

# On the resolution of linear systems

Laurence HALPERN

LAGA - Université Paris 13

Hanoi, February 2013

1 / 40

- 1 Direct methods
- 2 Stationary iterative methods
- 3 Non-Stationary iterative methods
- 4 Preconditioning

2 / 40

# Purpose

Solve  $AX = b$ .

- $A$  is a squared matrix,
- $b$  is a given righthand side, or a family of given righthand sides

3 / 40

# Description

$$\underbrace{\begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & -1 \\ 3 & 11 & 6 \end{pmatrix}}_A \underbrace{\begin{pmatrix} \\ \\ \end{pmatrix}}_X = \underbrace{\begin{pmatrix} 9 \\ 1 \\ 36 \end{pmatrix}}_b$$

$$\left( \begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 6 & 36 \end{array} \right) \rightarrow \left( \begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 3 & 9 \end{array} \right) \rightarrow \left( \begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 1 & 1 \end{array} \right)$$

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -3 & 1 & 1 \end{pmatrix}}_M \underbrace{\left( \begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 6 & 36 \end{array} \right)}_{(A|b)} = \underbrace{\left( \begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 1 & 1 \end{array} \right)}_{(U|Mb)}$$

5 / 40

$$Ax = b \iff Ux : MAx = Mb$$

$M$  is a **preconditioner**

$$M = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -3 & 1 & 1 \end{pmatrix} \longrightarrow L := M^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & -1 & 1 \end{pmatrix}$$

$$U = MA \iff A = LU, Ax = b \iff LUx = b$$

- ①  $LU$  decomposition  $\mathcal{O}(\frac{2n^3}{3})$  elementary operations.
- ② Solve  $Ly = b$   $\mathcal{O}(n^2)$  elementary operations.
- ③ Solve  $Ux = y$   $\mathcal{O}(n^2)$  elementary operations.

For  $P$  values of the righthand side,  $N_{op} \sim \frac{2n^3}{3} + P \times 2n^2$ .

## Theoretical results

**Theorem 1** Let  $A$  be an invertible matrix, with principal minors  $\neq 0$ . Then there exists a unique matrix  $L$  lower triangular with  $l_{ii} = 1$  for all  $i$ , and a unique matrix  $U$  upper triangular, such that  $A = LU$ . Furthermore  $\det(A) = \prod_{i=1}^n u_{ii}$ .

**Theorem 2** Let  $A$  be an invertible matrix. There exist a permutation matrix  $P$ , a matrix  $L$  lower triangular with  $l_{ii} = 1$  for all  $i$ , and a matrix  $U$  upper triangular, such that

$$PA = LU$$

# Sparse and banded matrices

$$\begin{array}{c}
 \text{p}=3 \\
 \longleftrightarrow \\
 \begin{array}{c}
 \text{q}=2 \\
 \updownarrow
 \end{array}
 \left( \begin{array}{ccccccc}
 2 & 1 & 0 & -1 & 0 & 0 & 0 \\
 -4 & 2 & 3 & 0 & 0 & 0 & 0 \\
 0 & -12 & 3 & 1 & 2 & 0 & 0 \\
 0 & 0 & -24 & 4 & -7 & 0 & 0 \\
 0 & 0 & -40 & 0 & 5 & 1 & 4 \\
 0 & 0 & 0 & 0 & -60 & 6 & -23 \\
 0 & 0 & 0 & 0 & 0 & -84 & 0
 \end{array} \right)
 \end{array}$$

A banded matrix, upper bandwidth  $p = 3$  and lower bandwidth  $q = 2$ , in total  $p + q + 1$  nonzero diagonals.

8 / 40

# Sparse and banded matrices

$$U = \left( \begin{array}{ccccccc}
 2 & 1 & 0 & -1 & 0 & 0 & 0 \\
 0 & 4 & 3 & -2 & 0 & 0 & 0 \\
 0 & 0 & 12 & -5 & 2 & 0 & 0 \\
 0 & 0 & 0 & -6 & -3 & 0 & 0 \\
 0 & 0 & 0 & 0 & 20 & 1 & 4 \\
 0 & 0 & 0 & 0 & 0 & 9 & -11 \\
 0 & 0 & 0 & 0 & 0 & 0 & -102.7
 \end{array} \right)$$

$$L = \left( \begin{array}{ccccccc}
 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -2 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & -3 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & -2 & 1 & 0 & 0 & 0 \\
 0 & 0 & -3.3 & 2.81 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & -3 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & -9.3 & 1
 \end{array} \right)$$

$L$  lowerbanded  $q = 2$ , and  $U$  upperbanded  $p = 3$ .

