Alternating approach for solving design optimization problems

A. Hamdi , A. Griewank , N. Gauger, E. Özkaya

Department of Mathematics
Humboldt Universitat zu Berlin

March 2008

Plan :

Plan :

1. Motivation
2. Problem statement
3. One-Shot approach
   - Doubly augmented Lagrangian
   - Suitable preconditionner B
   - Current choice of α and β
   - Low Rank update for B
4. Numerical tests

Plan :

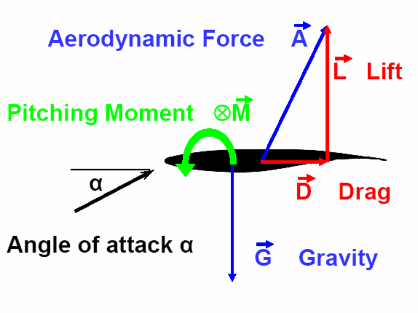
1. Motivation
2. Problem statement
3. One-Shot approach
   - Doubly augmented Lagrangian
   - Suitable preconditionner B
   - Current choice of α and β
   - Low Rank update for B
4. Numerical tests

Plan :

1. Motivation
2. Problem statement
3. One-Shot approach
   - Doubly augmented Lagrangian
   - Suitable preconditionner B
   - Current choice of α and β
   - Low Rank update for B
4. Numerical tests

Plan :

Motivation: Shape optimization in aerodynamics



Notations:

- u denotes a parametrisation of the wing shape.
- y denotes the state vector. Given u + PDE $\Longrightarrow y$.

Plan :

Problem statement:

Via the discretisation of PDE, we get the constraint $c(y,u)=0$.

Aim:

$$(P) \qquad \min_{y,u} f(y,u) \quad s.t \quad c(y,u)=0.$$

Assumptions:

- $\frac{\partial c}{\partial y}$ always invertible. IFT $\Longrightarrow$ given u, $\exists! y$ s.t $c(y,u)=0$.
- The user supplies $y_{k+1}=G(y_k,u)$. $G(y,u)=y \Leftrightarrow c(y,u)=0$.

$$\|G_y(y,u)\| = \|G_y^\top(y,u)\| \leq \rho < 1$$

$$\Downarrow$$

$$\|G(y_1,u)-G(y_2,u)\| \leq \rho \|y_1-y_2\|.$$

- Regularity: $G,f \in C^{2,1}$.

Plan :

One-Shot approach:

Lagrangian: $L(y,\bar{y},u) = f(y,u) + \bar{y}^\top (G(y,u) - y).$

$$\triangleright N(y,\bar{y},u) := f(y,u) + \bar{y}^\top G(y,u) \Longrightarrow L(y,\bar{y},u) = N(y,\bar{y},u) - \bar{y}^\top y.$$

First order necessary condition:

$$\begin{cases} \frac{\partial L}{\partial \bar{y}} = G(y^*,u^*) - y^* = 0 \\ \frac{\partial L}{\partial y} = N_y(y^*,\bar{y}^*,u^*)^\top - \bar{y}^* = 0 \\ \frac{\partial L}{\partial u} = N_u(y^*,\bar{y}^*,u^*)^\top = 0 \end{cases}$$

A one-shot step:

$$(OS) \begin{cases} y_{k+1} = G(y_k,u_k) & \text{primal feasibility} \\ \bar{y}_{k+1} = N_y(y_k,\bar{y}_k,u_k)^\top & \text{dual feasibility} \\ u_{k+1} = u_k - B_k^{-1} N_u(y_k,\bar{y}_k,u_k)^\top & \text{optimality} \end{cases}$$

Challenges and Questions:

General:

- Definition and computation of B for good local and global convergence.
- Ensuring consistent progress w.r.t a merit function.
- Taking into account extra cost caused by design update.
- Choice of Norms and Preconditionning.

Focus of this talk:

- ▷ Merit function L^a of doubly augmented Lagrangian type.
- ▷ Definition of B as a multiple of the Hessian of L^a w.r.t design.
- ▷ Optimal choice of weighting coefficients.
- ▷ Approximation of B using BFGS $\Longrightarrow$ Alternating approach.
- ▷ Criteria for choice between feasibility and design steps.

