November 5, 2010

Saddle Point Systems in Optimal Control

Martin Stoll

Computational Methods in Systems and Control Theory
Max Planck Institute for Dynamics of Complex Technical Systems

"I would rather have today's algorithms on yesterday's computers than vice versa." – Phillipe Toint (Namur)

Optimization with constraints

Consider the Quadratic Programming (QP) problem

$$\min_{x} \; x^T A x + x^T b$$
$$\text{s.t.} \; Bx = c$$

with $A \in \mathbb{R}^{n,n}$ and $B \in \mathbb{R}^{m,n}$.

Optimization with constraints

Consider the Quadratic Programming (QP) problem

$$\min_x \; x^T A x + x^T b$$
$$\text{s.t. } \; Bx = c$$

with $A \in \mathbb{R}^{n,n}$ and $B \in \mathbb{R}^{m,n}$. KKT conditions give

$$\underbrace{\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}}_{\mathcal{K}} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -b \\ c \end{bmatrix}$$

This is a so-called saddle point problem.

Properties of the saddle point system

$$\mathcal{K} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

- $\mathcal{K}$ is non-singular if B has full rank and A is positive on $\ker(B)$
- $\mathcal{K}$ is symmetric and indefinite
- $\mathcal{K}$ is typically poorly conditioned
- $\mathcal{K}$ has n positive and m negative eigenvalues

For more details see [BGL'05][1].

[1] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer, 14 (2005), pp. 1–137.

$$\mathcal{K} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

- $\mathcal{K}$ is non-singular if B has full rank and A is positive on $\ker(B)$
- $\mathcal{K}$ is symmetric and indefinite
- $\mathcal{K}$ is typically poorly conditioned
- $\mathcal{K}$ has n positive and m negative eigenvalues

For more details see [BGL'05][1].

Fact

In almost all (large, 3D) applications it is not feasible to factorize $\mathcal{K}$! What? No backslash?! Get out!

[1] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer, 14 (2005), pp. 1–137.

Iterative solvers for saddle point systems

Don't worry, we can save the day.

Krylov-subspace solvers

Iterative solvers can be applied.

- Only need (one) matrix vector multiplication with $\mathcal{K}$.
- Usually satisfy an optimality criterion for residual or error at the k-th step.
- Use space $\mathrm{span}\left\{r_0, \mathcal{K}r_0, \ldots, \mathcal{K}^{k-1}r_0\right\}$ with $r_0 = b - \mathcal{K}x_0$.

Iterative solvers for saddle point systems

Don't worry, we can save the day.

Krylov-subspace solvers

Iterative solvers can be applied.

- Only need (one) matrix vector multiplication with $\mathcal{K}$.
- Usually satisfy an optimality criterion for residual or error at the k-th step.
- Use space $\mathrm{span}\left\{r_0, \mathcal{K}r_0, \ldots, \mathcal{K}^{k-1}r_0\right\}$ with $r_0 = b - \mathcal{K}x_0$.

More bad news?

These methods might be incredibly slow! Depending on the eigenvalues of $\mathcal{K}$ (rule of thumb).

Iterative solvers for saddle point systems

Don't worry, we can save the day.

Krylov-subspace solvers

Iterative solvers can be applied.

- Only need (one) matrix vector multiplication with $\mathcal{K}$.
- Usually satisfy an optimality criterion for residual or error at the k-th step.
- Use space $\mathrm{span}\left\{r_0, \mathcal{K}r_0, \ldots, \mathcal{K}^{k-1}r_0\right\}$ with $r_0 = b - \mathcal{K}x_0$.

More bad news?

These methods might be incredibly slow! Depending on the eigenvalues of $\mathcal{K}$ (rule of thumb).

No, we need to use preconditioning. Solve

$$\mathcal{P}^{-1}\mathcal{K}x = \mathcal{P}^{-1}b.$$

Some general preconditioning results

In [MGW '00][2] it is shown that an "ideal" block-diagonal preconditioner

$$\mathcal{P}_{BD} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}$$

where $-S = -BA^{-1}B^T$ is the Schur-complement of $\mathcal{K}$ leads to the preconditioned system $\mathcal{P}_{BD}^{-1}\mathcal{K}$ having three distinct eigenvalues at 1 and $1 \pm \frac{\sqrt{5}}{2}$.