9 / 40

## Sparse and banded matrices with pivoting

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -0.5 & -0.17 & -0.05 & -0.21 & 0.025 & 0.0027 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} -4 & 2 & 3 & 0 & 0 & 0 & 0 \\ 0 & -12 & 3 & 1 & 2 & 0 & 0 \\ 0 & 0 & -40 & 0 & 5 & 1 & 4 \\ 0 & 0 & 0 & 4 & -10 & -0.6 & -2.4 \\ 0 & 0 & 0 & 0 & -60 & 6 & -23 \\ 0 & 0 & 0 & 0 & 0 & -84 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.275 \end{pmatrix}$$

10 / 40

## The permutation matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

11 / 40

## Stationary iterative methods

$$AX = b; \quad A = M - N; \quad MX = NX + b,$$

$$MX^{m+1} = NX^m + b.$$

Use  $A = D - E - F$ .

- ① Jacobi :  $M = D$  diagonal part of  $A$ .
- ② Gauss-Seidel :  $M = D - E$  lower part of  $A$ .
- ③ Relaxation :  $\hat{U}^{m+1}$  obtained by Gauss-Seidel,

$$X^{m+1} = \omega \hat{U}^{m+1} + (1 - \omega)X^m.$$

$$M = \frac{1}{\omega}D - E, \quad N = F + \frac{1 - \omega}{\omega}D - E$$

- ④ Richardson algorithm

$$X^{m+1} = X^m - \rho r^m = X^m - \rho(AX^m - b)$$

$$M = \frac{1}{\rho}I \quad \rho_{opt} = \frac{2}{\lambda_1 + \lambda_n}$$

13 / 40

## Stationary methods, continue

$$MX^{m+1} = NX^m + b \iff MX^{m+1} = (M - A)X^m + b$$

$$\iff X^{m+1} = (I - M^{-1}A)X^m + M^{-1}b$$

$$\iff \text{fixed point algorithm to solve } M^{-1}AX = M^{-1}b$$

Preconditioning

$$AX = b \iff M^{-1}AX = M^{-1}b$$

$$\iff X = (I - M^{-1}A)X + M^{-1}b$$

14 / 40

## Stationary methods, continue

$$\begin{aligned} \text{Error } e^m &:= X - X^m, \\ \text{Residual } r^m &:= b - AX^m = AX - AX^m = Ae^m. \end{aligned}$$

$$MX^{m+1} = NX^m + b$$

$$MX = NX + b$$

$$Me^{m+1} = Ne^m$$

$$e^{m+1} = M^{-1}Ne^m$$

$R = M^{-1}N$  is the iteration matrix

Useful alternative formula  $R = I - M^{-1}A$ .

15 / 40

## Fundamentals tools

$$X^{m+1} = RX^m + \tilde{b}, \quad e^{m+1} = Re^m, \quad R = M^{-1}N.$$

**Theorem** The sequence is convergent for any initial guess  $X^0$  if and only if  $\rho(R) < 1$ .

$\rho(R) = \max\{|\lambda|, \lambda \text{ eigenvalue of } A\}$  : convergence factor.

To reduce the initial error by a factor  $\epsilon$ , we need

$$\frac{\|e^m\|}{\|e^0\|} \leq \sim (\rho(R))^m \sim \epsilon$$

So we have  $M \sim \frac{\ln \epsilon}{\ln \rho(R)}$ .

Convergence rate =  $-\ln \rho(R)$  digits per iteration.

