УДК 519.614.2

ДВУСТОРОННИЙ МЕТОД НЬЮТОНА ДЛЯ ВЫЧИСЛЕНИЯ СПЕКТРАЛЬНЫХ ПРОЕКТОРОВ

К. В. Демьянко[1], **Ю. М. Нечепуренко**[2]

Предложен и обоснован эффективный метод ньютоновского типа для вычисления спектрального проектора, отвечающего подмножеству собственных значений большой разреженной матрицы, ближайших к заданной точке комплексной плоскости и отделенных от остальной части ее спектра. Обсуждаются результаты численных экспериментов с дискретным аналогом неэрмитового эллиптического оператора.

Ключевые слова: метод Ньютона, обратные итерации, тюнинг, инвариантное подпространство, спектральный проектор.

1. Введение. Настоящая статья посвящена вычислению спектрального проектора [1], отвечающего изолированному подмножеству собственных значений большой разреженной матрицы $A \in \mathbb{C}^{n\times n}$, ближайших к заданной точке комплексной плоскости. Без потери общности будем предполагать, что матрица A невырожденная и требуется найти спектральный проектор, отвечающий ее $p \ll n$ (с учетом кратности) минимальным по модулю собственным значениям. В противном случае можно рассматривать вместо исходной матрицы A матрицу $A-\sigma I$, где σ — некоторый сдвиг. Здесь и далее I означает единичную матрицу соответствующего порядка.

Пусть $\mathcal{X}_1$ и $\mathcal{X}_2 \subset \mathbb{C}^n$ — соответственно правое и левое инвариантные подпространства, отвечающие p минимальным по модулю собственным значениям матрицы A, а X_1 и $X_2 \in \mathbb{C}^{n\times p}$ — матрицы, столбцы которых образуют биортогональные базисы в этих подпространствах, т.е. $\mathcal{X}_l = \operatorname{span}(X_l)$ и

$$X_2^* X_1 = I. \tag{1}$$

Тогда имеем $AX_1 = X_1\Lambda$ и $A^*X_2 = X_2\Lambda^*$, где $\Lambda = X_2^* A X_1$. Спектр $\lambda(\Lambda)$ матрицы Λ состоит из p минимальных по модулю собственных значений матрицы A, а спектральный проектор (его также называют проектором Риса, или инвариантным проектором), отвечающий p минимальным по модулю собственным значениям матрицы A, может быть представлен в виде

$$\mathcal{P} = X_1 X_2^*. \tag{2}$$

Проектор (2) обычно не формируют явно, а используют (например, для спектральной редукции) в указанном выше малоранговом виде, т.е. как произведение двух прямоугольных матриц.

Матрицы X_1 и X_2, удовлетворяющие равенству (1), мы далее для краткости будем называть биортогональными. Для фиксированных подпространств $\mathcal{X}_1$ и $\mathcal{X}_2$ биортогональные базисы в этих подпространствах определены неоднозначно, однако проектор (2) от выбора базисов не зависит. Тем не менее, с точки зрения вычислительной устойчивости матрицы X_1 и X_2 предпочтительнее выбирать специальным образом, а именно так, чтобы $X_2^*X_2 = X_1^*X_1$, что обеспечивает равенство норм

$$\|\mathcal{P}\|_2 = \|X_1\|_2^2 = \|X_2\|_2^2. \tag{3}$$

Такие биортогональные матрицы мы будем называть сбалансированно биортогональными.

Для вычисления матриц X_1 и X_2 в данной работе предлагается использовать комбинацию двустороннего метода обратных итераций и метода Ньютона. Первый метод описан в разделе 2 и является модификацией метода обратных итераций с тюнингом, предложенного в работах [2–4]. Пусть матрицы $X_{1,0}, X_{2,0} \in \mathbb{C}^{n\times p}$ удовлетворяют условию биортогональности. Тогда метод двусторонних обратных итераций строит две последовательности матриц $X_{1,k}$ и $X_{2,k}$, $k = 1, 2, \ldots$, приближенно решая на каждом шаге

[1] Московский физико-технический институт (МФТИ), факультет проблем физики и энергетики, Институтский переулок, 9, 141700, Московская обл., г. Долгопрудный; аспирант, e-mail: kirill.demyanko@yandex.ru

[2] Институт вычислительной математики РАН, ул. Губкина, д. 8, 119333, Москва; вед. науч. сотр., e-mail: yumnech@yandex.ru

блочные системы вида $AX_{1,k+1} = X_{1,k}$ и $A^*X_{2,k+1} = X_{2,k}$ с помощью обобщенного метода минимальных невязок GMRES (Generalized Minimal RESidual [5]) с правым предобусловливанием и тюнингом; после решения этих систем выполняется биортогонализация найденных матриц $X_{1,k+1}$ и $X_{2,k+1}$.

После нескольких шагов метода двусторонних обратных итераций найденные приближенные инвариантные подпространства берутся в качестве начального приближения для двустороннего метода Ньютона, который выведен в разделе 3 по аналогии с методом Ньютона, предложенным и обоснованным в работах [6, 7]. Результаты численных экспериментов с предложенными нами двусторонними алгоритмами обсуждаются в разделе 4.

Отметим, что необходимые для нахождения спектрального проектора правое и левое инвариантные подпространства матрицы A могут быть вычислены и другими методами, такими как метод Арнольди, несимметричный метод Ланцоша или метод Якоби–Дэвидсона [8]. Однако предложенный двусторонний метод Ньютона, имея высокую скорость сходимости, является более простым, чем упомянутые выше методы и требует меньше дополнительной памяти. Так, для вычисления инвариантного подпространства методом Арнольди или несимметричным методом Ланцоша необходимо использовать их более дорогие блочные варианты, а для метода Якоби–Дэвидсона нужна процедура исчерпывания [9].

