# Acceleration I: Chebychev Polynomials

Monday, March 9, 2020 8:38 PM

Yulan

Recall the Richardson Iteration (Assme A is symetric)

$$x^{(t)} = x^{(t-1)} - (A x^{(t-1)} - b)$$

$$= (I - A) x^{(t-1)} + b$$

$$= \sum_{k=0}^{t} (I - A)^k b = p_t(A) \cdot b$$

We want $\| p_t(A) . b - x^* \| \leq \varepsilon \| x^* \|$

i.e. $\| (p_t(A) . A - I) x^* \| \leq \varepsilon \| x^* \|$

or $\| I - A p_t(A) \|_{op} \leq \varepsilon$.

i.e. $\| 1 - \lambda(A p_t(A)) \| \leq \varepsilon$

or $\| 1 - \lambda(A) \, p_t(\lambda(A)) \| \leq \varepsilon$

i.e. $\| 1 - \lambda p_t(\lambda) \| \leq \varepsilon \quad \forall$ eigenvalues $\lambda$ of $A$.

$\| 1 - x p_t(x) \| \leq \varepsilon \quad \forall \, x \in [\lambda_{min}, \lambda_{max}]$.

$q(x) = 1 - x p(x) \qquad q(0) = 1$

$q(x) = \left(1 - \frac{x}{\lambda_{max}(A)}\right)^t$ satisfies $\| q(x) \| \leq \left(1 - \frac{\lambda_{min}}{\lambda_{max}}\right)^t$

so $t = O(\lambda_{max} \log \frac{1}{\varepsilon})$ suffices

$q(x) = \quad \lambda_{max}(A))$

So $t = O\left(\frac{\lambda_{max}}{\lambda_{min}} \log \frac{1}{\varepsilon}\right)$ suffices.

$= O\left(\kappa \log \frac{1}{\varepsilon}\right)$.

This is the Richardson Iteration.

Q. Can we use lower degree?

We want a polynomial $q(x)$ $(x \to \infty \Rightarrow q(x) \to \infty)$ with $\|q(x)\|$ as small as possible for $x \in [-1,1]$ (say, after normalizing $[\lambda_{min}, \lambda_{max}] \to [-1,1]$).

Ans. Chebyshev polynomials!

$t^{th}$ C.P. is the degree $t$ poly s.t. $T_t(\cos\theta) = \cos(t\theta)$

$(T_t(x) = \cos(t\cos^{-1}(x))$ for $x \in [-1,1])$

$$\cos(\theta + \alpha) = \cos\theta\cos\alpha - \sin\theta\sin\alpha$$

$$\therefore \cos(t\theta) = \cos((t-1)\theta)\cos\theta - \sin((t-1)\theta)\sin\theta$$

$$\cos((t-2)\theta) = \cos((t-1)\theta)\cos\theta + \sin((t-1)\theta)\sin\theta$$

$$\Rightarrow \cos(t\theta) = 2\cos((t-1)\theta)\cos\theta - \cos((t-2)\theta)$$

$$T_t(x) = 2x\,T_{t-1}(x) - T_{t-2}(x) \quad (*)$$

$T_0(x) = 1 \qquad T_1(x) = x,\quad T_2(x) = 2x^2 - 1 \ \ldots$

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1 \ \dots$$

Note that $T_t(\cosh\theta) = \cosh(t \cosh^{-1}(\theta))$ also holds

since $\cosh\theta = \dfrac{e^{\theta} + e^{-\theta}}{2}$ $\quad \sinh\theta = \dfrac{e^{\theta} - e^{-\theta}}{2}$

satisfies $\cosh(\theta + \alpha) = \cosh(\theta)\cosh(\alpha) + \sinh(\theta)\sinh(\alpha)$

So we get $(*)$ again.

For $x \geq 1$, $\quad T_t(x) = \cosh(t \cosh^{-1}(x))$.

