ЛЕКЦИЯ 7
Глава 4. МЕТОДЫ ДЛЯ ЗАДАЧ УСЛОВНОЙ ОПТИМИЗАЦИИ И КОМПЛЕМЕНТАРНЫХ ЗАДАЧ

Содержание лекции

1. Методы для задач оптимизации с простыми ограничениями
   - Методы проекции градиента
   - Возможные направления и методы спуска
   - Методы условного градиента. Условные методы Ньютона

Задача оптимизации с прямым ограничением

Задача

$$f(x) \to \min, \quad x \in P, \tag{1}$$

где

- $P \subset \mathbb{R}^n$ — множество «простой структуры» (замкнутое и выпуклое; проекция $\pi_P(x)$ на него любой точки $x \in \mathbb{R}^n$ существует и единственна);
- $f : \mathbb{R}^n \to \mathbb{R}$ — дифференцируемая функция.

Точка $\bar{x} \in P$ является стационарной в задаче (1), если

$$\langle f'(\bar{x}), x - \bar{x} \rangle \geqslant 0 \quad \forall x \in P,$$

или, эквивалентно, для некоторого (для любого) $t > 0$

$$\pi_P(\bar{x} - tf'(\bar{x})) = \bar{x}.$$

Методы проекции градиента

Объединяя идею градиентных методов безусловной оптимизации с проектированием генерируемых приближений на допустимое множество условной задачи, получаем

Методы проекции градиента

$$x^0 \in P, \quad x^{k+1} = \pi_P(x^k - \alpha_k f'(x^k)), \quad k = 0, 1, \ldots \tag{2}$$

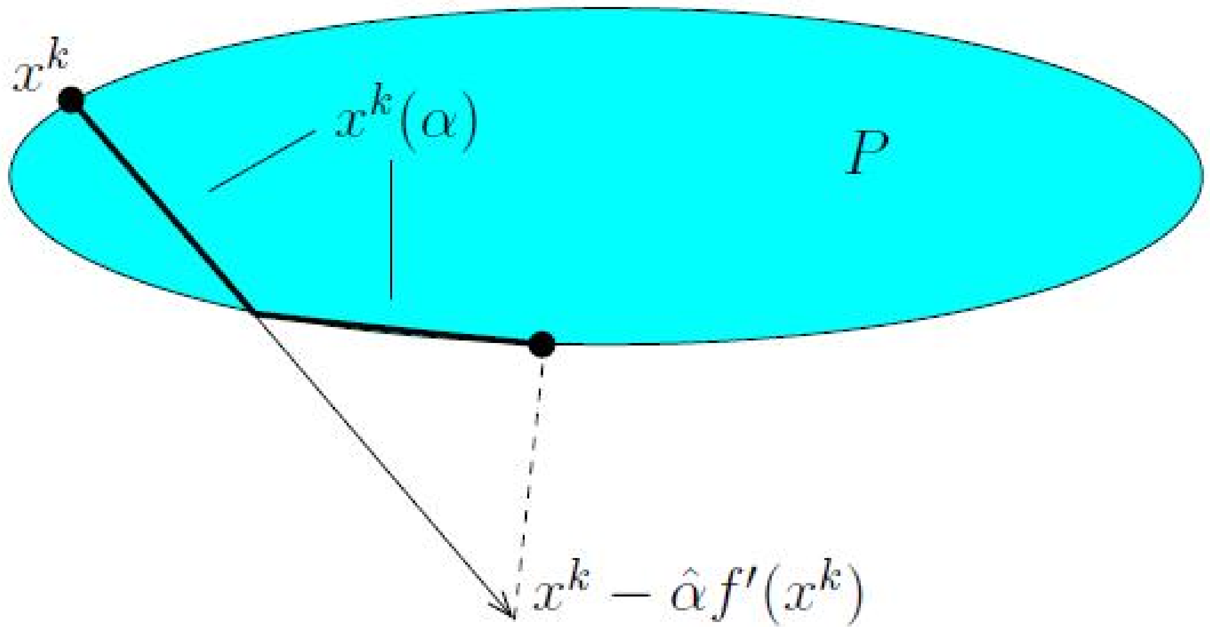
Процедуры одномерного поиска для выбора параметров длины шага $\alpha_k > 0$ используют функцию

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

где

$$x^k(\alpha) = \pi_P(x^k - \alpha f'(x^k)).$$

Методы проекции градиента



Процедуры одномерного поиска

Правило одномерной минимизации

Параметр $\alpha_k > 0$ выбирается как решение одномерной задачи оптимизации

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

либо

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

где $\hat{\alpha} > 0$ — параметр.

