Deflation techniques for distinct solutions of nonlinear PDEs

P. E. Farrell[1,2] Á. Birkisson[1] S. W. Funke[2]

[1]University of Oxford

[2]Simula Research Laboratory, Oslo

June 16, 2015

Section 1

Motivation

A central question in scientific computing

How can we compute multiple solutions of PDEs?

A central question for my talk

Why should we compute multiple solutions of PDEs?

A central question for my talk

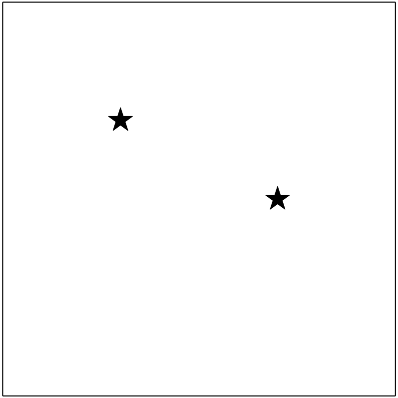
Why should we compute multiple solutions of PDEs?

Answer #1

Prediction.

A central question for my talk

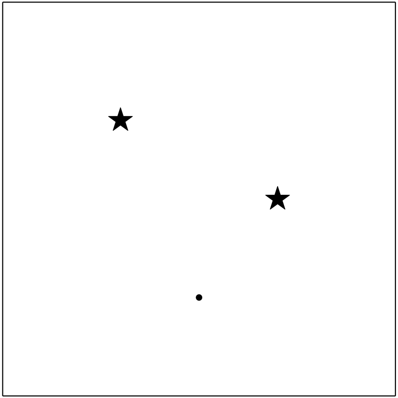
Why should we compute multiple solutions of PDEs?



A PDE with two unknown solutions.

A central question for my talk

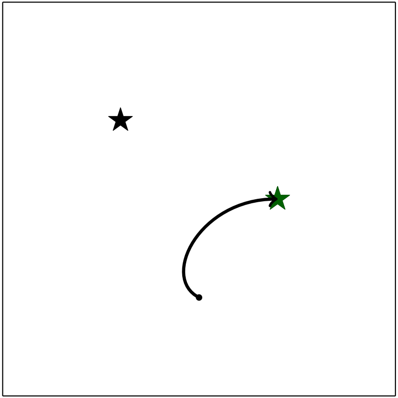
Why should we compute multiple solutions of PDEs?



Start from some initial guess.

A central question for my talk

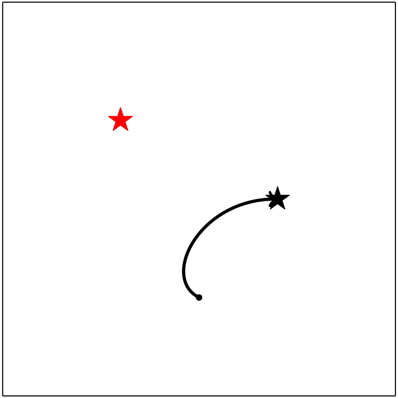
Why should we compute multiple solutions of PDEs?



We converge to one solution, our prediction.

A central question for my talk

Why should we compute multiple solutions of PDEs?



But nature has chosen another (unknown) solution!

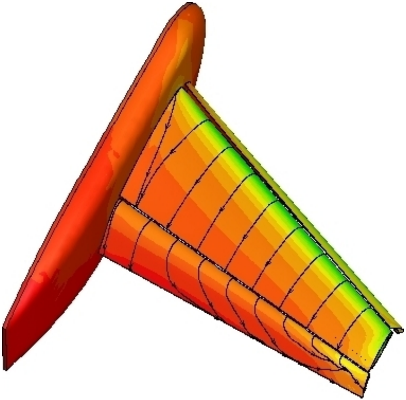
A central question for my talk

Why should we compute multiple solutions of PDEs?

The AIAA/NASA high lift prediction test case (Kamenetskiy et al., 2013).

A central question for my talk

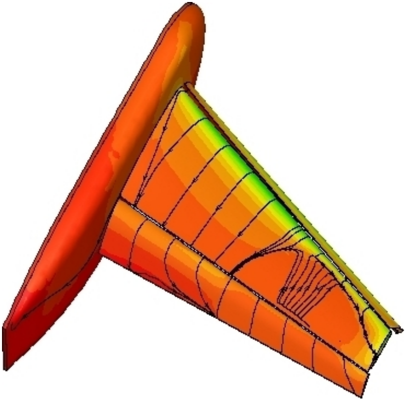
Why should we compute multiple solutions of PDEs?



The AIAA/NASA high lift prediction test case (Kamenetskiy et al., 2013).

A central question for my talk

Why should we compute multiple solutions of PDEs?



The AIAA/NASA high lift prediction test case (Kamenetskiy et al., 2013).

A central question for my talk

Why should we compute multiple solutions of PDEs?

We have encountered unexpected multiple solutions in both simple and complex configurations in computational fluid dynamics (CFD); this phenomenon is both extremely important and not well understood. It has serious implications for the use of CFD as a predictive tool.

— Venkat Venkatakrishnan
Computational Aerodynamic Optimization
Boeing Research & Technology

A central question for my talk

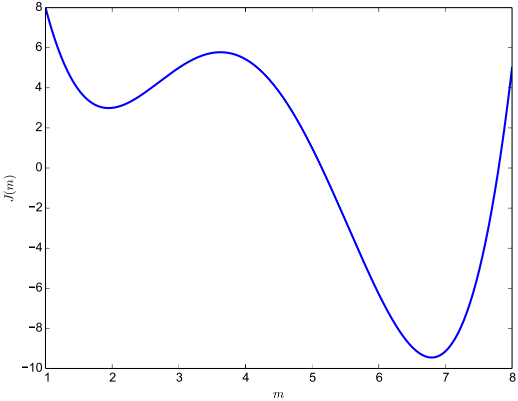
Why should we compute multiple solutions of PDEs?

Answer #2

Optimisation.

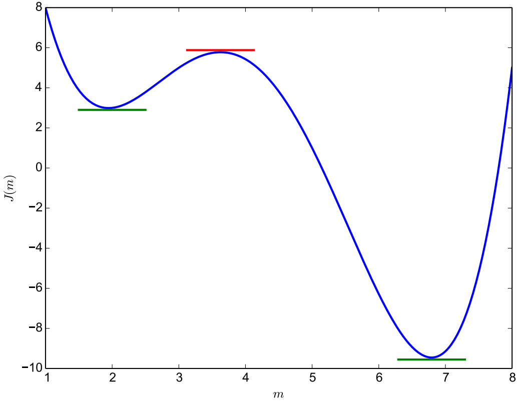
A central question for my talk

Why should we compute multiple solutions of PDEs?



A central question for my talk

Why should we compute multiple solutions of PDEs?



By solving $\nabla J = 0$, we can find a superset of the minima.

