Численная оптимизация

Функции многих переменных: условная оптимизация

26 ноября 2012 г.

$$
\begin{gathered}
f(x) \to \min, \quad x \in \mathbb{R}^n \\
g_j(x) \geqslant 0, \quad j = 1, \ldots, J \\
h_k(x) = 0, \quad k = 1, \ldots, K \\
x_i^{(l)} \leqslant x_i \leqslant x_i^{(u)}, \quad i = 1, \ldots, N
\end{gathered}
$$

Методы на основе преобразования задачи

Штрафная функция

$$P(x, R) = f(x) + \Omega(R, g(x), h(x))$$

R — набор штрафных параметров
Ω — штраф

Методы внешней точки: $x^{(i)}$ — недопустимые точки.
Методы внутренней точки (барьерные методы): $x^{(i)}$ — допустимые точки.

Квадратичный:

$$\Omega = R \cdot h(x)^2.$$

«Бесконечный»:

$$\Omega = 10^{20} \sum_{j \in \bar{J}} |g_j(x)|, \quad \bar{J} = \big\{ j \in \{1, \ldots, J\} \mid g_j(x) < 0 \big\}.$$

Логарифмический:

$$\Omega = -R \cdot \ln\big(g(x)\big).$$

Штраф, задаваемый обратной функцией:

$$\Omega = R \cdot g(x)^{-1}.$$

Штраф типа квадрата срезки:

$$\Omega = R \cdot \big\langle g(x) \big\rangle, \quad \langle \alpha \rangle = \begin{cases} \alpha^2, & \alpha < 0, \\ 0, & \alpha \geqslant 0. \end{cases}$$

$$f = (x_1 - 4)^2 + (x_2 - 4)^2 \to \min, \quad h = x_1 + x_2 - 5 = 0$$

Штрафная функция:

$$P = f + R \cdot h^2$$

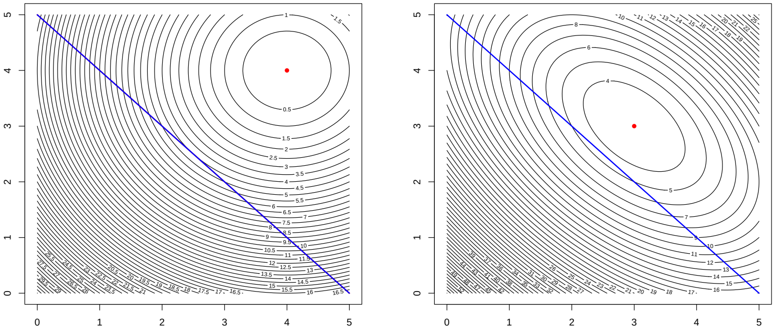
Решением системы уравнений

$$\frac{\partial P}{\partial x_1} = 0, \quad \frac{\partial P}{\partial x_2} = 0$$

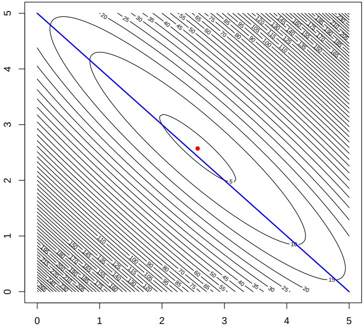
будет

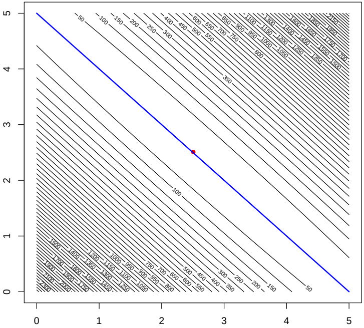
$$x_1 = \frac{4 + 5R}{1 + 2R}, \quad x_2 = \frac{4 + 5R}{1 + 2R}$$