Процедуры одномерного поиска

Правило Армихо

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

1. Проверяем неравенство
$$f(x^k(\alpha)) \leqslant f(x^k) + \varepsilon \langle f'(x^k), x^k(\alpha) - x^k \rangle.$$
2. Если неравенство не выполнено, то заменяем α на $\theta\alpha$ и переходим к п. 1. Иначе полагаем $\alpha_k = \alpha$.

Применяют также различные модификации правила Армихо (например, в духе правила Голдстейна).

Процедуры одномерного поиска

Правило постоянного параметра

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

Методы проекции градиента

Алгоритм 1

Выбираем $x^0 \in P$ и полагаем $k = 0$. Выбираем одно из трех правил одномерного поиска и необходимые для реализации этого правила параметры.

1. Вычисляем α_k в соответствии с выбранным правилом одномерного поиска.
2. Вычисляем x^{k+1} по формуле (2).
3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Методы проекции градиента

Если для некоторого k точка x^k оказывается стационарной в задаче (1), то $x^k = x^{k+1} = \ldots$, как бы ни выбирался параметр длины шага α_k:

$$\pi_P(x^k - tf'(x^k)) = x^k \quad \forall\, t > 0.$$

На практике алгоритм в этом случае останавливают (должно подразумеваться практическими правилами остановки).

Методы проекции градиента

Если $P = \mathbb{R}^n$, методы проекции градиента с указанными выше способами выбора α_k превращаются в соответствующие градиентные методы.
Теоретические свойства методов проекции градиента во многом аналогичны свойствам градиентных методов.

Глобальная сходимость

Теорема 1

Пусть множество $P \subset \mathbb{R}^n$ замкнуто и выпукло, а функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема на $\mathbb{R}^n$, и ее производная липшицева на $\mathbb{R}^n$ с константой $\ell > 0$. Пусть в случае использования в алгоритме 1 правила постоянного параметра он удовлетворяет условию $\bar{\alpha} < 2/\ell$.

Тогда любая предельная точка любой траектории $\{x^k\}$ алгоритма 1 является стационарной точкой задачи (1).

Если предельная точка существует, или если функция f ограничена снизу на P, то для любой ограниченной последовательности $\{t_k\} \subset \mathbb{R}_+$

$$\|\pi_P(x^k - t_k f'(x^k)) - x^k\| \to 0 \quad (k \to \infty).$$

Глобальная сходимость

Последнее утверждение означает сходимость по невязке необходимого условия первого порядка оптимальности.

Если, дополнительно, множество Лебега $L_{f,\,P}(f(x^0))$ ограничено, то $\{x^k\}$ имеет предельные точки, и

$$\operatorname{dist}(x^k,\, S_0 \cap L_{f,\,P}(f(x^0))) \to 0 \quad (k \to \infty),$$

где S_0 — множество стационарных точек задачи (1).

Скорость сходимости

Предположения

- Оценка расстояния до S_0:

$$\|\pi_P(x - f'(x)) - x\| \geqslant \gamma \operatorname{dist}(x, S_0) \quad \forall x \in U, \tag{3}$$

  где

$$U = \{x \in \mathbb{R}^n \mid \|\pi_P(x - f'(x)) - x\| \leqslant \delta\}, \tag{4}$$

  для некоторых $\delta > 0$ и $\gamma > 0$.
- Условие отделимости критических поверхностей уровня — в терминах стационарности для задачи (1).

Оба условия выполняются автоматически, если, например, (1) — задача квадратичного программирования, и $S_0 \neq \varnothing$.

