Методы оптимизации и их применение к решению обратных задач

Факультет Космических Исследований, МГУ

Москва

Основные обозначения и определения

Введем обозначения:

$$x^* = \underset{x \in X}{\operatorname{argmin}} f(x),\ X \subset H,\ (*)$$

$$f^* = f(x^*) = \min_{x \in X} f(x).$$

Теорема

Пусть X - компакт в H, тогда $f(x)$ - непрерывный X на функционал. Тогда существует точка глобального минимума $f(x)$ на X.

Выпуклость и сильная выпуклость

Выпуклость функции $f(x)$ означает, что $\forall x_1, x_2 \in X$, $\forall \lambda \in [0, 1]$:

$$f(\lambda)x_1 + (1-\lambda)x_2 \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

Сильная выпуклость, плюс к вышесказанному существует $\theta > 0$:

$$f(\lambda)x_1 + (1-\lambda)x^2 \leq \lambda f(x_1) + (1-\lambda)f(x_2) - \theta\lambda(1-\lambda)||x_1 - x_2||^2.$$

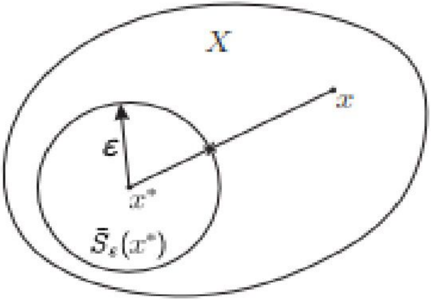
Теорема

Если $(*)$ - задача выпуклого программирования, то любая точка локального минимума является точкой глобального минимума.

Теорема

Пусть функционал $f(x)$ является выпуклым на всем пространстве H и дифференцируемым в точке $x^* \in H$. Если $f'(x) = 0$, то x^* - точка минимума $f(x)$ на H.

Локальный и глобальный минимум в случае выпуклых функций



Метод наименьших квадратов

$$\Phi(x) = ||Ax - b||^2 = (A^*Ax, x) - 2(A^*b, x) + (b, b),$$

легко получить

$$\Phi(x) = 2(A^*Ax - A^*b),$$

$$\Phi(x) = 2A^*A \geq 0.$$

Поскольку оператор A^*A - неотрицательно определен, то $\Phi(x)$ - выпулый функционал. Через равенство нулую градиента, получаем СЛАУ

$$A^*Ax = A^*b.$$

(достаточное условие минимума $\Phi'(x) = 0$, $\Phi''(x) > 0$).

Основной итерационный процесс

Определим последовательность:

$$x^{n+1} = x^n + \alpha_n h^n, \ n = 0, 1, 2..$$

Обозначим основные этапы алгоритма отпмизации:

1. Положить $n = 0$, задать x^0;
2. Проверить условия останова;
3. Вычислить α_n;
4. Вычислить x^{n+1};
5. Увеличить на единицу. Перейти к п. 2;

Методы нулевого/первого и более порядков.

Минимизирующая последовательность

Последовательность $\{x^n\}$ минимизирующая, если $f(x^n) \to f^* = \min_{x \in X} f(x)$ (сходимость по функционалу).

Сходимость по аргументу $x^n \to x^* = \operatorname{argmin}_{x \in X} f(x)$.

Различные виды сходимости

со скоростью геометрической прогрессии:

$$||x^{n+1} - x^*|| \leq q||x^n - x^*||, \; 0 \leq q \leq 1,$$

сверхлинейная сходимость:

$$||x^{n+1} - x^*|| \leq q_n||x^n - x^*||, \; q_n \to 0,$$

квадратичная скорость сходимости:

$$||x^{n+1} - x^*|| \leq C||x^n - x^*||^2$$

Критерии остановки

Могут применяться следующие критерии остановки процесса минимизации:

1. $$||x^{n+1} - x^*|| \leq \varepsilon_1,$$
2. $$|f(x^{n+1}) - f(x^n)| \leq \varepsilon_2,$$
3. $$|f'(x^n)| \leq \varepsilon_3.$$