A central question for my talk

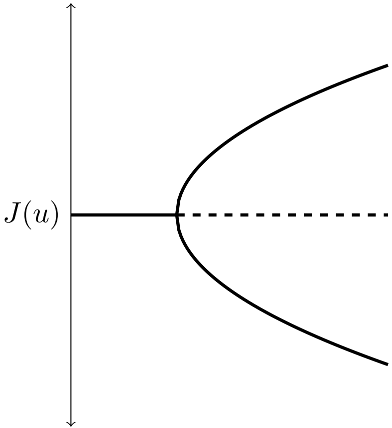
Why should we compute multiple solutions of PDEs?

Answer #3

Applications.

A central question for my talk

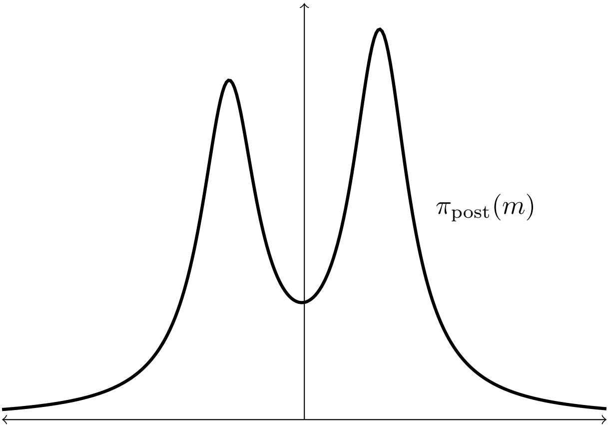
Why should we compute multiple solutions of PDEs?



Scalable tracing of bifurcation diagrams.

A central question for my talk

Why should we compute multiple solutions of PDEs?



Multimodal Bayesian inference.

Section 2

Deflation

The core idea

Deflation

Given

- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

The core idea

Deflation

Given

- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

construct a **new nonlinear problem** $\mathcal{G} : V \to Z$ such that:

The core idea

Deflation

Given

- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

construct a **new nonlinear problem** $\mathcal{G} : V \to Z$ such that:

- (Preservation of solutions.) $\mathcal{F}(\tilde{r}) = 0 \iff \mathcal{G}(\tilde{r}) = 0.$

The core idea

Deflation

Given

- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

construct a **new nonlinear problem** $\mathcal{G} : V \to Z$ such that:

- (Preservation of solutions.) $\mathcal{F}(\tilde{r}) = 0 \iff \mathcal{G}(\tilde{r}) = 0.$
- (Deflation property.) Newton's method applied to $\mathcal{G}$ will never converge to r again, starting from any initial guess.

The core idea

Deflation

Given

- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

construct a **new nonlinear problem** $\mathcal{G} : V \to Z$ such that:

- (Preservation of solutions.) $\mathcal{F}(\tilde{r}) = 0 \iff \mathcal{G}(\tilde{r}) = 0.$
- (Deflation property.) Newton's method applied to $\mathcal{G}$ will never converge to r again, starting from any initial guess.

Find more solutions, starting from the same initial guess.

The core idea

Deflation

Given

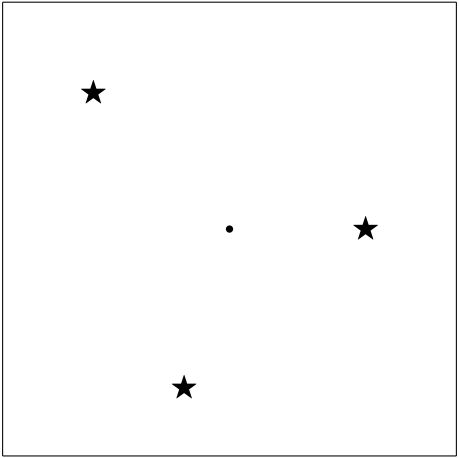
- a Fréchet differentiable residual $\mathcal{F} : V \to W$
- a solution $r \in V$, $\mathcal{F}(r) = 0$, $\mathcal{F}'(r)$ nonsingular
- $\tilde{r} \in V$, $\tilde{r} \neq r$

construct a **new nonlinear problem** $\mathcal{G} : V \to Z$ such that:

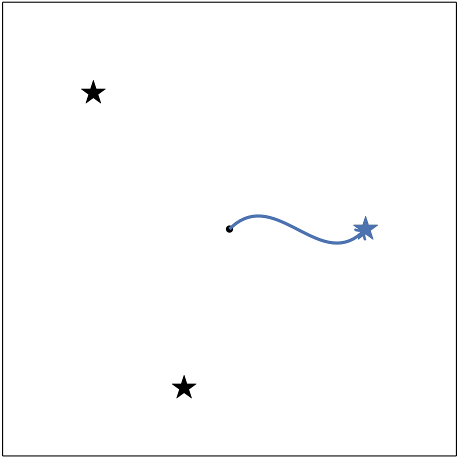
- (Preservation of solutions.) $\mathcal{F}(\tilde{r}) = 0 \iff \mathcal{G}(\tilde{r}) = 0$.
- (Deflation property.) Along any sequence converging to r, $||\mathcal{G}||_Z$ is bounded away from 0.

Find more solutions, starting from the same initial guess.

