Условный экстремум

Пусть M^k - гладкая k-мерная поверхность в $\mathbb{R}^n$. Пусть $f\colon \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируемая функция.

Опр: 1. Точка $p \in M^k$ является точкой локального условного максимума функции f на M^k, если:

$$\exists \mathcal{U}(p)\colon \forall x \in \mathcal{U}(p) \cap M^k,\ f(x) \le f(p)$$

Опр: 2. Точка $p \in M^k$ является точкой строгого локального условного максимума функции f на M^k, если:

$$\exists \mathcal{U}(p)\colon \forall x \in \mathcal{U}(p) \cap M^k, x \neq p,\ f(x) < f(p)$$

Опр: 3. Точка $p \in M^k$ является точкой локального условного минимума функции f на M^k, если:

$$\exists \mathcal{U}(p)\colon \forall x \in \mathcal{U}(p) \cap M^k,\ f(x) \ge f(p)$$

Опр: 4. Точка $p \in M^k$ является точкой строгого локального условного минимума функции f на M^k, если:

$$\exists \mathcal{U}(p)\colon \forall x \in \mathcal{U}(p) \cap M^k, x \neq p,\ f(x) > f(p)$$

Опр: 5. Точки локального условного максимума и минимума (в том числе строгого) называются точками локального условного экстремума.

Таким образом, мы изучаем на экстремум функцию f, но смотрим не на всём $\mathbb{R}^n$, а ограничиваемся значениями на M^k. Мы хотим научиться узнавать, чем характерны точки в которых функция f достигает своего максимума и минимума на поверхности M^k. Всего два основных утверждения.

Необходимое условие локального условного экстремума

Теорема 1. (Необходимое условие) Если p - точка локального условного экстремума на M^k, то:

$$\nabla f(p) \perp T_pM^k$$

В частности, если градиент $\nabla f(p) \neq 0$, то $T_pM^k \subset TS$, где TS это касательное пространство в точке p к $\{x\colon f(x) = f(p)\}$.

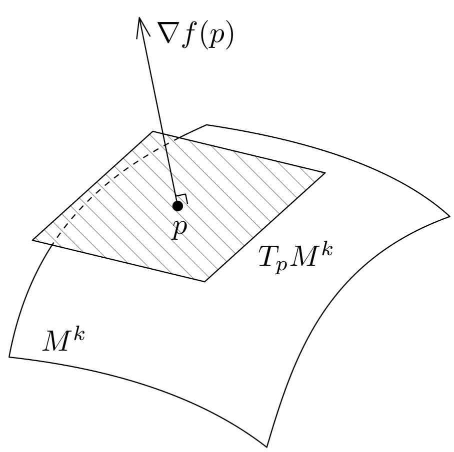

Рис. 1: Градиент функции f в точке p перпендикулярен касательной плоскости.

Идея: Если в точке p у функции f локальный условный экстремум и мы в ней посчитаем градиент, то он должен быть перпендикулярен поверхности $M^k \Rightarrow$ перпендикулярен касательной плоскости.

С другой стороны, градиент (невырожденный) это то что перпендикулярно множеству уровня. Если градиент f отличен от нуля, то в окрестности точки p множество уровня $\{x\colon f(x) = f(p)\}$ является гладкой $(n-1)$-мерной поверхностью и тогда свойства перпендикулярности означают, что касательная плоскость T_pM^k лежит среди всех касательных векторов $TS\big(f(p)\big)$ уровня $f(x) = f(p)$.

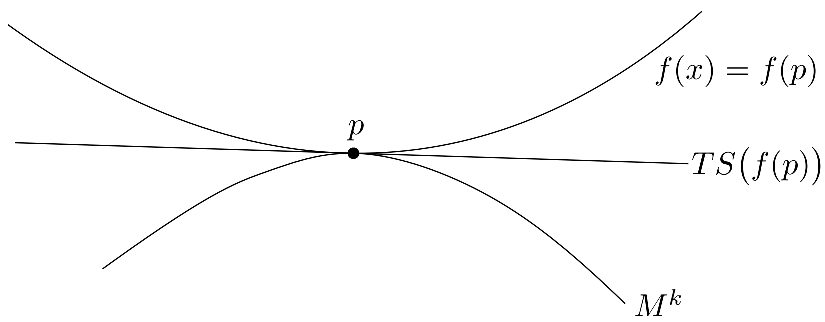

Рис. 2: Касательная плоскость T_pM^k к M^k в p содержится внутри касательного пространства $TS\big(f(p)\big)$.

