Метод сопряженных градиентов

Методы оптимизации

Александр Безносиков

Московский физико-технический институт

3 октября 2024

Снова к Коши

- Снова решаем систему линейных уравнений:

$$Ax = b.$$

  Ищем $x \in \mathbb{R}^d$

- $A \in \mathbb{R}^{d \times d}$ положительно определенная и $b \in \mathbb{R}^d$.

Сопряженные направления

Определение сопряженных направлений

Множество векторов $\{p_i\}_{i=0}^{n-1}$ будем называть сопряженным относительно положительно определенной матрицы A, если для любых $i \neq j \in \{0, \ldots n-1\}$ следует

$$p_i^T A p_j = 0.$$

Сопряженные направления: линейная независимость

Линейная независимость сопряженных направлений

Сопряженных векторы $\{p_i\}_{i=0}^{n-1}$ является линейно независимыми.

Сопряженные направления: линейная независимость

Доказательство

- От противного:

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$

$$0 = p_m^T A p_i = \sum_{i \neq j} \lambda_j p_m^T A p_j = \lambda_m p_m^T A p_m.$$

Вопрос: почему выполнен первый и последний переходы?

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$

$$0 = p_m^T A p_i = \sum_{i \neq j} \lambda_j p_m^T A p_j = \lambda_m p_m^T A p_m.$$

  Вопрос: почему выполнен первый и последний переходы? Из-за определения сопряженности.

- **Вопрос:** что получили?

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$

$$0 = p_m^T A p_i = \sum_{i \neq j} \lambda_j p_m^T A p_j = \lambda_m p_m^T A p_m.$$

  Вопрос: почему выполнен первый и последний переходы? Из-за определения сопряженности.

- **Вопрос:** что получили? $\lambda_m = 0$.

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:

$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$

- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$

$$0 = p_m^T A p_i = \sum_{i \neq j} \lambda_j p_m^T A p_j = \lambda_m p_m^T A p_m.$$

  Вопрос: почему выполнен первый и последний переходы? Из-за определения сопряженности.

- **Вопрос:** что получили? $\lambda_m = 0$.
- **Вопрос:** что из этого следует?

Сопряженные направления: линейная независимость

Доказательство

- От противного: пусть существует p_i такое, что существуют:
$$p_i = \sum_{i \neq j} \lambda_j p_j \text{ для некоторых } \lambda_j \in \mathbb{R}.$$
- Ударим определением: возьмем скалярное произведение с Ap_m, где $m \neq i$
$$0 = p_m^T A p_i = \sum_{i \neq j} \lambda_j p_m^T A p_j = \lambda_m p_m^T A p_m.$$
Вопрос: почему выполнен первый и последний переходы? Из-за определения сопряженности.
- **Вопрос:** что получили? $\lambda_m = 0$.
- **Вопрос:** что из этого следует? пробегаем по всем $m \neq i$ и получаем, что $\lambda_m = 0$, а значит $p_i = 0$. Противоречие.

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать?

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать? Если у нас есть d сопряженных векторов, то они формируют базис. Раскладываем решение:

$$x^* = \sum_{i=0}^{d-1} \lambda_i p_i.$$

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать? Если у нас есть d сопряженных векторов, то они формируют базис. Раскладываем решение:
$$x^* = \sum_{i=0}^{d-1} \lambda_i p_i.$$
- **Вопрос:** как найти λ_i?

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать? Если у нас есть d сопряженных векторов, то они формируют базис. Раскладываем решение:
$$x^* = \sum_{i=0}^{d-1} \lambda_i p_i.$$
- **Вопрос:** как найти λ_i? Возьмем скалярное произведение с Ap_j:
$$p_j^T A x^* = \sum_{i=0}^{d-1} \lambda_i p_j^T A p_i = \lambda_j p_j^T A p_j.$$
Здесь опять воспользовались определением сопряженности.

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать? Если у нас есть d сопряженных векторов, то они формируют базис. Раскладываем решение:

$$x^* = \sum_{i=0}^{d-1} \lambda_i p_i.$$

- **Вопрос:** как найти λ_i? Возьмем скалярное произведение с Ap_j:

$$p_j^T A x^* = \sum_{i=0}^{d-1} \lambda_i p_j^T A p_i = \lambda_j p_j^T A p_j.$$

  Здесь опять воспользовались определением сопряженности.
- Заметим, что $Ax^* = b$, тогда $p_j^T b = \lambda_j p_j^T A p_j$.

Сопряженные направления: как использовать?

- У нас есть какой-то базис. **Вопрос:** как его можно использовать? Если у нас есть d сопряженных векторов, то они формируют базис. Раскладываем решение:
$$x^* = \sum_{i=0}^{d-1} \lambda_i p_i.$$
- **Вопрос:** как найти λ_i? Возьмем скалярное произведение с Ap_j:
$$p_j^T A x^* = \sum_{i=0}^{d-1} \lambda_i p_j^T A p_i = \lambda_j p_j^T A p_j.$$
Здесь опять воспользовались определением сопряженности.
- Заметим, что $Ax^* = b$, тогда $p_j^T b = \lambda_j p_j^T A p_j$.
- Тогда
$$\lambda_j = \frac{p_j^T b}{p_j^T A p_j}.$$

Сопряженные направления: как использовать?

- **Вопрос:** а видно ли какие-то проблемы?

Сопряженные направления: как использовать?