Plan :

Doubly augmented Lagrangian:

Doubly augmented Lagrangian:

Descent on a merit function of doubly augmented Lagrangian type function:

$$L^a(y,\bar{y},u) = \frac{\alpha}{2}\|G(y,u)-y\|^2 + \frac{\beta}{2}\|N_y(y,\bar{y},u)^\top - \bar{y}\|^2 + N(y,\bar{y},u) - \bar{y}^\top y$$

Gradient of L^a: Let $\Delta G_y = I - G_y$.

$$\begin{bmatrix} \nabla_y L^a \\ \nabla_{\bar{y}} L^a \\ \nabla_u L^a \end{bmatrix} = \begin{bmatrix} -\alpha \Delta G_y^\top (G(y,u)-y) + (I+\beta N_{yy})(N_y(y,\bar{y},u)^\top - \bar{y}) \\ G(y,u) - y - \beta \Delta G_y (N_y(y,\bar{y},u)^\top - \bar{y}) \\ \alpha G_u^\top (G(y,u)-y) + \beta N_{yu}{}^\top (N_y(y,\bar{y},u)^\top - \bar{y}) + N_u(y,\bar{y},u)^\top \end{bmatrix}.$$

$\Longrightarrow$ we can use SOA of ADOL-C to compute adjoint derivatives

Aim: Derive conditions on α and β s.t L^a is an exact penalty function.

Properties of One-Shot steps:

Properties of One-Shot steps:

$$\triangleright \text{ Let : } s(y,\bar{y},u) = \begin{bmatrix} G(y,u) - y \\ N_y(y,\bar{y},u)^\top - \bar{y} \\ -B^{-1} N_u(y,\bar{y},u)^\top \end{bmatrix} \implies s = 0 \Leftrightarrow \nabla L = 0.$$

Proposition: gradient of the doubly augmented Lagrangian L^a is

$$\begin{bmatrix} \nabla_y L^a \\ \nabla_{\bar{y}} L^a \\ \nabla_u L^a \end{bmatrix} = -M\, s(y,\bar{y},u), \quad \text{where} \quad M = \begin{bmatrix} \alpha \Delta G_y^\top & -I - \beta N_{yy} & 0 \\ -I & \beta \Delta G_y & 0 \\ -\alpha G_u^\top & -\beta N_{uy} & B \end{bmatrix}.$$

$\implies$ If M is nonsingular, then $\nabla L^a = 0 \Leftrightarrow s = 0$.

Correspondence condition:

Corollary: There is a 1-1 correspondence between the stationary points of L^a and the roots of s if

$$det[\alpha\beta\Delta G_y^\top \Delta G_y - I - \beta N_{yy}] \neq 0$$

for which it is sufficient that $\alpha\beta(1-\rho)^2 > 1 + \beta\|N_{yy}\|$.

▷Consequence: Let $\theta_1 = \Delta G_y^{-\top}(N_{yy}\Delta G_y^{-1}G_u + N_{yu})$,

$$T = \begin{bmatrix} I & 0 & \theta_2 \\ \Delta G_y^{-\top}(N_{yy} + I/\beta) & I & \theta_1 \\ 0 & 0 & I \end{bmatrix}, \quad \text{and} \quad , \theta_2 = \Delta G_y^{-1} G_u .$$

$$\Longrightarrow T^\top \nabla^2 L^a T = diag[\alpha \Delta G_y^\top \Delta G_y - N_{yy} - I/\beta, \beta \Delta G_y \Delta G_y^\top, H]$$

▷$\nabla^2 L^a \succ 0 \Leftrightarrow$ the Projected Hessian H of (P) is s.t $H \succ 0$.

Plan :

Descent approach:

Theorem: $\Delta\bar{G}_y = \frac{1}{2}(\Delta G_y + \Delta G_y^\top)$, s yields descent of L^a for $B \succ\succ 0$ iff

$$\alpha\beta\Delta\bar{G}_y > (I + \frac{\beta}{2}N_{yy})(\Delta\bar{G}_y)^{-1}(I + \frac{\beta}{2}N_{yy}),$$

which is implied by $\sqrt{\alpha\beta}(1-\rho) > 1 + \frac{\beta}{2}\|N_{yy}\|$.