Скорость сходимости

Теорема 2

Пусть в дополнение к условиям теоремы 1 значение задачи (1) конечно, и выполнено (3) при U заданном в (4) и некоторых $\delta > 0$ и $\gamma > 0$. Пусть, кроме того, выполнено условие отделимости критических поверхностей уровня. Пусть в алгоритме 1 используется правило Армихо или правило постоянного параметра.

Тогда любая траектория $\{x^k\}$ алгоритма 1 сходится к некоторой стационарной точке $\bar{x}$ задачи (1).

Скорость сходимости последовательности $\{f(x^k)\}$ линейная, а по аргументу геометрическая.

Методы проекции градиента

Поскольку каждый одномерный поиск может требовать многократного вычисления проекции, методы проекции градиента могут представлять практический интерес лишь если вычисление проекции на P не связано с большими вычислительными затратами; в данном случае именно это имеется в виду под тем, что P должно быть множеством «простой структуры».
Например, если P полиэдр, то вычисление проекции — задача квадратичного программирования. Если P обобщенный параллелепипед, то проекция вычисляется по явной формуле.

Методы проекции градиента

Модификация метода проекции градиента

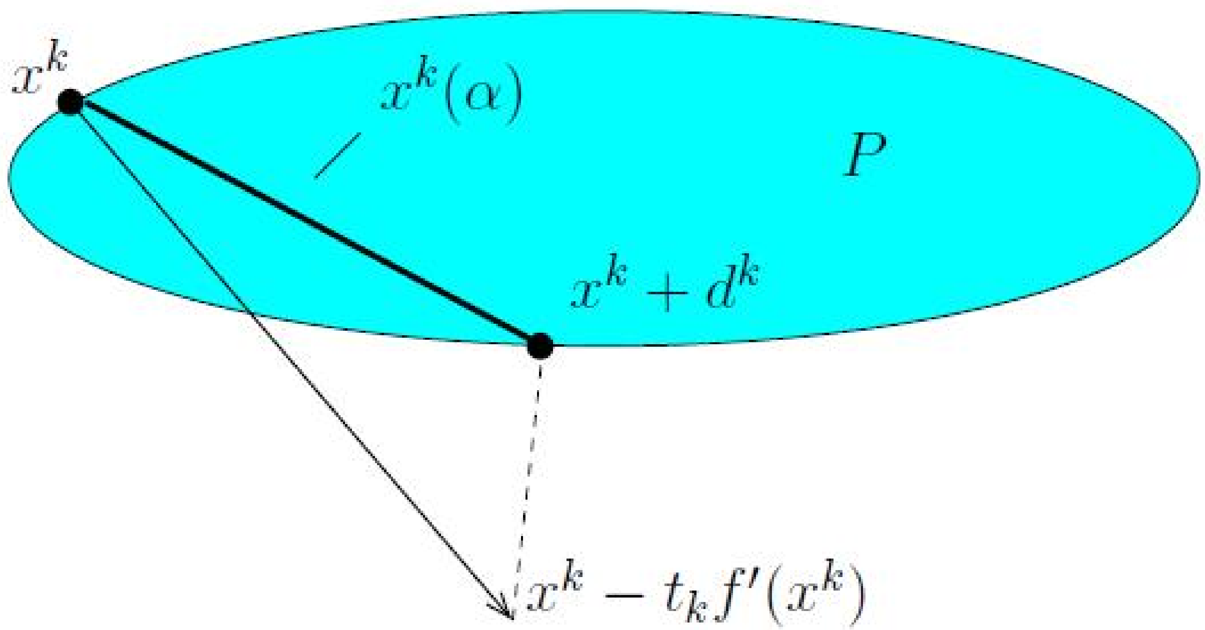
$$x^{k+1} = (1-\alpha_k)x^k + \alpha_k \pi_P(x^k - t_k f'(x^k)), \quad k = 0, 1, \ldots,$$

где $t_k > 0$, а α_k определяется согласно правилу одномерной минимизации, либо правилу Армихо при $\hat{\alpha} = 1$.
В правилах одномерного поиска нужно положить

$$x^k(\alpha) = (1-\alpha)x^k + \alpha \pi_P(x^k - t_k f'(x^k)).$$

Если последовательность $\{t_k\}$ отделена от нуля, то есть глобальная сходимость в смысле теоремы 1. *Линейной скорости сходимости нет.*

Методы проекции градиента



Методы проекции градиента

Теоретическое значение методов проекции градиента: многие другие методы с допустимыми траекториями и монотонно невозрастающими значениями целевой функции могут интерпретироваться как возмущения методов проекции градиента.