- **Вопрос:** а видно ли какие-то проблемы? Все хорошо кроме того, что мы сами придумали сопряженные направления, сами сказали, что они существуют, а как их получать в реальности пока непонятно.

Сопряженные направления: как использовать?

- **Вопрос:** а видно ли какие-то проблемы? Все хорошо кроме того, что мы сами придумали сопряженные направления, сами сказали, что они существуют, а как их получать в реальности пока непонятно.
- Начнем превращать рассуждения в некоторый итеративный метод:

$$\boxed{x^{k+1} = x^k + \alpha_k p_k.}$$

Т.е. предполагается, что мы на каждой итерации будем искать новое p_k и подбивать к нему α_k.

Метод сопряженных градиентов: α

Мы уже более менее поняли, как искать α, когда искали λ. **Вопрос:** $\lambda = \alpha$?

Метод сопряженных градиентов: α

Мы уже более менее поняли, как искать α, когда искали λ. **Вопрос:** $\lambda = \alpha$? Не всегда. Итеративная схема с α:

$$x^{k+1} = x^0 + \sum_{i=0}^{k} \alpha_i p_i.$$

Метод сопряженных градиентов: α

Мы уже более менее поняли, как искать α, когда искали λ. **Вопрос:** $\lambda = \alpha$? Не всегда. Итеративная схема с α:

$$x^{k+1} = x^0 + \sum_{i=0}^{k} \alpha_i p_i.$$

Итеративная схема с λ:

$$x^{k+1} = \sum_{i=0}^{k} \lambda_i p_i.$$

Метод сопряженных градиентов: α

Мы уже более менее поняли, как искать α, когда искали λ. **Вопрос:** $\lambda = \alpha$? Не всегда. Итеративная схема с α:

$$x^{k+1} = x^0 + \sum_{i=0}^{k} \alpha_i p_i.$$

Итеративная схема с λ:

$$x^{k+1} = \sum_{i=0}^{k} \lambda_i p_i.$$

Получается, что $\alpha_i = \lambda_i$, если $x^0 = 0$.

Метод сопряженных градиентов: α

Мы уже более менее поняли, как искать α, когда искали λ. **Вопрос:** $\lambda = \alpha$? Не всегда. Итеративная схема с α:

$$x^{k+1} = x^0 + \sum_{i=0}^{k} \alpha_i p_i.$$

Итеративная схема с λ:

$$x^{k+1} = \sum_{i=0}^{k} \lambda_i p_i.$$

Получается, что $\alpha_i = \lambda_i$, если $x^0 = 0$. Нужна формула поиска α, так как стартовать из 0 хорошо, но, возможно, у нас есть более близкий кандидат в качестве стартовой точки.

Метод сопряженных градиентов: α

- Можно разложить x^0 по базису и найти для него $\tilde{\lambda}_i$:

Метод сопряженных градиентов: α

- Можно разложить x^0 по базису и найти для него $\tilde{\lambda}_i$:

$$x^0 = \sum_{i=0}^{d-1} \tilde{\lambda}_i p_i, \text{ где } \tilde{\lambda}_i = \frac{p_i^T A x^0}{p_i^T A p_i}.$$

Метод сопряженных градиентов: α

- Можно разложить x^0 по базису и найти для него $\tilde{\lambda}_i$:

$$x^0 = \sum_{i=0}^{d-1} \tilde{\lambda}_i p_i, \text{ где } \tilde{\lambda}_i = \frac{p_i^T A x^0}{p_i^T A p_i}.$$

- Тогда справедливо следующее утверждение:

$$x^0 + \sum_{i=0}^{d-1} \alpha_i p_i = \sum_{i=0}^{d-1} \left(\frac{p_i^T A x^0}{p_i^T A p_i} + \alpha_i \right) p_i = \sum_{i=0}^{d-1} \lambda_i p_i = \sum_{i=0}^{d-1} \frac{p_i^T b}{p_i^T A p_i} p_i.$$

Метод сопряженных градиентов: α

- Можно разложить x^0 по базису и найти для него $\tilde{\lambda}_i$:

$$x^0 = \sum_{i=0}^{d-1} \tilde{\lambda}_i p_i, \text{ где } \tilde{\lambda}_i = \frac{p_i^T A x^0}{p_i^T A p_i}.$$

- Тогда справедливо следующее утверждение:

$$x^0 + \sum_{i=0}^{d-1} \alpha_i p_i = \sum_{i=0}^{d-1} \left(\frac{p_i^T A x^0}{p_i^T A p_i} + \alpha_i \right) p_i = \sum_{i=0}^{d-1} \lambda_i p_i = \sum_{i=0}^{d-1} \frac{p_i^T b}{p_i^T A p_i} p_i.$$

- Получаем

$$\alpha_k = \frac{p_k^T (b - A x^0)}{p_k^T A p_k}.$$

Метод сопряженных градиентов: α

- Результат уже нормальный, можно чуть-чуть еще докрутить:

$$p_k^T A(x^k - x^0) = 0.$$

Вопрос: почему?

Метод сопряженных градиентов: α

- Результат уже нормальный, можно чуть-чуть еще докрутить:

$$p_k^T A(x^k - x^0) = 0.$$

Вопрос: почему? $(x^k - x^0) = \sum_{i=0}^{k-1} \alpha_i p_i$, а p_i и p_k сопряженные относительно A.

