Symmetric iterative solvers for symmetric saddle-point problems

Sue Dollar and Nick Gould

Nonlinear programming problems

Many methods for solution of

$$\min_{x\in\mathbb{R}^n} F(x) \quad \text{subject to} \quad Ax = b, Cx \geq 0$$

involve solving a sequence of equality programming problems of the form

$$\min_{p\in\mathbb{R}^n} \frac{1}{2}p^T H p + g^T p \quad \text{subject to} \quad Ap = -d.$$

Nonlinear programming problems

Many methods for solution of

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{subject to} \quad Ax = b, Cx \geq 0$$

involve solving a sequence of equality programming problems of the form

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H p + g^T p \quad \text{subject to} \quad Ap = -d.$$

Karush-Kuhn-Tucker equations:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner
  - Not inertia revealing

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner
  - Not inertia revealing
- Projected PCG

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner
  - Not inertia revealing
- Projected PCG
  - Short-term recurrence

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner
  - Not inertia revealing
- Projected PCG
  - Short-term recurrence
  - Inertia revealing

Methods for solving KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

- Direct sparse methods
  - Black box
  - Inertia revealing
  - May have problems with very large problems
- MINRES
  - Short-term recurrence
  - Symmetric and positive-definite preconditioner
  - Not inertia revealing
- Projected PCG
  - Short-term recurrence
  - Inertia revealing
  - Uses constraint preconditioner $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$

Inertia revealing property

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H p + g^T p \quad \text{subject to} \quad Ap = -d. \quad \text{(EQP)}$$

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

Inertia revealing property

$$\min_{p\in\mathbb{R}^n} \frac{1}{2}p^T H p + g^T p \quad \text{subject to} \quad Ap = -d. \quad \text{(EQP)}$$

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

Theorem (Gould 1985): Let A have full row rank and Z be such that $AZ = 0$ and $\text{rank}(A^T, Z) = n$. Then

- (EQP) has a strong minimizer iff $Z^T H Z$ is positive definite;
- (EQP) has weak minimizer if $Z^T H Z$ is positive semi-definite with $Z^T H Z$ singular and equations consistent;
- Otherwise, (EQP) has no finite solution.

Inertia revealing property

$$\min_{p\in\mathbb{R}^n} \frac{1}{2}p^T Hp + g^T p \quad \text{subject to} \quad Ap = -d. \quad \text{(EQP)}$$

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

Theorem (Gould 1985): Let A have full row rank and Z be such that $AZ = 0$ and $\text{rank}(A^T, Z) = n$. Then

- (EQP) has a strong minimizer iff $Z^T HZ$ is positive definite;
- (EQP) has weak minimizer if $Z^T HZ$ is positive semi-definite with $Z^T HZ$ singular and equations consistent;
- Otherwise, (EQP) has no finite solution.

PPCG method derived by applying PCG to problem of form $Z^T HZp_z = r_z$ with preconditioner $Z^T GZ$

Inertia revealing property

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H p + g^T p \quad \text{subject to} \quad Ap = -d. \quad \text{(EQP)}$$

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \begin{bmatrix} p \\ -\lambda \end{bmatrix} = \begin{bmatrix} -g \\ -d \end{bmatrix}$$

Theorem (Gould 1985): Let A have full row rank and n_- and n_0 be the number of negative and zero eigenvalues of K. Then

- (EQP) has a strong minimizer iff $n_- = m$ and $n_0 = 0$;
- (EQP) has weak minimizer iff $n_- = m$, $n_0 > 0$ and equation consistent;
- Otherwise, (EQP) has no finite solution.

Requirements

We would like to form an iterative method that

- is a short-term recurrence scheme;
- is inertia revealing;
- performs similarly to MINRES.

Can we build a basis $\mathcal{U}_j$ for the Krylov subspace

$$\mathcal{K}_j\left(K, r_0\right) = \text{span}\left\{r_0, Kr_0, K^2 r_0, \ldots, K^j r_0\right\}$$

such that $U_j^T K U_j$ is block diagonal with 1x1 and 2x2 blocks?

