

Обработка астрометрических наблюдений

10. Решение нелинейных задач МНК

кто: Семенцов В.Н.

когда: 16 апреля, 2021

дано переопределенная система нелинейных уравнений
располагаем информацией о **точности** вычисления функций
надо найти устойчивое приближение к локальному
решению системы
желательно оценить глобальный оптимум

Задача безусловной нелинейной минимизации суммы квадратов

$$F(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^M f_i(\mathbf{x})^2 = \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|_2^2 \mapsto \min. \quad (1)$$

$\mathbf{x} = \{x_0, \dots, x_N\}$ — вектор искомых параметров.

Задача (1) возникает, например, когда есть модель $\phi(\mathbf{x}, t_i)$ и реальные измерения y_i , тогда $f_i(\mathbf{x}) = \phi_i(\mathbf{x}, t_i) - y_i$.

Обычно $M \gg N$.

Если модель составлена «правильно», можно ожидать, что величина $\|f(\mathbf{x})\|$ окажется «малой». При малом числе наблюдений хорошее согласование можно получить для любой модели.

Ньютоновские итерации

Классический ньютоновский метод решения задачи (1) принимает следующий вид. На k -й итерации решается

$$\nabla \mathbf{x}^{(k)} + H(\mathbf{x}^{(k)}) \mathbf{p}^{(k)} = 0 \quad (2)$$

относительно $\mathbf{p}^{(k)}$; затем положим

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{p}^{(k)}, \quad (3)$$

где $\alpha^{(k)}$ — некоторый шаг, такой, что $F(\mathbf{x}^{(k+1)}) < F(\mathbf{x}^{(k)})$. Если гессиан положительно определен для всех $\mathbf{x} \in E^n$, то итерация (2) является алгоритмом крутого спуска по норме

$$\|y\|_{H^{(k)}} = \left(y^T H^{(k)} y \right)^{1/2}, \quad (4)$$

и, следовательно, $\alpha^{(k)}$ всегда положительно. Численно устойчивым методом формирования оценки $\mathbf{p}^{(k)}$ в этом случае является разложение $H^{(k)}$ по методу Холецкого в виде

$$H^{(k)} = L^{(k)} D^{(k)} \left(L^{(k)} \right)^T, \quad (5)$$

Матричные производные нелинейной векторной функции

Матрица Якоби системы функций:

$$J(x) = \begin{pmatrix} \frac{\partial u_1}{\partial x_1}(x) & \frac{\partial u_1}{\partial x_2}(x) & \cdots & \frac{\partial u_1}{\partial x_n}(x) \\ \frac{\partial u_2}{\partial x_1}(x) & \frac{\partial u_2}{\partial x_2}(x) & \cdots & \frac{\partial u_2}{\partial x_n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial u_m}{\partial x_1}(x) & \frac{\partial u_m}{\partial x_2}(x) & \cdots & \frac{\partial u_m}{\partial x_n}(x) \end{pmatrix} \quad (6)$$

Матрица Гессе:

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \quad (7)$$

Матричные производные - 2

- Для нелинейной задачи наименьших квадратов:
- Градиент (1):

$$\nabla F = J(\mathbf{x})^T \mathbf{f}(\mathbf{x}), \quad (8)$$

- Матрица Гессе:

$$H(\mathbf{x}) = J(\mathbf{x})^T J(\mathbf{x}) + Q(\mathbf{x}), \quad (9)$$

где $Q(\mathbf{x}) = \sum_{i=1}^M f_i(\mathbf{x}) H_i(\mathbf{x})$.

Метод Гаусса-Ньютона

- Решаем задачу (1) аналогично методу Ньютона:

$$F(\mathbf{x}^{(k)} + \mathbf{p}) \approx F_{(k)} + \nabla F_{(k)}^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T H^{(k)} \mathbf{p}, \quad (10)$$

где $\mathbf{x}^{(k)}$ — очередная итерация в решении задачи (1).

- С учетом (8, 9) система (10) примет вид:

$$\left(J_{(k)}^T J_{(k)} + Q_{(k)} \right) \mathbf{p}_{(k)} = -J_{(k)}^T \mathbf{f}_{(k)} \quad (11)$$

- вблизи оптимума Q **приближается к нулевой матрице** и направление ньютоновского поиска можно **аппроксимировать** решением системы:

$$J_{(k)}^T J_{(k)} \mathbf{p}_{(k)} = -J_{(k)}^T \mathbf{f}_{(k)} \quad (12)$$

Метод Гаусса-Ньютона - 2

- Эта система (12) заведомо разрешима, содержит только первые производные \mathbf{f} и представляет собой оптимальный вектор в линейной задаче МНК:

$$\frac{1}{2} \|J_{(k)}\mathbf{p} + \mathbf{f}_{(k)}\|_2^2 \mapsto \min \quad (13)$$

- Когда норма $\|Q(\mathbf{x}^{(k)})\|$ близка к нулю и матрица Якоби $J_{(k)}$ имеет *полный столбцовый ранг* (столбцы линейно независимы, т.е. имеется актуальная зависимость от всех параметров \mathbf{x}), то направление, задаваемое алгоритмом Гаусса-Ньютона, мало отличается от ньютоновского.
- Метод может достигать квадратичной скорости сходимости, хотя использует только первые производные.

