DFG Research Center MATHEON
Mathematics for key technologies

# Definition and Evaluation of Projected Hessians for Piggyback Optimization

Andreas Griewank, Nicolas Gauger, Jan Riehme
Humboldt-Universität zu Berlin

Optimization in Aerodynamics, HU Berlin, 9. May 2005

# Table of Content

# Assumption:

$F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ composition of elemental $C^2$ functions

## ▷Consequence:

$$\mathbf{y} = F(\mathbf{x}) \;:\; \mathbb{R}^n \longrightarrow \mathbb{R}^m \text{ is } C^2 \quad \text{with Jacobian} \quad F'(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^{m\times n}$$

and

$$\dot{\mathbf{y}} = \dot{F}(\mathbf{x}, \dot{\mathbf{x}}) \equiv F'(\mathbf{x})\dot{\mathbf{x}} \in \mathbb{R}^n \quad \text{for any} \quad \dot{\mathbf{x}} \in \mathbb{R}^n$$

$$\bar{\mathbf{x}} = \bar{F}(\mathbf{x}, \bar{\mathbf{y}}) \equiv \bar{\mathbf{y}} F'(\mathbf{x}) \in \mathbb{R}^n \quad \text{for any} \quad \bar{\mathbf{y}} \in \mathbb{R}^m$$

$$\dot{\bar{\mathbf{x}}} = \dot{\bar{F}}(\mathbf{x}, \dot{\mathbf{x}}, \bar{\mathbf{y}}, \dot{\bar{\mathbf{y}}}) \equiv \bar{\mathbf{y}} F''(\mathbf{x})\,\dot{\mathbf{x}} + \dot{\bar{\mathbf{y}}} F'(\mathbf{x}) \in \mathbb{R}^n \quad \text{for any} \quad \dot{\bar{\mathbf{y}}} \in \mathbb{R}^m$$

# Complexity:

$$\frac{\max\left[\mathtt{OPS}(\bar{F}(\mathbf{x},\bar{\mathbf{y}}))\,,\ \mathtt{OPS}(\dot{F}(\mathbf{x},\dot{\mathbf{x}}))\right]}{\mathtt{OPS}(F(\mathbf{x}))} \quad\leq\quad 3 \quad\geq\quad \frac{\mathtt{CLOCK}(\bar{F}(\mathbf{x},\bar{\mathbf{y}}))}{\mathtt{CLOCK}(F(\mathbf{x}))}$$

and

$$\frac{\mathtt{OPS}(\dot{\bar{F}}(\mathbf{x},\bar{\mathbf{y}},\dot{\mathbf{x}},\dot{\bar{\mathbf{y}}}))}{\mathtt{OPS}(F(\mathbf{x}))} \quad\leq\quad 9 \quad\geq\quad \frac{\mathtt{CLOCK}(\dot{\bar{F}}(\mathbf{x},\bar{\mathbf{y}},\dot{\mathbf{x}},\dot{\bar{\mathbf{y}}}))}{\mathtt{CLOCK}(F(\mathbf{x}))}$$

BUT in worst case

$$\frac{\mathtt{OPS}(F'(\mathbf{x}))}{\mathtt{OPS}(F(\mathbf{x}))} \quad\sim\quad n \quad\sim\quad \frac{\mathtt{OPS}(\bar{\mathbf{y}}F''(\mathbf{x}))}{\mathtt{OPS}(F(\mathbf{x}))}$$

and it's hard to tell the good cases.

# Optimal Design Scenario

**• Problem:**

$$\text{Min } f(y,u) \quad \text{s.t.} \quad c(y,u) = 0$$

where $y \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$ are state and design variables

**• Available:**

Code for $f(y,u)$ and $G(y,u) \approx y - \left[\frac{\partial}{\partial y} c(y,u)\right]^{-1} c(y,u)$

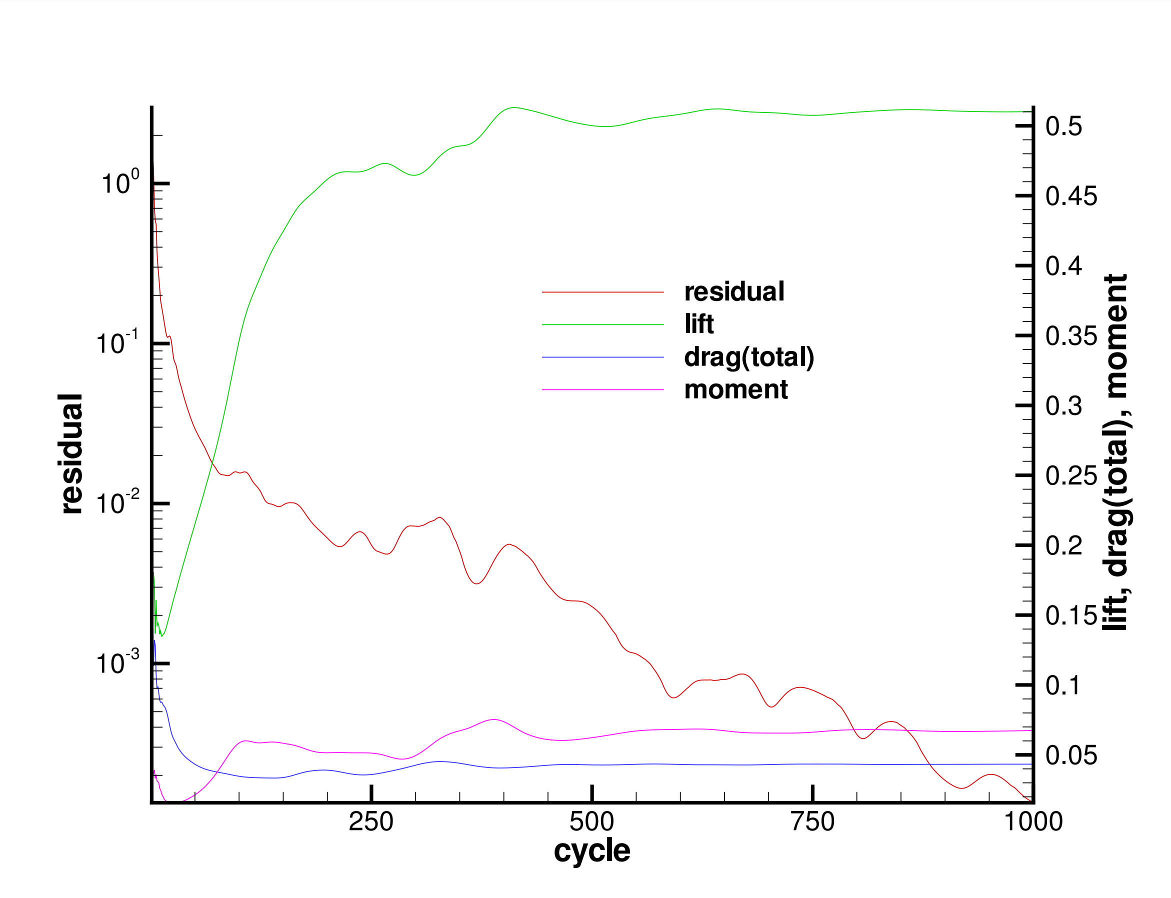
**• Assumption:**

$$G, f \in C^{2,1}(\mathbb{R}^{n+m}) \quad \text{and} \quad \|\tfrac{\partial}{\partial y} G(y,u)\| \leq \rho < 1$$

