On preconditioning PDE-constrained optimization problems with control constraints

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria.simoncini@unibo.it

On-going joint work with Margherita Porcelli and Mattia Tani

The problem - Our starting point: Herzog & Sachs, SIMAX 2010

Minimize $\frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}$

subject to

$$\begin{cases} -\Delta y - \boldsymbol{\beta} \cdot \nabla y = u & \text{in} \quad \Omega \\ \partial_{\vec{n}} y = 0 & \text{on} \quad \partial\Omega \end{cases}$$

$$u_a \leq u \leq u_b \quad \text{a.e. in} \quad \Omega$$

The problem - Our starting point: Herzog & Sachs, SIMAX 2010

Minimize $\frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}$

subject to

$$\begin{cases} -\Delta y - \boldsymbol{\beta} \cdot \nabla y = u & \text{in} \quad \Omega \\ \partial_{\vec{n}} y = 0 & \text{on} \quad \partial\Omega \end{cases} \tag{1}$$

$$u_a \leq u \leq u_b \quad \text{a.e. in} \quad \Omega$$

Key quantities:

- ν : parameter of control term in objective function
- (1): PDE $\rightarrow$ mesh parameter h
- (1): Convection term $\boldsymbol{\beta}$
- $u_a \leq u \leq u_b$: box constraints on control

Solution strategy

Using

$$\xi - \max\{0, \xi + (u - u_b)\} - \min\{0, \xi + (u_a - u)\} = 0$$

$\Rightarrow$ Semi-smooth Newton iteration with active set strategy

$$F(y, u, p, \mu) = 0, \qquad F'(y, u, p, \mu) = \begin{pmatrix} M & \cdot & L^T & \cdot \\ \cdot & \nu M & -M & I \\ L & -M & \cdot & \cdot \\ \cdot & \chi_{\mathcal{A}} & \cdot & \chi_{\mathcal{I}} \end{pmatrix}$$

(in the proper spaces)

$\chi_{\mathcal{A}}$ diagonal, $(\chi_{\mathcal{A}})_{ii} = 1$ if $i \in \mathcal{A}$ (set of active indeces for the control)

Note: L convection-diffusion operator, M inner product operator

The discrete problem

Newton iteration for the discrete problem leads to:

$$\begin{bmatrix} M & \cdot & L^T & \cdot \\ \cdot & \nu M & -M & P_{\mathcal{A}^k}^T \\ L & -M & \cdot & \cdot \\ \cdot & P_{\mathcal{A}^k} & \cdot & \cdot \end{bmatrix} \begin{bmatrix} y^{k+1} \\ u^{k+1} \\ p^{k+1} \\ \mu_{\mathcal{A}^k}^{k+1} \end{bmatrix} = \begin{bmatrix} My_d \\ 0 \\ 0 \\ P_{\mathcal{A}_b^k} u_b + P_{\mathcal{A}_a^k} u_a \end{bmatrix} \quad \Leftrightarrow \quad J_k x^{k+1} = b^k$$

$k = 1, 2, \ldots,$ where

$$\begin{aligned} \mathcal{A}^k &= \mathcal{A}_b^k \cup \mathcal{A}_a^k \\ \mathcal{A}_b^k &= \{i \mid \mu_i^k + (u_i^k - (u_b)_i) > 0\}, \ \mathcal{A}_a^k = \{i \mid \mu_i^k + ((u_a)_i - u_i^k) < 0\} \end{aligned}$$

and $P_{\mathcal{C}}$ rectangular matrix of rows of $\chi_{\mathcal{C}}$ belonging to the indices in $\mathcal{C}$

Note: dependence on mesh parameter omitted

State-of-the-art approaches. Herzog and Sachs, 2010

$$J_k = \left[\begin{array}{ccc|c} M & \cdot & L^T & \cdot \\ \cdot & \nu M & -M & P_{\mathcal{A}^k}^T \\ L & -M & \cdot & \cdot \\ \hline \cdot & P_{\mathcal{A}^k} & \cdot & \cdot \end{array}\right] = \begin{bmatrix} A & B_k^T \\ B_k & 0 \end{bmatrix}$$

Preconditioner:

$$\mathcal{K} = \begin{bmatrix} I & 0 \\ B_k \widehat{A}^{-1} & I \end{bmatrix} \begin{bmatrix} \widehat{A} & B_k^T \\ 0 & -\widehat{S} \end{bmatrix}$$

