[vcocP2.tex, 19.03.21]

4. Дифференцирование отображений в линейных нормированных пространствах

Если нужно вспомнить, что такое линейное пространство, нормированное пространство, линейный оператор, ограниченный оператор, непрерывный оператор ..., то можно заглянуть в книги [1] или [2]. Этот параграф написан в основном по книге [3].

4.1. Производные отображений

Пусть X, Y – линейные нормированные пространства, $x^0 \in X$, U – окрестность x^0 в X, отображение $f\,:\,U \to Y,\quad U \subset D(f)$.

1. Пусть $h \in X$. Если существует

$$\lim_{\lambda\downarrow 0} \frac{f(x^0+\lambda\, h)-f(x^0)}{\lambda},$$

то он называется *производной отображения f на элементе x^0 по направлению h* и обозначается $f'(x^0;h)$ (здесь $f'(x^0;h)$ – элемент пространства Y).

2. Если $\forall h \in X$ существует $f'(x^0;h)$, то отображение $\delta f(x^0;\cdot)\,:\,h\in X \mapsto f'(x^0;h)\in Y$ называется *первой вариацией отображения f на элементе x^0*, а при $f'(x^0;-h) = -f'(x^0;h)$ – первой вариацией по Лагранжу.

3. Если A – такое линейное непрерывное отображение (оператор), что

$$f(x^0+\lambda h) = f(x^0) + \lambda A h + o(\lambda) \ \text{ при } \ \lambda \to 0 \ \ \forall h \in X,$$

то A – *производная Гато* (или *слабая производная*) *отображения f на элементе x^0*; обозначается $f'_G(x^0)$.

Это определение равносильно тому, что первая вариация $\delta f(x^0;\cdot)$ является линейным и непрерывным отображением.

4. Если $A\,:\,X \to Y$ – такое линейное непрерывное отображение, что

$$f(x^0+h) = f(x^0) + Ah + \alpha(h)\,||h|| \quad \text{где} \quad \alpha(h)\to 0 \quad \text{при} \quad ||h||\to 0,$$

то A – *производная Фреше* (или *сильная производная*) *отображения f на элементе x^0*; обозначается $f'(x^0)$.

Отображение дифференцируемо по Фреше тогда и только тогда, когда

$$\forall\varepsilon>0 \ \ \exists\delta(\varepsilon)>0 \ \mid \ ||f(x^0+h)-f(x^0)-f'(x^0)h|| < \varepsilon||h|| \ \text{ при } \ ||h||<\delta.$$

5. Отображение $f:X\to Y$ называется *строго дифференцируемым* на элементе x^0, если найдется такое линейное непрерывное отображение $A\,:\,X\to Y$, что

$$\forall\varepsilon>0 \ \ \exists\delta(\varepsilon)>0 \ \mid \ ||f(x')-f(x'')-A(x'-x'')|| < \varepsilon||x'-x''|| \ \text{ при } \ ||x'-x^0||<\delta, \ \ ||x''-x^0||<\delta.$$

Ясно, что A – производная Фреше.

Если отображение дифференцируемо по Фреше на элементе x^0, то оно непрерывно на этом элементе.

Если отображение строго дифференцируемо на элементе x^0, то оно непрерывно в окрестности этого элемента.

Эти утверждения следуют непосредственно из определений.

Каждое определение дифференцируемости является более сильным, чем предыдущее. Сформулируем это так.

Теорема 1.
5. $\Rightarrow$ 4. $\Rightarrow$ 3. $\Rightarrow$ 2. $\Rightarrow$ 1., *но не наоборот.*

Пример 1. Пусть $f(x) = b \ \forall x \in X$. Тогда $f'(x^0) = ?$

Пример 2. Пусть A – линейный непрерывный оператор. У аффинного оператора $Ax + b$ производная Гато совпадает с A. Как частный случай имеем: производная линейного непрерывного оператора совпадает с ним самим.

Пример 3. Пусть $A : X \to Y$ – линейный непрерывный оператор, X – банахово пространство, Y – гильбертово пространство с естественной нормой $||y||^2 = (y, y)$. Вычислим производную функционала $f(x) = ||Ax - b||^2$:

$$f(x^0 + h) - f(x^0) = (Ax^0 + Ah - b, Ax^0 + Ah - b) - (Ax^0 - b, Ax^0 - b) =$$

$$= 2(Ax^0 - b, Ah) + (Ah, Ah) = 2(A^*(Ax^0 - b), h) + ||Ah||^2 \ldots$$

Пример 4. Дифференцируемо ли отображение

$$f[x(\cdot)] = \int\limits_c^d \Big(\int\limits_a^b x(\tau)k(\tau, t)\, d\tau - f(t)\Big)^2 dt$$

в пространствах C и L_2?

4.2. Правила дифференцирования

Легко проверить, что сумма отображений дифференцируема, если дифференцируемо каждое слагаемое. Скалярный множитель можно выносить за знак производной.

Теорема 2. (о суперпозиции отображений)
Пусть X, Y, Z – нормированные пространства, U – окрестность x^0 в X, V – окрестность y^0 в Y, $\varphi : U \to V$, $\varphi(x^0) = y^0$, $\psi : V \to Z$, $f = \psi \circ \varphi : U \to Z$ – суперпозиция отображений. Если ψ дифференцируемо по Фреше на y^0, а φ дифференцируемо на x^0 в смысле 1. – 4., то и f дифференцируемо в смысле 1. – 4. При этом

$$f'(x^0) = \psi'(y^0) \circ \varphi'(x^0) \quad \text{или} \quad f'(x^0; h) = \psi'(y^0)(\varphi'(x^0; h)).$$

Если отображения φ и ψ строго дифференцируемы, то и отображение f строго дифференцируемо.

