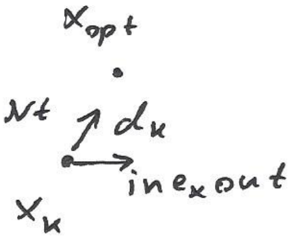
07. 10. 17 момо V

# Hessian Free Newton (HFN)

$f(x) \to \min\limits_{x \in \mathbb{R}^n}$, Newton: $x_{k+1} = x_k + \alpha_k d_k$

$d_k : H_k d_k = -g_k$

Идея HFN ① Решение СЛАУ через CG
② Решение СЛАУ неточно




$r_k = H_k d_k + g_k$

Критерий останова в CG:

$$\|r_k\| \le \eta_k \|g_k\|$$

не уменьшается монотонно ← форсирующая посл-ть

$H_k d = -g_k \iff \varphi(d) = \frac{1}{2} d^T H_k d + g_k^T d \to \min\limits_d$

Хотим $\langle d_k, g_k \rangle < 0$:

] $d_{start} = 0 \Rightarrow \varphi(d_k) \le \varphi(d_{start}) = 0 \Rightarrow$

$\varphi(d_k) = \frac{1}{2} d_k^T H_k d_k + g_k^T d_k \le 0 \Rightarrow$

$g_k^T d_k \le -\frac{1}{2} d_k^T H_k d_k < 0$

Глобальная сходимость линейная.

Локальная сходимость HFN

$f \in C_M^{2,2}$, $\nabla^2 f(x_{opt}) \succeq \mu I \succ 0$, $\alpha_k = 1$

$x_{k+1} = x_k + d_k$, $d_k = H_k^{-1}(r_k - g_k)$, $\|r_k\| \le \eta_k \|g_k\|$

$\|x_{k+1} - x_{opt}\| = \|x_k - x_{opt} + H_k^{-1}(r_k - g_k)\| \le$

$\le \|H_k^{-1}\| \|H_k(x_k - x_{opt}) - g_k + r_k\| \le \underbrace{\|H_k^{-1}\|}_{const} (\underbrace{\|H_k(x_k - x_{opt}) - g_k\|}_{const \|x_k - x_{opt}\|^2} + \|r_k\|) \le const\,(O(\|x_k - x_{opt}\|^2) + \eta_k \|g_k\|) \le$

$\le const\,(O(\|x_k - x_{opt}\|^2) + \eta_k O(\|x_k - x_{opt}\|))$

$\{\|g_k\| = \|\nabla f(x_k) - \overset{=0}{\nabla f(x_{opt})}\| \le \{f \in C_L^{1,1}\} \le L\|x_k - x_{opt}\|\}$ 1

① $\eta_k \le \eta < 1$ → линейная скорость

② $\eta_k \xrightarrow[k\to\infty]{} 0$ → сверхлинейная

типичный выбор $\eta_k = \min(\frac{1}{2}, \sqrt{\|g_k\|})$

③ $\eta_k = O(\|g_k\|)$ → квадратичная

типичный выбор $\eta_k = \min(\frac{1}{2}, \|g_k\|)$

$Ax = b \leftarrow x = CG(A[\cdot], b, x_0, \varepsilon)$

Схема HFN

$x_0, \varepsilon$

$g_0 = \nabla f(x_0)$

для $k = 0, 1, 2, \dots$

$\varepsilon_k = \min(\frac{1}{2}, \sqrt{\|g_k\|})\|g_k\|$

$d_k = CG(H_k[\cdot], -g_k, 0, \varepsilon_k)$

$\alpha_k = \arg\min\limits_{\alpha > 0} f(x_k + \alpha d_k)$ // Armijo / Wolf, $\alpha_{start} = 1$

$x_{k+1} = x_k + \alpha_k d_k$

$g_{k+1} = \nabla f(x_{k+1})$

если $\|g_{k+1}\|^2 < \varepsilon$, то стоп

Пример: логистическая регрессия

$\{x_i, y_i\}_{i=1}^N$, $x_i \in \mathbb{R}^D$, $y_i \in \{-1, +1\}$

$\hat{y}(x) = \mathrm{sgn}\, w^T x$, $f(w) = \frac{1}{N}\sum_{i=1}^N \ln(1 + \exp(-y_i w^T x_i)) + \lambda \|w\|_2^2 \to \min\limits_w$

