

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**

Факультет прикладной математики и информатики

Республиканский конкурс научных работ студентов высших учебных
заведений Республики Беларусь

Математика. Методы и алгоритмы вычислительной математики
математического моделирования для решения задач экономики, техники и
природоведения

**ИТЕРАЦИОННЫЕ ПРОЦЕССЫ УСТАНОВЛЕНИЯ
ДЛЯ ЖЕСТКИХ ЗАДАЧ**

Бондарь Иван Васильевич,
магистрант

Фалейчик Борис Викторович,
доцент кафедры
вычислительной математики,
канд. физ.-мат. наук

Минск 2012

РЕФЕРАТ

Работа, 28 с., 24 рис., 8 источников.

ПРИНЦИП УСТАНОВЛЕНИЯ, ЖЁСТКИЕ ЗАДАЧИ, МЕТОД РУНГЕ-КУТТЫ, СОБСТВЕННЫЕ ЗНАЧЕНИЯ, СПЕКТРАЛЬНЫЙ РАДИУС

Объект исследования – методы решения жёстких задач.

Цель работы – разработка вычислительного алгоритма для решения жёстких дифференциальных задач на основе принципа установления.

Методы исследования – методы численного анализа.

Результатом является алгоритм решения жёстких дифференциальных задач, в основе которого лежит принцип установления.

Областью применения является решение задач математической физики.

Работа выполнена при поддержке гранта БРФФИ №Ф12МВ-006.

СОДЕРЖАНИЕ

	Стр.
РЕФЕРАТ.....	2
ВВЕДЕНИЕ.....	4
ГЛАВА 1 ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ	5
1.1 Линейный случай	5
1.2 Уравнение установления	6
1.3 Выбор вспомогательного метода Рунге-Кутты	7
1.4 Спектральные свойства линейных задач	10
1.5 Скорость сходимости итерационного процесса.....	13
1.6 Нелинейный случай	14
ГЛАВА 2 ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ ДЛЯ ЛИ-	
НЕЙНЫХ ЗАДАЧ	17
2.1 Задача с собственными значениями из сектора	17
2.2 Случай наличия близких к нулю действительных частей соб-	
ственных значений	23
2.3 Линейная задача с разреженной матрицей.....	25
ЗАКЛЮЧЕНИЕ	27
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	28

ВВЕДЕНИЕ

Жесткие задачи исследуются примерно со второй половины 20-го века. Однако до сих пор трудно сформулировать точное определение жесткости. Исторически первым было определение данное Кёртиссом и Хиршфельдером в 1952 году: *жесткие уравнения – это уравнения, для которых определенные неявные методы дают лучший результат, обычно несравнимо более хороший чем явные методы.* Хотя это определение не претендует на полноту и уж тем более на строгость, оно в достаточной мере характеризует проблему: большинство явных методов не приспособлены к решению жестких задач.

Более полным является определение данное Ламбертом: *если численный метод с ограниченной областью абсолютной устойчивости, примененный к системе с произвольными начальными условиями вынужден использовать на некотором интервале интегрирования величину шага, которая чрезмерно мала по отношению к гладкости точного решения на этом интервале, тогда говорят что система является жесткой на этом интервале.*

При применении неявного метода к линейной задаче возникает система линейных уравнений, которую можно решать, вообще говоря, любым методом решения линейных систем. Но следует учитывать специфику: во-первых, система как правило большой размерности, что исключает применение прямых методов решения в силу чрезмерной трудоемкости и возможности накопления вычислительной погрешности; во-вторых, у возникшей системы спектр как правило комплексный, что не даёт эффективно применять итерационные процессы типа Якоби. Остаются методы, основанные на подпространствах Крылова.

В нелинейном случае для реализации неявных методов очень часто применяются методы ньютоновского типа. Но их применение, как известно, также сопряжено с процедурой решения больших СЛАУ, что в случае больших систем неприемлемо ввиду излишней трудоемкости. К тому же, для сходимости метода Ньютона необходимо выбрать достаточно близкое начальное приближение, что также является не самой простой задачей.

Учитывая всё вышесказанное, хотелось бы получить алгоритм решения жестких задач, который бы не требовал обращения матриц, был применим в нелинейном случае, являясь одновременно простым и удобным для реализации. Описанию такого алгоритма и посвящена данная работа.

ГЛАВА 1

ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Линейный случай

Рассмотрим задачу Коши для неоднородной линейной системы обыкновенных дифференциальных уравнений (ОДУ)

$$\begin{aligned} y'(t) &= Jy(t) + f(t), \\ y(t_0) &= y_0, \\ t &\in [t_0, t_0 + \tau], \\ y_0 &\in \mathbb{R}^N, \quad y : [t_0, t_0 + \tau] \rightarrow \mathbb{R}^N, \\ J &\in \mathbb{R}^N \times \mathbb{R}^N, \quad \tau \in [0, +\infty). \end{aligned} \tag{1.1}$$

Для нахождения приближения к $y(t_0 + \tau)$, $\tau > 0$ проинтегрируем её произвольным s -стадийным неявным методом типа Рунге-Кутты (базовым методом), представленным следующей таблицей Бутчера

$$\begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1s} \\ \dots & \dots & \dots & \dots \\ c_s & a_{s1} & \dots & a_{ss} \\ \hline & b_1 & \dots & b_s \end{array} \tag{1.2}$$

Здесь $A = (a_{i,j})_{i,j=1}^s$ — так называемая матрица Бутчера базового метода. Тогда

$$y(t_0 + \tau) \approx y_1 = y_0 + \tau \sum_{i=1}^s b_i k_i,$$

где $\{k_i\}_{i=1}^s$ находятся как решение следующей системы линейных алгебраических уравнений (СЛАУ)

$$k_i = J(y_0 + \tau \sum_{j=1}^s a_{ij} k_j) + f(t_0 + c_i \tau).$$

В дальнейшем будем пользоваться матричной записью этой СЛАУ:

$$\begin{aligned} (\tau A \otimes J - I)k + g &= 0, \\ g &= (g_1, g_2, \dots, g_s)^T, \quad g_i = f(t_0 + c_i \tau) + Jy_0, \quad i = 1, \dots, s, \\ k &= (k_1, k_2, \dots, k_s)^T, \quad k_i \in \mathbb{R}^N. \end{aligned} \tag{1.3}$$

Здесь \otimes обозначает кронекеровское произведение матриц [1], по определению которого получаем, что

$$G = \tau A \otimes J - I$$

– блочная матрица вида

$$\begin{pmatrix} -1 + \tau a_{11}J & \tau a_{12}J & \dots & \tau a_{1s}J \\ \tau a_{12}J & -1 + \tau a_{22}J & \dots & \tau a_{2s}J \\ \dots & \dots & \dots & \dots \\ \tau a_{s1}J & \tau a_{s2}J & \dots & -1 + \tau a_{ss}J \end{pmatrix}. \quad (1.4)$$

1.2 Уравнение установления

Рассмотрим вспомогательное уравнение

$$k' = (\tau A \otimes J - I)k + g = Gk + g = r(k), \quad (1.5)$$

которое в дальнейшем будем называть уравнением установления.

Очевидно, что точное решение уравнения (1.3) k^* будет являться стационарным решением (1.5). Для этого достаточно, чтобы спектр матрицы G целиком содержался в левой комплексной полуплоскости. Поэтому, если проинтегрировать (1.5) каким-нибудь численным методом, то можно получить приближение к решению (1.3).

Для решения (1.5) будем использовать явный метод Рунге-Кутты, задаваемый таблицей вида

$$\begin{array}{ccccccc} \alpha_{21} & & & & & & \\ \alpha_{31} & \alpha_{32} & & & & & \\ \dots & \dots & \dots & & & & \\ \alpha_{\sigma 1} & \alpha_{\sigma 2} & \dots & \alpha_{\sigma \sigma - 1} & & & \\ \hline \beta_1 & \beta_2 & \dots & \beta_{\sigma - 1} & \beta_{\sigma} & & \end{array} \quad (1.6)$$

Пусть ω - шаг по фиктивному времени. В результате получаем итерационный процесс вида

$$\begin{aligned} k^{l+1} &= \Phi(k^l) \\ \Phi(k) &= k + \omega \sum_{p=1}^{\sigma} \beta_p K_p(k), \\ K_p(k) &= G(k + \omega \sum_{q=1}^{p-1} \alpha_{pq} K_q(k)) + g. \end{aligned} \quad (1.7)$$

Учитывая специфику интегрирования уравнения установления (1.5), нужно выбрать ω , $\{\alpha_{ij}\}_{i,j=1}^{\sigma}$, $\{\beta_i\}_{i=1}^{\sigma}$.

1.3 Выбор вспомогательного метода Рунге-Кутты

Для выбора коэффициентов вспомогательного метода заметим, что процесс (1.7) представим в виде

$$k^{l+1} = R_\sigma(\omega G)k^l + P(\omega, G), \quad (1.8)$$

где R_σ – многочлен степени σ , называемый многочленом перехода (функцией устойчивости), P – функция, точный вид которой несущественен в данном случае.

Известно, что многочлен перехода во многом определяет свойства устойчивости метода интегрирования ОДУ. В нашем же случае он определяет свойства сходимости итерационного процесса. Мы заинтересованы в том, чтобы вспомогательный метод был устойчив на как можно большей области. В частности, область устойчивости

$$S = \{z \in \mathbb{C} : |R_\sigma(z)| < 1\}$$

должна содержать в себе спектр матрицы ωG [2].