**• Notation:**

$$N(\bar{y}, y, u) \equiv f(y,u) + \bar{y}\, G(y,u) \equiv \text{Lagrangian} + \bar{y} y,$$

where the Lagrangian is formed w.r.t. $c(y,u) \equiv G(y,u) - y = 0.$


residual
lift
drag(total)
moment
residual
$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
250
500
750
1000
cycle
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
lift, drag(total), moment

# Structured KKT System :

## Adjoint Equation:

$$0 \;=\; \bar{c}(y, \bar{y}, u) \;\equiv\; \bar{y}\, c_y(y, u) + f_y(y, u) \quad \textbf{with} \quad \bar{y}^T \in \mathbb{R}^n$$

## Reduced Gradient

$$\bar{u}(y, \bar{y}, u) \;\equiv\; \bar{y}\, c_u(y, u) + f_u(y, u) \quad \textbf{with} \quad \bar{u}^T \in \mathbb{R}^m$$

## Coupled System

$$\begin{bmatrix} c(y, u) \\ \bar{c}(y, \bar{y}, u) \\ \bar{u}(y, \bar{y}, u) \end{bmatrix} \;=\; 0$$

# Exact Subspace Iteration:

Normal Step

$$y \;=\; G(y,u) \;=\; y - c_y(y,u)^{-1}c(y,u)$$

Adjoint Step

$$\bar{y} \;=\; f_y(y,u)\,c_y(y,u)^{-1} \quad \text{and} \quad \bar{u} = \bar{y}\,c_u(y,u) + f_u(y,u)$$

Optimization Step

$$u \;=\; u - H^{-1}\bar{u}^T$$

with *preconditioner* $H \in \mathbb{R}^{m\times m}$ incorporating step-size.

(At best) linearly convergent:

$$\cdots \to \textbf{normal} \to \textbf{adjoint} \to \textbf{optim} \to \cdots$$

## Superlinearly convergent:

$$\cdots \to \textbf{normal} \to \textbf{adjoint} \to \textbf{optim} \to \textbf{normal} \to \cdots$$

if

$$H(1) \;=\; L_{uu} + Z^T L_{yu} + L_{uy} Z + Z^T L_{yy} Z \tag{1}$$

Most general substep sequence

$$\cdots \to [\textbf{normal}]^s \to [\textbf{normal}, \textbf{adjoint}]^r \to \textbf{optim} \to \cdots \tag{2}$$

for certain repetition numbers $r$ and $s$.

Hierarchical Approach

$$\rho^{(s+r)} \;\approx\; 0 \Longrightarrow \; H_k = \; H(1)$$

otherwise we need more conservative = larger $H_k$.
From now on *single-step* method

$$\cdots \to [\textbf{normal}, \textbf{adjoint}, \textbf{optim}] \to \cdots \tag{3}$$

# Single-step-one-shot = Piggy-Back Approach :

$$y_{k+1} = G(y_k, u_k) \longrightarrow \text{primal feasibility at } y_*$$

$$\bar{y}_{k+1} = N_y(y_k, \bar{y}_k, u_k) \longrightarrow \text{dual feasibility at } \bar{y}_*$$

$$u_{k+1} = u_k - H_k^{-1} N_u(y_k, \bar{y}_k, u_k) \longrightarrow \text{optimality at } u_*$$

where $N_u = \bar{y}\, G_u + f_u \approx$ reduced gradient

and $H_k$ is a suitable preconditioner

# Questions/Tasks for Piggy-Back:

- Avoid data objects larger than $\dim(y) \cdot \dim(u)$
- Analyse convergence of $y_k$ and $\bar{y}_k$ for fixed $u$
- Determine preconditioner $H_k$ for fast local convergence
- Evaluate/approximate $H_k$ by differentiation or updating
- Globalize by tradeoff between feasibility and optimality.

# Linear Convergence Result

$$\|G_y(y,u) - G_y(\tilde{y},u)\| \leq \nu \|y - \tilde{y}\| \geq \|f_y(y,u) - f_y(\tilde{y},u)\|$$

$$\Longrightarrow$$

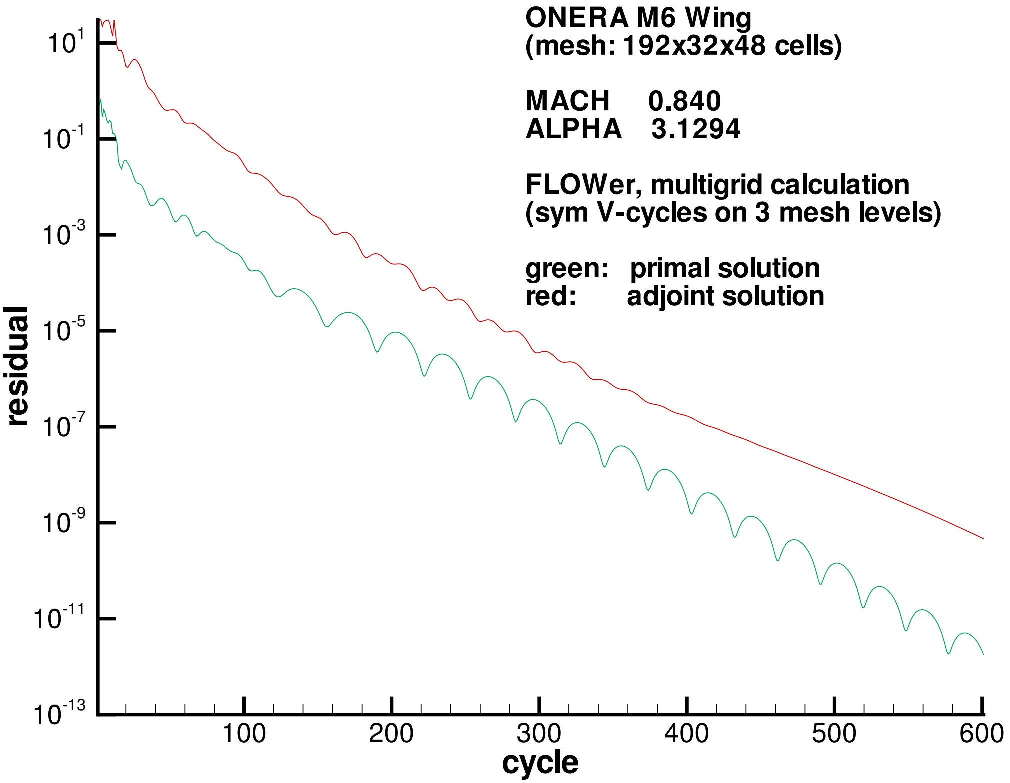
$$\overline{\lim} \sqrt[k]{\|\Delta y_k\|} \quad \leq \quad \varrho \quad \geq \quad \overline{\lim} \sqrt[k]{\|\Delta \bar{y}_k\|}$$

for $\Delta y_k = y_k - y(u) \quad , \quad \Delta \bar{y}_k = \bar{y}_k - \bar{y}(u)$

Proof: Based on monotonic reduction of

$$\|\Delta y_k\| + \omega \|\Delta \bar{y}_k\|$$

for $\omega$ small enough.