Геометрический смысл: Если размерность $TS\big(f(p)\big)$ и T_pM^k была бы одинаковая, то они бы совпадали $\Leftrightarrow$ происходит касание между поверхностями M^k и $\{x\colon f(x) = f(p)\}$. Но если размерность T_pM^k меньше размерности $TS\big(f(p)\big)$ (множество уровня функции f это гиперповерхность, а поверхность с условиями M^k уже меньшей размерности), то касательное пространство к M^k будет содержаться в касательном пространстве к множеству уровня.

Пусть мы в $\mathbb{R}^3$, можно поставить два дополнительных условия (два уравнения) и взять поверхность M^k в виде линий. Пусть M^k это линия $\Rightarrow k = 1$. Если на этой линии функция f достигает экстремума в точке p, то касательный вектор в этой точке к этой линии должен лежать в касательном пространстве к множеству уровня функции f.

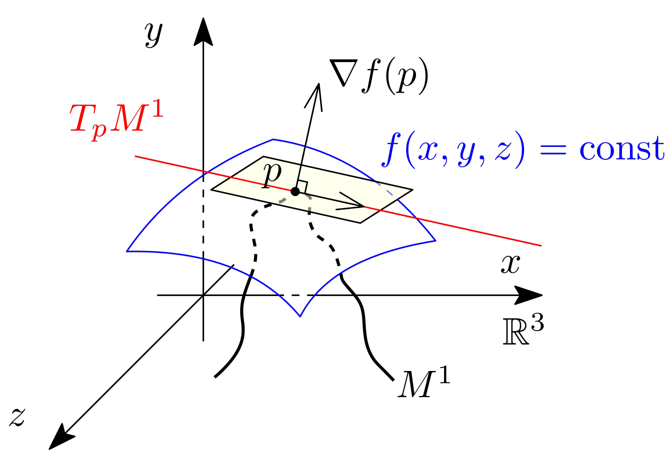

Рис. 3: Касательная плоскость к $M^1 \subset$ касательная плоскость к линии уровня $f(p)$.

Rm: 1. Это воплощение геометрического образа, которое мы видели на конкретных примерах по поиску условного экстремума в предыдущих лекциях.

☐ Пусть $v \in T_pM^k \Rightarrow \exists\, \gamma\colon \gamma(0) = p,\ \gamma(t) \in M^k,\ \dot\gamma(0) = v$. Рассмотрим $f\big(\gamma(t)\big)$ - функция одного аргумента, у которой в точке $t = 0$ - локальный экстремум, поскольку у функции f в точке p - локальный экстремум.

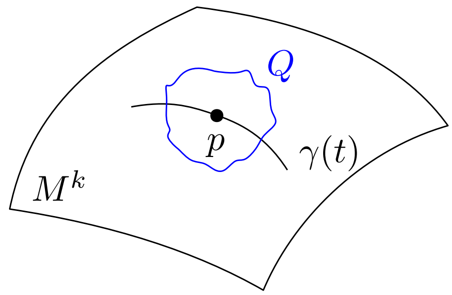

Рис. 4: В окрестности Q точки p, функция f имеет точку экстремума в p.

