

НАУЧНАЯ СТАТЬЯ

© Измаилов А.Ф., Усков Е.И., 2024

<https://doi.org/10.20310/2686-9667-2024-29-148-401-424>

УДК 519.6



Ускорение сходимости ньютоновских методов к особым решениям нелинейных уравнений

Алексей Феридович ИЗМАИЛОВ¹, Евгений Иванович УСКОВ²¹ ФГБОУ ВО «Московский государственный университет им. М. В. Ломоносова»
119991, ГСП-2, Российская Федерация, г. Москва, Ленинские горы, 1² ФГБОУ ВО «Тамбовский государственный университет им. Г. Р. Державина»
392000, Российская Федерация, г. Тамбов, ул. Интернациональная, 33

Аннотация. Рассматривается простейшая процедура экстраполяции, а именно, удвоение шага, для ускорения сходимости ньютоновских методов к особым решениям гладких нелинейных уравнений. Демонстрируется, что ускоряющий эффект этой процедуры для разных ньютоновских методов может быть разным. Для линейно-квадратичных уравнений приводятся теоретические результаты, дающие количественные оценки потенциального эффекта от экстраполяции для методов Ньютона, Левенберга–Марквардта, а также недавно предложенного ньютоновского метода с подзадачами линейного программирования, в определенном смысле объясняющие наблюдаемую разницу. Теоретически анализ основан на интерпретации этих методов как возмущенного метода Ньютона с соответствующими оценками на возмущения, а также на тонких результатах, количественно характеризующих шаг такого возмущенного метода и его локальную сходимость с линейной скоростью к особым решениям, в которых выполняется условие 2-регулярности по направлению из ядра первой производной. Также проводятся численные эксперименты для глобализаций указанных алгоритмов, снабженных выбором параметра длины шага, на двух тестовых наборах. Экспериментальные наблюдения подтверждают теоретические результаты, а также демонстрируют, что в случаях, когда уравнение содержит нелинейные и неквадратичные члены, эффект от экстраполяции выравнивается.

Ключевые слова: нелинейное уравнение, особое решение, метод Ньютона, метод Левенберга–Марквардта, ньютоновский метод с подзадачами линейного программирования, экстраполяция

Благодарности: Исследование выполнено за счет гранта Российского научного фонда (проект № 24-21-00015, <https://rscf.ru/project/24-21-00015/>).

Для цитирования: Измаилов А.Ф., Усков Е.И. Ускорение сходимости ньютоновских методов к особым решениям нелинейных уравнений // Вестник российских университетов. Математика. 2024. Т. 29. № 148. С. 401–424. <https://doi.org/10.20310/2686-9667-2024-29-148-401-424>

SCIENTIFIC ARTICLE

© A. F. Izmailov, E. I. Uskov, 2024

<https://doi.org/10.20310/2686-9667-2024-29-148-401-424>

Accelerating convergence of Newton-type methods to singular solutions of nonlinear equations

Alexey F. IZMAILOV¹, Evgeniy I. USKOV²

¹ Lomonosov Moscow State University

1 Leninskiye Gory, Moscow 119991, Russian Federation

² Derzhavin Tambov State University

33 Internatsionalnaya St., Tambov 392000, Russian Federation

Abstract. We consider the simplest extrapolation procedure, specifically doubling the step, intended for acceleration of convergence of Newton-type methods to singular solutions of smooth nonlinear equations. We demonstrate that the acceleration effect of this procedure can be different for different Newton-type methods. For linear-quadratic equations we provide theoretical results yielding quantitative estimates of the potential effect of extrapolation for the Newton method, for the Levenberg–Marquardt method, and for the recently proposed LP-Newton method, in some sense explaining the observed difference. Theoretical analysis relies on interpretation of these methods as a perturbed Newton method with the appropriate estimates of perturbations, as well as on sharp results yielding a quantitative characterization of a step of such perturbed method, and its local convergence at a linear rate to singular solutions satisfying the 2-regularity condition in a direction from the null space of the first derivative. Furthermore, we perform numerical experiments with globalized versions of the algorithms in question, equipped with choosing the stepsize parameter, on two sets of test problems. Experimental observations confirm the theoretical results, and also demonstrate that in cases when the equation contains nonlinear and nonquadratic terms, the effect of extrapolation is evened out.

Keywords: nonlinear equation, singular solution, Newton method, Levenberg–Marquardt method, LP-Newton method, extrapolation

Acknowledgements: The research was supported by the Russian Science Foundation (project no. 24-21-00015, <https://rscf.ru/project/24-21-00015/>).

Mathematics Subject Classification: 47J05, 65H10.

For citation: Izmailov A.F., Uskov E.I. Accelerating convergence of Newton-type methods to singular solutions of nonlinear equations. *Vestnik Rossiyskikh Universitetov. Matematika = Russian Universities Reports. Mathematics*, **29**:148 (2024), 401–424.

<https://doi.org/10.20310/2686-9667-2024-29-148-401-424> (In Russian, Abstr. in Engl.)

Введение

Известным средством ускорения сходимости ньютоновских методов к особым решениям нелинейных уравнений является экстраполяция [1, 2]. Однако, как демонстрируется в этой статье, ускоряющий эффект этой процедуры для разных ньютоновских методов может быть разным, и целью работы является теоретическое объяснение этой разницы для линейно-квадратичных уравнений, а также численный анализ эффекта от экстраполяции для глобализованных одномерным поиском методов Ньютона, Левенберга–Марквардта, а также ньютоновского метода с подзадачами линейного программирования (соответствующие ссылки на литературу приводятся ниже).

Используемые в статье обозначения вполне традиционны. Через $L^\perp$ обозначается ортогональное дополнение линейного подпространства L . Если не указано иное, будем считать, что $\|\cdot\|$ — евклидова норма; в некоторых случаях будет также использоваться ∞ -норма, которая для $u \in \mathbb{R}^p$ определяется как $\|u\|_\infty = \max\{|u_j| \mid j \in \{1, \dots, p\}\}$. Для заданных $\bar{u}, \bar{v} \in \mathbb{R}^p$, $\varepsilon > 0$ и $\delta > 0$ определим множество

$$K_{\varepsilon, \delta}(\bar{u}; \bar{v}) = \{u \in \mathbb{R}^p \mid \|u - \bar{u}\| \leq \varepsilon, \ ||\bar{v}\|(u - \bar{u}) - \|u - \bar{u}\|\bar{v}\| \leq \delta\|u - \bar{u}\|\|\bar{v}\|\}.$$

Через I обозначается единичная матрица, а через $\text{im } A$ и $\ker A$ — образ (множество значений) и ядро (множество нулей) матрицы (линейного оператора) A . Для симметричного билинейного отображения $B : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ и для заданного $v \in \mathbb{R}^p$ через $B[v]$ обозначается матрица линейного оператора $B[v, \cdot]$. Наконец, запись $O(t)u$ используется для произведения вида Au с матрицей $A = O(t)$ в случаях, когда другая специфика этой матрицы не имеет значения.

1. Постановка задачи и ньютоновские методы

Рассматриваются ньютоновские методы для решения нелинейного уравнения

$$\Phi(u) = 0 \tag{1.1}$$

с достаточно гладким отображением $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$. Такие методы по имеющемуся приближению $u \in \mathbb{R}^p$ генерируют следующее приближение как $u + v$, где $v \in \mathbb{R}^p$ есть решение подзадачи, характеризующей конкретный метод. Например:

- подзадача метода Ньютона (NM; см., например, [3, разд. 2.1.1]) — это линеаризованное уравнение (1.1), т. е.

$$\Phi(u) + \Phi'(u)v = 0; \tag{1.2}$$

- подзадача метода Левенберга–Марквардта (LM; происходит из работ [4, 5]; см. также, например, [6, разд. 10.2], а также обзор современных теорий сходимости в [7]) имеет вид

$$\frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{2}\sigma(u)\|v\|^2 \rightarrow \min, \tag{1.3}$$

с функцией $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}_+$, определяющей значения параметра регуляризации, что равносильно линейному уравнению

$$(\Phi'(u))^\top \Phi(u) + ((\Phi'(u))^\top \Phi'(u) + \sigma(u)I)v = 0; \tag{1.4}$$

- подзадача ньютоновского метода с подзадачами линейного программирования (LPN; [8]) имеет вид

$$\gamma \rightarrow \min, \quad \|\Phi(u) + \Phi'(u)v\| \leq \gamma \|\Phi(u)\|^2, \quad \|v\| \leq \gamma \|\Phi(u)\|, \quad (1.5)$$

относительно (v, γ) с $\gamma \in \mathbb{R}$, и если используется ∞ -норма, то это задача линейного программирования, что и дает название методу.

Как обсуждается ниже, все перечисленные методы могут интерпретироваться как возмущенный метод Ньютона (pNM) с подзадачей

$$\Phi(u) + (\Phi'(u) + \Omega(u))v = \omega(u), \quad (1.6)$$

где отображения $\Omega : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$ и $\omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$ характеризуют различные возмущения базового NM. Свойства базового NM и pNM вблизи неособых решений излагаются в [3, разд. 2.1.1]. Здесь решение $\bar{u}$ уравнения (1.1) называется неособым, если $\Phi'(\bar{u})$ — невырожденная матрица, и особым в противном случае.

Данная работа посвящена именно случаю особого решения $\bar{u}$. Ключевую роль при этом играет условие 2-регулярности отображения Φ в точке $\bar{u}$ по направлению $v \in \mathbb{R}^p$ [9, разд. 1.3.3], сводящееся в рассматриваемом здесь случае отображения из $\mathbb{R}^p$ в $\mathbb{R}^p$ к невырожденности линейного оператора $u \mapsto \Pi \Phi''(\bar{u})[v, u] : \ker \Phi'(\bar{u}) \rightarrow (\operatorname{im} \Phi'(\bar{u}))^\perp$, где Π — оператор ортогонального проектирования на $(\operatorname{im} \Phi'(\bar{u}))^\perp$. Важное свойство, которое легко проверить, состоит в том, что из 2-регулярности Φ в точке $\bar{u}$ по направлению v вытекает существование $\varepsilon > 0$ и $\delta > 0$ таких, что для $u \in K_{\varepsilon, \delta}(\bar{u}, v)$ матрица $\Phi'(u)$ невырождена, причем

$$(\Phi'(u))^{-1} = O(\|u - \bar{u}\|^{-1}) \quad (1.7)$$

при $u \rightarrow \bar{u}$. С использованием этого факта, следующий результат о локальной сходимости pNM с линейной скоростью был получен в [10, теорема 1] (см. также [11, теорема 3.1], где используются более слабые предположения гладкости), и является развитием результатов, полученных [1, 2] для базового NM. Для всякого $u \in \mathbb{R}^p$ введем однозначное разложение $u = u_1 + u_2$, $u_1 \in (\ker \Phi'(\bar{u}))^\perp$, $u_2 \in \ker \Phi'(\bar{u})$.