ONERA M6 Wing
(mesh: 192x32x48 cells)
MACH 0.840
ALPHA 3.1294
FLOWer, multigrid calculation
(sym V-cycles on 3 mesh levels)
green: primal solution
red: adjoint solution
residual
cycle
10^1
10^-1
10^-3
10^-5
10^-7
10^-9
10^-11
10^-13
100
200
300
400
500
600

Question: Do $\|\Delta y_k\|$ and $\|\Delta \bar{y}_k\|$ converge equally fast??

Answer: NO – Adjoints lag behind because:

$$(G, f) \in C^{2,1}, \quad N(\bar{y}, y, u) \equiv \bar{y}G(y, u) + f(y, u)$$

$$\Longrightarrow$$

$$\begin{bmatrix} \Delta y_{k+j} \\ \Delta \bar{y}_{k+j} \end{bmatrix} = \begin{bmatrix} G_y & 0 \\ N_{yy} & G_y^T \end{bmatrix}^j \begin{bmatrix} \Delta y_k \\ \Delta \bar{y}_k \end{bmatrix} + O\left(\|\Delta y_k\|^2 + \|\Delta \bar{y}_k\|^2\right)$$

$\Longrightarrow$ provided $G_y = T\Gamma T^{-1}, \quad D = \mathbf{diag}(T^T B T)$

$$\begin{bmatrix} G_y & 0 \\ N_{yy} & G_y^T \end{bmatrix}^j \sim \begin{bmatrix} \Gamma^j & 0 \\ jDT^{j-1} & \Gamma^j \end{bmatrix}$$

$$\Longrightarrow \|\Delta \bar{y}_{k+j}\| \approx j\varrho^{j-1}\|\Delta y_k\| + \varrho^j\|\Delta \bar{y}_k\| \Longrightarrow \|\Delta y_k\|/\|\Delta \bar{y}_k\| \sim 1/k \to 0.$$

# Consequence

Adjoints $\bar{y}_k$ and reduced gradient

$$\bar{u}_k \equiv N_u(\bar{y}_k, y_k, u_k) \quad \underset{k}{\longrightarrow} \quad \frac{df(y(u), u)}{du}$$

lag behind primal feasibility like $\frac{1}{k}$. For fixed $\dot{u}$ also

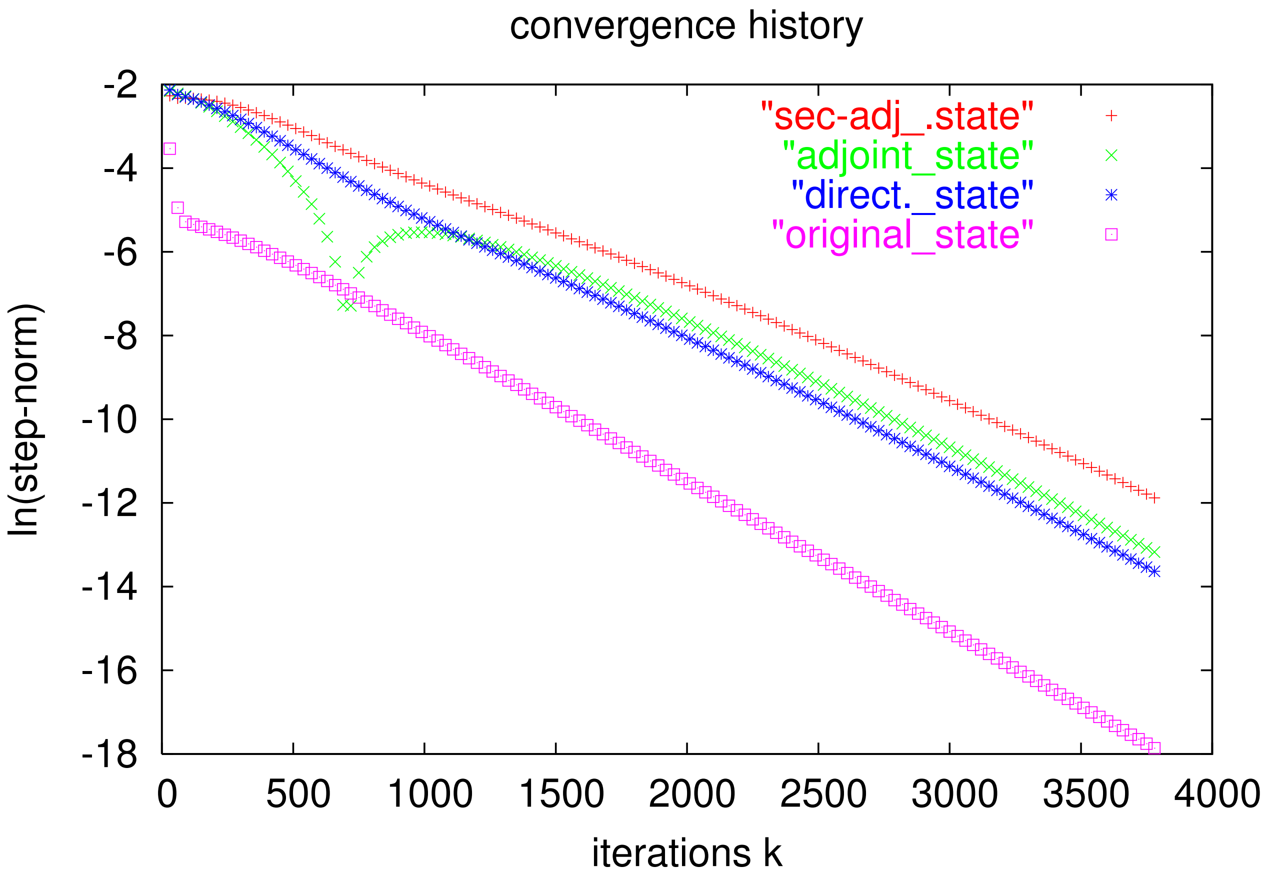
$$\dot{y}_{k+1} = \dot{G}(y_k, \dot{y}_k, \dot{u}) \equiv G_y(y_k, u_k)\dot{y}_k + G_u(x_k, u)\dot{u}$$