Can we build a basis $\mathcal{U}_j$ for the Krylov subspace

$$\mathcal{K}_j\left(K, r_0\right) = \text{span}\left\{r_0, Kr_0, K^2 r_0, \ldots, K^j r_0\right\}$$

such that $U_j^T K U_j$ is block diagonal with 1x1 and 2x2 blocks?
Original idea: Is it possible to find U_j such that

$$U_j^T K U_j = \begin{bmatrix} K_1 & & & & & \\ & \ddots & & & & \\ & & K_m & & & \\ & & & k_{m+1} & & \\ & & & & \ddots & \\ & & & & & k_{j-m+1} \end{bmatrix}$$

and K_i are 2x2 saddle-point systems with zero (2,2) block?

Can we build a basis $\mathcal{U}_j$ for the Krylov subspace

$$\mathcal{K}_j\left(K, r_0\right) = \text{span}\left\{r_0, K r_0, K^2 r_0, \ldots, K^j r_0\right\}$$

such that $U_j^T K U_j$ is block diagonal with 1x1 and 2x2 blocks?
Original idea: Is it possible to find U_j such that

$$U_j^T K U_j = \begin{bmatrix} K_1 & & & & & \\ & \ddots & & & & \\ & & K_m & & & \\ & & & k_{m+1} & & \\ & & & & \ddots & \\ & & & & & k_{j-m+1} \end{bmatrix}$$

and K_i are 2x2 saddle-point systems with zero (2,2) block?
No, but can detect at each stage whether we should form a 2x2 or 1x1 block.

Can we build a basis $\mathcal{U}_j$ for the Krylov subspace

$$\mathcal{K}_j\left(K, r_0\right) = \textsf{span}\left\{r_0, K r_0, K^2 r_0, \ldots, K^j r_0\right\}$$

such that $U_j^T K U_j$ is block diagonal with 1x1 and 2x2 blocks?
Original idea: Is it possible to find U_j such that

$$U_j^T K U_j = \begin{bmatrix} K_1 & & & & & \\ & \ddots & & & & \\ & & K_m & & & \\ & & & k_{m+1} & & \\ & & & & \ddots & \\ & & & & & k_{j-m+1} \end{bmatrix}$$

and K_i are 2x2 saddle-point systems with zero (2,2) block?
No, but can detect at each stage whether we should form a 2x2 or 1x1 block.
SYMMBK

Lanczos method

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} p \\ -\lambda \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} -g \\ -d \end{bmatrix}}_{b}$$

Lanczos method

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} p \\ -\lambda \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} -g \\ -d \end{bmatrix}}_{b}$$

Forms basis $\mathcal{Q}_j$ of $\mathcal{K}_j(K, r_0)$ such that

$$KQ_j - Q_jT_j = \gamma_{j+1}q_{j+1}e_{j+1}^T,$$

where

$$T_j = \begin{bmatrix} \delta_0 & \gamma_1 & & & \\ \gamma_1 & \delta_1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \delta_{j-1} & \gamma_j \\ & & & \gamma_j & \delta_j \end{bmatrix} = Q_j^T K Q_j.$$

Lanczos method

$$\begin{matrix} n \\ m \end{matrix} \underbrace{\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} p \\ -\lambda \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} -g \\ -d \end{bmatrix}}_{b}$$

Forms basis $\mathcal{Q}_j$ of $\mathcal{K}_j(K, r_0)$ such that

$$KQ_j - Q_jT_j = \gamma_{j+1} q_{j+1} e_{j+1}^T,$$

where

$$T_j = \begin{bmatrix} \delta_0 & \gamma_1 & & & \\ \gamma_1 & \delta_1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \delta_{j-1} & \gamma_j \\ & & & \gamma_j & \delta_j \end{bmatrix} = Q_j^T K Q_j.$$

At each iteration, solve $T_j v_j = Q_j^T b$ and set $y_j = Q_j v_j$.

SYMMBK (Chandra 1978)

Using Bunch-Parlett (1971), factor $T_j = L_j D_j L_j^T$, where D_j block diagonal with 1x1 and 2x2 blocks.

$$D_j = L_j^{-1} Q_j^T K Q_j L_j^{-T} = S_j^T K S_j$$

Vectors in S_j defined by short-term recurrence formula.