- Choice of B: $u = B^{-\frac{1}{2}}\tilde{u}$ where $B = B^{\frac{\top}{2}}B^{\frac{1}{2}} \implies \nabla\tilde{L}^a = -\tilde{M}\,\tilde{s}$

$$\text{where} \quad \tilde{M} = B_{12}^\top M\, B_{12},\; B_{12} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & B^{-\frac{1}{2}} \end{bmatrix} \text{ and } \tilde{s} = \begin{bmatrix} G - y \\ N_y^\top - \bar{y} \\ -N_{\tilde{u}}^\top \end{bmatrix}.$$

▷ Proposition: $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^\top \tilde{M} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \geq \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix}^\top D \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix}, \forall v_1, v_2 \in I\!R^n, v_3 \in I\!R^m.$

Search of B:

▷ Proposition: Let $\theta = \|N_{yy}\|$. $D \succ 0$ if

$$\sigma := 1 - \rho - \frac{(1+\frac{\theta}{2}\beta)^2}{\alpha\beta(1-\rho)} \geq (\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|)^2.$$

- $$(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\|_2 + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|_2)^2 \leq \left\| \begin{matrix} \sqrt{\alpha}G_{\tilde{u}} \\ \sqrt{\beta}N_{y\tilde{u}} \end{matrix} \right\|_2^2 = \left\| \begin{pmatrix} \sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu} \end{pmatrix} B^{-\frac{1}{2}} \right\|_2^2$$

$$\begin{pmatrix} \sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu} \end{pmatrix} = QR. \quad \left\| \begin{pmatrix} \sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu} \end{pmatrix} B^{-\frac{1}{2}} \right\|_2^2 = \|RB^{-\frac{1}{2}}\|_2^2 = \|RB^{-1}R^\top\|_2 = \sigma.$$

$$\Longrightarrow RB^{-1}R^\top = \sigma I \quad \Longrightarrow B = \frac{1}{\sigma}R^\top R.$$

▷ Lemma: Descent if $B = \frac{1}{\sigma}(\alpha G_u^\top G_u + \beta {N_{yu}}^\top N_{yu})$.

Plan :

Choice of α and β:

Current choice of α and β:

$$(\alpha,\beta) = \min_{\alpha,\beta} \frac{\alpha\|G_u\|_2^2 + \beta\|N_{yu}\|_2^2}{1-\rho-\frac{(1+\frac{\theta}{2}\beta)^2}{\alpha\beta(1-\rho)}} \geq \|B\|_2.$$

▷Result: If $B = B^\top \succ \frac{1}{\sigma}(\alpha G_u^\top G_u + \beta {N_{yu}}^\top N_{yu})$, then B yields descent on the doubly augmented Lagrangian L^a.

Current choice of B: Assuming that $N_{uu} \succeq 0$, we use

$$B = \frac{1}{\sigma}(\alpha G_u^\top G_u + \beta {N_{yu}}^\top N_{yu} + N_{uu}) \approx \frac{1}{\sigma}\nabla^2_{uu} L^a.$$

▷ the later approximation is exact at primal and dual feasibility.

Plan :

BFGS update for B:

Secant equation:

$$\begin{aligned} B\Delta u &= \frac{1}{\sigma}\nabla_{uu}L^a(y,\bar{y},u)\Delta u \\ &\approx \frac{1}{\sigma}(\nabla_u L^a(y,\bar{y},u+\Delta u) - \nabla_u L^a(y,\bar{y},u)). \end{aligned}$$

▷ Alternating Approach:

$$\begin{cases} \tau\Delta u^\top \nabla_u L^a < \Delta y^\top \nabla_y L^a + \Delta\bar{y}^\top \nabla_{\bar{y}} L^a < 0, & \text{pure design step} \\ \text{otherwise, do a pure feasibility step} \end{cases}$$

where $\tau \in]0,1]$.

Plan :

Numerical tests

Schulz/Bratu problem: $\min_{y,u} f(y,u)$ where

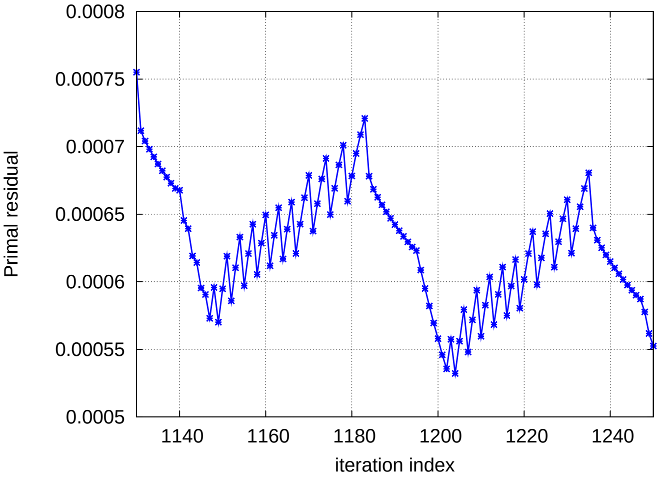
$$f(y,u) = \int_0^1 \left[\frac{\partial y}{x_2}(x_1,x_2)\Big|_{x_2=1} - 4 - \cos(2\pi x_1) \right]^2 + \kappa \left[u(x_1)^2 + u'(x_1)^2 \right] dx_1$$

