296.
20. Sarlos T. Improved approximation algorithms for large matrices via random projections // Proc. of the 47th Annual IEEE Symp. on Foundations of Computer Sci. — 2006. — P. 143–152.
21. Rokhlin V., Tygert M. A fast randomized algorithm for overdetermined linear least-squares regression // *PNAS*. — 2008. — **105**(36). — P. 13212–13217.

22. Tygert M. A fast algorithm for computing minimal-norm solutions to underdetermined systems of linear equations: Tech. rep. N09-48 // arXiv:0905.4745. — May 2009. — 13 p.
23. Niyogi P., Girosi F. On the relationship between generalization error, hypothesis complexity and sample complexity for radial basis functions // Neural Comput. — 1996. — **8**. — P. 819–842.
24. Hansen P. C. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems // Numer. Algorithms. — 1994. — **6**. — P. 1–35.
25. Akaike H. A new look at the statistical model identification // IEEE Trans. Automatic Control. — 1974. — **19**, N 6. — P. 716–723.
26. Mallows C.L. Some comments on C_p // Technometrics. — 1973. — **15**, N 4. — P. 661–675.
27. Hansen M., Yu B. Model selection and minimum description length principle // J. Amer. Statist. Assoc. — 2001. — **96**. — P. 746–774.
28. Sugiyama M., Ogawa H. Subspace information criterion for model selection // Neural Comput. — 2001. — **13**, N 8. — P. 1863–1889.
29. Rissanen J. Lectures on statistical modeling theory. — Tampere: Univ. of Technology, 2002. — P. 1–65. (<http://www.cs.tut.fi/~rissanen/>)
30. Rachkovskij D.A., Revunova E.G. Intelligent gamma-ray data processing for environmental monitoring // Intelligent Data Processing in Global Monitoring for Environment and Security. — Kiev–Sofia: ITHEA, 2011. — P. 136–157.
31. Забулонов Ю.Л., Лисиченко Г.В., Ревунова Е.Г. Стохастическая регуляризация обратной задачи восстановления параметров поля радиоактивности по данным аэро-гамма съемки // Сб. науч. тр. ИПМЭ НАН Украины «Проблемы моделирования в энергетике». — 2009. — Вып. 52. — С. 89–96.
32. Belge M., Kilmer M.E., Miller E.L. Efficient determination of multiple regularization parameters in a generalized L-curve framework // Inverse Problems. — 2002. — **18**. — P. 1161–1183.

Поступила 14.07.2010