Лекция 4

Методы поиска экстремума

Классификация методов математического программирования

В САПР основными методами оптимизации являются <u>*поисковые методы*</u>.

Поисковые методы основаны на ***пошаговом изменении*** управляемых параметров

$$X_{k+1} = X_k + \Delta X_k$$

В большинстве методов приращение ΔX_k вектора управляемых параметров вычисляется по формуле

$$\Delta X_k = h\, g(X_k)$$

X_k - значение вектора управляемых параметров на *k*-м шаге,

h - шаг,

$g(X_k)$ - направление поиска.

Следовательно, если выполняются *условия сходимости*, то реализуется пошаговое (*итерационное*) приближение к экстремуму.

Методы оптимизации классифицируют по ряду признаков:

В зависимости от числа управляемых параметров различают методы *одномерной* (*управляемый параметр единственный*) и *многомерной* (*размер вектора* X *не менее двух*) *оптимизации*.

Реальные задачи в САПР многомерны,
методы одномерной оптимизации играют вспомогательную роль на отдельных этапах многомерного поиска.

Различают методы *условной* и *безусловной оптимизации* по наличию или отсутствию ограничений.

Для реальных задач характерно наличие ограничений.

Однако задачи условной оптимизации с помощью специальных методов могут быть сведены к задачам без ограничений

(к безусловной оптимизации).

В зависимости от числа экстремумов различают задачи одно- и многоэкстремальные.

Локальный метод **ориентирован на определение какого-либо локального экстремума.**

Метод глобального поиска **– метод, результатом которого является глобальный экстремум.**

Удовлетворительные по вычислительной эффективности методы глобального поиска для общего случая отсутствуют и потому на практике в САПР используют методы поиска локальных экстремумов.

Методы нескольких порядков различают по использованию при поиске **производных целевой функции** по управляемым параметрам.

Если производные не используются, то имеет место метод ***нулевого порядка***, если используются первые или вторые производные, то соответственно метод ***первого*** или ***второго порядка***.

Методы первого порядка называют также <u>***градиентными***</u>, поскольку вектор первых производных $F(\mathbf{X})$ по $\mathbf{N}$ есть градиент целевой функции

$$grad\left(F\left(X\right)\right)=\left(\partial F/\partial x_1,\partial F/\partial x_2,\ldots,\partial F/\partial x_n\right)$$

Конкретные методы определяются следующими факторами:

1) ***способом вычисления*** направления поиска $g(X_k)$ в формуле $\Delta X_k = h\,\mathrm{g}\left(X_k\right)$;

2) ***способом выбора*** шага h;

3) ***способом определения*** окончания поиска.

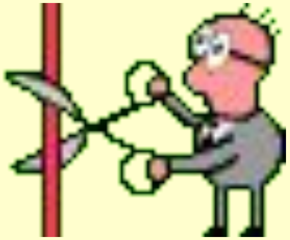
Определяющим фактором является первый.

Шаг может быть постоянным или выбираться исходя из одномерной оптимизации — поиска минимума целевой функции в выбранном направлении $g(X_k)$.

В последнем случае шаг будем называть оптимальным.

Правило окончания поиска:

если на протяжении r подряд идущих шагов траектория поиска остается в малой ε-окрестности текущей точки поиска X_k, то поиск следует прекратить.



<u>Условие окончания</u> поиска :

$$\left| X_k - X_{k-r} \right| < \varepsilon$$

Необходимые условия экстремума

В задачах <u>**безусловной оптимизации**</u> ***необходимые условия*** **представляют собой** <u>***равенство нулю градиента целевой функции***</u>

$$grad\ F(X) = 0$$

Базовая (общая) задача оптимизации ставится как задача математического программирования:

$$\underset{\mathbf{X}\in\mathbf{D}_x}{\mathbf{extr}}\, F\left(\mathbf{X}\right)$$

$$\mathbf{D}_x = \left\{\mathbf{X} \middle| \varphi\left(\mathbf{X}\right) > 0, \boldsymbol{\psi}\left(\mathbf{X}\right) = 0\right\}$$

где $F(X)$ — целевая функция, X — вектор управляемых (проектных) параметров, $\varphi(X)$ и $\psi(X)$ — функции-ограничения, D_x —допустимая область в пространстве управляемых параметров.

