Preconditioning Newton methods for Optimal Control Problems with Sparsity Constraints

Valeria Simoncini

Dipartimento di Matematica
Alma Mater Studiorum - Università di Bologna

Joint work with
Margherita Porcelli - Università di Firenze
Martin Stoll - Max Planck Inst. for Dynamics of Complex Techn. Systems, Magdeburg

The Constrained Optimal Control Problem

$$\min \quad J(y,u) = \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\|u\|_{L^1(\Omega)} \leftarrow \text{ sparsity constr.}$$

$$\text{over } (y,u) \in H^1_0(\Omega) \times L^2(\Omega)$$

$$\text{s.t.} \quad \mathcal{L}y = u \quad \text{in } \Omega \quad \leftarrow \text{ state equation}$$

$$\text{and} \quad a \le u \le b \text{ a.e. in } \Omega \quad \leftarrow \text{ box constraints}$$

- u and y are the control and state variables
- $y_d \in L^2(\Omega)$ is the desired state, $\Omega \subseteq \mathbb{R}^d$ with $d = 2, 3$
- $\mathcal{L}$ is a second-order linear elliptic differential operator
- Control box constraints: $a, b \in L^2(\Omega)$ and $a < 0 < b$
- Parameters: L^2-norm term $\alpha > 0$ and L^1-norm term $\beta > 0$.

Sparsity constraints in optimal control problems

Motivation

- Optimal control applications: provide information about the optimal location of control device and actuators [Stadler, COAP 2009] [Costa et al. Comput. Struct. 2007].

Main references:

- L^1-norm: Casas, Clacson, Kunish, Herzog, Stadler, Wachsmuth, 2009-2012
- Directional Sparsity $\|\cdot\|_{1,2}$ Herzog, Stadler and Wachsmuth SICON 2012.

based on semismooth Newton's approach [Hintermüller, Ito, Kunish SIOPT 2002].

None of these works takes into account discretization/implementation issues for the linear algebra phase (e.g. preconditioning).

Sparsity constraints in optimal control problems

Motivation

- Optimal control applications: provide information about the optimal location of control device and actuators [Stadler, COAP 2009] [Costa et al. Comput. Struct. 2007].

Main references:

- L^1-norm: Casas, Clacson, Kunish, Herzog, Stadler, Wachsmuth, 2009-2012
- Directional Sparsity $\|\cdot\|_{1,2}$ Herzog, Stadler and Wachsmuth SICON 2012.

based on semismooth Newton's approach [Hintermüller, Ito, Kunish SIOPT 2002].

None of these works takes into account discretization/implementation issues for the linear algebra phase (e.g. preconditioning).

Active-set interpretation

- Complementarity conditions for box constraints:

$$\Pi_{\mathcal{A}_a}(x-a) - c\Pi_{\mathcal{I}}\mu = 0$$

with

$$\mathcal{A}_a = \{i \mid (x_i - a_i) + c\mu_i < 0\} \qquad \text{and} \qquad \mathcal{I} = \{1,\ldots,n\} \setminus \mathcal{A}_a$$

$$x_i = a \text{ for } i \in \mathcal{A}_a \;\text{ and }\; \mu_i = 0 \text{ for } i \in \mathcal{I}$$

- Complementarity conditions for L^1-norm sparsity constraints:

$$\Pi_{\mathcal{A}_0}x - c(\Pi_{\mathcal{I}_+}(\mu-\beta) + \Pi_{\mathcal{I}_-}(\mu+\beta)) = 0.$$

with

$$\mathcal{A}_0 = \{i \mid x_i + c(\mu_i+\beta) \geq 0\} \cup \{i \mid x_i + c(\mu_i-\beta) \leq 0\}$$

$$\mathcal{I}_+ = \{i \mid x + c(\mu_i-\beta) > 0\} \qquad \text{and} \qquad \mathcal{I}_- = \{i \mid x_i + c(\mu_i+\beta) < 0\}$$

$$x_i = 0 \text{ for } i \in \mathcal{A}_0 \;\text{ and }\; \mu_i = -\beta \text{ for } i \in \mathcal{I}_- \text{ and } \mu_i = \beta \text{ for } i \in \mathcal{I}_+.$$

- $\Pi_{\mathcal{C}}$ is an $n \times n$ diagonal 0-1 matrix with 1s corresponding to $\mathcal{C}$.