При описании предложенных в данной работе алгоритмов мы будем использовать две вспомогательные процедуры. Первая процедура выполняет ортонормировку столбцов заданной полноранговой матрицы $W \in \mathbb{C}^{n\times r}$, вычисляя ее QR-разложение $W = QR$, где $Q \in \mathbb{C}^{n\times r}$ — унитарная прямоугольная матрица, такая, что $Q^*Q = I$, а $R \in \mathbb{C}^{r\times r}$ — верхняя треугольная матрица [10]. Результат работы этой процедуры мы будем записывать в виде $(Q, R) = \operatorname{ort}(W)$ либо $Q = \operatorname{ort}(W)$, если матрица R не требуется.

Вторая процедура выполняет биортогонализацию столбцов двух заданных полноранговых матриц W_1 и $W_2 \in \mathbb{C}^{n\times r}$ на основе сингулярного разложения $W_2^*W_1 = UDV^*$, где U и V — унитарные матрицы, а D — неотрицательно определенная диагональная матрица порядка r [10]. Вычислив U, V и D, эта процедура вычисляет биортогональные матрицы $V_1 = W_1VD^{-1/2}$ и $V_2 = W_2UD^{-1/2}$, если матрица D невырожденная, либо констатирует, что биортогонализация невозможна. Успешный результат работы этой процедуры мы будем записывать в виде $(V_1, V_2, U, D, V) = \operatorname{biort}(W_1, W_2)$ или, если матрицы U, V и D далее не требуются, — в виде $(V_1, V_2) = \operatorname{biort}(W_1, W_2)$. Для определенности будем предполагать далее, что диагональные элементы матрицы D упорядочены в порядке невозрастания.

Для получения из заданных матриц W_1 и W_2 сбалансированно биортогональных матриц достаточно перед биортогонализацией ортогонализовать каждую из них: $(V_1, V_2) = \operatorname{biort}\big(\operatorname{ort}(W_1), \operatorname{ort}(W_2)\big)$. Нетрудно проверить, что построенные таким образом матрицы V_1 и V_2 будут удовлетворять следующим равенствам: $\operatorname{span}(V_l) = \operatorname{span}(W_l)$, $V_l^*V_l = D^{-1}$, $V_2^*V_1 = I$.

Проектор (2), где X_1 и X_2 биортогональные матрицы, является спектральным проектором, отвечающим некоторому подмножеству собственных значений матрицы A, в том и только том случае, если он коммутирует с матрицей A. Поэтому в критериях останова предлагаемых двусторонних методов мы будем следить за величиной нормы коммутатора

$$E = A\mathcal{P} - \mathcal{P}A \tag{4}$$

либо за нормами спектральных невязок

$$R_1 = AX_1 - X_1\Lambda, \quad R_2 = A^*X_2 - X_2\Lambda^*, \tag{5}$$

где $\Lambda = X_2^*AX_1$, которые связаны с E равенствами $E = R_1X_2^* - X_1R_2^*$, $R_1 = EX_1$, $R_2 = -E^*X_2$. Если матрицы X_1 и X_2 сбалансированно биортогональные, то в силу (3) имеем: $\|R_l\|_2 \leqslant \|E\|_2\|\mathcal{P}\|_2^{1/2}$, $\|E\|_2 \leqslant \big(\|R_1\|_2 + \|R_2\|_2\big)\|\mathcal{P}\|_2^{1/2}$, т.е. малость нормы коммутатора влечет малость норм спектральных невязок и наоборот.

Для вычисления нормы матрицы E саму эту матрицу формировать не требуется (это привело бы к огромным вычислительным затратам), а достаточно воспользоваться алгоритмом $(Q_l, N_l) = \operatorname{ort}\big([R_l, X_l]\big)$, $\|E\|_2 = \left\| N_1 \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} N_2^* \right\|_2$, где матрицы N_l — верхние треугольные порядка $2p$. Заметим, что формировать матрицы Q_l также не требуется, а это сокращает вычислительные затраты. Кроме того, используя этот алгоритм, мы можем с небольшими дополнительными вычислительными затратами найти и нормы невязок, поскольку $\|R_l\|_2 = \big\|\widetilde{N}_l\big\|_2$, где $\widetilde{N}_l$ — главная подматрица порядка p матрицы N_l.

2. Двусторонний метод обратных итераций. Для нахождения достаточно точного начального приближения для двустороннего метода Ньютона предлагается использовать

Алгоритм 1 (двусторонний метод обратных итераций). *Задать $\varepsilon > 0$, $\rho > 0$, $\eta > 0$ и сбалансированно биортогональные матрицы $X_{1,0}$, $X_{2,0} \in \mathbb{C}^{n\times p}$. Для $k = 0, 1, \ldots$*

1) *вычислить* $\Lambda_k = X_{2,k}^* A X_{1,k}$ *и спектральные невязки* $R_{1,k} = AX_{1,k} - X_{1,k}\Lambda_k$, $R_{2,k} = A^* X_{2,k} - X_{2,k}\Lambda_k^*$, *проверить сходимость коммутатора* $E_k = R_{1,k}X_{2,k}^* - X_{1,k}R_{2,k}^*$: *если* $\|E_k\|_2 \leqslant \varepsilon$, *положить* $X_{1,\mathrm{out}} = X_{1,k}$, $X_{2,\mathrm{out}} = X_{2,k}$ *и остановиться;*

2) *решить относительно* $X_{1,k+1}$ *и* $X_{2,k+1}$ *системы* $AX_{1,k+1} = X_k$ *и* $A^* X_{2,k+1} = X_{2,k}$ *приближенно, так, чтобы*

$$\|X_{1,k} - AX_{1,k+1}\|_2 \leqslant \gamma_{1,k}, \quad \|X_{2,k} - A^* X_{2,k+1}\|_2 \leqslant \gamma_{2,k}, \tag{6}$$

где $\gamma_{l,k} = \min\{\rho, \eta\|R_{l,k}\|_2\}$;

3) *положить* $(X_{1,k+1}, X_{2,k+1}) = \mathrm{biort}\,(X_{1,k+1}, X_{2,k+1})$.