Поскольку функция одномерная, то в точке $t = 0$ верно следующее:

$$\frac{d}{dt}\bigg|_{t=0} f\big(\gamma(t)\big) = 0 \Rightarrow \langle \nabla f(p), \dot\gamma(0)\rangle = 0 \Rightarrow \langle \nabla f(p), v\rangle = 0 \Rightarrow \nabla f(p) \perp T_pM^k$$

■

Следствие 1. Если в условиях теоремы множество M^k в окрестности точки p задано системой уравнений: $F_{k+1}(x) = 0, \dots, F_n(x) = 0$, где градиенты $\nabla F_{k+1}(p), \dots, \nabla F_n(p)$ - линейно независимы, то:

$$\exists\, \lambda_{k+1}, \dots, \lambda_n \in \mathbb{R}\colon \nabla f(p) = \lambda_{k+1}\nabla F_{k+1}(p) + \ldots + \lambda_n \nabla F_n(p)$$

☐ По утверждению с прошлой лекции:

$$v \in T_pM^k \Leftrightarrow \left\{\begin{array}{ccc} \langle \nabla F_{k+1}(p), v\rangle & = & 0 \\ \vdots & \vdots & \vdots \\ \langle \nabla F_n(p), v\rangle & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{c} v \perp \nabla F_{k+1}(p) \\ \vdots \\ v \perp \nabla F_n(p) \end{array}\right.$$

Заметим, что $\dim(T_pM^k) = k$, число градиентов равно $(n-k)$, они линейно независимы и лежат в ортогональной плоскости $(T_pM^k)^\perp$. Из линейной алгебры мы знаем, что размерность ортогонального пространства равна $\dim(\mathbb{R}^n) - \dim(T_pM^k) = n - k$, значит $\nabla F_{k+1}(p), \dots, \nabla F_n(p)$ - это базис в $(T_pM^k)^\perp$.

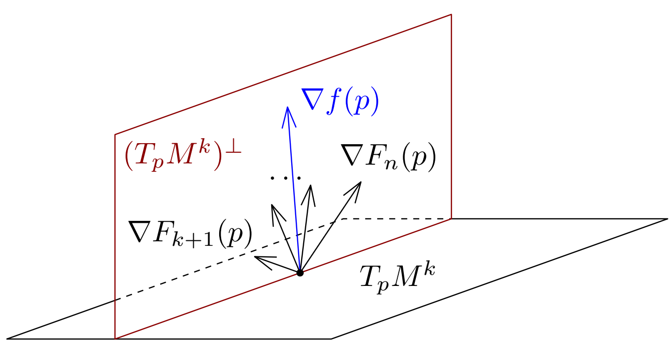

Рис. 5: Ортогональное пространство $(T_pM^k)^\perp$.

Поскольку $\nabla f(p) \perp T_pM^k \Rightarrow \nabla f(p) \in (T_pM^k)^\perp \Rightarrow \nabla f(p)$ является линейной комбинацией базисных векторов ортогонального пространства $\Rightarrow \exists\, \lambda_{k+1}, \dots, \lambda_n \in \mathbb{R}\colon \nabla f(p) = \lambda_{k+1}\nabla F_{k+1}(p) + \ldots + \lambda_n\nabla F_n(p)$. ■

Стандартная задача на условный экстремум

В обычной постановке задачи звучат так:

$$\begin{cases} f \to \text{extr!} \\ F_{k+1} = 0 \\ \vdots \\ F_n = 0 \end{cases}$$

где условия $F_{k+1} = 0, \dots, F_n = 0$ задают k-мерную поверхность. Нарисуем в точке p касательную плоскость, тогда градиенты $\nabla F_{k+1}(p), \dots, \nabla F_n(p)$ будут лежать в ортогональном дополнении и градиент функции f в точке p перпендикулярен касательному пространству.

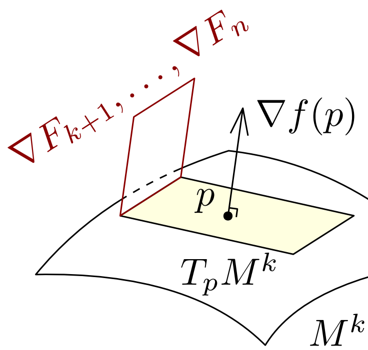

Рис. 6: Градиент функции f выражается через градиенты $\nabla F_{k+1}, \dots, \nabla F_n$.