$\nabla^2 f(w) = \frac{1}{N} X^T B X + \lambda I$, $B = \mathrm{diag}(b_1 \dots b_N)$

$b_i = \sigma(y_i w^T x_i)(1 - \sigma(y_i w^T x_i)) > 0$

$$\nabla^2 f(w) \cdot d = \frac{1}{N} X^T B X d + \lambda d =: p \iff \left[\begin{array}{l} p_1 = Xd \\ p_2 = B p_1 \\ p_3 = X^T p_2 \\ p = \frac{1}{N} p_3 + \lambda d \end{array}\right.$$

| | Вычисление $\nabla^2 f(w)$ | Последнее умножение |
|---|---|---|
| Сложность $\nabla^2 f(w) d$ | $O(ND^2) + O(D^2)$ | $O(ND)$ |

# Разностное дифференцирование

$$\nabla f(x+\varepsilon d) = \nabla f(x) + \nabla^2 f(x)\,\varepsilon d + O(\varepsilon^2)$$

① $$\underbrace{\nabla^2 f(x)\cdot d}_{\Theta_{true}} = \underbrace{\frac{\nabla f(x+\varepsilon d) - \nabla f(x)}{\varepsilon}}_{\Theta_1(\varepsilon)} + O(\varepsilon)$$

② $$\nabla^2 f(x)\, d = \frac{\nabla f(x+\varepsilon d) - \nabla f(x-\varepsilon d)}{2\varepsilon} + O(\varepsilon^2)$$

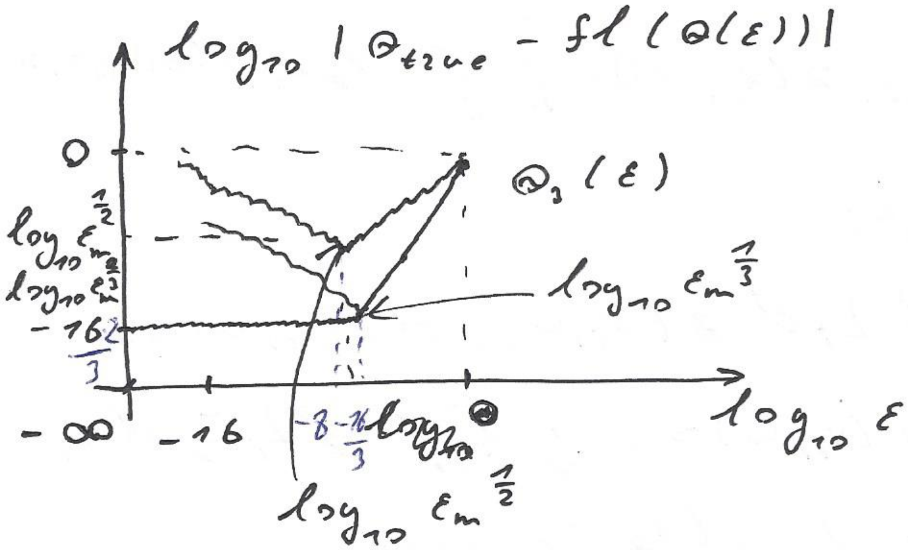
$fl(x) \approx x(1+\varepsilon_m)$, $\quad |fl(x) - x| \le |x|\,\varepsilon_m \le L_x\cdot\varepsilon_m$

$$|\Theta_{true} - fl(\Theta_2(\varepsilon))| \le \underset{\underset{L\varepsilon^2}{\wedge}}{|\Theta_{true} - \Theta_2(\varepsilon)|} + \underset{\underset{\frac{L_1\varepsilon_m + L_1\varepsilon_m}{2\varepsilon}}{\wedge}}{|\Theta_2(\varepsilon) - fl(\Theta_2(\varepsilon))|} \le$$

$$\le L\varepsilon^2 + \frac{L_1\varepsilon_m}{\varepsilon} \to \min_{\varepsilon}$$

$$const\cdot\varepsilon^2 + const\cdot\frac{\varepsilon_m}{\varepsilon} = const\,\varepsilon_m^{2/3}$$

$$2L\varepsilon - \frac{L_1\varepsilon_m}{\varepsilon^2} = 0 \Rightarrow \varepsilon = \left(\frac{L_1}{2L}\varepsilon_m\right)^{\frac{1}{3}}$$