На втором шаге алгоритма 1 необходимо решить две блочные системы линейных уравнений. Эти системы предлагается решать по столбцам с помощью метода GMRES, используя для ускорения вычислений правое предобусловливание: $AP_{1,k}^{-1}Z_{1,k} = X_{1,k}$, $X_{1,k+1} = P_{1,k}^{-1}Z_{1,k}$, $A^*P_{2,k}^{-1}Z_{2,k} = X_{2,k}$, $X_{2,k+1} = P_{2,k}^{-1}Z_{2,k}$. Анализ, аналогичный проведенному в [3], показывает, что количество итераций метода GMRES не будет зависеть от k, если в качестве критерия останова использовать (6) и выбирать $P_{1,k} = P + (A-P)X_{1,k}X_{2,k}^*$, $P_{2,k} = P^* + (A-P)^* X_{2,k}X_{1,k}^*$, где P — некоторое приближение (например, неполное LU-разложение [5]) матрицы A. В этом случае

$$(P_{1,k} - A)X_{1,k} = 0 \Rightarrow AP_{1,k}^{-1}(P_{1,k} - A)X_{1,k} = \big(I - AP_{1,k}^{-1}\big)AX_{1,k} = 0;$$

следовательно, $\big(I - AP_{1,k}^{-1}\big)X_{1,k} = -\big(I - AP_{1,k}^{-1}\big)R_{1,k}\Lambda_{1,k}^{-1} \to 0$ при $k \to \infty$ и $R_{1,k} \to 0$. В частности, это означает, что, решая систему $AP_{1,k}^{-1}Z_{1,k} = X_{1,k}$ относительно $Z_{1,k}$, в качестве начального приближения разумно выбрать правую часть, т.е. матрицу $X_{1,k}$. Аналогичное утверждение справедливо и для системы $A^*P_{2,k}^{-1}Z_{2,k} = X_{2,k}$.

Пусть разложение

$$P = \Pi^{-1}LU \approx A \tag{7}$$

является неполным LU-разложением матрицы A, где L и U соответственно нижняя и верхняя треугольные матрицы, а Π — матрица перестановок. Используя формулу Вудбери [10] и тот факт, что для матрицы перестановок справедливо равенство $\Pi^* = \Pi^{-1}$, можно вывести следующие формулы, удобные для умножения на вектор предобусловливателей:

$$P_{1,k}^{-1} = U^{-1}(I + V_{1,k}W_{1,k})L^{-1}\Pi, \quad P_{2,k}^{-1} = \Pi^* L^{-*}(I + V_{2,k}W_{2,k})U^{-*}, \tag{8}$$

где $W_{1,k} = X_{2,k}^* U^{-1}$, $W_{2,k} = X_{1,k}^* \Pi^* L^{-*}$, $V_{l,k} = \widetilde{V}_{l,k}\big(I - W_{l,k}\widetilde{V}_{l,k}\big)^{-1}$, $\widetilde{V}_{1,k} = UX_{1,k} - L^{-1}\Pi AX_{1,k}$ и $\widetilde{V}_{2,k} = L^*\Pi X_{2,k} - U^{-*}A^*X_{2,k}$.

Отметим, что на третьем шаге алгоритма 1 выполняется лишь несбалансированная биортогонализация матриц $X_{l,k+1}$, поскольку балансировка увеличивала бы суммарное количество итераций GMRES, необходимых для достижения требуемой точности. Этот достаточно тонкий эффект объясняется тем, что выбранный нами способ несбалансированной биортогонализации (с использованием сингулярного разложения) в случае, когда искомые собственные значения существенно отличаются по абсолютным величинам, превращает обратные итерации в алгоритм, в котором после каждой внешней итерации выполняется замена базисов в найденных приближенных правом и левом инвариантных подпространствах, делающая базисными приближенные правые и левые собственные векторы соответственно. В результате на следующей внешней итерации алгоритма 1 системы, отвечающие достаточно хорошо сошедшимся приближенным собственным векторам (обычно это векторы, отвечающие минимальным по модулю собственным значениям), решаются с требуемой точностью за 1–2 итерации GMRES благодаря тюнингу. Более подробное обоснование алгоритма 1, играющего в данной работе вспомогательную роль, будет опубликовано отдельно.

3. Двусторонний метод Ньютона. Пусть $X_l \in \mathbb{C}^{n\times p}$ $(l = 1, 2)$ — сбалансированно биортогональные матрицы, столбцы которых образуют базисы в соответственно правом и левом приближенных инвариантных подпространствах матрицы A, достаточно близких к точным, отвечающим ее p минимальным по модулю собственным значениям, отделенным от остальной части спектра, и $\mathcal{P}$ — проектор (2). Тогда будут малы по норме коммутатор E и спектральные невязки R_1 и R_2, определенные в (4) и (5) соответственно.

Рассмотрим матрицы

$$\widetilde{X}_l = X_l - \Phi_l, \tag{9}$$

где $\Phi_l \in \mathbb{C}^{n\times p}$ $(l = 1, 2)$ — решения системы уравнений

$$(I - \mathcal{P})(A\Phi_1 - \Phi_1\Lambda - R_1) = 0, \quad \mathcal{P}\Phi_1 = 0, \quad (I - \mathcal{P})^*(A^*\Phi_2 - \Phi_2\Lambda^* - R_2) = 0, \quad \mathcal{P}^*\Phi_2 = 0. \tag{10}$$

Можно показать, что при $\tau = \|E\|_2 \to 0$ матрица $\widetilde{X}_2^*\widetilde{X}_1$ невырожденная и проектор $\widetilde{\mathcal{P}} = \widetilde{X}_1\big(\widetilde{X}_2^*\widetilde{X}_1\big)^{-1}\widetilde{X}_2^*$ коммутирует с матрицей A с точностью τ^2, т.е.

$$\widetilde{E} = A\widetilde{\mathcal{P}} - \widetilde{\mathcal{P}}A = \mathcal{O}\big(\tau^2\big). \tag{11}$$

Следовательно, формулы (9), (10) описывают шаг ньютоновского уточнения исходных приближенных инвариантных подпространств $\operatorname{span}(X_1)$ и $\operatorname{span}(X_2)$.