Доказательство. Пусть отображение ψ дифференцируемо по Фреше. Тогда

$$\psi(y) = \psi(y^0) + \psi'(y^0)(y - y^0) + \alpha(y)||y - y^0||, \quad \lim_{y \to y^0} \alpha(y) = 0.$$

Если существует

$$\varphi'(x^0; h) = \lim_{\lambda \downarrow 0} \frac{\varphi(x^0 + \lambda h) - \varphi(x^0)}{\lambda},$$

то

$$\lim_{\lambda \downarrow 0} \frac{f(x^0 + \lambda h) - f(x^0)}{\lambda} = \lim_{\lambda \downarrow 0} \frac{\psi(\varphi(x^0 + \lambda h)) - \psi(y^0)}{\lambda} =$$

$$= \lim_{\lambda \downarrow 0} \frac{\psi'(y^0)(\varphi(x^0 + \lambda h) - y^0) + \alpha(\varphi(x^0 + \lambda h))\, ||\varphi(x^0 + \lambda h) - y^0||}{\lambda} = \psi'(y^0)\, \varphi'(x^0; h).$$

В курсе МА была доказана теорема о среднем (формула конечных приращений Лагранжа): если функция $f : [a, b] \to R^1$ непрерывна на $[a, b]$ и дифференцируема в (a, b), то найдется $c \in (a, b) \mid f(b) - f(a) = f'(c)(b - a)$.

Аналогичное утверждение можно доказать и для более общих отображений.

Отрезок в линейном пространстве определяется так: $[a, b] = \{x \mid x = a + t(b - a), \ 0 \le t \le 1\}$.

Теорема 3.
Пусть X, Y – нормированные пространства и открытое множество $U \subset X$ содержит отрезок $[a, b]$. Если отображение $f : U \to Y$ дифференцируемо по Гато на каждом элементе $x \in [a, b]$, то

$$||f(b) - f(a)|| \le \sup_{c \in (a, b)} ||f'_G(c)|| \cdot ||b - a||.$$

Доказательство [3].

По следствию из теоремы Хана-Банаха $\forall y \in Y \quad \exists y^* \in Y^* \quad | \quad ||y^*|| = 1$ и $y^*(y) = ||y||$. Пусть $y = f(b) - f(a)$. Тогда $\exists y^* \in Y^* \ | \ y^*[f(b) - f(a)] = ||f(b) - f(a)||$. Обозначим $\varphi(t) = y^*[f(a + t(b-a))]$. Так как y^* – линейный непрерывный функционал и отображение f дифференцируемо по Гато, то по теореме о суперпозиции $\varphi'(t) = y^*[f'_G(a + t(b-a))\,(b-a)] \ \ \forall t \in [0,1]$. Следовательно, $\varphi(1) - \varphi(0) = \varphi'(\theta), \ \ 0 < \theta < 1$. Тогда $||f(b) - f(a)|| = y^*[f(b) - f(a)] = \varphi(1) - \varphi(0) = y^*[f'_G(a + \theta(b-a))\,(b-a)] \leq ||f'_G(a + \theta(b-a))\,(b-a)|| \leq \ldots$

Следствие.
Пусть выполнены условия теоремы 3 и $\Lambda \in \mathcal{L}(X, Y)$. *Тогда*

$$||f(b) - f(a) - \Lambda(b-a)|| \leq \sup_{c \in (a,b)} ||f'_G(c) - \Lambda|| \cdot ||b - a||.$$

4.3. Дифференцирование в произведениях пространств

Пусть X, Y, Z – нормированные пространства, $x^0 \in X$, $\ U$ – окрестность x^0 в X, $\ f : X \to Y \times Z$, $U \subset D(f)$. Иными словами, $f = (g, h) \ | \ g : X \to Y, \ \ h : X \to Z$ – вектор-отображение, которое $x \in X$ ставит в соответствие $(g(x), h(x)) \in Y \times Z$.

Теорема 4.
Отображение $f = (g, h)$ *дифференцируемо на элементе* x^0 *в смысле 1. – 5. тогда и только тогда, когда отображения* g *и* h *дифференцируемы на* x^0 *в том же смысле. При этом*

$$f'(x^0; h) = (g'(x^0; h), h'(x^0; h)) \quad \ldots \quad f'(x^0) = (g'(x^0), h'(x^0)).$$

Доказательство: все следует из определений. •

Пусть X, Y, Z – нормированные пространства, $x^0 \in X$, $\ \ y^0 \in Y$, $\ \ U$ – окрестность (x^0, y^0) в $X \times Y$, $\ f : X \times Y \to Z$, $\ U \subset D(f)$.

Если отображение $x \mapsto f(x, y^0)$ дифференцируемо на элементе x^0 (по Гато, по Фреше или строго), то его производная называется *частной производной* по x отображения f на элементе (x^0, y^0) и обозначается $\dfrac{\partial f}{\partial x}(x^0, y^0)$. Аналогично определяется $\dfrac{\partial f}{\partial y}(x^0, y^0)$.

Теорема 5. (о полном дифференциале)
Пусть отображение $f : X \times Y \to Z$ *имеет на каждом элементе* $(x, y) \in X \times Y$ *частные производные по Гато* $\dfrac{\partial f}{\partial x}(x, y)$ *и* $\dfrac{\partial f}{\partial y}(x, y)$.