under the constraints

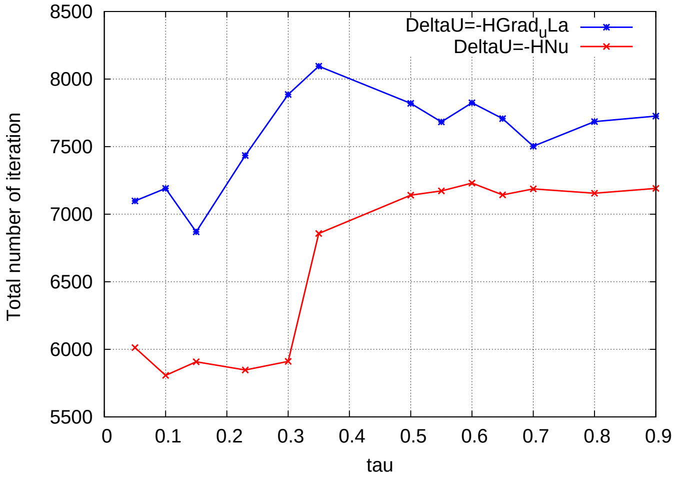
$$\begin{cases} \Delta y(x) + e^{y(x)} = 0 & x = (x_1,x_2) \in [0,1]^2 \\ y(0,x_2) = y(1,x_2) & x_2 \in [0,1] \\ y(x_1,0) = \sin(2\pi x_1) & x_1 \in [0,1] \\ y(x_1,1) = u(x_1) & x_1 \in [0,1] \end{cases}$$

Numerical tests: $h = 1/12.0, \theta = 1.0, \rho = 0.95$ and $\kappa = 1E-03$.

Primal residual behavior



Alternating Approach: Total number of iteration



One-Step One-Shot Method

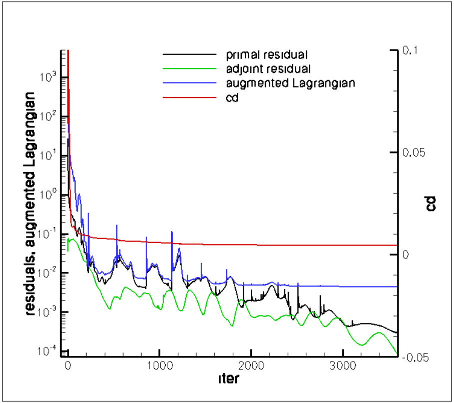


Figure: Drag (cd) reduction for 2D Euler case

RAE 2822 airfoil, $M = 0.73$, $\alpha = 2.0$. Mesh: 161×33 cells. 20 design variables (Hicks-Henne).

End

Thank you for your attention

Remarks:

- **Ellipsoid norm**: $A \in \mathbb{K}^{n,n}$, $\rho(A) = \max_{1 \le i \le n} \{|\lambda_i(A)|\}$.

  Then $\forall \varepsilon > 0, \exists W \in \mathbb{R}^{n,n}, W \succ 0$ s.t $|||v||| = \sqrt{v^\top W v}$ and
  $$|||A||| = \sup_{v \neq 0} \frac{|||Av|||}{|||v|||} \le \rho(A) + \varepsilon.$$
  Example

  ▷ $\rho(J^*) < 1 \implies$ the One-Shot iterations converge.

  $\Downarrow$

  Aim: Find B s.t $\frac{1-\rho}{1-\rho(J^*)}$ bounded (retardation).

- **Perturbation lemma**: $A \in \mathbb{K}^{n,n}$ nonsingular, δA perturbation of A s.t $\|A^{-1}\|\|\delta A\| < 1$. Then, $A + \delta A$ is nonsingular and
  $$\|(A+\delta A)^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\delta A\|}.$$

  ▷ $\Delta G_y = I - G_y$ is an invertible matrix and $\|\Delta G_y\| \ge 1 - \rho$.