Методы спуска

Пусть известно направление спуска такое что $f(x + \alpha x) < f(x)$. Пусть заданы x^n, h^n, необходимо выбрать α_n, такое что

$$f(x^n + \alpha_n h^n) = \min_{\alpha \geq 0} f(x^n + \alpha h^n).$$

Данную задачу не сложно решить в явном виде для квадратичного функционала:

$$f(x) = \frac{1}{2}(Ax, x) + (b, x) + (c, c).$$

Метод доверительной области

Рассмотрим приближенную модель с учетом ограниченного шага

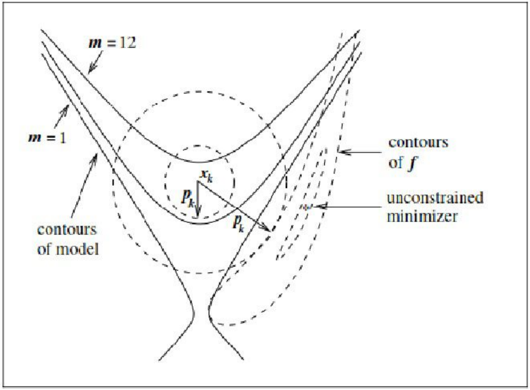
$$\min_{p \in \mathbb{R}} m_k(p) = \frac{1}{2} p^T B_k p + g_k^T p + f_k, \; ||p|| \leq \triangle_k,$$

важна величина близости модели к исходной функции

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)},$$

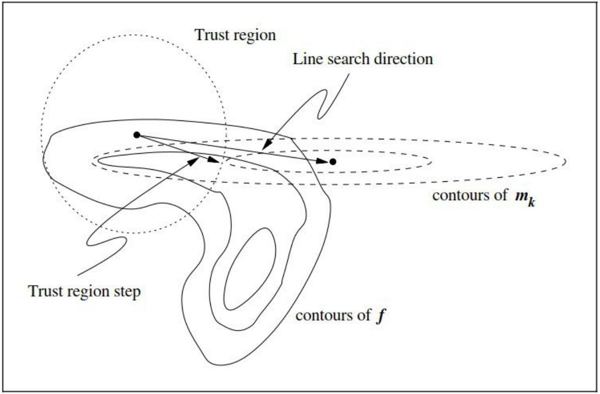
идеальный случай $\rho_k \sim 1$, если ρ_k - маленькое - уменьшаем область, если ρ_k близко к 1 и шаг p_k достигает границы - увеличиваем.

Контуры минимизации



Разница подходов: доверительная область - максимальный радиус, направление
Для методов спуска - направление, длина шага.

Контуры доверительной области



Методы спуска с условиями Вульфа и Голдштайна

Условие Вульфа

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k = l(\alpha),\ c_1 \in (0, 1),$$

в том числе условия кривизны

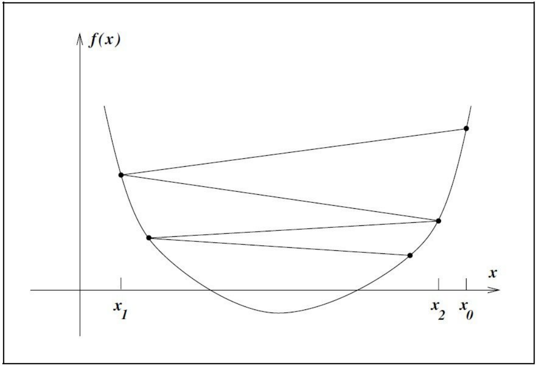
$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k,\ c_2 \in (c_1, 1),$$

Условие Гольдштайна

$$f(x_k) + (1 - c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f_k^T p_k,$$

