$\min_{x \in \mathbb{R}^n} f(x)$, $f: \mathbb{R}^n \to \mathbb{R}$ convex, differentiable, quasi-Self-Concordant:

$$D^3 f(x)[h,h,u] \leq M \|h\|_x^2 \|u\| \quad \forall x \in \mathbb{R}^n;\ h \in \mathbb{R}^n, u \in \mathbb{R}^n$$

- $\|h\|_x = \langle f''(x) h, h\rangle^{1/2}$ – local norm
- $\|u\| = \langle B h, h\rangle^{1/2}$ – global norm, $B = B^T \succ 0$.

Example logistic regression, $B = \sum_i a_i a_i^T$, $M = 1$.

Newton's Method with Gradient Regularization:

$$x_{k+1} = x_k - \left(f''(x_k) + M\|f'(x_k)\|_* B\right)^{-1} f'(x_k), \quad k \geq 0.$$

if f is non-convex: $f''(x_k) \not\succeq 0$

last time:

Theorem $k \geq 0$:

$$f(x_k) - f(x_{k+1}) \geq \langle f'(x_{k+1}), x_k - x_{k+1}\rangle \geq \frac{1}{2M}\left[\frac{\|f'(x_{k+1})\|_*}{\|f'(x_k)\|_*}\right]^2 \|f'(x_k)\|_*.$$

convexity

Global linear rate

Notation $F_k = f(x_k) - f^*$. $g_k = \|f'(x_k)\|_*$

Th. $\Rightarrow$

$$F_k - F_{k+1} \geq \frac{1}{2M}\left[\frac{g_{k+1}}{g_k}\right]^2 g_k$$

By convexity:

$$F_k = f(x_k) - f^* \leq \langle f'(x_k), x_k - x^*\rangle \leq g_k \mathcal{D}$$

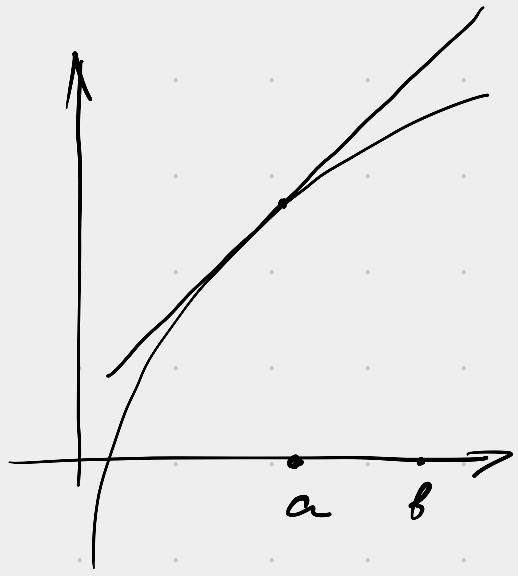
$$\mathcal{D} = \max\{\|x - x^*\| : f(x) \leq f(x_0)\}$$

Progress: $F_k - F_{k+1} \geq \frac{1}{2MD}\left[\frac{g_{k+1}}{g_k}\right]^2 F_k.$ $\approx$ linear rate

$$\Updownarrow$$

$$F_{k+1} \leq F_k \cdot \left(1 - \frac{1}{2MD}\left[\frac{g_{k+1}}{g_k}\right]^2\right)$$

② Concavity of logarithm



$$\log b \leq \log a + \frac{1}{a}(b-a) \qquad \forall a,b>0$$

$$\Updownarrow$$

$$\log a - \log b = \log \frac{a}{b} \geq \frac{1}{a}(a-b)$$

$$\log F_k - \log F_{k+1} \geq \frac{1}{F_k}(F_k - F_{k+1}) \geq \frac{1}{2MD}\left[\frac{g_{k+1}}{g_k}\right]^2$$

③ Telescoping:

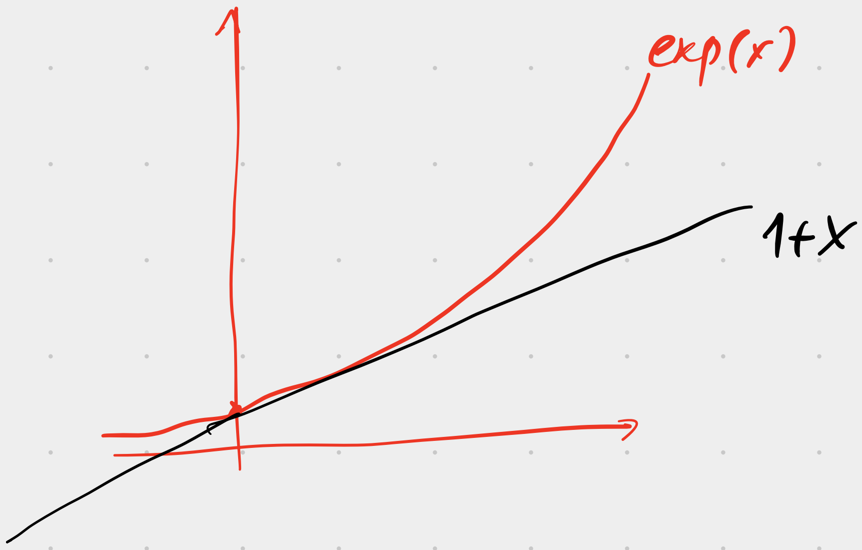
$$\log F_0 - \log F_k \geq \frac{k}{2MD} \underbrace{\frac{1}{k}\sum_{i=0}^{k-1}\left[\frac{g_{i+1}}{g_i}\right]^2}_{\text{arithmetic mean}} \geq$$

Jensen's ineq., using concavity of log

$$\geq \frac{k}{2MD} \underbrace{\left(\prod_{i=0}^{k-1}\left[\frac{g_{i+1}}{g_i}\right]^2\right)^{\frac{1}{k}}}_{\text{geometric mean}} = \frac{k}{2MD}\left[\frac{g_k}{g_0}\right]^{\frac{2}{k}} =$$

$$= \frac{k}{2MD}\exp\left[\frac{2}{k}\log\left[\frac{g_k}{g_0}\right]\right] \geq \frac{k}{2MD}\left[1 + \frac{2}{k}\log\left[\frac{F_k}{g_0 D}\right]\right].$$

exp(·) is convex function:



$\exp(x) \geq 1+x$

$$\log \frac{F_0}{F_n} \geq \frac{K}{2MD} \left[1 + \frac{2}{K} \log\left[\frac{F_K}{g_0 D}\right]\right]$$

1. consider: $\frac{2}{K}\log\frac{F_K}{g_0 D} \leq -\frac{1}{2} \iff F_K \leq \exp\left(-\frac{K}{4}\right)\cdot g_0 D.$

2. Otherwise: $\frac{2}{K}\log\frac{F_K}{g_0 D} \geq -\frac{1}{2} \Rightarrow \log\frac{F_0}{F_n} \geq \frac{K}{4MD} \iff F_K \leq F_0 \exp\left(-\frac{K}{4MD}\right)$

Finally, we get:

Theorem

$$f(x_n) - f^* \leq \overbrace{(f(x_0)-f^*)\exp\left(-\frac{K}{4MD}\right)}^{\leq \frac{\varepsilon}{2}} + \overbrace{\|f'(x_0)\|_* D\cdot\exp\left(-\frac{K}{4}\right)}^{\leq \frac{\varepsilon}{2}}.$$

To find $f(x_n) - f^* \leq \varepsilon$ it's enough:

$$K = O\left(MD \cdot \log\frac{f(x_0)-f^*}{\varepsilon}\right).$$

main complexity parameter

- Taking into account local quadratic convergence region:

$$K = O\left(MD + \log\log\frac{1}{\varepsilon}\right)$$

- Accelerated: $K = O\left((MD)^{\frac{2}{3}} + \log\log\frac{1}{\varepsilon}\right).$

Contracting - Point Acceleration

Fix a prox function: $d: \mathbb{R}^n \to \mathbb{R}$ - convex function.

Bregman Divergence: $\beta_d(x;y) = d(y) - d(x) - \langle d'(x), y-x \rangle \geq 0.$

Main lemma: For any $g: \mathbb{R}^n \to \mathbb{R}$ convex function,

$$v^+ = \arg\min_y \left[g(y) + \beta_d(v;y) \right]$$

$$\forall y: \quad g(y) + \beta_d(v;y) \geq g(v^+) + \beta_d(v;v^+) + \underbrace{\beta_d(v^+;y)}_{\geq 0}$$

Sequences of points:

$\{v_k\}_{k\geq 0}$ - "prox centers"

$\{x_k\}_{k\geq 0}$ - "main iterates"

Coefficients $A_k \geq 0 \nearrow$ $A_0 = 0$

A_0, A_1, A_2 $\qquad a_{k+1} = A_{k+1} - A_k \geq 0 \iff A_k = \sum_{i=1}^{k} a_i$

Define the rate of convergence

Contracting coeff: $\boxed{\gamma_k = \frac{a_{k+1}}{A_{k+1}}}$

let's try to satisfy: For every iteration $k \geq 0$:

$$(*) \quad \beta d(x_0; x) + A_k f(x) \geq \underbrace{\beta d(v_k; x)}_{\geq 0} + A_k f(x_k) \quad \forall x$$

Plug in $x := x^*$: $f(x_k) - f^* \leq \frac{\beta d(x_0; x^*)}{A_k}$.