Active-set interpretation

- Complementarity conditions for box constraints:

$$\Pi_{\mathcal{A}_a}(x-a) - c\Pi_{\mathcal{I}}\mu = 0$$

with

$$\mathcal{A}_a = \{i \mid (x_i - a_i) + c\mu_i < 0\} \qquad \text{and} \qquad \mathcal{I} = \{1,\dots,n\} \setminus \mathcal{A}_a$$

$$x_i = a \text{ for } i \in \mathcal{A}_a \quad \text{and} \quad \mu_i = 0 \text{ for } i \in \mathcal{I}$$

- Complementarity conditions for L^1-norm sparsity constraints:

$$\Pi_{\mathcal{A}_0} x - c(\Pi_{\mathcal{I}_+}(\mu-\beta) + \Pi_{\mathcal{I}_-}(\mu+\beta)) = 0.$$

with

$$\mathcal{A}_0 = \{i \mid x_i + c(\mu_i+\beta) \geq 0\} \cup \{i \mid x_i + c(\mu_i - \beta) \leq 0\}$$

$$\mathcal{I}_+ = \{i \mid x + c(\mu_i - \beta) > 0\} \qquad \text{and} \qquad \mathcal{I}_- = \{i \mid x_i + c(\mu_i+\beta) < 0\}$$

$$x_i = 0 \text{ for } i \in \mathcal{A}_0 \quad \text{and} \quad \mu_i = -\beta \text{ for } i \in \mathcal{I}_- \text{ and } \mu_i = \beta \text{ for } i \in \mathcal{I}_+.$$

- $\Pi_{\mathcal{C}}$ is an $n \times n$ diagonal 0-1 matrix with 1s corresponding to $\mathcal{C}$.

Optimality conditions for the optimal control problem with bound and sparsity constraints

The KKT system [Stadler COAP 2009]

The solution $(\bar{y}, \bar{u}) \in H_0^1(\Omega) \times L^2(\Omega)$ of the optimal control problem is characterized by the existence of $(\bar{p}, \bar{\mu}) \in H_0^1(\Omega) \times L^2(\Omega)$ such that

$$\mathcal{L}\bar{y} - \bar{u} = 0$$
$$\mathcal{L}^*\bar{p} + \bar{y} - y_d = 0$$
$$-\bar{p} + \alpha\bar{u} + \bar{\mu} = 0$$

$$F(u, \mu; c, \beta) := \bar{u} - \max(0, \bar{u} + c(\bar{\mu} - \beta)) - \min(0, \bar{u} + c(\bar{\mu} + \beta)) + \max(0, (\bar{u} - b) + c(\bar{\mu} - \beta)) + \min(0, (\bar{u} - a) + c(\bar{\mu} + \beta)) = 0$$

a.e. in Ω, with $c > 0$.

- The complementarity function F is nonlinear and semismooth $\Rightarrow$ Semismooth Newton's method for the KKT system, i.e. a Newton's method where the Jacobian of the system is obtained using generalized derivatives.

The semismooth Newton's method as active-set strategy

- Let us define the disjoint sets (defined *a.e. in* Ω)

$$
\begin{aligned}
\mathcal{A}_b &= \{x \in \Omega \mid c(\mu - \beta) + (u - b) > 0\} \\
\mathcal{A}_a &= \{x \in \Omega \mid c(\mu + \beta) + (a - u) < 0\} \\
\mathcal{A}_0 &= \{x \in \Omega \mid u + c(\mu + \beta) \geq 0\} \cup \{x \in \Omega \mid u + c(\mu - \beta) \leq 0\} \\
\mathcal{I}_+ &= \{x \in \Omega \mid u + c(\mu - \beta) > 0\} \cup \{x \in \Omega \mid c(\mu - \beta) + (u - b) \leq 0\} \\
\mathcal{I}_- &= \{x \in \Omega \mid u + c(\mu + \beta) < 0\} \cup \{x \in \Omega \mid c(\mu + \beta) + (u - a) \geq 0\}.
\end{aligned}
$$

Then

$$\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_a \cup \mathcal{A}_0$$

is the set of active constraints and the set of inactive constraints is

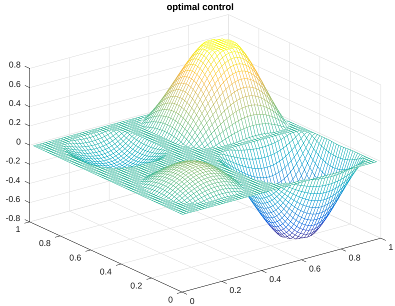
$$\mathcal{I} = \mathcal{I}_+ \cup \mathcal{I}_-.$$

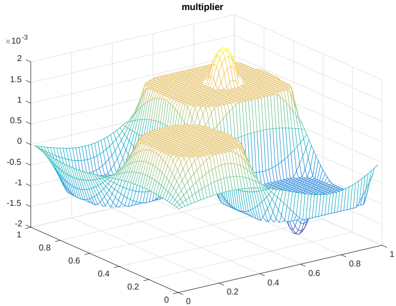
- The complementarity equation becomes

$$\chi_{\mathcal{A}_0} u + \chi_{\mathcal{A}_b}(u - b) + \chi_{\mathcal{A}_a}(u - a) - c(\chi_{\mathcal{I}_+}(\mu - \beta) + \chi_{\mathcal{I}_-}(\mu + \beta)) = 0$$

where $\chi_{\mathcal{C}}$ denotes the characteristic function of a generic $\mathcal{C}$.

