

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 29.07.2022 18:06:32
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6745d99a6ae00adc8e27b55cbe1e2dbd7c78

1.1. Обобщенный алгоритм

Задача машинного обучения заключается в восстановлении зависимостей по конечным выборкам данных (прецедентов). Пусть $(\mathbf{X}, t) = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$ – обучающая выборка, где $\mathbf{x}_i = (x_{i,1} \ x_{i,2} \ \dots \ x_{i,k}) \in \mathbb{R}^k$ – признаковое описание объекта, а $t \in T$ – значение скрытой компоненты (классовая принадлежность, значение прогноза, номер кластера и т.д.). при статистическом подходе к решению задачи машинного обучения предполагается, что обучающая выборка является выборкой из некоторой генеральной совокупности с плотностью $p(\mathbf{x}|t)$. Требуется восстановить $p(t|\mathbf{x})$, то есть знание о скрытой компоненте объекта по измеренным признакам.

1.2. Примеры частных алгоритмов

В классических постановках задач обучающая выборка $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ состоит из векторов $\mathbf{x}_i = (x_{i,1} \ x_{i,2} \ \dots \ x_{i,k})$ вещественнозначных признаков.

Задача классификации (распознавания образов) возникла из задачи машинного зрения. Образом вектора \mathbf{x} является переменная t , принимающая значение из конечного множества $T = \{1; 2; \dots; l\}$. Требуется построить алгоритм (классификатор), который по вектору признаков \mathbf{x} вернул бы метку \hat{l} класса или вектор оценок принадлежности (апостериорных вероятностей) $\{p(s|\mathbf{x})\}_{s=1}^l$ к каждому из классов.

Задача восстановления регрессии возникла при исследовании влияния одной группы непрерывных случайных величин на другую группу

непрерывных случайных величин. Образом вектора \mathbf{x} является непрерывная вещественнозначная переменная t . Требуется построить алгоритм (регрессор), который по вектору признаков \mathbf{x} вернул бы точечную оценку значения регрессии \hat{t} , доверительный интервал (t_-, t_+) или апостериорное распределение $p(t|\mathbf{x})$ на множестве значений регрессионной переменной.

Задача кластеризации (обучения без учителя) возникла из задачи группировки сложных объектов в единую структуру (кластер) с последующим выделением общих черт. Требуется построить алгоритм (кластеризатор), который разбил бы выборку на непересекающиеся группы (кластеры) $X = \bigcup_{k=1}^l C_k$, $C_k \subset \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $C_i \cap C_j = \emptyset$ ($j \neq i$, $1 \leq i, j \leq l$). В каждый кластер должны попасть объекты в некотором смысле похожие друг на друга.

Задача идентификации возникла из необходимости отделить объекты \mathbf{x} выборки, обладающие определённым свойством $\chi_A(\mathbf{x})=1$, от остальных. Особенностью задачи является то, что все объекты принадлежат одному классу, причём не существует возможности сделать репрезентативную выборку «остальные». Требуется построить алгоритм (идентификатор), который по вектору признаков \mathbf{x} определил бы наличие свойства A у объекта \mathbf{x} , либо вернул оценку степени его выраженности $p(\chi_A(\mathbf{x})=1|\mathbf{x})$.

Задача прогнозирования возникла при исследовании временных рядов и попытке предсказания их значений через какой-то промежуток времени. Вектор $\mathbf{x}_i = (x_{i,1} \ x_{i,2} \ \dots \ x_{i,k})$ представляет собой набор измерений, сделанных в момент времени t_i . Требуется построить алгоритм (предиктор), который вернул бы точечную оценку $\{\mathbf{x}_i\}_{i=n+1}^{n+m}$, доверительный интервал $\{\{\mathbf{x}_i^{(-)}, \mathbf{x}_i^{(+)}\}\}_{i=n+1}^{n+m}$ или апостериорное распределение вероятностей $p(\mathbf{x}_{n+1} \dots \mathbf{x}_{n+m} | \mathbf{x}_1 \dots \mathbf{x}_n)$ прогноза на заданную глубину m . В отличие от задачи

восстановления регрессии, прогноз осуществляется по времени, а не по признакам.

Задача извлечения знаний возникла при исследовании взаимозависимостей между косвенными показателями одного и того же явления. Требуется построить алгоритм, генерирующий набор объективных закономерностей между признаками, имеющих место в генеральной совокупности. Закономерности обычно имеют форму предикатов «если ..., то ...» и могут выражаться как в числовых, так и в текстовых терминах.

Метод Ньютона

Стандартным методом нахождения решений нелинейных систем вида

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0, \\ f_2(x_1, x_2, \dots, x_n) = 0, \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

является метод Ньютона, в котором последующее приближение

$\mathbf{x}^{(s+1)} = (x_1^{(s+1)}, x_2^{(s+1)}, \dots, x_n^{(s+1)})^T$ выражается через предыдущее приближение

$\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})^T$ итерационной схемой

$$\begin{pmatrix} x_1^{(s+1)} \\ x_2^{(s+1)} \\ \vdots \\ x_n^{(s+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(s)} \\ x_2^{(s)} \\ \vdots \\ x_n^{(s)} \end{pmatrix} - \begin{pmatrix} (\partial f_1 / \partial x_1)^{(s)} & (\partial f_1 / \partial x_2)^{(s)} & \dots & (\partial f_1 / \partial x_n)^{(s)} \\ (\partial f_2 / \partial x_1)^{(s)} & (\partial f_2 / \partial x_2)^{(s)} & \dots & (\partial f_2 / \partial x_n)^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ (\partial f_n / \partial x_1)^{(s)} & (\partial f_n / \partial x_2)^{(s)} & \dots & (\partial f_n / \partial x_n)^{(s)} \end{pmatrix}^{-1} \cdot \begin{pmatrix} f_1^{(s)} \\ f_2^{(s)} \\ \vdots \\ f_n^{(s)} \end{pmatrix},$$

где $f_i^{(s)} = f_i(x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})$, $\left(\frac{\partial f_i}{\partial x_j}\right)^{(s)} = \frac{\partial f_i(x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})}{\partial x_j}$, $1 \leq i, j \leq n$.

Метод множителей Лагранжа (второго порядка)

Задача минимизации $f(x_1, x_2, \dots, x_n) \rightarrow \min$ с ограничениями в форме равенств $h_i(x_1, x_2, \dots, x_n) = 0, 1 \leq i \leq m$ и с ограничениями в форме неравенств $g_i(x_1, x_2, \dots, x_n) \leq 0, 1 \leq i \leq l$ введением дополнительных переменных $z_i, 1 \leq i \leq l$ сводится к задаче минимизации с ограничениями только в форме равенств

$$f(x_1, x_2, \dots, x_n) \rightarrow \min;$$

$$h_i(x_1, x_2, \dots, x_n) = 0, 1 \leq i \leq m; g_i(x_1, x_2, \dots, x_n) + z_i^2 = 0, 1 \leq i \leq l.$$

Модифицированная функция Лагранжа для этой задачи имеет вид

$$L = f + \sum_{i=1}^m (\lambda_i h_i + \frac{1}{2} p h_i^2) + \sum_{i=1}^l (\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2) \rightarrow \min, \quad (A)$$

где $\lambda_i, 1 \leq i \leq m; \mu_i, 1 \leq i \leq l$ – множители Лагранжа; $p > 0$ – параметр квадратичного штрафа.