Действительно, p минимальных по модулю собственных значений матрицы A отделены от остальной части спектра, матрицы X_1 и X_2 сбалансированно биортогональные и подпространства $\operatorname{span}(X_1)$ и $\operatorname{span}(X_2)$ стремятся при $\tau \to 0$ к соответственно правому и левому инвариантным подпространствам, отвечающим этим собственным значениям. Поэтому существует такое достаточно малое $\tau_0 > 0$, что при $0 \leqslant \tau < \tau_0$ уравнения (10), представляющие собой уравнения Сильвестра относительно Φ_1 и Φ_2, однозначно разрешимы и справедливы оценки $\|\Phi_l\|_2 \leqslant c\|R_l\|_2$, $l = 1, 2$, с некоторой константой c, зависящей только от A, p и τ_0.

Это означает, что, поскольку $R_l = \mathcal{O}(\tau)$, справедливо асимптотическое равенство $\Phi_l = \mathcal{O}(\tau)$ и, следовательно,

$$\widetilde{X}_2^*\widetilde{X}_1 = I + \mathcal{O}\big(\tau^2\big), \tag{12}$$

т.е. матрицы $\widetilde{X}_1$ и $\widetilde{X}_2$ — биортогональные с точностью τ^2. Более того, учитывая, что $\mathcal{P}A = A\mathcal{P} + \mathcal{O}(\tau)$, $(I - \mathcal{P})R_1 = R_1$, $(I - \mathcal{P})^*R_2 = R_2$, из (10) следуют асимптотические равенства $R_1 = A\Phi_1 - \Phi_1\Lambda + \mathcal{O}\big(\tau^2\big)$, $R_2 = A^*\Phi_2 - \Phi_2\Lambda^* + \mathcal{O}\big(\tau^2\big)$. Кроме того, $\Lambda X_2^* = X_2^*AX_1X_2^* = X_2^*A + \mathcal{O}(\tau)$, $\quad X_1\Lambda = X_1X_2^*AX_1 = AX_1 + \mathcal{O}(\tau)$. Поэтому $E = R_1X_2^* - X_1R_2^* = A\Phi_1X_2^* - \Phi_1X_2^*A - X_1\Phi_2^*A + AX_1\Phi_2^* + \mathcal{O}\big(\tau^2\big)$ и

$$A\widetilde{X}_1\widetilde{X}_2^* - \widetilde{X}_1\widetilde{X}_2^*A = \mathcal{O}\big(\tau^2\big). \tag{13}$$

Из (12) и (13) следует (11).

Опишем теперь алгоритм решения системы уравнений (10). Пусть $\Lambda = QTQ^*$ — разложение Шура [10], где Q — унитарная, а T — верхняя треугольная матрицы. Умножив каждое из уравнений в (10) справа на матрицу Q и сделав замену переменных $\Phi_l := \Phi_lQ$ и $R_l := R_lQ$, получим систему уравнений точно такого же вида, что и (10), но с верхней треугольной матрицей T вместо Λ. Решив полученную систему относительно Φ_l, выполняем обратное преобразование: $\Phi_l := \Phi_lQ^*$.

Обозначим через Φ_{lj} и R_{lj} соответственно j-е столбцы матриц Φ_l и R_l, а через t_{ij} — (i, j)-й элемент матрицы T. В этих обозначениях система (10) с верхней треугольной матрицей T вместо Λ может быть записана в виде

$$(I - \mathcal{P})\,(A - t_{jj}I)\,\Phi_{1j} = \Omega_{1j}, \quad \mathcal{P}\Phi_{1j} = 0, \quad j = 1, 2, \ldots, p, \tag{14}$$

$$(I - \mathcal{P})^*\,(A - t_{jj}I)^*\,\Phi_{2j} = \Omega_{2j}, \quad \mathcal{P}^*\Phi_{2j} = 0, \quad j = 1, 2, \ldots, p, \tag{15}$$

где

$$\Omega_{11} = (I - \mathcal{P})R_{11}, \qquad \Omega_{1j} = (I - \mathcal{P})\left(R_{1j} + \sum_{i=1}^{j-1} t_{ij}\Phi_{1i}\right), \qquad j > 1,$$

$$\Omega_{2p} = (I - \mathcal{P})^*R_{2p}, \quad \Omega_{2j} = (I - \mathcal{P})^*\left(R_{2j} + \sum_{i=j+1}^{p} \overline{t}_{ji}\Phi_{2i}\right), \quad j < p.$$

Первые уравнения систем (14) и (15) будем решать, используя правое предобусловливание:

$$H_l\Gamma_{lj} = \Omega_{lj}, \quad \Phi_{lj} = L_l\Gamma_{1j}, \tag{16}$$

где $H_1 = (I - \mathcal{P})\,(A - t_{jj}I)\,L_1$, $H_2 = (I - \mathcal{P})^*\,(A - t_{jj}I)^*\,L_2$, а L_l — некоторые предобусловливающие матрицы. Применяя GMRES к (16), мы найдем приближенное решение $\widehat{\Gamma}_{lj}$, удовлетворяющее неравенству

$$\left\|\Omega_{lj} - H_l\widehat{\Gamma}_{lj}\right\|_2 \leqslant \delta\,\|R_l\|_2\,, \tag{17}$$