SYMMBK (Chandra 1978)

Using Bunch-Parlett (1971), factor $T_j = L_j D_j L_j^T$, where D_j block diagonal with 1x1 and 2x2 blocks.

$$D_j = L_j^{-1} Q_j^T K Q_j L_j^{-T} = S_j^T K S_j$$

Vectors in S_j defined by short-term recurrence formula.

Stability (Higham 1999): look ahead one Lanczos iteration before making decision whether new entry is in 1x1 or 2x2 pivot. No permutation required.

$$\begin{aligned} D_j &= \begin{bmatrix} D_{j-1} & \\ & d_j \end{bmatrix}, \\ D_j v_j &= S_j^T b, \quad y_j = S_j v_j, \\ y_j &= y_{j-1} + s_j d_j^{-1} s_j^T b. \end{aligned}$$

SYMMBK (Chandra 1978)

Using Bunch-Parlett (1971), factor $T_j = L_j D_j L_j^T$, where D_j block diagonal with 1x1 and 2x2 blocks.

$$D_j = L_j^{-1} Q_j^T K Q_j L_j^{-T} = S_j^T K S_j$$

Vectors in S_j defined by short-term recurrence formula.

Stability (Higham 1999): look ahead one Lanczos iteration before making decision whether new entry is in 1x1 or 2x2 pivot. No permutation required.

$$\begin{aligned} D_j &= \begin{bmatrix} D_{j-1} & \\ & d_j \end{bmatrix}, \\ D_j v_j &= S_j^T b, \quad y_j = S_j v_j, \\ y_j &= y_{j-1} + s_j d_j^{-1} s_j^T b. \end{aligned}$$

Note: Marcia (2007) uses Bunch-Marcia factorization - look ahead two Lanczos iterations. Does not need estimate of $\|K\|$.

SYMMLQ (Paige & Saunders 1975) uses $T_j = L_j W_j$. SYMMBK generally has favourable operation counts and but requires one extra vector to be stored.

For SPD problems, SYMMBK reduces to the CG method.

MINRES: $\min_{x_j \in \mathcal{K}_j} \|K x_j - b\|$ SYMBBK: $\|K x_j - b\| \le \|L\| \left\| \widehat{S}_j b \right\|, \ S = \left[S_j, \widehat{S}_j \right]$

