Lecture 14: End of sc nth opt

Ann.

* $\geq$ 10/14 hybrid in Siebel 1105
* hybrid lectures have <u>in-person</u> OH

Theorem. Suppose $\|J_w - J_v\| \leq \beta\|w-v\|$

$$\sigma_n(J_0) > 0, \quad \alpha \geq \frac{\sigma_1 \beta \sqrt{5000\, L(\alpha f(w(0)))}}{\sigma_n^3}$$

$$\{L(\alpha f(w(t))), L(\alpha f_0(u(t)))\} \leq L(\underbrace{\alpha f(w(0))}_{\alpha f(w(t))}) \exp\left(-\frac{\alpha^2 \sigma_n^2 t}{2}\right)$$

$$\{\|w(t) - w(0)\|, \|u(t) - u(0)\|\} \leq \frac{12\sigma_1 \sqrt{L(\alpha f(w(0)))}}{\alpha \sigma_n^2}$$

Proof.

Define "good ball" & exit time

$$T := \inf\{t > 0 : \|w(t) - w(0)\| > B\}, \qquad B = \frac{\sigma_n}{2\beta}.$$

Lemma ① $\forall t \in [0, T]$: $\frac{3}{2}\sigma_1 \geq \sigma_1(J_t) \geq \sigma_n(J_t) \geq \frac{\sigma_n}{2}$

Rem. $\dot w(t) = -\alpha J_t \nabla L(\alpha f(w(t)))$; $\frac{d\, \alpha f(w(t))}{dt} = -\alpha^2 \underbrace{J_t^T J_t}_{pd} \nabla L(\alpha \cdots)$

Lemma ② $\dot z(t) = -Q(t) \nabla L(z(t))$, $\lambda_{min}(Q(t)) \geq \lambda$ $\forall t \in [0, \tau]$

$\Rightarrow \forall t \in [0, \tau]$, $L(z(t)) \leq L(z(0)) \exp(-2\lambda t)$.

Rem. Suffices to be λ-sc along GF path.

$\dot w(t) = -\alpha J_t \nabla L(\alpha f(w(t)))$

Lemma ③ $\dot v(t) = -S(t) \nabla L(g(v(t)))$,
$\forall t \in [0, \tau]$ $A \geq \sigma_1(S(t)) \geq \sigma_n(S(t)) \geq B > 0$

$$\Rightarrow \|v(t) - v(0)\| \leq \frac{A\sqrt{2L(g(v(0)))}}{B^2}.$$

(will prove shortly)

$w(t)$, $\forall t \in [0, T]$: ① $\frac{3\sigma_1}{2} \geq \sigma_1(J_t) \geq \sigma_n(J_t) \geq \frac{1}{2}\sigma_n$

② $\Rightarrow L(\alpha f(w(t))) \leq L(\alpha f(w(0))) \exp\left(-2\alpha^2\left(\frac{\sigma_n}{2}\right)^2 t\right)$.

③ $\Rightarrow$
$$\|w(t) - w(0)\| \leq \frac{(3\alpha\sigma_1/2)\sqrt{2L(\alpha f(w(0)))}}{(\alpha\sigma_n/2)^2} = \underbrace{\boxed{\frac{\sigma_1 \sqrt{72\, L(\alpha f(w(0)))}}{\alpha \sigma_n^2}}}_{B'}.$$

Done if $B' \leq B = \frac{\sigma_n}{2\beta}$

suffices
$$\alpha > \frac{\beta \sigma_1 \sqrt{288\, L(\alpha f(w(0)))}}{\sigma_n^3}$$

Assumed this; thus $T = \infty$.

& ① ② ③ hold for all t.

Also hold with better constants for $u(t)$; Recall:

$$f_0(u) = f(x; w(0)) + J_0(u - w(0)),$$

$$\dot u(t) = \frac{d}{du} L(\alpha f_0(u(t))).$$

(**Rem.** Proof <u>degrades</u> for $w(t)$ as compared with $u(t)$.)

