

СХОДИМОСТЬ ВЫЧИСЛИТЕЛЬНЫХ АЛГОРИТМОВ В ВАРИАЦИОННЫХ ЗАДАЧАХ ИДЕНТИФИКАЦИИ ТИПА ПРОНИ

А.А. Ломов

Институт математики им. С.Л. Соболева СО РАН;

Новосибирский государственный университет

Россия, 630090, Новосибирск, пр. Академика Коптюга, 4; ул. Пирогова, 2

E-mail: lomov@math.nsc.ru; a.lomov@g.nsu.ru

Ключевые слова: разностные уравнения, идентификация коэффициентов, модифицированный метод Прони, вариационный метод идентификации, нелинейная задача на собственные значения, обратные итерации, сходимость.

Аннотация: Модифицированный метод Прони извлечения синусоид и экспонент из наблюдений формулируется как вариационная задача идентификации коэффициентов разностного уравнения. Последняя соответствует нелинейной задаче на собственные значения матрицы из наблюдений в переменной метрике. Исследуется сходимость трех вычислительных алгоритмов с большим радиусом сходимости, основанных на обратных итерациях с переменной матрицей.

1. Введение

Будем рассматривать задачу аппроксимации сеточной функции $x = [x[1]; \dots; x[N]] \in \mathbb{R}^N$ решениями $z = [z[1]; \dots; z[N]] \in \mathbb{R}^N$ разностного уравнения с вещественными коэффициентами γ_i

$$(1) \quad z[k+n] + \gamma_{n-1}z[k+n-1] + \dots + \gamma_0z[k] = 0, \quad k \in \overline{1, N-n}.$$

Ищутся начальные условия $z[1], \dots, z[n]$ и коэффициенты $\gamma_0, \dots, \gamma_{n-1}$, минимизирующие при условии (1) целевую функцию

$$(2) \quad J = \|x - z\|^2 = (x - z)^T (x - z) \rightarrow \min_{\gamma_i, z: (1)}.$$

Задача идентификации (2) относится к классу вариационных [4] и отличается сложным характером изоповерхностей функционала J [2, 13]. К ней сводится широко известный в литературе модифицированный метод Прони извлечения синусоид и экспонент из наблюдений [7–9, 14].

Известны вычислительные алгоритмы минимизации (2) с большими радиусами и скоростью сходимости [1, 6, 7, 10]. Они основаны на обратных итерациях с обновляемой обратной матрицей для нахождения минимального собственного вектора в нелинейной задаче на собственные значения специальных матриц, зависящих от наблюдений

x и коэффициентов γ_i . Опишем алгоритмы О70 [6], Е73 [1], О75 [7]. Уравнение (1) соответствует системе $G^T z = 0$ с треплицевой матрицей $G^T \doteq \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & 1 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & & & \gamma_0 & \gamma_1 & \dots & 1 \end{bmatrix}$. Минимизация по $z[1], \dots, z[n]$ в (2) приводит к выражению $J = x^T G C G^T x = \gamma^T V^T C V \gamma$, $C \doteq (G^T G)^{-1}$, $\gamma \doteq [\gamma_0, \dots, \gamma_{n-1}, 1]^T$, $V \doteq \begin{bmatrix} x[1] & \dots & x[n+1] \\ \vdots & & \vdots \\ x[N-n] & \dots & x[N] \end{bmatrix}$. Пусть $\lambda_1(Q)$ и $p_1(Q)$ — минимальное собственное число и соответствующий собственный вектор матрицы $Q \doteq V^T C V$. Эта матрица сформирована из наблюдений $V = V(x)$, и ввиду зависимости C от параметров γ_i возникает задача на собственные значения Q в «переменной метрике» (Riemannian SVD [11]). Предполагается, что число $\lambda_1(Q)$ некратное (это условие полноты наблюдения x).

Алгоритм О70 [6]:

$$(3) \quad \gamma_{[k+1]} = p_1(Q(\gamma_{[k]})), \quad k \geq 0.$$

Минимальный собственный вектор $p_1(Q)$ вычисляется обратными итерациями $\tau = Q^{-1}\gamma_{[k]}$, $\gamma_{[k+1]} = \tau/\|\tau\|$, $k \geq 0$, $p_1(Q) = \lim_{k \rightarrow \infty} \gamma_{[k]}$. Если использовать результат одного шага, получим алгоритм Е73 [1]:

$$(4) \quad \gamma_{[k+1]} = \tau/\|\tau\|, \quad \tau = (Q(\gamma_{[k]}))^{-1} \gamma_{[k]}, \quad k \geq 0.$$