Remarks:

- **Ellipsoid norm**: $A \in \mathbb{K}^{n,n}$, $\rho(A) = \max_{1 \le i \le n} \{|\lambda_i(A)|\}$.

  Then $\forall \varepsilon > 0, \exists W \in \mathbb{R}^{n,n}, W \succ 0$ s.t $|||v||| = \sqrt{v^\top W v}$ and

  $$|||A||| = \sup_{v \neq 0} \frac{|||Av|||}{|||v|||} \le \rho(A) + \varepsilon.$$

  Example

  ▷ $\rho(J^*) < 1 \implies$ the One-Shot iterations converge.

  $\Downarrow$

  Aim: Find B s.t $\frac{1-\rho}{1-\rho(J^*)}$ bounded (retardation).

- **Perturbation lemma**: $A \in \mathbb{K}^{n,n}$ nonsingular, δA perturbation of A s.t $\|A^{-1}\| \|\delta A\| < 1$. Then, $A + \delta A$ is nonsingular and

  $$\|(A + \delta A)^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|\delta A\|}.$$

  ▷ $\Delta G_y = I - G_y$ is an invertible matrix and $\|\Delta G_y\| \ge 1 - \rho$.

Remarks:

- **Ellipsoid norm**: $A \in \mathbb{K}^{n,n},\ \rho(A) = \max_{1 \le i \le n} \{|\lambda_i(A)|\}$.

  Then $\forall \varepsilon > 0, \exists W \in \mathbb{R}^{n,n}, W \succ 0$ s.t $|||v||| = \sqrt{v^\top W v}$ and
  $$|||A||| = \sup_{v \neq 0} \frac{|||Av|||}{|||v|||} \le \rho(A) + \varepsilon.$$
  Example

  ▷ $\rho(J^*) < 1 \implies$ the One-Shot iterations converge.

  $\Downarrow$

  Aim: Find B s.t $\frac{1-\rho}{1-\rho(J^*)}$ bounded (retardation).

- **Perturbation lemma**: $A \in \mathbb{K}^{n,n}$ nonsingular, δA perturbation of A s.t $\|A^{-1}\|\|\delta A\| < 1$. Then, $A + \delta A$ is nonsingular and
  $$\|(A+\delta A)^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\delta A\|}.$$

  ▷ $\Delta G_y = I - G_y$ is an invertible matrix and $\|\Delta G_y\| \ge 1 - \rho$.

Remarks:

- **Ellipsoid norm**: $A \in \mathbb{K}^{n,n}$, $\rho(A) = \max_{1 \le i \le n} \{|\lambda_i(A)|\}$.

  Then $\forall \varepsilon > 0, \exists W \in \mathbb{R}^{n,n}, W \succ 0$ s.t $|||v||| = \sqrt{v^\top W v}$ and

  $$|||A||| = \sup_{v \neq 0} \frac{|||Av|||}{|||v|||} \le \rho(A) + \varepsilon.$$

  Example

  ▷ $\rho(J^*) < 1 \implies$ the One-Shot iterations converge.

  $\Downarrow$

  Aim: Find B s.t $\frac{1-\rho}{1-\rho(J^*)}$ bounded (retardation).

- **Perturbation lemma**: $A \in \mathbb{K}^{n,n}$ nonsingular, δA perturbation of A s.t $\|A^{-1}\| \|\delta A\| < 1$. Then, $A + \delta A$ is nonsingular and

  $$\|(A + \delta A)^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|\delta A\|}.$$

  ▷ $\Delta G_y = I - G_y$ is an invertible matrix and $\|\Delta G_y\| \ge 1 - \rho$.

Matrices $\tilde{M}$ and D:

The matrix $\tilde{M}$:

$$\tilde{M} = \begin{bmatrix} \alpha \Delta \bar{G}_y & -I - \frac{\beta}{2} N_{yy} & -\frac{\alpha}{2} G_{\tilde{u}} \\ -I - \frac{\beta}{2} N_{yy} & \beta \Delta \bar{G}_y & -\frac{\beta}{2} N_{y\tilde{u}} \\ -\frac{\alpha}{2} G_{\tilde{u}}^\top & -\frac{\beta}{2} N_{y\tilde{u}}^\top & I \end{bmatrix}.$$

