1. О градиенте

1.1. Вычисление

Запишем задачу в стиле безусловной минимизации:

$$\underset{x\in\mathbb{R}^n}{min}\; f(x) = \frac{1}{2}\frac{(Ax,x)}{(x,x)} + \frac{(b,x)}{\sqrt{(x,x)}}. \tag{1.1}$$

Естественно задача имеет смысл при ненулевых x: $x \neq 0$.

Вычислим градиент функции:

$$\begin{aligned}
\nabla f(x) &= \nabla\left(\frac{1}{2}\frac{(Ax,x)}{(x,x)} + \frac{(b,x)}{\sqrt{(x,x)}}\right) = \\
&= \frac{Ax}{(x,x)} - \frac{x}{(x,x)}\left(\frac{(Ax,x)}{(x,x)} + \frac{(b,x)}{\sqrt{(x,x)}}\right) + \frac{b}{\sqrt{(x,x)}} = \\
&= \frac{\xi(x)}{\sqrt{(x,x)}}, \\
\xi(x) &= \frac{Ax}{\|x\|} - \frac{x}{\|x\|}\widehat{f}(x) + b, \\
\widehat{f}(x) &= \left(\frac{(Ax,x)}{(x,x)} + \frac{(b,x)}{\|x\|}\right).
\end{aligned} \tag{1.2}$$

Заметим, что $\xi(x)$ совпадает с направлением градиента, поскольку $\sqrt{(x,x)} > 0$. $\xi(x)$ опять же функция от единичного вектора.

Зададим $x_0 \neq \vec{0}$, обозначим $\xi(x_0) = \xi_0$ и попробуем двигаться в сторону антиградиента:

$$x_1 = x_0 - \gamma\xi_0 \tag{1.3}$$

Рассмотрим (x_1, x_1) :

$$\begin{gathered}
(x_1,x_1) = (x_0,x_0) - 2\gamma(x_0,\xi_0) + \gamma^2(\xi_0,\xi_0) = (x_0,x_0) + \gamma^2(\xi_0,\xi_0), \\
(x_0,\xi_0) = \frac{(Ax_0,x_0)}{\|x_0\|} - \frac{(x_0,x_0)}{\|x_0\|}\frac{(Ax_0,x_0)}{(x_0,x_0)} - \frac{(x_0,x_0)}{\|x_0\|}\frac{(b,x_0)}{\|x_0\|} + (b,x_0) = 0,
\end{gathered} \tag{1.4}$$

что означает, что градиент всегда ортогонален точке, в которой он взят. Теперь рассмотрим следующие величины, которые входят в выражение $f(x_1)$:

$$(Ax_1,x_1) = (Ax_0,x_0) - 2\gamma(Ax_0,\xi_0) + \gamma^2(A\xi_0,\xi_0). \tag{1.5}$$

Рассмотрим (Ax_0,ξ_0). Для начала выразим Ax_0 из выражения для градиента:

$$\begin{gathered}
Ax_0 = \|x_0\|\xi_0 + x_0\widehat{f}(x_0) - b\|x\| \\
(Ax_0,\xi_0) = \|x_0\|\|\xi_0\|^2 + \underbrace{(\xi_0,x_0)}_{=0,\ (1.4)}\widehat{f}(x_0) - (b,\xi_0)\|x_0\| = \|x_0\|\big(\|\xi_0\|^2 - (b,\xi_0)\big).
\end{gathered} \tag{1.6}$$

Отсюда легко заметить, что при $\xi_0 = b$ выполняется $(Ax_0, b) = 0$.

Не забываем про "линейную часть". Учитывая выражение для $\widehat{f}(x)$,

$$(b,x_1) = (b,x_0-\gamma\xi_0) = \|x_0\|\widehat{f}(x_0) - \frac{(Ax_0,x_0)}{\|x_0\|} - \gamma\Big(\|\xi_0\|\widehat{f}(\xi_0) - \frac{(A\xi_0,\xi_0)}{\|\xi_0\|}\Big) \tag{1.7}$$

Теперь соберем все в выражение для $f(x_1)$:

$$\begin{aligned}
f(x_1) = {} & \frac{1}{2}\frac{(Ax_0,x_0) - 2\gamma\|x_0\|\big[(\xi_0,\xi_0)-(b,\xi_0)\big] + \gamma^2\big[(A\xi_0,\xi_0)\big]}{(x_0,x_0)+\gamma^2(\xi_0,\xi_0)} + \\
& + \frac{\|x_0\|\widehat{f}(x_0) - \frac{(Ax_0,x_0)}{\|x_0\|} - \gamma\big[\|\xi_0\|\widehat{f}(\xi_0) - \frac{(A\xi_0,\xi_0)}{\|\xi_0\|}\big]}{\sqrt{(x_0,x_0)+\gamma^2(\xi_0,\xi_0)}}
\end{aligned} \tag{1.8}$$