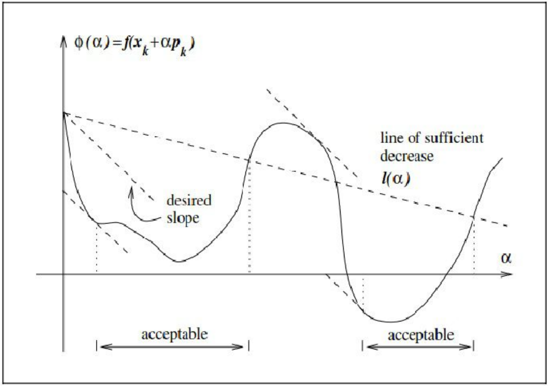
Контрпример

Рассмотрим последовательность $f(x_k) = 1/k$

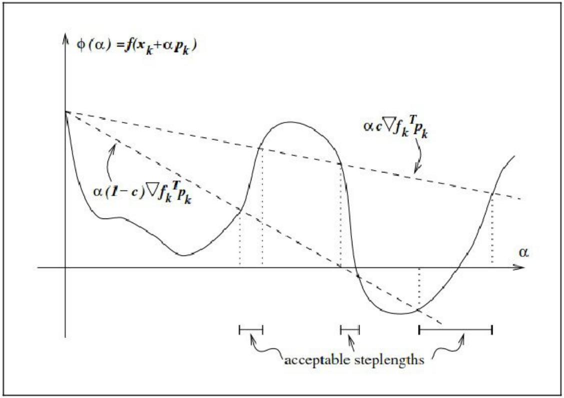


которая не сходится к минимуму $f(x^*) = -1$.

Условия Вульфа иллюстрация



Условия Гольдштайна, иллюстрация



Метод наискорейшего спуска

Пусть функционал имеет квадртоичный вид

$$f(x) = \frac{1}{2}x^T Qx - b^T x,$$

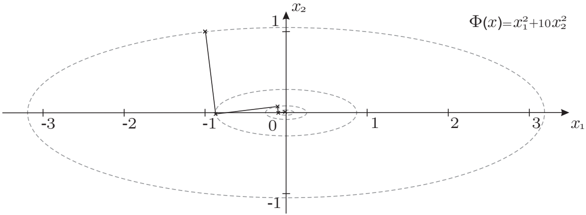
Матрица Q - симметрична и положительно определена, минимум соответствует решению уравнения $Qx = b$. Приравнивая к нулю производные функции $f(x_k - \alpha\nabla f_k)$, находим значение оптимального параметра

$$\alpha_k = \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k},$$

получим итерационный процесс

$$x_{k+1} = x_k - \left(\frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k} \right) \nabla f_k.$$

Метод наискорейшего спуска, иллюстрация



Метод Ньютона

Направление спуска для метода Ньютона

$$p_k^N = -\nabla^2 f_k^{-1} \nabla f_k,$$

1. матрица Гессе $\nabla^2 f_k$ не всегда положительно определена
2. p_k^N может не быть направлением уменьшения;

Возможные решения проблем заключаются в 1. модификации матрицы Гессе для обеспечения ее положительной определенности, 2. подход доверительных областей, в котором матрица $\nabla^2 f_k$ положительно определена. Таким образом, будет обеспечена квадратичная сходимость, при том, что параметр спуска $\alpha_k = 1$.

Квазиньютоновские методы

Рассмотрим направления спуска, достаточно положить $\alpha_k = 1$

$$p_k = -B_k^{-1}\nabla f_k,$$

Теорема

Пусть функционал f два раза непрерывно дифференцируем. Рассматривая итерации $x_{k+1} = x_k + \alpha_k p_k$, где α_k удовлетворяют условиям Вульфа. Если последовательность $\{x_k\}$ сходится к точке x^* для которой $\nabla f(x^*) = 0$ и $\nabla f(x^*)$ положительно определена, а также направления спуска удовлетворяют условиям

$$lim_{k\to\infty}\frac{||\nabla f_k + \nabla^2 f_k p_k||}{||p_k||} = 0,$$

то справедливо длины шага $\alpha_k = 1$ допустимы начаная с некоторого k_0 и если $\alpha_k = 1$ при $k > k_0$, сходимость к x^* сверхлинейная.