16 / 40

# Symmetric positive definite matrices

**Householder-John theorem** : Suppose  $A$  is positive. If  $M + M^T - A$  is positive definite, then  $\rho(R) < 1$ .

## Corollary

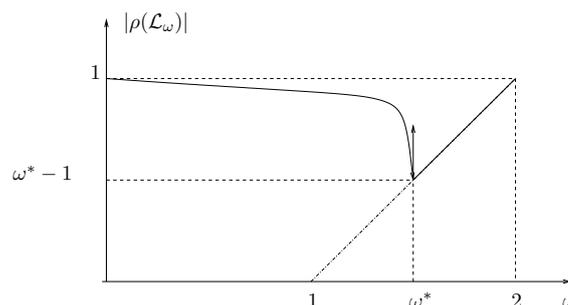
- ① If  $D + E + F$  is positive definite, then Jacobi converges.
- ② If  $\omega \in (0, 2)$ , then SOR converges.

17 / 40

# Tridiagonale matrices

- ①  $\rho(\mathcal{L}_1) = (\rho(J))^2$  : Jacobi Gauss-Seidel converge or diverge simultaneously. If convergent, Gauss-Seidel is twice as fast.
- ② Suppose the eigenvalues of  $J$  are real. Then Jacobi and SOR convergent ou converge or diverge simultaneously for  $\omega \in ]0, 2[$ .
- ③ Same assumptions, SOR has an optimal parameter

$$\omega^* = \frac{2}{1 + \sqrt{1 - (\rho(J))^2}}, \quad \rho(\mathcal{L}_{\omega^*}) = \omega^* - 1.$$



18 / 40

## Descent methods

The descent directions  $p_m$  are given. Define

$$X^{m+1} = X^m + \alpha_m p^m, \quad e^{m+1} = e^m - \alpha_m p^m, \quad r^{m+1} = r^m - \alpha_m A p^m.$$

**Theorem**  $X$  is the solution of  $AX = b \iff$  it minimizes over  $\mathbb{R}^N$  the functional  $J(y) = \frac{1}{2}(Ay, y) - (b, y)$ .

This is equivalent to minimizing  $G(y) = \frac{1}{2}(A(y - X), y - X)$

At each step minimize  $J$  in the direction of  $p_m$

$$\alpha_m = \frac{(p^m, r^m)}{(Ap^m, p^m)}, \quad (p^m, r^{m+1}) = 0$$

$$G(x^{m+1}) = G(x^m)(1 - \mu_m), \quad \mu_m = \frac{(r^m, p^m)^2}{(Ap^m, p^m)(A^{-1}r^m, r^m)}$$

20 / 40

## Steepest descent

$$p^m = r^m.$$

$$X^{m+1} = X^m + \alpha_m r^m, \quad e^{m+1} = e^m - \alpha_m r^m, \quad r^{m+1} = (I - \alpha_m A)r^m.$$

$$\alpha_m = \frac{\|r^m\|^2}{(Ar^m, r^m)}, \quad (r^m, r^{m+1}) = 0$$

$$G(x^{m+1}) = G(x^m) \left( 1 - \frac{\|r^m\|^4}{(Ar^m, r^m)(A^{-1}r^m, r^m)} \right) \leq \left( \frac{\kappa(A) - 1}{\kappa(A) + 1} \right)^2 G(x^m)$$

21 / 40

## Conjugate gradient

$$X^{m+1} = X^m + \alpha_m p^m, \quad (r^m, p^{m-1}) = 0.$$

Search  $p^m$  as  $p^m = r^m + \beta_m p^{m-1}$

$$G(x^{m+1}) = G(x^m)(1 - \mu_m)$$

$$\mu_m = \frac{(r^m, p^m)^2}{(Ap^m, p^m)(A^{-1}r^m, r^m)} = \frac{\|r^m\|^4}{(Ap^m, p^m)(A^{-1}r^m, r^m)}$$

Maximize  $\mu_m$ , or minimize

$$(Ap^m, p^m) = \beta_m^2 (Ap^{m-1}, p^{m-1}) + 2\beta_m (Ap^{m-1}, r^m) + (Ar^m, r^m)$$

$$\beta_m = -\frac{(Ap^{m-1}, r^m)}{Ap^{m-1}, p^{m-1}} \Rightarrow (Ap^{m-1}, p^m) = 0$$

$$(r^m, r^{m+1}) = 0, \quad \beta_m = \frac{\|r^m\|^2}{\|r^{m-1}\|^2}.$$

22 / 40

## Other properties

Choose  $p^0 = r^0$ . Then  $\forall m \geq 1$ , if  $r^i \neq 0$  for  $i < m$ .