Illustration of the active set approach





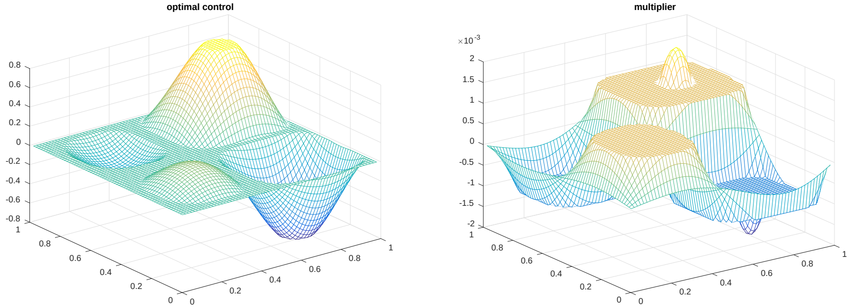
- $u = 0$ for $x \in \mathcal{A}_0$;
- $u = a$ for $x \in \mathcal{A}_a$;
- $u = b$ for $x \in \mathcal{A}_b$;

- $\mu = -\beta$ for $x \in \mathcal{I}_-$;
- $\mu = \beta$ for $x \in \mathcal{I}_+$.

$$\underbrace{\chi_{\mathcal{A}_0} u + \chi_{\mathcal{A}_b}(u-b) + \chi_{\mathcal{A}_a}(u-a)}_{u,\mathcal{A}} \underbrace{- c(\chi_{\mathcal{I}_+}(\mu-\beta) + \chi_{\mathcal{I}_-}(\mu+\beta))}_{\mu,\mathcal{I}} = 0$$

Generalized derivative: $\begin{bmatrix} 0 & \chi_{\mathcal{A}} & 0 & -c\chi_{\mathcal{I}} \end{bmatrix}$

Illustration of the active set approach



- $u = 0$ for $x \in \mathcal{A}_0$;
- $u = a$ for $x \in \mathcal{A}_a$;
- $u = b$ for $x \in \mathcal{A}_b$;

- $\mu = -\beta$ for $x \in \mathcal{I}_-$;
- $\mu = \beta$ for $x \in \mathcal{I}_+$.

$$\underbrace{\chi_{\mathcal{A}_0} u + \chi_{\mathcal{A}_b}(u-b) + \chi_{\mathcal{A}_a}(u-a)}_{u,\mathcal{A}} \underbrace{-c(\chi_{\mathcal{I}_+}(\mu-\beta) + \chi_{\mathcal{I}_-}(\mu+\beta))}_{\mu,\mathcal{I}} = 0$$

Generalized derivative: $\begin{bmatrix} 0 & \chi_{\mathcal{A}} & 0 & -c\chi_{\mathcal{I}} \end{bmatrix}$

kth iteration of the semismooth Newton's method for the KKT

- Assume that the initial point is "feasible" and that the Newton's equation is solved "exactly".
- Given the current iterate (y_k, u_k, p_k, μ_k), a step of the semismooth Newton's method applied to KKT system is

$$\begin{pmatrix} I & \cdot & \mathcal{L}^T & \cdot \\ \cdot & \alpha I & -I & I \\ \mathcal{L} & -I & \cdot & \cdot \\ \cdot & \chi_{\mathcal{A}_k} & \cdot & -c\chi_{\mathcal{I}_k} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta u \\ \Delta p \\ \Delta \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -F(u_k, \mu_k; c, \beta) \end{pmatrix}$$

and its symmetrization

$$\underbrace{\begin{bmatrix} I & \cdot & \mathcal{L}^T & \cdot \\ \cdot & \alpha I & -I & P_{\mathcal{A}_k}^T \\ \mathcal{L} & -I & \cdot & \cdot \\ \cdot & P_{\mathcal{A}_k} & \cdot & \end{bmatrix}}_{J_k} \underbrace{\begin{bmatrix} \Delta y \\ \Delta u \\ \Delta p \\ (\Delta \mu)_{\mathcal{A}_k} \end{bmatrix}}_{\Delta x} = \underbrace{\begin{bmatrix} 0 \\ -\chi_{\mathcal{I}_k}(\mu_{k+1} - \mu_k) \\ 0 \\ -P_{\mathcal{A}_k} F(u_k, \mu_k; c, \beta) \end{bmatrix}}_{f_k}$$

where $P_{\mathcal{A}}$ is the projection on the subspace defined by the active set $\mathcal{A}$.