где δ — заданная точность. Из вторых уравнений в (14) и (15) следует, что матрицы L_1 и L_2 должны удовлетворять равенствам $L_1 = (I - \mathcal{P})L_1$, $L_2 = (I - \mathcal{P})^*L_2$. С другой стороны, на каждом шаге GMRES поправка к начальному значению принадлежит крыловскому подпространству, сгенерированному матрицей H_l, которая при $l = 1$ удовлетворяет равенству $H_1 = (I - \mathcal{P})H_1$, а при $l = 2$ — равенству

$H_2 = (I-\mathcal{P})^* H_2$, с начальной невязкой $r_l^0 = \Omega_{lj} - H_l \widehat{\Gamma}_{lj}^0$, которая при $l=1$ удовлетворяет равенству $r_1^0 = (I-\mathcal{P}) r_1^0$, а при $l=2$ — равенству $r_2^0 = (I-\mathcal{P})^* r_2^0$. Это означает, что поправка при $l=1$ принадлежит подпространству $(I-\mathcal{P})\mathbb{C}^n$, а при $l=2$ — подпространству $(I-\mathcal{P})^*\mathbb{C}^n$. Следовательно, результат не изменится (в точной арифметике), если мы будем использовать матрицу $L_1(I-\mathcal{P})$ вместо L_1 и матрицу $L_2(I-\mathcal{P})^*$ вместо L_2. Таким образом, наиболее общий вид предобусловливающих матриц L_1 и L_2 — это $L_1 = (I-\mathcal{P})\widetilde{L}_1(I-\mathcal{P})$, $L_2 = (I-\mathcal{P})^*\widetilde{L}_2(I-\mathcal{P})^*$, где $\widetilde{L}_l$ — некоторые $n \times n$ матрицы, и проблема выбора предобусловливающих матриц L_l сводится к выбору матриц $\widetilde{L}_l$. В описанных в следующем разделе численных экспериментах мы использовали $\widetilde{L}_1 = P^{-1}$ и $\widetilde{L}_2 = P^{-*}$, где P — неполное LU-разложение (7) матрицы A, т.е.

$$L_1 = (I-\mathcal{P})U^{-1}L^{-1}\Pi(I-\mathcal{P}), \quad L_2 = L_1^*. \tag{18}$$

Отметим, что для наибольшей эффективности описанного алгоритма диагональные элементы формы Шура T для уравнения (14) должны быть расположены в порядке неубывания их абсолютных величин, а для уравнения (15) — в порядке невозрастания. Обычно разложение Шура заданной матрицы выполняется с помощью стандартного численного программного обеспечения в два этапа. Сначала, используя QR-алгоритм, находят разложение Шура, где диагональные элементы формы Шура расположены в некотором, априори непредсказуемом порядке. Затем с помощью отдельной процедуры из найденного разложения Шура получают разложения Шура с заданным расположением диагональных элементов формы Шура. Второй этап требует значительно меньших вычислительных затрат. Учитывая это, после решения системы (14) и перед решением системы (15) требуется переупорядочить диагональные элементы формы Шура.

Приведем формальное описание предлагаемого алгоритма в целом.

Алгоритм 2 (двусторонний метод Ньютона). *Задать $\varepsilon > 0$, $\delta > 0$ и сбалансированно биортогональные матрицы $X_{1,0}, X_{2,0} \in \mathbb{C}^{n\times p}$. Для $k = 0, 1, \ldots$*

1) *вычислить $\Lambda_k = X_{2,k}^* A X_{1,k}$ и спектральные невязки $R_{1,k} = AX_{1,k} - X_{1,k}\Lambda_k$, $R_{2,k} = A^* X_{2,k} - X_{2,k}\Lambda_k^*$; проверить сходимость коммутатора $E_k = R_{1,k}X_{2,k}^* - X_{1,k}R_{2,k}^*$: если $\|E_k\|_2 \leqslant \varepsilon$, положить $X_{1,\mathrm{out}} = X_{1,k}$, $X_{2,\mathrm{out}} = X_{2,k}$ и остановиться;*

2) *вычислить разложение Шура $\Lambda_k = Q_k T_k Q_k^*$ с диагональными элементами формы Шура T_k, расположенными в порядке неубывания их абсолютных величин; положить $R_{1,k} := R_{1,k}Q_k$; решить приближенно систему $(I-\mathcal{P}_k)(A\Phi_{1,k} - \Phi_{1,k}T_k) = (I-\mathcal{P}_k)R_{1,k}$, $\mathcal{P}_k\Phi_{1,k} = 0$, с $\mathcal{P}_k = X_{1,k}X_{2,k}^*$ относительно $\Phi_{1,k}$ и положить $\Phi_{1,k} := \Phi_{1,k}Q_k^*$;*

3) *используя ранее найденное разложение Шура, вычислить разложение Шура $\Lambda_k = Q_k T_k Q_k^*$ с диагональными элементами формы Шура T_k, расположенными в порядке невозрастания их абсолютных величин; положить $R_{2,k} := R_{2,k}Q_k$; решить приближенно систему $(I-\mathcal{P}_k)^*(A^*\Phi_{2,k} - \Phi_{2,k}T_k^*) = (I-\mathcal{P}_k)^*R_{2,k}$, $\mathcal{P}_k^*\Phi_{2,k} = 0$ относительно $\Phi_{2,k}$ и положить $\Phi_{2,k} := \Phi_{2,k}Q_k^*$;*

4) *выполнить ньютоновский шаг и сбалансированную биортогонализацию:*

$$(X_{1,k+1}, X_{2,k+1}) = \mathrm{biort}\,(\mathrm{ort}\,(X_{1,k} - \Phi_{1,k}), \mathrm{ort}\,(X_{2,k} - \Phi_{2,k})).$$

4. Численные эксперименты. В качестве тестовой задачи рассматривался эллиптический оператор

$$\frac{\partial}{\partial x}u + \frac{\partial}{\partial y}v + \mu\Delta \tag{19}$$