Форма $\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2$ допускает непосредственную минимизацию по z_i^2 . Перепишем ее в виде

$$\begin{aligned} \mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2 &= \frac{1}{2} p \left((g_i + z_i^2)^2 + 2(g_i + z_i^2) \frac{\mu_i}{p} + \left(\frac{\mu_i}{p} \right)^2 \right) - \frac{1}{2} p \left(\frac{\mu_i}{p} \right)^2 = \\ &= \frac{1}{2} p \left(g_i + z_i^2 + \frac{\mu_i}{p} \right)^2 - \frac{1}{2} p \left(\frac{\mu_i}{p} \right)^2. \end{aligned}$$

При любом z_i^2 квадрат $\left(g_i + z_i^2 + \frac{\mu_i}{p} \right)^2$ неотрицателен:

$$\left(g_i + z_i^2 + \frac{\mu_i}{p} \right)^2 \geq 0.$$

При $g_i + \frac{\mu_i}{p} \leq 0$ за счет выбора $z_i^2 = -\left(g_i + \frac{\mu_i}{p} \right) = \left| g_i + \frac{\mu_i}{p} \right|$ удается

добиться равенства квадрата $\left(g_i + z_i^2 + \frac{\mu_i}{p} \right)^2$ нулю, и в этом случае

$$\min_{z_i^2} \left(\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2 \right) = -\frac{\mu_i^2}{2p}.$$

Поскольку $z_i^2 \geq 0$, то $g_i + z_i^2 + \frac{\mu_i}{p} \geq g_i + \frac{\mu_i}{p}$, и при $g_i + \frac{\mu_i}{p} > 0$ имеет место

неравенство $\left(g_i + z_i^2 + \frac{\mu_i}{p} \right)^2 \geq \left(g_i + \frac{\mu_i}{p} \right)^2 > 0$, то самое большее, чего удастся

добиться за счет выбора $z_i^2 = 0$, – это равенства

$$\min_{z_i^2} \left(\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2 \right) = \frac{1}{2} p \left(g_i + \frac{\mu_i}{p} \right)^2 - \frac{\mu_i^2}{2p}.$$

В общем случае (при любом знаке $g_i + \mu_i/p$)

$$\min_{z_i^2} \left(\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2 \right) = \begin{cases} -\frac{\mu_i^2}{2p}, & g_i + \frac{\mu_i}{p} \leq 0, \\ \frac{1}{2} p \left(g_i + \frac{\mu_i}{p} \right)^2 - \frac{\mu_i^2}{2p}, & g_i + \frac{\mu_i}{p} > 0 \end{cases}$$

или

$$\min_{z_i^2} \left(\mu_i (g_i + z_i^2) + \frac{1}{2} p (g_i + z_i^2)^2 \right) = \frac{1}{2} p \left(\max \left(0; g_i + \frac{\mu_i}{p} \right) \right)^2 - \frac{\mu_i^2}{2p}.$$

Модифицированная функция Лагранжа преобразуется к виду

$$L = f + \sum_{i=1}^m \left(\lambda_i h_i + \frac{1}{2} p h_i^2 \right) + \sum_{i=1}^l \left(\frac{1}{2} p \left(\max \left(0; g_i + \frac{\mu_i}{p} \right) \right)^2 - \frac{\mu_i^2}{2p} \right) \rightarrow \min. \quad (B)$$

Это преобразование означает следующее. В *точке* условного минимума функции f будут иметь место равенства $g_i + z_i^2 = 0$, $1 \leq i \leq l$. Но в *процессе* минимизации значения $g_i + z_i^2 = 0$ могут отличаться от нуля как в положительную, так и в отрицательную стороны. Преобразованная функция Лагранжа (B), таким образом, является оценкой снизу для модифицированной функции Лагранжа (A), причем достижимой. А минимизация достижимой оценки снизу означает и минимизацию самой функции. Только (B) в отличие от (A) не содержит z_i^2 .

Необходимые условия минимума для преобразованной функции

Лагранжа

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^m (\lambda_j + p h_j) \frac{\partial h_j}{\partial x_i} + \sum_{j=1}^l p \max\left(0; g_j + \frac{\mu_j}{p}\right) \frac{\partial \max(0; g_j + \mu_j/p)}{\partial x_i} = 0, \quad 1 \leq i \leq n, \\ \frac{\partial L}{\partial \lambda_i} = h_i = 0, \quad 1 \leq i \leq m, \\ \frac{\partial L}{\partial \mu_i} = c \max\left(0; g_i + \frac{\mu_i}{p}\right) \frac{\partial \max(0; g_i + \mu_i/p)}{\partial \mu_i} - \frac{\mu_i}{p} = 0, \quad 1 \leq i \leq l. \end{array} \right.$$

С учетом

$$\max\left(0; g_j + \frac{\mu_j}{p}\right) \frac{\partial \max(0; g_j + \mu_j/p)}{\partial x_i} = \begin{cases} 0, & g_j + \frac{\mu_j}{p} \leq 0, \\ \left(g_j + \frac{\mu_j}{p}\right) \frac{\partial g_j}{\partial x_i}, & g_j + \frac{\mu_j}{p} > 0 \end{cases}$$

или то же самое

$$\max\left(0; g_j + \frac{\mu_j}{p}\right) \frac{\partial \max(0; g_j + \mu_j/p)}{\partial x_i} = \max\left(0; g_j + \frac{\mu_j}{p}\right) \frac{\partial g_j}{\partial x_i}$$

и

$$\max\left(0; g_i + \frac{\mu_i}{p}\right) \frac{\partial \max(0; g_i + \mu_i/p)}{\partial \mu_i} = \begin{cases} 0, & g_i + \frac{\mu_i}{p} \leq 0, \\ \left(g_i + \frac{\mu_i}{p}\right) \frac{1}{p}, & g_i + \frac{\mu_i}{p} > 0 \end{cases}$$

или то же самое

$$\max\left(0; g_i + \frac{\mu_i}{p}\right) \frac{\partial \max(0; g_i + \mu_i/p)}{\partial \mu_i} = \frac{1}{p} \max\left(0; g_i + \frac{\mu_i}{p}\right),$$

необходимые условия минимума принимают вид

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial x_i} + \sum_{j=1}^m (\lambda_j + p h_j) \frac{\partial h_j}{\partial x_i} + \sum_{j=1}^l \max(0; \mu_j + p g_j) \frac{\partial g_j}{\partial x_i} = 0, \quad 1 \leq i \leq n, \\ h_i = 0, \quad 1 \leq i \leq m, \\ \max(0; \mu_i + p g_i) - \mu_i = 0, \quad 1 \leq i \leq l. \end{array} \right. \quad (C)$$

Линеаризуем уравнения системы (C)

$$\begin{aligned}
& \frac{\partial f}{\partial x_i} + \sum_{j=1}^m (\lambda_j + p h_j) \frac{\partial h_j}{\partial x_i} + \sum_{j=1}^l \max(0; \mu_j + p g_j) \frac{\partial g_j}{\partial x_i} + \\
& + \sum_{k=1}^n \left(\frac{\partial^2 f}{\partial x_k \partial x_i} + \sum_{j=1}^m \left(p \frac{\partial h_j}{\partial x_k} \frac{\partial h_j}{\partial x_i} + (\lambda_j + p h_j) \frac{\partial^2 h_j}{\partial x_k \partial x_i} \right) + \right. \\
& \left. + \sum_{j=1}^l \left(\frac{\partial \max(0; \mu_j + p g_j)}{\partial x_k} \frac{\partial g_j}{\partial x_i} + \max(0; \mu_j + p g_j) \frac{\partial^2 g_j}{\partial x_k \partial x_i} \right) \right) (\hat{x}_k - x_k) + \\
& + \sum_{j=1}^m \frac{\partial h_j}{\partial x_i} (\hat{\lambda}_j - \lambda_j) + \sum_{j=1}^l \frac{\partial \max(0; \mu_j + p g_j)}{\partial \mu_j} \frac{\partial g_j}{\partial x_i} (\hat{\mu}_j - \mu_j) = 0, \quad 1 \leq i \leq n, \quad (D)
\end{aligned}$$

$$h_i + \sum_{k=1}^n \frac{\partial h_i}{\partial x_k} (\hat{x}_k - x_k) = 0, \quad 1 \leq i \leq m, \quad (E)$$

$$\begin{aligned}
& \max(0; \mu_i + p g_i) - \mu_i + \sum_{k=1}^n \frac{\partial \max(0; \mu_i + p g_i)}{\partial x_k} (\hat{x}_k - x_k) + \\
& + \left(\frac{\partial \max(0; \mu_i + p g_i)}{\partial \mu_i} - 1 \right) (\hat{\mu}_i - \mu_i) = 0, \quad 1 \leq i \leq m. \quad (F)
\end{aligned}$$