Теорема 1.1. Пусть $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$ дважды дифференцируемо вблизи $\bar{u} \in \mathbb{R}^p$, и его вторая производная липшицева относительно $\bar{u}$, т. е.

$$\Phi''(u) - \Phi''(\bar{u}) = O(\|u - \bar{u}\|)$$

при $u \rightarrow \bar{u}$. Пусть $\bar{u}$ является решением уравнения (1.1), причем Φ 2-регулярно в точке $\bar{u}$ по направлению $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$. Пусть для $\Omega : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$ и $\omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$ существует $\delta > 0$ такое, что оценки

$$\Omega(u) = O(\|u - \bar{u}\|), \quad \omega(u) = O(\|u - \bar{u}\|^2),$$

$$\Pi \Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2)$$

$$\Pi \omega(u) = O(\|u - \bar{u}\| \|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3)$$

выполняются для $u \in K_{\varepsilon, \delta}(\bar{u}; \bar{v})$ при $\varepsilon \rightarrow 0+$.

Тогда для любых $\hat{\varepsilon} > 0$ и $\hat{\delta} > 0$ найдутся $\varepsilon = \varepsilon(\bar{v}) > 0$ и $\delta = \delta(\bar{v}) > 0$ такие, что для всякой начальной точки $u^0 \in K_{\varepsilon, \delta}(\bar{u}; \bar{v})$ существует единственная последовательность

$\{u^k\} \subset \mathbb{R}^p$ такая, что для каждого k выполняется $u^{k+1} = u^k + v^k$, где v^k удовлетворяет (1.6), и для этой последовательности и для каждого k имеет место $u_2^k \neq \bar{u}_2$, $u^k \in K_{\varepsilon, \delta}(\bar{u}; \bar{v})$, $\{u^k\}$ сходится к $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ сходится к нулю монотонно,

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_2^{k+1} - \bar{u}_2\|} = O(\|u^k - \bar{u}\|)$$

при $k \rightarrow \infty$, и

$$\lim_{k \rightarrow \infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|} = \frac{1}{2}.$$

Характер сходимости в теореме 1.1 служит основой для техник ускорения сходимости ньютоновских методов к особым решениям [1, 2], одной из которых является экстраполяция, которая в своем простейшем варианте состоит в генерировании, параллельно с основной последовательностью $\{u^k\}$, вспомогательной последовательности $\{\hat{u}^k\}$ посредством удвоения шага (p)NM:

$$\hat{u}^{k+1} = u^k + 2v^k. \quad (1.8)$$

Как показано в [2, Theorem 4.1], для базового NM можно ожидать $\{\hat{u}^k\}$ линейной сходимости с асимптотическим общим частным $1/4$, в то время как для основной итерационной последовательности $\{u^k\}$ оценки в теореме 1.1 гарантируют такое свойство только с $1/2$. Заметим, что (1.8) не связано ни с какими дополнительными вычислительными затратами, кроме одного лишнего вычисления Φ на каждой итерации, что требуется для оценки качества полученного приближения $\hat{u}^{k+1}$. Кроме того, экстраполяция никак не влияет на основную последовательность $\{u^k\}$, и поэтому легко комбинируется с различными реализациями pNM, и в частности, с глобализациями сходимости методов этого класса, в том числе с рассматриваемыми в разд. 4. алгоритмами, использующими одномерный поиск для адаптивного выбора параметра длины шага. Однако, эффект от экстраполяции можно ожидать только при асимптотическом принятии полного ньютоновского шага, поскольку данная процедура основана на характере сходимости именно полношагового pNM; см. [12, 13].

Возвращаясь к базовому pNM, для метода LM, использующего $\sigma(u) = O(\|\Phi(u)\|^\theta)$ с фиксированным $\theta > 0$, на основе (1.7) в [10, разд. 3.1] показано, что при 2-регулярности Φ в точке $\bar{u}$ по направлению $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$ найдутся $\varepsilon > 0$ и $\delta > 0$ такие, что для $u \in K_{\varepsilon, \delta}(\bar{u}, \bar{v})$ (1.4) записывается в виде (1.6) с $\omega(\cdot) \equiv 0$ и

$$\Omega(u) = \sigma(u)((\Phi'(u))^{-1})^\top = O(\|u - \bar{u}\|^{-1} \|\Phi(u)\|^\theta) \quad (1.9)$$

при $u \rightarrow \bar{u}$. В [14, разд. 3.2] предложена альтернативная интерпретация подзадачи (1.3) метода LM как подзадачи (1.6) pNM, с $\Omega(\cdot) \equiv 0$ и соответствующей оценкой на ω . В отличие от приведенной выше интерпретации, явно основанной на форме (1.4) итерационной подзадачи метода LM, альтернативная интерпретация применима и в случае уравнения с дополнительным ограничением, но получаемые при этом оценки несколько слабее и не подходят для использования в разд. 3 ниже.

Для метода LPN, снова с использованием (1.7), в [10, разд. 3.2] показано, что в тех же предположениях найдутся $\varepsilon > 0$ и $\delta > 0$ такие, что для $u \in K_{\varepsilon, \delta}(\bar{u}, \bar{v})$ любое решение подзадачи (1.5) удовлетворяет (1.6) с $\Omega(\cdot) \equiv 0$ и некоторым ω , для которого

$$\|\omega(u)\| \leq \gamma(u) \|\Phi(u)\|^2 = O(\|u - \bar{u}\| \|\Phi(u)\|) \quad (1.10)$$

при $u \rightarrow \bar{u}$, где $\gamma(u)$ — оптимальное значение задачи (1.5).

2. Сравнительный эффект от экстраполяции

Начнем с простого примера, хорошо демонстрирующего эффекты, о которых идет речь в этом разделе.

Пример 2.1. Пусть $p = 1$, $\Phi(u) = u^2$. Единственным решением уравнения (1.1) является $\bar{u} = 0$, оно является особым, и Φ 2-регулярно в точке 0 по любому ненулевому направлению.

Итерационное уравнение (1.2) NM имеет вид $u^2 + 2uv = 0$, и при $u \neq 0$ его единственным решением является $v = -u/2$. Поэтому $u + 2v = 0$, т.е. NM с экстраполяцией за один шаг попадает в точное решение из любого начального приближения.

Далее, итерационное уравнение (1.4) метода LM имеет вид $2u^3 + (4u^2 + \sigma(u))v = 0$, и при $u \neq 0$ его единственным решением является $v = -2u^3/(4u^2 + \sigma(u))$. Если $\sigma(u) = |u|^{2\theta}$, то

$$u + 2v = \frac{\sigma(u)u}{(4u^2 + \sigma(u))} = \frac{|u|^{2\theta+1}}{(4u^2 + |u|^{2\theta})},$$

что при $\theta > 1$ дает оценки

$$u + 2v = \frac{1}{4}|u|^{2\theta-1} + o(|u|^{2\theta-1}) = O(|u|^{2\theta-1}),$$

что при $\theta \geq 2$ превращается в оценку $u + 2v = O(|u|^3)$.

Наконец, итерационная подзадача (1.5) метода LPN имеет вид

$$\gamma \rightarrow \min, \quad -\gamma u^4 \leq u^2 + 2uv \leq \gamma u^4, \quad -\gamma u^2 \leq v \leq \gamma u^2,$$

и, скажем, при $u > 0$ единственным решением этой задачи линейного программирования является $(v, \gamma) = (-u/(2+u), 1/(2u+u^2))$. Тогда $u + 2v = u^2/(2+u)$, что дает оценки

$$u + 2v = \frac{1}{2}u^2 + o(u^2) = O(u^2).$$

Функция в примере 2.1 является квадратичной скалярной функцией одной переменной. Обратимся теперь к случаю общего линейно-квадратичного отображения Φ : пусть

$$\Phi(u) = Au + \frac{1}{2}B[u, u], \quad (2.1)$$

где A — $(n \times n)$ -матрица, а $B : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ — симметричное билинейное отображение. Тогда $\Phi'(u) = A + B[u]$, и уравнение (1.1) имеет решение $\bar{u} = 0$, в котором $\Phi'(0) = A$, и, соответственно, это решение является особым, если матрица A вырождена. Условие 2-регулярности Φ в точке 0 по направлению $v \in \mathbb{R}^p$ при этом состоит в невырожденности линейного оператора $\hat{B}(v) : \ker A \rightarrow (\operatorname{im} A)^\perp$, $\hat{B}(v)u = \Pi B[v, u]$, где Π — оператор ортогонального проектирования на $(\operatorname{im} A)^\perp$.

Итерационная подзадача (1.2) NM принимает вид

$$(A + B[u])v = -Au - \frac{1}{2}B[u, u]. \quad (2.2)$$

Если $u \in \ker A$, то правую часть (2.2) можно заменить на $-(A + B[u])u/2$, и тогда уравнение (2.2) имеет решение $v = -u/2$, причем это решение единственно в случае невырожденности $A + B[u]$. При этом $u + 2v = 0$, т.е. экстраполяция немедленно дает точное

решение. В чисто квадратичном случае, когда $A = 0$, условие $u \in \ker A$ выполняется автоматически, и значит, экстраполяция дает точное решение за один шаг из любого начального приближения u , в котором матрица $B[u]$ невырождена.

Далее, итерационная подзадача (1.6) pNM принимает вид

$$(A + B[u] + \Omega(u))v = -Au - \frac{1}{2}B[u, u] + \omega(u), \quad (2.3)$$

и при $u \in \ker A$ правую часть можно заменить на $-(A + B[u] + \Omega(u))u - \Omega(u)u/2 + \omega(u)$. Если матрица $A + B[u] + \Omega(u)$ обратима, то единственным решением (2.3) является

$$v = -\frac{1}{2}u + (A + B[u] + \Omega(u))^{-1} \left(\frac{1}{2}\Omega(u)u + \omega(u) \right). \quad (2.4)$$

При $u \in \ker A$ имеет место оценка $\Phi(u) = O(\|u\|^2)$, и, предполагая 2-регулярность Φ в точке 0 по направлению $\bar{v} \in \ker A \setminus \{0\}$, для метода LM с $\theta > 1$, в силу (1.9), для $u \in K_{\varepsilon, \delta}(0, \bar{v})$ при достаточно малых $\varepsilon > 0$ и $\delta > 0$ имеем

$$\Omega(u) = O(\|u\|^{2\theta-1}), \quad (2.5)$$

и матрица $A + B[u] + \Omega(u)$ обратима, причем, аналогично (1.7),

$$(A + B[u] + \Omega(u))^{-1} = O(\|u\|^{-1}). \quad (2.6)$$

Тогда из (2.4) (и того, что $\omega(\cdot) \equiv 0$) получаем

$$v = -\frac{1}{2}u + O(\|u\|^{2\theta-1}) \quad (2.7)$$

при $u \rightarrow 0$. Поэтому экстраполяция дает приближение

$$u + 2v = O(\|u\|^{2\theta-1}), \quad (2.8)$$

и, в частности, при $\theta \geq 2$ имеет место оценка

$$u + 2v = O(\|u\|^3)$$

при $u \rightarrow 0$.

Для метода LPN аналогичным образом при $u \in \ker A$ из (1.10) выводится оценка

$$\omega(u) = O(\|u\|^3), \quad (2.9)$$

и в силу (2.4), (2.6) (и того, что $\Omega(\cdot) \equiv 0$)

$$v = -\frac{1}{2}u + O(\|u\|^2),$$

и следовательно,

$$u + 2v = O(\|u\|^2) \quad (2.10)$$

при $u \rightarrow 0$.

