ЛЕКЦИЯ 4
Глава 3. МЕТОДЫ ДЛЯ ЗАДАЧ БЕЗУСЛОВНОЙ ОПТИМИЗАЦИИ И НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Содержание лекции

1. Методы спуска
    - Общая схема методов спуска

Задача безусловной оптимизации

Задача

$$f(x) \to \min, \quad x \in \mathbb{R}^n, \tag{1}$$

где $f : \mathbb{R}^n \to \mathbb{R}$ — заданная функция.

Идея методов спуска

- Определяется направление, при движении по которому из текущего приближения функция f убывает.
- Делается шаг некоторой длины по этому направлению.
- Для полученного таким образом нового приближения процедура повторяется, и т.д.

Направления убывания

Определение 1

Вектор $d \in \mathbb{R}^n$ называется *направлением убывания* функции $f : \mathbb{R}^n \to \mathbb{R}$ в точке $\widetilde{x} \in \mathbb{R}^n$, если для любого достаточно малого $t > 0$ имеет место неравенство $f(\widetilde{x} + td) < f(\widetilde{x})$.

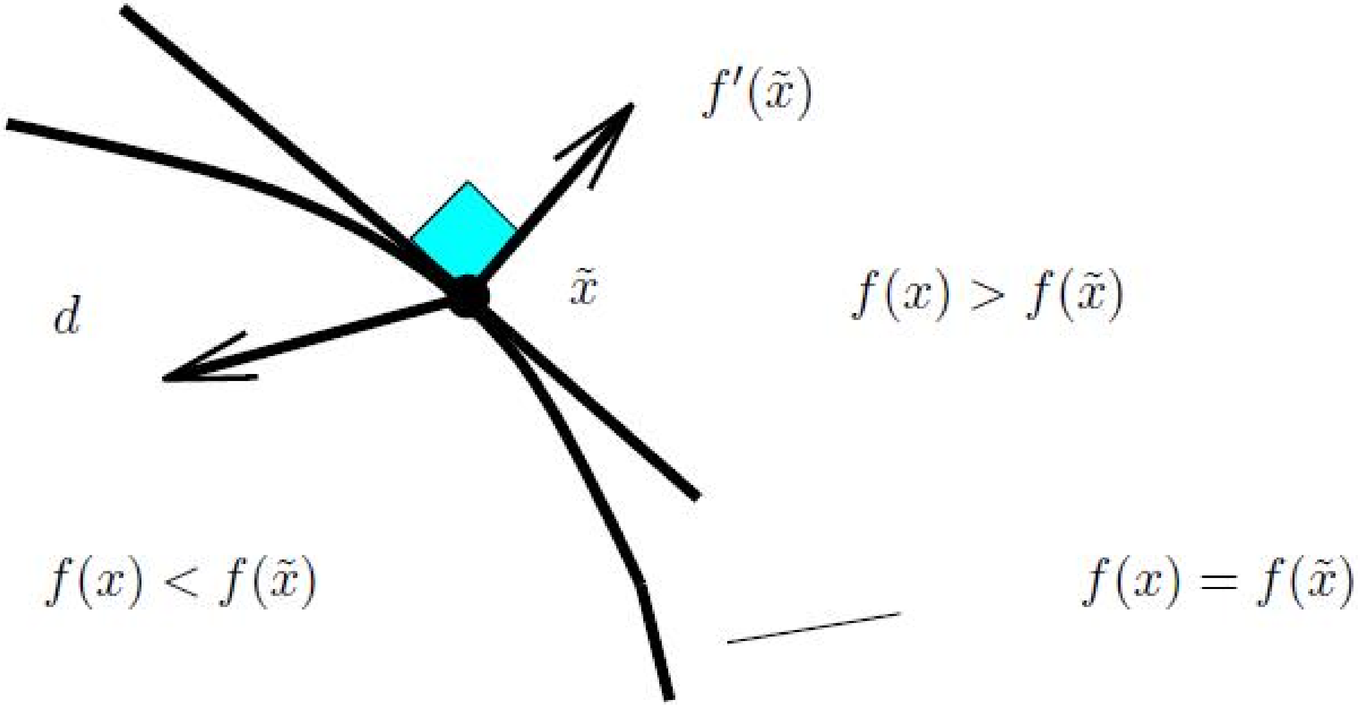
$\mathcal{D}_f(\widetilde{x})$ — множество направлений убывания функции f в точке $\widetilde{x} \in \mathbb{R}^n$ (всегда конус).