Если отображения $(x, y) \mapsto \dfrac{\partial f}{\partial x}(x, y)$ *и* $(x, y) \mapsto \dfrac{\partial f}{\partial y}(x, y)$ *непрерывны на элементе* (x^0, y^0), *то отображение* f *строго дифференцируемо на* (x^0, y^0) *и*

$$f'(x^0, y^0)(\xi, \eta) = \frac{\partial f}{\partial x}(x^0, y^0)\,\xi + \frac{\partial f}{\partial y}(x^0, y^0)\,\eta.$$

Доказательство. Пусть $\varepsilon > 0$. Выберем $\delta > 0$ так, чтобы
1) $\Pi = \{(x, y) \mid ||x - x^0|| < \delta, \ \ ||y - y^0|| < \delta\} \subset U$ и 2)

$$\Big|\Big|\frac{\partial f}{\partial x}(x, y) - \frac{\partial f}{\partial x}(x^0, y^0)\Big|\Big| < \varepsilon, \quad \Big|\Big|\frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(x^0, y^0)\Big|\Big| < \varepsilon.$$

Пусть (x', y'), $(x'', y'') \in U$. Тогда

$$\Delta = f(x', y') - f(x'', y'') - \frac{\partial f}{\partial x}(x^0, y^0)(x' - x'') - \frac{\partial f}{\partial y}(x^0, y^0)(y' - y'') =$$

$$= f(x', y') - f(x'', y') - \frac{\partial f}{\partial x}(x^0, y^0)(x' - x'') +$$

$$+f(x'', y') - f(x'', y'') - \frac{\partial f}{\partial y}(x^0, y^0)(y' - y'').$$

Элемент (x'', y') также принадлежит окрестности U, и отрезки $[(x', y'),\,(x'', y')]$, $[(x'', y'),\,(x'', y'')]$ лежат в U. Отображения $x \mapsto f(x, y')$ и $y \mapsto f(x'', y)$ дифференцируемы по Гато. По теореме о среднем

$$||\Delta|| \le \sup_{\xi \in [x', x'']} \left|\left| \frac{\partial f}{\partial x}(\xi, y') - \frac{\partial f}{\partial x}(x^0, y^0) \right|\right| \cdot ||x' - x''|| + \sup_{\eta \in [y', y'']} \left|\left| \frac{\partial f}{\partial y}(x', \eta) - \frac{\partial f}{\partial y}(x^0, y^0) \right|\right| \cdot ||y' - y''|| \le$$

$$\le \varepsilon\, ||x' - x''|| + \varepsilon\, ||y' - y''||. \;\bullet$$

4.4. Производные интегральных функционалов

Теперь можно строго доказать, что все рассмотренные в разделах 2 и 3 функционалы дифференцируемы, если соответствующие функции непрерывны и имеют непрерывные частные производные.

Рассмотрим функционал из простейшей задачи КВИ

$$I[x(\cdot)] = \int\limits_{t_0}^{t_1} L(t, x(t), x'(t))\; dt.$$

Чтобы обосновать дифференцируемость этого функционала, запишем его в виде суперпозиции нескольких отображений.

Пусть U – открытое множество в $R^3 = R^1 \times R^1 \times R^1$, $L \,:\, U \to R^1$, частные производные $\dfrac{\partial L}{\partial x}$ и $\dfrac{\partial L}{\partial y}$ непрерывны в U. Обозначим

$$V = \big\{ x(\cdot) \in C^1([t_0, t_1]) \;\;|\;\; (t, x(t), x'(t)) \in U \;\; \forall t \in [t_0, t_1] \big\}.$$

Пусть

$$D \,:\, x(\cdot) \mapsto x'(\cdot),$$

$$W \,:\, x(\cdot),\, y(\cdot) \mapsto (x(\cdot), y(\cdot)),$$

$$N \,:\, w(\cdot) \mapsto g(\cdot) \;\;|\;\; g(t) = L(t, w(t)),$$

$$J \,:\, g(\cdot) \mapsto \int\limits_{t_0}^{t_1} g(t)\; dt.$$

Тогда

$$I = J \circ N \circ W(I, D)$$

Аналогично, функционал

$$F[x(\cdot)] = \int\limits_a^b f\Big(t, x(t), \int\limits_a^b k(\tau, t, x(\tau))\; d\tau\Big)\; dt$$

можно представить в виде суперпозиции нескольких отображений.

5. Гладкие задачи с ограничениями в виде равенств и неравенств

Хорошо известно, что (в случае функций одной переменной) если в некоторой точке функция достигает экстремума и она дифференцируема, то производная в этой точке равна нулю. Аналогичные утверждения можно получить и в более общем случае.

5.1. Условия экстремума функционалов

Пусть X – нормированное (топологическое) пространство, $x^0 \in X$, U – окрестность x^0 в X, функционал $f : U \to R^1$, $U \subset D(f)$. Говорят, что элемент x^0 *доставляет минимум* функционалу $f(\cdot)$, если $f(x) \geq f(x^0) \quad \forall x \in U$. Обозначение: $x^0 \in \mathrm{loc}\ \min f(\cdot)$.

Теорема 1. (необходимое условие экстремума 1-го порядка)
Если $x^0 \in loc\ min f(\cdot)$ и функционал $f(\cdot)$ имеет на элементе x^0 или производную по направлению h, или 1-ю вариацию по Лагранжу, или производную Гато (Фреше), то или $f'(x^0;h) \geq 0$, или $\delta f(x^0;h) = 0 \quad \forall h \in X$, или $f'_G(x^0) = 0$ ($f'(x^0) = 0$).

Доказательство. Все следует из определений дифференцируемости, см. п. 4.1.

