MAT 228A Notes

Sam Fleischer

November 22, 2016

1 Conjugate Gradient Method

- Solving $Au = f$ where A is symmetric positive definite.
- A is symmetric if $A = A^T$
- A is positive definite if $y^T Ay > 0$ for all $y \neq 0$.
  - This means all eigenvalues are positive, real, and eigenvectors corresponding to different eigenvalues are orthogonal.
- We can solve using matrix-vector product.
- Given u, compute Au. We don't need A, only how to get Au.
- CG is an example of a Krylov method

CG is related to minimization. Suppose A is s.p.d., $N \times N$, and define the functional $\phi : \mathbb{R}^n \to \mathbb{R}$ by

$$\phi(u) = \frac{1}{2}u^T Au - u^T f$$

Solution to $Au = f$ is the minimizer of ϕ.

$$\boldsymbol{\nabla}\phi = \frac{1}{2}Au + \frac{1}{2}A^T u - f = Au - f$$

since $A = A^T$. So $\boldsymbol{\nabla}\phi = 0 \iff Au = f$. Is it a minimum or a maximum?

$$\boldsymbol{\nabla}\boldsymbol{\nabla}\phi = A$$

Because A is positive definite, we know ϕ is convex with positive concavity. That is, the solution to $Au = f$ is the minimizer of ϕ. We can use the method of steepest descents. Have u_k be the kth iterate. We want a method for generating u_{k+1}.

Remember $\boldsymbol{\nabla}\phi(u_k)$ points in the direction of greatest ascent. But

$$\boldsymbol{\nabla}\phi(u_k) = Au_k - f = -(f - Au_k) = -r_k$$

Thus we should move in the direction of the residual since that is the direction of greatest descent. So,

$$u_{k+1} = u_k + \alpha r_k$$

where α is a parameter. Actually, we want to choose α such that $\phi(u_{k+1}) = \min_\alpha \phi(u_k + \alpha r_k)$. How do we minimize?

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\phi(u + \alpha r)$$

and set equal to 0 to find α.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}\phi(u + \alpha r) &= \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\frac{1}{2}(u + \alpha r)^T A(u + \alpha r) - (u + \alpha r)^T f\right] \\
&= \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\left(\frac{1}{2}u^T Au - u^T f\right) + \left(\underbrace{\frac{1}{2}r^T Au + \frac{1}{2}u^T Ar}_{\text{equal since } A \text{ is symmetric}} - r^T f\right)\alpha + \left(\frac{1}{2}r^T Ar\right)\alpha^2\right] \\
&= r^T Au - r^T f + \alpha r^T Ar \\
&= r^T(Au - f) + \alpha r^T Ar \\
&= -r^T r + \alpha r^T Ar
\end{aligned}$$

Setting this equal to 0 gives

$$\alpha^* = \frac{r^T r}{r^T A r}$$

And we see the second derivative with respect to α is positive and thus α^* is a minimum. Thus,

$$u_{k+1} = u_k + \frac{r_k^T r_k}{r_k^T A r_k} u_k$$

The steepest descent algorithm:

- Initialize u_0 (guess)
- Loop through k
  - compute $r_k = f - Au_k$
  - check $\|r_k\|$ for stopping (stop if below a tolerance) Why is this the criterion? The residual is available, might as well use it.
  - compute $\alpha = \frac{r_k^T r_k}{r_k^T A r_k}$
  - $u_{k+1} = u_k + \alpha r_k$

How expensive?

- 2 matrix-vector products

$$r_{k+1} = f - Au_{k+1} = f - A(u_k + \alpha r_k) = f - Au_k - \alpha A r_k = r_k - \alpha A r_k \tag{1}$$

So a more efficient method (implementation) is

- Initialize u_0 and $r_0 = f - Au_0$
- Loop in k
  - check for stopping $\|r_k\|$
  - compute $w = Ar_k$
  - compute $\alpha = \frac{r_k^T r_k}{r^k T w}$
  - update $u_{k+1} = u_k + \alpha r_k$
  - update the residual $r_{k+1} = r_k - \alpha w$

How expensive?

- 1 matrix-vector product

Note that successive residuals are always orthogonal to each other. The shape of the level curves are related to the eigenvalues. There are special points on level curves where the residual points directly at the minimum. If we start at one of these special points, we are done after just one iteration. If we start at one of these special points, we know the residual r is proportional to the error. That is,

$$r = \lambda(u - v) \tag{2}$$

where u is the minimum and v is the starting guess. But

$$f - Av = \lambda(u - v) \tag{3}$$
$$A(u - v) = \lambda(u - v) \tag{4}$$

This tells us that the residual at these points are eigenvectors.

Suppose $\frac{\lambda_2}{\lambda_1} \approx 1$. Then the method converges quickly since the level curves are small perturbations of circles. Suppose $\frac{\lambda_2}{\lambda_1} \gg 1$. Then the two radii of the level curves are drastically different, producing a long skinny ellipse. If we start at the vertices, great, we are done in one step. But if we start really close the major vertex, we will take lots of tiny steps as we approach the minimum, i.e. very slow convergence. So the convergence rate depends on the ratio of the largest to smallest eigenvalue. Define κ as

$$\kappa = \|A\|_2 \left\|A^{-1}\right\|_2 \tag{5}$$

and call it the "condition number." For symmetric matrices A,

$$\kappa = \frac{\max_k |\lambda_k|}{\min_k |\lambda_k|} \tag{6}$$