Ньютоновские методы с модификацией Гессиана

Один из простых способов скорректировать матрицу с учетом минмального отрицательного собственного значения

1. положим $\beta > 0$
2. $\tau_0 = -\min(a_{ii}) + \beta$;
3. обновляем матрицу $B_k = \nabla^2 f(x_k) + \tau_k I$;
4. проверяем положительную определенность применением разложения Холецкого $L^T L = A + \tau_k I$;
5. увеличиваем $\tau_{k+1} = 2\tau_k$, повторяем предыдушую итерацию;
6. решаем систему $B_k p_k = -\nabla f(x_k)$;
7. проводим шаг минимизации $x_{k+1} = x_k + \alpha_k p_k$;

Алгоритм Бройдена — Флетчера — Гольдфарба — Шанно (BFGS)

Пусть получена дискретизация функционала

$$m_k(x) = \frac{1}{2}p^T B_k p + \nabla f_k^T p + f_k,$$

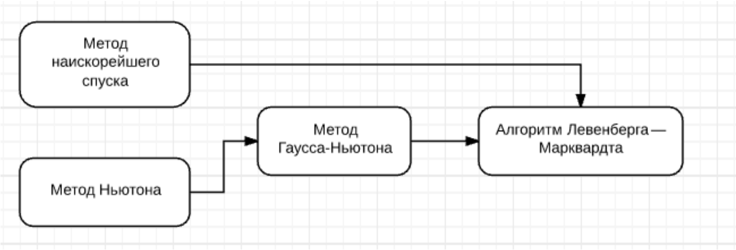
где вектор p используется в качестве направления спуска $x_{k+1} = x_k + \alpha_k p_k$ (α_k- можно найти например по методу Вульфа). Построим итерационный процесс для обновления матрицы Гессе:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}.(*)$$

Схема алгоритма BFGS

1. Задается начальное приближение x_0, точность ϵ;
2. While $||\nabla f_k|| > \epsilon$
   - 2.1 вычисляется направление $p_k = -H_k \nabla f_k$;
   - 2.2 $x_{k+1} = x_k + \alpha_k p_k$, где α_k ищется по направлению наискорейшего спуска;
   - 2.3 вычисляется H_{k+1} по формуле (*);
   - 2.4 следующий шаг $k = k + 1$;
3. end(while)

Идейная связь методов



Левенберга -Марквардта схема метода

Algorithm **Levenberg–Marquardt method**

begin

$k := 0;\quad \nu := 2;\quad \mathbf{x} := \mathbf{x}_0$

$\mathbf{A} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{J}(\mathbf{x});\quad \mathbf{g} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{r}(\mathbf{x})$

$found := (\|\mathbf{g}\|_\infty \le \varepsilon_1);\quad \mu := \tau * \max\{a_{ii}\}$

while (**not** *found*) **and** $(k < k_{\max})$

$\quad k := k+1;\quad$ Solve $(\mathbf{A} + \mu\mathbf{I})\mathbf{h}_{\mathrm{lm}} = -\mathbf{g}$

$\quad$ **if** $\|\mathbf{h}_{\mathrm{lm}}\| \le \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2)$

$\quad\quad found :=$ **true**

$\quad$ **else**

$\quad\quad \mathbf{x}_{\mathrm{new}} := \mathbf{x} + \mathbf{h}_{\mathrm{lm}}$

$\quad\quad \varrho := (F(\mathbf{x}) - F(\mathbf{x}_{\mathrm{new}}))/(L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{lm}}))$

$\quad\quad$ **if** $\varrho > 0$ $\qquad\qquad$ {step acceptable}

$\quad\quad\quad \mathbf{x} := \mathbf{x}_{\mathrm{new}}$

$\quad\quad\quad \mathbf{A} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{J}(\mathbf{x});\quad \mathbf{g} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{r}(\mathbf{x})$

$\quad\quad\quad found := (\|\mathbf{g}\|_\infty \le \varepsilon_1)$

$\quad\quad\quad \mu := \mu * \max\{\frac{1}{3}, 1 - (2\varrho - 1)^3\};\quad \nu := 2$

$\quad\quad$ **else**

$\quad\quad\quad \mu := \mu * \nu;\quad \nu := 2 * \nu$

end

Левенберга -Марквардта (доверительные интервалы)

Связь метода с методом доверительной области

$$\min_p \frac{1}{2}||J_k p + r_k||^2, \ ||p|| \leq \triangle_k,$$

когда достиагется граничное условие, задачу можно свести к

$$(J^T J + \lambda I)p = -J^T r.$$

Метод сопряженных градиентов, квадратичный случай

Рассмотрим квадратичный функционал

$$f(x) = \frac{1}{2}(Ax, x) + (b, x) + c.$$

Definition

Векторы h^1, h^2 называются сопряжнными относительно матрицы A, если $(Ah^1, h^2) = 0$.