Finding many solutions from the same guess

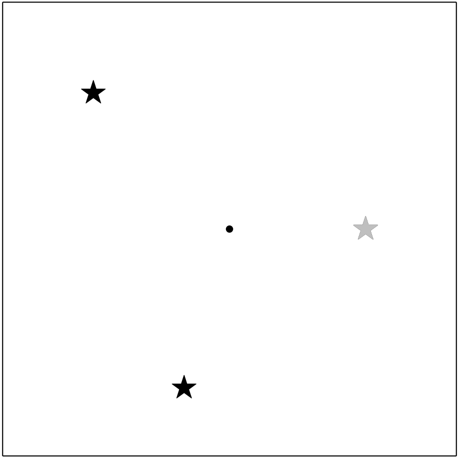


Finding many solutions from the same guess



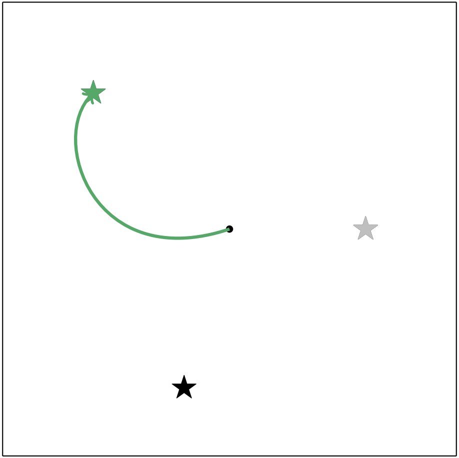
Step I: Newton from initial guess

Finding many solutions from the same guess



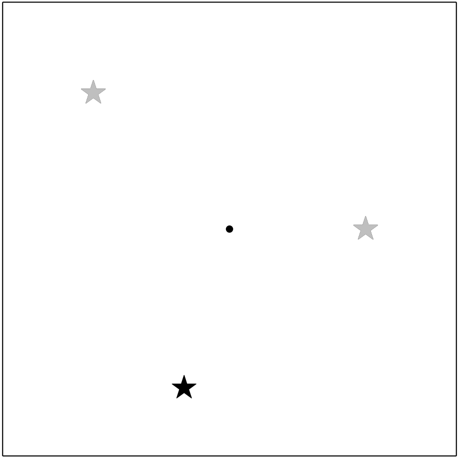
Step II: deflate solution found

Finding many solutions from the same guess



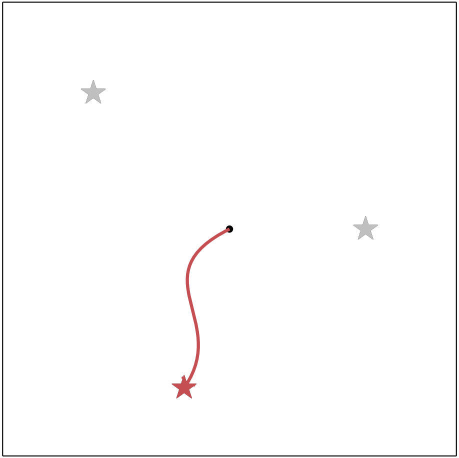
Step I: Newton from initial guess

Finding many solutions from the same guess



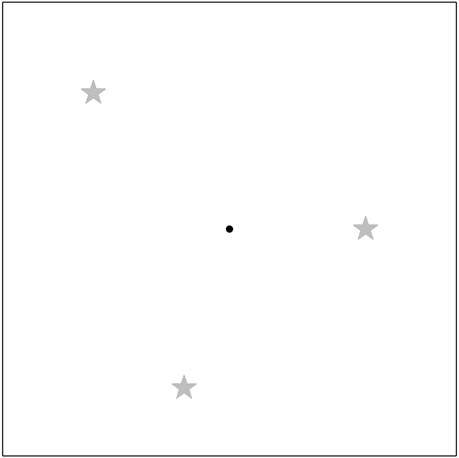
Step II: deflate solution found

Finding many solutions from the same guess



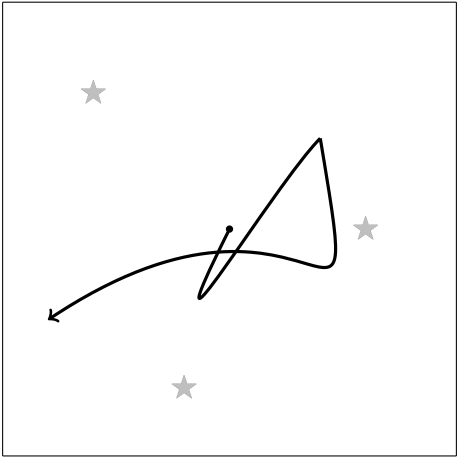
Step I: Newton from initial guess

Finding many solutions from the same guess



Step II: deflate solution found

Finding many solutions from the same guess



Step III: termination on nonconvergence

Finding many solutions from the same guess

Step III: termination on nonconvergence

Construction of deflated problems

A nonlinear transformation

$$\mathcal{G}(u) = \mathcal{M}(u; r)\mathcal{F}(u)$$

Construction of deflated problems

A nonlinear transformation

$$\mathcal{G}(u) = \mathcal{M}(u; r)\mathcal{F}(u)$$

A deflation operator

For $r \in V, u \in V \setminus \{r\}$, let $\mathcal{M}(u; r)$ be an invertible linear operator.
$\mathcal{M}(u; r) : W \to Z$ is a **deflation operator** if for any sequence $u_i \xrightarrow{U} r$

$$\liminf_{i \to \infty} ||\mathcal{M}(u_i; r)\mathcal{F}(u_i)||_Z > 0.$$

Construction of deflated problems

A nonlinear transformation

$$\mathcal{G}(u) = \mathcal{M}(u; r)\mathcal{F}(u)$$

A deflation operator

For $r \in V, u \in V \setminus \{r\}$, let $\mathcal{M}(u; r)$ be an invertible linear operator. $\mathcal{M}(u; r) : W \to Z$ is a **deflation operator** if for any sequence $u_i \xrightarrow{U} r$

$$\liminf_{i \to \infty} ||\mathcal{M}(u_i; r)\mathcal{F}(u_i)||_Z > 0.$$

Theorem (F., Birkisson, Funke 2014)

The following are deflation operators.

$$\mathcal{M}(u; r) = \frac{\mathcal{I}}{||u - r||}$$

Construction of deflated problems

A nonlinear transformation

$$\mathcal{G}(u) = \mathcal{M}(u; r)\mathcal{F}(u)$$

A deflation operator

For $r \in V, u \in V \setminus \{r\}$, let $\mathcal{M}(u; r)$ be an invertible linear operator. $\mathcal{M}(u; r) : W \to Z$ is a **deflation operator** if for any sequence $u_i \xrightarrow{U} r$

$$\liminf_{i \to \infty} ||\mathcal{M}(u_i; r)\mathcal{F}(u_i)||_Z > 0.$$

Theorem (F., Birkisson, Funke 2014)

The following are deflation operators.

$$\mathcal{M}(u; r) = \frac{\mathcal{I}}{||u - r||^p}$$

Construction of deflated problems

A nonlinear transformation

$$\mathcal{G}(u) = \mathcal{M}(u; r)\mathcal{F}(u)$$

A deflation operator

For $r \in V, u \in V \setminus \{r\}$, let $\mathcal{M}(u; r)$ be an invertible linear operator.
$\mathcal{M}(u; r) : W \to Z$ is a **deflation operator** if for any sequence $u_i \xrightarrow{U} r$

$$\liminf_{i \to \infty} ||\mathcal{M}(u_i; r)\mathcal{F}(u_i)||_Z > 0.$$

Theorem (F., Birkisson, Funke 2014)

The following are deflation operators.

$$\mathcal{M}(u; r) = \frac{\mathcal{I}}{||u - r||^p} + \alpha\mathcal{I}.$$

Prior work

Wilkinson (1963)

Deflation for polynomials, rounding error analysis

Prior work

Wilkinson (1963)

Deflation for polynomials, rounding error analysis

Brown and Gearhart (1971)

Generalisation to $f : \mathbb{R}^n \to \mathbb{R}^n$

Prior work

Wilkinson (1963)

Deflation for polynomials, rounding error analysis

Brown and Gearhart (1971)

Generalisation to $f : \mathbb{R}^n \to \mathbb{R}^n$

Levy and Gomez (1985)

Used deflation in the "tunnelling" method for global optimisation

Prior work

Wilkinson (1963)

Deflation for polynomials, rounding error analysis

Brown and Gearhart (1971)

Generalisation to $f : \mathbb{R}^n \to \mathbb{R}^n$

Levy and Gomez (1985)

Used deflation in the "tunnelling" method for global optimisation

This work

Generalisation to Banach spaces, shifting, applications, **preconditioning**

Section 3

Analysis

Newton–Krylov

A Newton step

$$P_F^{-1} J_F(u_i)\delta u_i = -P_F^{-1} F(u_i)$$

Newton–Krylov

A Newton step

$$P_F^{-1} J_F(u_i) \delta u_i = -P_F^{-1} F(u_i)$$

A deflated Newton step

$$P_G^{-1} J_G(u_i) \delta u_i = -P_G^{-1} G(u_i)$$

Newton–Krylov

A Newton step

$$P_F^{-1} J_F(u_i) \delta u_i = -P_F^{-1} F(u_i)$$

A deflated Newton step

$$P_G^{-1} J_G(u_i) \delta u_i = -P_G^{-1} G(u_i)$$

A problem

J_G is dense.