Запишем многочлен устойчивости в виде

$$R_\sigma(z) = 1 + \sum_{j=1}^{\sigma} a_j z^j,$$

где $z \in \mathbb{C}$. Коэффициенты $\{a_i\}_{i=1}^{\sigma}$, $a_i \in \mathbb{R}$ будем выбирать таким образом, чтобы

$$F(a_1, \dots, a_\sigma) = \int_0^1 \int_{\pi-\alpha}^{\pi+\alpha} |R_\sigma(\rho e^{i\varphi})|^2 d\varphi d\rho \longrightarrow \min. \quad (1.9)$$

Здесь α – некоторый заранее заданный угол, определяющий область устойчивости полинома R_σ [3],[4]. Вообще, α нужно выбирать исходя из представлений о спектре матрицы G , поскольку чем лучше будет приближен спектр, тем более эффективным будет метод. Параметр ω выбирается так, чтобы спектр матрицы ωG полностью содержался в области устойчивости. Для этого достаточно положить ω равным спектральному радиусу матрицы G . Примеры областей устойчивости полученных для значений угла α равных $\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{6}$ и $\sigma = 20$ можно увидеть на рисунке 1.1.

Метод, полученный с $\alpha = \frac{\pi}{2}$ будет применим для всех задач со спектром из левой комплексной полуплоскости, однако будет проигрывать по эффективности применимым методам с меньшими углами (см. Рис 1.2).

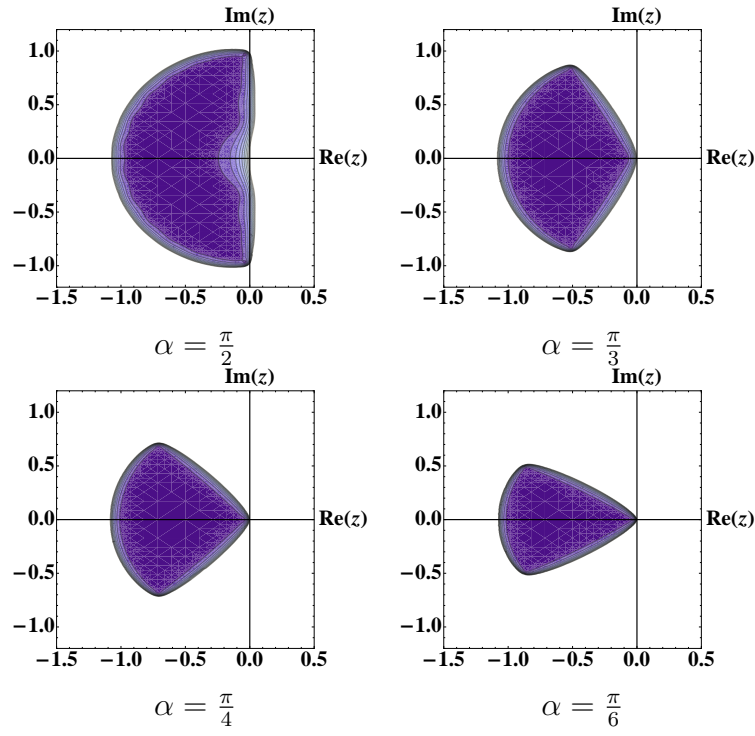


Рис. 1.1: Пример областей устойчивости.

Определим коэффициенты многочлена R_σ (здесь и далее полагаем $a_0 = 1$):

$$\begin{aligned}
|R_\sigma(\rho e^{i\varphi})|^2 &= \operatorname{Re}(R_\sigma(\rho e^{i\varphi}))^2 + \operatorname{Im}(R_\sigma(\rho e^{i\varphi}))^2 = \\
&= \left(1 + \sum_{j=1}^{\sigma} a_j \rho^j \cos(j\varphi)\right)^2 + \left(\sum_{j=1}^{\sigma} a_j \rho^j \sin(j\varphi)\right)^2 = \\
&= 1 + \sum_{p=1}^{\sigma} \rho a_p \cos(p\varphi) + \sum_{p=1}^{\sigma} \sum_{q=1}^{\sigma} \rho^{p+q} a_p a_q \cos(p\varphi) \cos(q\varphi) + \\
&\quad + \sum_{p=1}^{\sigma} \sum_{q=1}^{\sigma} \rho^{p+q} a_p a_q \sin(p\varphi) \sin(q\varphi) = \\
&= \left[a_p a_q \cos(p\varphi) \cos(q\varphi) + a_p a_q \sin(p\varphi) \sin(q\varphi) = \frac{1}{2} a_p a_q \cos((p-q)\varphi) \right] = \\
&= 1 + \sum_{p=1}^{\sigma} \rho a_p \cos(p\varphi) + \sum_{p=1}^{\sigma} \sum_{q=1}^{\sigma} \frac{1}{2} \rho^{p+q} a_p a_q \cos((p-q)\varphi)
\end{aligned}$$

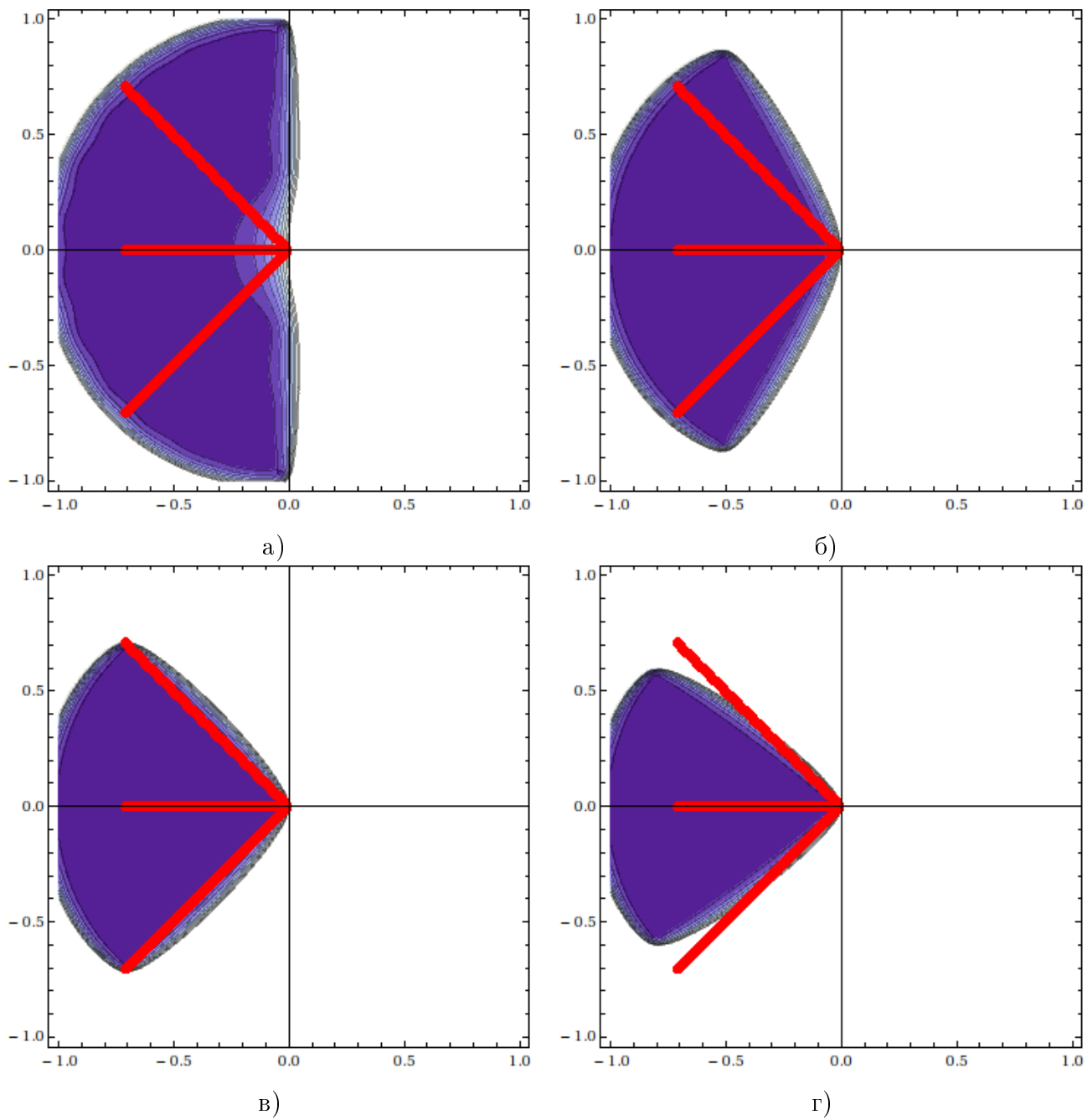


Рис. 1.2: Пример расположения спектра в области устойчивости вспомогательного метода: методы а, б, в применимы, причем в – наиболее эффективный; метод г не применим

Тогда:

$$F(a_1, \dots, a_\sigma) = \int_0^1 \int_{\pi-\alpha}^{\pi+\alpha} 1 + \sum_{p=1}^{\sigma} \rho a_p \cos(p\varphi) + \\ + \sum_{p=1}^{\sigma} \sum_{q=1}^{\sigma} \frac{1}{2} \rho^{p+q} a_p a_q \cos((p-q)\varphi) d\varphi d\rho \quad (1.10)$$

Чтобы найти минимум функционала (1.10), необходимо найти точку, в которой $\frac{\partial F}{\partial a_j} = 0, j = 1, \dots, \sigma$:

$$\frac{\partial F}{\partial a_j} = \sum_{q=1}^{\sigma} a_q \int_0^1 \int_{\pi-\alpha}^{\pi+\alpha} \rho^q \cos((q-j)\varphi) d\varphi d\rho + \int_0^1 \int_{\pi-\alpha}^{\pi+\alpha} \rho \cos(j\varphi) d\varphi d\rho = 0, \\ j = 1, \dots, \sigma. \quad (1.11)$$

Решив систему уравнений (1.11), мы и найдем искомые коэффициенты $\{a_i\}_{i=1}^{\sigma}$ определяющие многочлен $R(z)$.