Лемма 1

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема в точке $\widetilde{x} \in \mathbb{R}^n$ по направлению $d \in \mathbb{R}^n$, т.е. $f'(\widetilde{x}; d)$ существует и конечна. Тогда:

- если $d \in \mathcal{D}_f(\widetilde{x})$, то $f'(\widetilde{x}; d) \leqslant 0$;
- если $d \in \mathbb{R}^n$ удовлетворяет условию $f'(\widetilde{x}; d) < 0$, то $d \in \mathcal{D}_f(\widetilde{x})$.

Направления убывания



Общая схема методов спуска

Методы спуска

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k \in \mathcal{D}_f(x^k), \quad k = 0, 1, \ldots, \qquad (2)$$

где *параметры длины шага* $\alpha_k > 0$ выбираются так, чтобы выполнялось по крайней мере

$$f(x^{k+1}) < f(x^k). \qquad (3)$$

Если $\mathcal{D}_f(x^k) = \varnothing$, или если не представляется возможным найти $d^k \in \mathcal{D}_f(x^k)$, то процесс останавливают.
Если $d^k \in \mathcal{D}_f(x^k)$, то (3) выполняется для любого достаточно малого $\alpha_k > 0$.

Общая схема методов спуска

Конкретный метод спуска характеризуется:

- способом выбора направлений убывания d^k (с использованием некоторой приближенной модели функции f);
- процедурой *одномерного поиска* для выбора параметров длины шага (с использованием сужения самой функции f на луч, исходящий из точки x^k в направлении вектора d^k).

Выбор направления убывания

Согласно лемме 1, если f дифференцируема в точке x^k и $f'(x^k) \neq 0$, то $d^k = -f'(x^k) \in \mathcal{D}_f(x^k)$. Обычно методы с таким выбором направления убывания крайне неэффективны: это «последнее средство», когда более изощренные способы выбора не срабатывают.

Значительно большее практическое значение имеет выбор $d^k = -Q_k f'(x^k)$, где симметричная матрица $Q_k \in \mathbb{R}(n, n)$ задается «умными» способами.
Согласно лемме 1, если матрица Q_k положительно определена и $f'(x^k) \neq 0$, то $d^k = -Q_k f'(x^k) \in \mathcal{D}_f(x^k)$, поскольку