③ $$\nabla f(x+i\varepsilon d) = \nabla f(x) + \nabla^2 f(x)\, i\varepsilon d + O(\varepsilon^2) + iO(\varepsilon^3)$$

$$\mathrm{Im}[\nabla f(x+i\varepsilon d)] = \nabla^2 f(x)\,\varepsilon d + O(\varepsilon^3)$$

$$\nabla^2 f(x)\, d = \underbrace{\frac{\mathrm{Im}[\nabla f(x+i\varepsilon d)]}{\varepsilon}}_{\Theta_3(\varepsilon)} + O(\varepsilon^2)$$

? $f(x)$ бесконечная диф-ть

Гаусс - Ньютоновская аппроксимация гессиана

$\{x_i, y_i\}_{i=1}^N, \; x_i \in \mathbb{R}^D, \; f(w) = \frac{1}{N}\sum_{i=1}^N L(y_i, z(x_i, w)) +$

$+ \lambda R(w) \to \min_w$

Примеры:

| Задача | Ф-ия потерь | Прогноз |
| --- | --- | --- |
| Многомерная регрессия | $L(y,z) = \frac{1}{2}\|y - z\|_2^2$ | $z = Wx$ или $z = NN(x, w)$ |
| Классификация на 2 класса $y_i \in \{-1, +1\}$ | $L(y,z) = \ln(1 + e^{-yz})$ | $z = \operatorname{sgn}(w^T x)$ или $z = \operatorname{sgn}(NN(x,w))$ |
| Классификация на K классов | $L(y,z) = -\sum_{j=1}^K [y_j = j] z_j + \ln\left(\sum_{j=1}^K e^{z_j}\right)$ | — // — argmax |

$d f(w) = \frac{1}{N}\sum_i dL(y_i, z(x_i, w)) = \frac{1}{N}\sum_i \langle \nabla_z L_i, dz(x_i, w)\rangle$

$\nabla f(w) = \frac{1}{N}\sum_{i=1}^N J_{z_i}^T \nabla_z L_i$

$d^2 f(w) = \frac{1}{N}\sum_{i=1}^N \langle \nabla_z^2 L_i \, dz(x_i, w), dz(x_i, w)\rangle +$

$+ \frac{1}{N}\sum_{i=1}^N \langle \nabla_z L_i, d^2 z(x_i, w)\rangle = \frac{1}{N}\sum_{i=1}^N \langle \nabla_z^2 L_i J_{z_i} dw,$

$J_{z_i} dw\rangle + \frac{1}{N}\sum_i \langle \nabla_z L_i, d^2 z(x_i, w)\rangle$

$\nabla_{GN}^2 f(w) = \frac{1}{N}\sum_{i=1}^N J_{z_i}^T \nabla_z^2 L_i J_{z_i}$

Если $\nabla_z^2 L_i \succ 0$, то $\nabla_{GN}^2 f(w) \succeq 0$ ?

$\nabla_{LM}^2 f(w) = \nabla_{GN}^2 f(w) + \tau I \succ 0 \quad \forall \tau > 0$

Схема HFN + LM + trust region

$w_0$, $\varepsilon$

$g_0 = \nabla f(w_0)$, $\tau_0 = 1$

для $k = 0, 1, 2, \ldots$

$\varepsilon_k = \min\left(\frac{1}{2}, \sqrt{\|g_k\|}\right)\|g_k\|$

$d_k = CG\left(\{\nabla^2_{GN} f(w_k) + \tau_k I\}, -g_k, 0, \varepsilon_k\right)$

если $f(x_k + d_k) > f(x_k)$, тогда $\tau_k \leftarrow 4\tau_k$

$$p = \frac{f(x_k + d_k) - f(x_k)}{m_k(d_k) - m_k(0)}$$

$$m_k(d) = \frac{1}{2} d^T \left(\nabla^2_{GN} f(w_k) + \tau_k I\right) d + g_k^T d$$

если $p > \frac{3}{4}$, тогда $\tau_k \leftarrow \frac{\tau_k}{2}$

если $p < \frac{1}{4}$, тогда $\tau_k \leftarrow 2\tau_k$

$x_{k+1} = x_k + d_k$

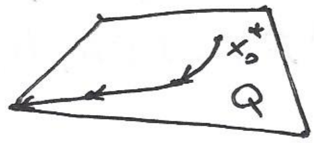
$g_{k+1} = \nabla f(w_{k+1})$

если $\|g_{k+1}\|^2 \le \varepsilon$, то выход

|| семинар ||

Самосогласованные функции

$\min\limits_{x \in Q} f(x)$




Барьерная ф-ия $F(x) \to \infty$ при $x \to x_0$

$\varphi_t(x) = t f(x) + F(x)$ 60-е

$\Downarrow x_t^* \to x^*$

$Q = \{x \in \mathbb{R}^n : \langle a_i, x\rangle \le b_i\}$ $\forall\, 1 \le i \le m$