По построенному множителю перехода восстановим таблицу явного метода, воспользовавшись способом описанным в [5], [6]: представим вспомогательный метод в виде суперпозиции двухстадийных и, возможно, одного одностадийного метода. Тогда

$$\Phi = \Phi_{\sigma'} \circ \Phi_{\sigma'-1} \circ \dots \circ \Phi_1, \quad \sigma' = \lceil \sigma/2 \rceil,$$

что соответствует разложению $R(z)$ на множители вида $(1 + \delta_m z)(1 + \delta'_m z)$, δ_m и $\delta'_m \in \mathbb{R}$ или δ_m и $\delta'_m \in \mathbb{C}$ и множитель $(1 + \gamma z)$, $\gamma \in \mathbb{R}$, появляющийся при нечётных σ .

При чётных σ все методы являются двухстадийными явными методами Рунге-Кутты:

$$\Phi_m(k) = k + \omega(\beta_{m,1}r(k) + \beta_{m,2}r(k + \omega\alpha_m r(k))),$$

где коэффициенты $\alpha_m, \beta_{m,1}, \beta_{m,2}$ удовлетворяют соотношению

$$(1 + \delta_m z)(1 + \delta'_m z) = 1 + (\beta_{m,1} + \beta_{m,2})z + \beta_{m,2}\alpha_m z^2.$$

При нечётных σ один из методов Φ_m представляет собой метод Эйлера:

$$\Phi_m(k) = k + \omega\gamma r(k)$$

1.4 Спектральные свойства линейных задач

Как было сказано ранее, определяющее влияние на сходимость итерационного процесса (1.7) имеет спектр матрицы ωG . Предположим, что нам известен спектр

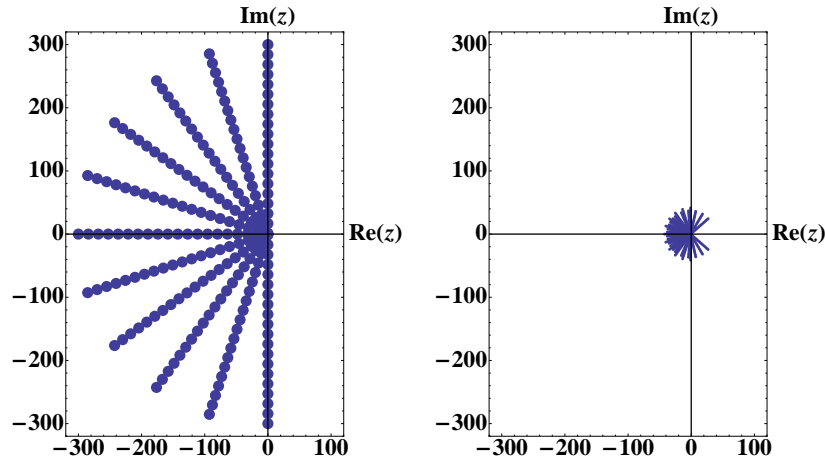


Рис. 1.3: Спектр исходной матрицы J (слева) и результирующей матрицы G (справа)

исходной матрицы J , и отследим каким может быть спектр результирующей матрицы

$$G = \tau(A \otimes J) - I.$$

По свойству кронекеровского произведения матриц [1], собственные значения $\nu_{i,j}$ матрицы G равны

$$\nu_{ij} = \tau\mu_j\lambda_i - 1, \quad (1.12)$$

где μ_i – собственные значения матрицы A , λ_j – собственные значения матрицы J . То есть, над спектром исходной матрицы системы (1.1) производятся операции масштабирования, поворота и параллельного переноса. Собственные значения матрицы ωG , очевидно, являются смасштабированными на единичный круг собственными значениями G . Схематически результат этих операции показан на рисунке 1.3. Здесь шаг по времени $\tau = 0.5$, в качестве матрицы A была взята матрица Бутчера неявного 3-х стадийного метода Радо IIA 5-го порядка:

$$A = \begin{pmatrix} \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{pmatrix} \quad (1.13)$$

Как видно из рисунка, описанные преобразования привели к выходу спектра матрицы G за пределы левой комплексной полуплоскости, что по определению плохо – процессы установления становится неприменимыми. Но даже если выход не произошел, нет гарантии что такая ситуация не возникнет при увеличении шага интегрирования по времени. На рисунке 1.4 продемонстрировано изменение спектра матрицы G при увеличении шага τ (в качестве исходной матрицы J взята такая же, как и на рисунке 1.3, матрица Бутчера базового метода – (1.13)).

Чтобы избежать описанного выше нежелательного явления, можно осуществить так называемую операцию переобусловливания [7],[8] умножив систему

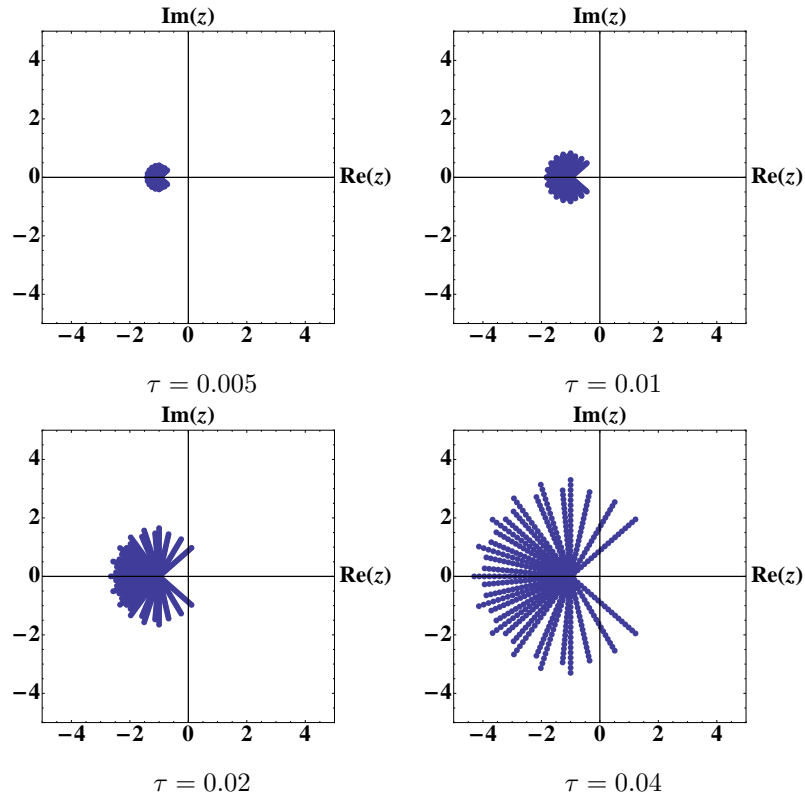


Рис. 1.4: Нежелательный выход за пределы левой комплексной полуплоскости при увеличении шага интегрирования τ .

(1.3) слева на $A^{-1} \otimes I$ (здесь $I \in \mathbb{R}^N \times \mathbb{R}^N$). Тогда система (1.3) примет вид:

$$\begin{aligned}
 (\tau I_s \otimes J - A^{-1} \otimes I_N)k + \tilde{g} &= 0, \\
 \tilde{g} = (\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_s)^T, \quad \tilde{g}_i &= (A^{-1} \otimes I_N)(f(t_0 + c_i \tau) + Jy_0), \quad i = 1, \dots, s, \\
 k = (k_1, k_2, \dots, k_s)^T, \quad k_i &\in \mathbb{R}^N,
 \end{aligned} \tag{1.14}$$

здесь I_s и I_N – единичные матрицы размерности $s \times s$ и $N \times N$ соответственно. Также стоит отметить, что точные решения (1.3) и (1.14) совпадают. По свойствам кронекеровского произведения, собственные значения матрицы

$$\tilde{G} = (\tau I_s \otimes J - A^{-1} \otimes I_N)$$

равны

$$\tilde{\nu}_{ij} = \tau \lambda_i - \frac{1}{\mu_j}.$$

Здесь отсутствует преобразование вращения, выполняются только масштабирование и параллельный перенос. Если все μ_j имеют положительные вещественные части (а это справедливо для большинства применяемых на практике методов Рунге–Кутты), то полностью отсутствует опасность выхода спектра матрицы \tilde{G} за пределы левой комплексной полуплоскости, что продемонстрировано на рис. 1.5.

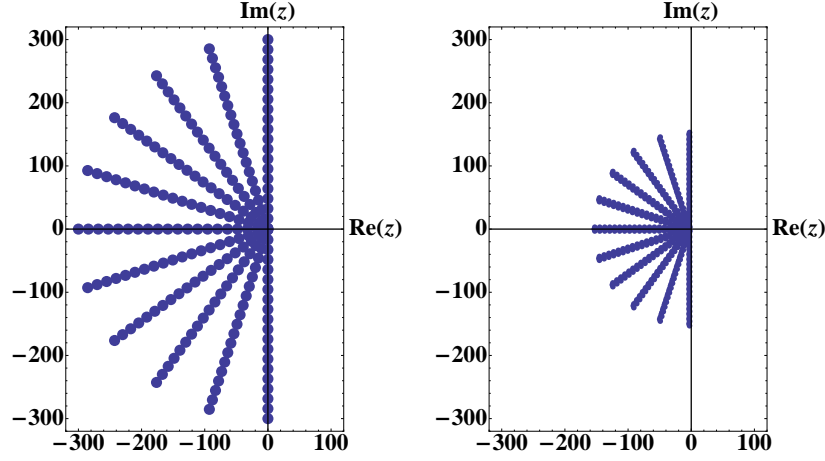


Рис. 1.5: Изменение спектра исходной матрицы (слева) в результате преобразований

Кроме вышеописанной пользы, операция переобусловливания также позволяет сократить объём вычислений. Поскольку

$$\tilde{G} = \begin{bmatrix} -\tilde{a}_{11}I_N + \tau J & -\tilde{a}_{12}I_N & \dots & -\tilde{a}_{1s}I_N \\ \tilde{a}_{12}I_N & -\tilde{a}_{22}I_N + \tau J & \dots & \tilde{a}_{2s}I_N \\ \dots & \dots & \dots & \dots \\ \tilde{a}_{s1}I_N & \tilde{a}_{s2}I_N & \dots & -\tilde{a}_{ss}I_N + \tau J \end{bmatrix}, \quad (1.15)$$

где $\{\tilde{a}_{ij}\}_{i,j=1}^s$ – элементы матрицы A^{-1} . Количество ненулевых элементов матрицы (1.15) в случае полной матрицы J в точности равно

$$n_{\tilde{G}} = sN^2 + (s^2 - s)N = s(N^2 + N(s - 1)),$$

в то время как у матрицы G $n_G = (sN)^2$. Таким образом

$$\frac{n_G}{n_{\tilde{G}}} = \frac{s^2 N^2}{s(N^2 + N(s - 1))} = \frac{s}{1 + (s - 1)/N} \rightarrow s, \quad N \rightarrow \infty,$$

что сулит выигрыш примерно в s раз при вычислениях и экономии памяти при хранении \tilde{G} в виде разреженной матрицы.