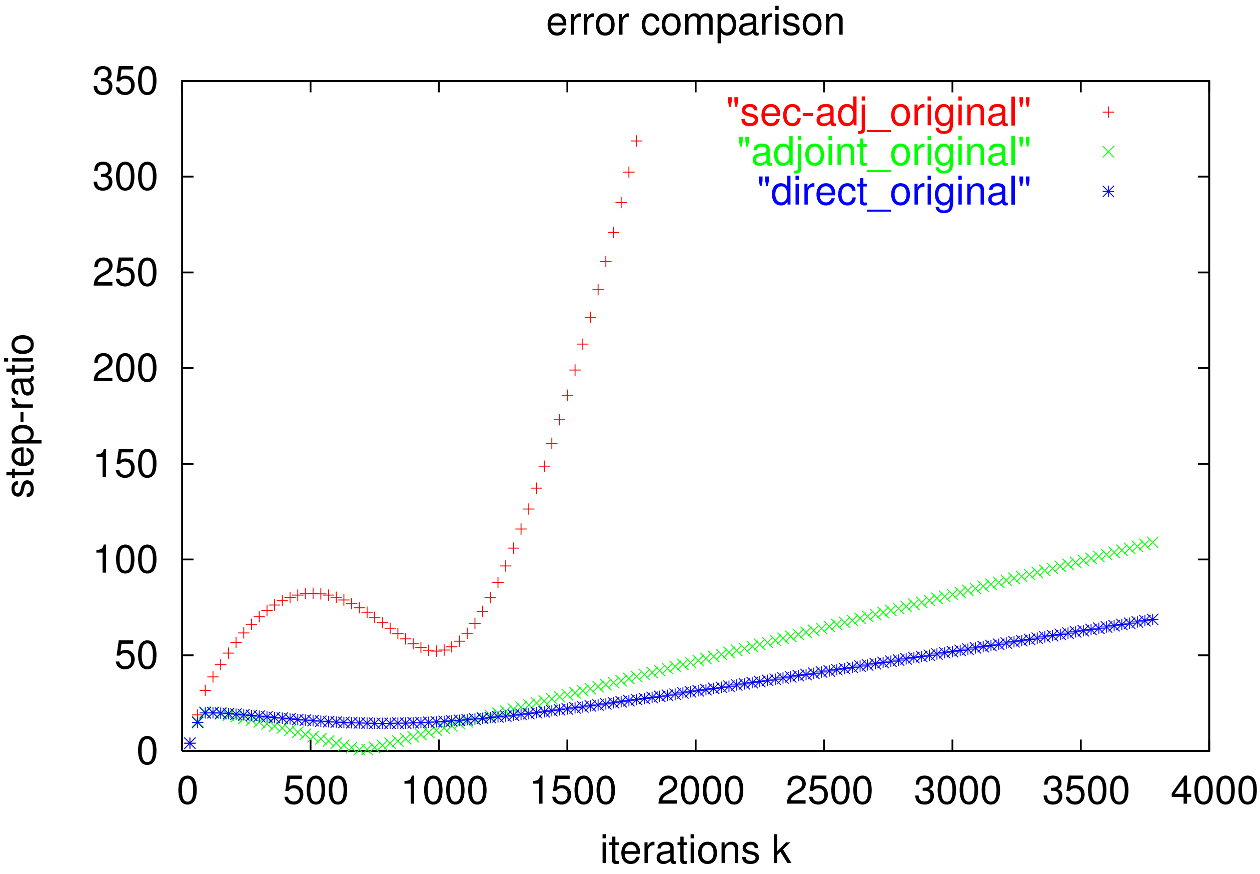
and

$$\dot{\bar{y}}_{k+1} \equiv \dot{\bar{G}}(\dot{y}_k, \bar{y}_k, \dot{y}_k, \dot{\bar{y}}_k, u, \dot{u}) = \dot{\bar{y}}_k G_y(y_k, u)...$$

converge, but lag behind according to

$$\|\Delta \dot{y}_k\| \sim k \|\Delta y_k\| \quad \text{and} \quad \|\Delta \dot{\bar{y}}_k\| \sim k^2 \|\Delta y_k\|$$


convergence history
"sec-adj_.state"
"adjoint_state"
"direct._state"
"original_state"
ln(step-norm)
-2
-4
-6
-8
-10
-12
-14
-16
-18
0
500
1000
1500
2000
2500
3000
3500
4000
iterations k


error comparison
"sec-adj_original"
"adjoint_original"
"direct_original"
step-ratio
iterations k
0
50
100
150
200
250
300
350
0
500
1000
1500
2000
2500
3000
3500
4000

# •Test Problem:

Min $f(y,u) \equiv \frac{1}{2}\int_0^1 [(y_n(\xi,1) - 2.2)^2 + \sigma(u(\xi)^2 + u'(\xi)^2)]d\xi$

s.t. $\Delta_x y(x) + e^{y(x)} = 0$ for $x \in [0,1]^2$

•Periodic boundary condition

$$y(0,\xi) = y(1,\xi) \quad \text{for } \xi \in [0,1]$$

•Dirichlet condition on lower edge

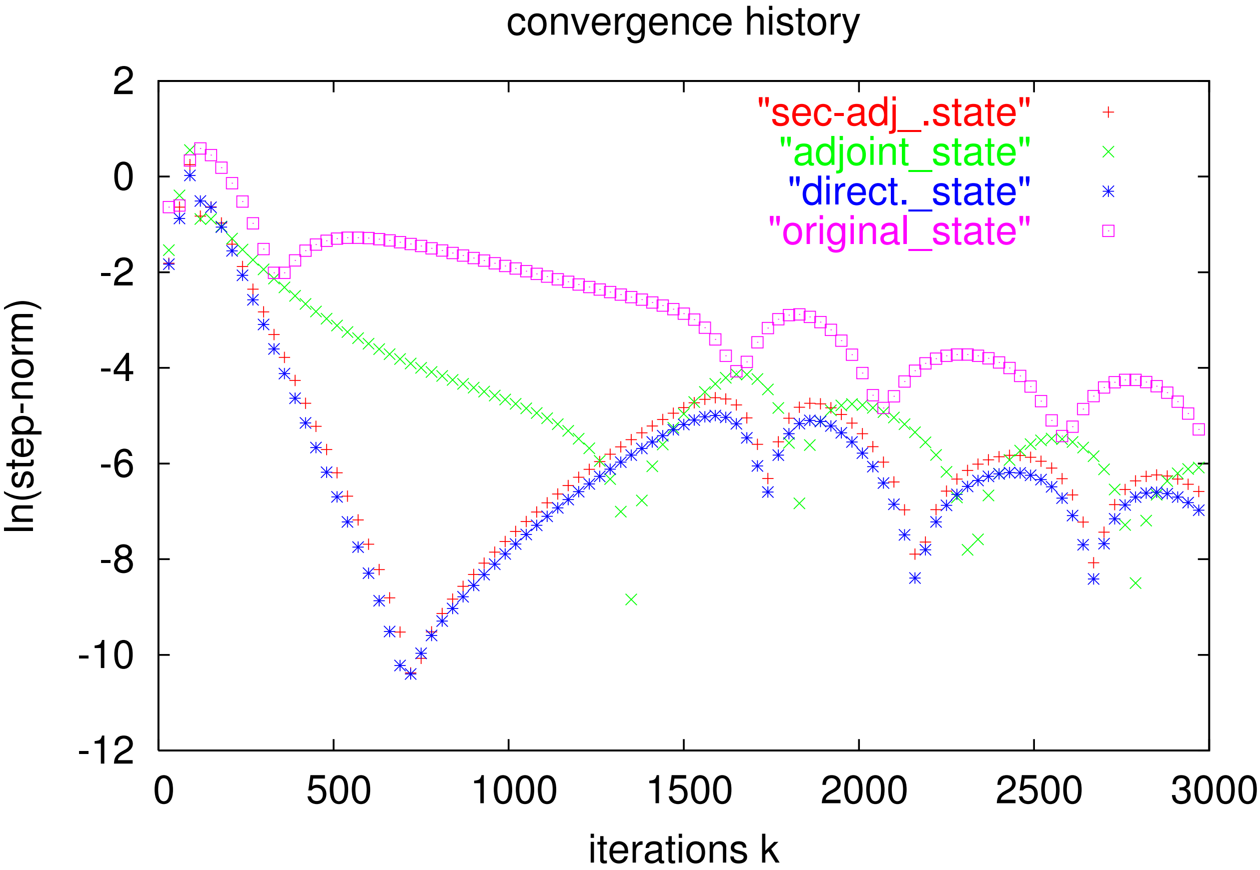
$$y(\xi,0) = \sin(2\pi\xi) \quad \text{for } \xi \in [0,1]$$