Preconditioning

Theorem (F., Birkisson, Funke, 2014).

Construct a P_G such that

$$\|P_G^{-1} J_G - I\| \leq s(\cdots)\|P_F^{-1} J_F - I\|$$

with $s(\cdots)$ well-behaved away from previous solutions.

Preconditioning

Theorem (F., Birkisson, Funke, 2014).

Construct a P_G such that

$$\|P_G^{-1} J_G - I\| \leq s(\cdots)\|P_F^{-1} J_F - I\|$$

with $s(\cdots)$ well-behaved away from previous solutions.

But ..

Good preconditioners don't need to control $\|P_F^{-1} J_F - I\|$.

Block-triangular factorisations

For example, if

$$J_F = \begin{bmatrix} A & B^T \\ C & 0 \end{bmatrix}$$

then

$$P_F^{-1} J_F = \begin{bmatrix} A^{-1} & 0 \\ 0 & (CA^{-1}B^T)^{-1} \end{bmatrix} \begin{bmatrix} A & B^T \\ C & 0 \end{bmatrix}$$

has three distinct eigenvalues (Murphy, Golub, Wathen, 2000).

A new bound

New theorem (F., 2015)

Suppose $P_F^{-1} J_F$ is diagonalisable. Then $P_G^{-1} J_G$ can be solved in **no more than twice as many Krylov iterations** as $P_F^{-1} J_F$.

A new bound

New theorem (F., 2015)

Suppose $P_F^{-1} J_F$ is diagonalisable. Then $P_G^{-1} J_G$ can be solved in **no more than twice as many Krylov iterations** as $P_F^{-1} J_F$.

Theorem (F., 2015)

Let A be diagonalisable and B be rank-one. Then $A + B$ has at most twice as many distinct eigenvalues as A.

A new bound

New theorem (F., 2015)

Suppose $P_F^{-1} J_F$ is diagonalisable. Then $P_G^{-1} J_G$ can be solved in **no more than twice as many Krylov iterations** as $P_F^{-1} J_F$.

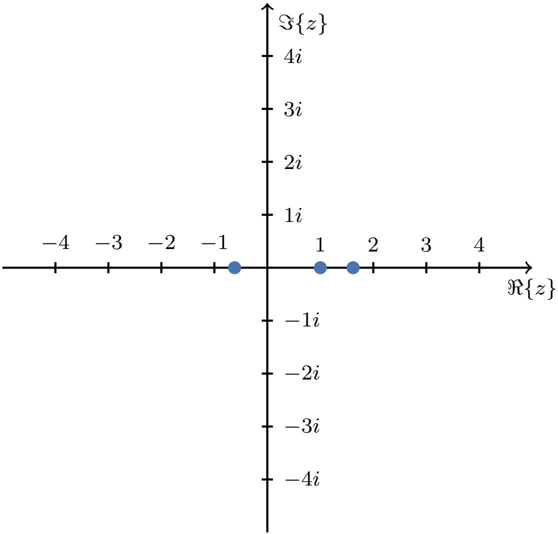
Theorem (F., 2015)

Let A be diagonalisable and B be rank-one. Then $A + B$ has at most twice as many distinct eigenvalues as A.

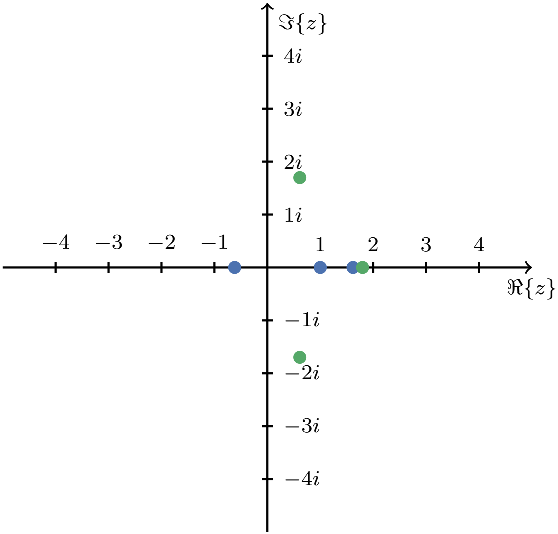
Theorem (F., 2015)

Let A be symmetric and B be nondefective rank-one. Then all but one of the eigenvalues of $A + B$ are interlaced with those of A.

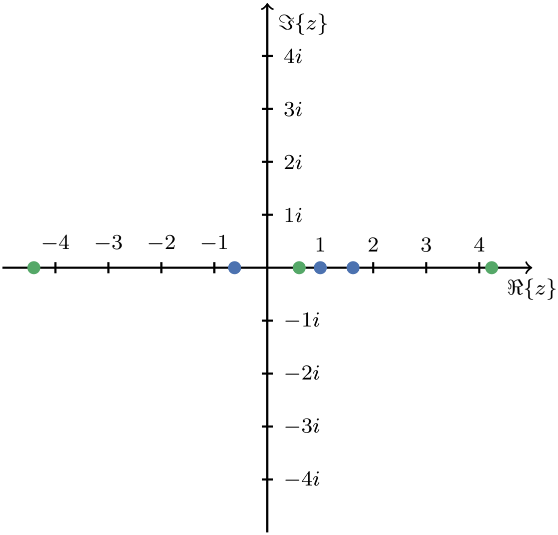
Eigenvalues after deflation



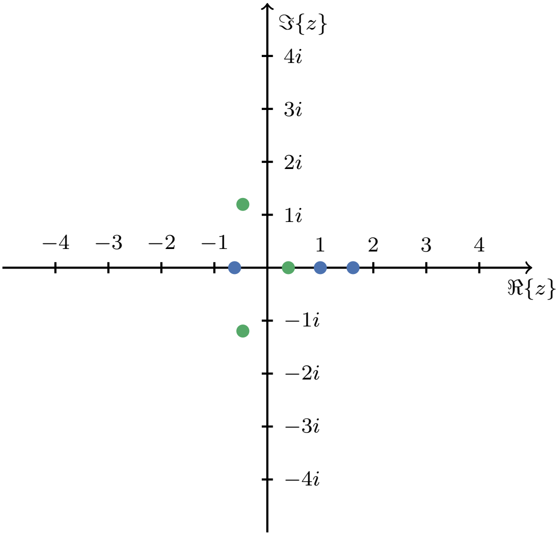
Eigenvalues after deflation



Eigenvalues after deflation



Eigenvalues after deflation



Section 4

Applications

The Yamabe problem

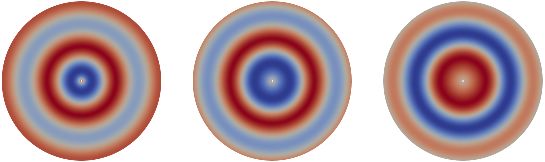
Application: differential geometry (Erway & Holst, 2011).