в области $(x,y) \in (0,1)\times(0,1)$ с граничными условиями типа Дирихле, где Δ — двумерный оператор Лапласа, $\mu = 5\times 10^{-4}$, а

$$u = \frac{\partial}{\partial y}\varphi, \quad v = -\frac{\partial}{\partial x}\varphi, \quad \varphi(x,y) = \frac{1}{4\pi}\cos\bigl(2\pi x^2\bigr)\cos\bigl(2\pi y^2\bigr). \tag{20}$$

Аппроксимация выполнялась методом центральных разностей второго порядка точности на равномерных сетках с одинаковым числом внутренних узлов $m = 200$, 300 или 400 по каждому направлению. После аппроксимации при каждом m получалась матрица A порядка $n = m^2$, являющаяся дискретным аналогом оператора (19), для которой вычислялся спектральный проектор $\mathcal{P}$ на ее правое инвариантное подпространство, отвечающее $p = 8$ минимальным по модулю собственным значениям. Матрицы $X_{1,0}$ и $X_{2,0}$ для алгоритма 2 вычислялись алгоритмом 1, стартовавшим со случайного, но фиксированного для каждого m подпространства размерности 8. Этот этап мы будем называть препроцессингом.

Блочные линейные системы в алгоритмах 1 и 2 решались приближенно с помощью GMRES с правым предобусловливанием. В предобусловливателях (8) и (18) использовалось неполное LU-разложение с параметром отсечения, равным 10^{-3}. Соответствующие количества ненулевых элементов в треугольных множителях L, U и матрице A в зависимости от m указаны в табл. 1. Максимальная размерность крыловского подпространства во всех экспериментах полагалась равной 50. В алгоритме 1 для критерия останова (6) параметры ρ и η всегда выбирались равными 10^{-4} и 10^{-2} соответственно. Для алгоритма 2 в критерии останова (17) использовались $\delta = 10^{-1}$, 10^{-2}, 10^{-3} или 10^{-4}.

Таблица 1
Количество ненулевых элементов nnz в матрицах A, L и U

m	$nnz(A)$	$nnz(L)$	$nnz(U)$
200	199 200	660 885	655 287
300	448 800	1 488 521	1 470 368
400	798 400	2 467 964	2 437 090

В качестве стартовых векторов для GMRES в алгоритме 1 выбирались столбцы матриц $X_{l,k}$, а в алгоритме 2 — нулевые векторы. Итерации GMRES останавливались либо при выполнении соответствующего критерия останова, либо при достижении максимальной размерности крыловского подпространства. В последнем случае предусматривался рестарт. Эксперименты показали, что количество итераций GMRES, необходимое для решения систем, не превышало 33, т.е. выполнять рестарты в ходе вычислений не потребовалось.

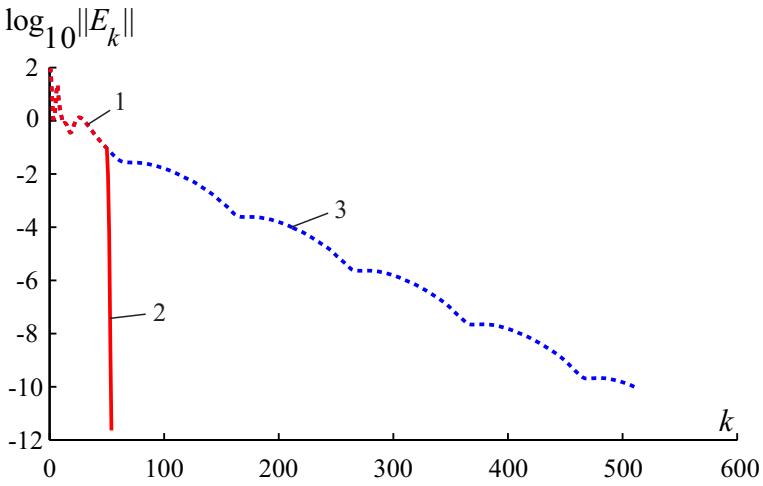

Рис. 1. Зависимость нормы коммутатора E_k от номера итерации k при $m = 200$. Алгоритм 1 — пунктиры 1 и 3, алгоритм 2 — сплошная линия 2

В численных экспериментах уточнение приближенного спектрального проектора прекращалось, когда норма коммутатора становилась меньше $\varepsilon = 10^{-10}$. В первом эксперименте сравнивалась скорость сходимости алгоритма 1 в чистом виде со скоростью сходимости комбинации алгоритмов 1 и 2 с $\delta = 10^{-4}$, где первый алгоритм использовался в качестве препроцессинга для второго. На рис. 1 представлена зависимость нормы коммутатора E_k от номера k итерации. Видно, что вычисление спектрального проектора с помощью комбинации алгоритмов 1 (красный пунктир 1) и 2 (красная сплошная линия 2) дает существенное ускорение по сравнению с использованием только алгоритма 1 (синий пунктир 3).

Таблица 2
Вычислительные затраты комбинации алгоритмов 1 и 2 при $\delta = 10^{-4}$

m	ε_{SI}	k_{SI}	$\|E_0\|_2$	S_{SI}	k_{NM}	$\|E_k\|_2$	S_{NM}	S_T
200	10^{-1}	50	9.20×10^{-2}	3890	4	2.28×10^{-12}	540	4430
	10^{-2}	112	9.87×10^{-3}	7616	3	1.98×10^{-12}	429	8045
	10^{-3}	149	9.55×10^{-4}	9604	2	7.29×10^{-12}	226	9830
300	10^{-1}	33	9.56×10^{-2}	3484	4	9.30×10^{-12}	692	4176
	10^{-2}	105	9.85×10^{-3}	8174	3	6.08×10^{-12}	561	8735
	10^{-3}	138	9.99×10^{-4}	10164	2	1.37×10^{-11}	302	10466
400	10^{-1}	59	9.92×10^{-2}	6219	3	5.26×10^{-11}	624	6843
	10^{-2}	97	9.22×10^{-3}	9001	3	3.37×10^{-11}	693	9694
	10^{-3}	168	9.63×10^{-4}	13699	2	3.83×10^{-11}	369	14068