Известно [12], что итерации О70 (3) не находят точного минимума (2). Напротив, алгоритм О75 [7] имеет неподвижную точку в нуле градиента J' . Верно выражение

$$J' = B(\gamma)\gamma = (Q - L^T L)\gamma,$$

где матрица L имеет треплицевую структуру $L^T = \begin{bmatrix} l_1 & \dots & l_{N-n} & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & & l_1 & \dots & l_{N-n} \end{bmatrix} \in \mathbb{R}^{(n+1) \times N}$, и $l \doteq [l_1; \dots; l_{N-n}] = C G^T x = C V \gamma$ — вектор множителей Лагранжа в задаче (2) (см. [4, 7]). Пусть $p_0(B)$ — собственный вектор матрицы $B(\gamma) = (Q - L^T L)$, соответствующий собственному числу с наименьшим модулем. Алгоритм О75 [7, (5.4)], [8], [9, (15)]:

$$(5) \quad \gamma_{[k+1]} = p_0(B(\gamma_{[k]})), \quad k \geq 0.$$

Для нахождения начального приближения может использоваться оценка ортогональной регрессии ($C = I$) [8]:

$$(6) \quad \gamma_{[0]} = p_1(V^T V).$$

Экспериментально показано [3, 7, 9, 13], что по скорости и радиусу сходимости при больших возмущениях итерации (3), (4), (5) существенно превосходят универсальные алгоритмы на основе оценок градиента и гессиана J . Теоретические обоснования сходимости обратных итераций с обновляемой обратной матрицей в задаче (2) затруднены сложным характером нелинейностей, и до сих пор они слабо изучены [13]. В [9] доказана устойчивость О75 в точке минимума. Для малых возмущений $\min J^{1/2} \ll \|x\|$ в [5] доказана глобальная сходимость О70 в малую окрестность точки минимума.

В настоящем докладе анонсируются результаты исследования глобальной сходимости Е73 (4) в малую окрестность минимума J и приводятся данные вычислительных экспериментов по сходимости О70, Е73, О75 при больших возмущениях.

2. Глобальная сходимость E73

Поведение итераций E73 (4) описывается следующими двумя теоремами. Определим нормы векторов и матриц: $\|x\| \doteq \sqrt{x^T x}$, $\|A\| \doteq \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$. Наблюдения $x = z_* + \eta$ порождены добавлением возмущения η к «истинной» траектории $G^T(\gamma_*)z_* = 0$, и выполнено условие полноты $\text{rank } V(z_*) = n$, равносильное некратности собственного числа $\lambda_1(Q_*) = 0$, $Q_* \doteq V(z_*)^T C V(z_*)$, $C > 0$.

Теорема 1. Пусть в области $\Gamma \subset \mathbb{R}^{n+1}$, содержащей истинную точку γ_* и точки $\gamma_{[k]}$, $k \geq k_*$ при некотором $k_* \geq 0$, выполнено условие на малость $\lambda_1(Q)$ и тем самым на малость нормы возмущения $\|\eta\|$:

$$(7) \quad \frac{\lambda_1}{\lambda_2 |\alpha_1|} c_6 + \frac{2n \cdot \sqrt{\lambda_1} \cdot \|C\| \cdot \|V\|}{\lambda_2 - \lambda_1} < 1,$$

$$\alpha_1 \doteq p_1^T \gamma, \quad c_6 \doteq 4n \|C\|^{1/2} + \left(2 + \frac{1}{|\alpha_1|}\right) \sqrt{n+1} c_{\max} + \left(1 + \frac{1}{|\alpha_1|}\right),$$

$$c_{\max} \doteq 2n \cdot \|C\| \cdot \|V\| \max_{i=2, n+1} \left(\frac{\sqrt{\lambda_i}}{\lambda_i - \lambda_{i-1}} \right) =$$

$$= 2n \cdot \|C\|^{1/2} \cdot \sqrt{\lambda_{n+1}} \cdot \max_{i=2, n+1} \left(\frac{\sqrt{\lambda_i}}{\lambda_i - \lambda_{i-1}} \right).$$

Тогда отображение $\gamma_{[k]} \rightarrow \gamma_{[k+1]}$ при $k \geq k_*$ в итерации (4) является сжимающим, следовательно, имеет единственную неподвижную точку в Γ .

Замечание 1. Для итераций O70 верна теорема 1 без первого слагаемого в левой части неравенства (7) [5].