Метод сопряженных градиентов: α

- Результат уже нормальный, можно чуть-чуть еще докрутить:

$$p_k^T A(x^k - x^0) = 0.$$

Вопрос: почему? $(x^k - x^0) = \sum_{i=0}^{k-1} \alpha_i p_i$, а p_i и p_k сопряженные относительно A.

- Тогда можно так:

$$\alpha_k = \frac{p_k^T (b - Ax^k)}{p_k^T A p_k} = -\frac{p_k^T r_k}{p_k^T A p_k}.$$

Здесь введено обозначение $r_k = Ax^k - b$.

Метод сопряженных градиентов: физический смысл α

- Рассмотрим шаг метода $x^{k+1} = x^k + \alpha_k p_k$, а также функцию

$$f(x) = \frac{1}{2} x^T A x - bx.$$

Метод сопряженных градиентов: физический смысл α

- Рассмотрим шаг метода $x^{k+1} = x^k + \alpha_k p_k$, а также функцию
$$f(x) = \frac{1}{2}x^T Ax - bx.$$
- **Вопрос:** что это за функция?

Метод сопряженных градиентов: физический смысл α

- Рассмотрим шаг метода $x^{k+1} = x^k + \alpha_k p_k$, а также функцию

$$f(x) = \frac{1}{2} x^T A x - bx.$$

- **Вопрос:** что это за функция? Ее минимизация эквивалентна поиску решения системы уравнений: $\nabla f(x^*) = Ax^* - b = 0$.

Метод сопряженных градиентов: физический смысл α

- Рассмотрим шаг метода $x^{k+1} = x^k + \alpha_k p_k$, а также функцию

$$f(x) = \frac{1}{2} x^T A x - bx.$$

- **Вопрос:** что это за функция? Ее минимизация эквивалентна поиску решения системы уравнений: $\nabla f(x^*) = Ax^* - b = 0$.
- Рассмотрим:

$$g(\alpha) = f(x^k + \alpha p_k).$$

Где у этой функции минимум по α^*?

Метод сопряженных градиентов: физический смысл α

- Рассмотрим шаг метода $x^{k+1} = x^k + \alpha_k p_k$, а также функцию

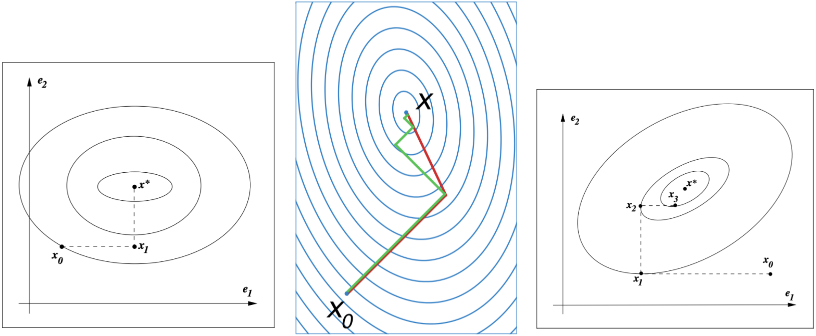
$$f(x) = \frac{1}{2} x^T A x - bx.$$

- **Вопрос:** что это за функция? Ее минимизация эквивалентна поиску решения системы уравнений: $\nabla f(x^*) = Ax^* - b = 0$.
- Рассмотрим:

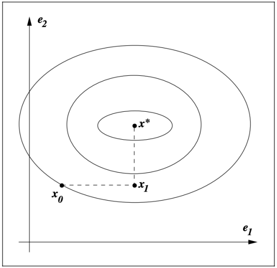
$$g(\alpha) = f(x^k + \alpha p_k).$$

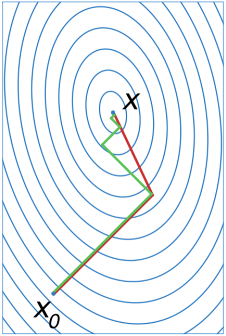
Где у этой функции минимум по α^*? $\alpha^* = \frac{p_k^T(b - Ax^k)}{p_k^T A p_k} = \alpha_k$. Эта и есть физика — минимизация вдоль p_k.

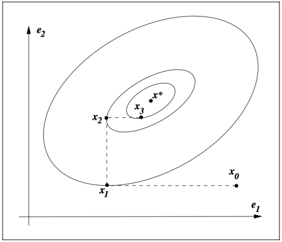
Метод сопряженных градиентов: физический смысл α



Метод сопряженных градиентов: физический смысл α

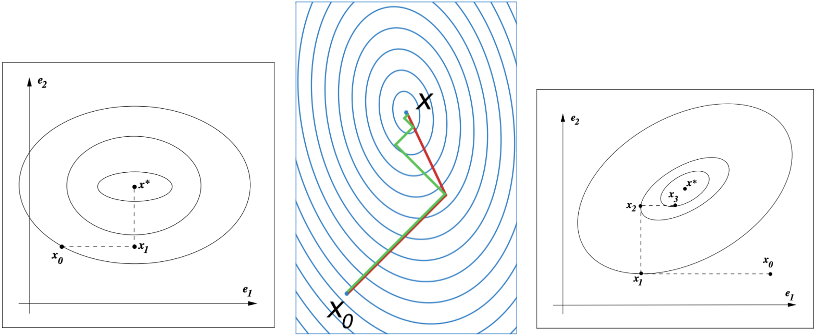






- На второй картинке показано, что сопряженные направления не ортогональны (в привычном смысле этого смысла слова).