В табл. 2 демонстрируются вычислительные затраты комбинации алгоритмов 1 и 2 при $\delta = 10^{-4}$. Для различных m и ε_{SI} в этой таблице указано (i) количество k_{SI} итераций алгоритма 1, необходимых для получения начальных матриц $X_{l,0}$ для алгоритма 2, для которых $\|E_0\|_2 \leqslant \varepsilon_{SI}$, (ii) количество k_{NM} итераций алгоритма 2 и суммарные количества S_{SI} и S_{NM} итераций GMRES, которые были выполнены в алгоритмах 1 и 2 соответственно, и (iii) $S_T = S_{SI} + S_{NM}$. Из этой таблицы видно, что хотя при

более точном начальном приближении метод Ньютона сходится за меньшее количество итераций k_{NM}, вычислительные затраты препроцессинга оказываются слишком большими, что в итоге приводит к росту общего числа S_T итераций GMRES. Таким образом, нет необходимости искать начальное приближение с высокой точностью. Выбор $\varepsilon_{SI} = 10^{-1}$ является оптимальным с точки зрения вычислительных затрат.

Отметим, что как в алгоритме 1, так и в алгоритме 2 в критериях останова GMRES вместо норм невязок можно использовать норму коммутатора. Соответствующий дополнительный численный эксперимент дал почти те же результаты, что и приведенные в табл. 2.

В следующем эксперименте при $m = 300$ была исследована сходимость алгоритма 2 при различных значениях ε_{SI} и δ. Соответствующие результаты представлены в табл. 3. Помимо выводов, сделанных из табл. 2, из табл. 3 можно заключить, что скорость сходимости алгоритма 2 растет с уменьшением величины δ, т.е. при более точном решении уравнений Сильвестра на каждой внешней итерации алгоритма 2.

Это иллюстрирует также рис. 2, где изображена зависимость нормы коммутатора E_k от номера k итерации алгоритма 2 при различных δ и ε_{SI}. Видно, что с ростом δ скорость сходимости алгоритма 2 уменьшается с квадратичной до линейной. Среднее количество итераций GMRES, необходимых для решения одного уравнения Сильвестра, равное $S_{NM}/(2pk_{NM})$, немного растет с уменьшением δ, однако значение $\delta = 10^{-4}$ является оптимальным с точки зрения вычислительных затрат, поскольку число внешних итераций (число шагов k_{NM} алгоритма 2) с уменьшением δ уменьшается сильнее.

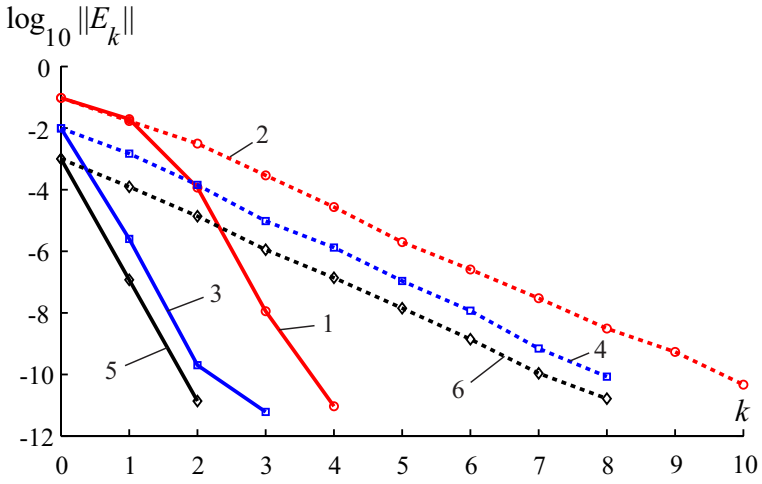

Рис. 2. Зависимость нормы коммутатора E_k от k при $m = 300$ и различных ε_{SI} и δ: 1) $\varepsilon_{SI} = 10^{-1}$, $\delta = 10^{-4}$; 2) $\varepsilon_{SI} = 10^{-1}$, $\delta = 10^{-1}$; 3) $\varepsilon_{SI} = 10^{-2}$, $\delta = 10^{-4}$; 4) $\varepsilon_{SI} = 10^{-2}$, $\delta = 10^{-1}$; 5) $\varepsilon_{SI} = 10^{-3}$, $\delta = 10^{-4}$; 6) $\varepsilon_{SI} = 10^{-3}$, $\delta = 10^{-1}$

Из табл. 3 видно, что при всех рассмотренных значениях δ величина $S_{NM}/(2pk_{NM})$ и максимальное количество $l_{\max}$ итераций GMRES при решении одного уравнения Сильвестра почти не зависели от ε_{SI}. Из этого можно сделать вывод, что при фиксированном δ вычислительные затраты одного шага алгоритма 2 не возрастают по мере его сходимости.

Таблица 3

Вычислительные затраты алгоритма 2 при $m = 300$

	$\varepsilon_{SI} = 10^{-1}$				$\varepsilon_{SI} = 10^{-2}$				$\varepsilon_{SI} = 10^{-3}$			
δ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-1}	10^{-2}	10^{-3}	10^{-4}
k_{NM}	10	6	5	4	8	4	3	3	8	4	3	2
S_{NM}	1074	830	814	692	800	490	451	561	807	498	451	302
$S_{NM}/(2pk_{NM}) \approx$	7	9	10	11	7	8	10	12	7	8	10	10
$l_{\max}$	18	21	24	26	18	21	24	27	18	22	24	26