[2] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput, 21 (2000), pp. 1969–1972.

Some general preconditioning results

In [MGW '00][2] it is shown that an "ideal" block-diagonal preconditioner

$$\mathcal{P}_{BD} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}$$

where $-S = -BA^{-1}B^T$ is the Schur-complement of $\mathcal{K}$ leads to the preconditioned system $\mathcal{P}_{BD}^{-1}\mathcal{K}$ having three distinct eigenvalues at 1 and $1 \pm \frac{\sqrt{5}}{2}$. For block-triangular preconditioner

$$\mathcal{P}_{BT} = \begin{bmatrix} A & 0 \\ B & -S \end{bmatrix}$$

the eigenvalues of $\mathcal{P}_{BT}^{-1}\mathcal{K}$ are given by 1. (Convergence in at most two iterations)

[2] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput, 21 (2000), pp. 1969–1972.

- Can use non-standard version[3] of Conjugate Gradients CG[4] with block-triangular preconditioner.
- Straightforward use of the preconditioned Minimal Residual Method MINRES[5] with block-diagonal preconditioner.

[3] J. H. Bramble and J. E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp, 50 (1988), pp. 1–17.

[4] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand, 49 (1952), pp. 409–436 (1953).

[5] C. C. Paige and M. A. Saunders, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal, 12 (1975), pp. 617–629.

A control model problem

with Andy Wathen (Oxford) and Tyrone Rees (University of British Columbia)

The functional to be minimized over a domain $\Omega \in \mathbb{R}^d$ with $d = 2, 3$ is given by

$$J(y, u) := \frac{1}{2} \|\mathrm{y} - \bar{\mathrm{y}}\|^2_{L^2(\Omega)} + \frac{\beta}{2} \|\mathrm{u}\|^2_{L^2(\Omega)},$$

subject to the state equation

$$-\triangle \mathrm{y} = \mathrm{u} \text{ in } \Omega$$

with y being the state, u the control and $\bar{y}$ the desired state.

A control model problem

with Andy Wathen (Oxford) and Tyrone Rees (University of British Columbia)

The functional to be minimized over a domain $\Omega \in \mathbb{R}^d$ with $d = 2, 3$ is given by

$$J(y, u) := \frac{1}{2} \|\mathrm{y} - \bar{\mathrm{y}}\|^2_{L^2(\Omega)} + \frac{\beta}{2} \|\mathrm{u}\|^2_{L^2(\Omega)},$$

subject to the state equation

$$-\triangle \mathrm{y} = \mathrm{u} \text{ in } \Omega$$

with y being the state, u the control and $\bar{y}$ the desired state. Additionally, we allow for the control to be bounded by so-called *box constraints*

$$\mathrm{u}_a(x) \leq \mathrm{u}(x) \leq \mathrm{u}_b(x) \text{ a.e in } \Omega$$

or the state

$$\mathrm{y}_a(x) \leq \mathrm{y}(x) \leq \mathrm{y}_b(x) \text{ a.e in } \Omega.$$

A control model problem

Discretize-then-Optimize

The inverse problem is discretized following a *discretize-then-optimize* strategy by using finite elements to get

$$\begin{cases} \min \frac{1}{2}\left(\mathbf{y}-\bar{\mathbf{y}}\right)^T M\left(\mathbf{y}-\bar{\mathbf{y}}\right)+\frac{\beta}{2}\mathbf{u}^T M\mathbf{u} \quad \text{s.t.} \\ K\mathbf{y}=M\mathbf{u}-f \\ \mathbf{u_a}\leq\mathbf{u}\leq\mathbf{u_b} \\ \mathbf{y_a}\leq\mathbf{y}\leq\mathbf{y_b} \end{cases}$$

with K the stiffness matrix, M the mass matrix and $\mathbf{y},\mathbf{u},\bar{\mathbf{y}}$ vectors representing the state, control and desired state.