Замечание. Эта теорема – аналог теоремы Ферма классического МА: если дифференцируемая функция имеет экстремум в точке, то ее производная обращается в нуль в этой точке.

Определим *вторые производные* отображений. Для производных по направлению

$$f''(x^0;h) = \lim_{\lambda \downarrow 0} \frac{f'(x^0 + \lambda\, h; h) - f'(x^0;h)}{\lambda}.$$

(Мы пишем одно h в производной, но на самом деле их два.)

Лемма 1. (формула Тейлора)
Если отображение $f(\cdot)$ имеет производную Фреше порядка n, то

$$f(x^0+h) = f(x^0)+f'(x^0)\,h+\frac{1}{2!}\,f''(x^0)\,(h,h)+\ldots+\frac{1}{n!}\,f^{(n)}(x^0)\,(h,h,\ldots,h)+\alpha(h)\,||h||^n, \quad \lim_{h\to 0}\alpha(h) = 0.$$

Доказательство проводится по индукции.

Для вариаций отображений

$$\delta^n f(x^0;h) = \frac{d^n}{d\lambda^n}\, f(x^0 + \lambda h)\Big|_{\lambda=0}.$$

Теорема 2. (необходимое условие экстремума 2-го порядка)
Пусть $x^0 \in loc\ min f(\cdot)$. Если функционал $f(\cdot)$ имеет на элементе x^0 или 2-ю вариацию по Лагранжу, или 2-ю производную Гато (Фреше), то или $\delta^2 f(x^0;h) \geq 0 \quad \forall h \in X$, или $f''_G(x^0)(h,h) \geq 0$ ($f''(x^0)(h,h) \geq 0$) $\quad \forall h \in X$.

Доказательство, как и в классическом анализе, основано на формуле Тейлора.

Теорема 3. (достаточное условие экстремума 2-го порядка)
Пусть функционал $f(\cdot)$ имеет на элементе x^0 2-ю производную Фреше. Если $f'(x^0) = 0$ и найдется такое число $\alpha > 0$, что

$$f''(x^0)(h,h) \geq \alpha\,||h||^2 \quad \forall h \in X,$$

то $x^0 \in loc\ min f(\cdot)$.

Доказательство также проводится с помощью формулы Тейлора.

Пример. Пусть X, Y – гильбертовы пространства, $A : X \to Y$ – линейный непрерывный оператор, $f(x) = ||Ax - y||^2 = (Ax - y, Ax - y)$. Вычислим производные этого функционала.

Так как

$$f(x+h) - f(x) = 2(Ax - y, Ah) + (Ah, Ah)$$

и

$$(Ax - y, Ah) = (A^*(Ax - y), h), \qquad (Ah, Ah) = ||Ah||^2 \leq ||A||^2 ||h||^2,$$

то можно считать, что $f'(x) = 2A^*(Ax - y)$. Тогда 2-я производная равна $2A^*A$.

5.2. Правило множителей Лагранжа

Рассмотрим наиболее простой случай, когда пространства конечномерные.

Пусть функции $f_j : R^n \to R^1$, $j = 0..m$ $(m < n)$. Рассмотрим конечномерную экстремальную задачу с ограничениями в виде равенств

$$f_0(x) \to \mathrm{extr}, \quad f_j(x) = 0, \;\; j = 1..m. \quad (x \in R^n)$$

Теорема 4.
Пусть функции $f_j(\cdot)$, $j = 0..m$, имеют непрерывные частные производные в окрестности точки x^0. Если x^0 – решение экстремальной задачи, то найдется такой ненулевой вектор $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ (множители Лагранжа), что для функции Лагранжа

$$\mathcal{L}(x, \lambda) = \sum_{j=0}^{m} \lambda_j f_j(x)$$

выполняется условие (условие стационарности)

$$\frac{\partial \mathcal{L}}{\partial x}(x^0, \lambda) = 0 \quad \text{или} \quad \sum_{j=0}^{m} \lambda_j f_j'(x^0) = 0 \quad \text{или} \quad \sum_{j=0}^{m} \lambda_j \frac{\partial f_j}{\partial x_k}(x^0) = 0, \;\; k = 1..n.$$

Доказательство. От противного. Предположим, что условие стационарности не выполняется:

$$\sum_{j=0}^{m} \lambda_j f_j'(x^0) \neq 0.$$

Тогда векторы $f_j'(x^0)$ линейно независимы и ранг матрицы

$$\left\{ \frac{\partial f_j}{\partial x_k}(x^0) \right\}_{j=0..m,\; k=1..n}$$

равен $m+1$. Следовательно, существует не равный нулю минор порядка $m+1$.

Можно считать, что этот минор расположен в левом верхнем углу матрицы. А еще можно считать, что $f_0(x^0) = 0$ (иначе введем новую искомую функцию $\tilde{f}(x) = f_0(x) - f_0(x^0)$).

Обозначим $\check{x} = (x_1, \ldots, x_{m+1})$ и $\hat{x} = (x_{m+2}, \ldots, x_n)$. В окрестности U фиксированной точки $\check{x}^0 = (x_1^0, \ldots, x_{m+1}^0)$ рассмотрим отображение

$$F : U \to R^{m+1} \;\; | \;\; F(\check{x}) = (f_0(\check{x}, \hat{x}^0), \ldots, f_m(\check{x}, \hat{x}^0)).$$

Это отображение непрерывно дифференцируемо и $F(\check{x}^0) = 0$.

По теореме об обратной функции в конечномерных пространствах существует обратное отображение F^{-1}, определенное в окрестности точки $0 \in R^{m+1}$ и дифференцируемое в этой окрестности. Для y из такой окрестности выполняется неравенство $|F^{-1}(y) - F^{-1}(0)| \leq L|y - 0|$ или $|F^{-1}(y) - \check{x}^0| \leq L|y|$.