Следовательно $\nabla f(p)$ должен выражаться через градиенты $\nabla F_{k+1}(p), \dots, \nabla F_n(p)$, то есть:

$$\nabla f(p) = \lambda_{k+1} \nabla F_{k+1}(p) + \ldots + \lambda_n \nabla F_n(p)$$

Метод множителей Лагранжа

Поиск условного экстремума можно представить в ином виде. Запишем функцию Лагранжа:

$$L(x, \lambda) = f(x) - \lambda_{k+1} F_{k+1}(x) - \ldots - \lambda_n F_n(x)$$

Необходимое условие условного экстремума:

$$\begin{cases} \dfrac{\partial L}{\partial x_i} = 0, \; i = \overline{1, n} \\ \dfrac{\partial L}{\partial \lambda_j} = 0, \; j = \overline{k+1, n} \end{cases}$$

где первые равенства получены из теоремы о достаточном признаке:

$$\frac{\partial L}{\partial x_i} = 0, \; i = \overline{1, n} \Leftrightarrow \nabla f(p) - \lambda_{k+1} \nabla F_{k+1}(p) - \ldots - \lambda_n \nabla F_n(p) = 0$$

и вторые равенства есть просто задание k-мерной поверхности:

$$\frac{\partial L}{\partial \lambda_j} = 0, \; j = \overline{k+1, n} \Leftrightarrow F_j(x) = 0, \; j = \overline{k+1, n}$$

Суть метода: хотим исследовать функцию f на экстремум при условии $F_{k+1} = 0, \dots, F_n = 0$, тогда необходимо составить функцию Лагранжа и действовать также как и раньше, исследуя функцию на обычный экстремум. Конкретнее, если хотим найти точки подозрительные на экстремум, то необходимо приравнять частные производные функции Лагранжа к нулю.

Задача Лагранжа: Пусть есть тонкая проволока, которая задается как $F(x) = 0$ и по ней без трения двигается бусина под действием некоей силы. Сила потенциальна (то есть она тянет куда-то бусину) и задается как ∇f. Изучаются точки равновесия, то есть точки в которых эта бусина будет в покое на заданной кривой.

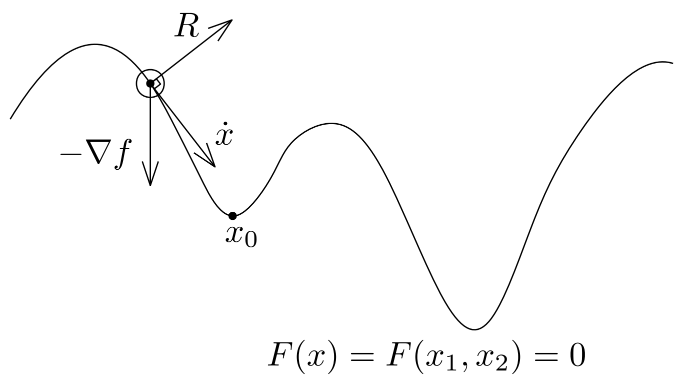

Рис. 7: Задача Лагранжа: ищем точки равновесия бусины на проволоке.

Движение бусины описывается следующим образом (по закону Ньютона): $\ddot{x} = -\nabla f + R$ (массу бусины считаем равной 1). Домножим на $\dot{x}$ (вектор скорости: всегда будет направлен по касательной к этой проволоке $\Rightarrow$ будет перпендикулярен силе рекции опоры R), тогда получим:

$$\langle \ddot{x}, \dot{x} \rangle = -\langle \nabla f, \dot{x} \rangle + \langle R, \dot{x} \rangle = -\langle \nabla f, \dot{x} \rangle \Leftrightarrow \frac{d}{dt}\left(\frac{\|\dot{x}\|^2}{2}\right) = -\frac{d}{dt} f\big(x(t)\big)$$