The first order or KKT conditions now result in the following linear system

$$\begin{bmatrix} M & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}^{(k)} \\ \mathbf{u}^{(k)} \\ \boldsymbol{\lambda}^{(k)} \end{bmatrix} = \begin{bmatrix} M\bar{\mathbf{y}} \\ 0 \\ f \end{bmatrix}.$$

The system matrix is in *saddle point form*.

Symmetric system can be solved with

- MINRES with block-diagonal preconditioning

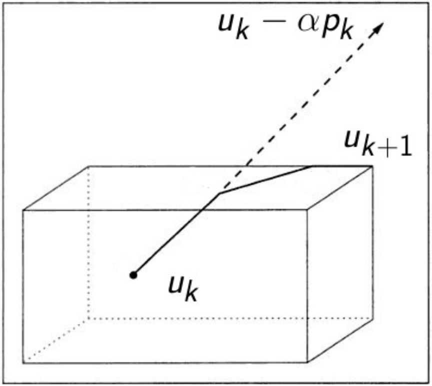
$$\mathcal{P}_{BD} = \begin{bmatrix} A_0 & 0 \\ 0 & S_0 \end{bmatrix}.$$

- Bramble-Pasciak CG with a block-triangular preconditioner

$$\mathcal{P}_{BT} = \begin{bmatrix} A_0 & 0 \\ B & -S_0 \end{bmatrix}.$$

where A_0 might be the Chebyshev semi-iteration for mass matrices and S_0 will involve two approximations to the PDE via an algebraic or geometric multigrid cycle.

Box constraints
Motivation



Projection of the search direction onto the admissible set.

The Newton method

Looks complicated? But only the blue bit is!

Algorithm 1 Primal dual active set method (PDAS)

Given $\mathcal{A}_+^{(0)}$ and $\mathcal{A}_-^{(0)}$
for $k = 0, 1, \ldots$ **do**
    Set $u_{(k)}$ on $\mathcal{A}_\pm^{(k)}$ and $\mu_{(k)} = 0$ on $\mathcal{I}^{(k)}$
    Solve saddle point system
    Compute $\mu_{(k)}$ on $\mathcal{A}_\pm^{(k)}$
    Compute $\mathcal{A}_\pm^{(k+1)}$
    if $\mathcal{A}_+^{(k)} = \mathcal{A}_+^{(k-1)}$, $\mathcal{A}_-^{(k)} = \mathcal{A}_-^{(k-1)}$, and $\mathcal{I}^{(k)} = \mathcal{I}^{(k-1)}$ **then**
        STOP (Algorithm converged)
    end if
end for

Numerical Results

Bound constraints

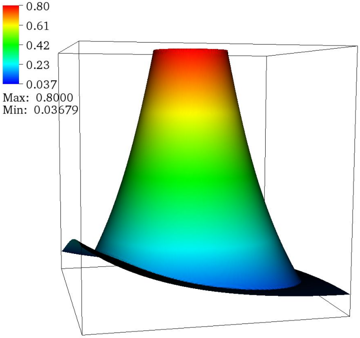


Figure: Computed Control

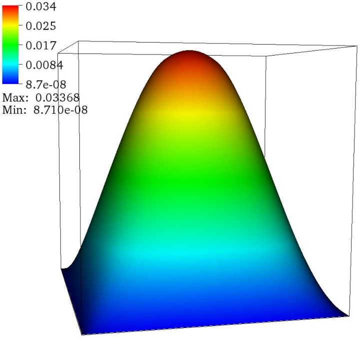


Figure: Computed State

Bound constraints on the state

The constraint $\mathbf{y_a} \leq \mathbf{y} \leq \mathbf{y_b}$ is significantly harder. Now use a Moreau-Yosida penalty function

$$J(y,u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Omega)} + \frac{1}{2\varepsilon}\|\max\{0, y - y_b\}\|^2_{L^2(\Omega)} + \frac{1}{2\varepsilon}\|\min\{0, y - y_a\}\|^2_{L^2(\Omega)},$$