- Fast local convergence [Stadler09] $\rightarrow$ globalization strategy is needed.

kth iteration of the semismooth Newton's method for the KKT

- Assume that the initial point is "feasible" and that the Newton's equation is solved "exactly".
- Given the current iterate (y_k, u_k, p_k, μ_k), a step of the semismooth Newton's method applied to KKT system is

$$\begin{pmatrix} I & \cdot & \mathcal{L}^T & \cdot \\ \cdot & \alpha I & -I & I \\ \mathcal{L} & -I & \cdot & \cdot \\ \cdot & \chi_{\mathcal{A}_k} & \cdot & -c\chi_{\mathcal{I}_k} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta u \\ \Delta p \\ \Delta \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -F(u_k, \mu_k; c, \beta) \end{pmatrix}$$

and its symmetrization

$$\underbrace{\begin{bmatrix} I & \cdot & \mathcal{L}^T & \cdot \\ \cdot & \alpha I & -I & P_{\mathcal{A}_k}^T \\ \mathcal{L} & -I & \cdot & \cdot \\ \cdot & P_{\mathcal{A}_k} & \cdot & \end{bmatrix}}_{J_k} \underbrace{\begin{bmatrix} \Delta y \\ \Delta u \\ \Delta p \\ (\Delta \mu)_{\mathcal{A}_k} \end{bmatrix}}_{\Delta x} = \underbrace{\begin{bmatrix} 0 \\ -\chi_{\mathcal{I}_k}(\mu_{k+1} - \mu_k) \\ 0 \\ -P_{\mathcal{A}_k} F(u_k, \mu_k; c, \beta) \end{bmatrix}}_{f_k}$$

where $P_{\mathcal{A}}$ is the projection on the subspace defined by the active set $\mathcal{A}$.

- Fast local convergence [Stadler09] $\rightarrow$ globalization strategy is needed.

Discretize-then-optimize: discretization by FE

$$\begin{bmatrix} M & 0 & K^T & 0 \\ 0 & \alpha M & -\overline{M}^T & MP_{\mathcal{A}_k}^T \\ K & -\overline{M} & 0 & 0 \\ 0 & P_{\mathcal{A}_k} M & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta u \\ \Delta p \\ (\Delta \mu)_{\mathcal{A}_k} \end{bmatrix} = \begin{bmatrix} 0 \\ -M\Pi_{\mathcal{I}_k}(\mu_{k+1} - \mu_k) \\ 0 \\ -P_{\mathcal{A}_k} MF(u_k, \mu_k; c, \beta) \end{bmatrix}$$

⋆ Poisson equation $\mathcal{L} = -\Delta$
$\Rightarrow$ $M(= \overline{M})$ and K are the lumped mass (diagonal) and stiffness matrices.

⋆ Convection-Diffusion equation $\mathcal{L} = -\epsilon\Delta + w \cdot \nabla$
$\Rightarrow$ $\bar{M}$ and K are the SUPG mass and stiffness matrices (unsym) and M is the lumped mass matrix (diag).

Since, M is diagonal, componentwise complementarity conditions still hold!

Preconditioning the sequence of Newton equations

$$J_k \Delta x = f_k \qquad (*)$$

where J_k is a 4x4 blocks saddle point matrix of dimension $3n_h + n_{\mathcal{A}_k}$

- Assume that Krylov subspace methods are used to solve the large and sparse Newton equations $\Rightarrow$ preconditioning is mandatory.

Objective

Solving the Newton's equations using effective optimal and robust preconditioners such that the number of iterations required to solve $(*)$ is low and (roughly) independent of the problem parameters $\alpha, \beta,\ h$.

Active-set preconditioners

$$J_k = \left[\begin{array}{cc|cc} M & 0 & K^T & 0 \\ 0 & \alpha M & -\bar{M} & M P_{\mathcal{A}_k}^T \\ \hline K & -\bar{M} & 0 & 0 \\ 0 & P_{\mathcal{A}_k} M & 0 & 0 \end{array}\right] = \begin{bmatrix} A & B_k^T \\ B_k & 0 \end{bmatrix}$$

- A block diagonal preconditioner $\mathcal{P}_k^{BDF}$

$$\mathcal{P}_k^{BDF} = \begin{bmatrix} A & 0 \\ 0 & \widehat{S}_k \end{bmatrix}$$

- An indefinite preconditioner $\mathcal{P}_k^{IPF}$

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

where $\widehat{S}_k \approx S_k = B_k A^{-1} B_k^T$ (active-set Schur complement)

- Proposed for bound-constrained optimal control problems and $\bar{M} = M$ in [Porcelli, Simoncini, Tani, SISC 2015] .