$x \leq -1$ $\quad T_t(x) = (-1)^t \cosh(t \cosh^{-1}(-x))$

Lemma. $\quad T_t(1+\gamma) \geq \frac{1}{2}(1 + \sqrt{2\gamma})^t \qquad \gamma > 0.$

Pf.

$$T_t(x) = \frac{1}{2}\left(e^{t\cosh^{-1}(x)} + e^{-t\cosh^{-1}(x)}\right)$$

$\cosh^{-1}(x) = \ln(x + \sqrt{x^2-1})$ for $x \geq 1$.

$$\geq \frac{1}{2}\left(x + \sqrt{x^2-1}\right)^t$$

$x = 1+\gamma$

$$= \frac{1}{2}\left(1 + \gamma + \sqrt{2\gamma + \gamma^2}\right)^t \geq \frac{1}{2}\left(1 + \sqrt{2\gamma}\right)^t.$$

---

[illegible] degree $t = O(\sqrt{\kappa} \log \frac{1}{\varepsilon})$.

**Thm.** $\exists\, q$ of degree $t = O(\sqrt{\kappa}\log\frac{1}{\varepsilon})$.

**Pf.** shift: $f(x) = \dfrac{\lambda_{max} + \lambda_{min} - 2x}{\lambda_{max} - \lambda_{min}}$

$$f(x) = \begin{cases} -1 & x = \lambda_{max} \\ 1 & x = \lambda_{min} \\ \dfrac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}} & x = 0. \end{cases}$$

$$q(f(x)) = \frac{T_t(f(x))}{T_t(f(0))} \qquad \text{s.t. } q(0) = 1.$$

$\forall x$: $f(x) \in [-1, 1]$ $\quad |T_t(f(x))| \leq 1.$ $\quad (\cos\theta)$.

$$T_t(f(0)) = T_t\left(1 + \frac{2}{\frac{\lambda_{max}}{\lambda_{min}} - 1}\right) = T_t\left(1 + \frac{2}{\kappa - 1}\right)$$

$$\geq \frac{1}{2}\left(1 + \sqrt{\frac{2}{\kappa - 1}}\right)^t$$

$$\therefore |q(x)| \leq \frac{2}{\left(1 + \sqrt{\frac{2}{\kappa-1}}\right)^t} \qquad \forall x \in [-1, 1]$$

i.e. $t = O(\sqrt{\kappa - 1} \cdot \log\frac{1}{\varepsilon})$ suffices.

---

Another, more general proof.

[illegible] degree $d$ s.t. [illegible]

**Thm.** $\forall\, s, d \quad \exists\, p(x)$ of degree $d$ s.t.

$$\max_{x \in [-1,1]} |p(x) - x^s| \le 2e^{-d^2/2s}$$

**Cor.** $\left(1 - \frac{x}{\lambda_{max}}\right)^s$ can be approximated to within $\varepsilon$ by $p(x)$ of degree $\sqrt{s \log \frac{1}{\varepsilon}}$

Since we use $s = O\left(k \log \frac{1}{\varepsilon}\right)$

$p(x)$ of degree $O\left(\sqrt{k} \log \frac{1}{\varepsilon}\right)$ suffices.

**Pf.** (of Thm). Let $Y_s = \sum_{i=1}^{s} Y_i \quad Y_i \sim \{-1, 1\}$ uniform.

$$\boxed{T_a(x) = T_{|a|}(x)}$$

$$\mathbb{E}\, T_{Z_s} = \frac{1}{2}\left(\mathbb{E}\, T_{Z_{s-1}+1} + \mathbb{E}\, T_{Z_{s-1}-1}\right)$$

But we know $x T_z(x) = \frac{1}{2}\left(T_{z+1}(x) + \frac{1}{2} T_{z-1}(x)\right)$

$$\Rightarrow \mathbb{E}\, T_{Z_s} = x\, \mathbb{E}\, T_{Z_{s-1}}(x) = x^s.$$

Let $p(x) = \mathbb{E}\, T_{Z_s}(x) \cdot \mathbb{1}_{|Z_s| \le d}$

Then $\max |p(x) - x^s| = \max \left|\mathbb{E}_Z\, T_Z(x) \cdot \mathbb{1}_{\ldots}\right|$

Then $\max_{x \in [-1,1]} \left| p(x) - x^s \right| = \max_{x \in [-1,1]} \left| \mathbb{E}_{z_s :} T_{z_s}(x) \cdot \mathbb{1}_{|z_s| > d} \right|$

$$= \max_{x \in [-1,1]} \mathbb{E}_{z_s} \left| T_{z_s}(x) \right| \cdot \mathbb{1}_{|z_s| > d}$$

$$\leq P_s \left( |z_s| > d \right)$$

$$\leq 2 e^{-\frac{d^2}{2s}} .$$

---