SYMMBK vs MINRES

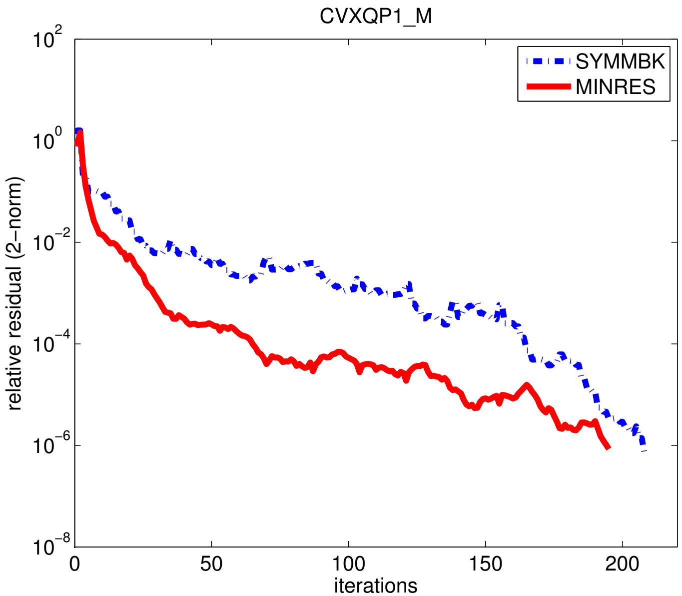
Matlab 2007a

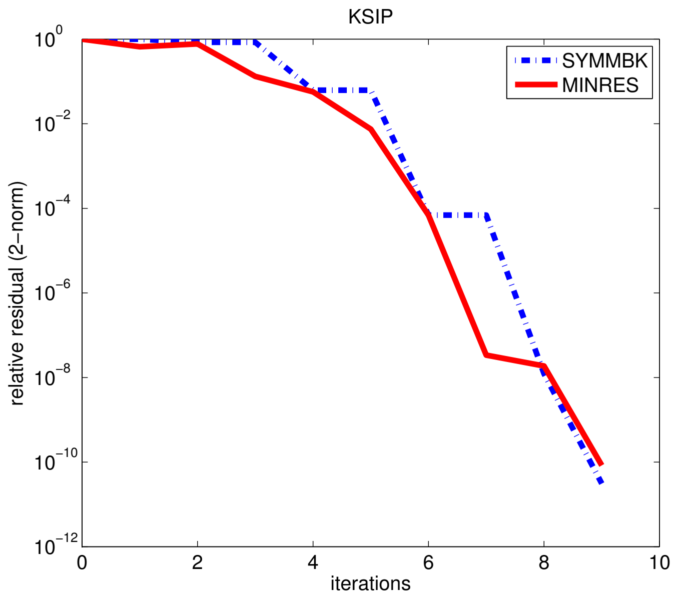
$$
\begin{aligned}
P &= \begin{bmatrix} H + A^T \overline{W} A & 0 \\ 0 & W \end{bmatrix}, \\
\gamma &= \text{normest}(A)^2/\text{normest}(H), \\
W &= \gamma I, \\
\overline{W} &= \text{diag}(w_1, w_2, \ldots, w_m) \\
w_i &= \begin{cases} 0 & \text{if row } i \text{ in } A \text{ is dense;} \\ \frac{1}{\gamma} & \text{otherwise.} \end{cases}
\end{aligned}
$$

(Rees & Greif, SISC 2007)

$$
\begin{aligned}
\lambda &= 1, \\
\lambda &= -1, \\
\lambda &\in (-1, 0).
\end{aligned}
$$

SYMMBK vs MINRES





CVXQP1_M ($n = 1000$, $m = 500$)

KSIP ($n = 1021$, $m = 1001$)

SYMMBK vs MINRES (cont.)

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Hx + g^T x \quad \text{subject to} \quad Ax = b, x \geq 0.$$

Predictor-corrector interior-point method (solve two KKT systems with same coefficient matrix each iteration)

KSIP ($n = 1021, m = 1001$)

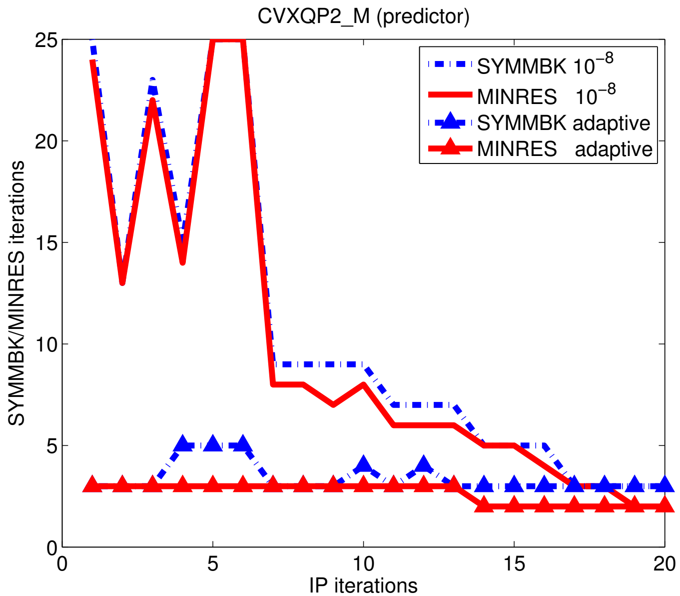
After 3 interior-point iterations (SYMMBK tolerance 10^{-2})

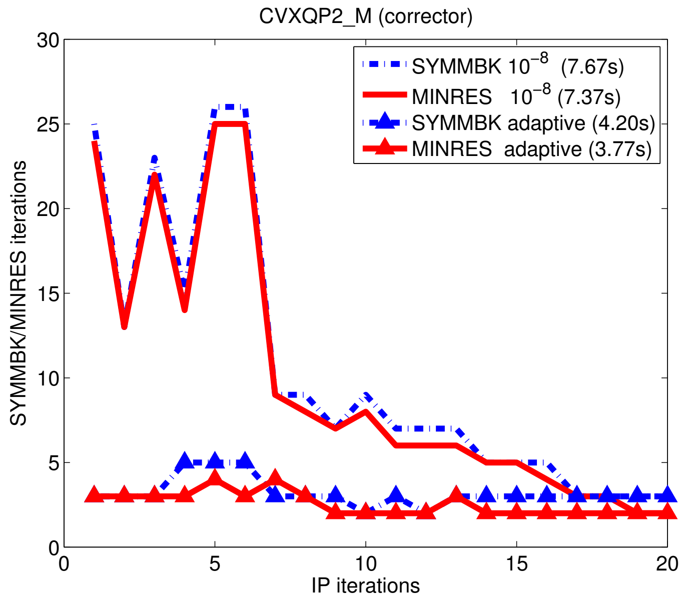
```
Warning:  too many negative eigenvalues found
> In symmbk2 at 201
In QP_MPC2 at 231
```