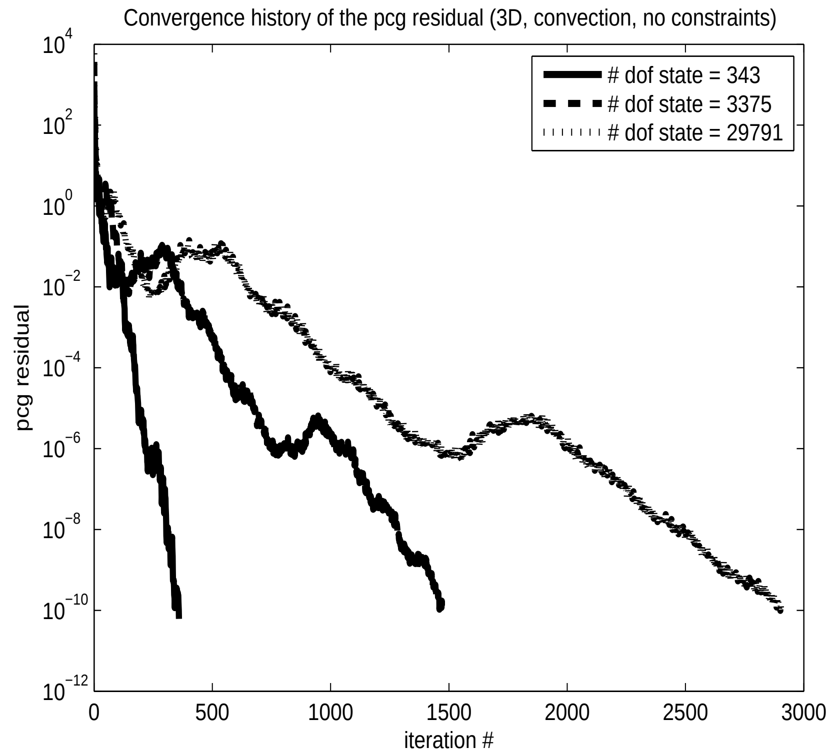
where $\widehat{A}(\sigma) \approx A$ and $\widehat{S}(\sigma, \tau) \approx S$ and are block diagonal

(from Schöberl and Zulehner 2007)

$\Rightarrow$ PCG in non-standard inner product

A 3D numerical experiment from Herzog and Sachs, 2010

Here $\boldsymbol{\beta} = (1000, 0, 0)$, $\nu = 10^{-2}$ First step of Newton ($\mathcal{A} = \emptyset$)



Problems reported in estimating preconditioner parameters σ, τ

Mesh dependence and (not shown) dependence on ν

Thanks to Roland Herzog for the codes!

State-of-the-art approaches. Stoll and Wathen, 2012

Reduction approach:

$$\begin{bmatrix} M & \cdot & -K^T \\ \cdot & \nu M^{\mathcal{A}_k,\mathcal{A}_k} & M^{\mathcal{A}_k,:} \\ -K & M^{:,\mathcal{A}_k} & \cdot \end{bmatrix}$$

(here K symmetric stiffness matrix, M diagonal)

Bramble-Pasciak preconditioner $\mathcal{K}$ in the inner product $\mathcal{H}$:

$$\mathcal{K} = \begin{bmatrix} A_0 & \cdot & \cdot \\ \cdot & A_1 & \cdot \\ -K & M & -S_0 \end{bmatrix} \qquad \mathcal{H} = \begin{bmatrix} M - A_0 & \cdot & \cdot \\ \cdot & \nu M - A_1 & \cdot \\ \cdot & \cdot & S_0 \end{bmatrix}$$

with $S_0 = \widehat{K} M^{-1} \widehat{K}^T \approx K M^{-1} K^T + \nu^{-1} M$ true Schur complement

(preconditioner projected in the reduced space)

A 3D numerical experiment (Stoll and Wathen, 2010)

(n, ν)	Newton it	$\lvert\mathcal{A}_a\rvert + \lvert\mathcal{A}_b\rvert$	PCG # its	PCG CPU
$(3375, 10^{-2})$	0	0	14	1.22
	1	741+2114	15	1.38
	3	188+2292	14	1.22
$(29791, 10^{-2})$	0	0	16	12.55
	1	6257+ 17298	20	4.79
	3	3729+ 19166	15	3.49
$(3375, 10^{-4})$	...	...	...	...
	8	947 + 2340	20	1.76
	9	987 + 2380	17	1.48
	10	987 + 2364	15	1.33
$(29791, 10^{-4})$	...	...	...	...
	13	10089+ 19670	22	5.30
	15	10121+ 19622	22	5.34
	16	10121+ 19590	16	3.92
$(3375, 10^{-6})$	...	...	...	...
	13	849 + 2300	136	12.17
	15	937 + 2340	99	8.92
$(29791, 10^{-6})$	...	...	...	...
	15	6544 + 21310	182	40.62
	17	6800 + 22068	135	30.46
	18	7353 + 22398	49	11.02