$R = \{0, 1, 10, 100\}$



$$R = \{0, 1, 10, 100\}$$





Методы на основе преобразования задачи

Пример: квадратичный штраф

$$(x_1 - 4)^2 + (x_2 - 4)^2 \to \min \qquad x_1 + x_2 - 5 = 0$$

```
f <- function(x) sum((x-c(4,4))^2) # f → min
h <- function(x) x[1]+x[2]-5       # h = 0

R <- 0
P <- function(x) f(x)+R*h(x)^2     # штрафная функция

x <- c(0,0)
X <- x
for (R in c(0,0.1,1,10,1000,10^6,10^15))
{
  x <- optim(x,P,method="BFGS")$par
  X <- cbind(X,x)
}
```

0 4 3.75 3 2.571429 2.50075 2.500001 2.5

0 4 3.75 3 2.571429 2.50075 2.500001 2.5

$$(x_1 - 4)^2 + (x_2 - 4)^2 \to \min, \quad 5 - (x_1 + x_2) \geqslant 0$$

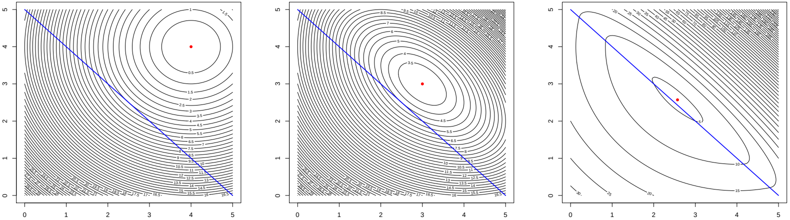
```
f <- function(x) sum((x-c(4,4))^2)
h <- function(x) 5-(x[1]+x[2])

R <- 0
P <- function(x) f(x)+R*ifelse(h(x)>=0,0,h(x)^2)

x <- c(0,0)
X <- x
for (R in c(0,0.1,1,10,1000,10^6,10^15))
{
  x <- optim(x,P,method="BFGS")$par
  X <- cbind(X,x)
}
```

```
0 4 3.75 3 2.571429 2.50075 2.499845 2.499845
0 4 3.75 3 2.571429 2.50075 2.499845 2.499845
```

Методы на основе преобразования задачи

Пример: штраф типа квадрата срезки

$$R = \{0.1, 1, 10\}$$



$$(x_1 - 4)^2 + (x_2 - 4)^2 \to \min, \quad 5 - (x_1 + x_2) \geqslant 0$$

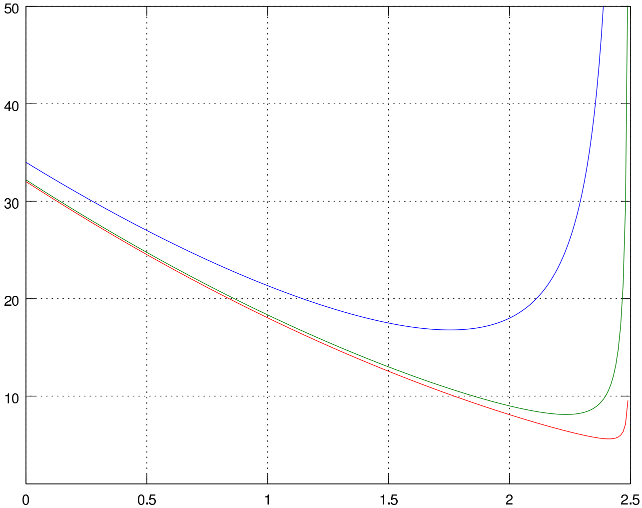
```
f <- function(x) sum((x-c(4,4))^2)
h <- function(x) 5-(x[1]+x[2])

R <- 0
P <- function(x) f(x)+R/h(x)

x <- c(0,0)
X <- x
for (R in c(1000,100,10,1,0.1,0.001))
{
  x <- optim(x,P,method="BFGS")$par
  X <- cbind(X,x)
}
```

```
0 -2.046769 2.500425 2.500425 2.500425 2.500425 3.999944
0 -2.046769 2.500425 2.500425 2.500425 2.500425 3.999944
```

Методы на основе преобразования задачи

Пример: штраф, задаваемый обратной функцией

$R = \{10, 1, 0.1\}$



Функция Лагранжа:

$$L(x,u,v) = f(x) - \sum_j u_j g_j(x) - \sum_k v_k h_k(x)$$

Если определённым образом модифицировать функцию Лагранжа, то полученная в результате функция будет достигать своего *безусловного минимума* в точке Куна–Таккера исходной задачи.