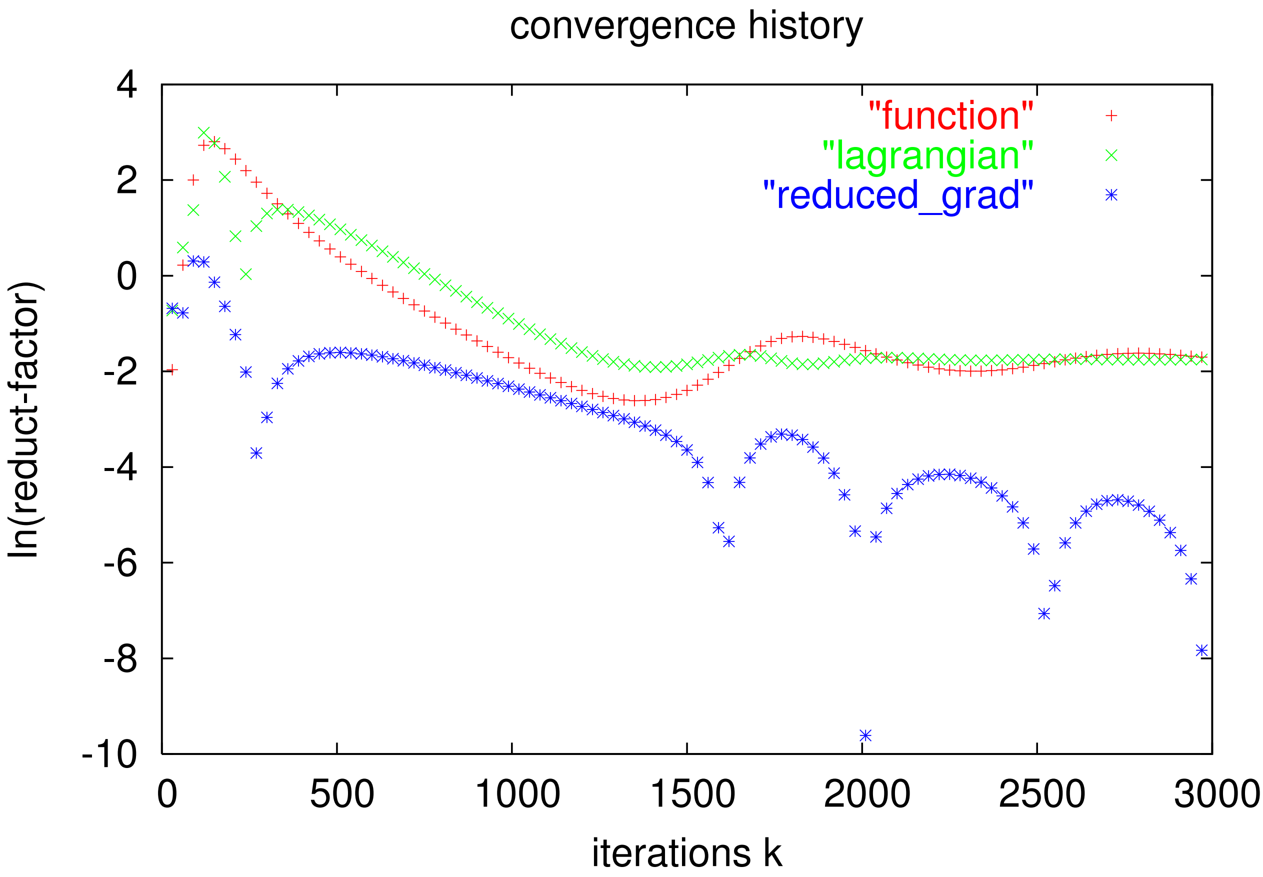
•Boundary control on upper edge

$$y(\xi,1) = u(\xi) \quad \text{for } \xi \in [0,1]$$

•Iteration Function

$G \equiv$ (nonlinear) Jacobi-method on 5 points descretization.


convergence history
"sec-adj_.state"
"adjoint_state"
"direct._state"
"original_state"
ln(step-norm)
iterations k
2
0
-2
-4
-6
-8
-10
-12
0
500
1000
1500
2000
2500
3000


convergence history
"function"
"lagrangian"
"reduced_grad"
ln(reduct-factor)
iterations k
4
2
0
-2
-4
-6
-8
-10
0
500
1000
1500
2000
2500
3000

# Spectral Analysis of Piggy-Back

$$\frac{\partial(y_{k+1}, \bar{y}_{k+1}, u_{k+1})}{\partial(y_k, \bar{y}_k, u_k)} = \begin{bmatrix} G_y & 0 & G_u \\ N_{yy} & G_y^T & N_{yu} \\ -H^{-1}N_{uy} & -H^{-1}G_u^T & I - H^{-1}N_{uu} \end{bmatrix}$$

has at $(y_*, \bar{y}_*, u_*)$ as eigenvalues $\lambda$ the roots of

$$P(\lambda) \equiv \det\left[H(\lambda) + (\lambda - 1)H\right]$$

where

$$\begin{aligned} H(\lambda) &\equiv [Z(\lambda)^T, I]\, \nabla^2{}_{(y,u)} N\, [Z(\lambda)^T, I]^T \\ Z(\lambda)^T &\equiv -G_u^T (G_y^T - \lambda I)^{-1} \end{aligned}$$