$$\langle f'(x^k), d^k \rangle = -\langle Q_k f'(x^k), f'(x^k) \rangle < 0.$$

Процедуры одномерного поиска

Правило одномерной минимизации

Параметр $\alpha_k > 0$ выбирается из условия

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k),$$

т.е. как решение одномерной задачи оптимизации

$$\varphi_k(\alpha) \to \min, \quad \alpha \in \mathbb{R}_+, \tag{4}$$

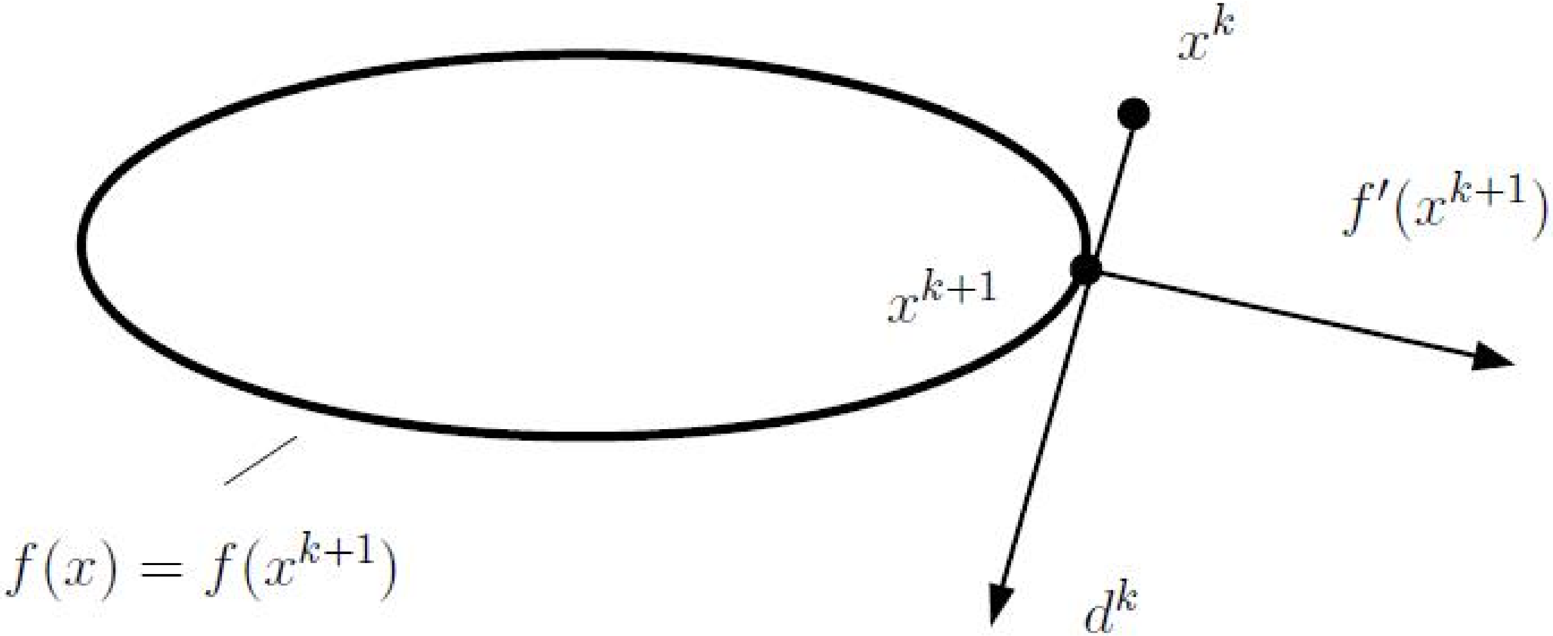
где

$$\varphi_k : \mathbb{R}_+ \to \mathbb{R}, \quad \varphi_k(\alpha) = f(x^k + \alpha d^k).$$

Если f дифференцируема в точке x^{k+1}, то

$$0 = \varphi_k'(\alpha_k) = \langle f'(x^k + \alpha_k d^k), d^k \rangle = \langle f'(x^{k+1}), d^k \rangle. \tag{5}$$

Процедуры одномерного поиска



Процедуры одномерного поиска

Для квадратичной функции

$$f(x) = \langle Bx, x\rangle + \langle a, x\rangle, \quad x \in \mathbb{R}^n,$$

где $B \in \mathbb{R}(n, n)$ — симметричная положительно определенная, $a \in \mathbb{R}^n$, одномерный минимум дается явной формулой:

$$\alpha_k = -\frac{\langle 2Bx^k + a, d^k\rangle}{2\langle Bd^k, d^k\rangle},$$

поскольку

$$\varphi_k(\alpha) = \langle Bd^k, d^k\rangle\alpha^2 + \langle 2Bx^k + a, d^k\rangle\alpha + \langle Bx^k, x^k\rangle + \langle a, x^k\rangle.$$

В общем же случае этот способ определения α_k неоправданно трудоемок.

Процедуры одномерного поиска

Часто задачу (4) заменяют задачей

$$\varphi_k(\alpha) \to \min, \quad \alpha \in [0, \widehat{\alpha}],$$

где $\widehat{\alpha} > 0$ — фиксированный параметр (выбор $\widehat{\alpha} = +\infty$ соответствует задаче (4)).

Процедуры одномерного поиска

Правило Армихо
(предполагает дифференцируемость f в точке x^k)

Фиксируем параметры $\widehat{\alpha} > 0$, ε, $\theta \in (0, 1)$. Полагаем $\alpha = \widehat{\alpha}$.