Тем самым доказано

Предложение 2.1. Пусть A — $(n \times n)$ -матрица, а $B : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ — симметричное билинейное отображение, и пусть для некоторого $\bar{v} \in \ker A \setminus \{0\}$ линейный оператор $\hat{B}(\bar{v})$ невырожден.

Тогда найдутся $\varepsilon > 0$ и $\delta > 0$ такие, что для $u \in K_{\varepsilon, \delta}(0, \bar{v}) \cap \ker A$, для уравнения (1.1) с линейно-квадратичным отображением Φ из (2.1) справедливо следующее:

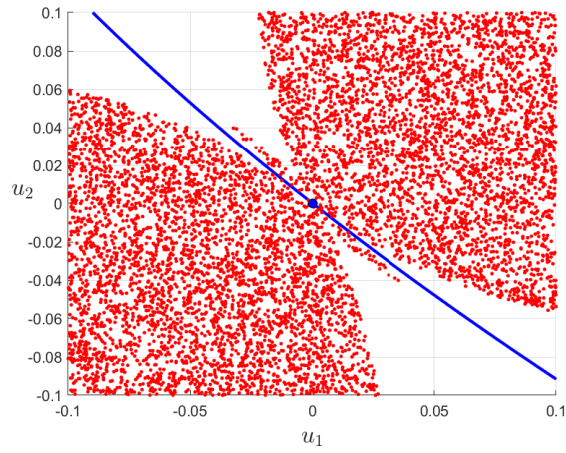
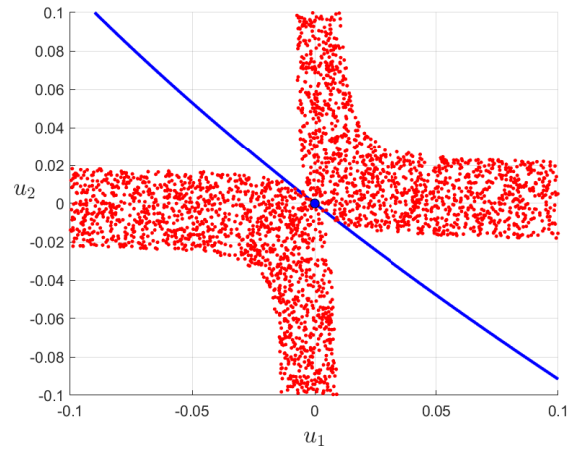
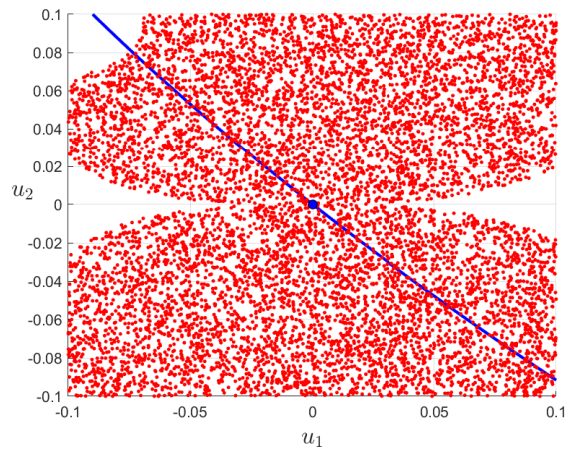
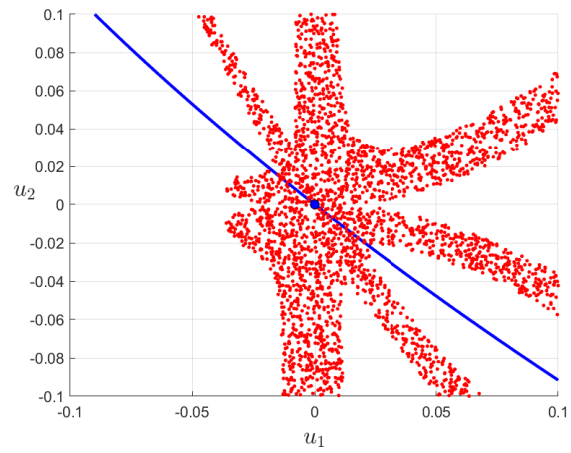
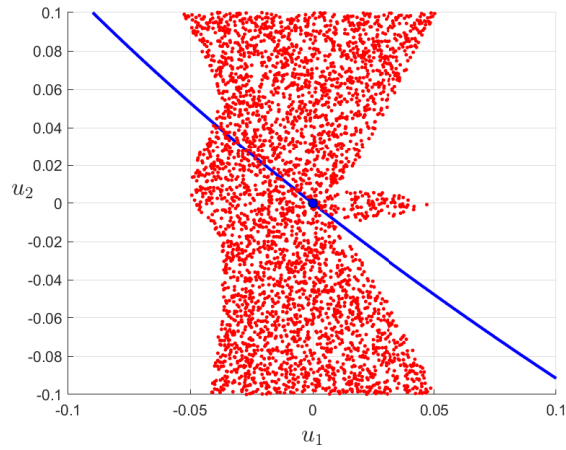
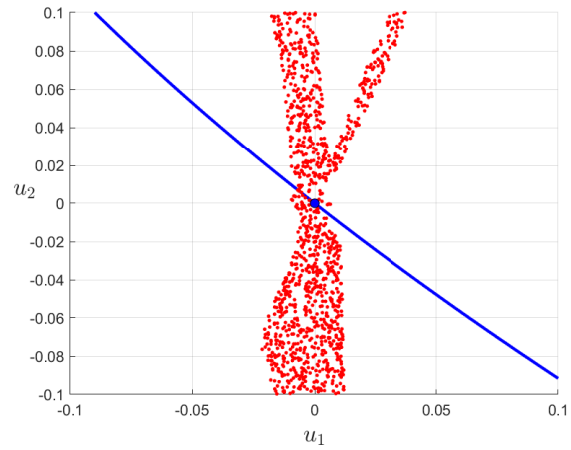
- NM с подзадачей (1.2) однозначно определяет направление v , и для него экстраполяция дает $u + 2v = 0$;
- метод LM с подзадачей (1.3), использующей $\sigma(u) = O(\|\Phi(u)\|^\theta)$ с фиксированным $\theta > 1$, однозначно определяет направление v , и для него экстраполяция дает приближение $u + 2v$, для которого выполняется оценка (2.8);
- метод LPN с подзадачей (1.5) определяет некоторое направление v , и для всякого такого направления экстраполяция дает приближение $u + 2v$, для которого выполняется оценка (2.10).

В частности, при $\theta > 1$ оценка (2.8) для метода LM лучше, чем оценка (2.10) для метода LPN, и это может служить объяснением большего эффекта экстраполяции для метода LM, чем для LPN.

В общем линейно-квадратичном случае (когда $A \neq 0$, и, соответственно, выполнение $u \in \ker A$ не является автоматическим), и даже просто в случае общего достаточно гладкого Φ , согласно [15, теорема 6.1] (см. также [13, лемма 7]), за один шаг NM «приближенно» попадает в $\ker A$ из асимптотически плотного звездного множества начальных точек $u^0 \in \mathbb{R}^p$, а именно, попадает в $K_{\varepsilon, \delta}(0, \bar{v})$ с соответствующим $\bar{v} = \bar{v}(u^0/\|u^0\|)$, при любых наперед заданных $\varepsilon > 0$ и $\delta > 0$. (Точного попадания в $\ker A$ может не быть даже для NM, даже в линейно-квадратичном случае, что демонстрируется примером 2.2 ниже, в котором $\dim \ker A = 1$; аналогичный эффект при $\dim \ker A = 2$ наблюдается в тесте Misc 10 в разд. 4.) Однако, как указано в комментариях в конце [10, разд. 2], гарантировать указанное свойство для общего pNM не удастся, если только не накладывать жесткие предположения на глобальное (не только вблизи ядра) поведение Ω и ω , а именно, $P\Omega(u) = O(\|u\|^2)$ и $P\omega(u) = O(\|u\|^3)$ для u из рассматриваемого звездного множества начальных приближений. Для метода LM эти предположения сводятся к оценке $\sigma(u) = O(\|u\|^3)$, что можно гарантировать только при $\theta \geq 3$. Для метода LPN гарантировать требуемое можно при выполнении оценки $\gamma(u) = O(\|u\|)$, а эта оценка ниоткуда не следует. Тем не менее, на практике попадание LM и LPN в $K_{\varepsilon, \delta}(0, \bar{v})$ с достаточно малыми $\varepsilon > 0$ и $\delta > 0$ обычно наблюдается, если не за одну, то за небольшое количество начальных итераций.

Пример 2.2. Пусть $p = 2$, $\Phi(u) = (u_1 + u_2^2/2, u_1 u_2 + u_2^2/2)$. Единственным решением уравнения (1.1) является $\bar{u} = 0$, оно является особым, $\ker \Phi'(0)$ натянуто на $\bar{v} = (0, 1)$, и Φ 2-регулярно в точке 0 по направлению $\bar{v}$.

На рис. 1 показана линия, задаваемая равенством $\det \Phi'(u) = 0$, а также области начальных точек, один шаг рассматриваемых методов из которых приводит в $K_{\varepsilon, \delta}(0, \bar{v})$ с $\varepsilon = 0.1$, при двух разных значениях δ . Эти области уменьшаются при уменьшении δ , но, по-видимому, остаются асимптотически плотными в точке 0.

(a) NM, $\delta = 0.05$ (b) NM, $\delta = 0.01$ (c) Метод LM, $\delta = 0.05$ (d) Метод LM, $\delta = 0.01$ (e) Метод LPN, $\delta = 0.05$ (f) Метод LPN, $\delta = 0.01$ Рис. 1. Пример 2.2: области попадания одного шага NM в $K_{\varepsilon, \delta}(0, \bar{v})$ с $\varepsilon = 0.1$

В соответствии со сказанным выше, далее будем рассматривать случай, когда $u^0 \in K_{\varepsilon, \delta}(0, \bar{v})$. По-прежнему будем использовать разложение всякого $u \in \mathbb{R}^p$ в сумму $u = u_1 + u_2$, где в линейно-квадратичном случае $u_1 \in (\ker A)^\perp$, $u_2 \in \ker A$.