Rows of $[Z(\lambda)^T, I]$ span tangent space of $\{G(y, u) = \lambda y\}$

# Contractivity in convex case

$$\lambda < 1 \iff H \succ 0 \quad \text{i.e. } H \text{ pos. def.}$$
$$\lambda > -1 \implies H \succ H(-1)/2$$

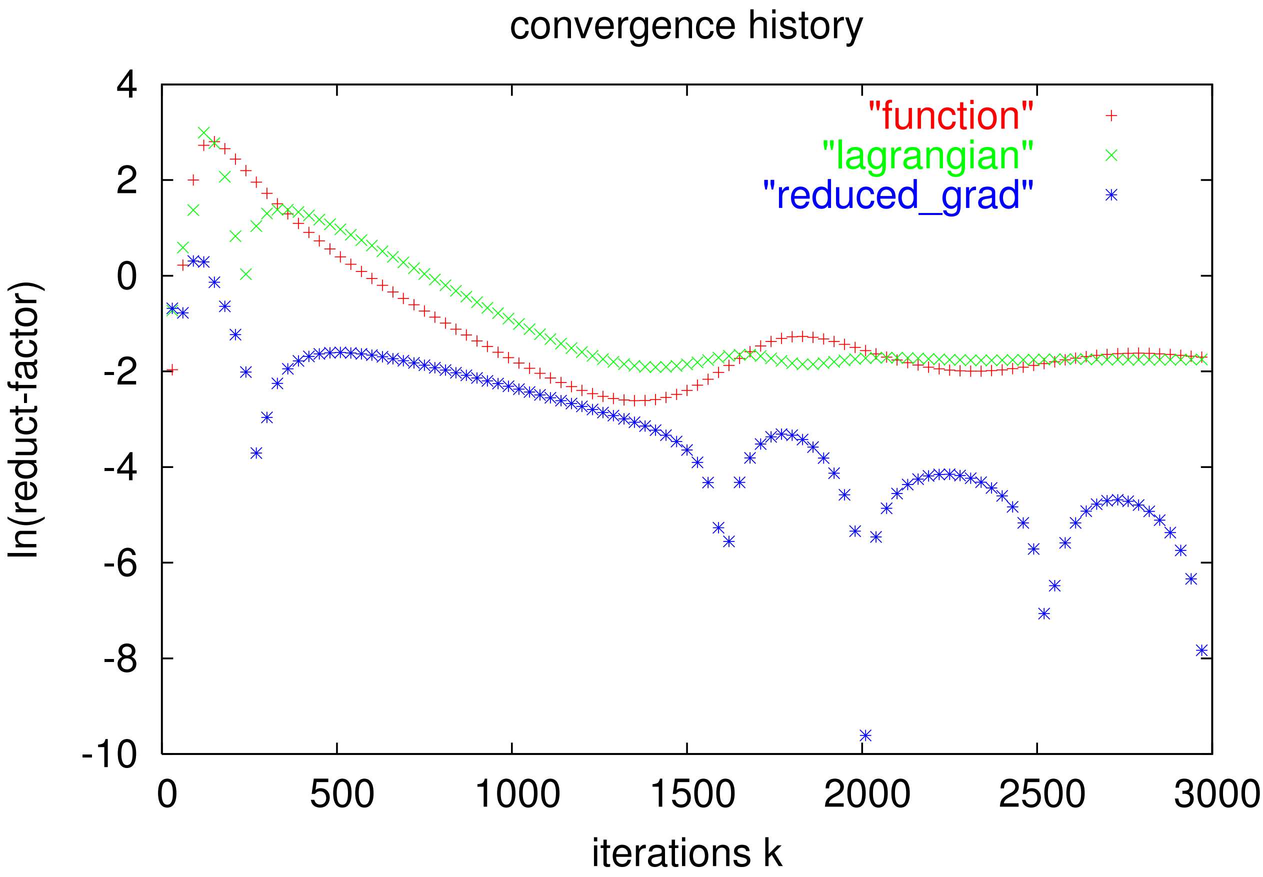
Numerical experience on test problem above:

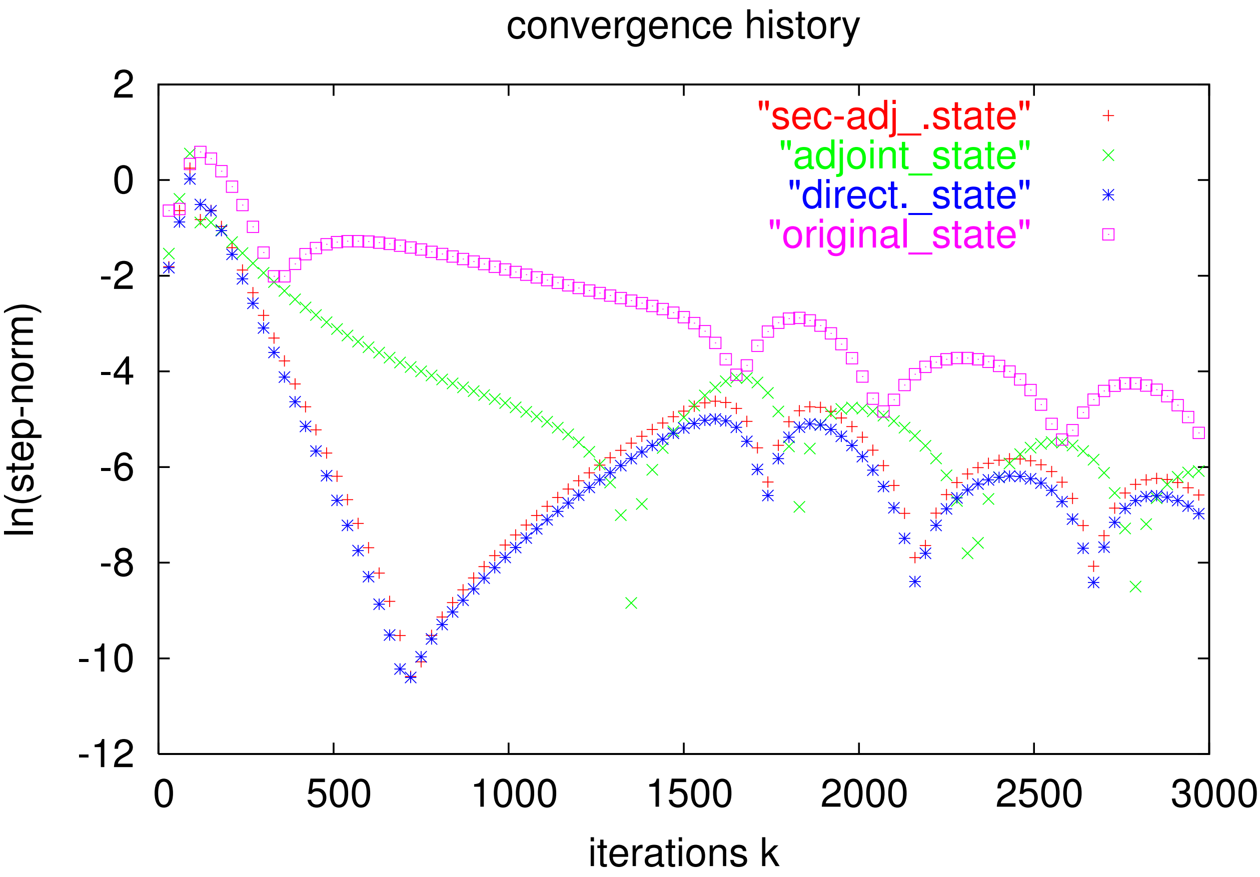
Reduced Hessian $H \equiv H(1) \implies$ Immediate Blow-up