Теорема 2. Все элементы $\gamma_{[k]}$ в итерациях (4) при $k \geq k_* = 1$ остаются в окрестности истинной точки γ_* :

$$(8) \quad \frac{\|\gamma_{[k]} - \gamma_*\|}{\|\gamma_*\|} \leq \frac{\alpha' \cdot \|C_*\| \cdot \|V_*\|}{\lambda_{2*}} \varepsilon + O(\varepsilon^2) \leq \frac{\alpha \cdot \|C_*\| \cdot \|V_*\|}{\lambda_{2*}} \varepsilon,$$

$\varepsilon \doteq \|V(\eta)\| = \|V - V_*\|$. Константа α' определена условием $\|C_{[k]}\| \leq \alpha' \|C_*\|$, $k \geq 0$, которое можно рассматривать как ограничение на начальное значение $\gamma_{[0]}$, а константа $\alpha > \alpha'$ определяется условием мажорирования слагаемых $O(\varepsilon^2)$ в (8).

Замечание 2. Теорема 2 гарантирует попадание в область (8) после первой итерации независимо от выполнения условий сжатия в начальной точке $\gamma_{[0]}$. Теорема 2 верна и для итераций O70 (3) [5], возможно, с другим значением константы α .

В следующей теореме устанавливается, при каких возмущениях ε условие сжатия (7) выполнено для значений $\gamma_{[k]}$, определяемых неравенством (8). Это гарантирует наличие неподвижной точки итераций (4) внутри области (8).

Теорема 3. Пусть выполнено $\varepsilon < \frac{1}{5n \|C_*\|^{1/2}} \cdot \frac{\lambda_2}{\alpha \|C_*\| \cdot \|V_*\| \cdot \|\gamma_*\|}$ и $\varepsilon < \|V_*\|$. Тогда следующее неравенство является достаточным для выполнения условия теоремы 1 в области (8): $\frac{\varepsilon}{\|V_*\|} < \frac{\varepsilon_*}{f_1^{1/2}}$, где $\varepsilon_* \doteq \frac{1}{\sqrt{1 + \frac{c_6}{|\alpha_1|}}} \frac{\sqrt{\omega^2 + \lambda_2 - \omega}}{\sqrt{2 \|C_*\|}}$, $\omega \doteq \frac{4n \|C_*\| \cdot \|V_*\|}{\sqrt{1 + \frac{c_6}{|\alpha_1|}}}$,

$$f_1 \doteq \|V_*\|^2 (\alpha_0 + \alpha_1 \|V_*\| + \alpha_2 \|V_*\|^2), \quad \alpha_0 \doteq 1 + \frac{2n(\sqrt{\lambda_2} \|C_*\|^{1/2} + \|C_*\| \cdot \|V_*\|)}{\lambda_2} \cdot \|V_*\|,$$

$$\alpha_1 \doteq \frac{2n(\sqrt{\lambda_2} \|C_*\|^{1/2} + 3 \|C_*\| \cdot \|V_*\|)}{\lambda_2}, \quad \alpha_2 \doteq \frac{4n \|C_*\|}{\lambda_2}.$$

3. Расчеты

В [5] опубликованы результаты расчетов по сходимости алгоритма O70 из разных начальных приближений, проверяющие утверждение теоремы 1 (замечание 1) для двух систем (1) второго порядка $n = 2$ со значениями коэффициентов $\gamma_{0,1}$ и корней $\zeta_{1,2}$ характеристических многочленов:

$$(9) \quad \gamma_0 = 0,980, \quad \gamma_1 = -1,883, \quad \zeta_{1,2} = 0,942 \pm 0,306i;$$

$$(10) \quad \gamma_0 = 0,957, \quad \gamma_1 = -1,956, \quad \zeta_1 = 0,989, \quad \zeta_2 = 0,967.$$

Далее приведем результаты экспериментов по сходимости алгоритмов O70, E73, O75 и комбинаций O70–O75, E73–НР при больших возмущениях для уравнения (1) порядка $n = 2$ со значениями коэффициентов $\gamma_0 = 0,941, \gamma_1 = -1,823$ и начальным приближением ОР (6). Сокращение «НР» обозначает универсальный алгоритм типа Ньютона–Рафсона с вычислением первой и оценкой второй производной целевой функции J (2). Для коэффициентов (9), (10) с начальным приближением ОР результаты качественно те же.

Длина отрезка наблюдения $N = 100$. Примеры наблюдений x и оценок \hat{z} траектории z_* при $\sigma = 0.5$ и $\sigma = 1.0$ показаны на рис. 1.