Пусть ε – достаточно малое число (любого знака). Пусть $y_\varepsilon = (\varepsilon, 0, \ldots, 0)$ и $\check{x}_\varepsilon = F^{-1}(y_\varepsilon)$. Тогда $F(\check{x}_\varepsilon) = (\varepsilon, 0, \ldots, 0)$ и $|\check{x}_\varepsilon - \check{x}^0| \leq L|\varepsilon|$.

Обозначим $x_\varepsilon = (\check{x}_\varepsilon, \hat{x}^0)$. Тогда $f_0(x_\varepsilon) = \varepsilon$ и $f_j(x_\varepsilon) = 0$, $j = 1..m$. Получилось, что x^0 не доставляет локальный экстремум в задаче, так как в окрестности этой точки есть точки x_ε, удовлетворяющие ограничениям в виде равенств, но значение $f_0(x_\varepsilon)$ может быть как положительным, так и отрицательным. •

Замечание 1. Производные можно понимать как по Гато или Фреше, так и как 1-е вариации по Лагранжу.

Замечание 2. Если векторы $f_j'(x^0)$, $j = 1..m$ линейно независимы, то $\lambda_0 \neq 0$.

Когда-то всегда полагали, что $\lambda_0 = 1$. Это не так. Есть простой контрпример: $x_1 \to \min$, $x_1^2 + x_2^2 = 0$. Очевидное решение – точка $(0, 0)$. Но …

Пример. [4], задача 3663.

$$xyz \to \mathrm{extr}, \;\; x + y + z = 0, \;\; x^2 + y^2 + z^2 = 1.$$

В случае бесконечномерных пространств дело обстоит существенно сложнее.

Пусть X, Y – нормированное пространство, отображение $F : X \to Y$, функционалы $f_j : X \to R^1$, $j = 0..m$. Рассмотрим экстремальную задачу с ограничениями типа равенств и неравенств:

$$f_0(x) \to \min, \quad f_j(x) \le 0, \ \ j = 1..m, \quad F(x) = 0, \quad x \in X.$$

Теорема 5.

Пусть X, Y – банаховы пространства, функционалы $f_j(\cdot)$, $j = 0..m$ и отображение $F(\cdot)$ строго дифференцируемы в окрестности точки x^0, множество значений оператора $F'(x^0)$ – замкнутое подпространство в Y.

Если x^0 – решение экстремальной задачи, то существуют такие вектор $\lambda \in R^{m+1}$ и функционал $y^* \in Y^*$ (множители Лагранжа), не равные нулю одновременно, что для функции Лагранжа

$$\mathcal{L}(x, \lambda, y^*) = \sum_{j=0}^{m} \lambda_j f_j(x) + < y^*, F(x) >$$

выполняются:
1) условие стационарности

$$\frac{\partial \mathcal{L}}{\partial x}(x^0, \lambda, y^*) = 0,$$

2) условия дополняющей нежесткости $\lambda_j f_j(x^0) = 0$, $j = 1..m$,
3) условия неотрицательности $\lambda_j \ge 0$, $j = 0..m$.

Дополнение к теореме. Если $f_j(x) = 0$, то условия 2) и 3) отсутствуют.

Доказательство можно найти в книге [5].

Пример (но все равно только конечномерный).

$$x_1^2 + x_2^2 + x_3^2 \to \min, \quad 2x_1 - x_2 + x_3 \le 5, \quad x_1 + x_2 + x_3 = 3.$$

Функция Лагранжа

$$\mathcal{L} = \lambda_0 (x_1^2 + x_2^2 + x_3^2) + \lambda_1 (2x_1 - x_2 + x_3 - 5) + \lambda_2 (x_1 + x_2 + x_3 - 3).$$

Условия стационарности

$$1) \qquad \lambda_0 2x_1 + 2\lambda_1 + \lambda_2 = 0, \quad \lambda_0 2x_2 - \lambda_1 + \lambda_2 = 0, \quad \lambda_0 2x_3 + \lambda_1 + \lambda_2 = 0,$$

условия дополняющей нежесткости

$$2) \qquad \lambda_1 (2x_1 - x_2 + x_3 - 5) = 0,$$

условия неотрицательности

$$3) \qquad \lambda_0 \ge 0, \quad \lambda_1 \ge 0$$

и ограничения задачи

$$4) \qquad 2x_1 - x_2 + x_3 \le 5, \quad x_1 + x_2 + x_3 = 3$$

образуют систему равенств и неравенств, которая имеет единственное решение

$$x_1 = x_2 = x_3 = -\lambda_2 = 1, \quad \lambda_1 = 0, \quad \lambda_0 = 0.5 \quad \text{или какое-то другое положительное число.}$$

Это не очевидно. Рассуждать нужно следующим образом.

Предположим сначала, что $\lambda_0 = 0$. Тогда из 1) следует, что $\lambda_1 = 0$, $\lambda_2 = 0$. Этого быть не может. Следовательно, $\lambda_0 \ne 0$. Удобно принять, что $\lambda_0 = 0.5$.

Тогда из 2) $2x_1 - x_2 + x_3 = 5$. Подставим сюда выражения x_1, x_2, x_3 через λ_1 и λ_2 из 1). Получим $-6\lambda_1 - 2\lambda_2 = 5$. Но из условия $x_1 + x_2 + x_3 = 3$ следует, что $-2\lambda_1 - 3\lambda_2 = 3 \ldots$

6. Оптимизация в бесконечномерных пространствах

Пусть $I : X \to R^1$, $x^0 \in X$, функционал $I(\cdot)$ дифференцируем по Фреше на x^0, то есть $I(x^0+h) = I(x^0) + I'(x^0)\, h + o(||h||)$. Здесь $I'(x^0)$ – тоже функционал, причем линейный и непрерывный. Его значение на элементе h обозначают также $< h,\, I'(x^0) >$.