1. Проверяем неравенство

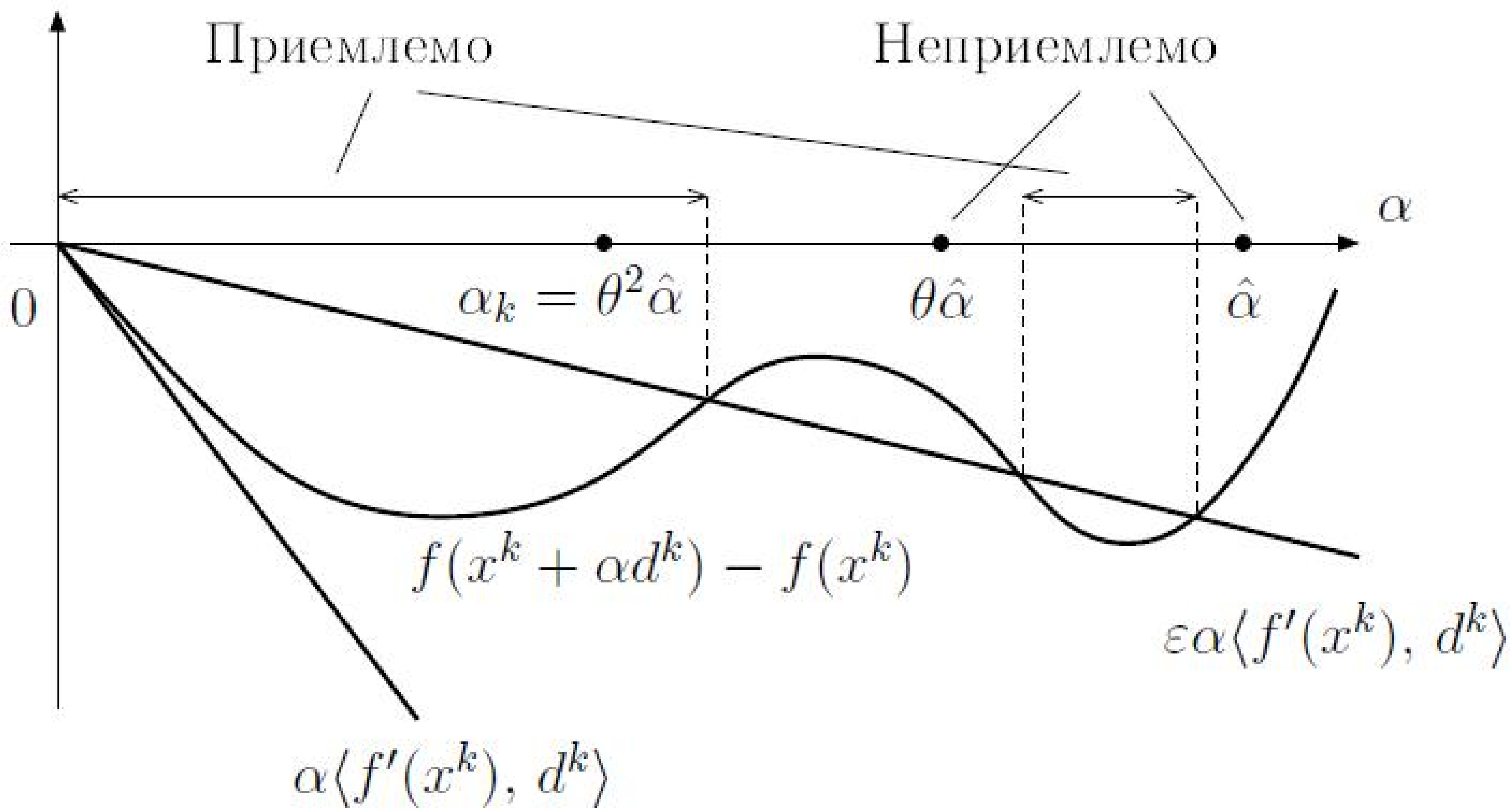
$$f(x^k + \alpha d^k) \leqslant f(x^k) + \varepsilon\alpha\langle f'(x^k), d^k\rangle. \tag{6}$$

2. Если (6) не выполнено, то заменяем α на $\theta\alpha$ и переходим к п. 1. Иначе полагаем $\alpha_k = \alpha$.

Процедуры одномерного поиска

Неравенство Армихо (6) означает, что реальное убывание значения целевой функции при шаге длины α по направлению d^k должно составлять как минимум заданную (определяемую выбором параметра ε) долю от «предсказанного» линейной моделью целевой функции убывания $\alpha\langle f'(x^k), d^k\rangle$. Параметр α_k вычисляется как первое из обладающих этим свойством чисел α, получаемых в результате дробления начального значения $\widehat{\alpha}$.