① $F(x) = -\sum\limits_{i=1}^{m} \ln(b_i - \langle a_i, x\rangle)$

② $F(x) = \sum\limits_{i=1}^{m} \frac{1}{(b_i - \langle a_i, x\rangle)^2}$

# Метод Ньютона

$$x^+ = x - (\nabla^2 f(x))^{-1} \nabla f(x)$$

$\underline{\text{Т}}$ $F \in C^2$

① $\|\nabla^2 F(y) - \nabla^2 F(x)\|_2 \le M \|y - x\|_2 \iff$

$|D^3 F(x)[h,h,h]| \le M \|h\|_2^3 \quad \forall x, \forall h$

② $\nabla^2 F(x^*) \succcurlyeq \mu I_n, \ \mu > 0 \quad \{\mu = \min \lambda \nabla F(x^*)\} \iff$

$D^2 F(x)[h,h] \ge \mu \|h\|_2^2 \quad \forall h$

Если $\|\nabla F(x_0)\|_2 \le \dfrac{2\mu^2}{M} \Rightarrow \|\nabla F(x^+)\| \le \dfrac{M}{2\mu^2} \|\nabla F(x)\|^2$

$V$ – вещ. лин. пр-во, $Q$ – открытое выпуклое мн-во в $V$, $F: Q \to \mathbb{R}$ $\smile$ $\in C^3$

$$\langle h, u \rangle_{F,x} = D^2 F(x)[h,u] = \langle \nabla^2 F(x) h, u \rangle$$

$x \in Q$

$\|h\|_{F,x} = \langle h, h \rangle_{F,x}^{1/2}$ – полунорма

$\underline{\text{О}}$ Самосогласованные ф-ии

$F: Q \to \mathbb{R}$ наз-ся самосогласованной с $M > 0$, если:

① (барьерное св-во) $\forall x_0 \in \partial Q$ $F(x) \to \infty$ при $x \to x_0$

② (согласованность производной)

$$|D^3 F(x)[h,h,h]| \le M \|h\|_{F,x}^3 \quad \forall x \in Q, \forall h \in V$$

Примеры:

I $F: (0, \infty) \to \mathbb{R}$, $F(t) = -\ln t$

① барьерное св-во $\partial(0,\infty) = \{0\}$

② $DF(t)[h] = -\dfrac{h}{t}$, $D^2 F(t)[h,h] = \dfrac{h^2}{t^2}$,

$D^3 F(t)[h,h,h] = -\dfrac{2h^3}{t^3}$, $\left|\dfrac{2h}{t^3}\right|^3 \le 2\left(\dfrac{h^2}{t^2}\right)^{\frac{3}{2}} = 2\dfrac{|h|^3}{t^3}$

$F$ яв-ся самосогласованной с параметром 2

II $F: (0,\infty) \to \mathbb{R}$, $F(t) = \frac{1}{t^p}$, $p > 0$ не с/с

III $F: V \to \mathbb{R}$, $A: V \to V$, $A \in S^n_+$, $b \in V$

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

$D^3 F(x) = 0$, самосогласована с $M = 0$

IV $Q = \{ x \in V : \frac{1}{2} \langle Ax, x \rangle + \langle b, x \rangle < c \}$

$F(x) = -\ln ( c - \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle )$

V $f: S^n_{++} \to \mathbb{R}$, $F(x) = -\ln \det (x)$

$DF(x)[H] = -\langle X^{-1}, H \rangle$, $D^2 F(x)[H, H] = \langle X^{-1} H X^{-1}, H \rangle =$