Метод Гаусса-Ньютона - 3

- «лобовая» реализация (12) с предварительным формированием $J^T J$ приведет к большим погрешностям в определении направления поиска, поскольку число обусловленности матрицы $J^T J$ — велико
- потеря точности при итерациях метода Гаусса-Ньютона — реальная угроза ещё и потому, что конкретные математические модели, лежащие в основе схемы метода, часто оказываются некорректно определенными
- методы решения (12) выбираются устойчивые, типа SVD или ортогональной факторизации (процесс Грама-Шмидта)
- решение (13) в случае дефекта ранга J и, следовательно, многозначности — следует выбирать с минимальной евклидовой нормой

Метод Гаусса-Ньютона - 4

- Каждая реализация метода Гаусса-Ньютона, в которой при линейной зависимости столбцов $J_{(k)}$ ищется решение (13) минимальной нормы, должна включать оценку ранга $J_{(k)}$ на каждом шаге.
- Это означает, что в процедуре факторизации, используемой для расчета направления итерации Гаусса-Ньютона, сам численный алгоритм в машинной реализации становится проблемно-зависимым, т.е. требуется вводить масштабирование и критерии «малости».
- Проблему легко увидеть на примере: пусть f — двумерная функция с компонентами f_1 и f_2 и ее матрица Якоби в некоторой точке равна:

$$J = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix},$$

где $\varepsilon \ll 1$, а значения f_1 и f_2 — порядка 1.

Метод Гаусса-Ньютона - 5

- Тогда допустив, что ранг J равен двум, мы получим направление Гаусса-Ньютона:

$$-\begin{pmatrix} f_1 \\ f_2/\varepsilon \end{pmatrix}$$

- А посчитав ε пренебрежимо малой, получим ранг J единица, а направление

$$-\begin{pmatrix} f_1 \\ 0 \end{pmatrix}$$

практически ортогональное первому, (которое, в свою очередь почти ортогонально градиенту)

- общего подхода **нет**
- если F «похожа» на плохо обусловленную квадратичную форму, то ранг лучше брать по максимуму, приближая к чистому методу Ньютона
- если матрица $J_{(k)}$ «почти вырождена», то ранг лучше занижать, чтобы итерации сильно не болтались

Метод Левенберга-Маркварта

- Направление поиска определяется как решение системы уравнений вида

$$\left(J_{(k)}^T J_{(k)} + \lambda_{(k)} I \right) \mathbf{p}_{(k)} = -J_{(k)}^T \mathbf{f}_{(k)} \quad (14)$$

$\lambda_{(k)} \geq 0$, т.е. некоторая смесь ньютоновского и градиентного поиска.

- Система (14) дает решение задачи (13) **при условии**

$$\|\mathbf{p}\|_2 \leq \Delta \quad (15)$$

где Δ связано с $\{\lambda_{(k)}\}$.

- Подобрав $\{\lambda_{(k)}\}$ достаточно большим, всегда можно добиться оптимизации, $F(\mathbf{x}_{(k)} + \mathbf{p}_{(k)}) < F_{(k)}$, но скорость сходимости упадет, как у чисто градиентного метода

Квазиньютоновские методы

- Методы Гаусса-Ньютона и Левенберга-Маркварта основаны на предположении о доминировании слагаемого $J_{(k)}^T J_{(k)}$. Это неверно для задач с **большой невязкой**
- в ЭТОТ МОМЕНТ может оказаться полезным задуматься **о качестве конкретной нелинейной модели**, лежащей в основе вычислений (и проблем с вычислениями)
- аппроксимация честной матрицы Гессе, требующей **вторых производных**, последовательностью приближений на основе **градиентного поиска**

Скорректированный метод Гаусса-Ньютона

- Рассмотрим SVD разложение матрицы Якоби на очередной итерации:

$$J_{(k)} = U \begin{pmatrix} S \\ 0 \end{pmatrix} V^T \quad (16)$$

- подставив выражение в итерацию метода Гаусса-Ньютона (11), получим систему:

$$(S^2 + V^T Q_{(k)} V) V^T \mathbf{p}_N = -S \bar{\mathbf{f}} \quad (17)$$

где вектор $\bar{\mathbf{f}}$ составлен из первых N компонент M -мерного вектора $U^T \mathbf{f}_{(k)}$.