The real 3×3 matrix D:

$$D = \begin{bmatrix} \alpha(1-\rho), & -1 - \frac{\beta}{2} \|N_{yy}\| & -\frac{\alpha}{2} \|G_{\tilde{u}}\| \\ -1 - \frac{\beta}{2} \|N_{yy}\| & \beta(1-\rho) & -\frac{\beta}{2} \|N_{y\tilde{u}}\| \\ -\frac{\alpha}{2} \|G_{\tilde{u}}\| & -\frac{\beta}{2} \|N_{y\tilde{u}}\| & 1 \end{bmatrix},$$

back

Contractivity of the Jacobian J^*:

Proposition: [A. Griewank] Unless they happen to coincide with eigenvalues of G_y, the eigenvalues of J^* solve

$$\det[P(\lambda)] = 0 \quad \text{where} \quad P(\lambda) = (\lambda - 1)B + H(\lambda),$$

$$H(\lambda) = Z(\lambda)^\top N_{xx} Z(\lambda),\ N_{xx} = \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix},\ Z(\lambda) = \begin{bmatrix} (\lambda I - G_y)^{-1} G_u \\ I \end{bmatrix}.$$

Remarks:

▷ $N_{xx} = N_{xx}^\top \succ 0 \implies H(\lambda)^\top = H(\lambda) \succ 0.$

▷ Choice : $B = B^\top \succ 0 \implies \lambda \geq 1$ can not be an eigenvalue of J^*.

▷ $(\lambda I - G_y)^{-1} = \dfrac{(\lambda I - G_y)^\sharp}{\det[\lambda I - G_y]} \implies (\lambda I - G_y)^{-1} G_u = O(|\lambda|^{-1})$ for $|\lambda| \uparrow$

$\implies (\lambda - 1)B$ is the dominate term of $P(\lambda)$ for $|\lambda| \uparrow \implies \lambda << -1$, $P(\lambda) \prec 0$. Choice: $P(-1) = -2B + H(-1) \prec 0 \implies B \succ \frac{1}{2} H(-1)$ back.

Contractivity of the Jacobian J^*:

Proposition: [A. Griewank] Unless they happen to coincide with eigenvalues of G_y, the eigenvalues of J^* solve

$$\det[P(\lambda)] = 0 \quad \text{where} \quad P(\lambda) = (\lambda - 1)B + H(\lambda),$$

$$H(\lambda) = Z(\lambda)^\top N_{xx} Z(\lambda),\ N_{xx} = \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix},\ Z(\lambda) = \begin{bmatrix} (\lambda I - G_y)^{-1} G_u \\ I \end{bmatrix}.$$

Remarks:

▷ $N_{xx} = N_{xx}^\top \succ 0 \implies H(\lambda)^\top = H(\lambda) \succ 0.$

▷ *Choice* : $B = B^\top \succ 0 \implies \lambda \geq 1$ can not be an eigenvalue of J^*.

▷ $(\lambda I - G_y)^{-1} = \dfrac{(\lambda I - G_y)^\sharp}{\det[\lambda I - G_y]} \implies (\lambda I - G_y)^{-1} G_u = O(|\lambda|^{-1})$ for $|\lambda| \uparrow$

$\implies (\lambda - 1)B$ is the dominate term of $P(\lambda)$ for $|\lambda| \uparrow \implies \lambda << -1$, $P(\lambda) \prec 0$. Choice: $P(-1) = -2B + H(-1) \prec 0 \implies B \succ \frac{1}{2} H(-1)$ back.

Contractivity of the Jacobian J^*:

Proposition: [A. Griewank] Unless they happen to coincide with eigenvalues of G_y, the eigenvalues of J^* solve

$$\det[P(\lambda)] = 0 \quad \text{where} \quad P(\lambda) = (\lambda - 1)B + H(\lambda),$$

$$H(\lambda) = Z(\lambda)^\top N_{xx} Z(\lambda),\; N_{xx} = \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix},\; Z(\lambda) = \begin{bmatrix} (\lambda I - G_y)^{-1} G_u \\ I \end{bmatrix}.$$

Remarks:

▷ $N_{xx} = N_{xx}^\top \succ 0 \implies H(\lambda)^\top = H(\lambda) \succ 0.$

▷ *Choice* : $B = B^\top \succ 0 \implies \lambda \geq 1$ can not be an eigenvalue of J^*.

▷ $(\lambda I - G_y)^{-1} = \dfrac{(\lambda I - G_y)^\sharp}{\det[\lambda I - G_y]} \implies (\lambda I - G_y)^{-1} G_u = O(|\lambda|^{-1})$ for $|\lambda| \uparrow$

$\implies (\lambda - 1)B$ is the dominate term of $P(\lambda)$ for $|\lambda| \uparrow \implies \lambda << -1$, $P(\lambda) \prec 0$. Choice: $P(-1) = -2B + H(-1) \prec 0 \implies B \succ \frac{1}{2} H(-1)$ back.