Модифицированная функция Лагранжа

$$P(x, \sigma^{(i)}, \tau^{(i)}) = f(x) + R\sum_{j=1}^{J}\left(\left\langle g_j(x) + \sigma_j^{(i)}\right\rangle^2 - \left(\sigma_j^{(i)}\right)^2\right) + \\ + R\sum_{k=1}^{K}\left(\left(h_k(x) + \tau_k^{(i)}\right)^2 - \left(\tau_k^{(i)}\right)^2\right)$$

$x^{(i)}$ — точка минимума $P(x, \sigma^{(i)}, \tau^{(i)})$
R — *постоянный* весовой (нормирующий) коэффициент

Пересчёт множителей при переходе к $(i+1)$-й итерации:

$$\sigma_j^{(i+1)} = \langle g_j(x) + \sigma_j^{(i)} \rangle, \quad j = 1, \ldots, J,$$
$$\tau_k^{(i+1)} = h_k(x) + \tau_k^{(i)}, \quad k = 1, \ldots, K$$

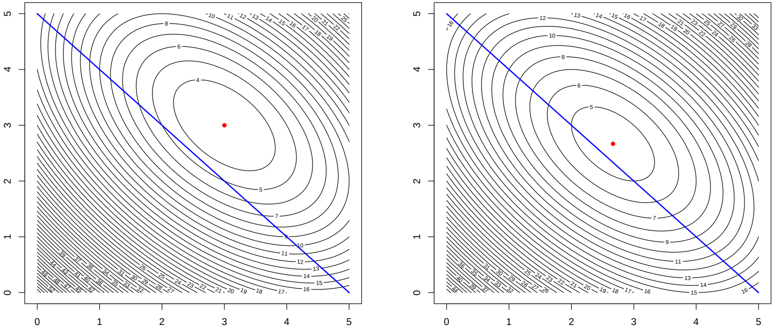
$$(x_1 - 4)^2 + (x_2 - 4)^2 \to \min, \quad x_1 + x_2 - 5 = 0$$

```
f <- function(x) sum((x-c(4,4))^2)
h <- function(x) x[1]+x[2]-5

x <- c(0,0);  X <- x

R <- 1; tau <- 0
P <- function(x) f(x)+R*((h(x)+tau)^2-tau^2)
for (i in 1:7)
{
  x <- optim(x,P,method="BFGS")$par
  X <- cbind(X,x)
  tau <- h(x)+tau
}
```

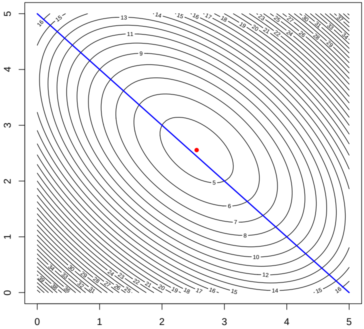
```
0 3 2.666667 2.555556 2.518519 2.506173 2.502058 2.500686
0 3 2.666667 2.555556 2.518519 2.506173 2.502058 2.500686
```

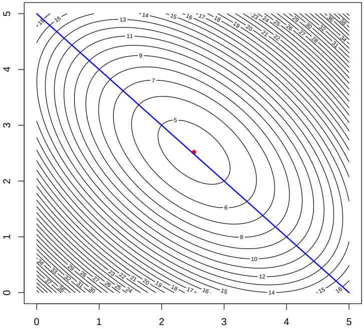
Методы на основе преобразования задачи

Пример



Пример





Методы прямого поиска

Подготовка задачи к решению:

1. Исключение ограничений в виде равенств.
2. Определение начальной допустимой точки.

