STSCI 6940:

Lecture 12 Date: 03/23

Lemma: $\frac{1}{\sqrt{n}} \| \hat{h}^{t+1} - h^{t+1} \|_2 \xrightarrow{P} 0$.

By induction. $\hat{h}^1 = h^1 = x^1$.

Assume that $\frac{1}{\sqrt{n}} \| \hat{h}^s - h^s \|_2 \xrightarrow[n \to +\infty]{P} 0 \quad \forall\, s \leq t$.

$$\hat{h}^{t+1} = A q^t - b_t q^{t-1}$$

$$= A q_{\parallel}^t + IA q_{\perp}^t - b_t q^{t-1}.$$

$$h^{t+1} = P_{t-1}^{\perp} A q_{\perp}^t + \sum_{s=0}^{t} \alpha_{s-1}^t h^s.$$

- $A q_{\parallel}^t$:

$$q_{\parallel}^t = \sum_{s=0}^{t-1} \alpha_s^t q^s.$$

$$A q_{\parallel}^t = \sum_{s=0}^{t-1} \alpha_s^t A q^s.$$

we have $\hat{h}^{s+1} = A q^s - b_s q^{s-1}$.

$$\Rightarrow A q_{\parallel}^t = \sum_{s=0}^{t-1} \alpha_s^t (\hat{h}^{s+1} + b_s q^{s-1}).$$

$$\hat{h}^{t+1} - h^{t+1} = P_{t-1} A q_{\perp}^t + \sum_{s=0}^{t-1} \alpha_s^t (\hat{h}^{s+1} + b_s q^{s-1}).$$
$$- b_t q^{t-1} - \sum_{s=0}^{t} \alpha_{s-1}^t h^s.$$

$$= P_{t-1} A q_{\perp}^t + \sum_{s=0}^{t-1} \alpha_s^t (\hat{h}^s - h^s)$$

$$+ \sum_{s=0}^{t-1} \alpha_s^t b_s \cdot q^{s-1} - b_t q^{t-1}.$$

Let's consider the term $P_{t-1} A q_\perp^t$:

$$P_{t-1} = Q_{t-1} (Q_{t-1}^T Q_{t-1})^{-1} Q_{t-1}^T.$$

$$P_{t-1} A q_\perp^t = Q_{t-1} (\quad)^{-1} Q_{t-1}^T A q_\perp^t.$$

- $A Q_{t-1} = A [q^0, \ldots, q^{t-1}]$

$$A q^s = \hat{h}^{s+1} + b_s q^{s-1}.$$

$$\Rightarrow A Q_{t-1} = \hat{H}_{t-1} + Q_{t-2} \cdot B_{t-1},$$

$$\hat{H}_{t-1} = [\hat{h}^1 | \ldots | \hat{h}^t]$$

$$Q_{t-1} = [q^0 | \ldots | q^{t-1}].$$

$$B_{t-1} = \mathrm{diag}(b_1, \ldots, b_{t-1})$$

$$\Rightarrow P_{t-1} A q_\perp^t = Q_{t-1} (Q_{t-1}^T Q_{t-1})^{-1} [\hat{H}_{t-1}^T q_\perp^t + B_{t-1} \underbrace{Q_{t-2}^T q_\perp^t}_{=0}].$$

$$= Q_{t-1} (\quad)^{-1} H_{t-1}^T q_\perp^t$$

$$+ \{Q_{t-1} (\quad)^{-1} (\hat{H}_{t-1} - H_{t-1}) q_\perp^t\} = A$$

$$\Rightarrow \hat{h}^{t+1} - h^{t+1} = Q_{t-1} (\quad)^{-1} H_{t-1}^T q_\perp^t + A$$

$$+ \sum_{s=0}^{t-1} \alpha_s^t b_s \cdot q^{s-1} - b_t q^{t-1}.$$

$$+ \{\sum_{s=0}^{t-1} \alpha_s^t (\hat{h}^s - h^s)\} = B$$

Assumption : $\frac{1}{n} Q_{t-1}^T Q_{t-1} \succcurlyeq c_0 I$, w.h.p.

$= (\langle q^s, q^{s'} \rangle)_{1 \le s, s' \le t-1}$

under ~~Ass~~umption ; $A \xrightarrow{P} 0$.

By induction. $B \longrightarrow 0$

$$\hat{h}^{t+1} - h^{t+1} = A + B + \sum_{s=1}^{t} c_s q^{s-1}$$

$$c_s = \left[(Q_{t-1}^T Q_{t-1})^{-1} H_{t-1}^T q_\perp^t\right]_s + \alpha_s^t b_s - b_t \mathbb{1}\{s=t\}.$$

$\frac{1}{\sqrt{n}} \| q^{s-1} \|_2$ will be bounded.