1.5 Скорость сходимости итерационного процесса

Проследим как будет изменяться ошибка на l -ой итерации: $\varepsilon^l = k^* - k^l$. Учитывая (1.8), получим:

$$\varepsilon^l = R_\sigma(\omega G)\varepsilon^{l-1}$$

Далее предположим, что у матрицы G имеется полный набор собственных векторов $\{\eta^i\}_{i=1}^N$. Тогда

$$\varepsilon^l = \sum_{i=1}^N \varepsilon_i^{l-1} R(\omega G) \eta^i = \sum_{i=1}^N \varepsilon_i^{l-1} R(\omega \nu_i) \eta^i,$$

здесь ν_i – собственное значение, соответствующее η^i . Таким образом,

$$\varepsilon_i^l = R_\sigma(\omega \nu_i) \varepsilon_i^{l-1} = (R(\omega \nu_i))^l \varepsilon_i^0.$$

Проанализируем последнее выражение. R_σ – многочлен перехода, и по построению $R_\sigma(0) = 1$. Учитывая непрерывность R_σ , получим что R_σ близок к 1 когда $\omega \nu_i$ близко к 0 (Рисунок 1.6). Это значит, что компоненты ошибки, соответствующие малым величинам $\omega \nu_i$, будут уменьшаться медленно. Нетрудно видеть, что характеристикой, в достаточной мере описывающей свойства сходимости, является так называемое “спектральное число обусловленности” матрицы J исходной системы, $\kappa = \rho(J)\rho(J^{-1})$ (ρ здесь – спектральный радиус). Чем больше эта величина, тем медленнее будет сходиться итерационный процесс.

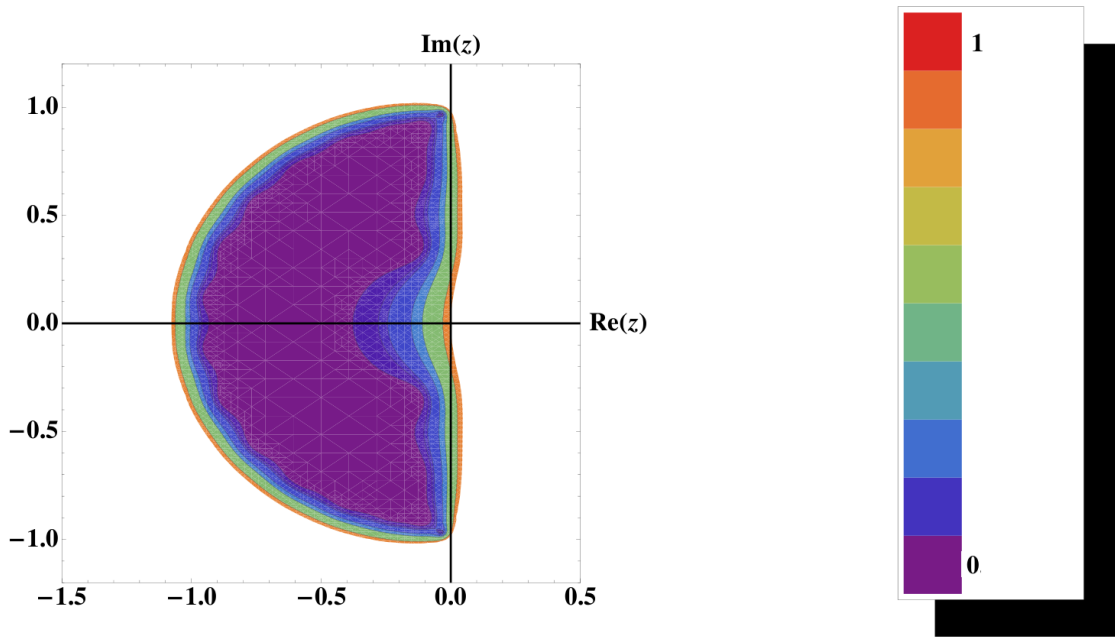


Рис. 1.6: Область устойчивости вспомогательного метода и линии уровня. Чем ближе к границе области устойчивости, тем ближе значения R_σ к 1.

1.6 Нелинейный случай

Рассуждения для нелинейного случая проходят во многом аналогично случаю линейному, поэтому остановимся только на различиях.

Рассмотрим систему нелинейных дифференциальных уравнений

$$\begin{aligned}
y'(t) &= f(t, y(t)), \\
y(t_0) &= y_0, \\
t &\in [t_0, t_0 + \tau], \quad \tau > 0 \\
y_0 &\in \mathbb{R}^N, \quad y : [t_0, t_0 + \tau] \rightarrow \mathbb{R}^N, \\
f &: \mathbb{R}^N \rightarrow \mathbb{R}^N.
\end{aligned} \tag{1.16}$$

Для интегрирования воспользуемся методом (1.2), причем в отличие от линейного случая применение запишем в симметричном виде:

$$\begin{aligned}
Y_i &= y_0 + \tau \sum_{j=1}^s a_{i,j} f(t_0 + c_j \tau, Y_j), \\
y(t_0 + \tau) &\approx y_1 = y_0 + \tau \sum_{j=1}^s b_j f(t_0 + c_j \tau, Y_j),
\end{aligned}$$

что в векторной форме представимо как

$$\begin{aligned}
Y &= e \otimes y_0 + \tau(A \otimes I)F(t_0, Y), \\
y_1 &= e \otimes y_0 + \tau(b^T \otimes I)F(t_0, Y).
\end{aligned} \tag{1.17}$$

Здесь $e = (1, \dots, 1)$, $e \in \mathbb{R}^s$, $Y = (Y_1, \dots, Y_s)^T$, $F(t, Y) = (f(t + c_1 \tau, Y_1), \dots, f(t + c_s \tau, Y_s))^T$. Уравнение установления для (1.17) имеет вид

$$Y(\theta)' = \tau(A \otimes I)F(t_0, Y(\theta)) - Y(\theta) + e \otimes y_0 = \tilde{r}(Y(\theta)). \tag{1.18}$$

Соответствующий ему процесс установления имеет вид аналогичный (1.7):

$$\begin{aligned}
Y^{l+1} &= \Phi(Y^l), \\
\Phi(Y) &= Y + \omega \sum_{p=1}^{\sigma} \beta_p K_p(Y), \\
K_p(Y) &= \tilde{r}(t_0, Y + \omega \sum_{q=1}^{p-1} \alpha_{pq} K_q(Y)).
\end{aligned} \tag{1.19}$$

Процесс (1.19) вообще говоря записать в форме (1.8) нельзя. Однако, если задача позволяет, можно провести линеаризацию и исследовать спектральные свойства уже для неё, полностью повторяя приведенные ранее рассуждения о конструировании вспомогательных методов.

Параметр ω полагаем таким, чтобы все собственные значения матрицы Якоби правой части уравнения (1.17) были по модулю меньше 1.

Переобусловливание происходит аналогично линейному случаю при помощи умножения на $A^{-1} \otimes I$. Результирующая система уравнений принимает вид

$$\sum_{j=1}^s \tilde{a}_{i,j} Y_j = \sum_{j=1}^s \tilde{a}_{i,j} y_0 + \tau f(t_0, Y_i), \quad i = 1, \dots, s.$$

Или в векторной форме

$$(A^{-1} \otimes I)Y = (A^{-1}e) \otimes y_0 + \tau F(t_0, Y). \quad (1.20)$$

Уравнение установления принимает вид

$$Y'(\theta) = (A^{-1}e) \otimes y_0 + \tau F(t_0, Y) - (A^{-1} \otimes I)Y(\theta).$$

ГЛАВА 2

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ ДЛЯ ЛИНЕЙНЫХ ЗАДАЧ

Эксперимент будем проводить следующим образом: зададим распределение собственных значений матрицы J в ком-нибудь секторе левой комплексной плоскости и применим к этой задаче неявный метод. Будем оценивать скорость реализации одного шага по времени. Для каждого эксперимента приведем время работы разработанного алгоритма и эти же параметры для известных методов GMRES и BICGSTAB, входящих в состав библиотеки deal.II [9].

В качестве базового метода использовали неявный метод Рунге-Кутты, с

$$A = \begin{pmatrix} \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{pmatrix},$$

$$b = \left(\frac{16-\sqrt{6}}{36}, \frac{16+\sqrt{6}}{36}, \frac{1}{9} \right),$$

$$c = \left(\frac{4-\sqrt{10}}{10}, \frac{4+\sqrt{10}}{10}, 1 \right),$$

порядок которого равен 5. Вычисления производятся пока норма невязки $Gk + g$ больше $\varepsilon = 10^{-4}$.