Метод комплексов

1. Сгенерировать $P \approx 2N$ допустимых точек.
2. Отразить точку с наибольшим значением целевой функции через центр тяжести остальных точек ($\alpha = 1.3$).
3. Сделать отражённую точку допустимой.
   1. Новая точка допустимая, и значение целевой функции в ней не совпадает с максимальным значением всей совокупности точек.
   2. Новая точка допустимая, и значение целевой функции в ней совпадает с максимальным значением всей совокупности точек: передвинуть точку на половину расстояния до ранее найденного центра тяжести.
   3. Новая точка недопустимая: уменьшаем в два раза расстояние до вычисленного раньше центра тяжести.
4. Процедура поиска продолжается, пока многогранник не будет стянут в центр тяжести в пределах заданной точности и/или пока разница значений функции в вершинах не станет достаточно малой.

Методы случайного поиска

Прямые выборочные процедуры:

- метод совместного поиска;
- метод серий (выборки с уменьшением интервала).

Комбинированные эвристические процедуры:

- адаптивный алгоритм случайного поиска с переменным шагом;
- комбинаторный эвристический алгоритм.

Выборка с уменьшением интервала

Дано: начальное допустимое решение x^0; оценка величины начального интервала l^0, где $x^0 - l^0/2 \leqslant x \leqslant x^0 + l^0/2$; параметр ε, определяющий уменьшение интервала $(0 < \varepsilon < 1)$. Q — количество серий, P — количество точек в серии. $q = 1$ — номер серии.

1. Для $i = 1, \ldots, N$ вычислить $x_i^p = x_i^{q-1} + r l_i^{q-1}$, где $r \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.
2.
   - x^p — недопустимая точка и $p < P$: повторить шаг 1.
   - x^p — допустимая точка: запомнить x^p и $f(x^p)$ и повторить шаг 1.
   - $p = P$: найти среди всех допустимых точек x^p точку с наименьшим значением $f(x^p)$ и положить её равной x^q.
3. Уменьшить интервал, полагая $l_i^q = (1 - \varepsilon) l_i^{q-1}$.
4. Если $q > Q$, закончить вычисления. В противном случае увеличить q и продолжать вычисления, начиная с шага 1.

Адаптивный алгоритм случайного поиска с переменным шагом

Даны параметры $\alpha_s = 1.618$, $\alpha_f = 0.618$ и $M = 3N$ и начальная допустимая точка x^0. Начальная величина шага α полагается равной 1, а m — число испытаний, не дающих улучшения, — принимается равным 0.

1. Получить случайный вектор d единичной длины и положить $x^1 = x^0 + \alpha d$.
2. Если x^1 — допустимая точка и $f(x^1) < f(x^0)$, положить $y = x^0 + \alpha_s(x^1 - x^0)$ и перейти к шагу 3. В противном случае принять $m = m + 1$ и перейти к шагу 4.
3. Если y — допустимая точка и $f(y) < f(x^0)$, положить $\alpha = \alpha_s \alpha$, $x^1 = y$ и перейти к шагу 5. В противном случае перейти к шагу 1.
4. Если $m > M$, положить $\alpha = \alpha_f \alpha$, $m = 0$ и перейти к шагу 5. В противном случае сразу перейти к шагу 5.
5. Перейти к шагу 1, если не выполнено условие окончания вычислений.

