Challenges for CLAPDE from Optimization:

A Personal View

Nick Gould (RAL)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c_{\mathcal{E}}(x) = 0$$

CLAPDE, University of Durham, July 2008

Generic quadratic optimization problem from CLAPDE

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) = 0 \quad \longleftarrow \textbf{discrete PDE}$$

Generic quadratic optimization problem from CLAPDE

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) = 0 \quad \longleftarrow$$ **discrete PDE**

$\Longrightarrow$ **EQP step subproblem** s_k from x_k

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} s^T H_k s + s^T g_k \quad \text{subject to} \quad J_k s + c_k = 0$$

$\longleftarrow$ **linearized PDE**

Generic quadratic optimization problem from CLAPDE

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) = 0 \quad \longleftarrow \textbf{discrete PDE}$$

$\Longrightarrow$ **EQP step subproblem** s_k from x_k

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2}s^T H_k s + s^T g_k \quad \text{subject to} \quad J_k s + c_k = 0$$

$\longleftarrow$ **linearized PDE**

- J_k **Jacobian** of constraints
- H_k symmetric but indefinite $\approx \nabla_{xx}\ell(x,y)$ **Hessian of Lagrangian**

Generic quadratic optimization problem from CLAPDE

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) = 0 \quad \longleftarrow$$ **discrete PDE**

$\Longrightarrow$ **EQP step subproblem** s_k from x_k

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} s^T H_k s + s^T g_k \quad \text{subject to} \quad J_k s + c_k = 0$$

$\longleftarrow$ **linearized PDE**

- J_k **Jacobian** of constraints
- H_k symmetric but indefinite $\approx \nabla_{xx} \ell(x, y)$ **Hessian of Lagrangian**

NB. If the PDE is nonlinear, this will influence H_k!

EQP is not equivalent to a saddle point system

$$s_k = \mathbf{arg} \underset{s \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2} s^T H_k s + s^T g_k \ \text{ subject to } \ J_k s + c_k = 0$$

$\not\Longrightarrow$ **saddle-point solution**

$$\begin{pmatrix} H_k & J_k^T \\ J_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ y_k \end{pmatrix} = - \begin{pmatrix} g_k \\ c_k \end{pmatrix}$$

unless $(\sim)$ $s^T H_k s > 0$ for all nonzero $s : J_k s = 0$ c.f., 2nd-order opt

EQP is not equivalent to a saddle point system

$$s_k = \arg \underset{s \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2} s^T H_k s + s^T g_k \ \text{ subject to } \ J_k s + c_k = 0$$

$\not\Longrightarrow$ **saddle-point solution**

$$\begin{pmatrix} H_k & J_k^T \\ J_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ y_k \end{pmatrix} = - \begin{pmatrix} g_k \\ c_k \end{pmatrix}$$

unless $(\sim)$ $s^T H_k s > 0$ for all nonzero $s : J_k s = 0$ c.f., 2nd-order opt

Tasks

- find iterative methods which can identify this situation

EQP is not equivalent to a saddle point system

$$s_k = \arg \underset{s \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} s^T H_k s + s^T g_k \ \text{ subject to } \ J_k s + c_k = 0$$

$\not\Longrightarrow$ **saddle-point solution**

$$\begin{pmatrix} H_k & J_k^T \\ J_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ y_k \end{pmatrix} = - \begin{pmatrix} g_k \\ c_k \end{pmatrix}$$

unless $(\sim)$ $s^T H_k s > 0$ for all nonzero $s : J_k s = 0$ c.f., 2nd-order opt

Tasks

- find iterative methods which can identify this situation
- if violated, want instead normalized s_k to minimize $s^T H_k s : J_k s = 0$, e.g., appropriate eigenvector

EQP is not equivalent to a saddle point system

$$s_k = \arg \underset{s \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2} s^T H_k s + s^T g_k \ \text{ subject to } \ J_k s + c_k = 0$$

$\not\Longrightarrow$ **saddle-point solution**

$$\begin{pmatrix} H_k & J_k^T \\ J_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ y_k \end{pmatrix} = - \begin{pmatrix} g_k \\ c_k \end{pmatrix}$$

unless $(\sim)$ $s^T H_k s > 0$ for all nonzero $s : J_k s = 0$ c.f., 2nd-order opt

Tasks

- find iterative methods which can identify this situation
- if violated, want instead normalized s_k to minimize $s^T H_k s : J_k s = 0$, e.g., appropriate eigenvector
- if satisfied, eigenvalue bounds?

EQP is not equivalent to a saddle point system

$$s_k = \arg \underset{s \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2} s^T H_k s + s^T g_k \ \text{ subject to } \ J_k s + c_k = 0$$

$\not\Longrightarrow$ **saddle-point solution**

$$\begin{pmatrix} H_k & J_k^T \\ J_k & 0 \end{pmatrix} \begin{pmatrix} s_k \\ y_k \end{pmatrix} = - \begin{pmatrix} g_k \\ c_k \end{pmatrix}$$

unless $(\sim)$ $s^T H_k s > 0$ for all nonzero $s : J_k s = 0$ c.f., 2nd-order opt

Tasks

- find iterative methods which can identify this situation
- if violated, want instead normalized s_k to minimize $s^T H_k s : J_k s = 0$, e.g., appropriate eigenvector
- if satisfied, eigenvalue bounds?