- Вторым слагаемым можно аналогично (12) пренебречь и получить скорректированный алгоритм Гаусса-Ньютона.
- На каждом шаге учитываются только направления с «большими» сингулярными числами.

Решение задач большой размерности

- квазиньютоновские методы дают плотное заполнение матрицы Гессе, кроме некоторых специальных случаев (трехдиагональные матрицы и т.п.)
- минимум квадратичной функции $\Phi(x) = c^T x + \frac{1}{2}x^T Gx$ можно искать методом **сопряженных градиентов**

$$\nabla\Phi(x_k) = c + Gx_k \quad (18)$$

Глобальный экстремум мультимодальной функции

- математически разработаны методы линейного и целочисленного программирования, **неприменимые в случае нелинейного МНК**
- глобальный Монте-Карло
- методы его «улучшения»

масштабирование

STUN

целевой функции с целью сгладить локальный рельеф (stochastic tunnelling) метод оптимизации, основанный на методе Монте-Карло, меняет целевую функцию случайным «скачком» от текущего вектора решения к другому с разницей в значении функции ΔE . Вероятность принятия такого скачка обычно выбирается равной $\min(1; \exp(-\beta \cdot \Delta E))$ (критерий Метрополиса) с соответствующим параметром β .

Глобальный экстремум мультимодальной функции - 2

STUN Общая идея STUN состоит в том, чтобы заменить случайное распределение адекватно конкретной задаче.

Эта цель достигается путем выборки методом Монте-Карло преобразованной функции, которая лишена этой медленной динамики. В «стандартной форме» преобразование выглядит

$f_{STUN} := 1 - \exp(-\gamma \cdot (E(x) - E_o))$, где E_o — минимальное значение функции, найденное до сих пор.

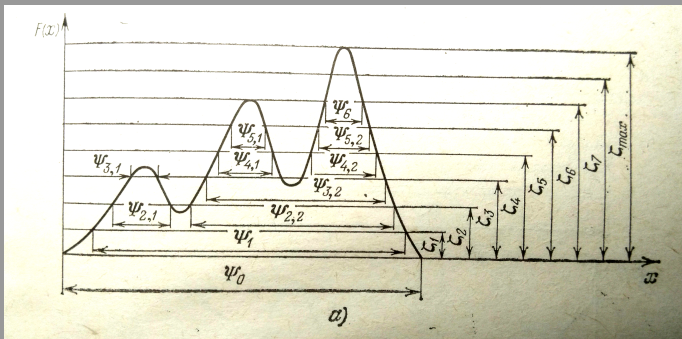
Это преобразование сохраняет локусы минимумов.

f_{STUN} затем используется вместо E в исходном алгоритме, дающем новую вероятность принятия $\min(1; \exp(-\beta \cdot \Delta f_{STUN}))$.

■ и т.д.

далее преобразование характеристик многомерной поверхности уровня к двумерному виду

Глобальный экстремум мультимодальной функции - 3 — методом Ψ - преобразования



- преобразование характеристик многомерной поверхности уровня к двумерному виду

Поиск экстремума методом Ψ -преобразования

Заменяем исходную целевую функцию F с ограничениями-неравенствами на функцию Φ без ограничений (включая их Φ в качестве штрафов)

- 1 Выбираем случайные x_1, \dots, x_N
- 2 Вычисляем значение функции $\Phi(x_1, \dots, x_N)$
- 3 Повторяем шаги 1, 2 всего s раз
- 4 Определяем $\sup \Phi_{\{s\}}$ и среднее $\langle \Phi \rangle_{\{s\}}$
- 5 Делим интервал $[\langle \Phi \rangle, \sup \Phi]$ на k частей с шагом $\Delta\zeta = \frac{\sup \Phi - \langle \Phi \rangle}{k}$ (практически $s \approx 10, k < s$)
- 6 Для $\nu = 1, 2, \dots, k$ вычисляются значения уровней $\zeta_\nu^k = \langle \Phi \rangle + (\nu - 1)\Delta\zeta$.
- 7 Генерируем случайный вектор $\mathbf{x} = (x_1, x_2, \dots, x_n)$.
- 8 Вычисляется значение функции $\Phi(\mathbf{x})$ в этой точке.
- 9 Проверяется условие $\Phi(\mathbf{x}) \geq \zeta_\nu$. Если НЕТ, то переход к п.13.