The Yamabe equation

$$
\begin{aligned}
-8\nabla^2 u - \frac{1}{10}u + \frac{1}{r^3}u^5 = 0 \quad &\text{in} \quad \Omega, \\
u = 1 \quad &\text{on } \partial\Omega.
\end{aligned}
$$

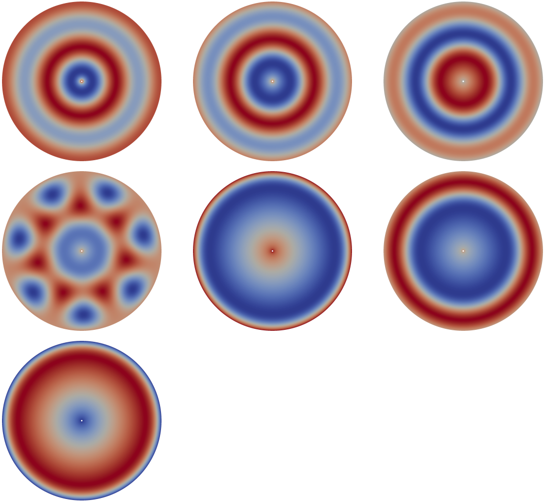
Discretisation: $\mathbb{P}_1$ finite elements.

Yamabe: solutions



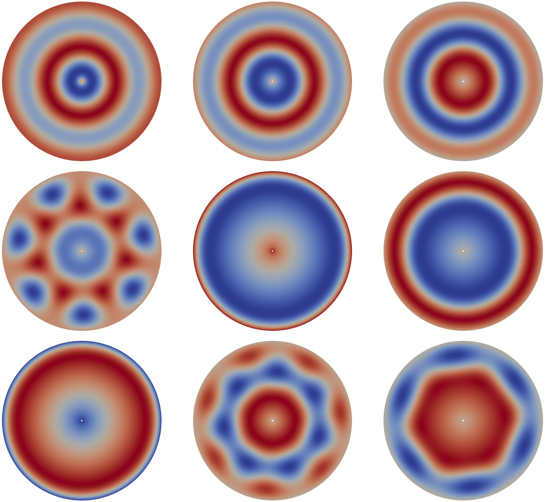
Solutions found using deflation from $u = 1$

Yamabe: solutions



Solutions found using deflation from $u = 1$

Yamabe: solutions



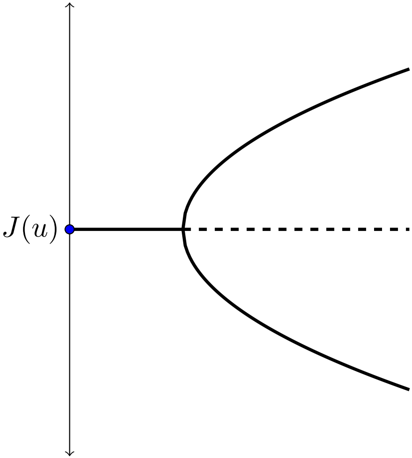
Solutions found using deflation from $u = 1$ and negation.

Yamabe: preconditioner performance

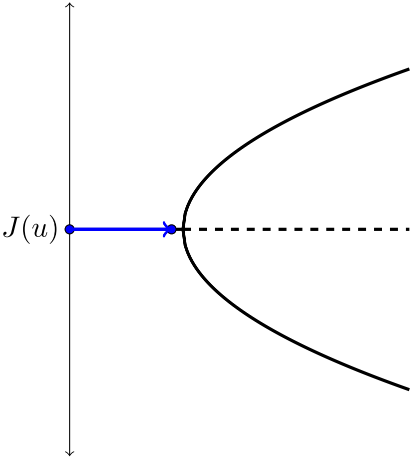
# of deflations	average Krylov iterations per solve
0	15.2
1	17.1
2	15.1
3	16.9
4	11.2
5	12.4
6	10.9
7	15.5
8	13.9

Good preconditioner performance up to $\sim$**2 billion dofs.**

Tracing bifurcation diagrams (classical)

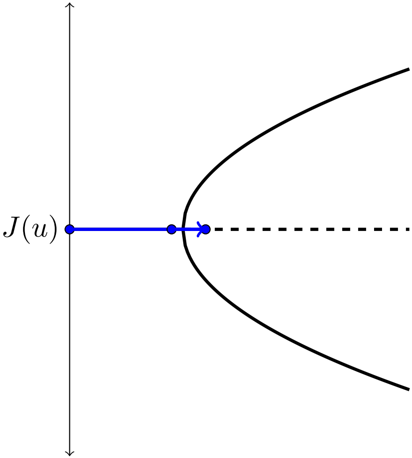


Tracing bifurcation diagrams (classical)



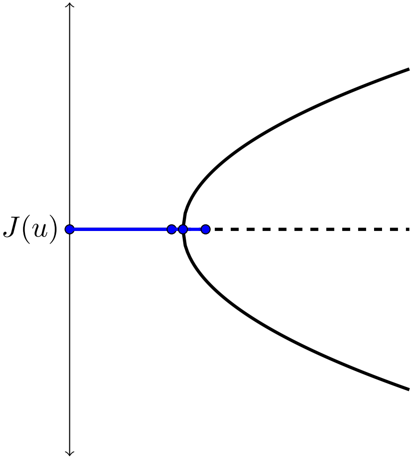
Step I: continuation

Tracing bifurcation diagrams (classical)



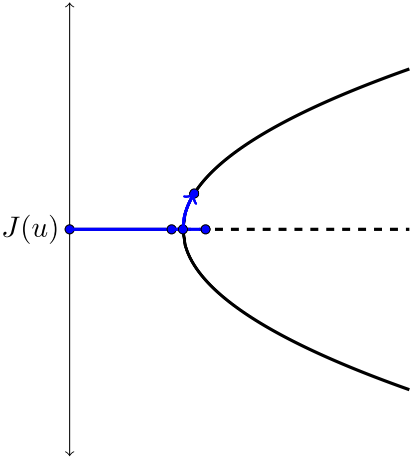
Step II: continuation

Tracing bifurcation diagrams (classical)



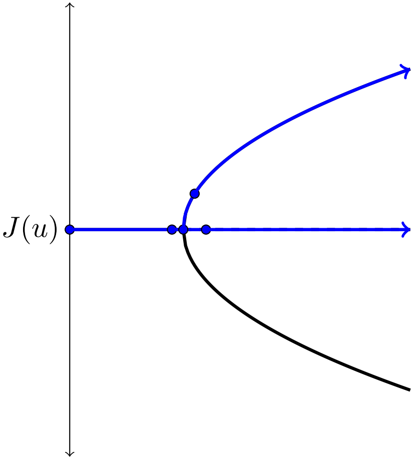
Step III: identify bifurcation point (**tricky**)

Tracing bifurcation diagrams (classical)