Перегруппируем слагаемые в уравнении (D) следующим образом

$$\begin{aligned}
& \frac{\partial f}{\partial x_i} + \sum_{j=1}^m \left(\hat{\lambda}_j - \lambda_j + \lambda_j + p \left(h_j + \sum_{k=1}^n \frac{\partial h_j}{\partial x_k} (\hat{x}_k - x_k) \right) \right) \frac{\partial h_j}{\partial x_i} + \\
& + \sum_{j=1}^l \left(\max(0; \mu_j + p g_j) + \sum_{k=1}^n \frac{\partial \max(0; \mu_j + p g_j)}{\partial x_k} (\hat{x}_k - x_k) + \right. \\
& \quad \left. + \frac{\partial \max(0; \mu_j + p g_j)}{\partial \mu_j} (\hat{\mu}_j - \mu_j) \right) \frac{\partial g_j}{\partial x_i} + \\
& + \sum_{k=1}^n \left(\frac{\partial^2 f}{\partial x_k \partial x_i} + \sum_{j=1}^m (\lambda_j + p h_j) \frac{\partial^2 h_j}{\partial x_k \partial x_i} + \sum_{j=1}^l \max(0; \mu_j + p g_j) \frac{\partial^2 g_j}{\partial x_k \partial x_i} \right) (\hat{x}_k - x_k) = 0, \\
& \quad 1 \leq i \leq n.
\end{aligned}$$

С учетом (E) и (F) оно упрощается

$$\frac{\partial f}{\partial x_i} + \sum_{k=1}^n \left(\frac{\partial^2 f}{\partial x_k \partial x_i} + \sum_{j=1}^m (\lambda_j + p h_j) \frac{\partial^2 h_j}{\partial x_k \partial x_i} + \sum_{j=1}^l \max(0; \mu_j + p g_j) \frac{\partial^2 g_j}{\partial x_k \partial x_i} \right) (\hat{x}_k - x_k) +$$

$$+ \sum_{j=1}^m \frac{\partial h_j}{\partial x_i} \hat{\lambda}_j + \sum_{j=1}^l \frac{\partial g_j}{\partial x_i} \hat{\mu}_j = 0, \quad 1 \leq i \leq n.$$

А уравнение (F) преобразуем в альтернативу

$$\begin{aligned} & \max(0; \mu_i + p g_i) - \mu_i + \sum_{k=1}^n \frac{\partial \max(0; \mu_i + p g_i)}{\partial x_k} (\hat{x}_k - x_k) + \\ & + \left(\frac{\partial \max(0; \mu_i + p g_i)}{\partial \mu_i} - 1 \right) (\hat{\mu}_i - \mu_i) = \begin{cases} \hat{\mu}_i = 0, & g_i + \mu_i / p \leq 0, \\ g_i + \sum_{k=1}^n \frac{\partial g_i}{\partial x_k} (\hat{x}_k - x_k) = 0, & g_i + \mu_i / p > 0, \end{cases} \\ & 1 \leq i \leq m. \end{aligned}$$

Таким образом, решение задачи минимизации $f(x_1, x_2, \dots, x_n) \rightarrow \min$ с ограничениями в форме равенств $h_i(x_1, x_2, \dots, x_n) = 0$, $1 \leq i \leq m$ и с ограничениями в форме неравенств $g_i(x_1, x_2, \dots, x_n) = 0$, $1 \leq i \leq l$ может быть найдено с помощью следующего итерационного процесса (метода множителей Лагранжа второго порядка)

$$\left\{ \begin{aligned} & \frac{\partial f}{\partial x_i} + \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{k=1}^m (\lambda_k + p h_k) \frac{\partial^2 h_k}{\partial x_i \partial x_j} + \sum_{k=1}^l \max(0; \mu_k + p g_k) \frac{\partial^2 g_k}{\partial x_i \partial x_j} \right) (\hat{x}_j - x_j) + \\ & + \sum_{j=1}^m \frac{\partial h_j}{\partial x_i} \hat{\lambda}_j + \sum_{j=1}^l \frac{\partial g_j}{\partial x_i} \hat{\mu}_j = 0, \quad 1 \leq i \leq n, \\ & h_i + \sum_{j=1}^n \frac{\partial h_i}{\partial x_j} (\hat{x}_j - x_j) = 0, \quad 1 \leq i \leq m, \\ & \left[\begin{aligned} & \hat{\mu}_i = 0, & \mu_i + p g_i \leq 0, \\ & g_i + \sum_{j=1}^n \frac{\partial g_i}{\partial x_j} (\hat{x}_j - x_j) = 0, & \mu_i + p g_i > 0, \end{aligned} \right. \quad 1 \leq i \leq l, \end{aligned} \right.$$

где λ_i , $1 \leq i \leq m$, μ_i , $1 \leq i \leq l$ – множители Лагранжа; $p > 0$ – параметр квадратичного штрафа.

2 Численные методы решения систем линейных алгебраических уравнений

2.1 Прямые, стационарные и нестационарные итерационные методы

Задача состоит в нахождении решения \mathbf{x} системы $\mathbf{Ax} = \mathbf{b}$ линейных алгебраических уравнений (СЛАУ). Методы решения разделяют на два направления – прямые и итерационные. В *прямых* методах для нахождения решения задачи заданной размерности число арифметических действий заранее предопределено. Это является достоинством прямых методов. В *итерационных* методах итерации продолжаются, пока не будет достигнута требуемая точность решения. Это достоинство итерационных методов – противовес недостатку прямых, в которых на получаемую точность повлиять нельзя.

Примером прямых методов является метод Гаусса. В прямых методах число арифметических действий, требуемое для решения СЛАУ с N неизвестными составляет $O(N^3)$. Примером итерационных методов, для которых одна итерация требует $O(N^2)$ действий, можно назвать метод Якоби. Но характерное число самих итераций может составлять M , которое в большей степени определяется спектральными свойствами матрицы СЛАУ, чем размерностью задачи N . В случае слабой зависимости M от N (именно такова действительность) существует пороговое значение N_0 , до которого преимущества за прямыми методами (и по вычислительной эффективности, и по обеспечиваемой точности), а свыше – за итерационными. Практически считается, что для типичных случаев до $N \leq 100$ итерационные методы не конкурируют с прямыми, а свыше $N \geq 200$ – прямые не конкурируют с итерационными. Интервал $100 < N < 200$ является переходным, и в нём определяющим являются характеристики матрицы СЛАУ [34], [35].

В анализе методов используются два вектора-индикатора – вектор погрешности и вектор невязки. Пусть $\mathbf{x}^{(s)}$ – приближение к точному решению \mathbf{x} на шаге s . Разность $\boldsymbol{\eta}^{(s)} = \mathbf{x}^{(s)} - \mathbf{x}$ между приближённым $\mathbf{x}^{(s)}$ и

точным \mathbf{x} решением – это погрешность. Невязка $\xi^{(s)}$ на приближении $\mathbf{x}^{(s)}$ – это вектор компенсации, добавляемый к левой части СЛАУ для точного выполнения уравнений $\mathbf{Ax}^{(s)} + \xi^{(s)} = \mathbf{b}$ при подстановке $\mathbf{x}^{(s)}$.

Чтобы указать вектор погрешности $\eta^{(s)}$, необходимо располагать точным решением \mathbf{x} , которое только отыскивается. Для указания вектора невязки $\xi^{(s)}$ достаточно располагать только самим приближением $\mathbf{x}^{(s)}$ к решению.

Вычтем из векторного уравнения $\mathbf{Ax}^{(s)} + \xi^{(s)} = \mathbf{b}$ векторное уравнение $\mathbf{Ax} = \mathbf{b}$ самой СЛАУ, то есть

$$\mathbf{A}(\mathbf{x}^{(s)} - \mathbf{x}) + \xi^{(s)} = \mathbf{b} - \mathbf{b} = \mathbf{A}\eta^{(s)} + \xi^{(s)} = \mathbf{0}.$$

Таким образом, получаем соотношение между векторами погрешности $\eta^{(s)}$ и невязки $\xi^{(s)}$

$$\xi^{(s)} = -\mathbf{A}\eta^{(s)},$$

показывающее, что отыскание погрешности по сравнению с невязкой равноценно самому решению СЛАУ (решив СЛАУ, мы располагали бы и точным решением).

Итерационные методы, как правило, имеют некоторый итерационный параметр (и даже не один) τ_s . Если от итерации к итерации параметр не изменяется, то такие методы называют *стационарными*, в противном случае – *нестационарными*.

2.2 Стационарные итерационные методы решения СЛАУ

Примеры стационарных итерационных методов решения СЛАУ:

1. Метод Зейделя;

2. Метод релаксации;
3. Метод простых итераций (простой градиентный спуск);
4. Чебышёвский итерационный метод.