Возможные направления

Идея: методы спуска с сохранением допустимости генерируемых приближений.

Определение 1

Вектор $d \in \mathbb{R}^n$ называется *возможным направлением* относительно множества $D \subset \mathbb{R}^n$ в точке $\tilde{x} \in D$, если для любого достаточно малого $t > 0$ выполняется $\tilde{x} + td \in D$.

$\mathcal{F}_D(\tilde{x})$ — множество возможных относительно множества D в точке $\tilde{x} \in D$ направлений (всегда конус).

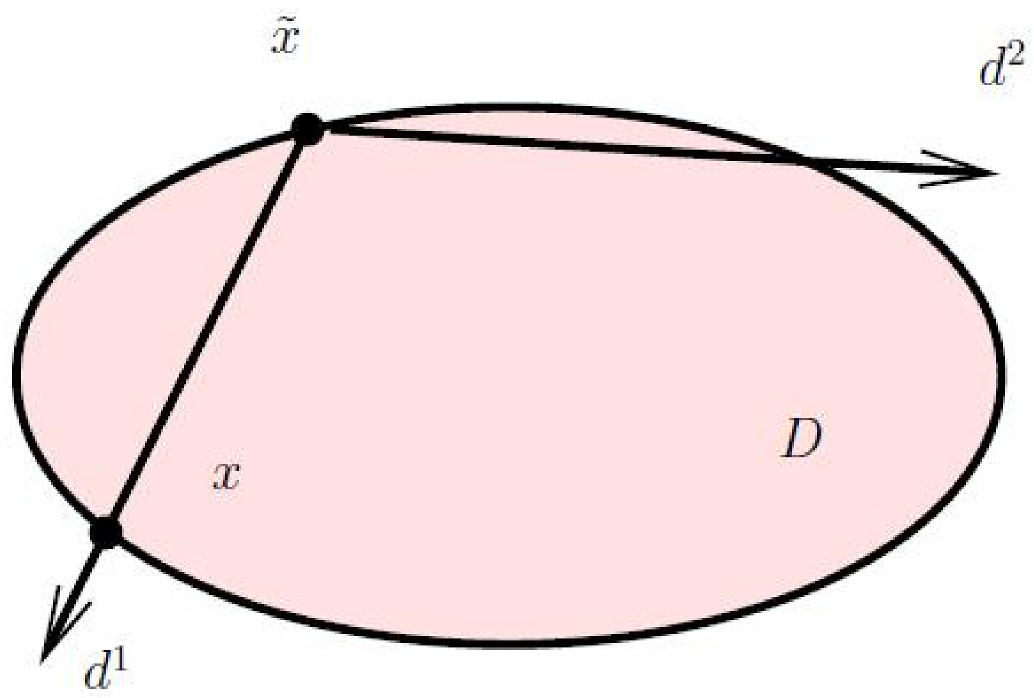
Возможные направления

Лемма 1

Пусть множество $D \subset \mathbb{R}^n$ выпукло.

Тогда

$$x - \tilde{x} \in \mathcal{F}_D(\tilde{x}) \quad \forall\, \tilde{x},\, x \in D.$$



Возможные направления

Лемма 2

Пусть

$$D = \{x \in \mathbb{R}^n \mid G(x) \leqslant 0\},$$

где отображение $G : \mathbb{R}^n \to \mathbb{R}^m$ дифференцируемо в точке $\tilde{x} \in D$.

Тогда:

- для любого $d \in \mathcal{F}_D(\tilde{x})$ выполнено $\langle g_i'(\tilde{x}), d \rangle \leqslant 0 \ \forall\, i \in A(\tilde{x})$;
- если $d \in \mathbb{R}^n$ удовлетворяет условию $\langle g_i'(\tilde{x}), d \rangle < 0 \ \forall\, i \in A(\tilde{x})$, то $d \in \mathcal{F}_D(\tilde{x})$.