- ①  $(r^m, p^i) = 0$  for  $i \leq m - 1$ .
- ②  $\text{vec}(r^0, \dots, r^m) = \text{vec}(r^0, Ar^0, \dots, A^m r^0)$ .
- ③  $\text{vec}(p^0, \dots, p^m) = \text{vec}(r^0, Ar^0, \dots, A^m r^0)$ .
- ④  $(p^m, Ap^i) = 0$  for  $i \leq m - 1$ .
- ⑤  $(r^m, r^i) = 0$  for  $i \leq m - 1$ .

**Definition** Krylov space  $\mathcal{K}_m = \text{vec}(r^0, Ar^0, \dots, A^{m-1}r^0)$ .

**Theorem**  $G(x^m) = \inf_{y \in x^0 + \mathcal{K}_m} G(y)$ .

23 / 40

## Final properties

**Theorem** Convergence in at most  $N$  steps (size of the matrix)

**Theorem** 
$$G(x^m) \leq 4 \left( \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^2 G(x^m)$$

24 / 40

## The algorithm

$$X^0 \text{ chosen, } p^0 = r^0 = b - AX^0.$$

While  $m < Niter$  or  $\|r^m\| \geq tol$ , do

$$\begin{aligned} \alpha_m &= \frac{\|r^m\|^2}{(Ap^m, p^m)}, \\ X^{m+1} &= X^m + \alpha_m p^m, \\ r^{m+1} &= r^m - \alpha_m Ap^m, \\ \beta_{m+1} &= \frac{\|r^{m+1}\|^2}{\|r^m\|^2}, \\ p^{m+1} &= r^{m+1} - \beta_{m+1} p^m. \end{aligned}$$

25 / 40

## 1-D Poisson problem

Poisson equation  $-u'' = f$  on  $(0, 1)$ ,

Dirichlet boundary conditions  $u(0) = g_g, u(1) = g_d$ .

Second order finite difference stencil.

$$(0, 1) = \cup(x_j, x_{j+1}), \quad x_{j+1} - x_j = h = \frac{1}{n+1}, \quad j = 0, \dots, n.$$

$$-\frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1}))}{h^2} \sim f(x_i), \quad i = 1, \dots, n$$

$$u_0 = g_g, \quad u_{n+1} = g_d.$$

$$|u_i - u(x_i)| \leq h^2 \frac{\sup_{x \in [a,b]} |u^{(4)}(x)|}{12}$$

26 / 40

## 1-D Poisson problem

Discrete unknowns  $U = {}^t(u_1, \dots, u_n)$ .

$$A = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \quad b = \begin{pmatrix} f_1 - \frac{g_g}{h^2} \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n - \frac{g_d}{h^2} \end{pmatrix}$$

The matrix  $A$  is symmetric definite positive.

Discrete problem to be solved is

$$AX = b$$

27 / 40

## Condition number and error

$$AX = b, \quad A\hat{X} = \hat{b}$$

Define  $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$ . If  $A$  is symmetric  $> 0$ ,  $\kappa(A) = \frac{\max \lambda_i}{\min \lambda_i}$ .

### Theorem

$$\frac{\|\hat{X} - X\|_2}{\|X\|_2} \leq \kappa(A) \frac{\|\hat{b} - b\|_2}{\|b\|_2}$$

and there is a  $b$  such that it is equal.