SYMMBK vs MINRES (cont.)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T H x + g^T x \quad \text{subject to} \quad Ax = b, x \geq 0.$$

Predictor-corrector interior-point method (solve two KKT systems with same coefficient matrix each iteration)



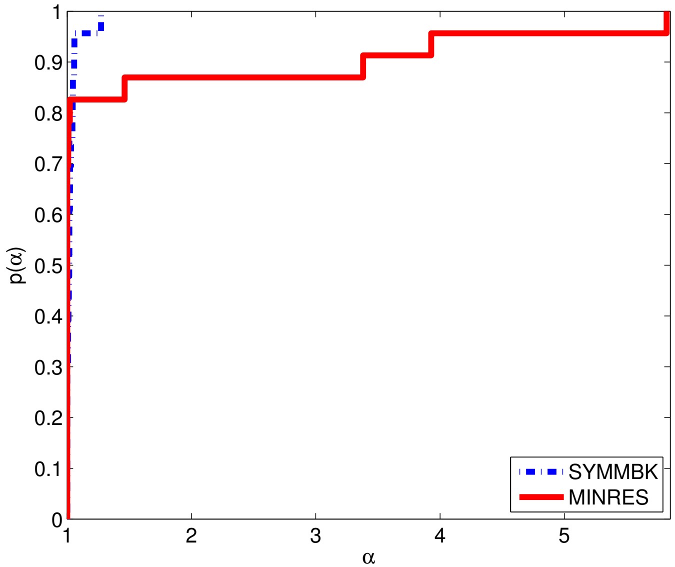


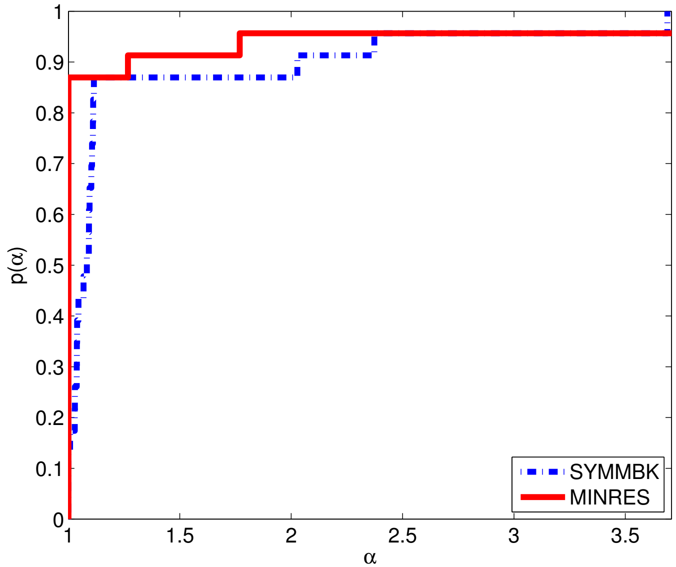
Adaptive tolerance $\min\{10^{-2}, \max\{0.01\mu, 10^{-10}\}\}$

SYMMBK vs MINRES (cont.)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T H x + g^T x \quad \text{subject to} \quad Ax = b, x \geq 0.$$

Predictor-corrector interior-point method (solve two KKT systems with same coefficient matrix each iteration)





Adaptive tolerance $\min\{10^{-3}, \max\{0.01\mu, 10^{-10}\}\}$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