2.1 Задача с собственными значениями из сектора

Для первого эксперимента мы взяли область распределения собственных значений как сектор кольца с минимальным и максимальным радиусом $\rho_{min} = 1$ и $\rho_{max} = 300$ соответственно, $\alpha = \frac{\pi}{3}$ (рисунок 2.1). Начальные условия $y_0 = (1, \dots, 1)$, функция правой части константная, $f(t) = (5, \dots, 5)$. Матрицы J генерируются следующим образом: сначала на диагональ матрицы устанавливаются блоки, соответствующие комплексно-сопряженным парам собственных значений, а затем применяем преобразование отражения. Таким образом, матрицы являются полными. Также стоит отметить, что если матрица исходной системы имеет размерность M , то после применения неявного метода размерность задачи составит уже $N = s M$, где s - число стадий базового метода (во всех наших экспериментах оно равно 3). Далее под N будет подразумеваться размерность матрицы G .

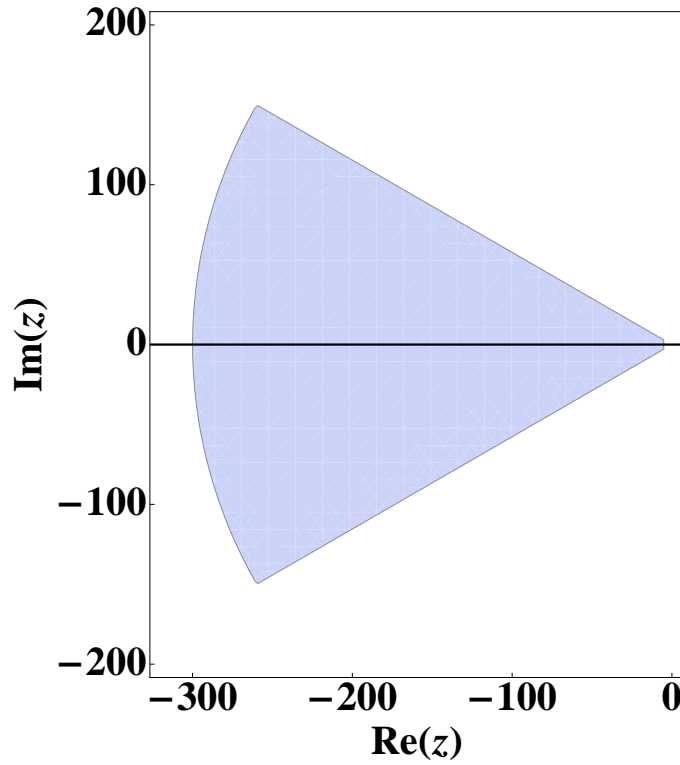


Рис. 2.1: Область распределения собственных значений матрицы J

Отследим эффективность различных алгоритмов при изменении только размерности, не затрагивая при этом основных спектральных характеристик ρ_{min} и ρ_{max} . Параметры эксперимента следующие: $\rho_{max} = 100$, $\rho_{min} = 1$, $\tau = 10$. Дополнительно для сравнения мы привели результаты метода, чья область устойчивости является сектором с углом $\alpha = \frac{\pi}{3}$. Результаты для не переобусловленного и переобусловленного случаев представлены на рисунках 2.2 и 2.3 соответственно.

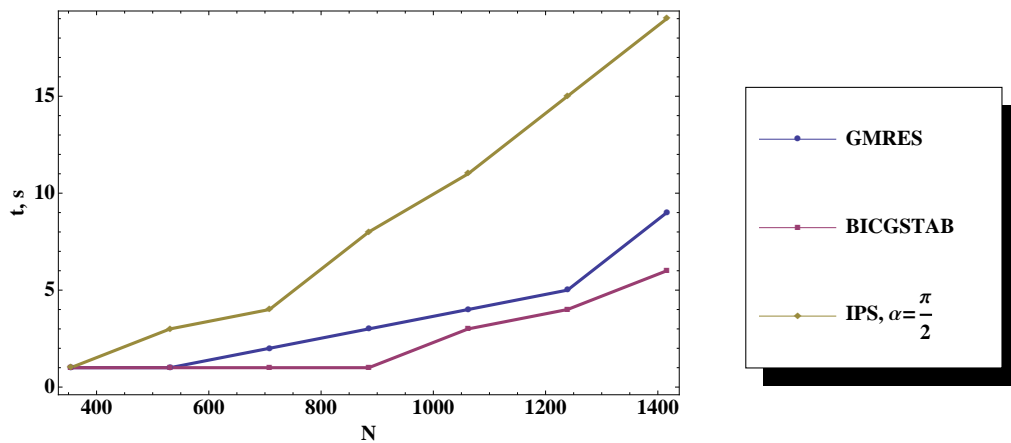


Рис. 2.2: Изменение трудоемкости с увеличением N (без переобусловливания)

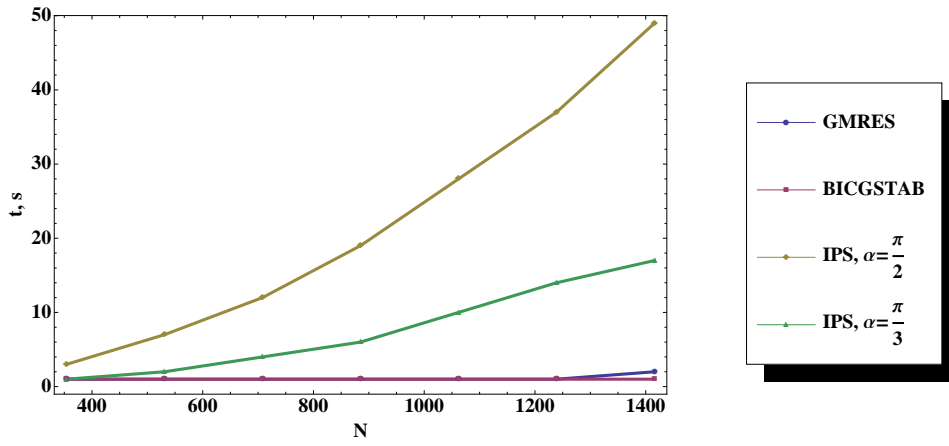


Рис. 2.3: Изменение трудоемкости с увеличением N (с переобусловливанием)

Положим теперь $N = 885$, оставляя остальные параметры численного эксперимента неизменными, и исследуем время решения каждым из приведенных алгоритмов при постепенном увеличении шага τ в случае с переобусловливанием (рис. 2.4) и без него (рис. 2.5).

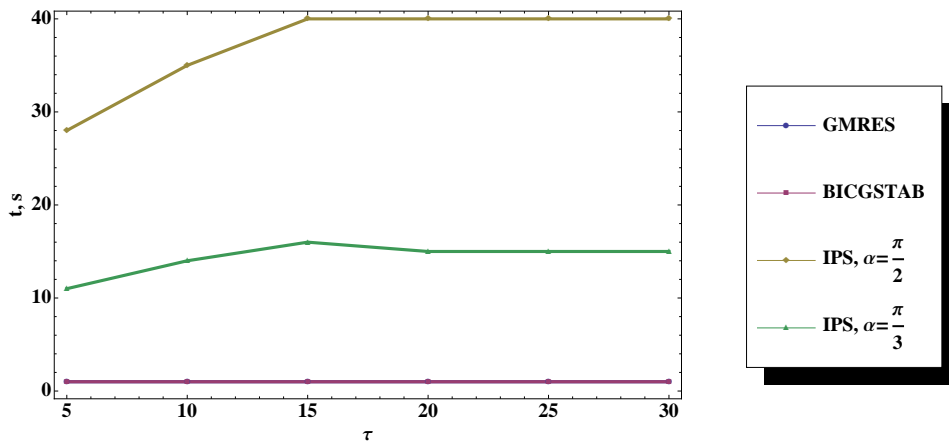


Рис. 2.4: Изменение трудоемкости с увеличением шага τ с 5 до 30 (с переобусловливанием)

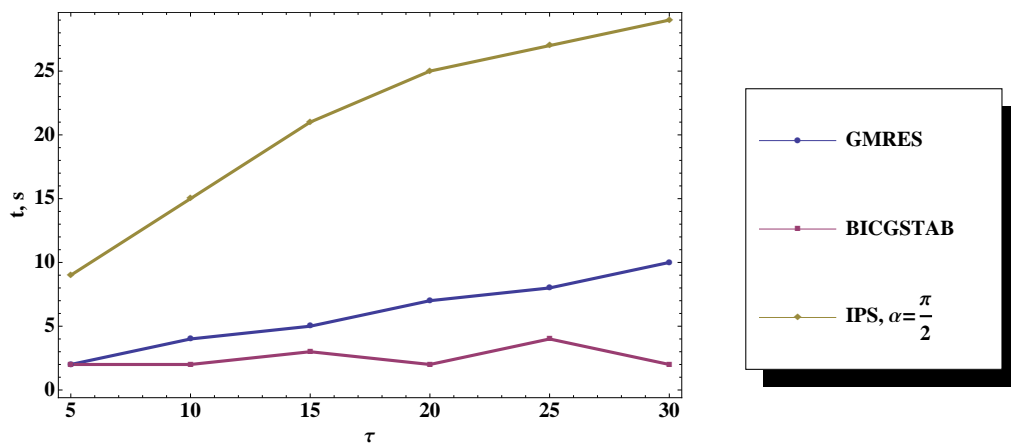


Рис. 2.5: Изменение трудоемкости с увеличением шага τ с 5 до 30 (без переобусловливания)

Следующий эксперимент демонстрирует влияние величины разброса собственных значений на скорость сходимости итерационных процессов: зафиксировав $N = 885$, $\tau = 10$, $\rho_{min} = 1$, будем изменять ρ_{max} от 10 до 10^3 . Результат представлен на рисунка 2.6 и 2.7.