Чебышёвский итерационный метод имеет черты и стационарных, и нестационарных. С одной стороны, итерационный параметр от итерации к итерации изменяется, но метод принципиально состоит из целого числа макроитераций, которые повторяются одна за другой в неизменном виде.

2.6 Соответствие между задачей минимизации квадратичного функционала и задачей решения СЛАУ с симметричной матрицей

Квадратичным функционалом называется следующая скалярная функция Φ векторного аргумента \mathbf{x}

$$\Phi = \frac{1}{2}(\mathbf{Ax}, \mathbf{x}) - (\mathbf{b}, \mathbf{x}) + c = \frac{1}{2} \left(\sum_{j=1}^n \sum_{k=1}^n a_{jk} x_j x_k \right) - \left(\sum_{i=1}^n b_i x_i \right) + c,$$

где $\mathbf{A} = (a_{i,j})_{n \times n}$ – матрица квадратичной формы,

$\mathbf{b} = (b_i)_n$ – вектор линейной формы,

c – число (скаляр).

Необходимыми условиями минимума

$$\Phi = \frac{1}{2}(\mathbf{Ax}, \mathbf{x}) - (\mathbf{b}, \mathbf{x}) + c \rightarrow \min \quad (22)$$

будет система равенств нулю частных производных Φ по всем неизвестным x_i ($1 \leq i \leq n$)

$$\begin{cases} \partial\Phi/\partial x_1 = a_{11}x_1 + \frac{1}{2}(a_{12} + a_{21})x_2 + \frac{1}{2}(a_{13} + a_{31})x_3 + \dots + \frac{1}{2}(a_{1n} + a_{n1})x_n - b_1 = 0, \\ \partial\Phi/\partial x_2 = \frac{1}{2}(a_{21} + a_{12})x_1 + a_{22}x_2 + \frac{1}{2}(a_{23} + a_{32})x_3 + \dots + \frac{1}{2}(a_{2n} + a_{n2})x_n - b_2 = 0, \\ \partial\Phi/\partial x_3 = \frac{1}{2}(a_{31} + a_{13})x_1 + \frac{1}{2}(a_{32} + a_{23})x_2 + a_{33}x_3 + \dots + \frac{1}{2}(a_{3n} + a_{n3})x_n - b_3 = 0, \\ \vdots \\ \partial\Phi/\partial x_n = \frac{1}{2}(a_{n1} + a_{1n})x_1 + \frac{1}{2}(a_{n2} + a_{2n})x_2 + \frac{1}{2}(a_{n3} + a_{3n})x_3 + \dots + a_{nn}x_n - b_n = 0, \end{cases}$$

которую можно представить в матричной форме

$$\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x} = \mathbf{b}. \quad (23)$$

Таким образом, решение задачи безусловной минимизации (22) сводится к решению СЛАУ (23) с симметричной матрицей. Верно и обратное, что решение СЛАУ с симметричной матрицей вида (23) является решением задачи безусловной минимизации функционала (22).

2.7 Метод простых итераций (простой градиентный спуск)

Система (23) выполняется точно при достижении минимума функционала Φ . Но на приближении $\mathbf{x}^{(s)}$ градиент $\nabla\Phi^{(s)}$ ещё не является нулевым вектором. Заметим, что $\nabla\Phi^{(s)}$ только знаком отличается от невязки

$$\nabla\Phi^{(s)} = -\xi^{(s)}.$$

Основу следующих итерационных методов (стационарных и нестационарных) составляет схема

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \xi^{(s)} = \mathbf{x}^{(s)} - \tau_s \nabla\Phi^{(s)},$$

которую можно рассматривать как схему оптимизационного метода градиентного спуска (поскольку речь идёт о минимизации функционала). Если τ_s не зависит от итерации ($\tau_s = \tau = \text{const}$), то итерационный метод называют *стационарным*. В противном случае – *нестационарным* итерационным методом.

Установим рекуррентное соотношение для невязок двух последовательных итераций

$$\xi^{(s+1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(s+1)} = \mathbf{b} - \mathbf{A}(\mathbf{x}^{(s)} + \tau_s \xi^{(s)}) = (\mathbf{b} - \mathbf{A}\mathbf{x}^{(s)}) - \tau_s \mathbf{A}\xi^{(s)} = \xi^{(s)} - \tau_s \mathbf{A}\xi^{(s)},$$

то есть

$$\xi^{(s+1)} = \xi^{(s)} - \tau_s \mathbf{A}\xi^{(s)} = (\mathbf{E} - \tau_s \mathbf{A})\xi^{(s)}.$$

Несложно доказывается, что все собственные числа симметричной (и Эрмитовой) матрицы являются вещественными числами. Для симметричной положительно определённой матрицы они будут вещественными и положительными. Как следствие, все собственные векторы будут состоять только из вещественных компонент. Собственные векторы, соответствующие различным собственным числам, будут ортогональны. То есть, если все собственные числа различны, то все собственные векторы попарно ортогональны и могут быть использованы как ортогональный базис. Если среди собственных чисел встречаются кратные, то и в этом случае может быть сформирован ортогональный базис из собственных векторов.

Если разложить невязки $\xi^{(s+1)}$ и $\xi^{(s)}$ по базису $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ из собственных векторов, то

$$\begin{aligned} \xi^{(s+1)} &= c_1^{(s+1)}\mathbf{e}_1 + c_2^{(s+1)}\mathbf{e}_2 + \dots + c_n^{(s+1)}\mathbf{e}_n = (\mathbf{E} - \tau_s \mathbf{A})\xi^{(s)} = \\ &= (\mathbf{E} - \tau_s \mathbf{A})(c_1^{(s)}\mathbf{e}_1 + c_2^{(s)}\mathbf{e}_2 + \dots + c_n^{(s)}\mathbf{e}_n) = \\ &= (1 - \tau_s \lambda_1)c_1^{(s)}\mathbf{e}_1 + (1 - \tau_s \lambda_2)c_2^{(s)}\mathbf{e}_2 + \dots + (1 - \tau_s \lambda_n)c_n^{(s)}\mathbf{e}_n. \end{aligned}$$

Поскольку разложение по базису единственно, то

$$c_1^{(s+1)} = (1 - \tau_s \lambda_1) c_1^{(s)}, c_2^{(s+1)} = (1 - \tau_s \lambda_2) c_2^{(s)}, \dots, c_n^{(s+1)} = (1 - \tau_s \lambda_n) c_n^{(s)}.$$

Продолжая процесс, получим

$$c_k^{(s+1)} = (1 - \tau_s \lambda_k)^{s+1} c_k^{(0)}, \quad 1 \leq k \leq n.$$

Видно, что если $|1 - \tau_s \lambda_k| > 1$, то компоненты невязки будут возрастать неограниченно, вместо того, чтобы стремиться к нулю. Поэтому для сходимости итерационного процесса $\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \xi^{(s)}$ необходимо выбрать τ_s таким, чтобы для всех собственных чисел выполнялось $|1 - \tau_s \lambda_k| \leq 1$. Это обеспечится при

$$0 \leq \tau \leq 2/\lambda_{\max}.$$

У семейства графиков $|1 - \tau_s \lambda_k| \leq 1$ ($1 \leq k \leq n$) существует ещё одна важная точка. Часто её называют оптимальным значением τ_{opt} . Однако, в зависимости от распределения коэффициентов невязки $c_k^{(0)}$, $1 \leq k \leq n$, в конкретных случаях можно указать значение τ , обеспечивающее гораздо более эффективную сходимость, но отличающееся от τ_{opt} . Правильно называть это значение минимаксным, потому что оно соответствует минимизации последствий наихудшего случая (который не обязательно реализуется). Это значение соответствует точке пересечения лучей $y = 1 - \tau \lambda_{\min}$ и $y = \tau \lambda_{\max} - 1$ и равно

$$\tau_m = 2/(\lambda_{\min} + \lambda_{\max}).$$

Для плохо обусловленной симметричной положительно определённой матрицы ($0 < \lambda_{\min} \ll \lambda_{\max}$) это значение расположено близко от критической границы $\tau_b = 2/\lambda_{\max}$.