$k=0$: $A_k = 0$, $x_0 = v_0$.

Assume it holds for $k \geq 0$. For next iterate:

$$\beta d(x_0; x) + A_{k+1} f(x) = \beta d(x_0; x) + A_k f(x) + \underbrace{a_{k+1} f(x)}_{\text{"progress of step"}} \geq$$

Induction

$$\geq \beta d(v_k; x) + A_k f(x_k) + a_{k+1} f(x)$$

Convexity of f

$$\geq \beta d(v_k; x) + A_{k+1} f(\gamma_k x + (1-\gamma_k) x_k) \longrightarrow \min$$

v_{k+1}

Assume:

$$v_{k+1} := \operatorname*{argmin}_x \left[\beta d(v_k; x) + A_{k+1} f(\gamma_k x + (1-\gamma_k) x_k) \right]$$

$$\geq \underbrace{\beta d(v_k; v_{k+1})}_{\geq 0} + A_{k+1} f(\underbrace{\gamma_k v_{k+1} + (1-\gamma_k) x_k}_{x_{k+1}}) + \underline{\beta d(v_{k+1}; x)}$$

Algorithm (Contracting-Point Scheme for Acceleration)

Initialization: $x_0 \in \mathbb{R}^n$, $A_0 = 0$, $x_0 = v_0$, $d(\cdot)$.

Iterate $k \geq 0$:

- Choose $a_{k+1} > 0$. Set $A_{k+1} = A_k + a_{k+1}$. Set $\gamma_k = \frac{a_{k+1}}{A_{k+1}}$.
- Form the contracted objective with Bregman regularizer:

$$h_k(x) = \underbrace{A_{k+1} f(\gamma_k x + (1-\gamma_k) x_k)}_{g_k(x)} + \beta_d(v_k; x)$$

- Compute $v_{k+1} \approx \underset{x}{\operatorname{argmin}}\, h_k(x)$

FGM: linearize $g_k(x)$ around point v_k

- Set $x_{k+1} = \gamma_k v_{k+1} + (1-\gamma_k) x_k$.

<u>Theorem</u> We assume $v_{k+1} = \underset{x}{\operatorname{argmin}}\, h_k(x)$:

$$f(x_k) - f^* \leq \frac{\beta_d(x_0; x^*)}{A_k}.$$

If $\|h_k'(v_{k+1})\|$ is "small" $\Rightarrow$ $f(x_k) - f^* \lesssim \frac{\beta_d(x_0; x^*)}{A_k} + \delta$.

Example How to select a_{k+1}?

$$d(x) = \frac{1}{2}\|x\|_2^2$$

$$h_k(x) = A_{k+1} f(\gamma_k x + (1-\gamma_k) x_k) + \frac{1}{2}\|x - v_k\|^2$$

① Strongly convex, $\mu = 1$

② f has Lipschitz gradient L_f:

$$h_k'(x) = A_{k+1}\gamma_k f'(\gamma_k x + (1-\gamma_k)x_k) + (x - v_k)$$

$$h_k''(x) = A_{k+1}\gamma_k^2 \underbrace{f''(\gamma_k x + (1-\gamma_k)x_k)}_{\preceq L_f I} + I$$

$$L_k = 1 + A_{k+1}\gamma_k^2 L_f = 1 + \frac{a_{k+1}^2}{A_{k+1}} L_f$$

③ Gradient Method on Strongly Convex Smooth Func?

$h_k(\cdot) \to \min \qquad h_k(\bar{x}) - h_k^* \leq \delta$

$$t = O\left(\frac{L_k}{\mu_k}\log\frac{1}{\delta}\right) = O\left(\underbrace{\left(1 + \frac{a_{k+1}^2}{A_{k+1}} L_f\right)}_{\text{let's make it const!}}\log\frac{1}{\delta}\right)$$

$$\frac{a_{k+1}^2}{A_{k+1}} L_f = 1 \quad \Rightarrow \quad A_k \approx \frac{k^2}{L_f}$$

Finally: For the outer: $f(x_k) - f^* \lesssim \frac{R^2}{A_k} = \frac{L_f R^2}{k^2}$

inner iteration: $\tilde{O}(1)$.

Second-order Methods

Acceleration of Cubic Newton.

$$d(x) = \frac{1}{3}\|x\|^3.$$

$$\Rightarrow \quad A_k \approx \frac{k^3}{L_2} \quad \Rightarrow \quad f(x_k) - f^* \lesssim \frac{L_2 R^3}{k^3}.$$