5. Выводы. В данной работе предложен и обоснован эффективный двусторонний метод Ньютона (алгоритм 2) для вычисления спектрального проектора, отвечающего изолированному подмножеству собственных значений большой разреженной матрицы. В качестве начального приближения для этого метода используются приближенные правое и левое инвариантные подпространства, найденные с помощью нескольких шагов двустороннего метода обратных итераций, основанного на GMRES с тюнингом (алгоритм 1). В двустороннем методе Ньютона на каждой внешней итерации необходимо решать уравнения Сильвестра, для чего предложено использовать специальный алгоритм, основанный на разложении Шура и GMRES. Приведенные результаты численных экспериментов позволяют сделать следующие выводы:

— если имеется хорошее начальное приближение и уравнения Сильвестра решаются достаточно точно, то двусторонний метод Ньютона сходится квадратично;

— нет смысла вычислять начальное приближение со слишком высокой точностью, поскольку это приводит к сильному увеличению вычислительных затрат на этапе препроцессинга и, как следствие, к значительному росту суммарных вычислительных затрат;

— при наличии достаточно хорошего начального приближения вычислительные затраты одного шага предложенного двустороннего метода Ньютона не возрастают по мере его сходимости.

Работа выполнена при финансовой поддержке РФФИ, код проекта 13–01–00350 и программы РАН "Современные проблемы теоретической математики", проект "Оптимизация вычислительных алгоритмов решения задач математической физики".

СПИСОК ЛИТЕРАТУРЫ

1. *Stewart G.W., Sun J.-G.* Matrix perturbation theory. San Diego: Academic Press, 1990.
2. *Freitag M.A., Spence A.* A tuned preconditioner for inexact inverse iteration applied to Hermitian eigenvalue problems // IMA J. Numer. Anal. 2008. **28**. 522–551.
3. *Robbe M., Sadkane M., Spence A.* Inexact inverse subspace iteration with preconditioning applied to non-Hermitian eigenvalue problems // SIAM J. Matrix Anal. Appl. 2009. **31**. 92–113.
4. *Xue F., Elman H.C.* Fast inexact subspace iteration for generalized eigenvalue problems with spectral transformation // Linear Algebra Appl. 2011. **435**. 601–622.
5. *Saad Y.* Iterative methods for sparse linear systems. Boston: PWS Publishing, 1996.
6. *El Khoury G., Nechepurenko Yu.M., Sadkane M.* Acceleration of inverse subspace iteration with Newton's method // J. Comput. Appl. Math. 2014. **259**. 205–215.
7. *Годунов С.К., Нечепуренко Ю.М.* Оценки скорости сходимости метода Ньютона для вычисления инвариантных подпространств // Журн. вычислит. матем. и матем. физики. 2002. **42**, № 6. 771–779.
8. *Bai Z., Demmel J., Dongarra J., Ruhe A., van der Vorst H.* Templates for the solution of algebraic eigenvalue problems: a practical guide. Philadelphia: SIAM, 2000.
9. *Hechme G., Nechepurenko Yu.M., Sadkane M.* Efficient methods for computing spectral projectors for linearized Navier–Stokes equations // SIAM J. Sci. Comput. 2008. **31**. 667–686.
10. *Голуб Дж., Ван Лоун Ч.* Матричные вычисления. М.: Мир, 1999.

Поступила в редакцию
04.02.2014

Bi-Newton's method for computing spectral projectors

K. V. Demyanko [1] **and Yu. M. Nechepurenko** [2]

[1] *Moscow Institute of Physics and Technology, Faculty of Problems of Physics and Energetics; pereulok Institutskii 9, Dolgoprudnyi, 141700, Russia; Graduate Student, e-mail: kirill.demyanko@yandex.ru*

[2] *Institute of Numerical Mathematics, Russian Academy of Sciences; ulitsa Gubkina 8, Moscow, 119333, Russia; Ph.D., Leading Scientist, e-mail: yumnech@yandex.ru*

Received February 4, 2014

Abstract: An efficient Newton-like method for computing the spectral projector associated with a separated group of eigenvalues near a specified shift of a large sparse matrix is proposed and justified. A number of numerical experiments with a discrete analogue of the non-Hermitian elliptic operator are discussed.

Keywords: Newton's method, inverse iterations, tuning, invariant subspace, spectral projector.

References

1. G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory* (Academic Press, San Diego, 1990).

2. M. A. Freitag and A. Spence, "A Tuned Preconditioner for Inexact Inverse Iteration Applied to Hermitian Eigenvalue Problems," IMA J. Numer. Anal. **28**, 522–551 (2008).

3. M. Robbe, M. Sadkane, and A. Spence, "Inexact Inverse Subspace Iteration with Preconditioning Applied to Non-Hermitian Eigenvalue Problems," SIAM J. Matrix Anal. Appl. **31**, 92–113 (2009).

4. F. Xue and H. C. Elman, "Fast Inexact Subspace Iteration for Generalized Eigenvalue Problems with Spectral Transformation," Linear Algebra Appl. **435**, 601–622 (2011).

5. Y. Saad, *Iterative Methods for Sparse Linear Systems* (PWS Publ., Boston, 1996).

6. G. El Khoury, Yu. M. Nechepurenko, and M. Sadkane, "Acceleration of Inverse Subspace Iteration with Newton's Method," J. Comput. Appl. Math. **259**, 205–215 (2014).

7. S. K. Godunov and Yu. M. Nechepurenko, "Bounds for the Convergence Rate of Newton's Method for Calculating Invariant Subspaces," Zh. Vychisl. Mat. Mat. Fiz. **42** (6), 771–779 (2002) [Comput. Math. Math. Phys. **42** (6), 739–746 (2002)].

8. Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide* (SIAM, Philadelphia, 2000).

9. G. Hechme, Yu. M. Nechepurenko, and M. Sadkane, "Efficient Methods for Computing Spectral Projectors for Linearized Navier–Stokes Equations," SIAM J. Sci. Comput. **31**, 667–686 (2008).

10. G. H. Golub and C. F. Van Loan, *Matrix Computations* (The John Hopkins Univ. Press, Baltimore, 1996; Mir, Moscow, 1999).