OK for constraint-preconditioned CG, but what else??

Residual reduction is not enough

- would like reduction in

$$(*) \qquad \|J_k s + c_k\| \quad \text{and/or} \quad \tfrac{1}{2} s^T H_k s + s^T g_k$$

Residual reduction is not enough

- would like reduction in

  $(*) \qquad \|J_k s + c_k\|$ and/or $\frac{1}{2} s^T H_k s + s^T g_k$

- minimum residual like-methods (MINRES, GMRES, QMR, . . .) aim for reduction in

$$\left\| \begin{pmatrix} H_k s + g_k + J_k^T y \\ J_k s + c_k \end{pmatrix} \right\|$$

Residual reduction is not enough

- would like reduction in

$$(*) \qquad \|J_k s + c_k\| \quad \text{and/or} \quad \tfrac{1}{2} s^T H_k s + s^T g_k$$

- minimum residual like-methods (MINRES, GMRES, QMR, . . .) aim for reduction in

$$\left\| \begin{pmatrix} H_k s + g_k + J_k^T y \\ J_k s + c_k \end{pmatrix} \right\|$$

- may not reduce $(*)$ for many iterations

Residual reduction is not enough

- would like reduction in

  $(*) \qquad \|J_k s + c_k\|$ and/or $\frac{1}{2} s^T H_k s + s^T g_k$

- minimum residual like-methods (MINRES, GMRES, QMR, . . .) aim for reduction in

$$\left\| \begin{pmatrix} H_k s + g_k + J_k^T y \\ J_k s + c_k \end{pmatrix} \right\|$$

- may not reduce $(*)$ for many iterations

- are there iterative methods which can ensure $(*)$ every iteration? Every pair of iterations??

Three basic computational components

Find $s_k = n_k + t_k$ where

- $J_k n_k + c_k \approx 0$
- $J_k^T y_k + g_k \approx 0$
- (approx min) $\frac{1}{2} t_k^T H_k t_k + t_k^T g_k : J_k t_k \approx 0$

Three basic computational components

Find $s_k = n_k + t_k$ where

- $J_k n_k + c_k \approx 0$
- $J_k^T y_k + g_k \approx 0$
- (approx min) $\frac{1}{2} t_k^T H_k t_k + t_k^T g_k : J_k t_k \approx 0$
- all need to be efficient and matrix free
- may need to "regularise" (trust-region/cubic regularisation??)
- can embed within globally convergent "funnel" framework

Three basic computational components

Find $s_k = n_k + t_k$ where

- $J_k n_k + c_k \approx 0$
- $J_k^T y_k + g_k \approx 0$
- (approx min) $\frac{1}{2} t_k^T H_k t_k + t_k^T g_k : J_k t_k \approx 0$
- all need to be efficient and matrix free
- may need to “regularise” (trust-region/cubic regularisation??)
- can embed within globally convergent “funnel” framework

C.f.— linesearch-based methods based on iterative saddle-point solution which (to my knowledge) use heuristic perturbations to H_k

Three basic computational components

Find $s_k = n_k + t_k$ where

- $J_k n_k + c_k \approx 0$
- $J_k^T y_k + g_k \approx 0$
- (approx min) $\frac{1}{2} t_k^T H_k t_k + t_k^T g_k : J_k t_k \approx 0$
- all need to be efficient and matrix free
- may need to “regularise” (trust-region/cubic regularisation??)
- can embed within globally convergent “funnel” framework

C.f.— linesearch-based methods based on iterative saddle-point solution which (to my knowledge) use heuristic perturbations to H_k

- how do non-trivial perturbations affect the excellent PDE-based preconditioners?

Multilevel methods

It is unclear how best to use multigrid in the PDE-optimization context

- apply linear multigrid to the EQP subproblem
- apply nonlinear multigrid/multilevel ideas
  - geometric (Toint, Gratton, Sartenaer, Mouffe, . . .)
  - algebraic

Auxiliary constraints

If there additional non-PDE side constraints on (e.g.) controls:

- extra equations
- simple bounds on variables
- general inequalities
- integer restrictions

how can we impose these without destroying PDE-specific structure (e.g.) preconditioners?

“Big” questions

- Krylov-based methods treat

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{J}^T \\ \boldsymbol{J} & \boldsymbol{0} \end{pmatrix}$$

as a generic matrix/operator
  - are there new methods which really exploit the zero block?
  - are there new methods which really exploit the substructure?

“Big” questions

- Krylov-based methods treat

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{J}^{\boldsymbol{T}} \\ \boldsymbol{J} & \mathbf{0} \end{pmatrix}$$

  as a generic matrix/operator
  - are there new methods which really exploit the zero block?
  - are there new methods which really exploit the substructure?

- Krylov-based methods obtain products

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{J}^{\boldsymbol{T}} \\ \boldsymbol{J} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix},$$

  i.e., $\boldsymbol{H}\boldsymbol{u}$, $\boldsymbol{J}\boldsymbol{u}$ and $\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{v}$
  - are there better methods without such strong ties, e.g., $\boldsymbol{H}\boldsymbol{u}$, $\boldsymbol{J}\boldsymbol{w}$ and $\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{v}$?

Over to you . . .

Thanks to Alison, Andy and David!