State-of-the-art approaches. Pearson and Wathen, 2012

Same problem as in Stoll-Wathen, but without bound constraints

$$\begin{bmatrix} M & \cdot & -K^T \\ \cdot & \nu M & M \\ -K & M & \cdot \end{bmatrix}$$

(K symmetric stiffness matrix, M diagonal)

Bramble-Pasciak preconditioner $\mathcal{K}$ in the inner product $\mathcal{H}$:

$$\mathcal{K} = \begin{bmatrix} A_0 & \cdot & \cdot \\ \cdot & A_1 & \cdot \\ -K & M & -S_0 \end{bmatrix} \qquad \mathcal{H} = \begin{bmatrix} M - A_0 & \cdot & \cdot \\ \cdot & \nu M - A_1 & \cdot \\ \cdot & \cdot & S_0 \end{bmatrix}$$

with $\widehat{S}_0 = (K + \frac{1}{\sqrt{\nu}} M) M^{-1} (K + \frac{1}{\sqrt{\nu}} M) \approx K M^{-1} K^T + \nu^{-1} M = S$

eig$(\widehat{S}_0^{-1} S) \subseteq [\frac{1}{2}, 1]$ $\quad \Rightarrow$ independence of ν

$(K + \frac{1}{\sqrt{\nu}} M) \approx (\widehat{K + \frac{1}{\sqrt{\nu}} M}) \Rightarrow$ independence of h

A different approach for the general setting

$$J_k = \left[\begin{array}{cc|cc} M & \cdot & L^T & \cdot \\ \cdot & \nu M & -M & P_{\mathcal{A}^k}^T \\ \hline L & -M & \cdot & \cdot \\ \cdot & P_{\mathcal{A}^k} & \cdot & \cdot \end{array}\right] = \begin{bmatrix} A & B_k^T \\ B_k & \cdot \end{bmatrix}$$

Consider the following usual "constraint" preconditioner:

$$\mathcal{K}_k = \begin{pmatrix} I & \cdot \\ B_k A^{-1} & I \end{pmatrix} \begin{pmatrix} A & \cdot \\ \cdot & -\widehat{S}_k \end{pmatrix} \begin{pmatrix} I & A^{-1}B_k^T \\ \cdot & I \end{pmatrix}, \quad \widehat{S}_k \approx S_k = B_k A^{-1} B_k^T$$

where the Schur complement S_k is

$$S_k = \frac{1}{\nu} \begin{pmatrix} \mathbf{S}_0 & -P_{\mathcal{A}^k}^T \\ -P_{\mathcal{A}^k} & S_{2,k} \end{pmatrix}, \qquad \begin{array}{l} \mathbf{S}_0 = \nu L M^{-1} L^T + M \\ S_{2,k} = P_{\mathcal{A}^k} M^{-1} P_{\mathcal{A}^k}^T \quad \text{diag.} \end{array}$$

Constraint preconditioner - Factorized Schur complement matrix

Schur Complement matrix

$$S_k = \frac{1}{\nu} \begin{bmatrix} I & -\Pi_k M P_{\mathcal{A}^k}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbb{S}_k & 0 \\ 0 & S_{2,k} \end{bmatrix} \begin{bmatrix} I & 0 \\ -P_{\mathcal{A}^k} M \Pi_k & I \end{bmatrix}$$

where $\Pi_k = P_{\mathcal{A}^k}^T P_{\mathcal{A}^k}$ is a diagonal projection matrix, and

$$\begin{aligned} \mathbb{S}_k &= \nu L^T M^{-1} L + M - \Pi_k M \Pi_k \\ &\approx \left(\sqrt{\nu} L + (I - \Pi_k) M\right)^T M^{-1} \left(\sqrt{\nu} L + (I - \Pi_k) M\right) \\ &= \mathbb{S}_k + \sqrt{\nu}\left(L(I - \Pi_k) + (I - \Pi_k) L^T\right) =: \widehat{\mathbb{S}}_k \end{aligned}$$

$\Rightarrow$ eig$(\widehat{\mathbb{S}}_k{}^{-1} \mathbb{S}_k) \subseteq [\frac{1}{2}, 1]$ (for L nonsym coercive, $\mathcal{A} = \emptyset$)