В общей задаче математического программирования необходимые условия экстремума (условия Куна-Таккера) формулируются:

для того, чтобы точка Q была экстремальной точкой выпуклой задачи математического программирования (ЗМП), необходимо наличие неотрицательных коэффициентов u_i, таких, что

$$u_i\,\varphi_i(\ ) = 0,\quad i = 1, 2, \ldots, m$$

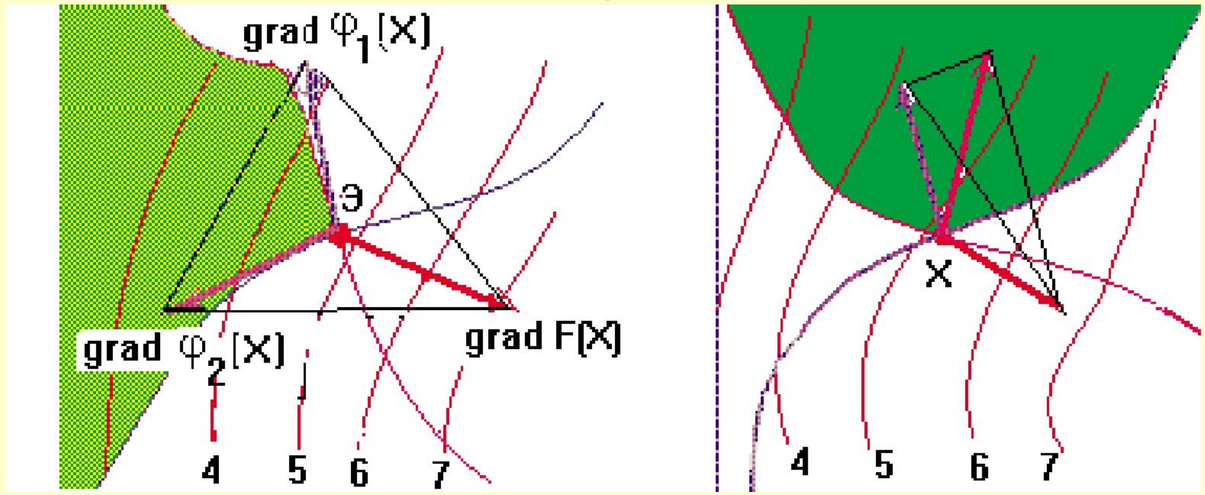
и при этом соблюдались ограничения задачи, а также выполнялось условие

$$\mathrm{grad}\,F(u) + \sum_{i=1}^{m} u_i\,\mathrm{grad}\,\varphi_i(a) + \sum_{j=1}^{L} a_j\,\mathrm{grad}\,\psi_j(\ ) = 0$$

где m — число ограничений типа неравенств, L— то же равенств, коэффициенты a_j>0.

Абстрактная формулировка условий имеет простой геометрический смысл

Рассмотрим случай с ограничениями только типа неравенств.



Если максимум находится внутри допустимой области R, то, выбирая все u_i=0, добиваемся выполнения $u_i\,\varphi_i(X)=0,\ i=1,2,\ldots,m$

если же точка максимума Q лежит на границе области R, то, как видно из левой части рисунка, эту точку всегда соответствующим подбором неотрицательных u_i можно поместить внутрь оболочки, натянутой на градиенты целевой функции $F(X)$ и функций-ограничений $\varphi_i(X)$.

Наоборот, если точка не является экстремальной, то условие

$$u_i \varphi_i(\ ) = 0, \quad i = 1, 2, \ldots, m,$$

нельзя выполнить при любом выборе положительных коэффициентов u_i (см. правую часть рисунка, где рассматриваемая точка N лежит вне выпуклой оболочки, натянутой на градиенты).