Согласно [11, лемма 3.1] для общего (не обязательно линейно-квадратично) случая, и даже всего лишь при сильной полугладкости Φ' в решении $\bar{u} = 0$, для $u \in K_{\bar{\varepsilon}, \bar{\delta}}(0, \bar{v})$ при малых $\bar{\varepsilon} > 0$ и $\bar{\delta} > 0$ направление v NM однозначно определяется и удовлетворяет оценкам

$$u_1 + v_1 = O(\|u_1\|\|u\|) + O(\|u\|^3), \quad (2.11)$$

$$u_2 + v_2 = \frac{1}{2}u_2 + O(\|u_1\|) + O(\|u\|^2) \quad (2.12)$$

при $u \rightarrow 0$. При этом $\|u + v\| \geq \gamma\|u\|$ с некоторым фиксированным $\gamma > 0$, и, поскольку $\|u_1\| = O(\|u\|)$,

$$\frac{\|u_1 + v_1\|}{\|u + v\|^2} = O\left(\frac{\|u_1 + v_1\|}{\|u\|^2}\right) = O\left(\frac{\|u_1\|}{\|u\|}\right) + O(\|u\|) = O(1), \quad (2.13)$$

т. е. $u_1 + v_1 = O(\|u + v\|^2)$. Далее, согласно [11, теорема 3.1], если $\varepsilon > 0$ и $\delta > 0$ достаточно малы, то для любого $u^0 \in K_{\varepsilon, \delta}(0, \bar{v})$ последовательность $\{u^k\}$ с шагами v^k NM однозначно определяется, сходится к 0, причем $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(0, \bar{v})$ для любого k , откуда и из оценки (2.13) следует, что $u_1^k = O(\|u^k\|^2)$ при $k \rightarrow \infty$. Тогда, снова привлекая (2.11),

$$\frac{\|u_1^k + v_1^k\|}{\|u^k + v^k\|^3} = O\left(\frac{\|u_1^k + v_1^k\|}{\|u^k\|^3}\right) = O\left(\frac{\|u_1^k\|}{\|u^k\|^2}\right) + O(1) = O(1),$$

т. е. имеет место оценка

$$u_1^k = O(\|u^k\|^3) \quad (2.14)$$

при $k \rightarrow \infty$, которая согласуется с [2, теорема 2.1, (iii)], где эта оценка приведена без доказательства.

Для рNM, согласно [11, лемма 3.1], оценки (2.11), (2.12) заменяются на

$$u_1 + v_1 = O(\|u_1\|\|u\|) + O(\|u\|\|\Omega(u)\|) + O(\|\omega(u)\|) + O(\|u\|^3), \quad (2.15)$$

$$u_2 + v_2 = \frac{1}{2}u_2 + O(\|u_1\|) + O(\|\Pi\Omega(u)\|) + O(\|u\|^{-1}\|\Pi\omega(u)\|) + O(\|u\|^2) \quad (2.16)$$

при $u \rightarrow 0$. Для метода LM, согласно (2.5), поскольку в этом случае $\omega(\cdot) \equiv 0$, эти оценки принимают вид

$$u_1 + v_1 = O(\|u_1\|\|u\|) + O(\|u\|^{2\theta}) + O(\|u\|^3),$$

$$u_2 + v_2 = \frac{1}{2}u_2 + O(\|u_1\|) + O(\|u\|^{2\theta-1}) + O(\|u\|^2),$$

что при $\theta \geq 3/2$ дает (2.11), (2.12). Для метода LPN, согласно (2.9), поскольку в этом случае $\Omega(\cdot) \equiv 0$, оценки (2.15), (2.16) сразу принимают вид (2.11), (2.12). Значит, и для этих методов в указанных случаях проходят рассуждения, проведенные выше для NM, и приводящие к оценке (2.14).

Следующая лемма уточняет в линейно-квадратичном случае оценку (1.7).

Лемма 2.1. Пусть A — $(n \times n)$ -матрица, а $B : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ — симметричное билинейное отображение, и пусть для некоторого $\bar{v} \in \mathbb{R}^p$ линейный оператор $\widehat{B}(\bar{v})$ невырожден.

Тогда найдутся $\bar{\varepsilon} > 0$ и $\bar{\delta} > 0$ такие, что для $u \in K_{\bar{\varepsilon}, \bar{\delta}}(0, \bar{v})$ матрица $A + B[u]$ невырождена, причем для $w \in \operatorname{im} A$ справедлива оценка

$$(A + B[u])^{-1}w = O(\|w\|) \quad (2.17)$$

при $w \rightarrow 0$.

Доказательство. Для заданных $u \in \mathbb{R}^p$ и $w \in \operatorname{im} A$, умножая левую и правую части уравнения

$$(A + B[u])v = w \quad (2.18)$$

на $I - \Pi$ и Π , получаем уравнения

$$(A + (I - \Pi)B[u])v_1 = w - (I - \Pi)B[u]v_2, \quad \Pi B[u]v_2 = -\Pi B[u]v_1. \quad (2.19)$$

Обозначим через $\widehat{A} : (\ker A)^\perp \rightarrow \operatorname{im} A$ сужение линейного оператора A на $(\ker A)^\perp$, а через $\mathcal{A}(u) : (\ker A)^\perp \rightarrow \operatorname{im} A$ — сужение линейного оператора $A + (I - \Pi)B[u]$ на $(\ker A)^\perp$. Тогда

$$\mathcal{A}(u) = \widehat{A} + O(\|u\|),$$

причем оператор $\widehat{A}$ обратим, откуда и, например, из [3, лемма A.6] следует, что для u достаточно близких к 0 оператор $\mathcal{A}(u)$ тоже обратим, причем

$$(\mathcal{A}(u))^{-1} = \widehat{A}^{-1} + O(\|u\|)$$

при $u \rightarrow 0$. Тогда первое уравнение в (2.19) при любом $v_2 \in \mathbb{R}^p$ имеет единственное решение

$$v_1 = (\mathcal{A}(u))^{-1}(w - (I - \Pi)B[u]v_2) = O(\|w\|) + O(\|u\|)v_2 \quad (2.20)$$

при $u \rightarrow 0$ и $w \rightarrow 0$.

Подставляя выражение из (2.20) во второе уравнение в (2.19), получаем

$$\mathcal{B}(u)v_2 = \Pi O(\|u\|\|w\|), \quad (2.21)$$

где линейный оператор $\mathcal{B}(u) : \ker A \rightarrow (\operatorname{im} A)^\perp$ в левой части имеет вид

$$\mathcal{B}(u) = \widehat{B}(u) + O(\|u\|^2).$$

Из невырожденности $\widehat{B}(\bar{v})$, снова применяя [3, лемма A.6], получаем существование $\bar{\varepsilon} > 0$ и $\bar{\delta} > 0$ таких, что для $u \in K_{\bar{\varepsilon}, \bar{\delta}}(0, \bar{v})$ оператор $\mathcal{B}(u)$ невырожден, причем

$$(\mathcal{B}(u))^{-1} = (\widehat{B}(u))^{-1} + O(1) = O(\|u\|^{-1})$$

при $u \rightarrow 0$. Отсюда следует, что (2.21), а значит, и второе уравнение в (2.19) имеет единственное решение

$$v_2 = O(\|w\|) \quad (2.22)$$

при $u \rightarrow 0$ и $w \rightarrow 0$.

Тем самым показано, что для $u \in K_{\bar{\varepsilon}, \bar{\delta}}(0, \bar{v})$ уравнение (2.18) имеет единственное решение $v = v_1 + v_2$, т.е. матрица $A + B[u]$ невырождена, причем из (2.20) и (2.22) вытекает оценка (2.17). $\square$

Для NM, согласно (2.2),

$$(A + B[u])v = -\frac{1}{2}(A + B[u])u - \frac{1}{2}Au,$$

и в предположениях леммы 2.1 для $u \in K_{\varepsilon, \bar{\delta}}(0, \bar{v})$

$$v = -\frac{1}{2}u - \frac{1}{2}(A + B[u])^{-1}Au = -\frac{1}{2}u + O(\|Au\|) = -\frac{1}{2}u + O(\|u_1\|)$$

при $u \rightarrow 0$. Поэтому, с учетом (2.14), если $u^0 \in K_{\varepsilon, \delta}(0, \bar{v})$ с достаточно малыми $\varepsilon > 0$ и $\delta > 0$, то

$$u^k + 2v^k = O(\|u_1^k\|) = O(\|u^k\|^3)$$

при $k \rightarrow \infty$.

Для метода LM, аналогично построениям выше, для $u \in K_{\varepsilon, \bar{\delta}}(0, \bar{v})$ выводится следующее обобщение оценки (2.7):

$$v = -\frac{1}{2}u + O(\|u\|^{2\theta-1}) + O(\|Au\|) = -\frac{1}{2}u + O(\|u\|^{2\theta-1}) + O(\|u_1\|)$$

при $u \rightarrow 0$, откуда и из (2.14) вытекает оценка

$$u^k + 2v^k = O(\|u^k\|^{2\theta-1}) + O(\|u_1^k\|)$$

при $k \rightarrow \infty$. Тогда, согласно (2.14),

$$u^k + 2v^k = O(\|u^k\|^{\min\{2\theta-1, 3\}}) \quad (2.23)$$

и, в частности, при $\theta \geq 2$

$$u^k + 2v^k = O(\|u^k\|^3) \quad (2.24)$$

при $k \rightarrow \infty$, как и для NM.

Наконец, для метода LPN аналогичным образом для $u \in K_{\varepsilon, \bar{\delta}}(0, \bar{v})$ выводится оценка

$$v = -\frac{1}{2}u + O(\|u\|^2) + O(\|Au\|) = -\frac{1}{2}u + O(\|u\|^2) + O(\|u_1\|)$$

при $u \rightarrow 0$, а значит, согласно (2.14), и оценка

$$u^k + 2v^k = O(\|u^k\|^2) \quad (2.25)$$

при $k \rightarrow \infty$.

Предложение 2.2. В предположениях предложения 2.1 найдутся $\varepsilon > 0$ и $\delta > 0$ такие, что для $u^0 \in K_{\varepsilon, \delta}(0, \bar{v})$, для уравнения (1.1) с линейно-квадратичным отображением Φ из (2.1) справедливо следующее:

- NM с подзадачей (1.2) однозначно определяет последовательности $\{v^k\}$ шагов и $\{u^k\}$ приближений, и для соответствующих экстраполированных приближений выполняется оценка (2.24) при $k \rightarrow \infty$;

- метод LM с подзадачей (1.3), использующей $\sigma(u) = O(\|\Phi(u)\|^\theta)$ с фиксированным $\theta \geq 3/2$, однозначно определяет последовательности $\{v^k\}$ шагов и $\{u^k\}$ приближений, и для соответствующих экстраполированных приближений выполняется оценка (2.23) при $k \rightarrow \infty$;
- метод LPN с подзадачей (1.5) определяет последовательности $\{v^k\}$ шагов и $\{u^k\}$ приближений, и для любых таких последовательностей, для соответствующих экстраполированных приближений выполняется оценка (2.25) при $k \rightarrow \infty$.

Полученные в этом разделе результаты существенно используют линейно-квадратичный характер отображения Φ . За пределами линейно-квадратичного случая оценка (2.24) для метода LM с $\theta = 2$, и даже для NM, может нарушаться, в том числе и при $u \in \ker A$, и эффект от экстраполяции для MN и методов LM и LPN при этом может выравниваться, что будет продемонстрировано в следующем разделе.

3. Глобализованные алгоритмы и численные результаты

Алгоритм 1 ниже, глобализующий сходимость NM, использует страховочные шаги градиентного метода для функции $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$,

$$\varphi(u) = \frac{1}{2} \|\Phi(u)\|^2, \quad (3.1)$$

градиент которой имеет вид

$$\varphi'(u) = (\Phi'(u))^\top \Phi(u). \quad (3.2)$$

Такие шаги используются в тех случаях, когда направление NM либо не удастся определить, либо оно оказывается «слишком длинным» в смысле нарушения теста (3.3). Этот алгоритм был предложен в [12, алгоритм 3.1], где были установлены свойства его глобальной сходимости.