Eigenvalues of  $A$  ( $h \times (n+1) = 1$ ).

$$\lambda_k = \frac{2}{h^2} \left(1 - \cos \frac{k\pi}{n+1}\right) = \frac{4}{h^2} \sin^2 \frac{k\pi h}{2}, \quad V_k = \left(\sin \frac{jk\pi}{n+1}\right)_{1 \leq i \leq n},$$

$$\kappa(A) = \frac{\sin^2 \frac{n\pi h}{2}}{\sin^2 \frac{\pi h}{2}} = \frac{\cos^2 \frac{\pi h}{2}}{\sin^2 \frac{\pi h}{2}} \sim \frac{4}{\pi^2 h^2}$$

28 / 40

## Comparison of the iterative methods

Algorithm	spectral radius $\rho(R)$	$n = 5$	$n = 30$
Jacobi	$\cos \pi h$	0.81	0.99
Gauss-Seidel	$(\rho(J))^2 = \cos^2 \pi h$	0.65	0.98
SOR	$\frac{1 - \sin \pi h}{1 + \sin \pi h}$	0.26	0.74
steepest descent	$\frac{K(A) - 1}{K(A) + 1}$	0.81	0.99
conjugate gradient	$\frac{\sqrt{K(A)} - 1}{\sqrt{K(A)} + 1}$	0.51	0.86

Reduction factor for one digit  $M \sim -\frac{1}{\text{Log}_{10} \rho(R)}$ . For  $n = 30$ ,

$n$	Jacobi	Gauss-Seidel	SOR	St Des	CG
10	56	28	4	56	8
100	4759	2380	37	4759	74

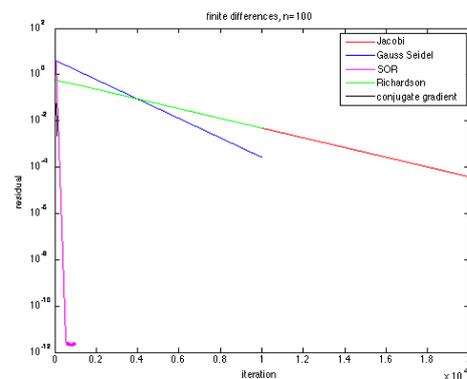
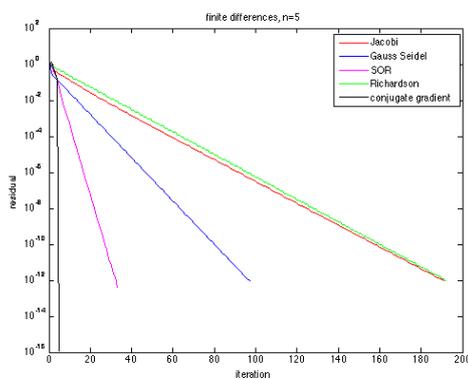
29 / 40

# Asymptotic behavior

Algorithm	spectral radius
Jacobi	$1 - \frac{\pi^2}{2} h^2,$
Gauss-Seidel	$1 - \pi^2 h^2,$
SOR	$1 - 2\pi h$
gradient	$1 - \pi h,$
conjugate gradient	$1 - \frac{\pi h}{2}.$

30 / 40

# Convergence history



31 / 40

## Number of elementary operations

Gauss elimination	$n^2$
optimal overrelaxation	$n^{3/2}$
FFT	$n \ln_2(n)$
conjugate gradient	$n^{5/4}$
multigrid	$n$

Asymptotic order of the number of elementary operations needed to solve the  $1 - D$  problem as a function of the number of grid points

32 / 40

## Purpose

Take the system  $AX = b$ , with  $A$  symmetric definite positive, and the conjugate gradient algorithm. The speed of convergence of the algorithm deteriorates when  $\kappa(A)$ . The purpose is to replace the problem by another system, better conditioned. Let  $M$  be a symmetric regular matrix. Multiply the system on the left by  $M^{-1}$ .

$$AX = b \iff M^{-1}AX = M^{-1}b \iff (M^{-1}AM^{-1})MX = M^{-1}b$$

Define

$$\tilde{A} = M^{-1}AM^{-1}, \quad \tilde{X} = MX, \quad \tilde{b} = M^{-1}b,$$

and the new problem to solve  $\tilde{A}\tilde{X} = \tilde{b}$ . Since  $M$  is symmetric,  $\tilde{A}$  is symmetric definite positive. Write the conjugate gradient algorithm for this "tilde" problem.