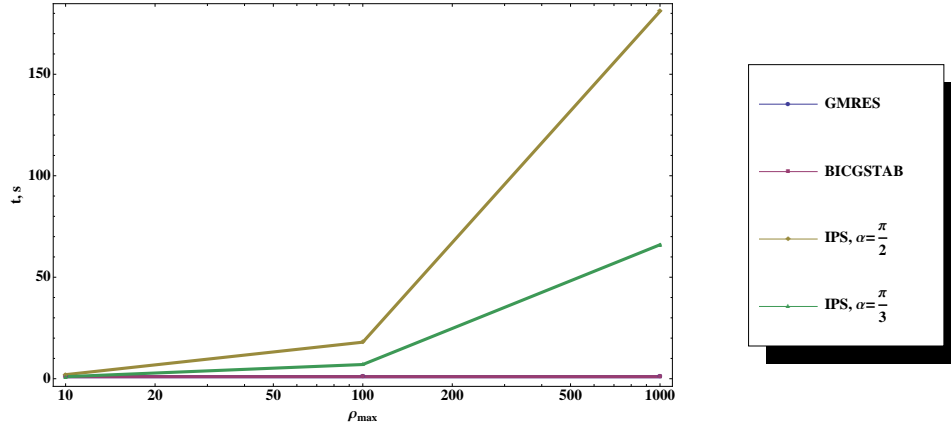


Рис. 2.6: Изменение трудоемкости с увеличением ρ_{max} с 10 до 10^3 (с переобусловливанием)

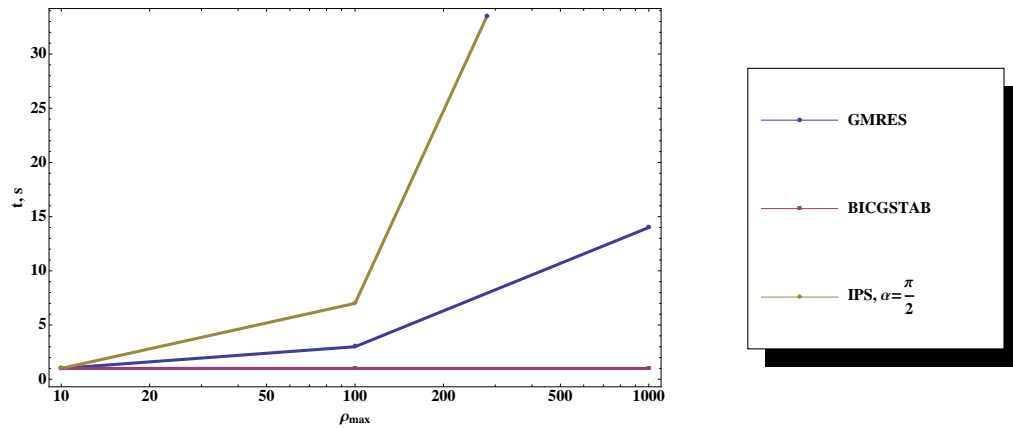


Рис. 2.7: Изменение трудоемкости с увеличением ρ_{max} с 10 до 10^3 (без переобусловливания)

Исследуем влияние изменения ρ_{min} на скорость сходимости. Для этого зафиксируем $N = 531$, $\tau = 10$, $\rho_{max} = 100$ и будем изменять ρ_{min} в диапазоне от 10^{-4} . Графики зависимости времени работы от ρ_{min} на рисунках 2.8 и 2.9.

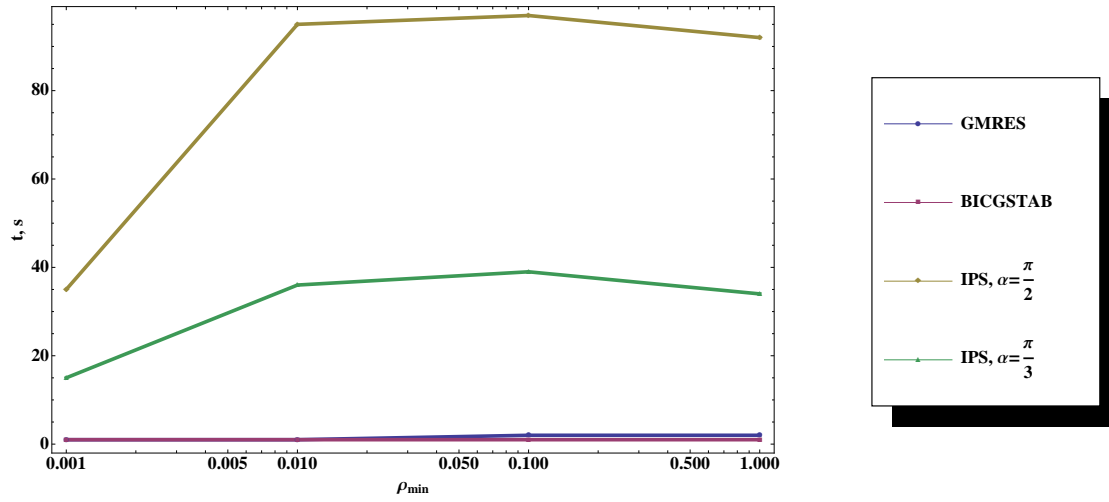


Рис. 2.8: Изменение трудоемкости с увеличением ρ_{min} с 10^{-3} до 1 (с переобусловливанием)

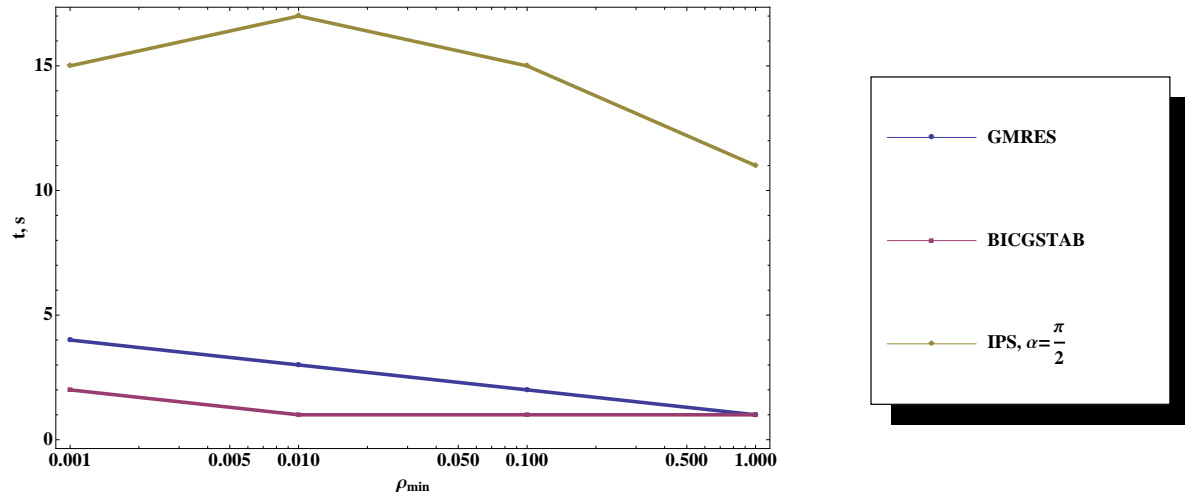


Рис. 2.9: Изменение трудоемкости с увеличением ρ_{min} с 10^{-3} до 1 (без переобусловливания)

Приведём также графики (рисунки 2.10 и 2.11) скорости решения системы с $\rho_{min} = 1$ и $\rho_{max} = 10$ и $\tau = 10$. N изменяется в пределах от 354 до 4602.

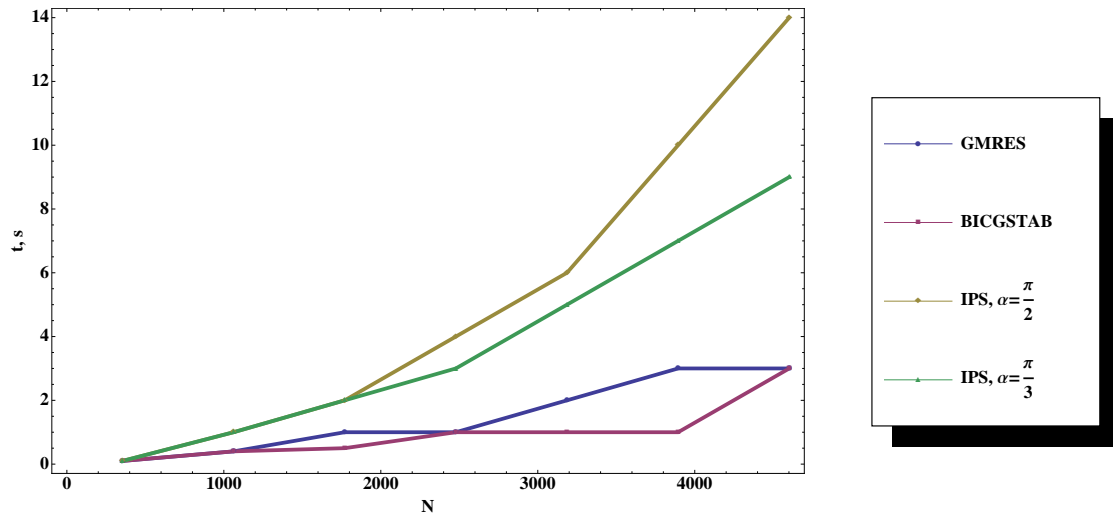


Рис. 2.10: Изменение трудоемкости с увеличением N с 354 до 4602 (с переобусловливанием)