The active-set Schur complement

$$S = \frac{1}{\alpha}\begin{bmatrix} I & -\bar{M}\Pi_{\mathcal{A}}M^{-1}P_{\mathcal{A}}^T \\ 0 & I \end{bmatrix}\begin{bmatrix} \mathbb{S} & 0 \\ 0 & P_{\mathcal{A}}MP_{\mathcal{A}}^T \end{bmatrix}\begin{bmatrix} I & 0 \\ -P_{\mathcal{A}}M^{-1}\Pi_{\mathcal{A}}\bar{M}^T & I \end{bmatrix},$$

where $\mathbb{S}$ is the Schur complement of S,

$$\mathbb{S} = \alpha K M^{-1} K^T + \bar{M}(I - \Pi_{\mathcal{A}})M^{-1}\bar{M}^T$$

The active-set Schur complement approximation

$$\widehat{\mathbb{S}} := (\sqrt{\alpha}K + \bar{M}(I - \Pi_{\mathcal{A}}))M^{-1}(\sqrt{\alpha}K + \bar{M}(I - \Pi_{\mathcal{A}}))^T \quad \approx \quad \mathbb{S}$$

$$\Rightarrow \quad \widehat{S} \approx S$$

From now on the index k is omitted.

- $\widehat{\mathbb{S}} = \mathbb{S} + \sqrt{\alpha}(K(I - \Pi_{\mathcal{A}}) + (I - \Pi_{\mathcal{A}})K^T)$
- If $\mathcal{A} = \{1, \ldots, n\} \Rightarrow \widehat{\mathbb{S}} = \mathbb{S} \Rightarrow S = \widehat{S}$ (exact $\mathcal{P}_k^{IPF}$!)

Spectral properties of the approximation $\widehat{\mathbb{S}}$ ($\bar{M} = M$)

- $\lambda \in \lambda(\widehat{\mathbb{S}}^{-1}\mathbb{S})$ satisfies

$$\frac{1}{2} \leq \lambda \leq \zeta^2 + (1 + \zeta)^2$$

with $\zeta = \|M^{\frac{1}{2}} \left(\sqrt{\alpha}K + M(I - \Pi_{\mathcal{A}})\right)^{-1} \sqrt{\alpha}KM^{-\frac{1}{2}}\|$.

Moreover, if $K + K^T \succ 0$, then for $\alpha \to 0$, ζ is bounded by a constant independent of α;

[Porcelli, Simoncini, Tani, SISC 2015]

Spectral analysis of the preconditioners ($\bar{M} = M$)

Assume that $\widehat{\mathbb{S}}_k$ is nonsingular. Then

$$\lambda(J_k, \mathcal{P}_k^{IPF}) \in \{1\} \cup \Lambda\left(\widehat{\mathbb{S}}_k^{-1}\mathbb{S}_k\right),$$

and

$$\lambda\left(J_k,\ \mathcal{P}_k^{BDF}\right) \in \left\{1, \frac{1 \pm \sqrt{5}}{2}\right\} \cup \left\{\frac{1}{2}\left(1 \pm \sqrt{1+4\sigma^2}\right) \mid \sigma^2 \in \Lambda\left(\widehat{\mathbb{S}}_k^{-1}\mathbb{S}_k\right)\right\}$$

using [Fischer et al. BIT 1988]

Reduced KKT system formulations

3×3 formulation and preconditioner

$$\begin{bmatrix} M & K^T & 0 \\ K & -\frac{1}{\alpha}\bar{M}M^{-1}\bar{M}^T & \frac{1}{\alpha}\bar{M}P_{\mathcal{A}}^T \\ 0 & \frac{1}{\alpha}P_{\mathcal{A}}\bar{M}^T & -\frac{1}{\alpha}P_{\mathcal{A}}MP_{\mathcal{A}}^T \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta p \\ (\Delta\mu)_{\mathcal{A}} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$$

$$\mathcal{P}^{BDF} = \begin{bmatrix} M & 0 \\ 0 & \widehat{S} \end{bmatrix} \qquad \mathcal{P}^{IPF} = \begin{bmatrix} I & 0 \\ [K;0]\,M^{-1} & I \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & -\widehat{S} \end{bmatrix} \begin{bmatrix} I & M^{-1}\,[K^T 0] \\ 0 & I \end{bmatrix}$$

2×2 formulation and preconditioner

$$\begin{bmatrix} M & K^T \\ K & -\frac{1}{\alpha}\bar{M}M^{-1}(I-\Pi_{\mathcal{A}})\bar{M}^T \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta p \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

$$\mathcal{P}^{BDF} = \begin{bmatrix} M & 0 \\ 0 & \frac{1}{\alpha}\widehat{\mathbb{S}} \end{bmatrix} \qquad \mathcal{P}^{IPF} = \begin{bmatrix} I & 0 \\ KM^{-1} & I \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & -\frac{1}{\alpha}\widehat{\mathbb{S}} \end{bmatrix} \begin{bmatrix} I & M^{-1}K^T \\ 0 & I \end{bmatrix}$$

Sparsity constraint

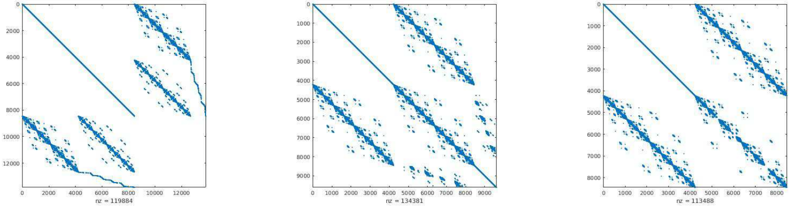


Figure : From left to right: 4×4, 3×3 and 2×2.