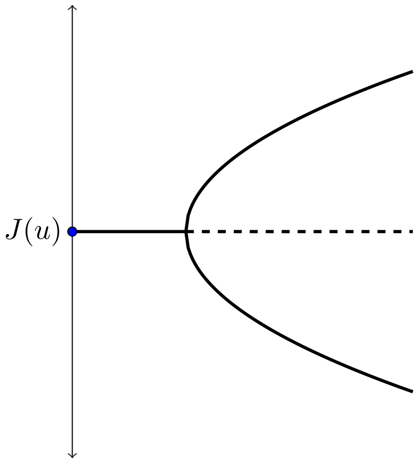
Step IV: compute eigenvectors (**expensive**) and switch (**tricky**)

Tracing bifurcation diagrams (classical)

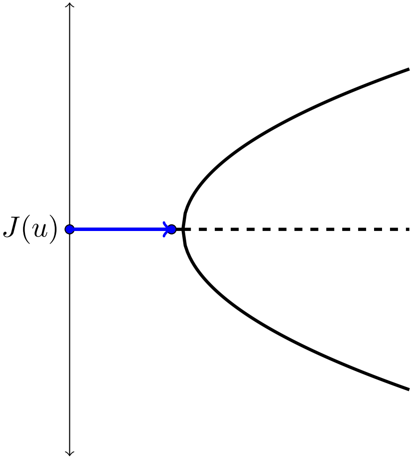


Step V: continuation on branches

Tracing bifurcation diagrams (deflation)

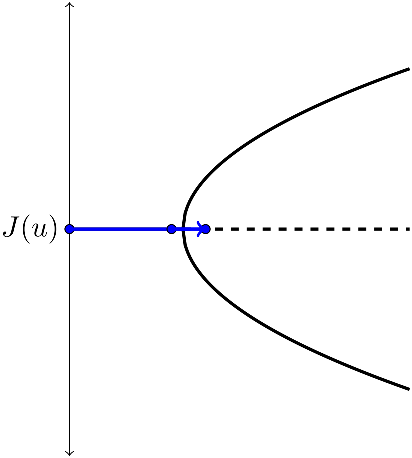


Tracing bifurcation diagrams (deflation)



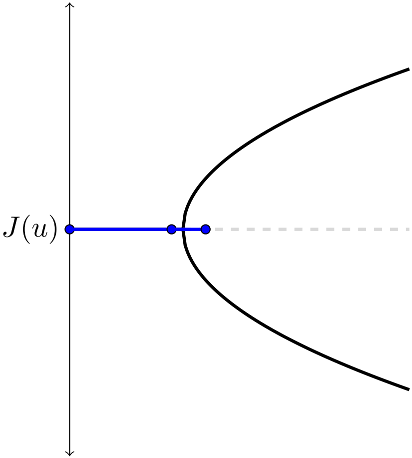
Step I: continuation

Tracing bifurcation diagrams (deflation)



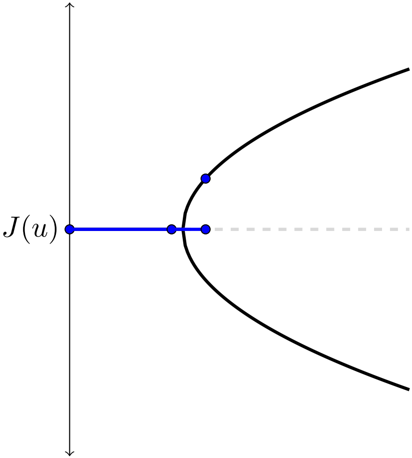
Step II: continuation

Tracing bifurcation diagrams (deflation)



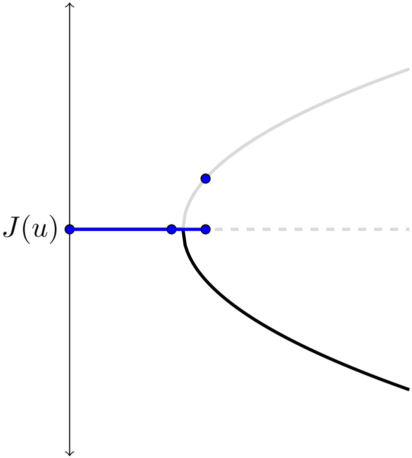
Step III: deflate

Tracing bifurcation diagrams (deflation)



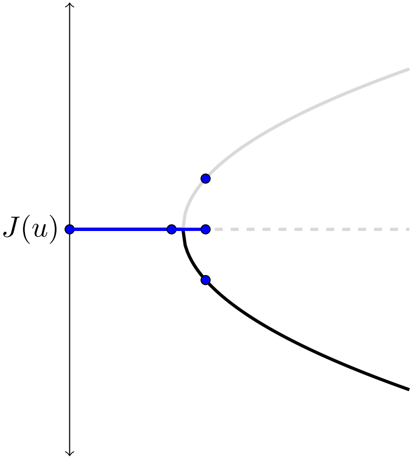
Step III+: solve deflated problem

Tracing bifurcation diagrams (deflation)



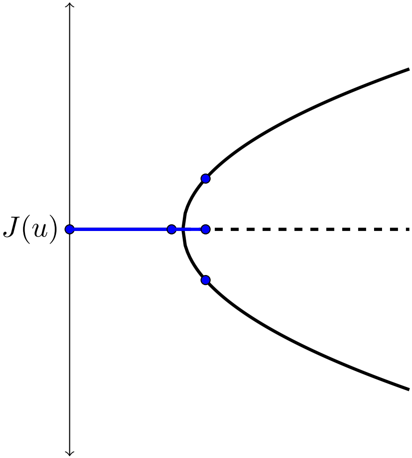
Step III: deflate

Tracing bifurcation diagrams (deflation)



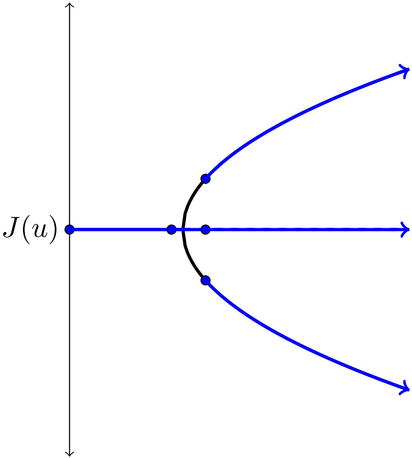
Step III+: solve deflated problem

Tracing bifurcation diagrams (deflation)



Step IV: continuation on branches

Tracing bifurcation diagrams (deflation)



Step IV: continuation on branches

Hyperelastic buckling

Application: buckling of a column under loading.