Алгоритм 1. Фиксируем параметры $C > 0$, $\tau > 0$, $\rho \in (0, 1)$ и $\varkappa \in (0, 1)$. Выбираем $u^0 \in \mathbb{R}^p$ и полагаем $k = 0$.

1. Если $\Phi(u^k) = 0$, стоп.
2. Вычисляем v^k как решение линейного уравнения (1.2). Если v^k не удастся вычислить, или v^k нарушает неравенство

$$\|v^k\| \leq \max\{C, 1/\|\Phi(u^k)\|^\tau\}, \quad (3.3)$$

переходим к шагу 4.

3. Полагаем $\alpha = 1$. Если выполняется неравенство

$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \rho\alpha)\|\Phi(u^k)\|, \quad (3.4)$$

полагаем $\alpha_k = \alpha$ и переходим к шагу 6. В противном случае заменяем α на $\varkappa\alpha$ и проверяем снова неравенство (3.4) до тех пор, пока оно не выполнится, после чего полагаем $\alpha_k = \alpha$ и переходим к шагу 6.

4. Полагаем $v^k = -\varphi'(u^k)$, где функция φ определена в (3.1) (см. (3.2)). Если $v^k = 0$, стоп.
5. Полагаем $\alpha = 1$. Если выполняется неравенство Армихо

$$\varphi(u^k + \alpha v^k) \leq \varphi(u^k) - \rho \alpha \|v^k\|^2, \quad (3.5)$$

полагаем $\alpha_k = \alpha$. В противном случае заменяем α на $\varkappa \alpha$ и проверяем снова неравенство (3.5) до тех пор, пока оно не выполнится, после чего полагаем $\alpha_k = \alpha$.

6. Полагаем $u^{k+1} = u^k + \alpha_k v^k$, увеличиваем k на 1 и переходим к шагу 1.

В отличие от NM, метод LM может быть глобализован «чистым» образом, без каких-либо гибридных элементов и, в частности, без необходимости использования каких-либо страховочных шагов.

Алгоритм 2. Фиксируем параметры $\theta > 0$, $\rho \in (0, 1)$ и $\varkappa \in (0, 1)$. Выбираем $u^0 \in \mathbb{R}^p$ и полагаем $k = 0$.

1. Если $\Phi(u^k) = 0$, стоп.
2. Полагаем $\sigma(u^k) = \|\Phi(u^k)\|^\theta$, и вычисляем v^k как решение линейного уравнения (1.4). Если $v^k = 0$, стоп.
3. Полагаем $\alpha = 1$. Если выполняется неравенство

$$\varphi(u^k + \alpha v^k) \leq \varphi(u^k) - \frac{1}{2} \rho \sigma(u^k) \alpha \|v^k\|^2, \quad (3.6)$$

где функция φ определена в (3.1), полагаем $\alpha_k = \alpha$. В противном случае заменяем α на $\varkappa \alpha$ и проверяем снова неравенство (3.6) до тех пор, пока оно не выполнится, после чего полагаем $\alpha_k = \alpha$.

4. Полагаем $u^{k+1} = u^k + \alpha_k v^k$, увеличиваем k на 1 и переходим к шагу 1.

Алгоритм 2 (с $\theta = 2$) и анализ его глобальной сходимости были предложены в [16]; см. также [7, алгоритм 6.2, теорема 6.2].

Метод LPN также допускает «чистую» глобализацию сходимости, разработанную в [17, алгоритм 1, теорема 3.1]. Алгоритм 3 ниже основан на одномерном поиске для функции $f : \mathbb{R}^p \rightarrow \mathbb{R}$,

$$f(u) = \|\Phi(u)\|_\infty. \quad (3.7)$$

Алгоритм 3. Фиксируем параметры $\rho \in (0, 1)$ и $\varkappa \in (0, 1)$. Выбираем $u^0 \in \mathbb{R}^p$ и полагаем $k = 0$.

1. Если $\Phi(u^k) = 0$, стоп.
2. Вычисляем (v^k, γ_k) как решение задачи линейного программирования (1.5), в которой используется ∞ -норма. Если для величины

$$\Delta_k = -f(u^k)(1 - \gamma_k f(u^k)),$$

где функция f определена в (3.7), выполняется $\Delta_k = 0$, стоп.

3. Полагаем $\alpha = 1$. Если выполняется неравенство

$$f(u^k + \alpha v^k) \leq f(u^k) + \rho \alpha \Delta_k, \quad (3.8)$$

полагаем $\alpha_k = \alpha$. В противном случае заменяем α на $\kappa \alpha$ и проверяем снова неравенство (3.8) до тех пор, пока оно не выполнится, после чего полагаем $\alpha_k = \alpha$.

4. Полагаем $u^{k+1} = u^k + \alpha_k v^k$, увеличиваем k на 1 и переходим к шагу 1.

В экспериментах, результаты которых представлены ниже, алгоритм 3 был также снабжен всеми предложенными в [17, разд. 5] модификациями, улучшающими его поведение, кроме немоного одномерного поиска. В процедурах одномерного поиска использовались параметры $\rho = 0.01$ и $\kappa = 0.5$. В алгоритме 1 использовались $C = 10^7$ и $\tau = 2$.

Запуск завершался с успехом в случае выполнения для некоторого $k \leq 100$ условия

$$\|\Phi(u^k)\| \leq 10^{-8},$$

либо условия

$$\min\{\|\Phi(u^k)\|, \|\Phi(u^{k-1} + 2v^{k-1})\|\} \leq 10^{-8} \quad (3.9)$$

в случае использования экстраполяции. В остальных случаях запуск считался неудачным, как и при выполнении для некоторого k условия

$$\|(\Phi'(u^k))^T \Phi(u^k)\| \leq 10^{-20}$$

для алгоритмов 1 и 2, и

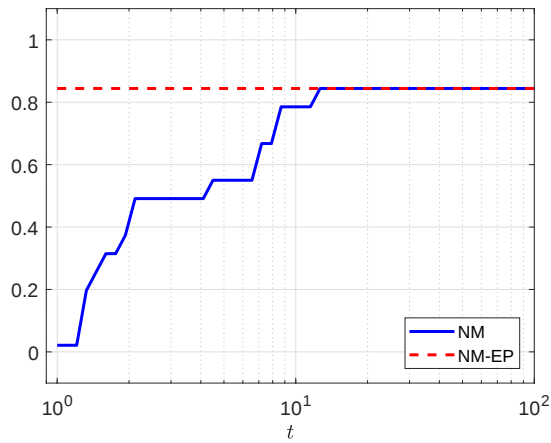
$$|\Delta_k| \leq 10^{-16}$$

для алгоритма 3, либо условия $\alpha \|v^k\| \leq 10^{-16}$ в процессе одномерного поиска на соответствующих шагах алгоритмов.

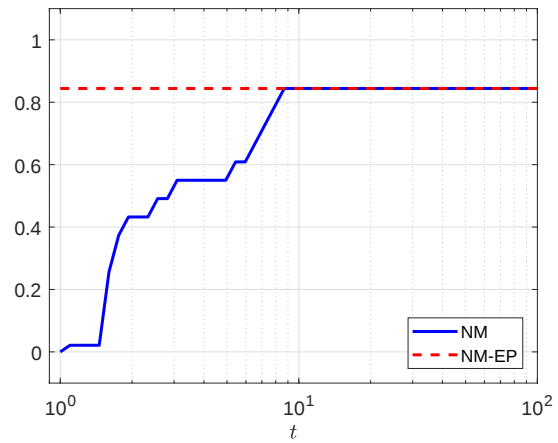
В алгоритме 2 правило выбора параметра регуляризации было модифицировано следующим образом, исключая слишком большие значения этого параметра вдали от решений: $\sigma(u) = \min\{1, \|\Phi(u)\|^\theta\}$, причем использовалось значение $\theta = 2$.

Эксперимент проводился на всех задачах из тестового набора MGH [18], для которых известны точные решения, за исключением трех задач с линейными отображениями Φ . Некоторые из выбранных таким образом задач имеют больше уравнений, чем переменных, и в таких случаях лишние уравнения удалялись (всегда линейные, за исключением теста Beale, в котором выбор уравнения для удаления не влияет существенным образом на поведение алгоритмов). Информация о полученном в результате тестовом наборе содержится в [12, таблица 1]. Далее, выбранные задачи трансформировались посредством техники из [19] таким образом, чтобы известное решение $\bar{u}$ оставалось решением модифицированного уравнения, но при этом выполнялось $\text{rank } \Phi'(\bar{u}) = p - 1$ (за исключением трех задач Powell singular, Extended Powell singular, и Variably dimensioned, которые имеют известные особые решения в своей оригинальной форме). Задачи Variably dimensioned и Brown almost-linear использовались в двух версиях: с $p = 10$ и с $p = 500$.

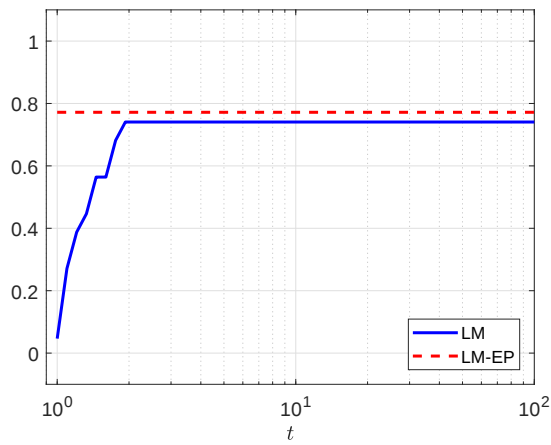
Запуски сравниваемых алгоритмов осуществлялись из 100 одинаковых случайных начальных точек, равномерно распределенных в l_∞ -шаре (гиперкубе) радиуса 1 с центром в решении.