Процедуры одномерного поиска



Процедуры одномерного поиска

Лемма 2

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема в точке $x^k \in \mathbb{R}^n$. Тогда если элемент $d^k \in \mathbb{R}^n$ удовлетворяет

$$\langle f'(x^k), d^k \rangle < 0, \tag{7}$$

то неравенство Армихо (6) имеет место для любого достаточно малого $\alpha > 0$.

Доказательство

Для достаточно малого $\alpha > 0$

$$f(x^k + \alpha d^k) - f(x^k) = \langle f'(x^k), \alpha d^k \rangle + o(\alpha) =$$

$$= \varepsilon\alpha\langle f'(x^k), d^k \rangle + (1-\varepsilon)\alpha\langle f'(x^k), d^k \rangle + o(\alpha) =$$

$$= \varepsilon\alpha\langle f'(x^k), d^k \rangle + \alpha \left((1-\varepsilon)\langle f'(x^k), d^k \rangle + \frac{o(\alpha)}{\alpha} \right) \leqslant$$

$$\leqslant \varepsilon\alpha\langle f'(x^k), d^k \rangle$$

(поскольку второе слагаемое в предпоследней строке отрицательно), а это и есть требуемое.

Процедуры одномерного поиска

При выполнении (7) выбор α_k по правилу Армихо гарантирует выполнение условия монотонного убывания значения целевой функции (3).
Более того, неравенство (6) при $\alpha = \alpha_k$ дает количественную характеристику того, *насколько* $f(x^{k+1})$ должно быть меньше $f(x^k)$, чтобы можно было обосновать сходимость.
Доказательство сходимости существенно упрощается, когда удается установить, что количество дроблений для определения α_k конечно *равномерно по k*.

Процедуры одномерного поиска

Лемма 3

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема на $\mathbb{R}^n$, и ее производная липшицева на $\mathbb{R}^n$ с константой $\ell > 0$.
Тогда если для некоторых $x^k,\ d^k \in \mathbb{R}^n$ выполнено (7), то неравенство (6) имеет место для любого $\alpha \in (0,\ \bar{\alpha}_k]$, где

$$\bar{\alpha}_k = -\frac{2(1-\varepsilon)\langle f'(x^k),\ d^k\rangle}{\ell\|d^k\|^2} > 0. \tag{8}$$

Если удается оценить сверху константу Липшица ℓ, то формула (8) может использоваться для явного вычисления подходящих параметров длины шага.

Процедуры одномерного поиска

Доказательство

По формуле Ньютона–Лейбница для любого $\alpha > 0$

$$f(x^k + \alpha d^k) - f(x^k) - \langle f'(x^k), \alpha d^k \rangle =$$

$$= \int_0^1 \langle f'(x^k + t\alpha d^k) - f'(x^k), \alpha d^k \rangle \, dt \leqslant$$

$$\leqslant \alpha \| d^k \| \int_0^1 \| f'(x^k + t\alpha d^k) - f'(x^k) \| \, dt \leqslant$$

$$\leqslant \ell \alpha^2 \| d^k \|^2 \int_0^1 t \, dt = \frac{\ell}{2} \alpha^2 \| d^k \|^2.$$

Процедуры одномерного поиска

Доказательство (завершение)

Тогда для любого $\alpha \in (0, \bar{\alpha}_k]$

$$f(x^k + \alpha d^k) - f(x^k) \leqslant \langle f'(x^k), \alpha d^k \rangle + \frac{\ell}{2}\alpha^2 \|d^k\|^2 =$$

$$= \alpha \left(\langle f'(x^k), d^k \rangle + \frac{\ell}{2}\alpha \|d^k\|^2 \right) \leqslant \varepsilon\alpha \langle f'(x^k), d^k \rangle,$$

а это и есть требуемое.

Процедуры одномерного поиска

Если в предположениях этой леммы

$$\frac{\langle f'(x^k),\, d^k \rangle}{\|d^k\|^2} \leqslant \delta \tag{9}$$

при некотором $\delta < 0$, не зависящем от k, то

$$\bar{\alpha}_k \geqslant -\frac{2}{\ell}(1 - \varepsilon)\delta,$$

а значит,

$$\alpha_k \geqslant \check{\alpha}$$

при некотором $\check{\alpha} > 0$, не зависящем от k.

Если $d^k = -f'(x^k)$, то (9) выполняется автоматически при $\delta = -1$.