Если $X = R^n$ (конечномерное пространство), то $I'(x^0)$ – градиент. Это вектор (по определению), который показывает направление наибольшего возрастания функционала (функции n переменных. Можно построить последовательность элементов x^k, сходящуюся к решению x^0 экстремальной задачи на минимум: x^1 – начальное приближение, $x^{k+1} = x^k - \alpha_k\, I'(x^k)$, $\ k = 1, 2, \ldots$.

6.1. Градиентные методы

В бесконечномерном пространстве все не так просто.

Предположим, что X – гильбертово пространство (вещественное или даже комплексное). По теореме Ф. Рисса каждому линейному непрерывному функционалу $f(\cdot)$ на X соответствует такой единственный элемент $y \in X$, что $< x,\, f >= (x,\, y)$. Кроме того, $||f|| = ||y||$. Так вот, в бесконечномерном пространстве *градиентом функционала* $f(\cdot)$ называют этот самый элемент y. В дальнейшем вместо y будем писать $I(\cdot)$. Поэтому для решения экстремальных задач в бесконечномерных пространствах можно использовать такие же алгоритмы, что и в конечномерном случае. Рассмотрим некоторые из них.

1. *Градиентный метод* используется при решении экстремальных задач без ограничений

$$I(x) \to \min, \quad x \in X,$$

здесь X – гильбертово пространство.

Алгоритм такой: выбирается начальный элемент x^1, а следующие элементы минимизирующей последовательности вычисляются по формуле

$$x^{k+1} = x^k - \alpha_k\, I'(x^k), \quad k = 1, 2, ...,$$

пока не будет достигнута требуемая точность. Здесь $I'(x^k)$ – градиент функционала, вычисленный на элементе x^k.

Числа α_k могут быть определены различными способами, в том числе:

– *примитивно*: из условия монотонности $I(x^{k+1}) < I(x^k)$. Например, можно задать некоторое значение α, положить $\alpha_k = \alpha$, а если на k-м шаге условие монотонности не выполняется, уменьшать это число;

– *априорно*: так, что

$$\alpha_k > 0, \quad \sum_{k=1}^{+\infty} \alpha_k = +\infty, \quad \sum_{k=1}^{+\infty} \alpha_k^2 < +\infty.$$

Например, $\alpha_k = 1/(k^\lambda)$, $\ 1/2 < \lambda < 1$;

– *оптимально*: α_k – решение задачи

$$I(x^k - \alpha_k\, I'(x^k)) \to \min$$

(метод скорейшего спуска).

2. *Метод проекции градиента* применяют в случае, когда имеются ограничения:

$$I(x) \to \min, \quad x \in \mathcal{X} \subset X,$$

предполагается, что $\mathcal{X}$ – выпуклое замкнутое подмножество в X.

На каждом шаге алгоритма новое приближение к решению задачи нужно спроектировать на множество ограничений,

$$x^{k+1} = P_{\mathcal{X}}[x^k - \alpha_k\, I'(x^k)], \quad k = 1, 2, ...$$

Здесь $P_{\mathcal{X}}[x]$ – проекция элемента x на множество $\mathcal{X}$, то есть такой элемент $p \in \mathcal{X}$, что

$$||x - p|| = \inf_{q \in \mathcal{X}} ||x - q||.$$

Например, если $\mathcal{X} = \{x \in X \mid ||x - \tilde{x}|| \le R\}$ – замкнутый шар, то

$$P_{\mathcal{X}}(x) = \Big\{ \, ||x - \tilde{x}|| \le R \, : \, x; \quad \text{иначе} \quad : \, \tilde{x} + r\,(x - \tilde{x})/||x - \tilde{x}|| \Big\}.$$

Часто используют также

3. *Метод условного градиента*

$$x^{k+1} = x^k + \alpha_k\,(\overline{x}^k - x^k), \quad 0 \le \alpha_k \le 1,$$

где $\overline{x}^k$ – решение задачи

$$< \overline{x}^k - x^k, \, I'(x^k) >= \inf_{x \in \mathcal{X}} < x - x^k, \, I'(x^k) >$$

($\mathcal{X}$ – ограниченное выпуклое замкнутое множество).

4. *Метод Ньютона-Канторовича*, когда элемент $\overline{x}^k$ ищется из условия $I_k(\overline{x}^k) = \inf\limits_{x \in \mathcal{X}} I_k(x)$, где

$$I_k(x) =< x - x^k, \, I'(x^k) > + \frac{1}{2} < x - x^k, \, I''(x^k)(x - x^k) > .$$

Если $\mathcal{X} = X$ и оператор $I''(x^k)$ имеет обратный, то $x^{k+1} = x^k - \alpha_k\,(I''(x^k))^{-1} I'(x^k)$.

Доказаны теоремы о сходимости градиентных методов и получены оценки скорости сходимости [6], [7].

6.2. Задача оптимального управления с линейным уравнением состояний и квадратичным функционалом

Рассмотрим упрощенный одномерный вариант одной задачи оптимального управления [6], с. 31.