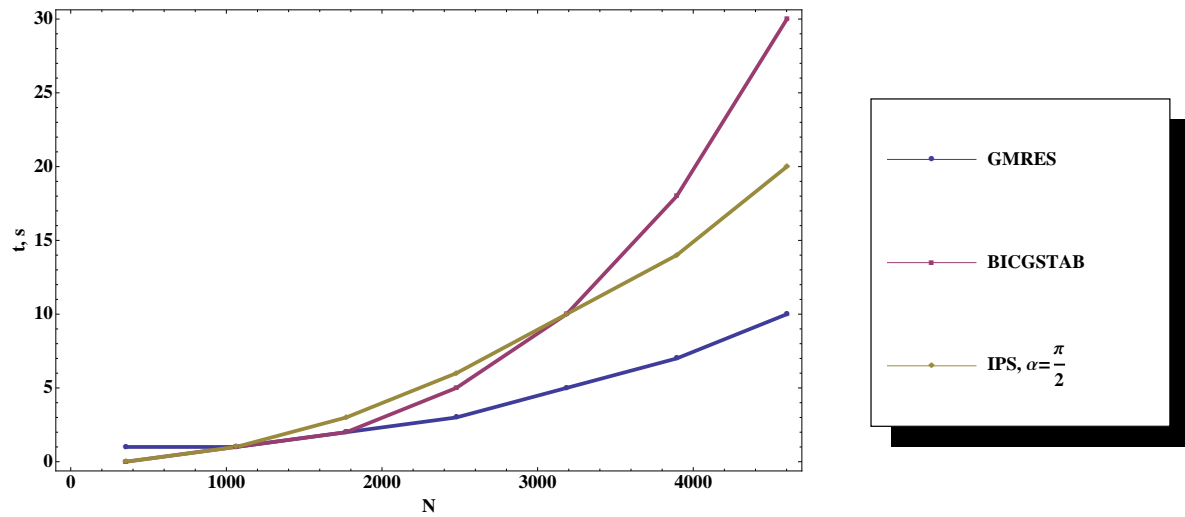


Рис. 2.11: Изменение трудоемкости с увеличением N с 354 до 4602 (без переобусловливания)

Исследуем влияние изменения шага по времени для задачи с малым отношением $\frac{\rho_{min}}{\rho_{max}}$. Параметры следующие: $\rho_{min} = 1000$, $\rho_{max} = 2000$, $\tau = 10$. Результаты представлены на рисунках 2.12 и 2.13

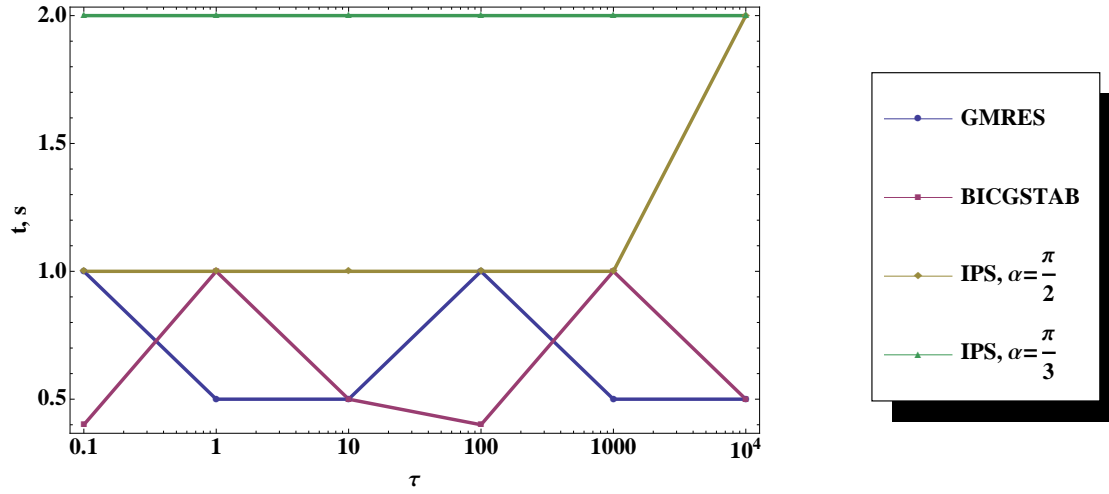


Рис. 2.12: Изменение трудоемкости с увеличением τ с 0.1 до 10^5 (с переобусловливанием)

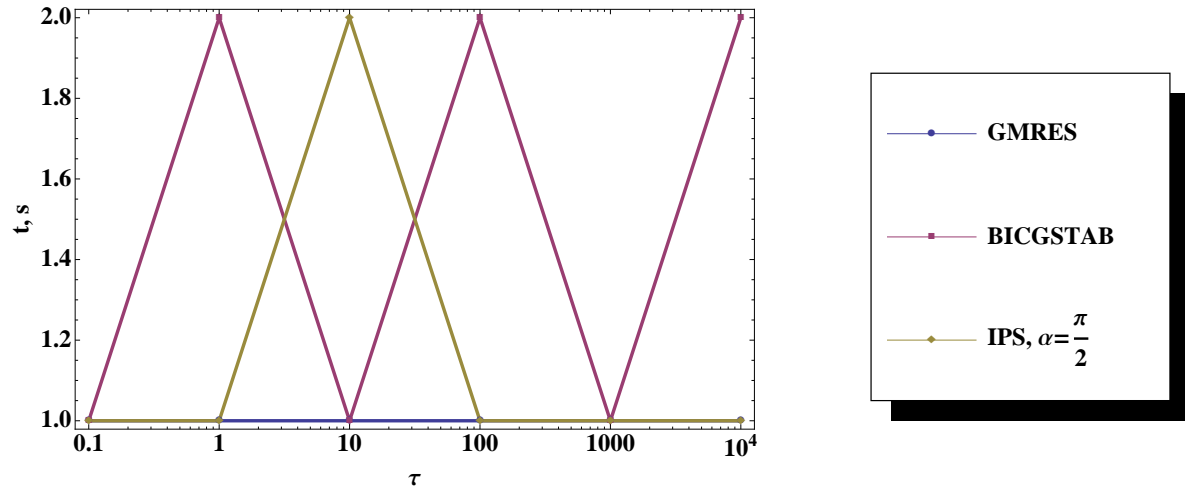


Рис. 2.13: Изменение трудоемкости с увеличением τ с 0.1 до 10^5 (без переобусловливания)

Таким образом можно заключить, что чем ближе ρ_{min} к ρ_{max} , тем быстрее работают методы, причем изменение шага τ фактически не влияет на трудоемкость отыскания решения, что в свою очередь, является очень полезным свойством при решении жёстких задач.

2.2 Случай наличия близких к нулю действительных частей собственных значений

Предыдущие эксперименты демонстрировали превосходство методов GMRES и BICGSTAB над разрабатываемым. Однако, в случае, когда собственные значения матрицы J находятся вблизи мнимой оси (в данном эксперименте это величины порядка 10^{-5} , см. рис. 2.14), получается кардинально иная картина. На графике

скорость работы метода BICGSTAB не отражена, поскольку этот метод вообще не сходится.

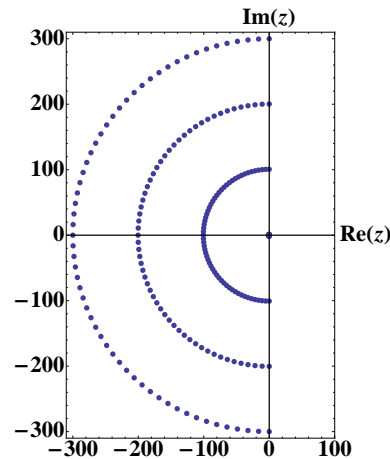


Рис. 2.14: Собственные значения матрицы J . $N = 236$

Чтобы это продемонстрировать, зададим параметры эксперимента следующим образом: $\tau = 10$, $\rho_{min} = 1$, $\rho_{max} = 300$ и будем увеличивать N . Результаты представлены на рисунке 2.15.

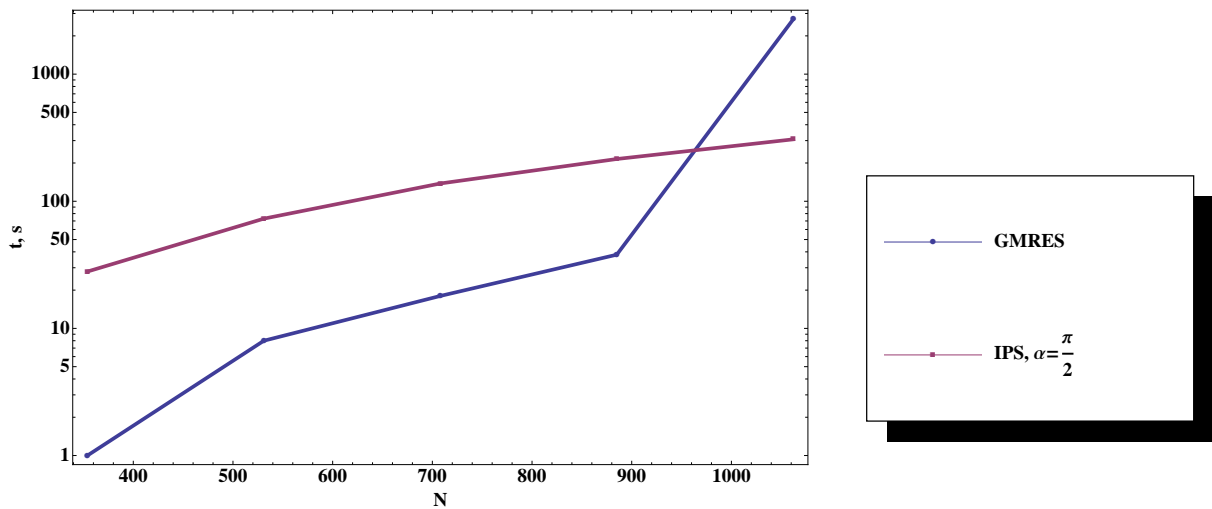


Рис. 2.15: Изменение трудоемкости с увеличением N с 354 до 1062

Стоит отметить, что метод GMRES на задаче с $N = 1062$ сходится более чем за 2700 секунд, но точное время работы было решено не устанавливать поскольку даже этот результат достаточно характеризует картину.