Поделим числитель из знаменатель каждого слагаемого на (x_0,x_0) и $\sqrt{(\xi_0,\xi_0)}$:

$$
\begin{aligned}
f(x_1) = &\frac{1}{2}\frac{\frac{(Ax_0,x_0)}{(x_0,x_0)} - 2\gamma\frac{1}{\|x_0\|}\big[(\xi_0,\xi_0)-(b,\xi_0)\big] + \gamma^2\big[\frac{(A\xi_0,\xi_0)}{(x_0,x_0)}\big]}{1+\gamma^2 t_0^2} + \\
&+ \frac{\widehat{f}(x_0) - \frac{(Ax_0,x_0)}{(x_0,x_0)} - \gamma\big[t_0\widehat{f}(\xi_0) - \frac{(A\xi_0,\xi_0)}{\|\xi_0\|\cdot\|x_0\|}\big]}{\sqrt{1+\gamma^2 t_0^2}}, \\
t_0 = &\frac{\|\xi_0\|}{\|x_0\|}.
\end{aligned} \tag{1.9}
$$

Причесываем дальше:

$$
\begin{aligned}
f(x_1) &= \frac{\frac{1}{2}\frac{(Ax_0,x_0)}{(x_0,x_0)} - \gamma\frac{t_0}{\|\xi_0\|}\big[(\xi_0,\xi_0)-(b,\xi_0)\big] + \frac{1}{2}\gamma^2 t_0^2\big[\frac{(A\xi_0,\xi_0)}{(\xi_0,\xi_0)}\big]}{1+\gamma^2 t_0^2} + \\
&+ \frac{\widehat{f}(x_0) - \frac{(Ax_0,x_0)}{(x_0,x_0)} - \gamma\big[t_0\widehat{f}(\xi_0) - t_0\frac{(A\xi_0,\xi_0)}{\|\xi_0\|\cdot\|\xi_0\|}\big]}{\sqrt{1+\gamma^2 t_0^2}} \\
f(x_1) &= \frac{\frac{1}{2}\frac{(Ax_0,x_0)}{(x_0,x_0)} - \gamma\frac{t_0}{\|\xi_0\|}\big[(\xi_0,\xi_0)-(b,\xi_0)\big] + \frac{1}{2}\gamma^2 t_0^2\big[\frac{(A\xi_0,\xi_0)}{(\xi_0,\xi_0)}\big]}{1+\gamma^2 t_0^2} + \\
&+ \frac{\frac{(b,x_0)}{\|x_0\|} - \gamma t_0\big[\frac{(b,\xi_0)}{\|\xi_0\|}\big]}{\sqrt{1+\gamma^2 t_0^2}}.
\end{aligned} \tag{1.10}
$$

Теперь рассмотрим разность $f(x_1)-f(x_0)$:

$$
\begin{aligned}
&f(x_1)-f(x_0) = f(x_1) - \frac{1}{2}\frac{(Ax_0,x_0)}{(x_0,x_0)} - \frac{(b,x_0)}{\|x_0\|} = \\
&= \frac{\frac{1}{2}\gamma^2 t_0^2\Big(\frac{(A\xi_0,\xi_0)}{(\xi_0,\xi_0)} - \frac{(Ax_0,x_0)}{(x_0,x_0)}\Big) - \gamma t_0\Big(\|\xi_0\| - \frac{(b,\xi_0)}{\|\xi_0\|}\Big)}{1+\gamma^2 t_0^2} + \frac{\frac{(b,x_0)}{\|x_0\|} - \gamma t_0\frac{(b,\xi_0)}{\|\xi_0\|}}{\sqrt{1+\gamma^2 t_0^2}} - \frac{(b,x_0)}{\|x_0\|}
\end{aligned} \tag{1.11}
$$

Стоит исследовать это выражение как функцию от γ, чтобы понять как выбирать шаг.

1.2. Необходимое условие экстремума первого порядка

Необходимое условие первого порядка говорит нам о том, что в оптимальной точке градиент необходимо равен нулю. Попробуем исследовать это условие. Будем исследовать равенство $\xi(x)=\vec{0}$, поскольку направления этого вектора с градиентом совпадают.