---

copy/paste:

Lemma ③ $\dot v(t) = -S(t) \nabla L(g(v(t)))$,
$\forall t \in [0, \tau]$ $A \geq \sigma_1(S(t)) \geq \sigma_n(S(t)) \geq B > 0$

$$\Rightarrow \|v(t) - v(0)\| \leq \frac{A\sqrt{2L(g(v(0)))}}{B^2}.$$

Pf. $\forall t \in [0, \tau]$

$$\begin{aligned}
\|v(t) - v(0)\| &= \left\|\int_0^t \dot v(s)\, ds\right\| \leq \int_0^t \|\dot v(r)\|\, dr \\
&= \int_0^t \|S(r) \nabla L(g(v(r)))\|\, dr \\
&\leq A \int_0^t \|\nabla L(g(v(r)))\|\, dr = A \int_0^t \|g(v(r)) - y\|\, dr \\
&= A \int_0^t \sqrt{2L(g(v(r)))}\, dr \\
&\leq A \int_0^t \sqrt{2L(g(v(0))) \exp(-2B^2 r)}\, dr \\
&= A\sqrt{2L(g(v(0)))} \int_0^t \exp(-B^2 r)\, dr \\
&= A \cdot \sqrt{2L(g(v(0)))} \cdot \left.\frac{\exp(-B^2 r)}{-B^2}\right|_0^t \\
&\leq \frac{A\sqrt{2L(g(v(0)))}}{B^2}.
\end{aligned}$$

Rem ("scale")

$$w \mapsto \alpha f(w/\alpha) =: g_\alpha(w)$$

$$\nabla g_\alpha(w) = \frac{\alpha}{\alpha} \nabla f(w/\alpha) = \nabla f(w/\alpha)$$

$$\nabla^2 g_\alpha(w) = \frac{1}{\alpha} \nabla^2 f(w/\alpha)$$

Rem check theorem for shallow case.

$$f(w) = \sum_j \underbrace{s_j}_{\text{fixed } \pm 1} \sigma(w_j^T x)$$

Suppose β_0-smooth

$$\|J_w - J_v\|^2 = \left\| \begin{bmatrix} - \nabla f(x_1; w)^T - \\ | \\ - \nabla f(x_n; w)^T - \end{bmatrix}^T - \begin{bmatrix} - & v & - \end{bmatrix} \right\|^2$$

$$= \sum_{i,j} \|s_j \sigma'(w_j^T x_i) x_i - s_j \sigma'(v_j^T x_i) x_i\|^2$$

$$= \sum_{i,j} \|x_i\|^2 \cdot |s_j|^2 \cdot \underbrace{\left(\sigma'(w_j^T x_i) - \sigma'(v_j^T x_i)\right)^2}_{\leq \beta_0^2 (w_j^T x_i - v_j^T x_i)^2 \leq \beta_0^2 \|w_j - v_j\|^2 \cdot \|x_i\|^2}$$

$$\leq \sum_{i,j} \|x_i\|^4 \cdot 1 \cdot \beta_0^2 \|w_j - v_j\|^2 = \|X\|_F^4 \beta_0^2 \|w - v\|_F^2$$

$\Rightarrow \beta_0 \cdot \|X\|_F^2$-smooth

$$\sigma_n(J_0^T J_0)^2 = \lambda_{min}(J_0^T J_0)$$

$$(J_0^T J_0)_{ij} = \nabla f(x_i; w_0)^T \nabla f(x_j; w_0) = \sum_{k=1}^m s_k^2 \sigma'(w_k^T x_i) \sigma'(w_k^T x_j) \cdot x_i^T x_j$$

$$\approx m \cdot x_i^T x_j \, \mathbb{E}\, \sigma'(w^T x_i) \sigma'(w^T x_j)$$

$$\Rightarrow \begin{aligned} \sigma_n &\approx \sqrt{m}\, \sigma_n\begin{pmatrix}\text{expected} \\ \text{gram matrix}\end{pmatrix} \\ \sigma_1 &\approx \sqrt{m}\, \sigma_1\begin{pmatrix}\text{expected} \\ \text{gram matrix}\end{pmatrix}\end{aligned}$$

$\Rightarrow$ calculations on theorem $\Rightarrow \alpha \geq \frac{1}{\sqrt{m}}$