Учет ограничений *типа равенств* очевиден, если добавляется сумма

$$\sum_{j=1}^{L} a_j \psi_j(\ )$$

Методы поиска условных экстремумов

Широко известен ***метод множителей Лагранжа***, ориентированный на поиск экстремума при наличии ограничений типа равенств

$\psi(X) = 0$,

т.е. на решение задачи

$$\underset{\mathbf{X}\in\mathbf{R}}{\mathbf{extr}}\, F(\mathbf{X}),$$

где $\mathbf{R} = \{\mathbf{X} \mid \psi(\mathbf{X}) = 0\}$

<u>Суть метода</u> заключается в преобразовании задачи условной оптимизации в задачу безусловной оптимизации с помощью образования новой целевой функции

$$\Phi(\mathbf{X},\boldsymbol{\Lambda}) = F(\mathbf{X}) + \sum_{i=1}^{L} \lambda_i \psi_i(\mathbf{X})$$

где $\Lambda = (\lambda_1, \lambda_2, \ldots \lambda_n)$ **- вектор множителей Лагранжа,**
L **- число ограничений.**

Необходимые условия экстремума функции $\Phi(X)$:

$$\partial\Phi(\mathbf{X},\mathbf{\Lambda})/\partial\mathbf{X} = \partial F(\mathbf{X})/\partial\mathbf{X} + \sum \lambda_i \partial\psi_i(\mathbf{X})/\partial\mathbf{X} = 0;$$

$$\partial\Phi(\mathbf{X},\mathbf{\Lambda})/\partial\mathbf{\Lambda} = \psi(\mathbf{X}) = 0.$$

Система содержит** $n+L$ **алгебраических уравнений, где** n **- размерность пространства управляемых параметров. Её решение - искомые координаты экстремальной точки и значения множителей Лагранжа.

При численном решении системы (алгоритмические модели) возникают трудности. Поэтому в САПР основными методами решения ЗМП являются

методы штрафных функций и проекции градиента.

Основная идея *методов штрафных функций* — преобразование задачи условной оптимизации в задачу безусловной оптимизации путем формирования новой целевой функции $\Phi(\boldsymbol{N})$ введением в исходную целевую функцию $F(\boldsymbol{X})$ специальным образом выбранной функции штрафа $S(\boldsymbol{X})$:

$$\Phi(N) = F(X) + r\,S(X)$$

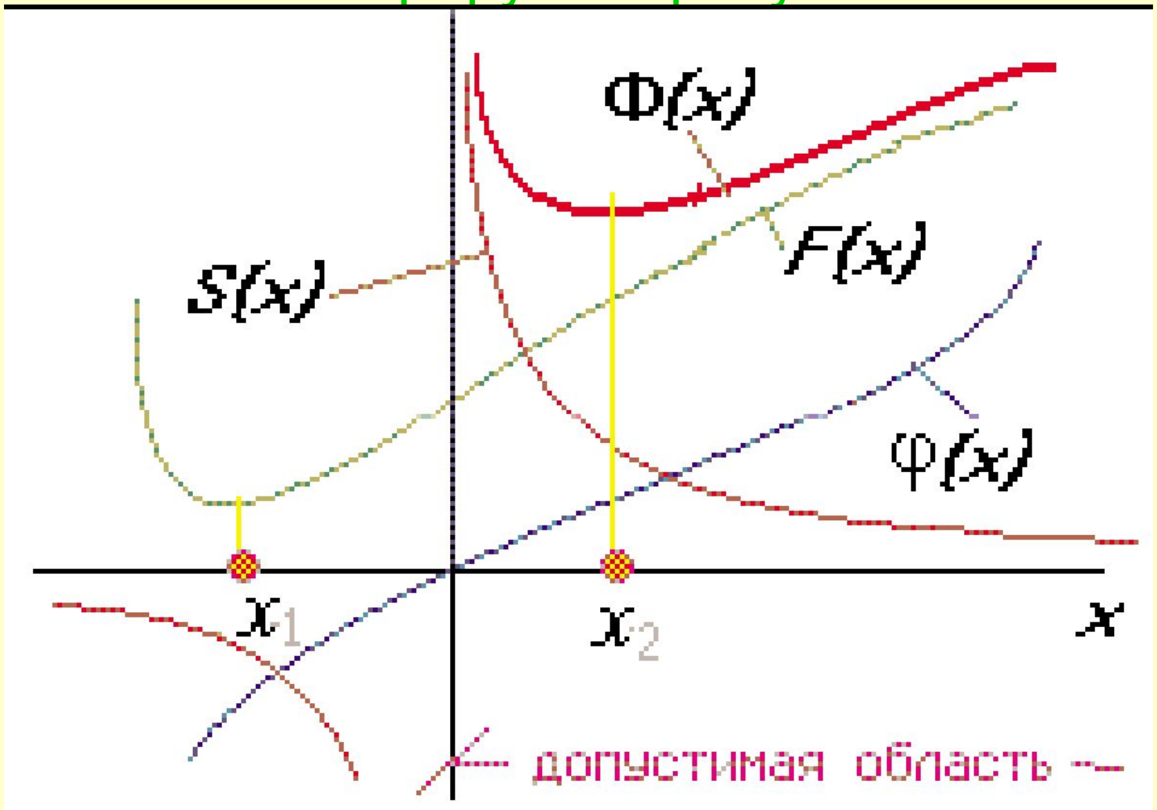
где *r* - множитель, значения которого можно изменять в процессе оптимизации.