2.8 Аналогия методов решения СЛАУ и методов решения ОДУ

Итерационную схему методов решения СЛАУ

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \xi^{(s)} = \mathbf{x}^{(s)} + \tau_s (\mathbf{b} - \mathbf{Ax}^{(s)})$$

можно рассматривать как схему Эйлера первого порядка (чисто явную схему)

$$(\mathbf{x}^{(s+1)} - \mathbf{x}^{(s)})/\tau_s = \mathbf{b} - \mathbf{Ax}^{(s)}$$

для решения системы обыкновенных дифференциальных уравнений (ОДУ)

$$d\mathbf{x}/dt = \mathbf{b} - \mathbf{Ax}.$$

Если у системы ОДУ существует стационарное решение $d\mathbf{x}/dt = \mathbf{0}$, то оно будет решением СЛАУ $\mathbf{Ax} = \mathbf{b}$. Верно и обратное. Решение СЛАУ можно рассматривать как достижение системой ОДУ стационарного состояния. Методы нахождения решений стационарных задач с помощью решения соответствующих нестационарных задач называют методами установления.

Соответственно, если физической системе, описываемой системой ОДУ $d\mathbf{x}/dt = \mathbf{b} - \mathbf{Ax}$, для выхода на стационарный режим требуется время T , то и для получения решения СЛАУ при помощи итерационной схемы

$\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s (\mathbf{b} - \mathbf{Ax}^{(s)})$ потребуется число итераций M такое, что $\tau_0 + \tau_1 + \dots + \tau_{M-1} \geq T$.

Для метода простых итераций ограничение на итерационный параметр диктуется условием $0 < \tau < 2/\lambda_{\max}$. И чем больше λ_{\max} у матриц из однотипного семейства, тем больше потребуется итераций.

2.12 Нестационарные итерационные методы решения СЛАУ

Примеры нестационарных итерационных методов решения СЛАУ:

1. Метод скорейшего спуска
2. Метод минимальных невязок
3. Метод сопряжённых направлений

В основе рассматриваемых методов лежит следующий итерационный процесс

$$\mathbf{x}^{(s+1)} \leftarrow \mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}.$$

При этом в качестве направления $\mathbf{d}^{(s)}$ в методах скорейшего спуска и минимальных невязок используется вектор невязки $\boldsymbol{\xi}^{(s)} = \mathbf{b} - \mathbf{Ax}^{(s)}$ на приближении $\mathbf{x}^{(s)}$ к решению СЛАУ $\mathbf{Ax} = \mathbf{b}$, а в методе сопряжённых направлений так называемое *сопряжённое* направление. Итерационный параметр τ_s изменяется от итерации к итерации. Отсюда и прилагательное «нестационарные», характеризующее группу методов. Между собой методы различаются идеей, лежащей в основе получения оптимального значения τ_s .

2.13 Метод скорейшего спуска

Итак, задачи (22) и (23) эквивалентны:

$$\Phi = \frac{1}{2}(\mathbf{Ax}, \mathbf{x}) - (\mathbf{b}, \mathbf{x}) \rightarrow \min, \quad \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x} - \mathbf{b} = \mathbf{0}.$$

Выразим значение $\Phi^{(s+1)}$ квадратичного функционала на $(s+1)$ -м приближении $\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \xi^{(s)}$ к решению СЛАУ (23), где $\xi^{(s)} = \mathbf{b} - \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x}^{(s)}$ – вектор невязки на приближении $\mathbf{x}^{(s)}$

$$\begin{aligned} \Phi^{(s+1)} &= \frac{1}{2}(\mathbf{Ax}^{(s+1)}, \mathbf{x}^{(s+1)}) - (\mathbf{b}, \mathbf{x}^{(s+1)}) = \\ &= \frac{1}{2}(\mathbf{A}(\mathbf{x}^{(s)} + \tau_s \xi^{(s)}), \mathbf{x}^{(s)} + \tau_s \xi^{(s)}) - (\mathbf{b}, \mathbf{x}^{(s)} + \tau_s \xi^{(s)}) = \\ &= \frac{1}{2}(\mathbf{Ax}^{(s)}, \mathbf{x}^{(s)}) - (\mathbf{b}, \mathbf{x}^{(s)}) + \frac{1}{2}\tau_s(\mathbf{A}\xi^{(s)}, \mathbf{x}^{(s)}) + \\ &\quad + \frac{1}{2}\tau_s(\mathbf{Ax}^{(s)}, \xi^{(s)}) - \tau_s(\mathbf{b}, \xi^{(s)}) + \frac{1}{2}\tau_s^2(\mathbf{A}\xi^{(s)}, \xi^{(s)}) = \\ &= \Phi^{(s)} + \frac{1}{2}\tau_s(\xi^{(s)}, \mathbf{A}^T \mathbf{x}^{(s)}) + \frac{1}{2}\tau_s(\mathbf{Ax}^{(s)}, \xi^{(s)}) - \tau_s(\mathbf{b}, \xi^{(s)}) + \frac{1}{2}\tau_s^2(\mathbf{A}\xi^{(s)}, \xi^{(s)}) = \\ &= \Phi^{(s)} + \tau_s\left(\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x}^{(s)} - \mathbf{b}, \xi^{(s)}\right) + \frac{1}{2}\tau_s^2(\mathbf{A}\xi^{(s)}, \xi^{(s)}) = \\ &= \Phi^{(s)} - \tau_s(\xi^{(s)}, \xi^{(s)}) + \frac{1}{2}\tau_s^2(\mathbf{A}\xi^{(s)}, \xi^{(s)}). \end{aligned}$$

Видим, что при положительно определённой матрице \mathbf{A} значение $\Phi^{(s+1)}$ представляет собой квадратичную функцию τ_s

$$\Phi^{(s+1)} = \Phi^{(s)} - \tau_s(\xi^{(s)}, \xi^{(s)}) + \frac{1}{2}\tau_s^2(\mathbf{A}\xi^{(s)}, \xi^{(s)}) = \frac{1}{2}\alpha\tau_s^2 - \beta\tau_s + \gamma,$$

минимальное значение которой достигается в вершине соответствующей параболы

$$\frac{\partial \Phi^{(s+1)}}{\partial \tau_s} = -(\xi^{(s)}, \xi^{(s)}) + \tau_s (\mathbf{A}\xi^{(s)}, \xi^{(s)}) = \alpha \tau_s - \beta = 0,$$

то есть при оптимальном значении

$$\tau_s = \frac{(\xi^{(s)}, \xi^{(s)})}{(\mathbf{A}\xi^{(s)}, \xi^{(s)})}.$$

Заметим, что требования положительной определённости и симметрии матрицы СЛАУ существенны.

Заметим также, что невязка $\xi^{(s+1)}$ следующим образом выражается через невязку $\xi^{(s)}$

$$\begin{aligned} \xi^{(s+1)} &= \mathbf{b} - \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x}^{(s+1)} = \mathbf{b} - \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)(\mathbf{x}^{(s)} + \tau_s \xi^{(s)}) = \\ &= \mathbf{b} - \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\mathbf{x}^{(s)} - \frac{1}{2}\tau_s(\mathbf{A} + \mathbf{A}^T)\xi^{(s)} = \xi^{(s)} - \frac{1}{2}\tau_s(\mathbf{A} + \mathbf{A}^T)\xi^{(s)}, \end{aligned}$$

то есть изменение невязки определяется следующим процессом

$$\xi^{(s+1)} = \xi^{(s)} - \frac{1}{2}\tau_s(\mathbf{A} + \mathbf{A}^T)\xi^{(s)}.$$

Для симметричной $\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ матрицы скалярное произведение $(\xi^{(s+1)}, \xi^{(s)})$ равно

$$(\xi^{(s+1)}, \xi^{(s)}) = (\xi^{(s)}, \xi^{(s)}) - \tau_s \left(\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)\xi^{(s)}, \xi^{(s)} \right) = (\xi^{(s)}, \xi^{(s)}) - \tau_s (\mathbf{A}\xi^{(s)}, \xi^{(s)}) = 0$$

при оптимальном значении $\tau_s = \frac{(\xi^{(s)}, \xi^{(s)})}{(\mathbf{A}\xi^{(s)}, \xi^{(s)})}$, то есть невязки $\xi^{(s+1)}$ и $\xi^{(s)}$ ортогональны.