Процедуры одномерного поиска

Правило постоянного параметра

Фиксируем не зависящее от k число $\bar{\alpha} > 0$, и полагаем $\alpha_k = \bar{\alpha}$.

Это правило не может быть эффективным (отсутствует адаптация длины шага) и применяется только в тех случаях, когда вычисление значений целевой функции очень трудоемко.

Теоретический анализ методов с *достаточно малым* постоянным параметром длины шага сводится к анализу методов, использующих правило Армихо.

Процедуры одномерного поиска

Существуют более изощренные практические правила выбора параметра длины шага.

Правило Голдстейна

Параметр $\alpha_k > 0$ выбирается из условия

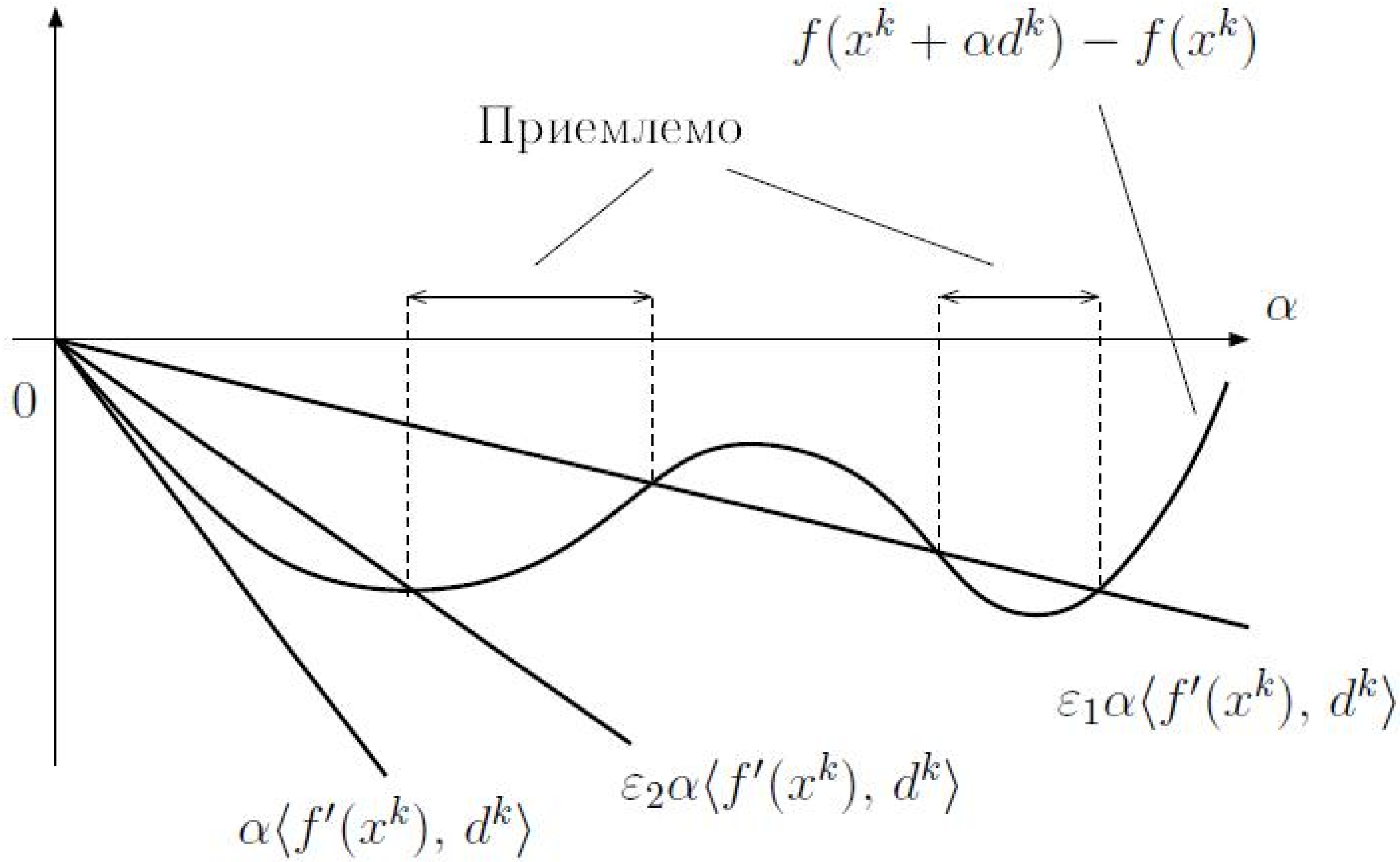
$$\varepsilon_1 \leqslant \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha \langle f'(x^k), d^k \rangle} \leqslant \varepsilon_2 \tag{10}$$

при фиксированных $\varepsilon_1, \varepsilon_2 \in (0, 1)$, $\varepsilon_1 < \varepsilon_2$.