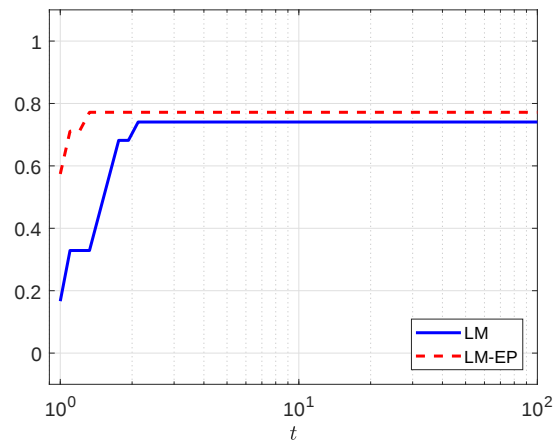
(a) Алгоритм 1, по количеству итераций



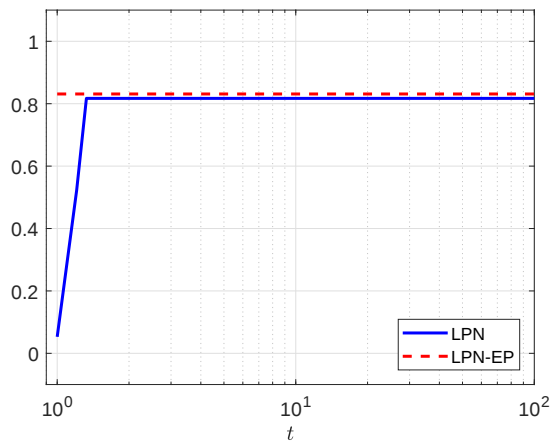
(b) Алгоритм 1, по времени запуска



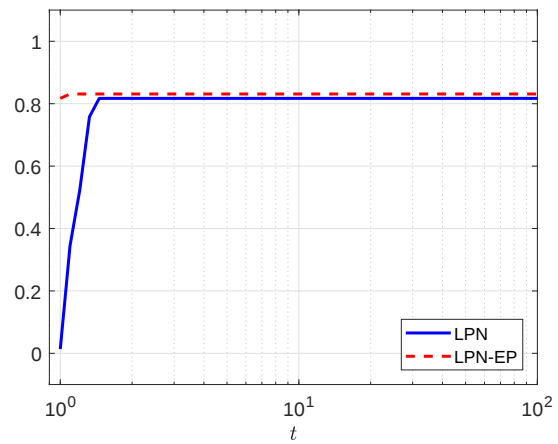
(c) Алгоритм 2, по количеству итераций



(d) Алгоритм 2, по времени запуска

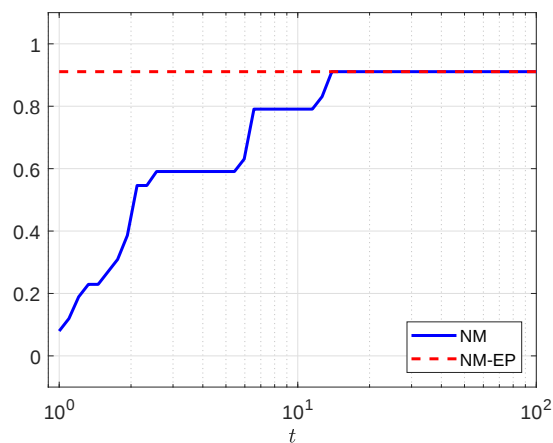


(e) Алгоритм 3, по количеству итераций

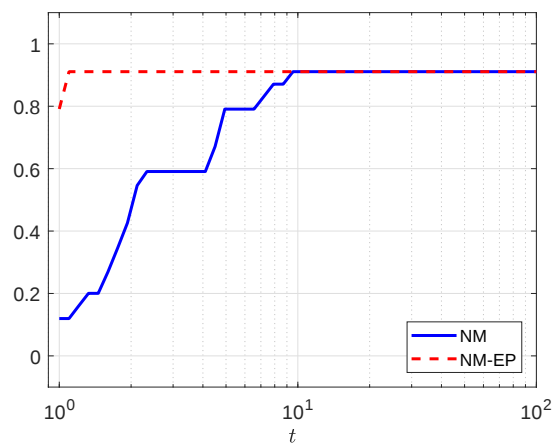


(f) Алгоритм 3, по времени запуска

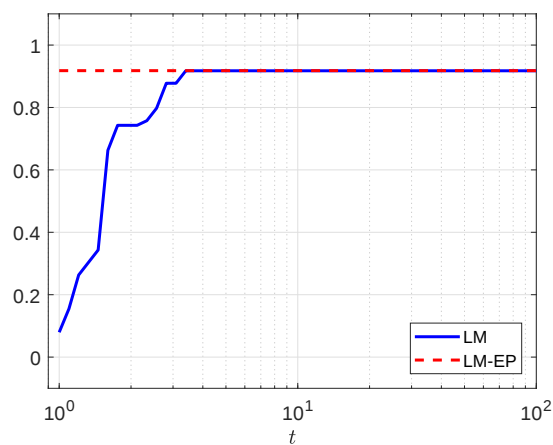
Рис. 2. Сравнение на наборе MGH



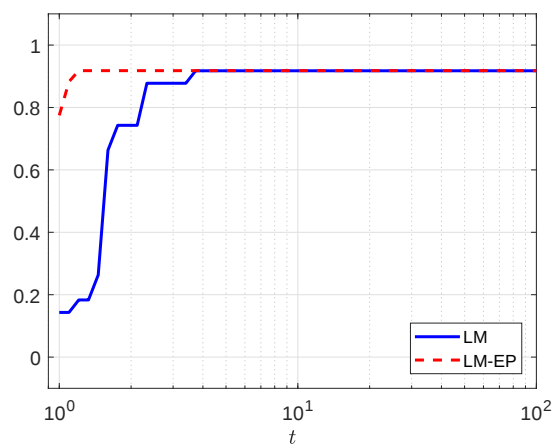
(a) Алгоритм 1, по количеству итераций



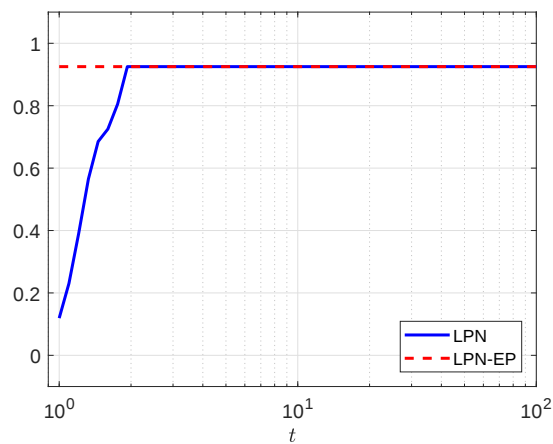
(b) Алгоритм 1, по времени запуска



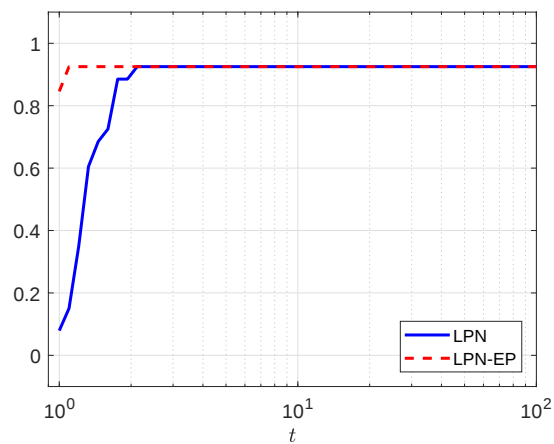
(c) Алгоритм 2, по количеству итераций



(d) Алгоритм 2, по времени запуска



(e) Алгоритм 3, по количеству итераций



(f) Алгоритм 3, по времени запуска

Рис. 3. Сравнение на наборе Misc

Результаты представлены в форме так называемых performance profiles, изначально предложенных в [20], и адаптированных в [21] на случай многократных запусков для каждой тестовой задачи, с использованием средних характеристик эффективности и процентов успешных запусков. На рис. 2 для каждого алгоритма показан график функции, значение которой в t по существу соответствует доле задач в наборе, для которых средний показатель эффективности для данного алгоритма был хуже наилучшего не более, чем в t раз. При этом считается, что результат любого неудачного запуска в бесконечное число раз хуже любого другого. Значения функции на правом конце (при больших t) соответствует доле успешных запусков алгоритма, а на левом — доле задач, для которых его средний показатель эффективности был наилучшим. В данном сравнении в роли показателей эффективности использовались количество итераций и время запуска.

Представленные результаты демонстрируют, что эффект от экстраполяции для метода LM слабее, чем для NM, а для метода LPN слабее, чем для LM.

Эксперимент на тестовом наборе MGN был дополнен аналогичным экспериментом на наборе Misc(ellaneous), состоящим из небольших систем нелинейных уравнений с особыми решениями, собранных из различных источников (первые 11 тестов взяты из [22, разд. 5]). А именно, набор состоит из следующих задач.

Misc 1. Это уравнение (1.1) из примера 2.1, с $p = 1$, $\Phi(u) = u^2$; изолированное особое решение $\bar{u} = 0$.

Misc 2. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1, 2u_2^2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 3. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2 - u_2^2, u_1 u_2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 4. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 + u_2, -u_1 - u_2 + u_1 u_2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 5. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2, u_2^2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 6. Уравнение (1.1) с $p = 2$, $\Phi(u) = (2(u_1 - u_2^2), u_2^2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 7. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1(u_1^2 + u_2), u_2(1 + u_2))$; изолированное особое решение $\bar{u} = (0, 0)$ (имеются также неособые решения $(0, -1)$, $(1, -1)$, $(-1, -1)$).

Misc 8. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 + u_2^2, 3u_1 u_2 / 2 + u_2^2(1 + u_2))$; изолированное особое решение $\bar{u} = (0, 0)$ (имеется также неособое решение $(-4, 2)$).

Misc 9. Уравнение (1.1) с $p = 3$, $\Phi(u) = (u_1 + u_2 + u_3 - 1, u_1^3 / 5 + u_2^2 / 2 - u_3 + u_3^2 / 2 + 1/2, u_1 + u_2 + u_3^2 / 2 - 1/2)$; изолированное особое решение $\bar{u} = (0, 0, 1)$ (имеется также другое особое решение $(-5/2, 5/2, 1)$).

Misc 10. Уравнение (1.1) с $p = 3$, $\Phi(u) = (u_1 + u_1 u_2 + u_2^2, -2u_1 + u_1^2 + u_2^2, u_1 + u_3^2)$; изолированное особое решение $\bar{u} = (0, 0, 0)$.

Misc 11 (дискретизованное H -уравнение Чандрасекара). Уравнение (1.1) с выбираемым значением p (в экспериментах использовалось $p = 5$),

$$\Phi_i(u) = u_i - \frac{i}{2n} u_i \sum_{j=1}^n \frac{u_j}{i+j} - 1, \quad i \in \{1, \dots, p\};$$

для $p = 5$ приближенное изолированное особое решение

$$\bar{u} = (1.359753, 1.688205, 2.005894, 2.318350, 2.627810).$$

Misc 12. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 + au_2^2/2, u_2^2/2)$ с вещественным параметром a (в экспериментах использовалось $a = \sqrt{15}$); изолированное особое решение $\bar{u} = (0, 0)$.

Misc 13. Это уравнение (1.1) из примера 2.2, с $p = 2$, $\Phi(u) = (u_1 + au_2^2/2, u_1u_2 + u_2^2/2)$ с вещественным параметром a (в экспериментах использовалось $a = 1$, как и в примере 2.2); изолированное особое решение $\bar{u} = (0, 0)$ (имеется также неособое решение $(-1/(2a), 1/a)$).

Misc 14. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2 + u_2^3, u_1u_2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 15. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 + u_1u_2 + u_2^2, u_1^2 - 2u_1 + u_2^2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 16. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2 - u_2, u_1^2 + u_2^2)$; изолированное особое решение $\bar{u} = (0, 0)$.

Misc 17. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2 - u_2^2, 3u_1^2 - 3u_2^2)$; неизолитованное решение $\bar{u} = (0, 0)$.

Misc 18. Уравнение (1.1) с числом переменных $p = 5$, и с числом уравнений, равным 4, $\Phi(u) = (u_1 + u_2 + u_3^2 + u_4^2 + u_5^2 - 2, u_1 - u_2 + u_3^2 + u_4^2 + u_5^2, -u_3^2 + u_4^2 + u_5^2, u_3^2 + u_4^2 - u_5^2)$; неизолитованное решение $\bar{u} = (1, 1, 0, 0, 0)$. Для этой задачи НМ работает как метод Гаусса–Ньютона, использующий решение минимальной нормы итерационного уравнения (1.2).