(à la Pearson-Wathen)

Overall constraint preconditioner

$$\mathcal{K}_k = \begin{pmatrix} I & \cdot \\ B_k A^{-1} & I \end{pmatrix} \begin{pmatrix} A & \cdot \\ \cdot & -\widehat{S}_k \end{pmatrix} \begin{pmatrix} I & A^{-1} B_k^T \\ \cdot & I \end{pmatrix},$$

with

$$\widehat{S}_k = \frac{1}{\nu} \begin{bmatrix} I & -\Pi_k M P_{\mathcal{A}^k}^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \widehat{\mathbb{S}}_k & 0 \\ 0 & P_{\mathcal{A}^k} M^{-1} P_{\mathcal{A}^k}^T \end{bmatrix} \begin{bmatrix} I & 0 \\ -P_{\mathcal{A}^k} M \Pi_k & I \end{bmatrix}$$

and

$$\widehat{\mathbb{S}}_k = \left(\sqrt{\nu} L + (I - \Pi_k) M\right)^T M^{-1} \left(\sqrt{\nu} L + (I - \Pi_k) M\right)$$

A 3D numerical experiment

From Herzog and Sachs, 2010:

$$\text{Minimize } \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{L^2(\Omega)}$$

subject to

$$\begin{cases} -\Delta y - \boldsymbol{\beta} \cdot \nabla y = u & \text{in} \quad \Omega \\ \partial_{\vec{n}} y = 0 & \text{on} \quad \partial\Omega \end{cases}$$

$$0 \le u \le 2.5 \quad \text{a.e. in} \quad \Omega$$

with $\Omega = [-1, 1]^3 \subset \mathbb{R}^3$

$\boldsymbol{\beta} = (\beta, 0, 0)$, $\beta \in \{10, 100, 1000\}$

$\nu \in \{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}\}$

and $\quad y_d = 1$ for $|x_1| \le 0.5$ and $y_d = -2$ elsewhere

A 3D numerical experiment. Final Newton iteration.

$\beta = (10, 0, 0)$

(n, ν)	Newton it	$\vert\mathcal{A}_a\vert + \vert\mathcal{A}_b\vert$	PGMRES avg # its	PGMRES tot CPU	nonlin res
$(3375, 10^{-2})$	3	2687	10	1.52	9.20e-19
$(29791, 10^{-2})$	3	22459	10	13.79	3.17e-19
$(3375, 10^{-4})$	12	600+ 2763	17	6.99	3.97e-17
$(29791, 10^{-4})$	12	6726+ 22989	16	63.19	4.71e-19
$(3375, 10^{-6})$	17	604+ 2763	40	21.33	2.61e-14
$(29791, 10^{-6})$	36	6838+ 22953	41	456.78	5.00e-09
$(3375, 10^{-8})$	18	604+ 2763	68	41.42	4.75e-14
$(29791, 10^{-8})$	40	6802+ 22989	104	1284.44	3.67e-10