Левое неравенство — неравенство Армихо (6) при $\varepsilon = \varepsilon_1$; оно обеспечивает достаточное убывание значения целевой функции.

Правое неравенство препятствует выбору слишком малых параметров длины шага.

Процедуры одномерного поиска



Процедуры одномерного поиска

Другая реализация той же идеи:

Правило Вулфа

Параметр $\alpha_k > 0$ выбирается из условий

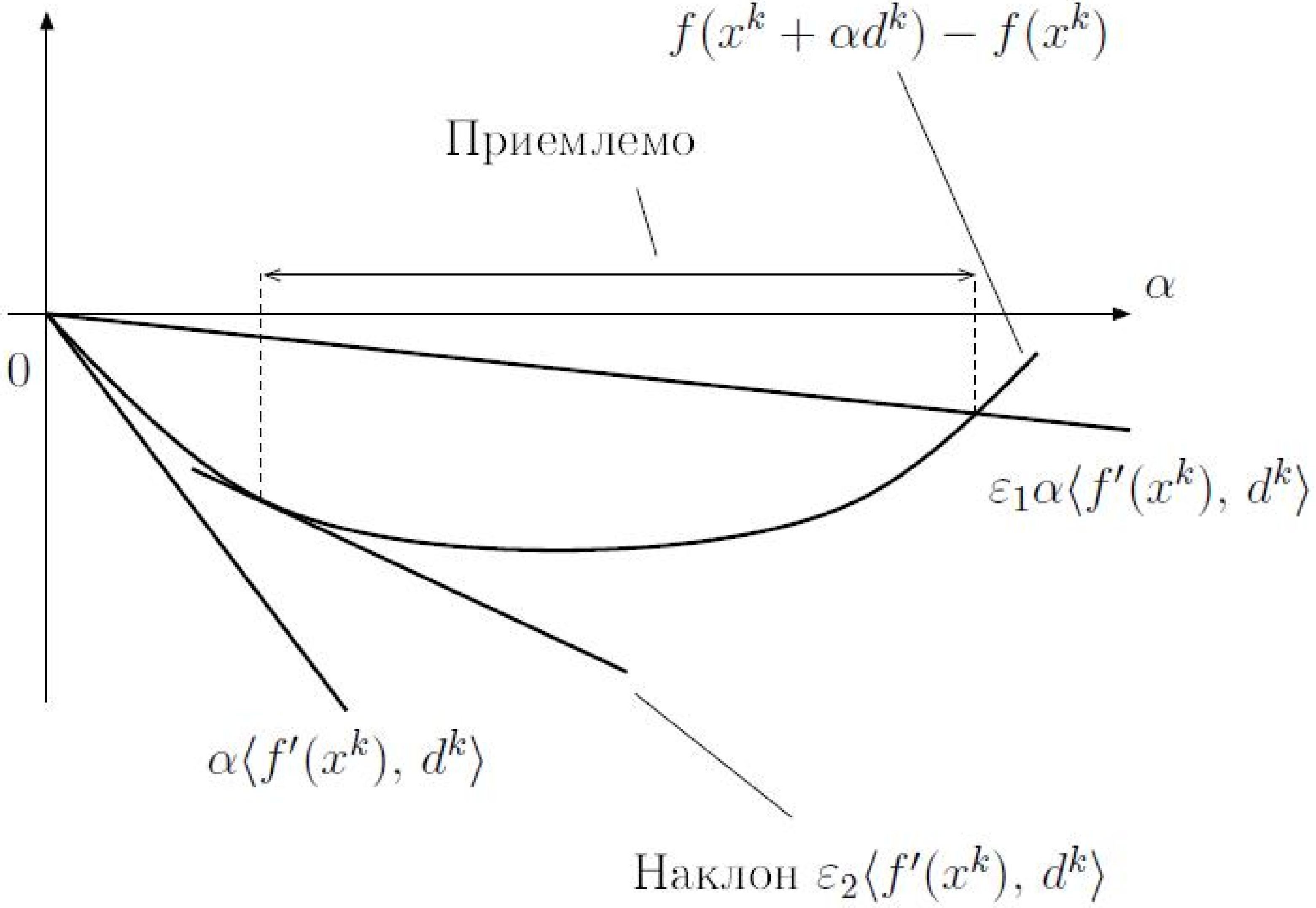
$$f(x^k + \alpha d^k) \leqslant f(x^k) + \varepsilon_1 \alpha \langle f'(x^k), d^k \rangle, \tag{11}$$