Compressible neo-Hookean hyperelasticity

Define the potential energy

$$\Pi = \int_\Omega \psi(u)\ \mathrm{d}x - \int_\Omega B \cdot u\ \mathrm{d}x - \int_{\partial\Omega} T \cdot u\ \mathrm{d}s.$$

Then

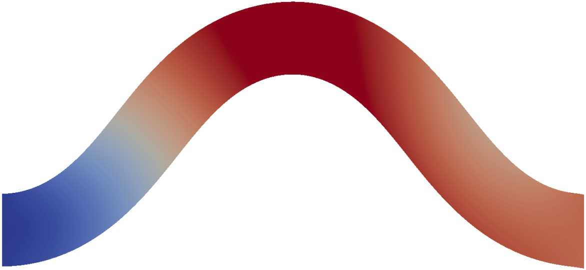
$$\begin{aligned} \Pi'(u; v) &= 0 && \forall\ v \in V, \\ u_0 &= 0 && \text{on } x = 0, \\ u_0 &= -\text{load} && \text{on } x = L, \\ u_1 &= 0 && \text{on } x = L. \end{aligned}$$

Discretisation: $[\mathbb{P}_1]^2$ finite elements.

Hyperelastic buckling: some solutions

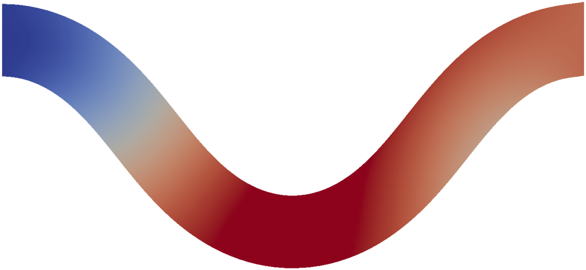
$7/13$ solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions



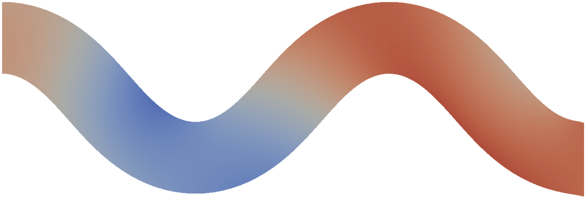
7/13 solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions



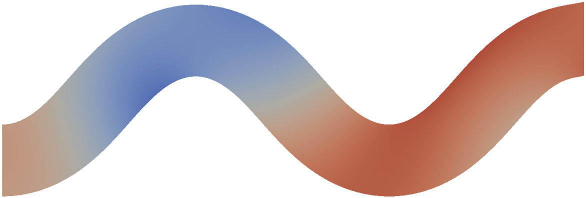
7/13 solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions



$7/13$ solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions

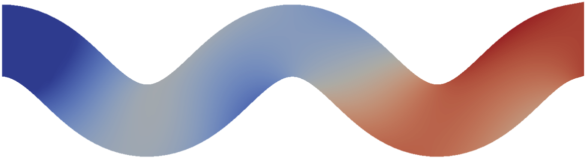


7/13 solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions

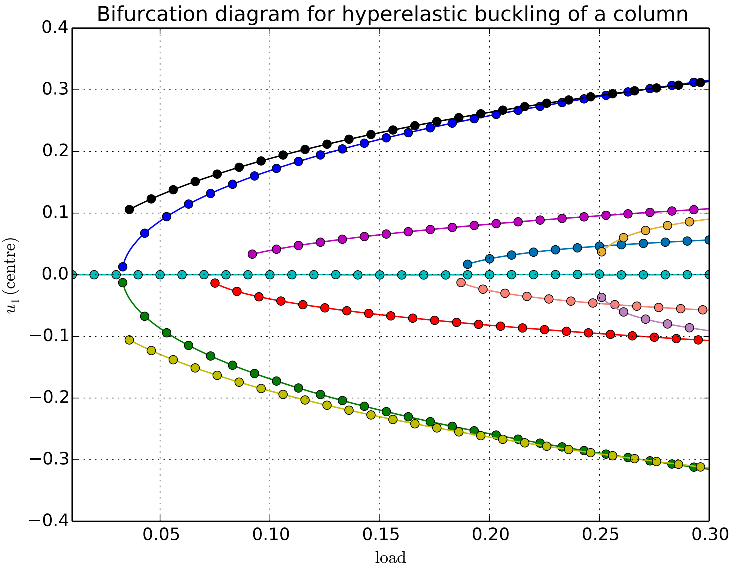
$7/13$ solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: some solutions



7/13 solutions of the problem for $\mathrm{load} = 0.3$.

Hyperelastic buckling: bifurcation diagram



Deflation vs. global optimisation

Global optimisation techniques

Computes **global minima** for problems of **small dimension** (~ 10).

Deflation vs. global optimisation

Global optimisation techniques

Computes **global minima** for problems of **small dimension** (~ 10).

Deflation + local optimisation

Computes **some minima** for problems of **arbitrary dimension**.

Equality-constrained optimisation problems

Multiple solutions of optimality conditions $\leftrightarrow$ multiple candidate optima

PDE-constrained optimisation problem

$$\begin{aligned} \underset{y \in H_0^1,\ u \in L^2}{\text{minimise}} \quad & \frac{1}{2}\int_\Omega (y - y_A)^2 + \frac{\beta}{2}\int_\Omega u^2 \\ \text{subject to} \quad & -\nabla^2 y = u \quad \text{in } \Omega. \end{aligned}$$

Equality-constrained optimisation problems

Multiple solutions of optimality conditions $\leftrightarrow$ multiple candidate optima

PDE-constrained optimisation problem

$$\begin{aligned} \underset{y \in H_0^1,\ u \in L^2}{\text{minimise}} \quad & \frac{1}{2}\int_\Omega (y - y_A)^2 (y - y_B)^2 + \frac{\beta}{2}\int_\Omega u^2 \\ \text{subject to} \quad & -\nabla^2 y = u \qquad \text{in } \Omega. \end{aligned}$$

Equality-constrained optimisation problems

Multiple solutions of optimality conditions $\leftrightarrow$ multiple candidate optima

PDE-constrained optimisation problem

$$\begin{aligned} \underset{y \in H_0^1,\ u \in L^2}{\text{minimise}} \quad & \frac{1}{2}\int_\Omega (y - y_A)^2 (y - y_B)^2 + \frac{\beta}{2}\int_\Omega u^2 \\ \text{subject to} \quad & -\nabla^2 y = u \qquad \text{in } \Omega. \end{aligned}$$

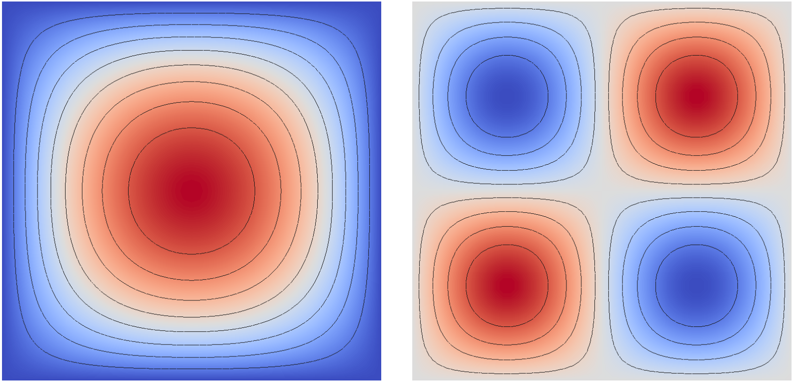
Karush–Kuhn–Tucker optimality conditions

$$\nabla \mathcal{L} = 0.$$

Discretisation: $[\mathbb{P}_1]^3$ finite elements.