2.14 Метод минимальных невязок

Среди методов оптимизации можно выделить два подхода. Первый связан непосредственно с оптимизацией самой целевой функции. Примером может служить метод скорейшего спуска. Второй связан со стремлением добиться равенства нулю квадрата модуля градиента целевой функции. Можно сказать, что в своём развитии первый подход приводит к так называемым двойственным методам. Второй подход приводит к так называемым методам штрафных функций.

Как правило, квадрат модуля градиента целевой функции оказывается более чувствительным к отклонению от нулевого значения, чем сама целевая функция к отклонению от своего оптимального значения. Это приводит к тому, что методы второго типа, как правило, оказываются эффективнее методов первого типа.

Таким образом, в основе метода минимальных невязок лежит тот же самый итерационный процесс

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \boldsymbol{\xi}^{(s)},$$

где $\boldsymbol{\xi}^{(s)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(s)}$, который лежит в основе метода скорейшего спуска. Но оптимальное значение τ_s определяется из условия достижения минимума квадратом модуля градиента (невязки) на следующем шаге

$$\begin{aligned} \left| \boldsymbol{\xi}^{(s+1)} \right|^2 &= \left| \mathbf{b} - \mathbf{A}\mathbf{x}^{(s+1)} \right|^2 = \left| \mathbf{b} - \mathbf{A}(\mathbf{x}^{(s)} + \tau_s \boldsymbol{\xi}^{(s)}) \right|^2 = \left| \mathbf{b} - \mathbf{A}\mathbf{x}^{(s)} - \tau_s \mathbf{A}\boldsymbol{\xi}^{(s)} \right|^2 = \\ &= \left| \boldsymbol{\xi}^{(s)} - \tau_s \mathbf{A}\boldsymbol{\xi}^{(s)} \right|^2 = \left(\boldsymbol{\xi}^{(s)} - \tau_s \mathbf{A}\boldsymbol{\xi}^{(s)}, \boldsymbol{\xi}^{(s)} - \tau_s \mathbf{A}\boldsymbol{\xi}^{(s)} \right) = \\ &= \left(\boldsymbol{\xi}^{(s)}, \boldsymbol{\xi}^{(s)} \right) - \tau_s \left(\boldsymbol{\xi}^{(s)}, \mathbf{A}\boldsymbol{\xi}^{(s)} \right) - \tau_s \left(\mathbf{A}\boldsymbol{\xi}^{(s)}, \boldsymbol{\xi}^{(s)} \right) + \tau_s^2 \left(\mathbf{A}\boldsymbol{\xi}^{(s)}, \mathbf{A}\boldsymbol{\xi}^{(s)} \right) = \\ &= \left(\boldsymbol{\xi}^{(s)}, \boldsymbol{\xi}^{(s)} \right) - 2\tau_s \left(\mathbf{A}\boldsymbol{\xi}^{(s)}, \boldsymbol{\xi}^{(s)} \right) + \tau_s^2 \left(\mathbf{A}\boldsymbol{\xi}^{(s)}, \mathbf{A}\boldsymbol{\xi}^{(s)} \right). \end{aligned}$$

В данном случае также получаем квадратичную функцию с графиком – параболой. Если равенство нулю не достигается, то, по крайней мере, минимальное положительное значение достигается в вершине параболы

$$\frac{\partial}{\partial \tau_s} |\xi^{(s+1)}|^2 = -2(\mathbf{A}\xi^{(s)}, \xi^{(s)}) + 2\tau_s (\mathbf{A}\xi^{(s)}, \mathbf{A}\xi^{(s)}) = 0,$$

то есть при оптимальном значении

$$\tau_s = \frac{(\mathbf{A}\xi^{(s)}, \xi^{(s)})}{(\mathbf{A}\xi^{(s)}, \mathbf{A}\xi^{(s)})}.$$

Почему невязка была названа градиентом, видно из системы необходимых условий минимума (23).

Заметим также, что в методе минимальных невязок нигде не возникло требования симметрии матрицы.

2.15 Метод сопряжённых направлений

Методы скорейшего спуска и минимальных невязок характеризуются линейной скоростью сходимости. Смысл термина «линейная скорость сходимости» станет понятен, если изобразить график убывания модуля невязки от итерации к итерации в координатах с логарифмическим масштабом вертикальной оси. На таком графике снижению невязки в 100 раз будет соответствовать убывание графика ниже уровня ‘-2’, а снижению в 1000 раз – ниже уровня ‘-3’ и т.д. За исключением лучшего линейная скорость сходимости является приемлемой. Но если есть возможность ускорения сходимости, то стремление к этому является естественным. Линейной скоростью сходимости, в общем, характеризуются методы оптимизации первого порядка. Доведение скорости сходимости до предела, граничащего

со скоростью сходимости методов второго порядка, выражается в методе сопряжённых направлений, являющегося усовершенствованием метода скорейшего спуска.

С другой стороны, если изобразить траекторию сходимости (методов скорейшего спуска и минимальных невязок) в пространстве приближений к решению, то для СЛАУ с достаточно большим числом обусловленности она будет выглядеть как пила (причём как пила не с несколькими десятками зубьев, а с несколькими сотнями или даже тысячами зубьев). Можно также сравнить эту траекторию с траекторией движения шарика по поверхности цилиндрического жёлоба с небольшим наклоном оси жёлоба в однородном поле тяжести (как дождевые водостоки под кровлями домов). В этом движении можно выделить колебательные движения шарика вокруг оси жёлоба и медленный дрейф вдоль его оси. В данном случае возникает стремление сделать сходимость более монотонной. И если движение на следующей итерации практически противоположно движению на предыдущей итерации, то по двум последовательным итерациям можно получить представление о направлении этого медленного дрейфа. Поэтому за направление $\mathbf{d}^{(s)}$ принимается линейная комбинация направления скорейшего спуска $\xi^{(s)}$ и направление предыдущей итерации $\mathbf{d}^{(s-1)}$, то есть
$$\mathbf{d}^{(s)} = \xi^{(s)} + \alpha_s \mathbf{d}^{(s-1)}.$$

Оказывается, что если положить в основу определения параметра τ_s условие минимизации значения квадратичного функционала вдоль направления $\mathbf{d}^{(s)}$, а в основу определения параметра α_s – условие А-ортогональности направления $\mathbf{d}^{(s)}$ предыдущему направлению $\mathbf{d}^{(s-1)}$, то есть $(\mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s-1)}) = 0$, то направление скорейшего спуска $\xi^{(s+1)}$ на следующей итерации будет (евклидово) ортогонально не только направлению $\mathbf{d}^{(s)}$, но также всем предыдущим направлениям $\mathbf{d}^{(s-1)}$, $\mathbf{d}^{(s-2)}$, ..., $\mathbf{d}^{(0)}$, и направление

$\mathbf{d}^{(s)}$ будет \mathbf{A} -ортогонально также всем предыдущим направлениям $\mathbf{d}^{(s-1)}$, $\mathbf{d}^{(s-2)}$, ..., $\mathbf{d}^{(0)}$. Докажем это.

Уточнение приближения $\mathbf{x}^{(s+1)}$ на шаге $s+1$ определяется следующим образом $\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}$. Значение τ_s определяется из условия минимизации значения $\Phi^{(s+1)}$ квадратичного функционала

$$\begin{aligned} \Phi^{(s+1)} &= \frac{1}{2} (\mathbf{A}\mathbf{x}^{(s+1)}, \mathbf{x}^{(s+1)}) - (\mathbf{b}, \mathbf{x}^{(s+1)}) = \\ &= \frac{1}{2} (\mathbf{A}(\mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}), \mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}) - (\mathbf{b}, \mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}) = \\ &= \Phi^{(s)} + \tau_s \left(\frac{1}{2} (\mathbf{A} + \mathbf{A}^T) \mathbf{x}^{(s)} - \mathbf{b}, \mathbf{d}^{(s)} \right) + \frac{1}{2} \tau_s^2 (\mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s)}) = \\ &= \Phi^{(s)} - \tau_s (\boldsymbol{\xi}^{(s)}, \mathbf{d}^{(s)}) + \frac{1}{2} \tau_s^2 (\mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s)}) \rightarrow \min, \end{aligned}$$