So it suffices to $c_s \xrightarrow{P} 0$. $\forall s \le t$.

$$R = \left(\frac{1}{n} Q_{t-1}^T Q_{t-1}\right)^{-1} \text{"of order 1"}$$

we have

$$\left[(Q_{t-1}^T Q_{t-1})^{-1} H_{t-1}^T q_\perp^t\right]_s$$

$$= \frac{1}{n} \sum_{r=0}^{t} R_{sr} \langle h^r, q_\perp^t \rangle.$$

$$= \sum_{r=0}^{t} R_{sr} \cdot \frac{1}{n} \langle h^r, q^t - \sum_{\ell=1}^{t-1} \alpha_\ell^t q^\ell \rangle$$

- we know from Proposition 2 (Geometry of iterates):

$$\frac{1}{n} \langle h^r, q^s \rangle = \frac{1}{n} \langle h^r, f_s(h^s) \rangle.$$

$$\simeq \mathbb{E}[Z^r \cdot f_s(Z^s)].$$

By Gaussian integration by parts:

$$\mathbb{E}[Z^r \cdot f_s(Z^s)] = \mathbb{E}[Z^r \cdot Z^s] \cdot \mathbb{E}[f_s'(Z^s)]$$

$$\simeq \frac{1}{n} \langle h^r, h^s \rangle \cdot b_s$$

$$\simeq \frac{1}{n} \langle q^{r-1}, q^{s-1} \rangle \cdot b_s$$

Therefore,

$$\left[(Q_{t-1}^T Q_{t-1})^{-1} H_{t-1}^T q_\perp^t\right]_s$$

$$= \sum_{r=0}^{t} R_{rs} \left(\frac{1}{n} \langle h^r, q^t \rangle - \sum_{\ell=1}^{t-1} \alpha_\ell^t \frac{1}{n} \langle h^r, q^\ell \rangle \right)$$

$$\simeq \sum_{r=0}^{t} R_{rs} \left\{ \frac{1}{n} \langle q^{r-1}, q^{t-1} \rangle \cdot b_t - \sum_{\ell=1}^{t-1} \alpha_\ell^t \frac{1}{n} \langle q^{r-1}, q^{\ell-1} \rangle b_\ell \right\}$$

Recall

$$R = \left(\frac{1}{n} Q_{t-1}^T Q_{t-1}\right)^{-1}$$

$$(R^{-1})_{r,s} = \frac{1}{n} \langle q^{r-1}, q^{s-1} \rangle.$$

$$= \sum_{r=0}^{t} R_{rs} (R^{-1})_{r,t} \cdot b_t - \sum_{\ell=1}^{t-1} \alpha_\ell^t b_\ell \sum_{r=0}^{t} R_{rs} (R^{-1})_{r,\ell}.$$

$$= \sum_{r=0}^{t} \mathbb{1}\{r = t\}\, b_t - \sum_{\ell} \alpha_\ell^t b_\ell \mathbb{1}\{s = \ell\}$$

$$= b_t \mathbb{1}\{s = t\} - \alpha_s^t b_s.$$

$$\Rightarrow \quad c_s \simeq 0 \quad \forall s < t \text{ by induction.}$$

$$\hat{h}^{t+1} - h^{t+1} = A + B + \sum_{s=1}^{t} c_s q^{s-1}$$

Therefore $\frac{1}{\sqrt{n}} \| \hat{h}^{t+1} - h^{t+1} \|_2 \xrightarrow{P} 0$.

- We can get of the assumption that $Q_{t-1}^T Q_{t-1}$ is invertible by perturbing the functions f_t

$$\tilde{f}_t = f_t + \varepsilon g_t .$$

$g_t \sim N(0, 1)$.

is indep of everything else.

- Can prove now that $Q_{t-1}^T Q_{t-1}$ is invertible w.h.p.
  $\Rightarrow$ apply previous argument.

Then let $\varepsilon \to 0$.

Generalizations:

1. Can take the non-linearities f_t to depend on the entire history of the algorithm:

$$f_t : \mathbb{R}^{t+1} \to \mathbb{R}$$

$f_t(x^0, \ldots, x^t)$ instead of $f_t(x^t)$.

(applied component-wise).

SE:

$$\frac{1}{n} \sum_{i=1}^{t} \phi(x_i^0, \ldots, x_i^t) \to \mathbb{E}\, \phi(Z^0, \ldots, Z^t).$$

$$\mathbb{E}[Z^{t+1} \cdot Z^{s+1}] = \mathbb{E}\{ f_t(Z^0, \ldots, Z^t) \cdot f_s(Z^0, \ldots, Z^s) \}$$

2. We can take $f_t : \mathbb{R}^n \to \mathbb{R}$

There is an appropriate generalization of SE.
(Berthier - Montanari - Nguyen 2017).

3. $A \in \mathbb{R}^{m \times n}$. The iteration becomes:

$$\begin{cases} u^{t+1} = A f_t(u^t) - \_\_\_\_ \\ v^t = A^T g_t(u^t) - \_\_\_\_\_ , \end{cases}$$

There is an appropriate generalization of SE.
(Javanmard - Montanari 2013)

4. A doesn't have to be gaussian.

A_{ij}, independent, centered, variance $\frac{1}{n}$, and

$\to$ $\sqrt{n} A_{ij}$ is subgaussian, $\forall$ i, j.

SE evolution still holds but the proof is more challenging.

- If (f_t) are polynomials the SE holds.
  [Bayati, Montanari, Lelarge 2015].

- SE holds for general Lipschitz (f_t)

  [Chen - Lam 2020].

Proof works by interpolation $A_{ij}^{(t)} = \sqrt{t} A_{ij} + \sqrt{1-t}\, G_{ij}$.

- Many applications:

$\left\{ \begin{array}{l} \text{Rank-one matrix estimation} \\ \text{Lasso, generalized linear models,} \\ \text{Compressed sensing, matrix completion,} \\ \text{optimization of random functions,} \\ \text{Maximum likelihood estimation, ---} \end{array} \right\}.$

- If $A = \frac{\lambda}{n} x x^T + w$

$$x^{t+1} = A\, m^t - b_t\, m^{t-1}.$$

$$x^{t+1} = \lambda\, x \cdot \underbrace{\frac{\langle x, m^t \rangle}{n}}_{\mu_t} + w\, m^t - b_t\, m^{t-1}$$