Постановка задачи:

$$I = [x(t_1) - x_1]^2 \to \min,$$

$$x'(t) = a(t)\,x(t) + b(t)\,u(t) + f(t), \quad t_0 < t < t_1, \quad x(t_0) = x_0,$$

$$u(t) \in \mathcal{U} \quad \forall t \in [t_0, t_1].$$

Здесь $u(\cdot)$ – управление, $x(\cdot)$ – состояние (зависит от управления, $x(\cdot) = x[u(\cdot)](\cdot)$. Поэтому $I = I[u(\cdot)]$.

Смысл задачи в следующем: задано начало траектории $x = x(t)$; нужно найти такое управление, что соответствующая ему траектория заканчивается наиболее близко к заданной точке.

Чтобы применить метод проекции градиента, нужно научиться вычислять градиент!

План такой: дадим приращение $\Delta u(\cdot)$ управлению $u(\cdot)$ и исследуем соответствующее приращение функционала

$$\Delta I = I[u(\cdot) + \Delta u(\cdot)] - I[u(\cdot)].$$

Если окажется, что

$$\Delta I = \int\limits_{t_0}^{t_1} G(\tau)\,\Delta u(\tau)\,d\tau + o(||\Delta u(\cdot)||),$$

то функция $G(\cdot)$ и есть градиент функционала.

Обозначим приращение состояния

$$\Delta x(\cdot) = x[u(\cdot) + \Delta u(\cdot)](\cdot) - x[u(\cdot)](\cdot).$$

Эта функция является решением задачи Коши

$$\Delta x'(t) = a(t)\,\Delta x(t) + b(t)\,\Delta u(t), \quad x(t_0) = 0.$$

Тогда (как в теории ДУ)

$$\Delta x(t) = \int\limits_{t_0}^{t} \big[a(\tau)\,\Delta x(\tau) + b(\tau)\,\Delta u(\tau)\big]\,d\tau$$

и

$$|\Delta x(t)| \le A \int\limits_{t_0}^{t} |\Delta x(\tau)|\,d\tau + B \int\limits_{t_0}^{t} |\Delta u(\tau)|\,d\tau \le B\,|\Delta u(\tau)|\,e^{A(t-t_0)}$$

(по лемме об интегральных неравенствах, $A = \max|a(t)|,\ B = \max|b(t)|$).

Рассмотрим теперь приращение функционала

$$\Delta I = \{x[u(\cdot)+\Delta u(\cdot)](t_1) - x_1\}^2 - \{x[u(\cdot)](t_1) - x_1\}^2 =$$

$$= \{x[u(\cdot)+\Delta u(\cdot)](t_1) - x[u(\cdot)](t_1) - 2\,x_1\}\,\Delta x(t_1) =$$

$$= 2\{x[u(\cdot)](t_1) - x_1\}\,\Delta x(t_1) + [\Delta x(t_1)]^2$$

(в квадратных скобках вычли и добавили $x[u(\cdot)](t_1)$). Второе слагаемое в силу полученного выше неравенства есть $o(||\Delta u(\cdot)||)$.

Теперь преобразуем первое слагаемое.

Пусть $y(\cdot)$ – решение задачи Коши

$$y'(t) = -a(t)\,y(t), \quad y(t_1) = 2\{x[u(\cdot)](t_1) - x_1\}$$

(почему именно так, скоро станет ясно). Тогда первое слагаемое

$$y(t_1)\,\Delta x(t_1) = y(t_1)\,\Delta x(t_1) - y(t_0)\,\Delta x(t_0) = \int\limits_{t_0}^{t_1} [y(\tau)\,\Delta x(\tau)]'\,d\tau =$$

$$= \int\limits_{t_0}^{t_1} [y'(\tau)\,\Delta x(\tau) + y(\tau)\,\Delta x'(\tau)]\,d\tau = \int\limits_{t_0}^{t_1} y(\tau)\,b(\tau)\,\Delta u(\tau)\,d\tau.$$

Следовательно, градиент функционала равен $y(\cdot)\,b(\cdot)$.

Алгоритм решения задачи оптимального управления такой:
задаем начальное приближение $u(\cdot)$;
находим $x(\cdot)$ как решение задачи Коши и, следовательно, $x(t_1)$;
находим $y(\cdot)$ как решение задачи Коши;
вычисляем градиент функционала $I'[u(\cdot)](\cdot) = y(\cdot)\,b(\cdot)$;
находим новое управление $u(\cdot)$ по методу проекции градиента
и, если точность вычислений не достигнута, возвращаемся на второй шаг.

6.3. Оптимальное управление температурой стержня

Пусть вдоль отрезка $[0,l]$ оси s расположен тонкий проводящий тепло стержень, $x(s,t)$ – температура в точке (в сечении) стержня с координатой s в момент времени t. Рассмотрим экстремальную задачу [6]

$$\frac{\partial x}{\partial t} = a^2\,\frac{\partial^2 x}{\partial s^2} + f(s,t), \quad (s,t) \in Q = \{0 < s < l,\ 0 < t < T\},$$

$$\frac{\partial x}{\partial s}(0,t) = 0, \quad \frac{\partial x}{\partial s}(l,t) = \nu\,[p(t) - x(l,t)], \quad 0 < t \le T,$$

$$x(s,0) = \varphi(s), \quad 0 \le s \le l,$$

$$\int\limits_{0}^{l} [x(s,T) - y(s)]^2\,ds \to \min.$$

Условия этой задачи таковы: температура стержня (состояние) – решение уравнения теплопроводности, левый его конец теплоизолирован, только через правый конец происходит обмен теплом с окружающей средой, начальное распределение температуры вдоль стержня задано. Нужно найти такое управление – плотность источников тепла внутри стержня $f(s,t)$ и температуру окружающей среды $p(t)$, что в момент времени T распределение температуры вдоль стержня будет наиболее близким к заданному распределению.