2.3 Линейная задача с разреженной матрицей

Матрица у представленных выше задач являлась полной. Рассмотрим теперь задачу с ленточной матрицей. Положим

$$y'(t) = Jy(t) + f(t), \quad t \in [0, 0.15], \quad M \in \mathbb{R}^N \times \mathbb{R}^N,$$

$$f_i = t \exp(10t),$$

$$y(0) = (1, \dots, 1)^T,$$

$$J = \begin{pmatrix} -3 & 1 & 0 & & 0 \\ -4 & -12 & 4 & & \\ & \ddots & \ddots & \ddots & \\ & & -(N-1)^2 & -3(N-1)^2 & (N-1)^2 \\ 0 & & & -N^2 & -3N^2 \end{pmatrix}.$$

Спектр матрицы J этой задачи при различных N представлен на рисунке 2.16. Учитывая область распределения собственных значений, можно воспользоваться методом с областью устойчивости в секторе с углом $\frac{\pi}{3}$. Шаг по времени $\tau = 0.15$. Результаты работы представлены на рисунках 2.17 и 2.18.

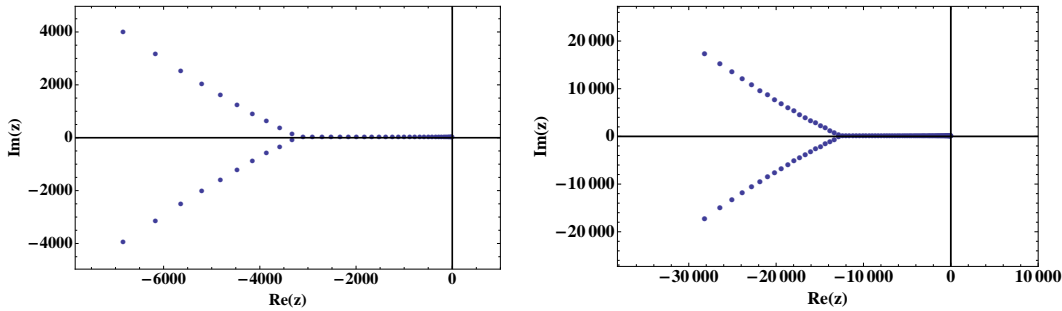


Рис. 2.16: Спектр матрицы J для $N = 50$ и $N = 100$

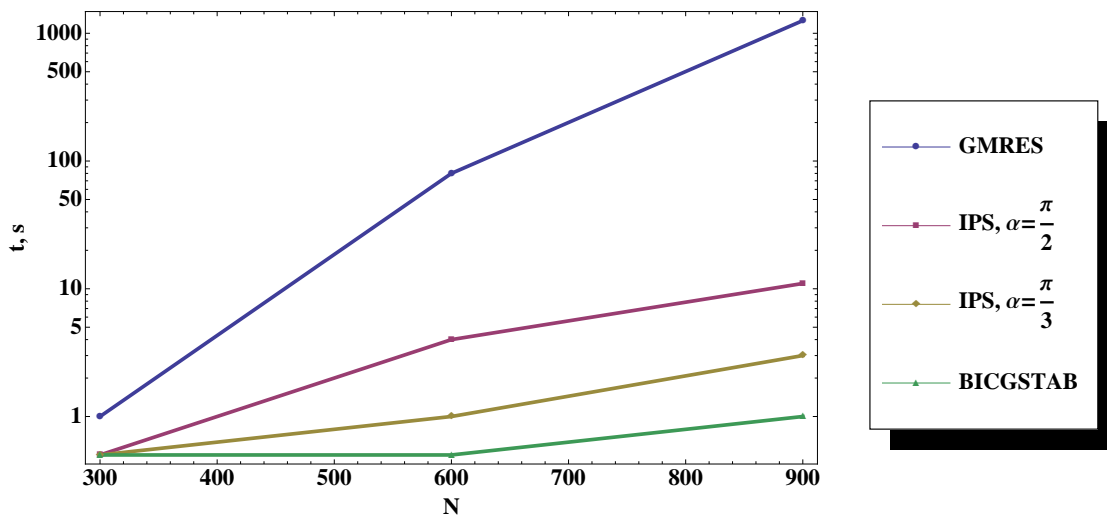


Рис. 2.17: Зависимость времени работы от размерности N . Случай с переобуславливанием.

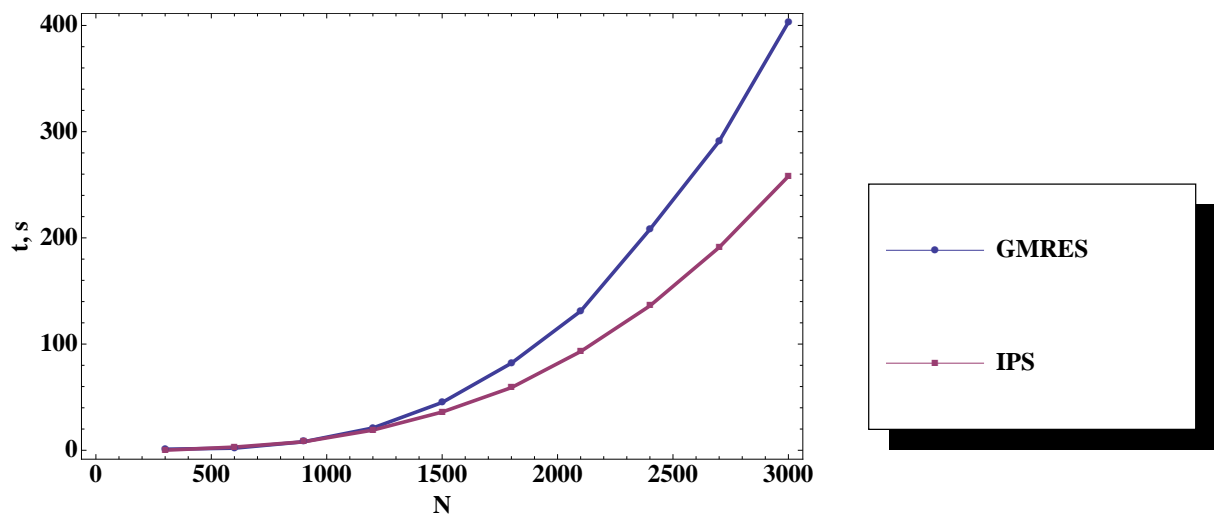


Рис. 2.18: Зависимость времени работы от размерности N . Случай без переобусловливания.

На графике 2.17 отражены данные только для N от 300 до 900, поскольку даже при 900 вычисления занимают около 20 минут и при увеличении размерности это время ещё возрастет.

ЗАКЛЮЧЕНИЕ

В работе представлен подход к реализации неявных методов для жёстких систем с на основе метода установления. Приведены расчётные формулы и данные численного эксперимента.

Рассмотрена и реализована возможность расширения области применимости методов, основанных на принципе установления при помощи переобусловливания.

Разработанный метод применен для решения тестовых линейных задач с заданным спектром. По результатам можно сказать, что разработанный метод подходит для решения задач со спектром из левой комплексной полуплоскости, причем в отдельных случаях работает лучше чем некоторые известные методы. При этом стоит особо отметить, что предлагаемые итерационные процессы естественным образом применимы в нелинейном случае, тогда как «решатели» GMRES и BICGSTAB — нет.

Низкая эффективность метода некоторых тестовых задачах может быть объяснена следующим образом: во-первых, в силу технических ограничений расчеты производились на матрицах относительно небольшой размерности, во-вторых реализация метода может быть не самой оптимальной.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Деккер К., Вервер Я. Устойчивость методов Рунге—Кутты для жестких нелинейных дифференциальных уравнений/ Пер. с англ.— М.: Мир, 1988.
2. Фалейчик Б. В., Бондарь И. В. Реализация неявных методов для жестких задач методом установления// Theoretical and Applied Aspects of Cybernetics. Proceedings of the International Scientific Conference of Students and Young Scientists – Kyiv: Bukrek, 2011. С. 297-299.
3. Faleichik B. V. Explicit Implementation of Collocation Methods for Stiff Systems with Complex Spectrum // Journal of Numerical Analysis, Industrial and Applied Mathematics. Vol. 5
4. Фалейчик Б. В., Бондарь И. В. Реализация неявных методов Рунге-Кутты с использованием принципа установления // Аналитические методы анализа и дифференциальных уравнений: Тез. докл. междунар. конф. 12-17 сент. 2011г, Минск, Беларусь. С. 146-147
5. Хайпер Э., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи./ Пер. с англ. — М.: Мир, 1999. — 685 с.
6. Фалейчик Б. В. Реализация неявных методов для жестких задач с использованием обобщенных итераций Пикара // Тр. 6-й международной конференции «Аналитические методы анализа и дифференциальных уравнений»: в двух томах – Т.1 Математический анализ. — Минск: Институт математики НАН Беларуси, 2012. С. 131–135.
7. Фалейчик Б. В., Бондарь И. В. Реализация неявных методов Рунге-Кутты для больших жестких систем//Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях: XV Республиканская научная конференция студентов и аспирантов “Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях”, 26-28 марта 2012 г.: [материалы]: в 2 ч. Ч.1/редкол. : Демиденко О.М. – Гомель: ГГУ им. Ф. Скорины, 2012. С. 175-176
8. Бондарь И. В. Итерационные процессы установления для жестких линейных задач //Тр. 69-й ежегодной научной конференции студентов и аспирантов БГУ: допущено в печать.
9. Интернет адрес: <http://www.dealii.org>