Комбинаторный эвристический алгоритм

1. Построить случайную допустимую начальную точку x^0 и положить $F_{\min} = f(x^0)$. Для каждой координаты i, $i = 1, \ldots, N$, выполнить следующую последовательность вычислений.
2. Провести оптимизацию по i-й координате, зафиксировав остальные.
    1. Выбрать случайным образом возможные значения i-й координаты для нахождения q дополнительных допустимых точек с лучшим значением целевой функции по сравнению с текущей базовой точкой. Если такие точки получить не удаётся, повторить шаг 2 для координаты $i + 1$.
    2. Определить наилучшее из q допустимых решений и положить значение целевой функции $T_{\min}$.
    3. Произвести «упреждающий» поиск.
        1. Для каждого из q допустимых решений, найденных на шаге 2.1, провести случайных выбор одного из q возможных значений координаты $(i + 1)$ для определения допустимого значения этой координаты, дающего лучшее значение целевой функции по сравнению с $T_{\min}$.
        2. Выбрать наилучшую из q допустимых точек. Зафиксировать значение координаты i, соответствующее этой точке как оптимальное.
    4. Если $i = N$, перейти к шагу 3. В противном случае выполнить шаг 2 для координаты $(i + 1)$.
3. Провести случайный поиск для определения наилучшего значения координаты N при фиксированных значениях других координат, соответствующих текущим базовым точкам. Найденную точку принять на новую базисную точку, а значение целевой функции в ней — за новое значение $F_{\min}$.
4. Перейти к шагу 2 с $i = 1$, если не выполнены условия окончания вычислений.

Функция $f(x)$ аппроксимируется *линейной* функцией

$$\tilde{f}(x; x^0) = f(x^0) + \nabla f(x^0) \cdot (x - x^0).$$

x^0 — точка линеаризации

Алгоритм Франка–Вульфа

1. Вычислить $\nabla f(x^{(i)})$. Если $\|\nabla f(x^{(i)})\| < \varepsilon$, то прекратить вычисления.
2. Решить задачу линейного программирования:
$$\nabla f(x^{(i)})y \to \min, \qquad Ay \leqslant b, \quad y \geqslant 0.$$
3. $\alpha^{(i)} = \arg \min\limits_{0 \leqslant \alpha \leqslant 1} f\left(x^{(i)} + \alpha \cdot (y^{(i)} - x^{(i)})\right)$
4. $x^{(i+1)} = x^{(i)} + \alpha^{(i)} \cdot (y^{(i)} - x^{(i)})$
5. Прекратить вычисления, если
$$\|x^{(i+1)} - x^{(i)}\| \leqslant \delta \|x^{(i+1)}\|,$$
$$\left\|f(x^{(i+1)}) - f(x^{(i)})\right\| \leqslant \varepsilon \left\|f(x^{(i+1)})\right\|.$$

Вместо

$$
\begin{gathered}
f(x) \to \min, \\
g_j(x) \geqslant 0, \quad j = 1, \dots, J \\
h_k(x) = 0, \quad k = 1, \dots, K \\
x_i^{(l)} \leqslant x_i \leqslant x_i^{(u)}, \quad i = 1, \dots, N
\end{gathered}
$$

на каждом шаге решается

$$
\begin{gathered}
f(x^{(i)}) + \nabla f(x^{(i)})(x - x^{(i)}) \to \min, \\
g_j(x^{(i)}) + \nabla g_j(x^{(i)})(x - x^{(i)}) \geqslant 0, \quad j = 1, \dots, J \\
h_k(x^{(i)}) + \nabla h_k(x^{(i)})(x - x^{(i)}) = 0, \quad k = 1, \dots, K \\
x_i^{(l)} \leqslant x_i \leqslant x_i^{(u)}, \quad i = 1, \dots, N
\end{gathered}
$$

Метод допустимых направлений

$$f(x) \to \min, \qquad g_j(x) \geqslant 0, \quad j = 1, \ldots, J$$

$x^{(1)}$ — начальная точка, удовлетворяющая ограничениям:

$$g_j(x^{(1)}) \geqslant 0, \quad j = 1, \ldots, J$$

- $g_j(x^{(1)}) = 0$ — связывающее (активное) ограничение
- $g_j(x^{(1)}) > 0$ — несвязывающее ограничение

Метод допустимых направлений

Вектор d определяет подходящее направление поиска, если:

- d — направление спуска:

$$\nabla f(x^{(1)})d < 0;$$

- точки луча

$$x(\alpha) = x^{(1)} + \alpha d, \quad \alpha \geqslant 0,$$

  являются допустимыми по крайней мере на небольшом расстоянии от $x^{(1)}$.