Общая задача оптимизации

Задача

$$f(x) \to \min, \quad x \in D,$$

где, вообще говоря, $D \subset \mathbb{R}^n$ не предполагается ни замкнутым, ни выпуклым, а функция $f : \mathbb{R}^n \to \mathbb{R}$ гладкой.

Методы спуска

Методы спуска

$$x^0 \in D,$$

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k), \quad k = 0, 1, \ldots, \quad (5)$$

где параметры длины шага $\alpha_k > 0$ выбираются так, чтобы выполнялось

$$f(x^{k+1}) < f(x^k), \quad x^{k+1} \in D. \quad (6)$$

Если $\mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k) = \varnothing$, или если не представляется возможным найти $d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k)$, то процесс останавливают.

Если $d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k)$, то (6) выполняется для любого достаточно малого $\alpha_k > 0$.

Методы спуска

Конкретный метод спуска характеризуется

- способом выбора возможных направлений убывания;
- процедурой одномерного поиска для выбора параметров длины шага.

Если $D = \mathbb{R}^n$, то (5), (6) совпадает со схемой методов спуска для задач безусловной оптимизации.

Методы спуска

Если $D = P$ — замкнуто и выпукло, то второе условие в (6) (условие допустимости) выполнено $\forall \alpha_k \in [0, \hat{\alpha}_k]$, где

$$\hat{\alpha}_k = \sup_{\alpha \geqslant 0:\, x^k + \alpha d^k \in P} \alpha > 0. \tag{7}$$

При этом процедуры одномерного поиска — те же, что для задач безусловной оптимизации, но с дополнительным ограничением $\alpha_k \leqslant \hat{\alpha}_k$.
Все упрощается, если $\hat{\alpha}_k$ можно явно вычислить (например, если P — полиэдр).

Методы спуска

В случае выпуклого и замкнутого допустимого множества модифицированный метод проекции градиента — метод спуска.

Далее: еще два примера реализации методов спуска для условных задач с простыми ограничениями, которые приведут к идее ньютоновских методов для задач условной оптимизации.

Методы условного градиента

Пусть $D = P$ — замкнуто, выпукло и ограничено.

Методы условного градиента

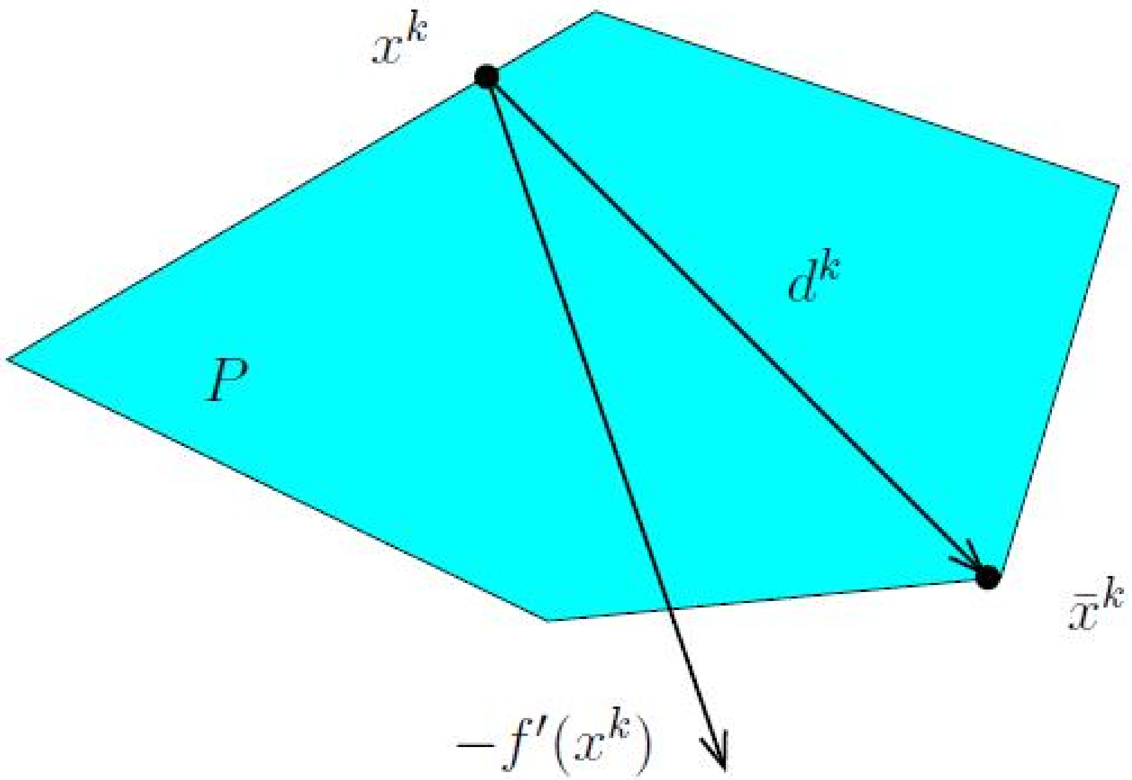
Это схема методов спуска (5) при $d^k = \bar{x}^k - x^k$, где для текущего $x^k \in P$ точка $\bar{x}^k$ — решение задачи