Поиск экстремума методом Ψ -преобразования (2)

- 10** Вычисляются и накапливаются $y_\nu = [\Phi(x) - \zeta_\nu]^l$ (l см. п. 13).
- 11** Вычисляются $x_j [\Phi(x) - \zeta_\nu]^l$ и накапливаются в $\sigma_{j\nu}$ ($j = 1, 2, \dots, n$)
- 12** пп. 9, 10, 11 выполняются для всех $\nu = (1, 2, \dots, k)$.
- 13** пп. 7–12 повторяются s раз. В результате

$$\text{получаются } \langle y_\nu \rangle = \sum_{i=1}^{\xi_\nu} [\Phi(x) - \zeta_\nu]_i^l, \quad (\nu = 1, 2, \dots, k),$$

$$\langle \sigma_{j\nu} \rangle = \sum_{i=1}^{\xi_\nu} \left\{ x_j [\Phi(x) - \zeta_\nu]_i^l \right\}, \quad (\nu = 1, 2, \dots, k),$$

($j = 1, 2, \dots, n$).

Здесь ξ_ν — число точек случайной выборки, где $\Phi(x) \geq \zeta_\nu$,

$l = 1, 2, \dots$ управляет желаемой точностью решения задачи. Обычно $l = 1$, при $l = 3$ точность выше.

Поиск экстремума методом Ψ -преобразования (3)

- 14 Вычисляем $\Psi_\nu = \Psi(\zeta_\nu) = \frac{\langle y_\nu \rangle}{s}$, $x_{j\nu} = x_j(\zeta_\nu) = \frac{\langle \sigma_{j\nu} \rangle}{\langle y_\nu \rangle}$, где $\nu = 1, 2, \dots, k$, $j = 1, 2, \dots, n$.
- 15 м.н.к.-аппроксимация значений Ψ_1, \dots, Ψ_k параболой
- 16 вычисляется ζ^* — наименьший положительный корень параболы из п.15
- 17 м.н.к.-аппроксимация значений $x_{j1}, \dots, x_{j\nu}, \dots, x_{jk}$ параболой ($j = 1, \dots, n$). Определяются коэффициенты $\beta_0^j, \beta_1^j, \beta_2^j$
- 18 Положение глобального экстремума вычисляется как: $x_j^* = \beta_0^j + \beta_1^j \zeta^* + \beta_2^j \zeta^{*2}$
- 19 Вычисляется значение $\Phi(\mathbf{x})$ при подстановке $x_1^*, x_2^*, \dots, x_n^*$.
- 20 Это значение сравнивается с глобальным экстремумом ζ^* . Совпадение в пределах ошибки (априорной точности вычисления функции) означает, что задача решена правильно.

Поиск экстремума методом Ψ -преобразования (4)

В противном случае начинается **шаманство**, смысл которого в исследовании локального поведения целевой функции в окрестности экстремума и увеличение статистической выборки. Конкретно:

21

Даем ζ^* приращение $\Delta\zeta$ и строим две последовательности векторов.

$x^k = x(\zeta_k)$, где $\zeta_k = \zeta^* + k\Delta\zeta$ ($k = 1, 2, \dots$) и

$x^i = x(\zeta_i)$, где $\zeta_i = \zeta^* - i\Delta\zeta$ ($i = 1, 2, \dots$)

Значения k и i увеличиваются, пока выполняются условия:

$\Phi(x^k) \geq \Phi(x^{k-1})$ и $\Phi(x^i) \geq \Phi(x^{i-1})$ соответственно.

Пусть m_k и m_i — максимальные значения, при которых эти условия выполняются. Тогда максимум функции:

$$\max \{ \Phi(x^{m_i}), \Phi(x^{m_k}) \} = \Phi(x^m)$$

Поиск экстремума методом Ψ -преобразования (5)

22

В случае, если в результате выполнения п.21 удовлетворительного решения нет, необходимо в окрестности точки $x_1^*, x_2^*, \dots, x_n^*$ из п.18 провести дополнительные статистические испытания. Т.е. повторить пп. 1–20 в более узкой области.

Дополнительная литература

- 1 Гилл Ф., Мюррей М., Райт М. *Практическая оптимизация*, М. “Мир”, 1985
- 2 В.К. Чичинадзе *Решение невыпуклых нелинейных задач оптимизации*, М. “Наука”, 1983, 256 с.

Задания к семинару

- 1 Построить случайный генератор распределения Максвелла
- 2 Построить Вашим стандартным генератором (или URAND из Форсайта) равномерно распределенных случайных чисел набор 13-мерных векторов и посмотреть распределение их проекций на плоскость (7,8)-й координат.

спасибо за внимание