Contractivity of the Jacobian J^*:

Proposition: [A. Griewank] Unless they happen to coincide with eigenvalues of G_y, the eigenvalues of J^* solve

$$\det[P(\lambda)] = 0 \quad \text{where} \quad P(\lambda) = (\lambda - 1)B + H(\lambda),$$

$$H(\lambda) = Z(\lambda)^\top N_{xx} Z(\lambda),\; N_{xx} = \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix},\; Z(\lambda) = \begin{bmatrix} (\lambda I - G_y)^{-1} G_u \\ I \end{bmatrix}.$$

Remarks:

▷ $N_{xx} = N_{xx}^\top \succ 0 \implies H(\lambda)^\top = H(\lambda) \succ 0.$

▷ *Choice* : $B = B^\top \succ 0 \implies \lambda \geq 1$ can not be an eigenvalue of J^*.

▷ $(\lambda I - G_y)^{-1} = \dfrac{(\lambda I - G_y)^\sharp}{\det[\lambda I - G_y]} \implies (\lambda I - G_y)^{-1} G_u = O(|\lambda|^{-1})$ for $|\lambda| \uparrow$

$\implies (\lambda - 1)B$ is the dominate term of $P(\lambda)$ for $|\lambda| \uparrow \implies \lambda << -1$, $P(\lambda) \prec 0$. Choice: $P(-1) = -2B + H(-1) \prec 0 \implies B \succ \frac{1}{2} H(-1)$ *back*.

Practical implementation

- Given routine: input

$$eval(\overset{\downarrow}{y},\overset{\downarrow}{u},\underset{\downarrow}{z},\underset{\downarrow}{f})$$

output

where $z = G(y,u)$ and $f = f(y,u)$.

- Applying an AD tool yields:

$$beval(\underset{\downarrow}{\overset{(0)}{by}},\overset{\downarrow}{y},\underset{\downarrow}{\overset{(0)}{bu}},\overset{\downarrow}{u},\overset{\downarrow}{bz},\underset{\downarrow}{z},\overset{\downarrow}{bf},\underset{\downarrow}{f})$$

where

$$by = G_y(y,u)^\top bz + f_y(y,u)^\top bf$$
$$bu = G_u(y,u)^\top bz + f_u(y,u)^\top bf$$

Practical implementation

Initialisation: $y = y_0$ and $u = u_0$.

Repeat:

- *dbeval* $(\overset{(0)}{\underset{\downarrow}{bu}}, \overset{\downarrow}{u}, \overset{\downarrow}{du}, \overset{(0)}{\underset{\downarrow}{dbu}}, \overset{(0)}{\underset{\downarrow}{by}}, \overset{\downarrow}{y}, \overset{\downarrow}{dy}, \overset{(0)}{\underset{\downarrow}{dby}}, \overset{\downarrow}{bz}, \underset{\downarrow}{z}, \underset{\downarrow}{dz}, \overset{\downarrow}{dbz}, \overset{\downarrow}{bf}, \underset{\downarrow}{f}, \underset{\downarrow}{df})$.

- Calculation of B_k from dz, dbu and dby.
- Line-search: $(y_k, \bar{y}_k, u_k) \rightsquigarrow \eta(z_k - y_k, bz_k - by_k, -B_k^{-1}u_k)$.
- $y = z$, $bz = by$.

until : $\|z - y\| \simeq 0 \simeq \|by - bz\| \simeq 0 \simeq \|bu\|$.

Retardation ratios for various precisions

Figure: Retardation ratios for various precisions

Ellipsoid norm: example

- $A = \begin{bmatrix} 0 & 10^8 \\ 0 & 0 \end{bmatrix}$. $\lambda_1(A) = \lambda_2(A) = 0 \Longrightarrow \rho(A) = 0.$

- $AA^\top = \begin{bmatrix} 10^{16} & 0 \\ 0 & 0 \end{bmatrix}$. $\lambda_i(AA^\top) = \{0, 10^{16}\}, i = 1,2 \Longrightarrow \|A\| = \sqrt{\rho(AA^\top)} = 10^8.$

