Overview

- Symmetric positive definite matrices
- Gradient descent
- Conjugate gradient

- Consider $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ with $\boldsymbol{A}\in\mathbb{R}^{m\times m}$, symmetric positive definite, (spd)
- Recast linear system as an optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^m}\varphi(\boldsymbol{x}),\ \varphi(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}-\boldsymbol{x}^T\boldsymbol{b}=\frac{1}{2}x_i A_{ij} x_j - x_i b_i$$

- Stationarity condition $\nabla\varphi(\boldsymbol{x})=\boldsymbol{0}$, $\nabla\varphi=\left(\begin{array}{ccc}\partial\varphi/\partial x_1 & \dots & \partial\varphi/\partial x_m\end{array}\right)$, a row vector

$$(\nabla\varphi(\boldsymbol{x}))_k=\frac{\partial\varphi}{\partial x_k}=\varphi_{,k}=\frac{1}{2}x_{i,k}A_{ij}x_j+\frac{1}{2}x_iA_{ij}x_{j,k}-x_{i,k}b_i$$

$$(\nabla\varphi(\boldsymbol{x}))_k=\frac{1}{2}\delta_{ik}A_{ij}x_j+\frac{1}{2}x_iA_{ij}\delta_{jk}-\delta_{ik}b_i=\frac{1}{2}A_{kj}x_j+\frac{1}{2}x_iA_{ik}-b_k$$

$$(\nabla\varphi(\boldsymbol{x}))_k=\frac{1}{2}A_{kj}x_j+\frac{1}{2}A_{ki}x_i-b_k=A_{ki}x_i-b_k=0\Rightarrow\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

- Solution of quadratic optimization function gives solution to linear system

- Construct sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \ldots$ to minimize $\varphi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, by

$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \alpha_n \nabla\varphi(\boldsymbol{x}_{n-1}) = \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{r}_{n-1} = \boldsymbol{x}_{n-1} + \alpha_n (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{n-1})$$

- Go along gradient to mininum of restriction $\varphi_{n-1}(\alpha_n) = \varphi(\boldsymbol{x}_{n-1} + \alpha_n(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{n-1}))$

$$\frac{\mathrm{d}\varphi_{n-1}}{\mathrm{d}\alpha_n} = \nabla\varphi(\boldsymbol{x}_n)\frac{\mathrm{d}\boldsymbol{x}_n}{\mathrm{d}\alpha_n} = -\boldsymbol{r}_n^T\boldsymbol{r}_{n-1} = 0,$$

until residual at new position becomes orthogonal to that at old position

$$\boldsymbol{r}_n^T\boldsymbol{r}_{n-1} = 0 \Rightarrow [\boldsymbol{b}^T - (\boldsymbol{x}_{n-1}^T + \alpha_n \boldsymbol{r}_{n-1}^T)\boldsymbol{A}]\,\boldsymbol{r}_{n-1} = 0 \Rightarrow \alpha_n = \frac{\boldsymbol{r}_{n-1}^T \boldsymbol{r}_{n-1}}{\boldsymbol{r}_{n-1}^T\boldsymbol{A}\boldsymbol{r}_{n-1}}$$

Algorithm

$\boldsymbol{x}_0 = \boldsymbol{0}$, $\boldsymbol{r}_0 = \boldsymbol{b}$,
for $n = 1, 2, 3, \ldots$
$\quad\alpha_n = (\boldsymbol{r}_{n-1}^T \boldsymbol{r}_{n-1}) / (\boldsymbol{r}_{n-1}^T\boldsymbol{A}\boldsymbol{r}_{n-1})$
$\quad\boldsymbol{x}_n = \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{r}_{n-1}$
$\quad\boldsymbol{r}_n = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n = \boldsymbol{r}_{n-1} - \alpha_n \boldsymbol{A}\boldsymbol{r}_{n-1}$

- Hestenes & Stiefel, 1952: choose a different search direction $\boldsymbol{p}_n$ such that

$$\boldsymbol{r}_n^T \boldsymbol{r}_{n-1} = 0, \boldsymbol{p}_n^T \boldsymbol{A} \boldsymbol{p}_{n-1} = 0,$$

i.e., residuals are orthogonal, search directions are "conjugate" ($\boldsymbol{A}$-weighted scalar product)