Метод сопряженных градиентов: физический смысл α



- На второй картинке показано, что сопряженные направления не ортогональны (в привычном смысле этого смысла слова).
- На третьей картинке направления не являются сопряженным относительно A из-за этого возникают проблемы.

Метод сопряженных градиентов: физический смысл α

Физический смысл p

Если $\{p_i\}_{i=0}^k$ сопряженные направления, то для любого $k \geq 0$ и $i \leq k$ справедливо:

$$r_{k+1}^T p_i = 0 \text{ то же самое, что } \langle \nabla f(x^{k+1}), p_i \rangle = 0.$$

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:**

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 A p_0 = r_0 + \alpha_0 A p_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T A p_0}$:

$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T A p_0 = 0.$$

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 A p_0 = r_0 + \alpha_0 A p_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T A p_0}$:
$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T A p_0 = 0.$$
- **Предположение:** пусть предположение верно для всех $i \leq k$.

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 Ap_0 = r_0 + \alpha_0 Ap_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T Ap_0}$:
$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T Ap_0 = 0.$$
- **Предположение:** пусть предположение верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$.

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 Ap_0 = r_0 + \alpha_0 Ap_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T Ap_0}$:
$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T Ap_0 = 0.$$
- **Предположение:** пусть предположение верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Рассмотрим:
$$r_{k+1} = Ax^{k+1} - b = Ax^k - b + \alpha_k Ap_k = r_k + \alpha_k Ap_k.$$

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 Ap_0 = r_0 + \alpha_0 Ap_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T Ap_0}$:
$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T Ap_0 = 0.$$
- **Предположение:** пусть предположение верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Рассмотрим:
$$r_{k+1} = Ax^{k+1} - b = Ax^k - b + \alpha_k Ap_k = r_k + \alpha_k Ap_k.$$

  Откуда в силу выбора α_k
$$p_k^T r_{k+1} = p_k^T r_k + \alpha_k p_k^T Ap_k = 0.$$

  Для $i < k$:
$$p_i^T r_{k+1} = p_i^T r_k + \alpha_k p_i^T Ap_k = 0.$$

Вопрос: почему?

Метод сопряженных градиентов: физический смысл α

Доказательство

- По индукции. **База:** $r_1 = Ax^1 - b = Ax^0 - b + \alpha_0 A p_0 = r_0 + \alpha_0 A p_0$, в силу выбора $\alpha_0 = -\frac{p_0^T r_0}{p_0^T A p_0}$:
$$p_0^T r_1 = p_0^T r_0 + \alpha_0 p_0^T A p_0 = 0.$$
- **Предположение:** пусть предположение верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Рассмотрим:
$$r_{k+1} = Ax^{k+1} - b = Ax^k - b + \alpha_k A p_k = r_k + \alpha_k A p_k.$$

  Откуда в силу выбора α_k
$$p_k^T r_{k+1} = p_k^T r_k + \alpha_k p_k^T A p_k = 0.$$

  Для $i < k$:
$$p_i^T r_{k+1} = p_i^T r_k + \alpha_k p_i^T A p_k = 0.$$

Вопрос: почему? В силу индукции и сопряженности.

Метод сопряженных градиентов: p

- Пора уже искать p.
- **Вопрос:** что хотим потребовать от техники поиска p (вспомните, например, процедуру Грама-Шмидта)?

Метод сопряженных градиентов: p

- Пора уже искать p.
- **Вопрос:** что хотим потребовать от техники поиска p (вспомните, например, процедуру Грама-Шмидта)? «Дешевизну» подсчета p_k:

$$p_k = -r_k + \beta_k p_{k-1},$$

где β_k – некоторый коэффициент. Чтобы найти p_k нужно знать только p_{k-1} и r_k, а старые r_i и p_i можно уже забыть (они учтены в x^k).

Метод сопряженных градиентов: p

- Пора уже искать p.
- **Вопрос:** что хотим потребовать от техники поиска p (вспомните, например, процедуру Грама-Шмидта)? «Дешевизну» подсчета p_k:

$$p_k = -r_k + \beta_k p_{k-1},$$

где β_k – некоторый коэффициент. Чтобы найти p_k нужно знать только p_{k-1} и r_k, а старые r_i и p_i можно уже забыть (они учтены в x^k).
- **Вопрос:** как найти β_k?

Метод сопряженных градиентов: p

- Пора уже искать p.
- **Вопрос:** что хотим потребовать от техники поиска p (вспомните, например, процедуру Грама-Шмидта)? «Дешевизну» подсчета p_k:

$$p_k = -r_k + \beta_k p_{k-1},$$

где β_k – некоторый коэффициент. Чтобы найти p_k нужно знать только p_{k-1} и r_k, а старые r_i и p_i можно уже забыть (они учтены в x^k).
- **Вопрос:** как найти β_k? Сопряженность p_k и p_{k-1}:

$$0 = p_{k-1}^T A p_k = -p_{k-1}^T A r_k + \beta_k p_{k-1}^T A p_{k-1},$$

Метод сопряженных градиентов: p

- Пора уже искать p.
- **Вопрос:** что хотим потребовать от техники поиска p (вспомните, например, процедуру Грама-Шмидта)? «Дешевизну» подсчета p_k:

$$p_k = -r_k + \beta_k p_{k-1},$$

где β_k – некоторый коэффициент. Чтобы найти p_k нужно знать только p_{k-1} и r_k, а старые r_i и p_i можно уже забыть (они учтены в x^k).
- **Вопрос:** как найти β_k? Сопряженность p_k и p_{k-1}:

$$0 = p_{k-1}^T A p_k = -p_{k-1}^T A r_k + \beta_k p_{k-1}^T A p_{k-1},$$

откуда

$$\beta_k = \frac{p_{k-1}^T A r_k}{p_{k-1}^T A p_{k-1}}.$$

Метод сопряженных градиентов

Алгоритм 1 Метод сопряженных градиентов

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $r_0 = Ax_0 - b$, $p_0 = -r_0$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**
2: $\quad \alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$
3: $\quad x^{k+1} = x^k + \alpha_k p_k$
4: $\quad r_{k+1} = Ax^{k+1} - b$
5: $\quad \beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$
6: $\quad p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$
7: **end for**

Выход: x^K

Метод сопряженных градиентов

Алгоритм 2 Метод сопряженных градиентов

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $r_0 = Ax_0 - b$, $p_0 = -r_0$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $r_{k+1} = Ax^{k+1} - b$

5: $\beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$

6: $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$

7: **end for**

Выход: x^K

Вопрос: а почему вдруг градиентов то?

Метод сопряженных градиентов

Алгоритм 3 Метод сопряженных градиентов

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $r_0 = Ax_0 - b$, $p_0 = -r_0$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $r_{k+1} = Ax^{k+1} - b$

5: $\beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$

6: $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$

7: **end for**

Выход: x^K

Вопрос: а почему вдруг градиентов то? $r_k = Ax^k - b = \nabla f(x^k)$. Это стоит запомнить.

Метод сопряженных градиентов: доказательство

- **Вопрос:** а может мы уже доказали оценку на сходимость?

Метод сопряженных градиентов: доказательство

- **Вопрос:** а может мы уже доказали оценку на сходимость? Близки к этому, мы знаем, что если все $\{p_i\}$ сопряженные направления, то нам хватит d шагов, чтобы восстановить коэффициенты для x^* в базисе из $\{p_i\}$.

Метод сопряженных градиентов: доказательство

- **Вопрос:** а может мы уже доказали оценку на сходимость? Близки к этому, мы знаем, что если все $\{p_i\}$ сопряженные направления, то нам хватит d шагов, чтобы восстановить коэффициенты для x^* в базисе из $\{p_i\}$.
- **Вопрос:** Знаем ли, что все $\{p_i\}$ сопряженные?

Метод сопряженных градиентов: доказательство

- **Вопрос:** а может мы уже доказали оценку на сходимость? Близки к этому, мы знаем, что если все $\{p_i\}$ сопряженные направления, то нам хватит d шагов, чтобы восстановить коэффициенты для x^* в базисе из $\{p_i\}$.
- **Вопрос:** Знаем ли, что все $\{p_i\}$ сопряженные? Нет, мы знаем только, что p_k и p_{k-1} сопряжены в силу подбора β_k. Надо показать более широкое утверждение:

$$\text{Для любого } k \geq 1 \text{ для любого } i < k \text{ выполняется } p_k^T A p_i = 0.$$

Метод сопряженных градиентов: доказательство

- Идем по индукции.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^k$ сопряженные.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^{k}$ сопряженные.
- **Переход:** Докажем для $k+1$.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^{k}$ сопряженные.
- **Переход:** Докажем для $k+1$. p_{k+1} и p_k сопряжены в силу подбора β_{k+1}.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^k$ сопряженные.
- **Переход:** Докажем для $k+1$. p_{k+1} и p_k сопряжены в силу подбора β_{k+1}. Рассмотрим $i < k$:

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i = -r_{k+1}^T A p_i.$$

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^k$ сопряженные.
- **Переход:** Докажем для $k+1$. p_{k+1} и p_k сопряжены в силу подбора β_{k+1}. Рассмотрим $i < k$:

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i = -r_{k+1}^T A p_i.$$

Вопрос: почему валиден второй переход?

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^k$ сопряженные.
- **Переход:** Докажем для $k+1$. p_{k+1} и p_k сопряжены в силу подбора β_{k+1}. Рассмотрим $i < k$:

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i = -r_{k+1}^T A p_i.$$

Вопрос: почему валиден второй переход? в силу предположения индукции и того, что $i < k$.

Метод сопряженных градиентов: доказательство

- Идем по индукции.
- **База:** p_0 и p_1 сопряжены в силу подбора β_1.
- **Предположение:** пусть для $k \geq 1$ все $\{p_i\}_{i=0}^k$ сопряженные.
- **Переход:** Докажем для $k+1$. p_{k+1} и p_k сопряжены в силу подбора β_{k+1}. Рассмотрим $i < k$:

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i = -r_{k+1}^T A p_i.$$

  Вопрос: почему валиден второй переход? в силу предположения индукции и того, что $i < k$.
- Осталось показать, что $-r_{k+1}^T A p_i = 0$. Запомним это.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\mathrm{span}\{r_0, \ldots r_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$ и $\mathrm{span}\{p_0, \ldots p_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \text{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \text{span}\{r_0, \ldots A^k r_0\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \text{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \text{span}\{r_0, \ldots A^k r_0\}$. Тогда $Ap_k \in \text{span}\{Ar_0, \ldots A^{k+1} r_0\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\mathrm{span}\{r_0, \ldots r_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$ и $\mathrm{span}\{p_0, \ldots p_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \mathrm{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \mathrm{span}\{r_0, \ldots A^k r_0\}$. Тогда $Ap_k \in \mathrm{span}\{Ar_0, \ldots A^{k+1} r_0\}$.А зная, что $r_{k+1} = r_k + \alpha_k A p_k$, получаем одно включение $r_{k+1} \in \{r_0, \ldots A^{k+1} r_0\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\mathrm{span}\{r_0, \ldots r_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$ и $\mathrm{span}\{p_0, \ldots p_k\} = \mathrm{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \mathrm{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \mathrm{span}\{r_0, \ldots A^k r_0\}$. Тогда $Ap_k \in \mathrm{span}\{Ar_0, \ldots A^{k+1} r_0\}$.А зная, что $r_{k+1} = r_k + \alpha_k A p_k$, получаем одно включение $r_{k+1} \in \{r_0, \ldots A^{k+1} r_0\}$. Откуда $\mathrm{span}\{r_0, \ldots r_{k+1}\} \subseteq \mathrm{span}\{r_0, \ldots A^{k+1} r_0\}$, но нам нужно равенство.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\mathrm{span}\{r_0, \dots r_k\} = \mathrm{span}\{r_0, \dots A^k r_0\}$ и $\mathrm{span}\{p_0, \dots p_k\} = \mathrm{span}\{r_0, \dots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \mathrm{span}\{r_0, \dots A^k r_0\}$ и $p_k \in \mathrm{span}\{r_0, \dots A^k r_0\}$. Тогда $Ap_k \in \mathrm{span}\{Ar_0, \dots A^{k+1} r_0\}$.А зная, что $r_{k+1} = r_k + \alpha_k A p_k$, получаем одно включение $r_{k+1} \in \{r_0, \dots A^{k+1} r_0\}$. Откуда $\mathrm{span}\{r_0, \dots r_{k+1}\} \subseteq \mathrm{span}\{r_0, \dots A^{k+1} r_0\}$, но нам нужно равенство. Заметим, что из второго предположения: $A^{k+1} r_0 = A(A^k r_0) \in \mathrm{span}\{Ap_0, \dots Ap_k\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \text{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \text{span}\{r_0, \ldots A^k r_0\}$. Тогда $Ap_k \in \text{span}\{Ar_0, \ldots A^{k+1} r_0\}$.А зная, что $r_{k+1} = r_k + \alpha_k A p_k$, получаем одно включение $r_{k+1} \in \{r_0, \ldots A^{k+1} r_0\}$. Откуда $\text{span}\{r_0, \ldots r_{k+1}\} \subseteq \text{span}\{r_0, \ldots A^{k+1} r_0\}$, но нам нужно равенство. Заметим, что из второго предположения: $A^{k+1} r_0 = A(A^k r_0) \in \text{span}\{Ap_0, \ldots Ap_k\}$. Так как $(r_{i+1} - r_i)/\alpha_i = Ap_i$ получаем, что $A^{k+1} r_0 \in \text{span}\{r_0, \ldots r_{k+1}\}$.

Метод сопряженных градиентов: доказательство

- Докажем, что для $k \geq 0$ выполнено $\text{span}\{r_0, \ldots r_k\} = \text{span}\{r_0, \ldots A^k r_0\}$ и $\text{span}\{p_0, \ldots p_k\} = \text{span}\{r_0, \ldots A^k r_0\}$.
- По индукции. **База:** следует из инициализации.
- **Предположение:** пусть верно для всех $i \leq k$.
- **Переход:** докажем для $k+1$. Используя предположение: $r_k \in \text{span}\{r_0, \ldots A^k r_0\}$ и $p_k \in \text{span}\{r_0, \ldots A^k r_0\}$. Тогда $Ap_k \in \text{span}\{Ar_0, \ldots A^{k+1} r_0\}$.А зная, что $r_{k+1} = r_k + \alpha_k A p_k$, получаем одно включение $r_{k+1} \in \{r_0, \ldots A^{k+1} r_0\}$. Откуда $\text{span}\{r_0, \ldots r_{k+1}\} \subseteq \text{span}\{r_0, \ldots A^{k+1} r_0\}$, но нам нужно равенство. Заметим, что из второго предположения: $A^{k+1} r_0 = A(A^k r_0) \in \text{span}\{Ap_0, \ldots Ap_k\}$. Так как $(r_{i+1} - r_i)/\alpha_i = Ap_i$ получаем, что $A^{k+1} r_0 \in \text{span}\{r_0, \ldots r_{k+1}\}$. Отсюда $\text{span}\{r_0, \ldots A^{k+1} r_0\} \subseteq \text{span}\{r_0, \ldots r_{k+1}\}$. Включение в обе стороны доказано.

Метод сопряженных градиентов: доказательство

- Остался еще переход для второй части.