Среди методов штрафных функций различают *методы внутренней и внешней точки*.

Согласно <u>методам внутренней точки</u> (методам *барьерных функций*) исходную для поиска точку можно выбирать только внутри допустимой области, а для <u>методов внешней точки</u> как внутри, так и вне допустимой области (в ней функции целевая и ограничений должны быть определены).

Ситуация появления барьера у целевой функции $\Phi(x)$ и соотношение между условным в точке x_2 и безусловным в точке x_1 минимумами $F(x)$ в простейшем одномерном случае

иллюстрируется рисунком



Примеры штрафных функций:

1) для метода внутренней точки при ограничениях $\varphi_i(X) > 0$

$$S(X) = \sum_{i=1}^{m} \left(1/\phi_i(\mathrm{X})\right),$$

где m - число ограничений типа неравенств;

2) для метода внешней точки при таких же ограничениях

$$S(X) = \sum_{i=1}^{m} \left(\min\{0, \varphi_i(X)\}\right)^2$$

здесь штраф сводится к включению в $\Phi(\mathbf{N})$ суммы квадратов активных (т.е. нарушенных) ограничений;

3) в случае ограничений типа равенств $\psi_i(X) = 0$

$$S(X) = \sum_{i=1}^{L} \left(\psi_i(\mathrm{X})\right)^2$$

Чем больше коэффициент r, тем точнее решение задачи, однако при больших r может ухудшаться ее обусловленность.

Поэтому в начале поиска обычно выбирают умеренные значения r, увеличивая их в окрестностях экстремума.

Основной вариант
метода проекции градиента
ориентирован на задачи математического программирования с ограничениями типа равенств.

Поиск при выполнении ограничений осуществляется в подпространстве $(n\text{-}m)$ измерений,

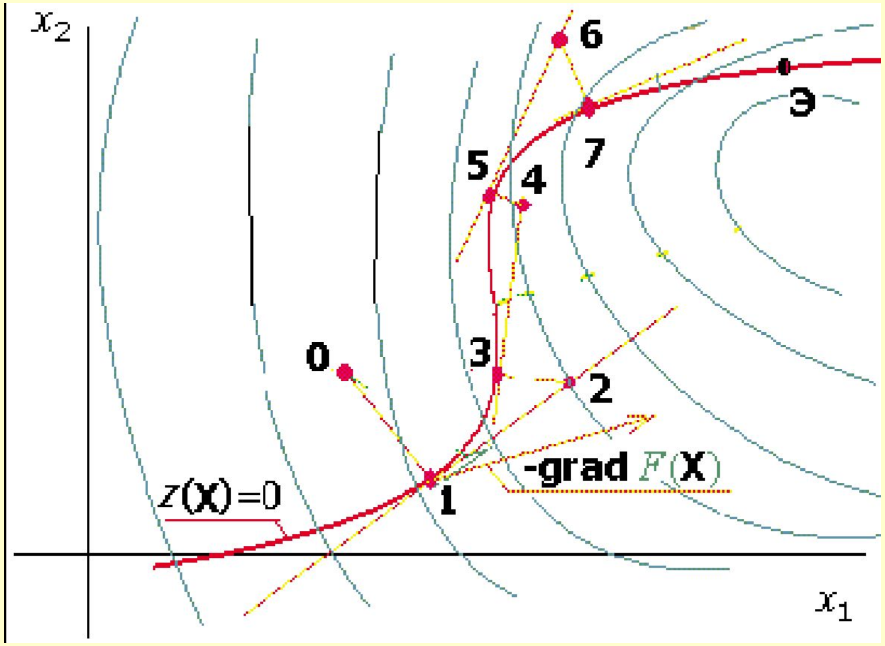
где n - число управляемых параметров,

m - число ограничений.