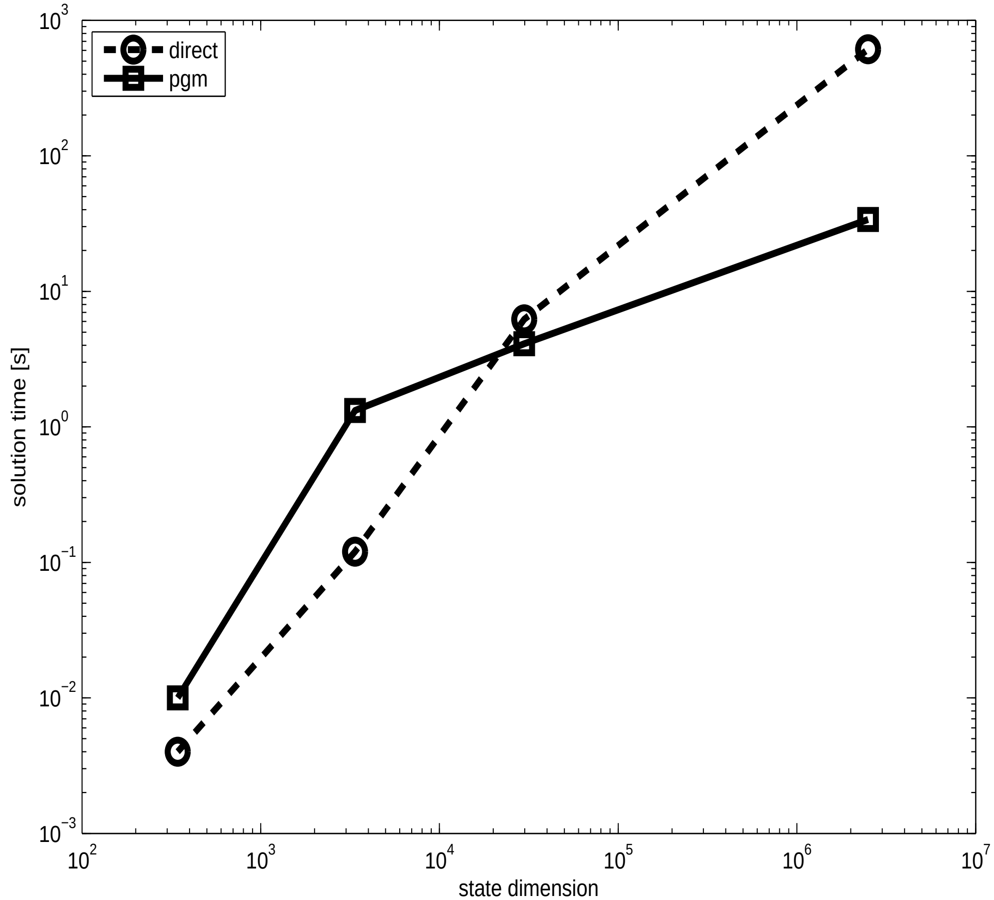
$\beta = (1000, 0, 0)$

(n, ν)	Newton it	$\lvert\mathcal{A}_a\rvert + \lvert\mathcal{A}_b\rvert$	PGMRES avg # its	PGMRES tot CPU	nonlin res
$(3375, 10^{-2})$	1	3375	4	0.46	2.59e-17
$(29791, 10^{-2})$	1	29543	4	-	3.31e-17
$(3375, 10^{-4})$	1	3375	8	0.63	3.38e-15
$(29791, 10^{-4})$	1	29543	9	-	4.06e-15
$(3375, 10^{-6})$	3	4+3371	13	1.23	7.89e-09
$(29791, 10^{-6})$	3	1093+28698	13	-	6.85e-09
$(3375, 10^{-8})$	7	3375	27	6.26	1.14e-12
$(29791, 10^{-8})$	7	1093+28698	29	-	6.58e-09

- : not reported (apparent problem with multigrid solver)

Sort of "sanity check"

Iterative vs direct (sparse) solution



here $\beta = 10$, $\nu = 10^{-2}$, no control box constraints

Ongoing analysis

⋆ Claim: for $\mathcal{A} \neq \emptyset$,

$$\mathrm{eig}(\widehat{\mathbb{S}}_k{}^{-1}\mathbb{S}_k) \subseteq [\tfrac{1}{2}, \gamma] \quad \text{with} \quad \gamma = \mathcal{O}(1)$$

independent of h, β (numerical evidence), with possible mild dependence on ν

(this will lead to a similar result for the whole preconditioned problem)

Ongoing analysis

- Claim: for $\mathcal{A} \neq \emptyset$,

$$\text{eig}(\widehat{\mathbb{S}}_k{}^{-1}\mathbb{S}_k) \subseteq [\tfrac{1}{2}, \gamma] \quad \text{with} \quad \gamma = \mathcal{O}(1)$$

independent of h, β (numerical evidence), with possible mild dependence on ν

(this will lead to a similar result for the whole preconditioned problem)

- Mixed (state and control) box constraints:

$$y_a \leq \varepsilon u + y \leq y_b$$

(see, e.g., Pearson, Stoll, Wathen 2014)

Ongoing analysis

⋆ Claim: for $\mathcal{A} \neq \emptyset$,

$$\text{eig}(\widehat{\mathbb{S}}_k{}^{-1}\mathbb{S}_k) \subseteq [\tfrac{1}{2}, \gamma] \quad \text{with} \quad \gamma = \mathcal{O}(1)$$

independent of h, β (numerical evidence), with possible mild dependence on ν

(this will lead to a similar result for the whole preconditioned problem)

⋆ Mixed (state and control) box constraints:

$$y_a \leq \varepsilon u + y \leq y_b$$

(see, e.g., Pearson, Stoll, Wathen 2014)

⋆ Sparsity constraints, e.g., $\|u\|_{L^1(\Omega)}$ instead of (or in addition to) $\|u\|_{L^2(\Omega)}$