Last Course Lecture - Chebyshev acceleration.

Suppose we convert $Ax=b$ to some iteration $x_{i+1} = Rx_i + c$, and we get $\{x_i\} \to x = A^{-1}b$.

Given a bunch of steps $x_1, \dots, x_m$, we want to know whether they can be combined to yield

$$y_m = \sum_{i=1}^{m} \gamma_{m,i}\, x_i$$

which is somehow closer to x.

Note. $\sum_{i=1}^{m} \gamma_{m,i} \equiv 1$, since $x_i = x$ is a plausible iteration.

Now we can write

$$y_m - x = \left(\sum \gamma_{m,i} x_i\right) - x$$

$$= \sum \gamma_{m,i} (x_i - x)$$

$$= \sum \gamma_{m,i} R^i (x_0 - x)$$

$$= p_m(R)(x_0 - x)$$

where $p_m(R) = \sum_{i=0}^{m} \gamma_{m,i} R^i$ is a polynomial of degree m in R with $p(1) = \sum \gamma_{m,i} = 1$. Now we would like

$$p_m(R) = 0,$$

so guessing that $p_m(R)$ = characteristic polynomial of R would work. There are two problems.

a) We don't know the characteristic polynomial.

b) R is very large ($\sim 10^6 \times \sim 10^6$) so this takes too many steps to get started anyway.

So let's think about what we do know about R. We ~~know~~ might know

1) the eigenvalues of R are real

2) the eigenvalues lie in $(-\rho, \rho) \subsetneq [-1, 1)$.

In that case, we want to choose a polynomial p_m so that

i) $p_m(1) = 1$

ii) $\max_{x \in (-\rho, \rho)} |p_m(x)|$ is as small as possible.

By the spectral mapping theorem, that makes eigenvals of $p_m(R)$ as small as possible!

But these properties are exactly what we wanted (so long ago!) from the Chebyshev polynomials. We recall

Definition. The mth Chebyshev ~~approximation~~ polynomial $T_m(x)$ is defined by

$$T_m(x) = 2x\,T_{m-1}(x) - T_{m-2}(x)$$

where $T_0(x) = 1$ and $T_1(x) = x$.

Note that $T_m(1) = 2T_{m-1}(1) - T_{m-2}(1)$ so it is easy to show $T_m(1) = 1$ by induction.

So suppose we know ρ.
(Really, this is just an <u>overestimate</u> of $\rho(R)$, so we could guess.) Set

$$p_m(x) = T_m(x/\rho) \big/ T_m(1/\rho).$$

Clearly $p_m(1) = 1$. ~~But if $x < \rho$, we can observe $T_m($ Now suppo~~
Now we claim $|T_m(x)| \leq 1$ if $|x| \leq 1$ (again, this follows by induction), so for $x \in (-\rho, \rho)$ we have

$$|p_m(x)| < 1/T_m(1/\rho).$$

But for $\rho = \frac{1}{1+\epsilon}$, this is $1/T_m(1+\epsilon)$, which turns out to be ~~gigantic!~~ tiny!

For instance, if

	$\epsilon = 10^{-4}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-2}$
$m = 1000$	$\sim 10^{-5}$	$\sim 10^{-19}$	$\sim 10^{-61}$

So now how do we implement?

For convenience, let $\mu_m = \frac{1}{T_m(1/\rho)}$ so $p_m(R) = \mu_m T_m(R/\rho)$. Now

$$T_m(1/\rho) = (2/\rho)\, T_{m-1}(1/\rho) - T_{m-2}(1/\rho)$$

by the recurrence for Chebyshev polynomials. So

$$\frac{1}{\mu_m} = \frac{2}{\rho\, \mu_{m-1}} - \frac{1}{\mu_{m-2}}$$

Then once we get started,

$$y_m - x = p_m(R)(x_0 - x)$$
$$= \mu_m T_m\left(\frac{R}{\rho}\right)(x_0 - x)$$

$$= \nu_m \left[2\frac{R}{\rho} T_{m-1}\left(\frac{R}{\rho}\right)(x_0 - x) - T_{m-2}\left(\frac{R}{\rho}\right)(x_0 - x) \right]$$

$$= \nu_m \left[2\frac{R}{\rho} \frac{P_{m-1}(R/\rho)(x_0 - x)}{\nu_{m-1}} - \frac{P_{m-2}(R/\rho)(x_0 - x)}{\nu_{m-2}} \right]$$

$$= \nu_m \left[2\frac{R}{\rho} \frac{y_{m-1} - x}{\nu_{m-1}} - \frac{y_{m-2} - x}{\nu_{m-2}} \right]$$

Now we don't know x, so this is <u>not</u> yet a useful formula. But we can compute by pulling out y_{m-1}, y_{m-2} terms

$$y_m = \frac{2\nu_m}{\nu_{m-1}} \frac{R}{\rho} y_{m-1} - \frac{\nu_m}{\nu_{m-2}} y_{m-2} + d_m$$

where

$$d_m = x - \frac{2\nu_m}{\nu_{m-1}} \left(\frac{R}{\rho}\right) x + \frac{\nu_m}{\nu_{m-2}} x$$

Now x is a fixed point of the

iteration, so $x = Rx + c$, or $Rx = x - c$.

So this

$$= x - \frac{2\mu_m}{\mu_{m-1}}\left(\frac{x-c}{\rho}\right) + \frac{\mu_m}{\mu_{m-2}}x$$

$$= \underbrace{\cancel{\mu_m\left(\frac{1}{\mu_m} - \frac{2}{\rho\mu_{m-1}} + \frac{1}{\mu_{m-2}}\right)}}_{0} x + \frac{2\mu_m}{\rho\mu_{m-1}}c$$

$$= \frac{2\mu_m}{\rho\mu_{m-1}}c.$$

Thus the Chebyshev iteration is, at long last,

$$\mu_0 = 1,\ \mu_1 = \rho,\ y_0 = x_0,\ y_1 = Rx_0 + c$$

and

$$\mu_m = 1\Big/\left(\frac{2}{\rho\mu_{m-1}} - \frac{1}{\mu_{m-2}}\right)$$

$$y_m = \frac{2\mu_m}{\rho\mu_{m-1}}Ry_{m-1} - \frac{\mu_m}{\mu_{m-2}}y_{m-2} + \frac{2\mu_m}{\rho\mu_{m-1}}c.$$

Note that we're really still only doing one matrix multiply per iteration, so we are about as fast as another iterator!

$\boxed{\text{Thank You!}}$