$\varepsilon = 10^{-2}$: $W = \begin{bmatrix} 10^{-20} & 0 \\ 0 & 1 \end{bmatrix}, \forall\, v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, |||v||| = \sqrt{v^\top W v} = \sqrt{10^{-20} v_1^2 + v_2^2}.$

$$Av = \begin{bmatrix} 10^8 v_2 \\ 0 \end{bmatrix}, \quad |||Av||| = \sqrt{(Av)^\top W (Av)} = \sqrt{10^{-4} v_2^2} = 10^{-2}|v_2|.$$

$$\Longrightarrow |||A||| = \sup_{v \neq 0} \frac{10^{-2}|v_2|}{\sqrt{10^{-20} v_1^2 + v_2^2}} \leq 10^{-2} = \rho(A) + \varepsilon$$

back

Jacobian

$$J^* = J\Big|_{(y^*,\bar{y}^*,u^*)} = \begin{bmatrix} G_y & 0 & G_u \\ N_{yy} & G_y^\top & N_{yu} \\ -B^{-1}N_{uy} & -B^{-1}G_u^\top & I - B^{-1}N_{uu} \end{bmatrix}, \quad B_k \longrightarrow B.$$

Reduced Hessian:

Null space: $Ns := \{w \;\; s.t \;\; \nabla(G-y)w = 0\}$

$$\nabla(G-y)w = \begin{bmatrix} G_y - I \\ Gu \end{bmatrix} \begin{bmatrix} w1 \\ w2 \end{bmatrix} = 0 \implies w = \begin{bmatrix} (I-G_y)^{-1}G_u \\ I \end{bmatrix}.$$

$\Longrightarrow$ Reduced Hessian H:

$$H = \begin{bmatrix} (I-G_y)^{-1}G_u \\ I \end{bmatrix}^{\top} \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix} \begin{bmatrix} (I-G_y)^{-1}G_u \\ I \end{bmatrix}.$$

back

Optimal B

Rescaling: $u = \tilde{u}B^{-1/2}, G_{\tilde{u}} = G_u B^{-1/2}, N_{y\tilde{u}} = N_{yu}B^{-1/2}$.

Lemma : *Let* $D_C = \begin{bmatrix} D & -d \\ -d^\top & 1 \end{bmatrix}$ *where*

$$D = \begin{bmatrix} \alpha(1-\rho) & -\delta \\ -\delta & \beta(1-\rho) \end{bmatrix}, \delta = 1 + \frac{c}{2}\beta \text{ and } d = \begin{bmatrix} \frac{\alpha}{2}\|G_{\tilde{u}}\| \\ \frac{\beta}{2}\|N_{y\tilde{u}}\| \end{bmatrix}.$$

D_C is positive definite $\Longrightarrow$ s is a descent direction.

Remark:

$$\begin{bmatrix} I & 0 \\ d^\top D^{-1} & 1 \end{bmatrix} \begin{bmatrix} D & -d \\ -d^\top & 1 \end{bmatrix} \begin{bmatrix} I & D^{-1}d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & 1 - d^\top D^{-1} d \end{bmatrix}$$

$$\Longrightarrow \quad 1 - d^\top D^{-1} d \succ 0.$$

Optimal B

The condition $1 - d^\top D^{-1} d \succ 0$ is given by

$$\left(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|\right)^2 < (1-\rho) - \frac{\delta^2}{\alpha\beta(1-\rho)}.$$

Firstly,

$$\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\|_2 + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|_2 \leq \max\{\sqrt{\alpha}\|G_{\tilde{u}}\|_2, \sqrt{\beta}\|N_{y\tilde{u}}\|_2\} \leq \left\|\begin{matrix}\sqrt{\alpha}G_{\tilde{u}}\\ \sqrt{\beta}N_{y\tilde{u}}\end{matrix}\right\|_2.$$

Secondly,

$$\begin{bmatrix}\sqrt{\alpha}G_u\\ \sqrt{\beta}N_{yu}\end{bmatrix} = QR \Longrightarrow \left\|\begin{matrix}\sqrt{\alpha}G_{\tilde{u}}\\ \sqrt{\beta}N_{y\tilde{u}}\end{matrix}\right\|_2^2 = \|QRB^{-\frac{1}{2}}\|_2^2 = \|QRB^{-1}R^\top Q^\top\|_2.$$

Let $\sigma = \|QRB^{-1}R^\top Q^\top\|$, $RB^{-1}R^\top = \sigma I \Longrightarrow B = \frac{1}{\sigma}R^\top R$