34 / 40

## The algorithm for $\tilde{A}$

$$\tilde{X}^0 \text{ given, } \tilde{p}^0 = \tilde{r}^0 = \tilde{b} - \tilde{A}\tilde{X}^0.$$

While  $m < Niter$  or  $\|\tilde{r}^m\| \geq tol$ , do

$$\begin{aligned} \alpha_m &= \frac{\|\tilde{r}^m\|^2}{(\tilde{A}\tilde{p}^m, \tilde{p}^m)}, \\ \tilde{X}^{m+1} &= \tilde{X}^m + \alpha_m \tilde{p}^m, \\ \tilde{r}^{m+1} &= \tilde{r}^m - \alpha_m \tilde{A}\tilde{p}^m, \\ \beta_{m+1} &= \frac{\|\tilde{r}^{m+1}\|^2}{\|\tilde{r}^m\|^2}, \\ \tilde{p}^{m+1} &= \tilde{r}^{m+1} - \beta_{m+1} \tilde{p}^m. \end{aligned}$$

Now define

$$p^m = M^{-1}\tilde{p}^m, \quad X^m = M^{-1}\tilde{X}^m, \quad r^m = M\tilde{r}^m,$$

and replace in the algorithm above.

35 / 40

## The algorithm for $A$

$$Mp^0 = M^{-1}r^0 = M^{-1}b - M^{-1}AM^{-1}MX^0 \iff \begin{cases} p^0 = M^{-2}r^0, \\ r^0 = b - AX^0. \end{cases}$$

$$\|\tilde{r}^m\|^2 = (M^{-1}r^m, M^{-1}r^m) = (M^{-2}r^m, r^m)$$

$$\text{Define } \boxed{z^m = M^{-2}r^m}. \text{ Then } \boxed{\beta_{m+1} = \frac{(z^{m+1}, r^{m+1})}{(z^m, r^m)}}.$$

$$(\tilde{A}\tilde{p}^m, \tilde{p}^m) = (M^{-1}AM^{-1}Mp^m, Mp^m) = (Ap^m, p^m)$$

$$\Rightarrow \boxed{\alpha_m = \frac{(z^m, r^m)}{(Ap^m, p^m)}}.$$

$$MX^{m+1} = MX^m + \alpha_m Mp^m \iff \boxed{X^{m+1} = X^m + \alpha_m p^m}.$$

$$M^{-1}r^{m+1} = M^{-1}r^m - \alpha_m M^{-1}AM^{-1}Mp^m \iff \boxed{r^{m+1} = r^m - \alpha_m Ap^m}.$$

$$Mp^{m+1} = M^{-1}r^{m+1} - \beta_{m+1} Mp^m \iff \boxed{p^{m+1} = z^{m+1} - \beta_{m+1} p^m}. \quad 36 / 40$$

## The algorithm for $A$

Define  $C = M^2$ .

$$X^0 \text{ given, } r^0 = b - AX^0, \text{ solve } Cz^0 = r^0, \quad p^0 = z^0.$$

While  $m < Niter$  or  $\|r^m\| \geq tol$ , do

$$\begin{aligned} \alpha_m &= \frac{(z^m, r^m)}{(Ap^m, p^m)}, \\ X^{m+1} &= X^m + \alpha_m p^m, \\ r^{m+1} &= r^m - \alpha_m Ap^m, \\ \beta_{m+1} &= \frac{(z^{m+1}, r^{m+1})}{(z^m, r^m)}, \\ \text{solve } Cz^{m+1} &= r^{m+1}, \\ p^{m+1} &= z^{m+1} - \beta_{m+1} p^m. \end{aligned}$$

37 / 40

## How to choose $C$

$C$  must be chosen such that

- ①  $\tilde{A}$  is better conditioned than  $A$ ,
- ②  $C$  is easy to invert.

Use an iterative method such that  $A = C - N$  with symmetric  $C$ . For instance it can be a symmetrized version of SOR, named SSOR, defined for  $\omega \in (0, 2)$  by

$$C = \frac{1}{\omega(2-\omega)}(D - \omega E)D^{-1}(D - \omega F).$$

Notice that if  $A$  is symmetric definite positive, so is  $D$  and its coefficients are positive, then its square root  $\sqrt{D}$  is defined naturally as the diagonal matrix of the square roots of the coefficients. Then  $C$  can be rewritten as

$$C = SS^T, \quad \text{with } S = \frac{1}{\sqrt{\omega(2-\omega)}}(D - \omega E)D^{-1/2},$$

yielding a natural Cholewsky decomposition of  $C$ .