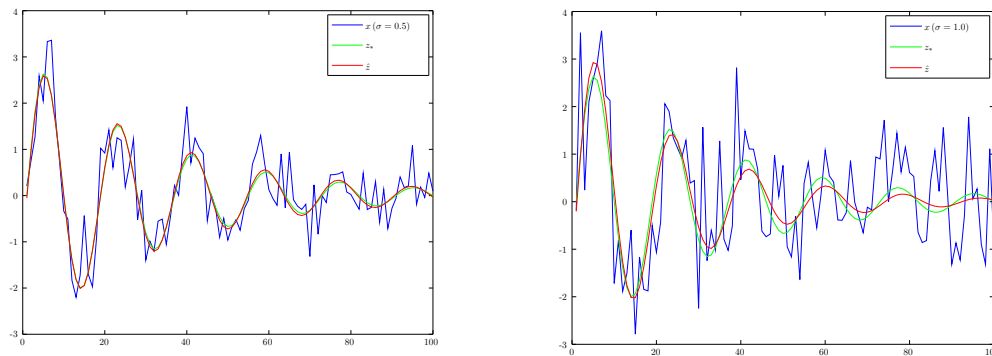


Рис. 1. Наблюдения x и оценки \hat{z} траектории z_* при $\sigma = 0.5$ и $\sigma = 1.0$

Для каждой из $M = 100$ реализаций случайных возмущений в наблюдениях производился запуск пяти алгоритмов из начального приближения ОР (6) (таблица 1). Считается, что алгоритм сошелся к искомой оценке $\hat{\gamma}_{0,1}$ коэффициентов $\gamma_{0,1}$, если для соответствующей оценки $\hat{z} \doteq [I - GCG^T]x$, $G = G(\hat{\gamma}_{0,1})$ «истинной» траектории z_* выполнено $\|\hat{z} - z_*\|/\|z_*\| \leq 0,4$.

Таблица 1. Частоты сходимости

σ	O75	O70	O70–O75	E73	E73–НР
0.5	48%	66%	69%	100%	100%
1.0	19%	33%	36%	77%	79%

4. Заключение

Теоремы 1–3 гарантируют сходимость алгоритмов О70, Е73 из любого начального приближения при условии полноты наблюдений x и малой нормы возмущений η . Для алгоритма О70 и системы (9) теоретический допустимый относительный уровень возмущений имеет порядок 10^{-5} [5]. Расчеты при практически значимых больших возмущениях с начальным значением ОР показывают существенное преимущество итераций Е73 перед О70 и О75.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (19-01-00754).

Список литературы

1. Егоршин А.О., Будянов В.П. Сглаживание сигналов и оценивание динамических параметров в автоматических системах с помощью ЦВМ // Автометрия. 1973. № 1. С. 78-82.
2. Костин В.И. О точках экстремума одной функции // Управляемые системы. 1984. Т. 24. С. 35-42.
3. Хасина Е.М., Ломов А.А. Audio Files Compression with the STLS-ESM method (Сжатие аудио-файлов методом STLS-ESM) // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2015. № 5 (229). С. 88-96.
4. Егоршин А.О. Об отслеживании параметров экстремума в вариационной задаче идентификации // Вестн. НГУ. Сер. матем., мех., информ. 2011. Т. 11, Вып. 3. С. 95-113.
5. Ломов А.А. О сходимости алгоритма с обратными итерациями в модифицированном методе Прони // Сибирские электронные математические известия. 2018. Т. 15. С. 1513-1529. <http://semr.math.nsc.ru/v15/p1513-1529.pdf>
6. Osborne M.R. A class of nonlinear regression problems // Data Representation / Eds. R. S. Anderssen and M. R. Osborne. St. Lucia: University of Queensland Press, 1970. P. 94-101.
7. Osborne M.R. Some special nonlinear least squares problems // SIAM J. Numer. Anal. 1975. Vol. 12. P. 571-592.
8. Smyth G.K. MPRONY MATLAB Function: Fits a sum of exponential functions to data by a modified Prony method. 1990. http://read.pudn.com/downloads72/sourcecode/math/262881//mprony.m_.htm
9. Osborne M.R., Smyth G.K. A modified Prony algorithm for fitting functions defined by difference equations // SIAM J. Sci. Statist. Comput. 1991. Vol. 12. P. 362-382.
10. Moor De B. Structured total least squares and L_2 approximation problems // Linear Algebra Appl. 1993. Vol. 188-189. P. 163-207.
11. Moor De B. The Riemannian singular value decomposition // SVD and Signal Processing. III. Algorithms, Architectures and Applications / Eds. M. Moonen and B. DeMoor. Amsterdam, New York: Elsevier, 1995. P. 61-78.
12. Osborne M.R., Smyth G.K. A Modified Prony Algorithm for Exponential Function Fitting // SIAM Journal of Scientific Computing. 1995. Vol. 16. P. 119-138.
13. Petersson J., Holmström K. A review of the parameter estimation problem of fitting positive exponential sums to empirical data // Applied Mathematics and Computation. 2002. Vol. 126, No. 1. P. 31-61.
14. Pereyra V., Scherer G. Exponential data fitting // Exponential Data Fitting and Its Applications / V. Pereyra, G. Scherer (Eds.) Bentham Science Publishers, 2010. P. 1-26.