Метод сопряженных градиентов: доказательство

- Остался еще переход для второй части.
- По правилу подсчета p_{k+1}:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{p_0, \ldots p_k, r_{k+1}\}.$$

Метод сопряженных градиентов: доказательство

- Остался еще переход для второй части.
- По правилу подсчета p_{k+1}:

$$\mathrm{span}\{p_0, \ldots p_{k+1}\} = \mathrm{span}\{p_0, \ldots p_k, r_{k+1}\}.$$

- По второму предположению индукции:

$$\mathrm{span}\{p_0, \ldots p_{k+1}\} = \mathrm{span}\{r_0, \ldots A^k r_0, r_{k+1}\}.$$

Метод сопряженных градиентов: доказательство

- Остался еще переход для второй части.
- По правилу подсчета p_{k+1}:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{p_0, \ldots p_k, r_{k+1}\}.$$

- По второму предположению индукции:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{r_0, \ldots A^k r_0, r_{k+1}\}.$$

- По первому предположению:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{r_0, \ldots r_k, r_{k+1}\}.$$

Метод сопряженных градиентов: доказательство

- Остался еще переход для второй части.
- По правилу подсчета p_{k+1}:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{p_0, \ldots p_k, r_{k+1}\}.$$

- По второму предположению индукции:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{r_0, \ldots A^k r_0, r_{k+1}\}.$$

- По первому предположению:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{r_0, \ldots r_k, r_{k+1}\}.$$

- По только что доказанному:

$$\text{span}\{p_0, \ldots p_{k+1}\} = \text{span}\{r_0, \ldots, A^{k+1} r_0\}.$$

Метод сопряженных градиентов: доказательство

- Возвращаемся к $-r_{k+1}^T A p_i = 0$ для $i < k$.

Метод сопряженных градиентов: доказательство

- Возвращаемся к $-r_{k+1}^T A p_i = 0$ для $i < k$.
- Теперь мы знаем, что
$$p_i \in \text{span}\{r_0, \ldots, A^i r_0\}.$$

Метод сопряженных градиентов: доказательство

- Возвращаемся к $-r_{k+1}^T A p_i = 0$ для $i < k$.
- Теперь мы знаем, что
$$p_i \in \mathrm{span}\{r_0, \ldots, A^i r_0\}.$$
- Откуда
$$Ap_i \in \mathrm{span}\{Ar_0, \ldots, A^{i+1} r_0\}.$$

Метод сопряженных градиентов: доказательство

- Возвращаемся к $-r_{k+1}^T A p_i = 0$ для $i < k$.
- Теперь мы знаем, что
$$p_i \in \mathrm{span}\{r_0, \ldots, A^i r_0\}.$$
- Откуда $$Ap_i \in \mathrm{span}\{Ar_0, \ldots, A^{i+1} r_0\}.$$
- Из только, что доказанного
$$Ap_i \in \mathrm{span}\{Ar_0, \ldots, A^{i+1} r_0\} \subseteq \mathrm{span}\{p_0, \ldots, p_{i+1}\}.$$

Метод сопряженных градиентов: доказательство

- Возвращаемся к $-r_{k+1}^T A p_i = 0$ для $i < k$.
- Теперь мы знаем, что
$$p_i \in \text{span}\{r_0, \ldots, A^i r_0\}.$$
- Откуда $$Ap_i \in \text{span}\{Ar_0, \ldots, A^{i+1} r_0\}.$$
- Из только, что доказанного
$$Ap_i \in \text{span}\{Ar_0, \ldots, A^{i+1} r_0\} \subseteq \text{span}\{p_0, \ldots, p_{i+1}\}.$$
- Но все p_j для j от 0 до i ортогональны r^{k+1} в силу то, что $\{p_j\}$ сопряженные в силу предположения индукции. А значит получили то, что нужно.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем d итераций.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем d итераций.

Эквивалентно минимизации сильной выпуклой квадратичной задачи.

Метод сопряженных градиентов: сходимость

- **Вопрос:** то, что получили плохо или хорошо?

Метод сопряженных градиентов: сходимость

- **Вопрос:** то, что получили плохо или хорошо? На самом деле не очень. И метод сопрояженных градиентов в момент своего появления 70 лет назад столкнулся с этой проблемой. Т.е. для *точного* решения системы уравнений конкурентен, но не особо популярен.

Метод сопряженных градиентов: сходимость

- **Вопрос:** то, что получили плохо или хорошо? На самом деле не очень. И метод сопрояженных градиентов в момент своего появления 70 лет назад столкнулся с этой проблемой. Т.е. для *точного* решения системы уравнений конкурентен, но не особо популярен.
- Ключевое слово в предыдущем абзаце «*точного*». Метод сопряженных градиентов можно останавливать раньше, он итеративный. А это уже интереснее.

Метод сопряженных градиентов: сходимость

- **Вопрос:** то, что получили плохо или хорошо? На самом деле не очень. И метод сопрояженных градиентов в момент своего появления 70 лет назад столкнулся с этой проблемой. Т.е. для *точного* решения системы уравнений конкурентен, но не особо популярен.
- Ключевое слово в предыдущем абзаце «*точного*». Метод сопряженных градиентов можно останавливать раньше, он итеративный. А это уже интереснее.
- Есть особенности сходимости, которые делают метод еще быстрее.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем r итераций, где r – число уникальных собственных значений матрицы.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем r итераций, где r – число уникальных собственных значений матрицы.

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d имеет следующую оценку сходимости:

$$\|x^k - x^*\|_A \leq 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k \|x^0 - x^*\|_A.$$

Здесь $\|x\|_A^2 = x^T A x$ и $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем r итераций, где r – число уникальных собственных значений матрицы.

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d имеет следующую оценку сходимости:

$$\|x^k - x^*\|_A \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \|x^0 - x^*\|_A.$$

Здесь $\|x\|_A^2 = x^T A x$ и $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$.