38 / 40

## How to choose $C$ , continue

Define furthermore  $M$  as the square root of  $C$  ( $M$  will never be used), i.e. diagonalize  $C$  as  $P\Lambda P^T$ , with  $\Lambda_{ii} > 0$  the eigenvalues of  $C$ , then define the symmetric square root of  $C$  as  $M = P\sqrt{\Lambda}P^T$ . Notice that  $\tilde{\lambda}$  is an eigenvalue of  $\tilde{A}$  associated to the eigenvector  $\tilde{z}$  if and only if

$$M^{-1}AM^{-1}\tilde{z} = \tilde{\lambda}\tilde{z} \iff M^{-2}AM^{-1}\tilde{z} = \tilde{\lambda}M^{-1}\tilde{z}$$

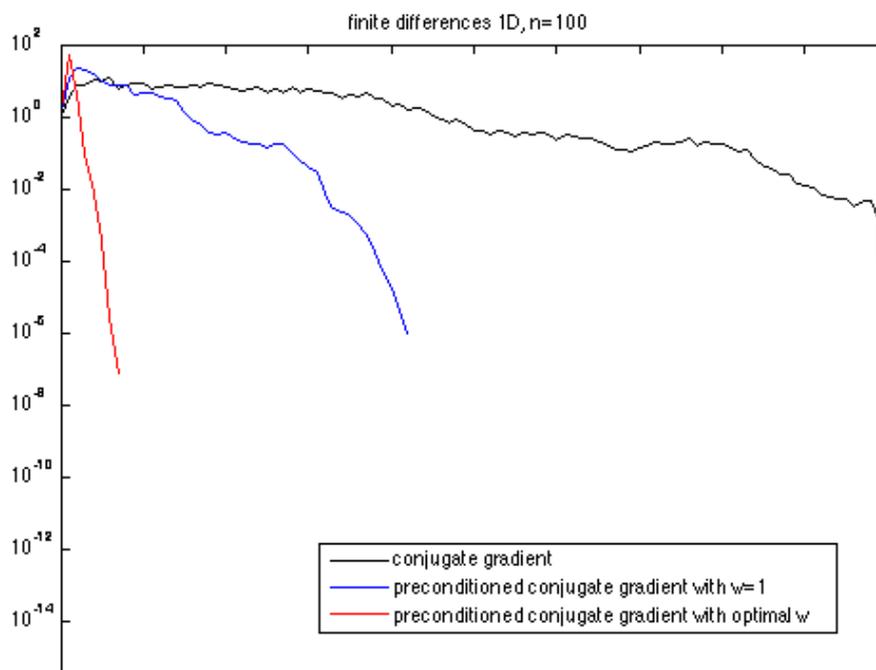
if and only if  $\tilde{\lambda}$  is an eigenvalue of  $C^{-1}A$  associated to the eigenvector  $M^{-1}\tilde{z}$ . The speed of convergence of the iterative method is measured by the spectral radius of  $C^{-1}N$ ,  $\rho(C^{-1}N) < 1$ . Note  $\mu_i$  the eigenvalues of  $C^{-1}N$ . Since  $C^{-1}N = I - C^{-1}A$ , the eigenvalues of  $C^{-1}A$  are equal to  $1 - \mu_i \in [1 - \rho(C^{-1}N), 1 + \rho(C^{-1}N)]$ .

Therefore  $\kappa(\tilde{A}) \leq \frac{1 + \rho(C^{-1}N)}{1 - \rho(C^{-1}N)}$ , and the smallest  $\rho(C^{-1}N)$ , the smallest the condition number of  $\tilde{A}$ .

39 / 40

## Comparison

For the 1-D finite differences matrix and  $n = 100$ , we compare the convergence of the conjugate gradient and the preconditioning by SSOR with  $\omega = 1$  and with the optimal parameter. The gain even with  $\omega = 1$  is striking.



40 / 40