$= \langle I_n, Z^2 \rangle$, $Z = X^{-\frac{1}{2}} H X^{-\frac{1}{2}}$

$D^3 F(x)[H, H, H] = -2 \langle X^{-1} H X^{-1} H X^{-1}, H \rangle = -2 \langle I_n, Z^3 \rangle$

$$\underbrace{\left( \sum_{i=1}^n \lambda_i^3 \right)^{\frac{1}{3}}}_{\|\lambda\|_3} \leq \underbrace{\left( \sum_{i=1}^n \lambda_i^2 \right)^{\frac{1}{2}}}_{\|\lambda\|_2}, \quad \|\lambda\|_p \leq \|\lambda\|_2, \; p > 2$$

<u>Утв.</u> Сохранение самосогласованности

$F: Q \to \mathbb{R}$ – самосогласована с пар-м $M$

$G: S \to \mathbb{R}$ – с/с с пар. $N$

1. $\alpha > 0$ $\alpha F: Q \to \mathbb{R}$ с/с с пар. $\frac{M}{\sqrt{\alpha}}$
2. $F + G: Q \cap S \to \mathbb{R}$ с/с с пар. $\max \{ N, M \}$
3. $A: V \to W$ – аффинное преобразование

   $G \circ A : A^{-1}(S) \to \mathbb{R}$ с/с с пар. $N$
4. $\Phi: Q \times S \to \mathbb{R}$ $\Phi(x, y) = F(x) + G(y)$ с/с $\max\{M, N\}$

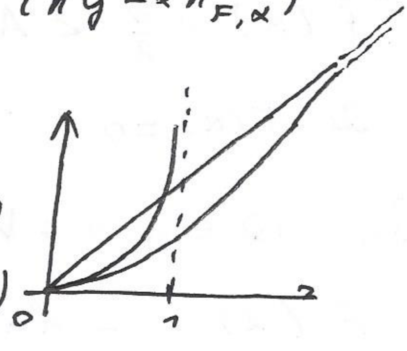
$F(x) = -\ln \det ( B - x_1 A_1 - \ldots - x_n A_n )$

$\{ x \in \mathbb{R}^n : x_1 A_1 + \ldots + x_n A_n < B \}$ стандартная самосогласованность

Утв $F: Q \to \mathbb{R}$ самосогл. с/к, $x, y \in Q$

$$f(x) + \langle \nabla F(x), y-x \rangle + w(\|y-x\|_{F,x}) \leq$$
$$\leq F(y) \leq F(x) + \langle \nabla F(x), y-x \rangle + w_*(\|y-x\|_{F,x})$$

$\|y-x\|_{F,x} < 1$



$w: [0,\infty) \to \mathbb{R}$, $w(t) = t - \ln(1+t)$

$w_*: [0,1) \to \mathbb{R}$, $w_*(s) = -s - \ln(1-s)$

$\| \quad F(y) \leq F(x) + \langle \nabla F(x), y-x \rangle + \frac{1}{2}\|y-x\|^2$

$\| \quad F(y) \geq F(x) + \langle \nabla F(x), y-x \rangle + \frac{\mu}{2}\|y-x\|^2$

Утв Эллипсоид Дикина, $x \in Q$

$\{y \in V : \|y-x\|_{F,x} < 1\} \subseteq Q$

(1) Демпфированный метод Ньютона

$$x_{k+1} = x_k - \frac{1}{1+\lambda_F(x_k)} (\nabla^2 F(x_k))^{-1} \nabla F(x_k)$$

$$\lambda_F(x) = \langle (\nabla^2 F(x_k))^{-1} \nabla F(x_k), \nabla F(x_k) \rangle^{\frac{1}{2}}$$

$$\|x_{k+1} - x_k\|_{F,x_k} = \frac{\lambda_F(x_k)}{1+\lambda_F(x_k)} < 1$$

скорость сх-ти:

$$F(x_{k+1}) \leq F(x_k) - w(\lambda_F(x_k)), \quad \forall k \geq 0$$

(2) $x_{k+1} = x_k - (\nabla^2 F(x_k))^{-1} \nabla F(x_k)$

Утв Если $\lambda_F(x_k) < 1 \Rightarrow \lambda_F(x_{k+1}) \leq \left(\frac{\lambda_F(x_k)}{1-\lambda_F(x_k)}\right)^2$

Как только $\dfrac{\lambda_F(x_k)}{(1-\lambda_F(x_k))^2} < 1 \Leftrightarrow \lambda_F(x_k) \leq \frac{3-\sqrt{5}}{2} \approx 0,38$

$\lambda_F(x_{k+1}) \leq C \lambda_F^2(x_k)$

квадратичная сх-ть