PDE-constrained problem

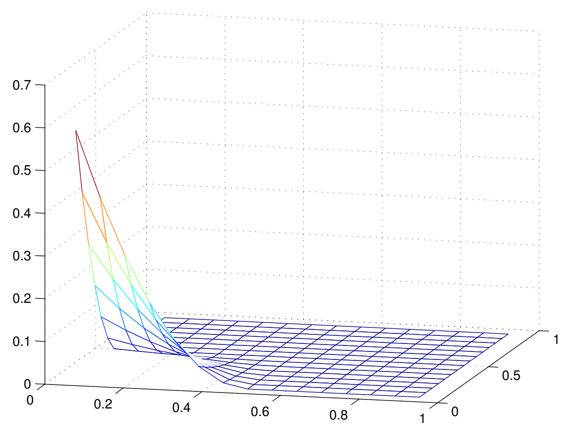
$$\min_{\mathtt{u},\mathtt{f}} \frac{1}{2} \|\mathtt{u} - \widehat{\mathtt{u}}\|_2^2 + \beta \|\mathtt{f}\|_2^2$$

subject to

$$\begin{aligned} -\nabla^2 \mathtt{u} &= \mathtt{f} \text{ in } \Omega = [0,1]^2 \\ \mathtt{u} &= \widehat{\mathtt{u}} \text{ on } \delta\Omega, \end{aligned}$$

where

$$\begin{aligned} \widehat{\mathtt{u}} &= \begin{cases} 16(x-\frac{1}{2})^2(y-\frac{1}{2})^2 & \text{if } (x,y) \in [0,\frac{1}{2}]^2 \\ 0 & \text{otherwise.} \end{cases} \\ \beta &= 0.01 \end{aligned}$$



PDE-constrained problem

Using bilinear **Q1** elements:

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & M & K^T \\ -M & K & 0 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} 2\beta M & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & KM^{-1}K \end{bmatrix}$$

PDE-constrained problem

Using bilinear **Q1** elements:

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & M & K^T \\ -M & K & 0 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} 2\beta M & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & KM^{-1}K \end{bmatrix}$$

(Rees, Dollar & Wathen, 2008 Tech Report)

$$\begin{aligned} \lambda &= 1, \\ \frac{1}{2}\left(1+\sqrt{5+\frac{2\alpha_1 h^4}{\beta}}\right) \leq \quad \lambda &\leq \frac{1}{2}\left(1+\sqrt{5+\frac{2\alpha_2}{\beta}}\right), \\ \frac{1}{2}\left(1-\sqrt{5+\frac{2\alpha_2}{\beta}}\right) \leq \quad \lambda &\leq \frac{1}{2}\left(1-\sqrt{5+\frac{2\alpha_1 h^4}{\beta}}\right). \end{aligned}$$

PDE-constrained problem

Using bilinear **Q1** elements:

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & M & K^T \\ -M & K & 0 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} 2\beta \widetilde{M} & 0 & 0 \\ 0 & \widetilde{M} & 0 \\ 0 & 0 & \widetilde{K} M^{-1} \widetilde{K} \end{bmatrix}$$

h	n	SYMMBK(10^{-6})	SYMMBK(10^{-12})	MINRES(10^{-6})	MINRES(10^{-12})
2^{-2}	27	0.02 (7)	0.04 (12)	0.02 (7)	0.04 (12)
2^{-3}	147	0.03 (7)	0.05 (14)	0.03 (7)	0.05 (14)
2^{-4}	675	0.06 (9)	0.08 (14)	0.06 (9)	0.09 (14)
2^{-5}	2883	0.12 (7)	0.22 (14)	0.12 (7)	0.23 (14)
2^{-6}	11907	0.66 (9)	0.99 (14)	0.67 (9)	1.05 (14)
2^{-7}	48487	2.97 (9)	4.96 (16)	3.04 (9)	5.05 (16)
2^{-8}	195075	14.1 (9)	26.4 (18)	15.6 (9)	25.3 (17)
2^{-9}	783363	71.8 (11)	119 (20)	71.1 (11)	122 (20)

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Conclusions

- Would like to track inertia of saddle-point problem

Conclusions

- Would like to track inertia of saddle-point problem
- SYMMBK allows us to do this

Conclusions

- Would like to track inertia of saddle-point problem
- SYMMBK allows us to do this
- In our examples, SYMMBK performs similarly to MINRES

Conclusions

- Would like to track inertia of saddle-point problem
- SYMMBK allows us to do this
- In our examples, SYMMBK performs similarly to MINRES
- GALAHAD