Метод допустимых направлений

Точки вдоль d будут допустимыми, если для всех связывающих в $x^{(1)}$ ограничений выполняется условие

$$g_j(x^{(1)}) + \nabla g_j(x^{(1)})(x - x^{(1)}) \geqslant 0.$$

Так как

$$g_j(x^{(1)}) = 0 \text{ и } x - x^{(1)} = \alpha d, \quad \alpha \geqslant 0$$

то

$$\nabla g_j(x^{(1)}) d \geqslant 0$$

для всех связывающих ограничений.

$x^{(t)}$ — допустимая точка, $I^{(t)}$ — множество индексов ограничений, активных в пределах заданной погрешности

$$I^{(t)} = \{j \mid 0 \leqslant g_j(x^{(t)}) \leqslant \varepsilon,\ j = 1, \ldots, J\}.$$

1. Решить задачу линейного программирования:

$$\theta \to \max$$
$$\begin{cases} \nabla f(x^{(t)})d \leqslant -\theta, & \\ \nabla g_j(x^{(t)})d \geqslant \theta, & j \in I^{(t)}, \\ -1 \leqslant d_i \leqslant 1, & i = 1, \ldots, n. \end{cases}$$

$d^{(t)}$ и $\theta^{(t)}$ — полученные решения.

2 Если $\theta^{(t)} \leqslant 0$, то закончить итерацию. Иначе найти

$$\bar{\alpha} = \min\{\alpha \mid g_j(x^{(t)} + \alpha d^{(t)}), j = 1, \dots, J, \text{ и } \alpha \geqslant 0\}.$$

Если не существует $\bar{\alpha} > 0$, то $\bar{\alpha} = \infty$.

3 Найти $\alpha^{(t)}$:

$$\alpha^{(t)} = \arg\min_{0 \leqslant \alpha \leqslant \bar{\alpha}} f(x^{(t)} + \alpha d^{(t)}).$$

$x^{(t+1)} = x^{(t)} + \alpha^{(t)} d^{(t)}$.

1. В точке $x^{(t)}$ при заданном $\varepsilon^{(t)} > 0$ определить $I^{(t)}$ и выполнить шаг 1 основного алгоритма.
2. - Если $\theta^{(t)} \geqslant \varepsilon^{(t)}$, положить $\varepsilon^{(t+1)} \geqslant \varepsilon^{(t)}$ и продолжать одномерный поиск как в основном методе.
   - Если $\theta^{(t)} < \varepsilon^{(t)}$, то положить $\varepsilon^{(t+1)} = \frac{1}{2}\varepsilon^{(t)}$ и продолжать одномерный поиск как в основном методе.
   - Если $\theta^{(t)} < 0$, то точка Куна–Таккера найдена.

Активные ограничения не используются.

$$\theta \to \max$$
$$\begin{cases} \nabla f(x^{(t)})d \leqslant -\theta, & \\ \nabla g_j(x^t)d \geqslant \theta, & j = 1, \ldots, J, \\ -1 \leqslant d_i \leqslant 1, & i = 1, \ldots, n. \end{cases}$$

$$f(x) \to \min, \quad a^\top x = b$$

$$\forall s \in \mathbb{R}^N : s = s' + s''$$

$$s' = \gamma a, \quad a^\top s'' = 0$$

$$a^\top s = a^\top(\gamma a) + a^\top s'' = a^\top(\gamma a)$$

$$\gamma = (a^\top a)^{-1} a^\top s$$

$$s'' = s - a(a^\top a)^{-1} a^\top s = \underbrace{\left(\mathsf{I} - a(a^\top a)^{-1} a^\top\right)}_{\mathsf{P}} s$$

P — проекционная матрица (симметричная и положительно определённая)

$$Ax = b \qquad a_k^\top x = b_k, \quad k = 1, \ldots, K$$

$$s = s' + s'' = A^\top \gamma + s''$$

A — матрица со строками $a_k^\top$

$$\gamma = (AA^\top)^{-1} As$$

$$s'' = \big(\mathsf{I} - A^\top (AA^\top)^{-1} A\big) s = \mathsf{P} s$$

Свойства проекций:

1. $s = -\mathsf{P} \cdot \nabla f$ — направление спуска.
2. Если $s = 0$, то точка $x^{(t)}$ удовлетворяет необходимым условиям Лагранжа.
3. Вектор множителей Лагранжа задаётся выражением

$$q = (AA^{\top})^{-1} A \cdot \nabla f.$$

Алгоритм

Вычислить P в предположении, что векторы a_k линейно независимы, заданы допустимая точка $x^{(t)}$ и допустимая погрешность сходимости $\varepsilon > 0$.