что эквивалентно ортогональности следующего направления $\boldsymbol{\xi}^{(s+1)} = \boldsymbol{\xi}^{(s)} - \tau_s \mathbf{A}\mathbf{d}^{(s)}$ направлению $\mathbf{d}^{(s)}$:

$$\begin{aligned} (\boldsymbol{\xi}^{(s+1)}, \mathbf{d}^{(s)}) &= (\mathbf{b} - \mathbf{A}\mathbf{x}^{(s+1)}, \mathbf{d}^{(s)}) = (\mathbf{b} - \mathbf{A}(\mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}), \mathbf{d}^{(s)}) = \\ &= ((\mathbf{b} - \mathbf{A}\mathbf{x}^{(s)}) - \tau_s \mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s)}) = (\boldsymbol{\xi}^{(s)}, \mathbf{d}^{(s)}) - \tau_s (\mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s)}) = 0, \end{aligned}$$

откуда (здесь учитывается симметрия матрицы \mathbf{A} , то есть $\frac{1}{2}(\mathbf{A} + \mathbf{A}^T) = \mathbf{A}$)

$$\tau_s = \frac{(\boldsymbol{\xi}^{(s)}, \mathbf{d}^{(s)})}{(\mathbf{A}\mathbf{d}^{(s)}, \mathbf{d}^{(s)})}.$$

За направление $\mathbf{d}^{(s)}$ принимается комбинация $\mathbf{d}^{(s)} = \boldsymbol{\xi}^{(s)} + \alpha_s \mathbf{d}^{(s-1)}$ невязки $\boldsymbol{\xi}^{(s)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(s)}$ и направления $\mathbf{d}^{(s-1)}$ предыдущего шага (для стартового шага считается $\mathbf{d}^{(0)} = \boldsymbol{\xi}^{(0)}$). Значение α_s определяется из условия \mathbf{A} -ортогональности (ортогональности относительно скалярного произведения, порождаемого симметричной $\mathbf{A}^T = \mathbf{A}$ положительно

определённой $\forall \mathbf{x} \neq \mathbf{0} : (\mathbf{Ax}, \mathbf{x}) > 0$ матрицей \mathbf{A}) направления $\mathbf{d}^{(s)}$ направлению $\mathbf{d}^{(s-1)}$ (из \mathbf{A} -ортогональности следует линейная независимость $\mathbf{d}^{(s)}$ и $\mathbf{d}^{(s-1)}$):

$$(\mathbf{Ad}^{(s)}, \mathbf{d}^{(s-1)}) = (\mathbf{A}(\xi^{(s)} + \alpha_s \mathbf{d}^{(s-1)}), \mathbf{d}^{(s-1)}) = (\mathbf{A}\xi^{(s)}, \mathbf{d}^{(s-1)}) + \alpha_s (\mathbf{Ad}^{(s-1)}, \mathbf{d}^{(s-1)}) = 0,$$

откуда

$$\alpha_s = -\frac{(\mathbf{A}\xi^{(s)}, \mathbf{d}^{(s-1)})}{(\mathbf{Ad}^{(s-1)}, \mathbf{d}^{(s-1)})}.$$

Таким образом, последовательность вычислений следующая:

$$\xi^{(s)} = \mathbf{b} - \mathbf{Ax}^{(s)}, \quad \alpha_s = -\frac{(\mathbf{A}\xi^{(s)}, \mathbf{d}^{(s-1)})}{(\mathbf{Ad}^{(s-1)}, \mathbf{d}^{(s-1)})}, \quad \mathbf{d}^{(s)} = \xi^{(s)} + \alpha_s \mathbf{d}^{(s-1)}, \quad \tau_s = \frac{(\xi^{(s)}, \mathbf{d}^{(s)})}{(\mathbf{Ad}^{(s)}, \mathbf{d}^{(s)})},$$

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(s)} + \tau_s \mathbf{d}^{(s)}.$$

Шаг $s = 0$: $\mathbf{x}^{(0)}$ – начальное приближение,

$$\xi^{(0)} = \mathbf{b} - \mathbf{Ax}^{(0)}, \quad \mathbf{d}^{(0)} = \xi^{(0)}, \quad \tau_0 = \frac{(\xi^{(0)}, \mathbf{d}^{(0)})}{(\mathbf{Ad}^{(0)}, \mathbf{d}^{(0)})}, \quad \mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \tau_0 \mathbf{d}^{(0)}.$$

τ_0 определяется из условия $(\xi^{(1)}, \mathbf{d}^{(0)}) = 0$.

Шаг $s = 1$:

$$\xi^{(1)} = \mathbf{b} - \mathbf{Ax}^{(1)}, \quad \alpha_1 = -\frac{(\mathbf{A}\xi^{(1)}, \mathbf{d}^{(0)})}{(\mathbf{Ad}^{(0)}, \mathbf{d}^{(0)})}, \quad \mathbf{d}^{(1)} = \xi^{(1)} + \alpha_1 \mathbf{d}^{(0)}, \quad \tau_1 = \frac{(\xi^{(1)}, \mathbf{d}^{(1)})}{(\mathbf{Ad}^{(1)}, \mathbf{d}^{(1)})},$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \tau_1 \mathbf{d}^{(1)}.$$

α_1 определяется из этого условия $(\mathbf{Ad}^{(1)}, \mathbf{d}^{(0)}) = 0$. τ_1 определяется из условия $(\xi^{(2)}, \mathbf{d}^{(1)}) = 0$. Кроме того,

$$\left(\xi^{(2)}, \mathbf{d}^{(0)}\right) = \left(\xi^{(1)} - \tau_1 \mathbf{A} \mathbf{d}^{(1)}, \mathbf{d}^{(0)}\right) = \left(\xi^{(1)}, \mathbf{d}^{(0)}\right) - \tau_1 \left(\mathbf{A} \mathbf{d}^{(1)}, \mathbf{d}^{(0)}\right) = 0.$$

Шаг $s = 2$:

$$\xi^{(2)} = \mathbf{b} - \mathbf{A} \mathbf{x}^{(2)}, \quad \alpha_2 = -\frac{\left(\mathbf{A} \xi^{(2)}, \mathbf{d}^{(1)}\right)}{\left(\mathbf{A} \mathbf{d}^{(1)}, \mathbf{d}^{(1)}\right)}, \quad \mathbf{d}^{(2)} = \xi^{(2)} + \alpha_2 \mathbf{d}^{(1)}, \quad \tau_2 = \frac{\left(\xi^{(2)}, \mathbf{d}^{(2)}\right)}{\left(\mathbf{A} \mathbf{d}^{(2)}, \mathbf{d}^{(2)}\right)},$$

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} + \tau_2 \mathbf{d}^{(2)}.$$

α_2 определяется из этого условия $\left(\mathbf{A} \mathbf{d}^{(2)}, \mathbf{d}^{(1)}\right) = 0$. τ_2 определяется из условия $\left(\xi^{(3)}, \mathbf{d}^{(2)}\right) = 0$. Кроме того (здесь используется самосопряжённость (симметрия) \mathbf{A}),

$$\begin{aligned} \left(\mathbf{A} \mathbf{d}^{(2)}, \mathbf{d}^{(0)}\right) &= \left(\mathbf{A} \left(\xi^{(2)} + \alpha_2 \mathbf{d}^{(1)}\right), \mathbf{d}^{(0)}\right) = \left(\mathbf{A} \xi^{(2)}, \mathbf{d}^{(0)}\right) + \alpha_2 \left(\mathbf{A} \mathbf{d}^{(1)}, \mathbf{d}^{(0)}\right) = \\ &= \left(\xi^{(2)}, \mathbf{A} \mathbf{d}^{(0)}\right) + 0 = \tau_0^{-1} \left(\xi^{(2)}, \xi^{(0)} - \xi^{(1)}\right) = \tau_0^{-1} \left(\xi^{(2)}, \mathbf{d}^{(0)} - \left(\mathbf{d}^{(1)} - \alpha_1 \mathbf{d}^{(0)}\right)\right) = \\ &= \tau_0^{-1} (1 + \alpha_1) \left(\xi^{(2)}, \mathbf{d}^{(0)}\right) - \tau_0^{-1} \left(\xi^{(2)}, \mathbf{d}^{(1)}\right) = 0, \end{aligned}$$

$$\left(\xi^{(3)}, \mathbf{d}^{(k)}\right) = \left(\xi^{(2)} - \tau_1 \mathbf{A} \mathbf{d}^{(2)}, \mathbf{d}^{(k)}\right) = \left(\xi^{(2)}, \mathbf{d}^{(k)}\right) - \tau_1 \left(\mathbf{A} \mathbf{d}^{(2)}, \mathbf{d}^{(k)}\right) = 0, \quad k = 0; 1.$$