Theorem

Если $\{h^k\}$ взаимно сопряженные относительно матрицы A, $k = 0, .., m-1$, α_k находится по наискорейшему спуску, то $f(x^m) = \min_{x \in X_m} f(x)$, где $X_m = \{x : x = x^0 + \sum \lambda_k h^k, \}$ - афинное многообразие.

Метод сопряженных градиентов, нелинейный случай

Рассмотрим алгоритм в форме Флетчера-Ривса:

1. Вычисляем антиградиент в начальной точке $d_0 = -\nabla f(x_0)$;
2. осуществляем наискорейший спуск $f(x_i + \alpha_i d_i)$;
3. переход в новую точку $x_{i+1} = x_i + \alpha_i d_i$;
4. вычисляется антиградиент $r_{i+1} = -\nabla f(x_{i+1})$;
5. коэффициент $\beta_k = \frac{(r_{k+1}, r_{k+1})}{(r_k, r_k)}$
6. вычисление нового направления $d_{i+1} = r_{i+1} + \beta_{i+1} d_i$.

Условная оптимизация

Решается задача

$$\min_{x \in \mathbb{R}^n} f(x)$$

с ораничениями

$$\begin{cases} c_i(x) = 0,\ i \in \mathcal{E}, \\ c_i(x) \geq 0,\ i \in \mathcal{I}, \end{cases}$$

Вводим функцию Лагранжа

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x),$$

Условия Каруша — Куна — Таккера

Теорема

Пусть есть локальное решение и функции непрерывно дифференцируемы, выполенно (LICQ) (градиенты активных ограничений и градиенты ограничений неравенств линейно независимы) в точке $x*$. Тогда существуют значения множителей Лагранжа, так что выполнены следующие условия:

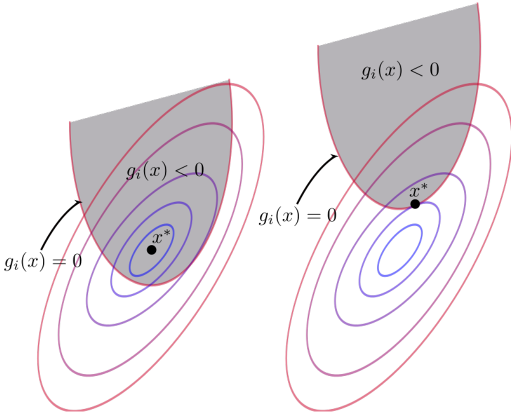
$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

$$c_i(x^*) = 0, \; i \in \mathbb{E},$$

$$c_i(x^*) \geq 0, \; i \in \mathbb{I},$$

$$\lambda_i^* \geq 0, \; i \in \mathbb{I},$$

$$\lambda_i^* c_i(x^*) = 0, \; i \in \mathbb{E} \cup \mathbb{I}.$$



$g_i(x) < 0$
x^*
$g_i(x) = 0$
$g_i(x) < 0$
x^*
$g_i(x) = 0$

Последовательное квадратичное программирование (SQP)

Минимизируем $\min f(x)$ при условии $c(x) = 0$.
Пусть $A(x) = [\nabla c_1(x), .., \nabla c_m(x)]$, Система ККТ первого рода:

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ c(x) \end{pmatrix} = \theta.$$

(Необходимые условия локального минимума для задач с ограничениями исследуются давно. Одним из основных остаётся предложенный Лагранжем перенос ограничений в целевую функцию. Условия Куна-Таккера тоже выведены из этого принципа.)

Последовательное квадратичное программирование (SQP)

якобиан задается формулой

$$F'(x,\lambda) = \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix},$$

для приращений $x_{k+1} = x_k + p_k$, $\lambda_{k+1} = \lambda_k + p_\lambda$ можно получить систему Ньютона-ККТ

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}_k & -A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} p_k \\ p_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f_k + A_k^T \lambda_k \\ -c_k \end{pmatrix}$$

Метод сопряженного градиента для решения задач на специальных множествах

Минимизируем функционал

$$\Phi(z) = ||Az - u_\delta||^2,$$

на множестве ограниченны монотонных $Z\downarrow$, выпуклых $\check{Z}$, выпуклых монотонных $\check{Z}\downarrow$

$$\Phi'(z) = 2(A^*Az - A^*u_\delta),$$

условие Липшица

$$||\Phi'(z_1) - \Phi'(z_2)|| \leq 2||A||^2||z_1 - z_2||,$$

позволяет применить метод условного градиента.