Experiments: implementation issues

- Compared Matlab implementations:

AS-GMRES-IPF	Active-set method with linear solver GMRES + $\mathcal{P}_k^{IPF}$
AS-MINRES-BDF	Active-set method with linear solver MINRES + $\mathcal{P}_k^{BDF}$

- Preconditioners via AMG (HSL-MI20)
  - $\mathcal{P}_k^{IPF}$ and $\mathcal{P}_k^{BDF}$: solving with $L_k = (\sqrt{\alpha}K + \bar{M}(I - \Pi_{\mathcal{A}_k}))$ (and L_k^T) for $\widehat{\mathbb{S}}_k$
- FEM matrices from the open source FE library deal.II
- Relative residual for inner stopping criterion, $\mathrm{tol}_{\mathrm{inner}} = 10^{-10}$
- Stopping test for the outer iteration: $\|F(u_k, \mu_k; c, \beta)\| \leq 10^{-8}$
- Semismooth monotone line-search strategy employed [Kanzow, OMS 2014].

Poisson state equation with FDs ($M = \bar{M} = I$)

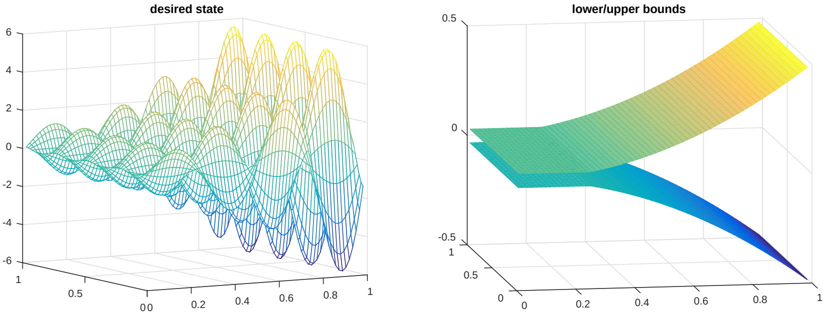


Figure : **Left**: Desired State $y_d = \sin(2\pi x)\sin(2\pi y)\exp(2x)/6$

Right: nonlinear lower/upper bounds $b = \begin{cases} 0.5(0.25)^2 & \text{if } x \leq 0.25 \\ 0.5x^2 & \text{else} \end{cases}, \quad a = -b.$

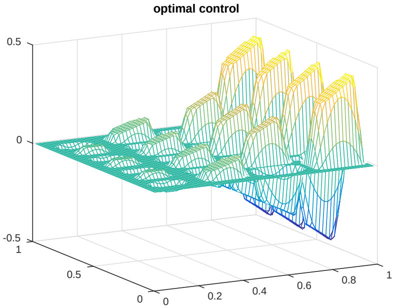
- 2D: $N = 2^p$ with $p = 7, 8, 9 \Rightarrow n = 16384, 65536, 262144$;

Poisson state equation with FDs (2D case)

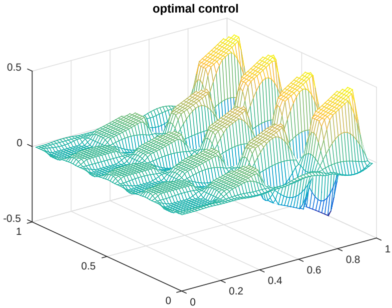
β	p	α	AS-GMRES-IPF LI (NLI)	AS-GMRES-IPF TCPU	AS-MINRES-BDF LI (NLI)	AS-MINRES-BDF TCPU	% u =0
10^{-3}	7	10^{-2}	8.5(2)	5.9	17.5(2)	5.8	47.9
		10^{-4}	8.7(3)	7.0	18.0(3)	6.0	47.9
		10^{-6}	7.3(3)	4.9	15.3(3)	5.0	47.9
	8	10^{-2}	9.0(2)	10.0	19.0(2)	13.2	48.6
		10^{-4}	10.0(2)	12.7	21.5(2)	15.3	48.6
		10^{-6}	6.7(3)	11.6	13.7(3)	17.0	48.6
	9	10^{-2}	9.5(2)	33.7	17.5(2)	45.9	48.7
		10^{-4}	9.6(3)	56.0	19.7(3)	82.0	48.7
		10^{-6}	9.3(3)	48.0	19.0(3)	73.3	48.7
$2 \cdot 10^{-3}$	7	10^{-2}	10.0(2)	6.1	21.0(2)	6.8	71.0
		10^{-4}	11.0(2)	5.6	23.0(2)	5.8	71.0
		10^{-6}	5.0(3)	4.7	8.0(3)	3.7	71.0
	8	10^{-2}	11.0(2)	14.1	23.0(2)	18.5	71.4
		10^{-4}	11.0(2)	14.6	25.0(2)	18.9	71.4
		10^{-6}	8.0(3)	15.8	18.0(3)	20.7	71.4
	9	10^{-2}	11.0(2)	40.2	23.0(2)	60.2	71.5
		10^{-4}	12.0(4)	98.5	25.0(3)	99.8	71.5
		10^{-6}	10.3(4)	74.9	21.5(4)	113.8	71.5