Algorithm

$\boldsymbol{x}_0 = \boldsymbol{0}$, $\boldsymbol{r}_0 = \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{r}_0$
for $n = 1, 2, 3, \ldots$
$\alpha_n = (\boldsymbol{r}_{n-1}^T \boldsymbol{r}_{n-1}) / (\boldsymbol{p}_{n-1}^T \boldsymbol{A} \boldsymbol{p}_{n-1})$
$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{p}_{n-1}$
$\boldsymbol{r}_n = \boldsymbol{r}_{n-1} - \alpha_n \boldsymbol{A} \boldsymbol{p}_{n-1}$
$\beta_n = (\boldsymbol{r}_n^T \boldsymbol{r}_n) / (\boldsymbol{r}_{n-1}^T \boldsymbol{r}_{n-1})$
$\boldsymbol{p}_n = \boldsymbol{r}_n + \beta_n \boldsymbol{p}_{n-1}$

```
octave] m=128; x=zeros(m,1); b=ones(m,1);
octave] maxiter=1000; r=b; p=r; D=diag(1:m);
octave] H=randn(m); [Q,R]=qr(H); A=Q*D*Q';
octave] for n=1:maxiter
          nrmr2 = r'*r;
          alpha = nrmr2/(p'*A*p);
          x = x + alpha*p; r = r - alpha*A*p;
          beta = r'*r/nrmr2;
          p = r + beta*p;
```

- Conjugate gradient (CG) is an optimal Krylov method

$$\mathcal{K}_n = \langle \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \rangle = \langle \boldsymbol{p}_0, \ldots, \boldsymbol{p}_{n-1} \rangle = \langle \boldsymbol{r}_0, \ldots, \boldsymbol{r}_{n-1} \rangle = \langle \boldsymbol{b}, \ldots, \boldsymbol{A}^{n-1} \boldsymbol{b} \rangle$$

$$\boldsymbol{x}_n = \arg\min_{\boldsymbol{y} \in \mathcal{K}_n} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{A}}, \boldsymbol{e}_n = \boldsymbol{x} - \boldsymbol{x}_n, \|\boldsymbol{e}_n\|_{\boldsymbol{A}} \leqslant \|\boldsymbol{e}_{n-1}\|_{\boldsymbol{A}}, \|\boldsymbol{z}\|_{\boldsymbol{A}} = (\boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z})^{1/2}.$$

- As in GMRES for general matrices, CG solves a polynomial approximation problem

$$\min_{p_n \in P_n} \|p_n(\boldsymbol{A})\, \boldsymbol{e}_0\|_{\boldsymbol{A}}$$

- Similar to GMRES, convergence is given by norm of characteristic polynomial of reduced Hessenberg matrix on spectrum of $\boldsymbol{A}$

$$\frac{\|\boldsymbol{e}_n\|_{\boldsymbol{A}}}{\|\boldsymbol{e}_0\|_{\boldsymbol{A}}} = \inf_{p \in P_n} \frac{\|p_n(\boldsymbol{A})\, \boldsymbol{e}_0\|_{\boldsymbol{A}}}{\|\boldsymbol{e}_0\|_{\boldsymbol{A}}} \leqslant \inf_{p \in P_n} \max_{\lambda \in \Lambda(\boldsymbol{A})} |p(\lambda)|$$

- Rate of convergence

$$\frac{\|\boldsymbol{e}_n\|_{\boldsymbol{A}}}{\|\boldsymbol{e}_0\|_{\boldsymbol{A}}} \leqslant 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n = 2r^n$$

Example: Crank-Nicolson $u_t = u_{xx}$, $m \gg 1$, say $m = 256\pi \Rightarrow \kappa = 256$, $r = 0.88$, $2r^{100} \cong 10^{-5}$

$$\kappa = \left|\frac{\lambda_{\max}}{\lambda_{\min}}\right| = \frac{1 + 4\sigma \sin^2\left(\frac{m\pi}{2(m+1)}\right)}{1 + 4\sigma \sin^2\left(\frac{\pi}{2(m+1)}\right)} \cong \frac{1 + 4\sigma}{1 + \sigma \frac{\pi^2}{(m+1)^2}} \cong 1 + 4\sigma = 1 + \frac{2\,\delta t}{\delta x^2} \cong (\delta x)^{-1} = \frac{m}{\pi}$$