При этом движение осуществляется в направлении проекции градиента целевой функции $F(\boldsymbol{X})$ на гиперплоскость, касательную к гиперповерхности ограничений (точнее к гиперповерхности пересечения гиперповерхностей ограничений).

Поиск минимума начинают со спуска из исходной точки на гиперповерхность ограничений. Далее выполняют шаг в указанном выше направлении (шаг вдоль гиперповерхности ограничений). Поскольку этот шаг может привести к заметному нарушению ограничений, вновь повторяют спуск на гиперповерхность ограничений и т.д.

Идею метода поясним для случая поиска в двумерном пространстве при одном ограничении $\psi(\boldsymbol{X}) = 0$.



На рисунке это ограничение представлено жирной линией, а целевая функция — совокупностью более тонких линий равного уровня.

Спуск обычно осуществляют по нормали к гиперповерхности ограничений (в данном случае к линии ограничения).

Условие окончания поиска основано на сопоставлении значений целевой функции в двух последовательных точках, получаемых после спуска на гиперповерхность ограничений.

Рассмотрим получение аналитических выражений для направлений спуска и движения вдоль гиперповерхности ограничений.

Спуск

Необходимо из текущей точки поиска **B** попасть в точку **A**, являющуюся ближайшей к **B** точкой на гиперповерхности ограничений, т.е. решить задачу:

min |**B-A**|

при условии $\psi(\mathbf{X})=0$, которое после линеаризации в окрестностях точки **B** имеет вид

$$\psi(\mathbf{B}) + \left(\mathbf{grad}\psi(\mathbf{B})\right)^{\mathrm{T}}(\mathbf{A} - \mathbf{B}) = 0$$

Используем метод множителей Лагранжа, обозначая **A-B=U** и учитывая, что минимизация расстояния равнозначна минимизации скалярного произведения **U** на **U**, получаем

$$\Phi(\mathbf{A}) = \mathbf{U}^{\mathrm{T}}\mathbf{U} + \lambda\left(\psi(\mathbf{B}) + (\mathbf{grad}\psi(\mathbf{B}))^{\mathrm{T}}\mathbf{U}\right)$$

$$\partial\Phi/\partial\mathbf{A} = 2\mathbf{U} + \lambda(\mathbf{grad}\psi(B)) = 0$$

$$\partial\Phi/\partial\lambda = \psi(\mathbf{B}) + (\mathbf{grad}\psi(B))^{T}\mathbf{U} = 0$$

тогда из второго выражения получаем

$$\mathbf{U} = -0{,}5\lambda(\mathbf{grad}\psi(\mathbf{B}))$$

подставляя его в третье выражение, имеем

$$\psi(\mathbf{B}) - 0{,}5\lambda(\mathbf{grad}\psi(\mathbf{B}))^{\mathrm{T}}\mathbf{grad}\psi(\mathbf{B}) = 0$$

откуда $\lambda = \left(0{,}5(\mathbf{grad}\psi(\mathbf{B}))^{\mathrm{T}}\mathbf{grad}\psi(\mathbf{B})\right)^{-1}\psi(\mathbf{B})$

Подставляя λ во второе выражение, находим

$$\mathbf{U} = -\mathbf{grad}\psi(\mathbf{B})\left((\mathbf{grad}\psi(\mathbf{B}))^{\mathrm{T}}(\mathbf{grad}\psi(\mathbf{B}))\right)^{-1}\psi(\mathbf{B})$$

Движение вдоль гиперповерхности ограничений

Шаг в гиперплоскости **D**, касательной к гиперповерхности ограничений, следует сделать в направлении вектора **S**, на котором целевая функция уменьшается в наибольшей мере при заданном шаге *h*.

Уменьшение целевой функции при переходе из точки **A** в новую точку **C** подсчитывают, используя формулу линеаризации *F*(**X**) в окрестностях точки **A**:

$$F(\mathbf{C}) - F(\mathbf{A}) = h(\mathbf{grad}F(\mathbf{A}))^{\mathrm{T}}\,\mathbf{S}$$

где $\mathrm{grad}\,F(A)^T S$ - приращение *F*(**X**), которое нужно минимизировать, варьируя направления **S**.

$$\min F(\mathbf{C}) = \min\left(\left(\mathbf{grad}F(\mathbf{A})\right)^T \mathbf{S}\right)$$

где вариация **S** осуществляется в пределах гиперплоскости **D**; **grad** $\psi(A)$ и **S** — ортогональные векторы.