Получаем закон сохранения энергии:

$$\frac{\|\dot{x}\|^2}{2} + f(x) = \text{const}$$

Представим, что взяли точку минимума потенциальной энергии x_0 ($f(x)$ трактуется как потенциальная энергия). Может ли бусина уехать из этой точки? Нет, так как необходима положительная скорость для этого $\Rightarrow$ необходимо уменьшить f, чтобы сумма оставалась константой, но мы и так взяли минимум. Точка минимума является точкой равновесия для бусины, а это означает что силы должны быть равны:

$$-\nabla f(x_0) + R(x_0) = 0$$

Сила реакции опоры действует перпендикулярно линии $F(x) = 0 \Rightarrow R = \lambda \nabla F$ и получим:

$$\nabla f(x_0) = \lambda \nabla F(x_0)$$

Достаточное условие локального условного экстремума

Теорема 2. (Достаточное условие) Пусть поверхность M^k в окрестности точки p задана системой уравнений $F_{k+1} = 0, \ldots, F_n = 0$, где F_i - дважды непрерывно дифференцируемые функции, f тоже дважды непрерывно дифференцируемая функция и в точке p выполняется необходимое условие локального условного экстремума. Пусть $L(x, \lambda) = L(x)$ - это соответствующая функция Лагранжа. Тогда:

1) Если $d^2L(p, h) > 0$, $\forall h \neq 0$, $h \in T_pM^k$, то p - точка строгого локального условного минимума функции f;
2) Если $d^2L(p, h) < 0$, $\forall h \neq 0$, $h \in T_pM^k$, то p - точка строгого локального условного максимума функции f;
3) Если $\exists\, v, h \in T_pM^k$: $d^2L(p, v) > 0$, $d^2L(p, h) < 0$, то p не является точкой локального условного экстремума;

Rm: 2. Проверять положительную или отрицательную определенность необходимо не на всех векторах, а только на касательных пространствах.

В точке p знак приращения функции определяется вторым дифференциалом, но приращения функции надо брать вдоль заданной поверхности M^k, а это все равно что взять касательные приращения (брать только касательные вектора). И таким образом, мы смотрим как меняется функция вдоль касательных направлений.

Rm: 3. Отметим также, что для изучения свойств f мы рассматриваем новую функцию L. Почему так? Почему бы не использовать достаточные условия локального экстремума:

$$f(p + h) - f(p) = \langle \nabla f(p), h \rangle + \tfrac{1}{2} d^2 f(p, h) + \ldots$$

Пусть $h \in T_pM^k$, тогда:

$$f(p + h) - f(p) = 0 + + \tfrac{1}{2} d^2 f(p, h) + \ldots$$

Почему так нельзя? Так нельзя поскольку $p + h \notin M^k$. Из-за того, что поверхность это нелинейный объект и приходится переходить к функции Лагранжа. Ровно по этой же причине второй дифференциал не является инвариантным при замене координат.

☐ Выпишем функцию Лагранжа $L(x)$:

$$L(x) = L(x, \lambda) = f(x) - \lambda_{k+1} F_{k+1}(x) - \ldots - \lambda_n F_n(x)$$

и заметим что на M^k верно $f(x) = L(x)$, поскольку если $x \in M^k \Rightarrow F_{k+1}(x) = 0, \ldots, F_n(x) = 0$. Следовательно p - точка экстремума L на $M^k \Leftrightarrow p$ - точка экстремума f на M^k. Заметим также, что:

$$\frac{\partial L}{\partial x_i}(p) = 0,\ i = \overline{1, n}$$

из необходимого условия локального условного экстремума.

Мы ранее обсуждали, что поверхность M^k в окрестности точки p можно задать параметрически:

$$\begin{cases} x_1 & = & x_1(u_1, \ldots, u_k) \\ \vdots & \vdots & \vdots \\ x_n & = & x_n(u_1, \ldots, u_k) \end{cases}$$

где $u_1, \ldots, u_k$ находятся в окрестности нуля W и $x(0) = p$. Обозначим $L\big(x(u)\big) = H(u)$. Точка p - точка локального экстремума L на $M^k \Leftrightarrow u = 0$ - точка локального экстремума H.