1. Вычислить $s^{(t)} = -\mathrm{P} \cdot \nabla f$.
2. Если $\|s^{(t)}\| \leqslant \varepsilon$, то вычислить v:
$$v = (AA^\top)^{-1} A \cdot \nabla f$$
и закончить вычисления. В противном случае продолжить вычисления.
3. Определить максимальную длину шага
$$\alpha_{\max} = \min \left\{ \max \left(0, \frac{b_k - a_k^\top x^{(t)}}{a_k^\top s^{(t)}} \text{ или } \infty, \text{ если } a_k^\top s^{(t)} = 0 \right); k = \overline{1, K} \right\}$$
4. Решить задачу одномерного поиска
$$f(x^{(t)} + \alpha s^{(t)}) \to \min, \quad 0 \leqslant \alpha \leqslant \alpha_{\max}.$$
5. Положить $x^{(t+1)} = x^{(t)} + \alpha_{opt} s^{(t)}$ и перейти к шагу 1.

$I^{(t)}$ — множество активных в точке $x^{(t)}$ ограничений

$w^{(t)}$ — недопустимая точка, достигнутая при перемещении из $x^{(t)}$ вдоль направления $s^{(t)}$

Направление, перпендикулярное к поверхностям ограничений:

$$s = A(w^{(t)})^\top \gamma$$

где столбцами $A(w^{(t)})^\top$ являются градиенты $\nabla h_k(w^{(t)})$, $k = 1, \dots, K$, а также $\nabla g_l(w^{(t)})$, $l \in I^{(t)}$

h — вектор ограничений–равенств и активных ограничений–неравенств

$$h(w^{(t)} + A^\top \gamma) = h(w^{(t)}) + A(w^{(t)})A(w^{(t)})^\top \gamma = 0$$

$$\nabla_\gamma h = \frac{dh}{dx}\frac{dx}{d\gamma}$$

$$\gamma = -(AA^\top)^{-1} h$$

$$w^{(t+1)} - w^{(t)} = -A^\top (AA^\top) h$$

Алгоритм

Дано: $x^{(0)}$ — допустимая точка, $\varepsilon_1 > 0$ — допустимая погрешность сходимости, ε_2 — погрешность определения активных ограничений

1. Определить активное множество в $x^{(t)}$
2. Вычислить P и $s = -\mathrm{P} \cdot \nabla f(x^t)$
3. Если $\|s^t\| > \varepsilon_1$, то перейти к шагу 4.
   В противном случае вычислить множители Лагранжа $(u, v) = (AA^\top)^{-1} A \cdot \nabla f$ и найти $u_m = \min\{u_l : l \in I^{(t)}\}$.
   Если $|u_m| \leqslant \varepsilon_1$, то закончить вычисления. В противном случае исключить ограничение m из $I^{(t)}$ и перейти к шагу 2.
4. Определить такую максимальную длину шага $\alpha_{\max}$, что $g_l\big(w(\alpha)\big) \geqslant 0$ для всех $l \notin I^{(t)}$.
5. - Используя ограничивающий поиск по α, $0 \leqslant \alpha \leqslant \alpha_{\max}$, определить α_1, α_2, α_3 и $w(\alpha_1)$, $w(\alpha_2)$ и $w(\alpha_3)$, которые локализуют минимум $f(x)$ на кривой $h\big(w(\alpha)\big) = 0$.
   - При помощи квадратичной интерполяции оценить α.