Equality-constrained optimisation problems

Multiple solutions of optimality conditions $\leftrightarrow$ multiple candidate optima



2 minima of 7 stationary points, found from $(y, u, \lambda) = (0, 0, 0)$.

Complementarity problems

Complementarity problems arise with inequality constraints.

Canonical complementarity problem in $\mathbb{R}^n$

Given a residual $F : \mathbb{R}^n \to \mathbb{R}^n$, a lower bound $l \in \mathbb{R}^n_\infty$ and an upper bound $u \in \mathbb{R}^n_\infty$, find $x \in \mathbb{R}^n$ such that exactly one of the conditions

$$\begin{aligned} l_i < x_i < u_i &\text{ and } F_i(x) = 0; \\ l_i = x_i \qquad &\text{ and } F_i(x) > 0; \\ x_i = u_i &\text{ and } F_i(x) < 0; \end{aligned}$$

holds for each i.

Complementarity problems

Complementarity problems arise with inequality constraints.

Canonical complementarity problem in $\mathbb{R}^n$

Given a residual $F : \mathbb{R}^n \to \mathbb{R}^n$, a lower bound $l \in \mathbb{R}^n_\infty$ and an upper bound $u \in \mathbb{R}^n_\infty$, find $x \in \mathbb{R}^n$ such that exactly one of the conditions

$$\begin{aligned} l_i < x_i < u_i &\text{ and } F_i(x) = 0; \\ l_i = x_i \quad\quad &\text{ and } F_i(x) > 0; \\ x_i = u_i &\text{ and } F_i(x) < 0; \end{aligned}$$

holds for each i.

Theorem (F., Croci, 2015)

Deflation also applies to complementarity problems.

Topology optimisation constrained by the Stokes equations

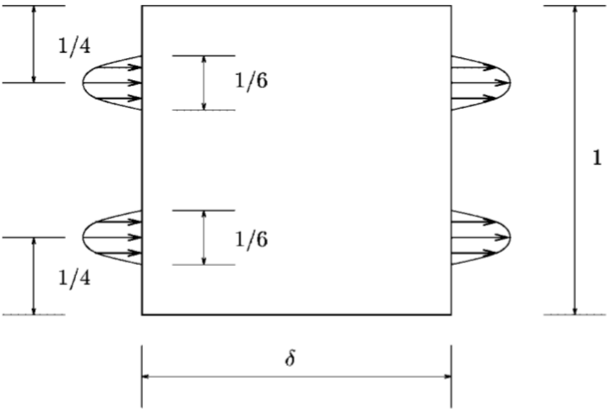


Figure 10. Design domain for the double pipe example.

What is the best pipe that connects inflow to outflow?

Stokes: governing PDE

We wish to minimise the dissipated power in the fluid

$$J = \frac{1}{2}\int_{\Omega} \alpha(\rho) u \cdot u + \frac{1}{2}\mu \int_{\Omega} \nabla u : \nabla u$$

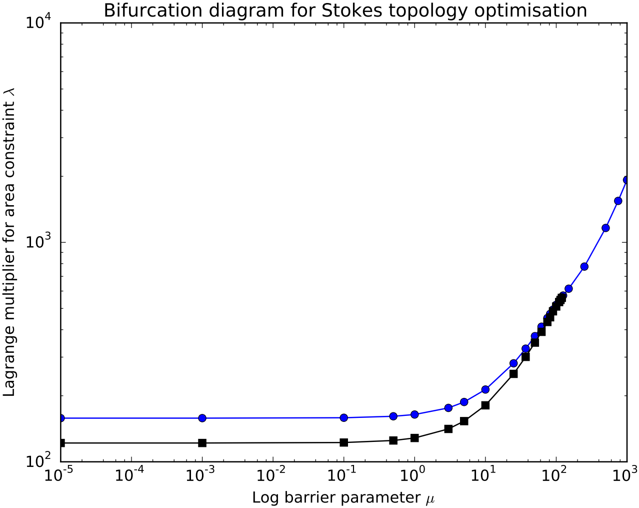
subject to the Stokes equations with a permeability term:

$$\begin{aligned} \alpha(\rho)u - \mu\nabla^2 u + \nabla p = 0 \qquad & \text{in } \Omega, \\ \nabla \cdot u = 0 \qquad & \text{in } \Omega, \\ u = b \qquad & \text{on } \delta\Omega, \\ \rho(x) \in [0,1] \quad & \text{a.e. in } \Omega, \\ \int_{\Omega} \rho \leq V. & \end{aligned}$$

Configuration and nonuniqueness: Borrvall and Petersson (2003).

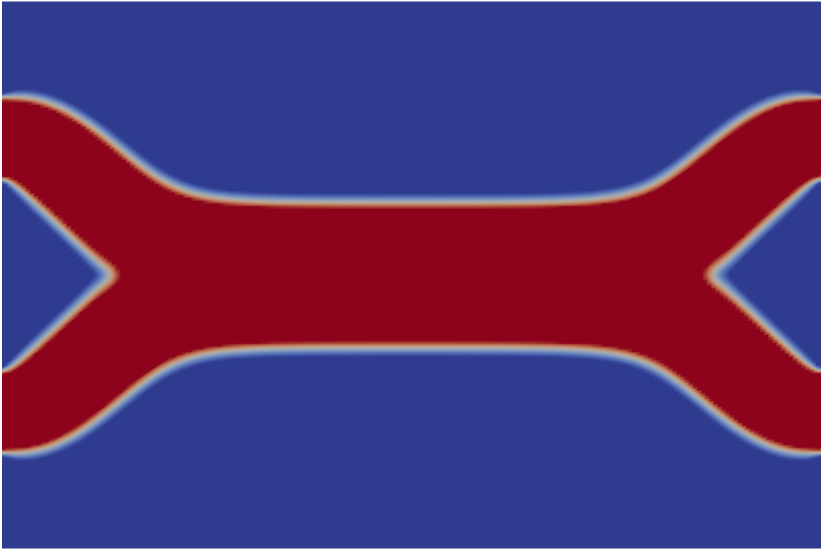
Discretisation: $[\mathbb{P}_2]^2$–$\mathbb{P}_1$.

Stokes: bifurcation diagram



Stokes: two solutions

Stokes:



Conclusions

- Multiple solutions of PDEs are **ubiquitous and important**.

Conclusions

- Multiple solutions of PDEs are ubiquitous and important.
- Deflation is a **useful technique for computing** them.

Conclusions

- Multiple solutions of PDEs are ubiquitous and important.
- Deflation is a useful technique for computing them.
- **Deflation and continuation** are natural complements.

Conclusions

- Multiple solutions of PDEs are ubiquitous and important.
- Deflation is a useful technique for computing them.
- Deflation and continuation are natural complements.
- There are **interesting applications** in:
  - nonlinear PDEs,
  - tracing bifurcation diagrams,
  - multimodal Bayesian inference,
  - and large-scale optimisation with constraints.