Следовательно, минимизацию этого выражения необходимо выполнять при ограничениях

$$\left(\mathbf{grad}\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{S} = 0$$

$$\mathbf{S}^{\mathrm{T}}\mathbf{S} = 1$$

Последнее ограничение говорит о том, что при поиске направления движения, вектор **S** должен лишь указывать это направление, т.е. его длина несущественна (пусть **S** — единичный вектор).

Для решения

$$\min F(\mathrm{C}) = \min\left(\left(\mathrm{grad}F(\mathrm{A})\right)^T \mathrm{S}\right)$$

используем метод множителей Лагранжа.

$$\Phi(\mathbf{S}, \lambda, q) = \left(\mathbf{grad}F(\mathbf{A})\right)^{\mathrm{T}} \mathbf{S} + \lambda\left(\mathbf{grad}\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{S} + q\left(\mathbf{S}^{\mathrm{T}}\mathbf{S} - 1\right)$$

где λ и q — множители Лагранжа;

$$\partial\Phi/\partial\mathbf{S} = \mathbf{grad}F(\mathbf{A}) + \lambda\left(\mathbf{grad}\psi(\mathbf{A})\right) + q\mathbf{S} = 0$$

$$\partial\Phi/\partial\lambda = \left(\mathbf{grad}F(\mathbf{A})\right)^{\mathrm{T}} \mathbf{S} = 0$$

$$\partial\Phi/\partial q = \mathbf{S}^{\mathrm{T}}\mathbf{S} - \mathbf{1} = 0$$

Из второго выражения следует, что

$$\mathbf{S} = -\left(\mathbf{grad}F(\mathbf{A}) + \lambda\mathbf{grad}\psi(\mathbf{A})\right)/q$$

подставляя **S** в третье выражение, получаем

$$\left(\mathbf{grad}\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}F(\mathbf{A}) + \lambda\left(\mathbf{grad}\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\psi(\mathbf{A}) = 0$$

откуда

$$\lambda = -\left[\left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,\psi(\mathbf{A})\right]^{-1} \left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,F(\mathbf{A}),$$

$$\mathbf{S} = -\left\{\mathbf{grad}\,F(\mathbf{A}) - \mathbf{grad}\,\psi(\mathbf{A})\left[\left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,\psi(\mathbf{A})\right]^{-1} \left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,F(\mathbf{A})\right\} \Big/ q =$$

$$= -\left\{\mathbf{E} - \mathbf{grad}\,\psi(\mathbf{A})\left[\left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,\psi(\mathbf{A})\right]^{-1} \left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}}\right\} \mathbf{grad}\,F(\mathbf{A}) \Big/ q\,.$$

Таким образом, матрица

$$P = \mathbf{E} - \mathbf{grad}\,\psi(\mathbf{A})\left[\left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}} \mathbf{grad}\,\psi(\mathbf{A})\right]^{-1} \left(\mathbf{grad}\,\psi(\mathbf{A})\right)^{\mathrm{T}}$$

представляет собой проектирующую матрицу, а вектор **S**, рассчитанный по верхнему выражению, — проекцию градиента $\mathbf{grad}\,F(\mathbf{A})$ на гиперповерхность ограничений.

Частным случаем применения метода проекции градиента являются <u>задачи оптимизации с максиминным критерием</u>. Для поиска экстремума функции минимума

$$\max_{\mathbf{X}} \min_{j} Z_j(\mathbf{X})$$

где Z_j — нормированная величина j-го выходного параметра y_j,

удобно применять метод проекции градиента.

В качестве ограничений задачи в исходной постановке фигурируют только прямые ограничения

$$x_{\max i} < x_i < x_{\min i}$$

Здесь $x_{\max i}$ и $x_{\min i}$ — граничные значения допустимого диапазона варьирования параметра x_i.

В процессе поиска, если минимальной является функция $Z_q(\mathbf{X})$ и траектория поиска пересекает гребень

$$Z_q(\mathbf{X}) - Z_k(\mathbf{X}) = 0$$

то поиск продолжается в направлении проекции градиента функции $Z_q(\mathbf{X})$ на гиперповерхность этого гребня.

Спасибо за внимание!