Предположим, выполняются соотношения:

$$\left(\mathbf{A} \mathbf{d}^{(j)}, \mathbf{d}^{(j-1)}\right) = \left(\mathbf{A} \mathbf{d}^{(j)}, \mathbf{d}^{(j-2)}\right) = \dots = \left(\mathbf{A} \mathbf{d}^{(j)}, \mathbf{d}^{(0)}\right) = 0$$

и

$$\left(\xi^{(j+1)}, \mathbf{d}^{(j)}\right) = \left(\xi^{(j+1)}, \mathbf{d}^{(j-1)}\right) = \dots = \left(\xi^{(j+1)}, \mathbf{d}^{(0)}\right) = 0.$$

Шаг $s = j + 1$:

$$\xi^{(j+1)} = \mathbf{b} - \mathbf{A} \mathbf{x}^{(j+1)}, \quad \alpha_{j+1} = -\frac{\left(\mathbf{A} \xi^{(j+1)}, \mathbf{d}^{(j)}\right)}{\left(\mathbf{A} \mathbf{d}^{(j)}, \mathbf{d}^{(j)}\right)}, \quad \mathbf{d}^{(j+1)} = \xi^{(j+1)} + \alpha_{j+1} \mathbf{d}^{(j)},$$

$$\tau_{j+1} = \frac{\left(\xi^{(j+1)}, \mathbf{d}^{(j+1)}\right)}{\left(\mathbf{A} \mathbf{d}^{(j+1)}, \mathbf{d}^{(j+1)}\right)}, \quad \mathbf{x}^{(j+2)} = \mathbf{x}^{(j+1)} + \tau_{j+1} \mathbf{d}^{(j+1)}.$$

α_{j+1} определяется из условия $(\mathbf{A}\mathbf{d}^{(j+1)}, \mathbf{d}^{(j)}) = 0$. τ_{j+1} определяется из условия $(\xi^{(j+2)}, \mathbf{d}^{(j+1)}) = 0$.

Докажем выполнение $(\mathbf{A}\mathbf{d}^{(j+1)}, \mathbf{d}^{(k)}) = 0$ для $0 \leq k \leq j-1$, $(\xi^{(j+2)}, \mathbf{d}^{(k)}) = 0$ для $0 \leq k \leq j$.

$$\begin{aligned}
(\mathbf{A}\mathbf{d}^{(j+1)}, \mathbf{d}^{(k)}) &= (\mathbf{A}(\xi^{(j+1)} + \alpha_{j+1}\mathbf{d}^{(j)}), \mathbf{d}^{(k)}) = \\
&= (\mathbf{A}\xi^{(j+1)}, \mathbf{d}^{(k)}) + \alpha_{j+1}(\mathbf{A}\mathbf{d}^{(j)}, \mathbf{d}^{(k)}) = (\xi^{(j+1)}, \mathbf{A}\mathbf{d}^{(k)}) + 0 = \\
&= \tau_k^{-1}(\xi^{(j+1)}, \xi^{(k)} - \xi^{(k+1)}) = \tau_k^{-1}(\xi^{(j+1)}, (\mathbf{d}^{(k)} - \alpha_k\mathbf{d}^{(k-1)}) - (\mathbf{d}^{(k+1)} - \alpha_{k+1}\mathbf{d}^{(k)})) = \\
&= \tau_k^{-1}(\xi^{(j+1)}, (1 + \alpha_{k+1})\mathbf{d}^{(k)} - \alpha_k\mathbf{d}^{(k-1)} - \mathbf{d}^{(k+1)}) = \\
&= \tau_k^{-1}(1 + \alpha_{k+1})(\xi^{(j+1)}, \mathbf{d}^{(k)}) - \tau_k^{-1}\alpha_k(\xi^{(j+1)}, \mathbf{d}^{(k-1)}) - \tau_k^{-1}(\xi^{(j+1)}, \mathbf{d}^{(k+1)}) = 0.
\end{aligned}$$

Тогда и для $0 \leq k \leq j$

$$(\xi^{(j+2)}, \mathbf{d}^{(k)}) = (\xi^{(j+1)} - \tau_j\mathbf{A}\mathbf{d}^{(j+1)}, \mathbf{d}^{(k)}) = (\xi^{(j+1)}, \mathbf{d}^{(k)}) - \tau_j(\mathbf{A}\mathbf{d}^{(j+1)}, \mathbf{d}^{(k)}) = 0.$$

Из условия \mathbf{A} -ортогональности направления $\mathbf{d}^{(s)} \neq \mathbf{0}$ всем предыдущим направлениям $\mathbf{d}^{(s-1)}, \mathbf{d}^{(s-2)}, \dots, \mathbf{d}^{(0)}$ следует линейная независимость $\mathbf{d}^{(s)}$ со всеми предыдущим направлениям $\mathbf{d}^{(s-1)}, \mathbf{d}^{(s-2)}, \dots, \mathbf{d}^{(0)}$.

Докажем это. Скалярное произведение $(\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(s)})$ для ненулевого вектора $\mathbf{d}^{(s)} \neq \mathbf{0}$ и положительно определённой матрицы \mathbf{A} строго положительно $(\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(s)}) > 0$.

Предположим, направление $\mathbf{d}^{(s)} \neq \mathbf{0}$ линейно зависимо от направлений $\mathbf{d}^{(s-1)}, \mathbf{d}^{(s-2)}, \dots, \mathbf{d}^{(0)}$. Это означает, что направление $\mathbf{d}^{(s)}$ может быть выражено в виде линейной комбинации $\mathbf{d}^{(s-1)}, \mathbf{d}^{(s-2)}, \dots, \mathbf{d}^{(0)}$

$$\mathbf{d}^{(s)} = \beta_1\mathbf{d}^{(s-1)} + \beta_2\mathbf{d}^{(s-2)} + \dots + \beta_s\mathbf{d}^{(0)},$$

где среди чисел $\beta_1, \beta_2, \dots, \beta_s$ обязательно есть ненулевые. Но тогда

$$\begin{aligned} (\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(s)}) &= (\mathbf{d}^{(s)}, \mathbf{A}(\beta_1\mathbf{d}^{(s-1)} + \beta_2\mathbf{d}^{(s-2)} + \dots + \beta_s\mathbf{d}^{(0)})) = \\ &= \beta_1(\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(s-1)}) + \beta_2(\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(s-2)}) + \dots + \beta_s(\mathbf{d}^{(s)}, \mathbf{A}\mathbf{d}^{(0)}) = 0, \end{aligned}$$

что противоречит предположениям $\mathbf{d}^{(s)} \neq \mathbf{0}$ и положительной определённости матрицы \mathbf{A} .

Это означает следующее. При решении СЛАУ с n неизвестными мы ищем решение в пространстве n измерений. Но в пространстве n измерений не может быть линейно независимых направлений больше n . Но $\mathbf{d}^{(1)}$ линейно независимо с $\mathbf{d}^{(0)}$, $\mathbf{d}^{(2)}$ линейно независимо с $\mathbf{d}^{(0)}$ и $\mathbf{d}^{(1)}$, $\mathbf{d}^{(3)}$ линейно независимо с $\mathbf{d}^{(0)}$, $\mathbf{d}^{(1)}$ и $\mathbf{d}^{(2)}$ и т.д. Как минимум, $\mathbf{d}^{(n)}$ уже не может быть линейно независимо с $\mathbf{d}^{(0)}$, $\mathbf{d}^{(1)}$, \dots , $\mathbf{d}^{(n-1)}$, потому что это будет уже $(n+1)$ -е линейно независимое направление. Противоречие исключается, если отказаться от $\mathbf{d}^{(n)} \neq \mathbf{0}$.

Но $\mathbf{d}^{(n)} = \mathbf{0}$ означает, что нет направления дальнейшего уточнения $\mathbf{x}^{(n)}$. Таким образом, не более чем за n итераций метод сопряжённых направлений достигает решения СЛАУ с n неизвестными (и n уравнениями).