Projected Hessian $H \equiv H(-1) \implies$ Full-step Convergence

$$
\begin{bmatrix}
34.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -3.62 \\
-27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 \\
-1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 \\
-0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 \\
-0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 \\
-0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 \\
-0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 & -0.07 \\
-0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 & -0.15 \\
-0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 & -0.38 \\
-0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 & -1.11 \\
-1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 58.90 & -27.62 \\
-3.62 & -1.11 & -0.38 & -0.15 & -0.07 & -0.06 & -0.07 & -0.15 & -0.38 & -1.11 & -27.62 & 34.90
\end{bmatrix}
$$

$$
\begin{bmatrix}
69.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -9.54 \\
-33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 \\
4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 \\
-2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 \\
1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 \\
-1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 \\
1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 & -1.46 \\
-1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 & 1.88 \\
1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 & -2.81 \\
-2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 & 4.82 \\
4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 93.81 & -33.54 \\
-9.54 & 4.82 & -2.81 & 1.88 & -1.46 & 1.34 & -1.46 & 1.88 & -2.81 & 4.82 & -33.54 & 69.81
\end{bmatrix}
$$


convergence history
"function"
"lagrangian"
"reduced_grad"
ln(reduct-factor)
iterations k
4
2
0
-2
-4
-6
-8
-10
0
500
1000
1500
2000
2500
3000


convergence history
"sec-adj_.state"
"adjoint_state"
"direct._state"
"original_state"
ln(step-norm)
iterations k
2
0
-2
-4
-6
-8
-10
-12
0
500
1000
1500
2000
2500
3000

# Implicit calculation of $H(\lambda)$ by additional iterations:

$$Z_{k+1} = [G_y(y_k, u_k) Z_k + G_u(y_k, u_k)] / \lambda$$

$$Z_*(1) \equiv \frac{dy_*}{du}, \qquad \bar{Z}_*(1) \equiv \frac{d\bar{y}_*}{du} \qquad H_*(1) \equiv \frac{d\bar{u}_*}{du}$$

$$\bar{Z}_{k+1} = [G_y(y_k, u_k)^T \bar{Z}_k + N_{yy}(y_k, \bar{y}_k, u_k) Z_k + N_{yu}(y_k, \bar{y}_k, u_k)] / \lambda$$

$$H_{k+1} = \bar{Z}_k G_u(y_k, u_k) + N_{uy}(y_k, \bar{y}_k, u_k) Z_k + N_{uu}(y_k, \bar{y}_k, u_k)$$

Convergence with contractive factor $\|G_y\|/|\lambda| = \varrho/|\lambda|$.

# Tentative explanation:

$$Z(1) \;=\; (I - G_y)^{-1} G_u$$

is rich in *monotonic* modes, i.e. eigenvectors with eigenvalues close to $1$.
lor

$$Z(-1) \;=\; (I + G_y)^{-1} G_u$$

is rich in *alternating* modes, i.e. eigenvectors with eigenvalues close to $-1$. Also the steps are likely to be predominantly in the alternating directions.

$$\begin{bmatrix} \Delta y_k \\ \Delta z_k \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & -\lambda \end{bmatrix} \begin{bmatrix} \Delta y_{k-1} \\ \Delta z_{k-1} \end{bmatrix}$$

with $0 < \lambda < 1$ yields

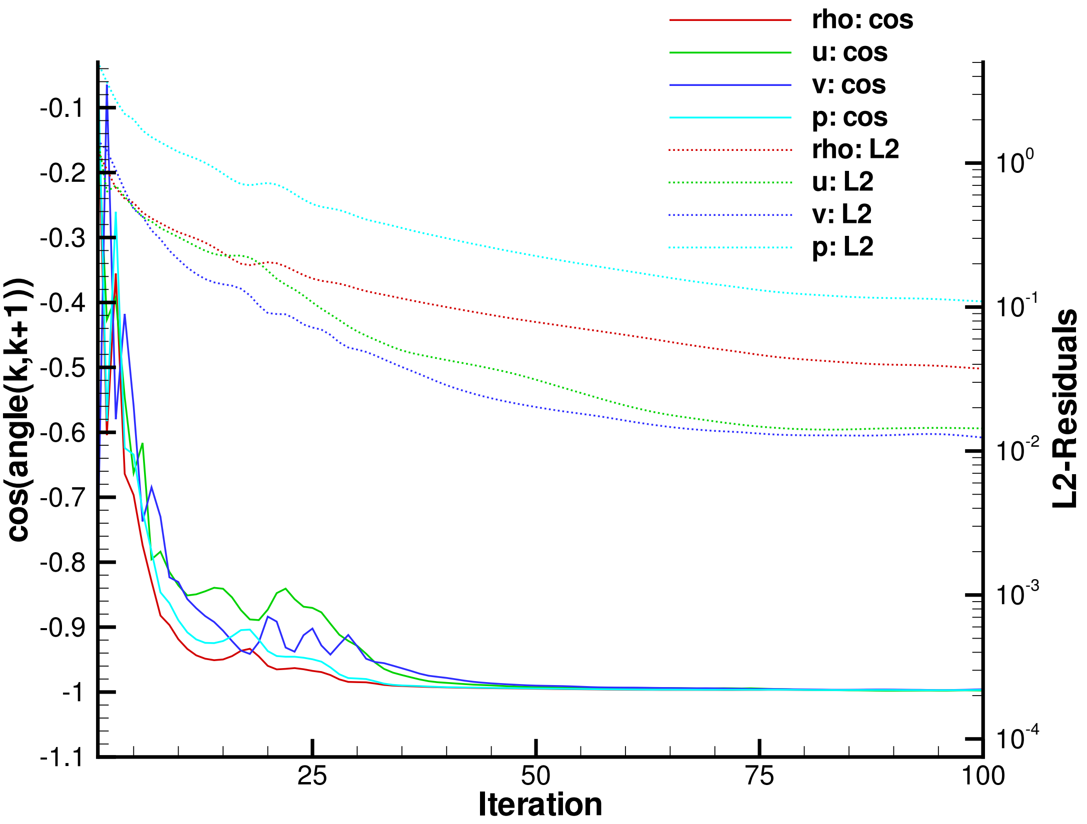
$$\frac{|y_k - y_{k-1}|}{|z_k - z_{k-1}|} = \frac{|y_0|}{|z_0|} \, \frac{1-\lambda}{1+\lambda}$$

It makes sense that the curvature of the Lagrangian with respect to the alternating modes is more critical than that with respect to the monotonic modes.
If all eigenvalues of $G_y$ were real the alternating modes could be eliminated by considering one double step

$$G^2(y,u) \;\equiv\; G(G(y,u),u))$$

as a single iteration. The resulting $H(-1)$ would be much smaller.