Дополнительные ограничения на управление такие:

$$p_{min} \le p(t) \le p_{max}, \quad \iint\limits_Q |f(s,t)|^2\, ds\, dt \le F.$$

Будем считать, что управление $u = (p, f)$ – непрерывные функции (хотя они могут быть интегрируемы с квадратом; если имеются точки разрыва, то нужно рассматривать обобщенное решение уравнение теплопроводности). Скалярное произведение в пространстве управлений

$$(u_1,\, u_2) = \int\limits_0^T p_1(t)\, p_2(t)\, dt + \iint\limits_Q f_1(s,t)\, f_2(s,t)\, ds\, dt.$$

Докажем, что функционал в этой задаче оптимального управления дифференцируем и вычислим его градиент.

Приращению управления $\Delta u = (\Delta p,\, \Delta f)$ соответствуют приращение состояния Δx и приращение функционала ΔI.

Приращение состояния является решением задачи

$$\frac{\partial(\Delta x)}{\partial t} = a^2\, \frac{\partial^2(\Delta x)}{\partial s^2} + \Delta f, \quad \frac{\partial(\Delta x)}{\partial s}(0,t) = 0, \quad \frac{\partial(\Delta x)}{\partial s}(l,t) = \nu\,[\Delta p - \Delta x(l,t)], \quad \Delta x(s,0) = 0.$$

Приращение функционала

$$\Delta I = \int\limits_0^l [x(s,T) + \Delta x(s,T) - y(s)]^2\, ds - \int\limits_0^l [x(s,T) - y(s)]^2\, ds =$$

$$= \int\limits_0^l 2[x(s,T) - y(s)]\, \Delta x(s,T)\, ds + \int\limits_0^l (\Delta x)^2(s,T)\, ds.$$

Можно показать, что второе слагаемое есть $o(||\Delta u||)$.

Пусть функция $\psi(s,t)$ – решение вспомогательной задачи

$$\frac{\partial \psi}{\partial t} = -a^2\, \frac{\partial^2 \psi}{\partial s^2}, \quad \frac{\partial \psi}{\partial s}(0,t) = 0, \quad \frac{\partial \psi}{\partial s}(l,t) = -\nu\, \psi(l,t), \quad \psi(s,T) = 2[x(s,T) - y(s)].$$

Первое слагаемое в приращении функционала нужно представить в виде

$$\int\limits_0^T G(t)\, \Delta p(t)\, dt + \iint\limits_Q H(s,t)\, \Delta f(s,t)\, ds\, dt,$$

тогда градиент – пара функций $(G(t),\, H(s,t))$.

Сделаем следующее:

$$\int\limits_0^l 2[x(s,T) - y(s)]\, \Delta x(s,T)\, ds = \int\limits_0^l \psi(s,T)\, \Delta x(s,T)\, ds = \int\limits_0^l \Big(\int\limits_0^T \frac{\partial}{\partial t}(\psi\, \Delta x)\, dt\Big)\, ds =$$

$$= \int\limits_0^l \int\limits_0^T \Big[\frac{\partial \psi}{\partial t}\, \Delta x + \psi\, \frac{\partial(\Delta x)}{\partial t}\Big]\, dt\, ds = \int\limits_0^l \int\limits_0^T \Big[-a^2\, \frac{\partial^2 \psi}{\partial s^2}\, \Delta x \mp a^2\, \frac{\partial \psi}{\partial s}\, \frac{\partial(\Delta x)}{\partial s} + \psi\, a^2\, \frac{\partial^2(\Delta x)}{\partial s^2} + \psi\, \Delta f\Big]\, dt\, ds =$$

$$= a^2 \int\limits_0^T \Big(\int\limits_0^l \frac{\partial}{\partial s} \Big[-\frac{\partial \psi}{\partial s}\, \Delta x + \psi\, \frac{\partial(\Delta x)}{\partial s} \Big]\, ds \Big)\, dt + \iint\limits_Q \psi\, \Delta f\, ds\, dt.$$

Так как

$$\int\limits_0^l \frac{\partial}{\partial s} \Big[-\frac{\partial \psi}{\partial s}\, \Delta x + \psi\, \frac{\partial(\Delta x)}{\partial s} \Big]\, ds = -\frac{\partial \psi}{\partial s}\, \Delta x + \psi\, \frac{\partial(\Delta x)}{\partial s} \Big|_{s=0}^{s=l} = \ldots = \psi(l,t)\, \nu\, \Delta p,$$

то градиент равен $(a^2\, \nu\, \psi(l,t),\ \psi(s,t))$.

Таким образом, алгоритм решения задачи оптимального управления следующий:
задаем начальное приближение $p(t)$, $f(s,t)$;
решаем краевую задачу и находим $x(s,t)$ и $x(s,T)$;
решаем краевую задачу и находим $\psi(s,t)$;
вычисляем градиент функционала $I'[u(\cdot)](\cdot) = (a^2\, \nu\, \psi(l,t),\ \psi(s,t))$;
находим новое управление $p(t)$, $f(s,t)$ (шаг по градиенту и проекция);
если точность вычислений не достигнута, возвращаемся на второй шаг.

Литература

1. Треногин В.А. Функциональный анализ.
2. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа.
3. Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление.
4. Демидович Б.П. Сборник задач и упражнений по математическому анализу.
5. Галеев Э.М., Тихомиров В.М. Краткий курс теории экстремальных задач.
6. Васильев Ф.П. Методы решения экстремальных задач.
7. Канторович Л.В., Акилов Г.П. Функциональный анализ.