Misc 19. Уравнение (1.1) с $p = 3$ из [1, приложение] с параметрами $k = 1$, $m = 1$; изолированное особое решение $\bar{u} = (0, 0, 0)$.

Misc 20. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1^2 + u_2^2 - 2, e^{u_1-1} + u_2^2 - 2)$; неособое решение $\bar{u} = (1, -1)$, которое трансформировалось в особое посредством упомянутой выше в связи с набором MGN процедуры из [19] (у исходного уравнения имеется также другое неособое решение $(1, 1)$).

Misc 21. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 - 10, u_1u_2 - 5000)$; неособое решение $\bar{u} = (5000, 10)$, которое трансформировалось в особое процедурой из [19].

Misc 22. Уравнение (1.1) с $p = 2$, $\Phi(u) = (e^{u_1+u_2^2} - 1, u_1 + u_2 - \sin(3(u_1 + u_2)))$; изолированное особое решение $\bar{u} = (0, 0)$.

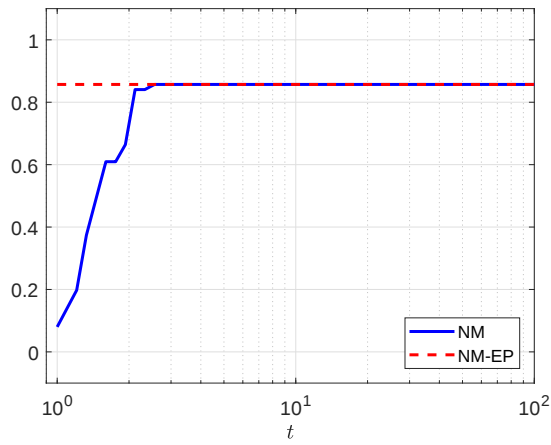
Misc 23. Уравнение из Misc 22, к решению которого дополнительно применялась процедура из [19].

Misc 24. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1, 1 - u_1/(u_1 + 0.1) + 2u_2^2)$; неособое решение $\bar{u} = (5000, 10)$, которое трансформировалось в особое процедурой из [19].

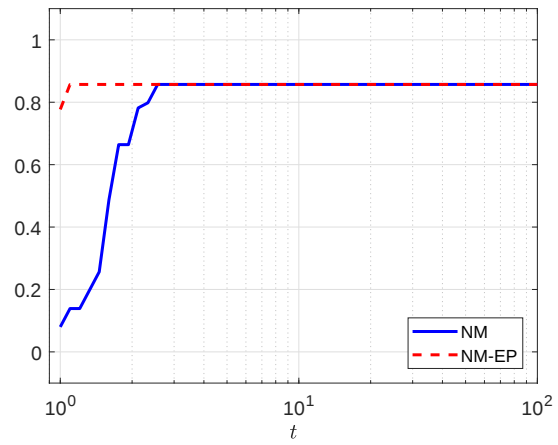
Misc 25. Уравнение (1.1) с $p = 2$, $\Phi(u) = (u_1 + u_2^2, 2(u_1 - 1)u_2)$; неособое решение $\bar{u} = (0, 0)$, которое трансформировалось в особое процедурой из [19].

Результаты эксперимента на наборе Misc представлены на рис. 3, и выводы из них аналогичны выводам для MGN: эффект от экстраполяции для НМ больше, чем для метода LM, а для метода LM больше, чем для метода LPN.

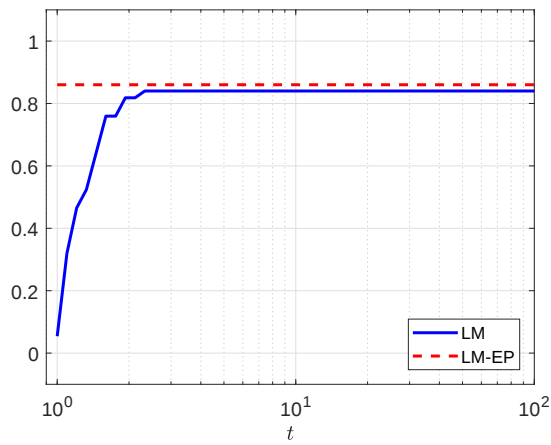
Дальнейший эксперимент был направлен на демонстрацию того, что за пределами линейно-квадратичного случая эффект от экстраполяции для НМ и методов LM и LPN выравнивается, что согласуется с рассмотрениями в разд. 3. Для этого из наборов MGN и Misc были отобраны задачи, в которых отображение Φ не является линейно-квадратичным, а именно, Freudenstein and Roth, Beale, Helical valley, Gulf research and development, Box three-dimensional, Biggs EXP6, Trigonometric, Brown almost-linear;



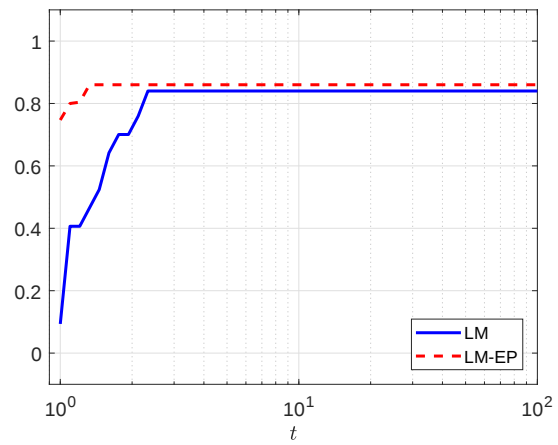
(a) Алгоритм 1, по количеству итераций



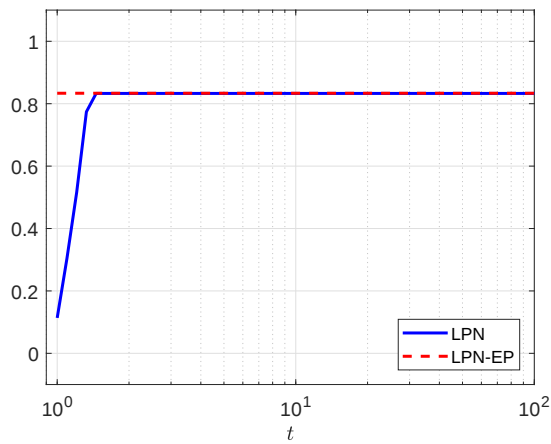
(b) Алгоритм 1, по времени запуска



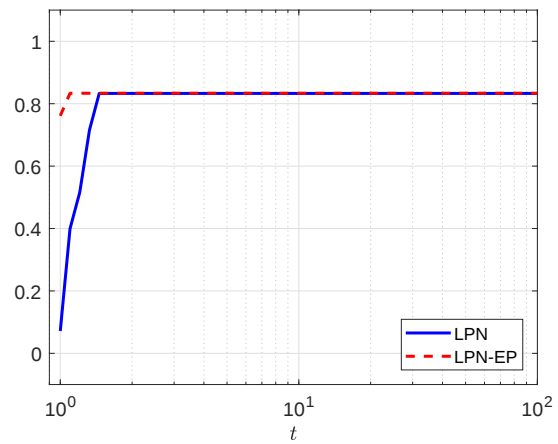
(c) Алгоритм 2, по количеству итераций



(d) Алгоритм 2, по времени запуска



(e) Алгоритм 3, по количеству итераций



(f) Алгоритм 3, по времени запуска

Рис. 4. Сравнение на задачах, не являющихся линейно-квадратичными

Misc 7-9, Misc 14, Misc 20, Misc 22-24. Результаты на рис. 4 показывают, что на данной выборке тестов сравниваемые алгоритмы имеют схожий эффект от экстраполяции, хотя для метода LPN эффект все же меньше, чем для других алгоритмов.

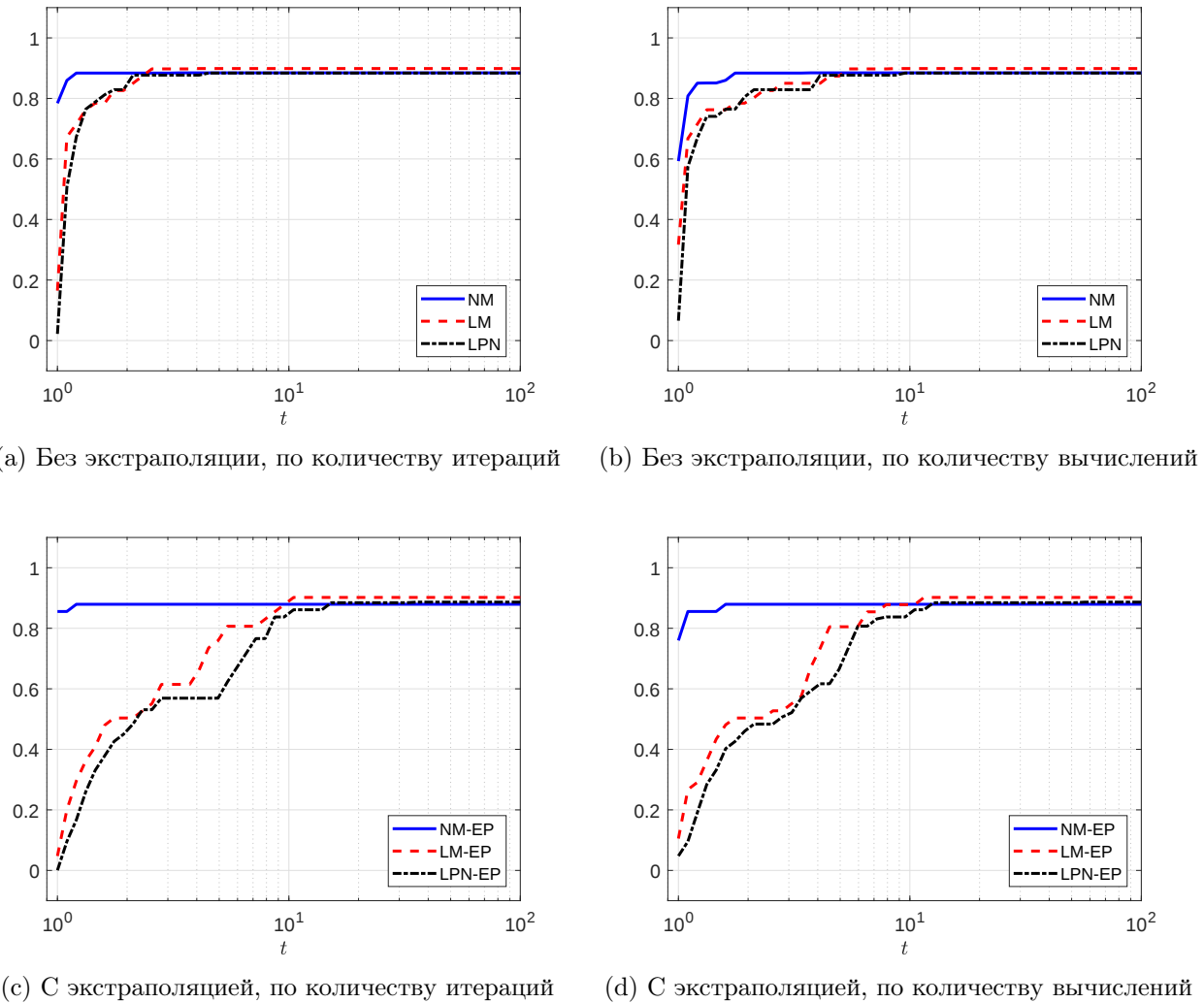


Рис. 5. Сравнение на полном тестовом наборе

На рис. 5 приведенные выше результаты дополнены сравнением трех алгоритмов 1, 2 и 3 между собой. Сравнение по времени здесь не приводится, поскольку метод LPN для рассматриваемых здесь уравнений без дополнительных ограничений неизбежно проигрывает по этому показателю, так как решение его подзадач (задач линейного программирования) является гораздо более затратным, чем решение подзадач NM и метода LM (систем линейных уравнений). Вместо этого приводится сравнение по другой важной характеристике, а именно, по количеству вычислений Φ . Важное наблюдение состоит в том, что по этой характеристике методы с экстраполяцией могут быть существенно менее эффективны, чем по количеству итераций, поскольку на каждой итерации требуют по крайней мере одного лишнего вычисления Φ для проверки условия останова (3.9).