- LI: average number of Linear Iters; NLI is the total number of NonLinear Iters
- TCPU: Total elapsed CPU time (sec.)

Poisson state equation with FDs (2D case)



$\beta = 10^{-3}$



$\beta = 10^{-4}$

Experimental study of parameters

$$\mathcal{L}\bar{y} - \bar{u} = 0, \qquad \mathcal{L}^*\bar{p} + \bar{y} - y_d = 0, \qquad -\bar{p} + \alpha\bar{u} + \bar{\mu} = 0$$

$$F(u,\mu;c,\beta) := \bar{u} - \max(0, \bar{u} + c(\bar{\mu} - \beta)) - \min(0, \bar{u} + c(\bar{\mu} + \beta)) + \max(0, (\bar{u} - b) + c(\bar{\mu} - \beta)) + \min(0, (\bar{u} - a) + c(\bar{\mu} + \beta)) = 0$$

2D problem (GMRES w/indef precond): $\alpha = 10^{-4}$, $\beta = 10^{-4}$, $p = 7$

$$c = c_{fact}/\alpha$$

c_{fact}	LI (NLI)	CPU	TCPU
0.001	15.8(*)	4.9	487.8
0.1	15.9(*)	4.7	465.0
0.2	15.9(*)	4.9	486.3
0.5	16.2(5)	4.1	20.9
1	16.0(5)	4.1	20.8
2	16.2(7)	4.1	29.2
5	16.6(13)	4.5	59.1
10	16.8(18)	4.3	79.0
100	17.1(74)	4.5	337.0

Convection-Diffusion state equation with FE (1)

$$-\epsilon\Delta y + w \cdot \nabla y = u$$

with $w = (2y(1-x^2), -2x(1-y^2))$

- SUPG discretization for $\bar{M}, K \in \mathbb{R}^{n\times n}$

$$n = 4225, \qquad \beta = 10^{-4}$$

AS-GMRES-IPF

	$\epsilon = 1$		$\epsilon = 0.5$		$\epsilon = 0.1$	
α	LI (NLI)	BT	LI (NLI)	BT	LI (NLI)	BT
10^{-1}	14.0(1)	0	15.0(1)	0	14.7(3)	2
10^{-2}	15.5(2)	1	15.7(3)	2	17.9(8)	15
10^{-3}	15.3(6)	5	16.9(9)	12	21.5(26)	69
10^{-4}	16.9(11)	21	19.3(15)	34	26.7(48)	165
10^{-5}	22.9(20)	50	26.5(24)	80	36.6(98)	463

- LI: average number of Linear Iterations
- NLI: total number of NonLinear Iterations
- BT: total number of Back-Tracking steps in the line-search strategy

Convection-Diffusion state equation with FE (2)

AS-GMRES-IPF

$n = 16641, \beta = 10^{-4}, \epsilon = 1$

	4 × 4		3 × 3		2 × 2	
α	LI (NLI)	TCPU	LI (NLI)	TCPU	LI (NLI)	TCPU
10^{-1}	17.3(3)	6.5	17.3(3)	6.0	17.3(3)	5.9
10^{-3}	22.3(24)	67.1	21.6(21)	55.2	22.3(24)	61.6
10^{-5}	38.1(69)	371.6	38.7(73)	384.5	37.8(69)	322.5

$n = 66049, \beta = 10^{-4}, \epsilon = 1$

	4 × 4		3 × 3		2 × 2	
α	LI (NLI)	TCPU	LI (NLI)	TCPU	LI (NLI)	TCPU
10^{-1}	19.7(3)	30.5	19.7(3)	28.8	19.7(3)	28.0
10^{-3}	24.4(21)	**281.6**	24.3(21)	**266.6**	20.9(21)	**250.0**
10^{-5}	38.7(58)	1360.3	39.3(52)	1099.1	38.7(58)	1104.9

- LS needed for convergence in 80% of the runs.