Метод условного градиента

1

$$(\Phi'(z^k), \bar{z}^k) = \min_{z \in M} (\Phi'(z^k), z),$$

точка $\bar{z}^k$ принадлежит границе множества M, в случае когда M - ограниченнй замкнутый выпуклый многогранник (возможно перебрать вершины, если они известны).

2

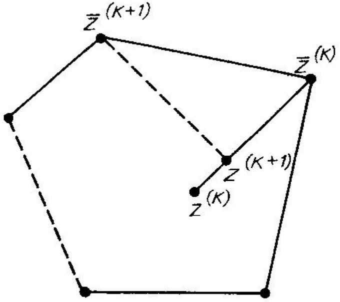
построим

$$z^{k+1} = z^k + \lambda_k(\bar{z}^k - z^k),$$

где $\lambda_k \in [0, 1]$ является решением одномерной задачи

$$\Phi(z^{k+1}) = \Phi(z^k + \lambda_k(\bar{z}^k - z^k)) = \min_{\lambda \in [0,1]} \Phi(z^k + \lambda(\bar{z}^k - z^k)).$$

Иллюстрация метода условного градиента



Метод сопряженных градиентов для решения оз 1

Будем рассматривать задачи на множествах простой структуры

$$\breve{M} \downarrow \begin{cases} z_{i-1} - 2z_i + z_{i+1} \leq 0, \\ z_{i+1} - z_i \leq 0, \\ z_i \geq 0, \end{cases}$$

Данное ограничение легко представимо в виде

$$Fz \leq g,$$

выделим множество активных ограничений $I(z)$ для которых выполнено

$$\sum_{j=1}^{n} F_{ij} = g_i.$$

Алгоритм метода СГ с ограничениями 1

Исходный квадратичный функционал представим в виде

$$\phi(z) = (z, Qz) + (d, z) + e,$$

1. минимизацию начинаем с произвольной точки z^0, число активных ограничений $m = 0$;
2. если $k = n$, то решение найдено, останов;
3. направление спуска для $k = 0$, $p^k = grad\phi(z^k)$, $p^k = -grad\phi(z^k) + \frac{||grad\phi(z^k)||}{||grad\phi(z^{k-1})||} p^{k-1}$,
4. вычисляется величина оптимального шага вдоль заданного направления p^k:

$$a_k = \frac{1}{2} \frac{(grad\phi(z^k), p^k)}{(Qp^k, p^k)},$$

Алгоритм метода СГ с ограничениями 2

5. вычисляется величина a_{max} максимально возможного шага вдоль направления p^k, не выводящего за пределы множества Y;
6. если $a_k \leq a_{max}$, $z^{k+1} = z^k + a_k p^k$, $k = k + 1$, переход на пункт, иначе $z^{k+1} = z^k + a_{max} p^k$, переходим на следующий шаг;
7. появилось новое активное ограничение $I(z)$, $m = m + 1$;
8. вычисляем проектор P_I на подпространство R_{n-m} $(F_I z = 0)$, по формуле $P_I = E - F_I^*(F_I F_I^*)^{-1} F_I$;
9. повторяем шаг , только в качестве начальной точки берется и всюду вместе берется их проекция, минимум по методу сг находится за $n - m$ шагов
10. если минимум на многобразии найден и $m = 0$,если минимум найден $m > 0$ то переходим к слудющему шагу, если минимум не найден $a_k > 0$, то переходим к шагу..

Алгоритм метода СГ с ограничениями 3

11. вычислим набор из m параметров по формуле

$$u^0 = (F_I F_I^*)^{-1} F_I grad \phi$$

12. если $u_i^0 \geq 0$, то найдено решение задачи;
13. если $u_i^0 < 0$ для каждого i исключаем из $I(z)$, переходим на шаг..., положив $m = m - 1$.