Метод сопряженных градиентов: сходимость

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d находит точное решение за не более чем r итераций, где r – число уникальных собственных значений матрицы.

Теорема о сходимости метода сопряженных градиентов

Метод сопряженных градиентов для решения системы линейных уравнений с квадратной положительно определенной матрицей размера d имеет следующую оценку сходимости:

$$\|x^k - x^*\|_A \leq 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k \|x^0 - x^*\|_A.$$

Здесь $\|x\|_A^2 = x^T A x$ и $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$.

Метод сопряженных градиентов

Алгоритм 4 Метод сопряженных градиентов (классическая версия)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $r_0 = Ax_0 - b$, $p_0 = -r_0$ количество итераций K

1: **for** $k = 0, 1, \dots, K-1$ **do**

2: $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $r_{k+1} = r_k + \alpha_k A p_k$

5: $\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

6: $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$

7: **end for**

Выход: x^K

Метод сопряженных градиентов

Алгоритм 5 Метод сопряженных градиентов (классическая версия)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $r_0 = Ax_0 - b$, $p_0 = -r_0$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $r_{k+1} = r_k + \alpha_k A p_k$

5: $\beta_{k+1} = \frac{r_{\mathbf{k+1}}^T r_{\mathbf{k+1}}}{r_k^T r_k}$

6: $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$

7: **end for**

Выход: x^K

Вспомним, что градиент то «зашит» в $r_k = Ax^k - b = \nabla f(x^k)$.

Метод сопряженных градиентов для произвольной задачи

Алгоритм 6 Метод сопряженных градиентов (Флетчер - Ривс)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $p_0 = -\nabla f(x_0)$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\alpha_k = ?$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $\beta_{k+1} = \frac{\langle \nabla f(x^{k+1}), \nabla f(x^{k+1}) \rangle}{\langle \nabla f(x^k), \nabla f(x^k) \rangle}$

5: $p_{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} p_k$

6: **end for**

Выход: x^K

Метод сопряженных градиентов для произвольной задачи

Алгоритм 7 Метод сопряженных градиентов (Флетчер - Ривс)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $p_0 = -\nabla f(x_0)$ количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\alpha_k = ?$

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $\beta_{k+1} = \frac{\langle \nabla f(x^{k+1}), \nabla f(x^{k+1}) \rangle}{\langle \nabla f(x^k), \nabla f(x^k) \rangle}$

5: $p_{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} p_k$

6: **end for**

Выход: x^K

Вопрос: как искать шаг α_k?

Метод сопряженных градиентов для произвольной задачи

Алгоритм 8 Метод сопряженных градиентов (Флетчер - Ривс)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $p_0 = -\nabla f(x_0)$ количество итераций K

1: **for** $k = 0, 1, \dots, K-1$ **do**
2: $\alpha_k = ?$
3: $x^{k+1} = x^k + \alpha_k p_k$
4: $\beta_{k+1} = \frac{\langle \nabla f(x^{k+1}), \nabla f(x^{k+1}) \rangle}{\langle \nabla f(x^k), \nabla f(x^k) \rangle}$
5: $p_{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} p_k$
6: **end for**

Выход: x^K

Вопрос: как искать шаг α_k? Мы хотим минимизировать вдоль направления p_k, получаем одномерную функцию зависящую от α. Вспомним про бинпоиск и золотое сечение.

Метод сопряженных градиентов для произвольной задачи

Алгоритм 9 Метод сопряженных градиентов (Полак - Рибьер)

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, $p_0 = -\nabla f(x_0)$ количество итераций K

1: **for** $k = 0, 1, \dots, K-1$ **do**

2: $\alpha_k =$ Линейый поиск

3: $x^{k+1} = x^k + \alpha_k p_k$

4: $\beta_{k+1} = \frac{\langle \nabla f(x^{k+1}), \nabla f(x^{k+1}) - \nabla f(x^k) \rangle}{\langle \nabla f(x^k), \nabla f(x^k) \rangle}$

5: $p_{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} p_k$

6: **end for**

Выход: x^K

Метод сопряженных градиентов для произвольной задачи

- Обобщения работают неплохо, но гарантии в теории далеко не самые оптимистичные.

Метод сопряженных градиентов для произвольной задачи

- Обобщения работают неплохо, но гарантии в теории далеко не самые оптимистичные.
- Лучше иногда делать «рестарты». В данном случае «рестарты» предполагают иногда брать $\beta_k = 0$, забывая историю. **Вопрос:** итерация какого метода тогда получится?

Метод сопряженных градиентов для произвольной задачи

- Обобщения работают неплохо, но гарантии в теории далеко не самые оптимистичные.
- Лучше иногда делать «рестарты». В данном случае «рестарты» предполагают иногда брать $\beta_k = 0$, забывая историю. **Вопрос:** итерация какого метода тогда получится? Градиентный спуск.

Метод сопряженных градиентов для произвольной задачи

- Обобщения работают неплохо, но гарантии в теории далеко не самые оптимистичные.
- Лучше иногда делать «рестарты». В данном случае «рестарты» предполагают иногда брать $\beta_k = 0$, забывая историю. **Вопрос:** итерация какого метода тогда получится? Градиентный спуск.
- Подходит как метод «стартер», которым из начальной неизвестной точки, можно близко, но не совсем точно подойти к решению.