$$\langle f'(x^k + \alpha d^k), d^k \rangle \geqslant \varepsilon_2 \langle f'(x^k), d^k \rangle. \tag{12}$$

Левая часть (12) равна $\varphi'_k(\alpha)$, поэтому при выполнении (7) условие (12) означает, что величина $\varphi'_k(\alpha)$ не должна быть «слишком отрицательной».

Любой локальный минимум φ_k на $\mathbb{R}_+$ удовлетворяет (12).

Процедуры одномерного поиска



Процедуры одномерного поиска

Если вычисление градиента функции f не слишком трудоемко, правило Вулфа признается наиболее эффективным известным правилом одномерного поиска.
Важное свойство этого правила связано с квазиньютоновскими методами.

Ключ к реализации: нарушение (11) говорит о том, что текущее пробное значение α нужно уменьшать, а нарушение (12) — что увеличивать.

Процедуры одномерного поиска

Реализация правила Вулфа

Фиксируем параметры ε_1, $\varepsilon_2 \in (0, 1)$, $\varepsilon_1 < \varepsilon_2$. Полагаем $\check{\alpha} = \widehat{\alpha} = 0$. Выбираем начальное пробное значение $\alpha > 0$.

1. Проверяем выполнение неравенств (11) и (12). Если оба они выполнены, то переходим к п. 6.
2. Если нарушено (11), то полагаем $\widehat{\alpha} = \alpha$ и переходим к п. 5.
3. Если нарушено (12), то полагаем $\check{\alpha} = \alpha$.
4. Если $\widehat{\alpha} = 0$, то выбираем новое пробное значение $\alpha > \check{\alpha}$ («экстраполяция») и переходим к п. 1.
5. Выбираем новое пробное значение $\alpha \in (\check{\alpha}, \widehat{\alpha})$ («интерполяция») и переходим к п. 1.
6. Полагаем $\alpha_k = \alpha$.

Процедуры одномерного поиска

Сначала реализуются шаги «экстраполяции», пока $\widehat{\alpha}$ не станет положительным.
Затем выполняются шаги «интерполяции»; при этом $\widehat{\alpha}$ может только уменьшаться, оставаясь положительным, а $\check{\alpha}$ — только увеличиваться, оставаясь меньше $\widehat{\alpha}$.

В отличие от правила Армихо, здесь *начальное пробное значение α может увеличиваться.*

Возможная реализация «экстраполяции» и «интерполяции»

Фиксируем параметры $\theta_1 > 1$, $\theta_2 \in (0, 1)$.
При «экстраполяции» заменяем α на $\theta_1\alpha$, а при «интерполяции» полагаем $\alpha = (1 - \theta_2)\check{\alpha} + \theta_2\widehat{\alpha}$.

Процедуры одномерного поиска

Лемма 4

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ непрерывно дифференцируема и ограничена снизу на $\mathbb{R}^n$.
Тогда если для некоторых x^k, $d^k \in \mathbb{R}^n$ выполнено (7), то реализующая правило Вулфа процедура, в которой $\check{\alpha} \to +\infty$ в случае бесконечного числа шагов «экстраполяции», и $(\widehat{\alpha} - \check{\alpha}) \to 0$ в случае бесконечного числа шагов «интерполяции», будет конечной.

Методы спуска

В последнее время часто используются на практике

Методы с немонотонным одномерным поиском

При выборе α_k значение $f(x^k + \alpha_k d^k)$ сравнивается не с $f(x^k)$, а с максимальным (либо средним) значением функции f за некоторое фиксированное число предшествующих итераций.

Позволяют делать более длинные шаги, допуская даже увеличение значения целевой функции на некоторых итерациях.