Bound constraints on the state

The constraint $\mathbf{y_a} \leq \mathbf{y} \leq \mathbf{y_b}$ is significantly harder. Now use a Moreau-Yosida penalty function

$$J(y,u) := \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Omega)} + \frac{1}{2\varepsilon}\|\max\{0, y-y_b\}\|^2_{L^2(\Omega)} + \frac{1}{2\varepsilon}\|\min\{0, y-y_a\}\|^2_{L^2(\Omega)},$$

which gives the following Newton systems

$$\begin{bmatrix} M+\varepsilon^{-1}G_{\mathcal{A}}MG_{\mathcal{A}} & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}^{(k+1)} \\ \mathbf{u}^{(k+1)} \\ \boldsymbol{\lambda}^{(k+1)} \end{bmatrix} = rhs$$

State constraints, some pictures

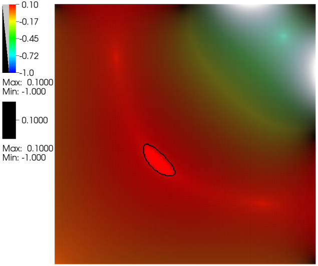


Figure: Computed State

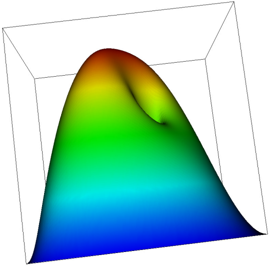


Figure: Computed Control

Time-dependent problems

Minimize

$$J_1(y,u) := \frac{1}{2}\int_{\Omega_1} (\mathrm{y(x,T)} - \bar{\mathrm{y}}(\mathrm{x}))^2 + \frac{\beta}{2}\int_0^T \int_{\Omega_2} (\mathrm{u(x,t)})^2$$

or

$$J_2(y,u) := \frac{1}{2}\int_0^T \int_{\Omega_1} (\mathrm{y(x,t)} - \bar{\mathrm{y}}(\mathrm{x,t}))^2 + \frac{\beta}{2}\int_0^T \int_{\Omega_2} (\mathrm{u(x,t)})^2$$

subject to

$$y_t - \triangle y = u$$

in $\Omega \times [0, T]$, with boundary conditions $y = 0$ on the spatial boundary $\partial\Omega$ and initial condition $y(x,0) = y_0(x)$.

and again a saddle point problem

We now get a GIGANTIC saddle point system

$$\begin{bmatrix} \tilde{\mathcal{M}} & 0 & -\mathcal{K}^T \\ 0 & \beta\tau\mathcal{M} & \tau\mathcal{M} \\ -\mathcal{K} & \tau\mathcal{M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \tilde{\mathcal{M}}\bar{\mathbf{y}} \\ 0 \\ d \end{bmatrix}$$

with

$$\underbrace{\begin{bmatrix} M+\tau K & & & & \\ -M & M+\tau K & & & \\ & -M & M+\tau K & & \\ & & \ddots & \ddots & \\ & & & -M & M+\tau K \end{bmatrix}}_{\mathcal{K}} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} - \tau\mathcal{M}\mathbf{u} = -d$$

coming from the backward Euler scheme

$$M\mathbf{y}_k + \tau K\mathbf{y}_k = M\mathbf{y}_{k-1} + \tau M\mathbf{u}_k.$$

Some results for time-dependent control

State and control at $t = 1$.

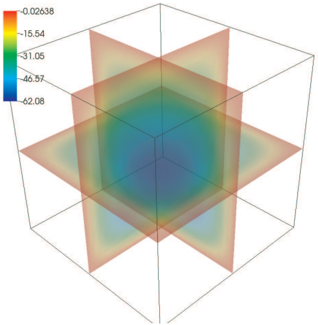


Figure: Computed State

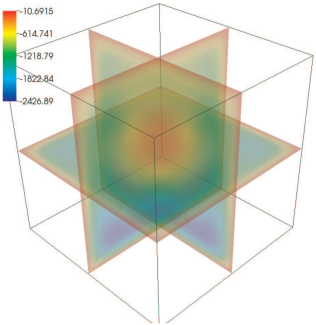


Figure: Computed Control

Conclusions

- Saddle point systems are everywhere!
- Each problem needs special attention.
- We need to take the structure into account.

More information on
http://www.mpi-magdeburg.mpg.de/people/stollm
or come to S3.10.

Thank you!

The plan to increase productivity by canceling coffee breaks flopped.