Conclusions

- Preconditioned semismooth Newton's method satisfactorily handles L^1 norm sparsity constraints
- Preliminary numerical experiments have showed good performance wrto different parameters

Current work

- Inexact (semi-residual based) semismooth Newton's method for optimal control with L^1 term;
- Spectral properties of the Schur complement approximation for different state equation (CD, Stokes).
- Different sparsity constraints (see [Herzog et a. SICON 2014])

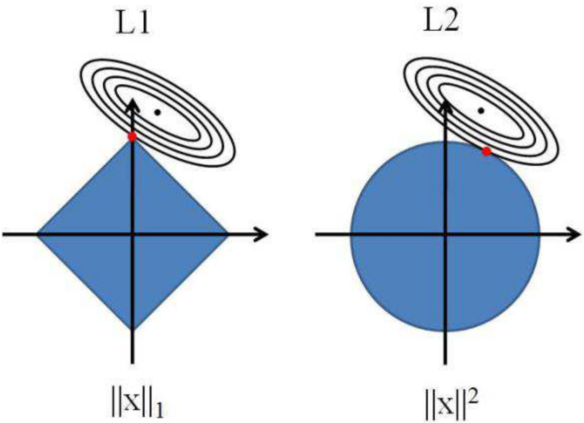
Conclusions

- Preconditioned semismooth Newton's method satisfactorily handles L^1 norm sparsity constraints
- Preliminary numerical experiments have showed good performance wrto different parameters

Current work

- Inexact (semi-residual based) semismooth Newton's method for optimal control with L^1 term;
- Spectral properties of the Schur complement approximation for different state equation (CD, Stokes).
- Different sparsity constraints (see [Herzog et a. SICON 2014])

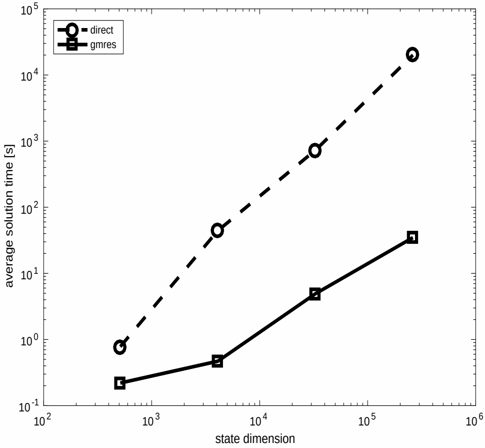
Illustration of the L^1norm penalty



Pictures from Tianyi Zhou's Research Blog

Sort of "sanity check"

Iterative vs direct (sparse) solution ("backslash")



here $\alpha = 10^{-6}, \quad \beta = 10^{-4}$

Sparsity constraint

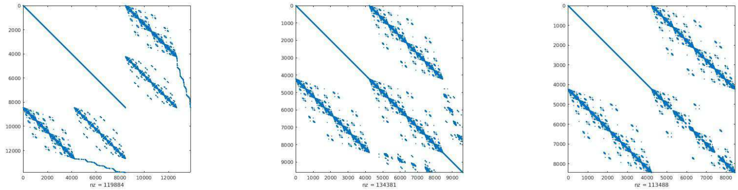


Figure : From left to right: 4×4, 3×3 and 2×2.

Poisson state equation with FDs (3D case)

- 3D: $N = 2^p$ with $p = 4, 5, 6 \Rightarrow n = 4096, 32768, 262144$.

$$\beta = 10^{-4} \quad u = 0 \approx 20\%$$

p	α	AS-GMRES-IPF 4×4 LI (NLI)	TCPU	AS-GMRES-IPF 3×3 LI (NLI)	TCPU	AS-GMRES-IPF 2×2 LI (NLI)	TCPU
4	10^{-2}	10.0(1)	0.5	10.0(1)	0.5	10.0(1)	0.4
	10^{-4}	13.0(2)	1.1	13.0(2)	1.0	13.0(2)	0.9
	10^{-6}	5.0(3)	0.7	5.0(3)	0.7	5.0(3)	0.6
5	10^{-2}	9.0(1)	2.1	9.0(1)	1.9	9.0(1)	1.8
	10^{-4}	14.0(2)	7.2	14.0(2)	6.3	14.0(2)	5.8
	10^{-6}	10.5(2)	5.0	10.5(2)	4.4	10.5(2)	4.1
6	10^{-2}	10.0(1)	18.1	10.0(1)	16.2	10.0(1)	17.1
	10^{-4}	14.0(2)	83.6	14.0(2)	75.7	14.0(2)	76.1
	10^{-6}	14.7(3)	83.0	14.7(3)	73.6	14.7(3)	68.2