$$\langle f'(x^k), x - x^k \rangle \to \min, \quad x \in P, \tag{8}$$

с линеаризованной целевой функцией (решение задачи (8) существует в силу теоремы Вейерштрасса).

Такой d^k — *условный антиградиент* функции f в точке x^k относительно множества P.

Методы условного градиента



Методы условного градиента

Поскольку $x^k \in P$, значение задачи (8)

$$v_k = \langle f'(x^k), \bar{x}^k - x^k \rangle \leqslant \langle f'(x^k), x^k - x^k \rangle = 0.$$

Если $v_k = 0$, то

$$\langle f'(x^k), x - x^k \rangle \geqslant v_k = 0 \quad \forall x \in P,$$

т.е. x^k — стационарная точка в задаче (1).
Если же $v_k < 0$, то $d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_P(x^k)$ (в силу лемм 3.1.1 и 1).

Для таких d^k выполняется $\hat{\alpha}_k = 1$, что делает возможной простую реализацию процедур одномерного поиска.

Методы условного градиента

Методы условного градиента имеют практический смысл лишь если вспомогательная задача (8) (с линейной целевой функцией) легко решается (например, если P полиэдр, то (8) — задача линейного программирования).

Есть глобальная сходимость и (для f выпуклой на P) неулучшаемая арифметическая оценка скорости сходимости метода.
В целом, такие методы крайне неэффективны.

Методы условного градиента

Теоретическое значение методов условного градиента: это простейшая реализация фундаментальной идеи линеаризации (принимающая окончательную форму в методах возможных направлений, где линеаризуется не только целевая функция, но и ограничения; не рассматриваем).

Развитие идеи линеаризации — использование более точной квадратичной аппроксимации целевой функции.

Условные методы Ньютона

Условные методы Ньютона

Это схема (5) при $d^k = \bar{x}^k - x^k$, где для текущего $x^k \in P$ точка $\bar{x}^k$ — решение задачи

$$\langle f'(x^k), x - x^k\rangle + \frac{1}{2}\langle f''(x^k)(x - x^k), x - x^k\rangle \to \min, \quad x \in P. \quad (9)$$

Условные методы Ньютона имеют практический смысл лишь если вспомогательная задача (9) (с квадратичной целевой функцией) существенно проще исходной задачи (1). Например, если P полиэдр, то (9) — задача квадратичного программирования.

Условные методы Ньютона

Если множество P задается функциональными ограничениями, то их можно линеаризовать в текущей точке x^k.

Однако, при этом значительно эффективнее использовать другой квадратичный член во вспомогательной задаче: для глобальной сходимости он может задаваться любой положительно определенной матрицей, а для высокой скорости сходимости он должен отражать «кривизну» не только целевой функции, но и ограничений (методы последовательного квадратичного программирования в разд. 4.3, 5.3).