Для алгоритмов с одномерным поиском эффект от экстраполяции во многом определяется асимптотическим принятием полного шага, поскольку эта техника основана на характере сходимости к особым решениям именно полношаговых ньютоновских методов.

Таблица 1. Процент среднего количества последних полных шагов

Тест	NM	LM	LPN
Rosenbrock	98.67	49.70	70.87
Freudenstein and Roth	99.44	98.95	99.75
Brown badly scaled	32.14	NaN	100.00
Beale	99.08	100.00	99.93
Helical valley	95.77	33.95	49.63
Gulf research and development	24.05	13.64	17.39
Box three-dimensional	90.71	28.46	44.85
Powell singular	100.00	100.00	100.00
Wood	99.23	94.97	98.16
Biggs EXP6	83.20	4.20	6.91
Extended Rosenbrock	93.48	26.68	87.02
Extended Powell singular	100.00	100.00	100.00
Variably dimensioned, $p = 10$	100.00	100.00	99.82
Variably dimensioned, $p = 500$	100.00	100.00	62.91
Trigonometric	NaN	53.00	NaN
Brown almost-linear, $p = 10$	98.24	100.00	95.34
Brown almost-linear, $p = 500$	100.00	98.27	65.41
Misc 1	100.00	100.00	100.00
Misc 2	100.00	100.00	100.00
Misc 3	100.00	100.00	100.00
Misc 4	98.65	100.00	100.00
Misc 5	100.00	100.00	100.00
Misc 6	100.00	100.00	100.00
Misc 7	97.89	99.74	99.18
Misc 8	98.57	100.00	100.00
Misc 9	97.75	100.00	98.77
Misc 10	98.50	100.00	100.00
Misc 11	100.00	100.00	100.00
Misc 12	100.00	100.00	100.00
Misc 13	96.75	99.33	99.50
Misc 14	99.86	100.00	100.00
Misc 15	98.43	100.00	100.00
Misc 16	97.94	100.00	100.00
Misc 17	87.50	100.00	100.00
Misc 18	99.76	100.00	100.00
Misc 19	89.61	83.42	86.06
Misc 20	89.69	55.38	79.58
Misc 21	100.00	NaN	13.21
Misc 22	98.07	100.00	99.07
Misc 23	13.29	18.83	13.12
Misc 24	96.56	52.57	57.29
Misc 25	76.92	73.04	84.40

В этих экспериментах единственный тест, на котором иногда наблюдались случаи неполного последнего шага (для всех трех алгоритмов), — это Misc 23, в котором $\Phi'(\bar{u}) = 0$, но 2-регулярности не имеет места ни по какому направлению. Таблица 1 содержит информацию о проценте среднего количества последних полных шагов на один успешный запуск, от общего количества итераций.

Интересное наблюдение состоит в том, что по времени алгоритм 2 показывает лучший результат на примерно 70% задач (по указанным выше причинам эти результаты здесь не приводятся), а по количеству вычислений Φ и особенно по количеству итераций такой процент гораздо ниже; см. рис. 5a, 5b. Возможное объяснение состоит в том, что по средним значениям принимаемых алгоритмами параметров длины шага алгоритм 2 также оказывается лучшим на сравнимой доле задач, как и по среднему количеству последних полных шагов; см. таблицу 1.

В заключение кратко опишем результаты эксперимента с методом LM, в котором в правиле для параметра регуляризации $\sigma(u) = \min\{\bar{\sigma}, \|\Phi(u)\|^2\}$ использовались разные значения $\bar{\sigma}$, а именно, наряду с 1, рассматривались значения 10^{-1} , 10^{-5} и 10^{-10} . На данных тестовых наборах версии метода LM с меньшими значениями $\bar{\sigma}$ (что в принципе позволяет алгоритму 2 вдали от решений генерировать и принимать более длинные шаги, оказываются несколько более эффективными, но при этом несколько менее робастными, однако разница невелика. Вместе с тем, выбор $\bar{\sigma} = 1$ значительно проигрывает по времени рассматриваемым альтернативам.

References

- [1] A. Griewank, *Analysis and Modification of Newton's Method at Singularities*, PhD Thesis, Australian National University, Canberra, 1980.
- [2] A. Griewank, "On solving nonlinear equations with simple singularities or nearly singular solutions", *SIAM Review*, **27** (1985), 537–563.
- [3] A. F. Izmailov, M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2014, 573 pp.
- [4] K. Levenberg, "A method for the solution of certain non-linear problems in least squares", *Quarterly of Applied Mathematics*, **2** (1944), 164–168.
- [5] D. W. Marquardt, "An algorithm for least squares estimation of non-linear parameters", *SIAM J.*, **11** (1963), 431–441.
- [6] Дж. Дэннис, Р. Шнабель, *Численные методы безусловной оптимизации и решение нелинейных уравнений*, Мир, М., 1988; англ. ориг.: J. E. Dennis, Jr., R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, 1983.
- [7] A. Fischer, A. F. Izmailov, M. V. Solodov, "The Levenberg–Marquardt method: an overview of modern convergence theories and more", *Computational Optimization and Applications*, **89**:1 (2024), 33–67.
- [8] F. Facchinei, A. Fischer, M. Herrich, "An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions", *Mathematical Programming*, **146** (2014), 1–36.
- [9] А. Ф. Измаилов, А. А. Третьяков, *2-регулярные решения нелинейных задач. Теория и численные методы*, Физматлит, М., 1999. [A. F. Izmailov, A. A. Tret'yakov, *2-Regular Solutions of Nonlinear Problems. Theory and Numerical Methods*, Fizmatlit Publ., Moscow, 1999 (In Russian)].
- [10] A. F. Izmailov, A. S. Kurennoy, M. V. Solodov, "Critical solutions of nonlinear equations: local attraction for Newton-type methods", *Mathematical Programming*, **167**:2 (2018), 355–379.
- [11] A. Fischer, A. F. Izmailov, M. Jelitte, "Behavior of Newton-type methods near critical solutions of nonlinear equations with semismooth derivatives", *Journal of Optimization Theory and*

- Applications*, 2023, New Trends in Optimization, Control, and Their Applications: A Special Issue Dedicated to Boris Sh. Mordukhovich.
- [12] A. Fischer, A. F. Izmailov, M. V. Solodov, “Accelerating convergence of the globalized Newton method to critical solutions of nonlinear equations”, *Computational Optimization and Applications*, **78**:1 (2021), 273–286.
 - [13] A. Fischer, A. F. Izmailov, M. V. Solodov, “Unit stepsize for the Newton method close to critical solutions”, *Mathematical Programming*, **187**:1–2 (2021), 697–721.
 - [14] A. Fischer, A. F. Izmailov, M. Jelitte, “Newton-type methods near critical solutions of piecewise smooth nonlinear equations”, *Computational Optimization and Applications*, **80**:2 (2021), 587–615.
 - [15] A. Griewank, “Starlike domains of convergence for Newton’s method at singularities”, *Numerische Mathematik*, **35** (1980), 95–111.
 - [16] A. Fischer, P. K. Shukla, “A Levenberg–Marquardt algorithm for unconstrained multicriteria optimization”, *Operations Research Letters*, **36** (2008), 643–646.
 - [17] A. Fischer, M. Herrich, A. F. Izmailov, M. V. Solodov, “A globally convergent LP-Newton method”, *SIAM J. on Optimization*, **26** (2016), 2012–2033.
 - [18] J. J. Moré, B. S. Garbow, K. E. Hillstom, “Testing unconstrained optimization software”, *ACM Transactions of Mathematical Software*, **7** (1981), 17–41.
 - [19] R. B. Schnabel, P. D. Frank, “Tensor methods for nonlinear equations”, *SIAM J. on Numerical Analysis*, **21** (1984), 815–843.
 - [20] E. D. Dolan, J. J. More, “Benchmarking optimization software with performance profiles”, *Mathematical Programming*, **91** (2002), 201–213.
 - [21] A. F. Izmailov, M. V. Solodov, E. I. Uskov, “Combining stabilized SQP with the augmented Lagrangian algorithm”, *Computational Optimization and Applications*, **62** (2015), 405–429.
 - [22] А. Ф. Измаилов, “О новых реализациях 2-фактор-метода”, *Журнал вычислительной математики и математической физики*, **55**:6 (2015), 933–946; англ. пер.: A. F. Izmailov, “New implementations of the 2-factor method”, *Computational Mathematics and Mathematical Physics*, **55**:6 (2015), 922–934.

Информация об авторах

Измаилов Алексей Ферилович, доктор физико-математических наук, профессор кафедры исследования операций, Московский государственный университет им. М. В. Ломоносова, г. Москва, Российская Федерация. E-mail: izmaf@cs.msu.ru
ORCID: <https://orcid.org/0000-0001-9851-0524>

Усков Евгений Иванович, кандидат физико-математических наук, научный сотрудник, Тамбовский государственный университет им. Г. Р. Державина, г. Тамбов, Российская Федерация. E-mail: euskov@cs.msu.ru
ORCID: <https://orcid.org/0000-0002-3639-0317>

Конфликт интересов отсутствует.

Для контактов:

Измаилов Алексей Ферилович
 E-mail: izmaf@cs.msu.ru

Поступила в редакцию 30.05.2024 г.
 Поступила после рецензирования 10.09.2024 г.
 Принята к публикации 22.11.2024 г.

Information about the authors

Alexey F. Izmailov, Doctor of Physical and Mathematical Sciences, Professor of the Operations Research Department, Lomonosov Moscow State University, Moscow, Russian Federation. E-mail: izmaf@cs.msu.ru
ORCID: <https://orcid.org/0000-0001-9851-0524>

Evgeniy I. Uskov, Candidate of Physical and Mathematical Sciences, Researcher, Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: euskov@cs.msu.ru
ORCID: <https://orcid.org/0000-0002-3639-0317>

There is no conflict of interests.

Corresponding author:

Alexey F. Izmailov
 E-mail: izmaf@cs.msu.ru

Received 30.05.2024
 Reviewed 10.09.2024
 Accepted for press 22.11.2024