Таким образом, задача о локальном условном экстремуме свелась к задаче об обычном локальном экстремуме функции H в координатах $u_1, \ldots, u_k$. Рассмотрим следующее:

$$\frac{\partial H}{\partial u_m}(0) = \sum_i \frac{\partial L}{\partial x_i}(p) \cdot \frac{\partial x_i}{\partial u_m}(0) = \sum_i 0 \cdot \frac{\partial x_i}{\partial u_m}(0) = 0$$

Следовательно для H выполнено необходимое условие. Рассмотрим вторые производные H:

$$\frac{\partial^2 H}{\partial u_l \partial u_m}(u) = \sum_{j,i} \frac{\partial^2 L}{\partial x_j \partial x_i}\big(x(u)\big) \cdot \frac{\partial x_j}{\partial u_l}(u) \cdot \frac{\partial x_i}{\partial u_m}(u) + \sum_i \frac{\partial L}{\partial x_i}\big(x(u)\big) \cdot \frac{\partial^2 x_i}{\partial u_l \partial u_m}(u)$$

Тогда в точке 0 мы получим:

$$\frac{\partial^2 H}{\partial u_l \partial u_m}(0) = \sum_{j,i} \frac{\partial^2 L}{\partial x_j \partial x_i}\big(p\big) \cdot \frac{\partial x_j}{\partial u_l}(0) \cdot \frac{\partial x_i}{\partial u_m}(0) + 0 = \sum_{j,i} \frac{\partial^2 L}{\partial x_j \partial x_i}\big(p\big) \cdot \frac{\partial x_j}{\partial u_l}(0) \cdot \frac{\partial x_i}{\partial u_m}(0)$$

Rm: 4. Обычно про такой эффект говорят, что в критической точке второй дифференциал инвариантен.

Запишем тогда второй дифференциал функции H:

$$d^2H(0, v) = \sum_{j,i} \frac{\partial^2 L}{\partial x_i \partial x_j}(p) \cdot \sum_{l,m} \frac{\partial x_i}{\partial u_m} \cdot \frac{\partial x_j}{\partial u_l} \cdot v_m \cdot v_l = d^2L(p, h),\ h = \frac{\partial x}{\partial u_1} \cdot v_1 + \ldots + \frac{\partial x}{\partial u_k} \cdot v_k$$

где из предыдущей лекции мы знаем, что касательное пространство натянуто на вектора $\frac{\partial x}{\partial u_1}, \ldots, \frac{\partial x}{\partial u_k}$. Следовательно верно, что:

$$h = \frac{\partial x}{\partial u_1} \cdot v_1 + \ldots + \frac{\partial x}{\partial u_k} \cdot v_k \in T_pM^k,\ \forall v_l$$

Таким образом, $d^2H(0, v) > 0 \Leftrightarrow d^2L(p, h) > 0$ и $d^2H(0, v) < 0 \Leftrightarrow d^2L(p, h) < 0$. Тем самым, утверждение вытекает из обычной теоремы о локальном экстремуме применяемой к функции H. ■

Пример: Пусть на плоскости задана функция $f(x, y)$, необходимо найти её экстремум на оси x:

$$\begin{cases} f \to \text{extr!} \\ y = 0 \end{cases}$$

Что в этом случае делаем? Подставляем $y = 0$ и исследуем функцию $f(x, 0)$.

В последнем шаге доказательства мы делаем нечто похожее: исследуем на экстремумы функцию L на поверхности. Поверхность задана параметрически через $u_1, \ldots, u_n$. Соответственно мы переводим функцию L в функцию $H(u) = L\big(x(u)\big)$ и задача свелась к обычной задаче локального экстремума.

Можно было поступить аналогично и с функцией $f \Rightarrow$ получить функцию $f\big(x(u)\big)$, но проблема будет в том, что мы не знаем что из себя представляет второй дифференциал $(d_u^2 f)$ такой функции, так как там будут вторые производные координатных отображений (опять же наглядная демонстрация неинвариантности второго дифференциала).

Поэтому использование функции L с заменой координат дает нужный результат, поскольку есть понимание что $d_x^2 L = d_u^2 L$.