Working Hypothesis:
Interesting iterations have complex eigenvalues.


rho: cos
u: cos
v: cos
p: cos
rho: L2
u: L2
v: L2
p: L2
cos(angle(k,k+1))
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
-1.1
25
50
75
100
Iteration
$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
L2-Residuals

# Implementation at the Software Level:

User supplied routine

$$\begin{array}{c}\texttt{input:}\\ \text{step}(\overset{\downarrow}{u}, \overset{\downarrow}{y}, \underset{\downarrow}{z}, \underset{\downarrow}{f})\\ \texttt{output:}\end{array}$$

where:

z= G(u,y)

f=f(u,y)

Basic Iteration:

init(u,z); y=0

$while(\|y - z\| \gg 0)$

y=z

step(u,y,z,f)

use(z,f)

Applying an AD tool in reverse mode:

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

with:

$$\begin{aligned} bu &= G_u(y,u)^T bz + f_u(y,u)^T bf \\ by &= G_y(y,u)^T bz + f_y(y,u)^T bf \end{aligned}$$

Coupled basic and adjoint Iteration:

$$\text{init}(u, z, by); y = 0; bz = 0$$

$$while(\|z - y\| + \|by - bz\| \gg 0)$$

$$y = z; bz = by; bu = 0; by = 0, bf = 1$$

$$\text{bstep}(bu, u, by, y, bz, z, bf, f)$$

$$\text{use}(z, f, bu, by)$$

One more differentiation in forward mode yields:.

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

Coupled basic with first and second adjoint :

$$\text{init}(u, z, by, dz, dby); \quad y = 0; bz = 0; dby = 0$$

$$\begin{aligned} & while(\|z - y\| + \|by - bz\| + \|dz - dy\| + \|dby - dbz\| \gg 0) \\ & \quad y = z; bz = by, by = 0; bu = 0; dbu = 0; dby = 0 \\ & \quad dy = dz/\lambda, dbz = dby/\lambda \\ & \quad \text{dbstep}(bu, u, du, dbu, by, y, dy, dby, bz, z, dz, dbz, bf, fdf) \end{aligned}$$

$$\text{use}(z, f, bu, by, dbu, dby)$$

# Simplified Implementation:

User supplied routine

`input:`

$\text{step}(\overset{\downarrow}{u}, y1 \ldots y9, \underset{\downarrow}{f})$

`output:`

where:

y1 . . . y9= G(u,y1. . . y9)

f=f(u,y)

Basic Iteration:

init(u)

$while(???)$

step(u,y1 . . . y9,f)

use(f)

## Simplified reverse mode:

$$\text{bstep}(\underset{\downarrow}{\overset{(0)}{bu}}, \overset{\downarrow}{u}, by1 \ldots by9, y1 \ldots y9, \overset{\downarrow}{bf}, \underset{\downarrow}{f})$$

Coupled basic and adjoint Iteration:

$$\text{init}(u);$$

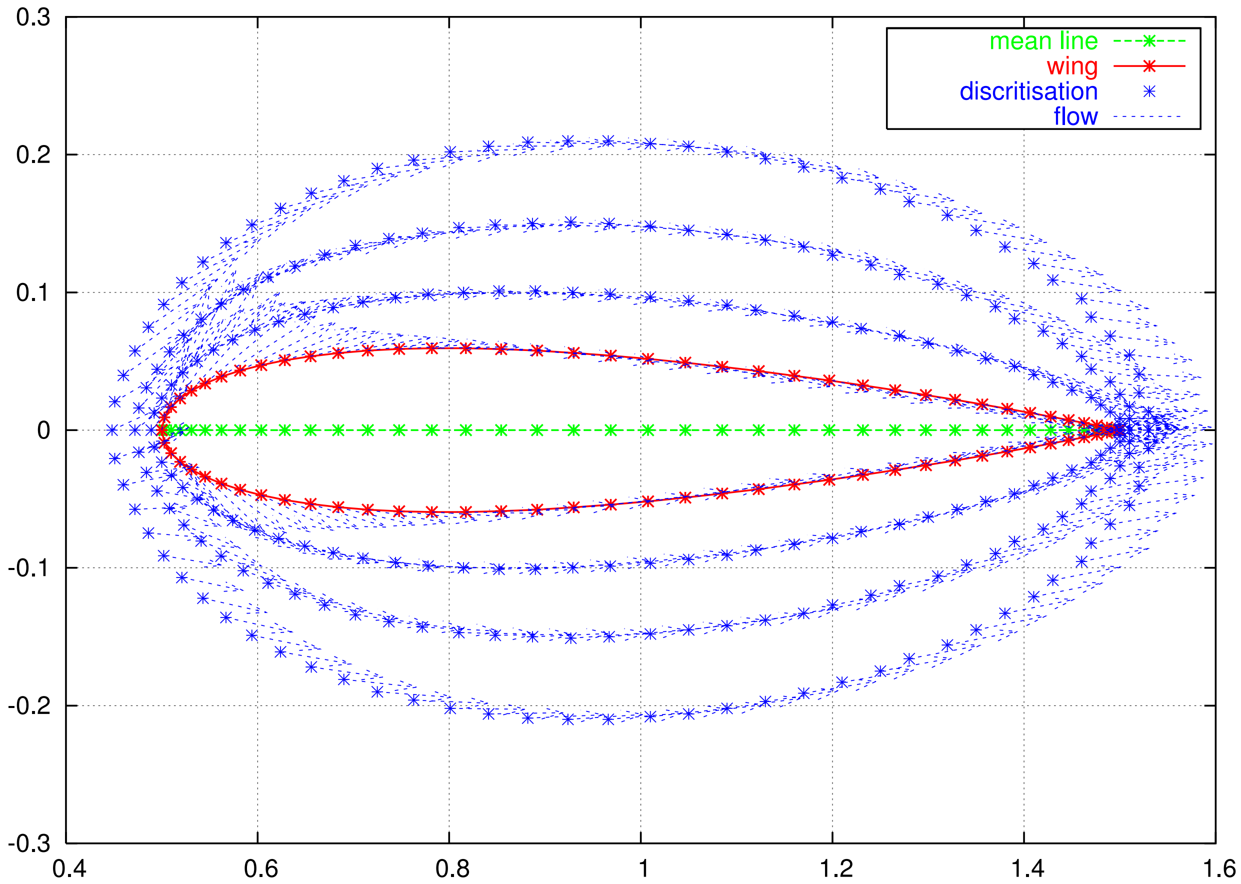
$$while(???)do$$

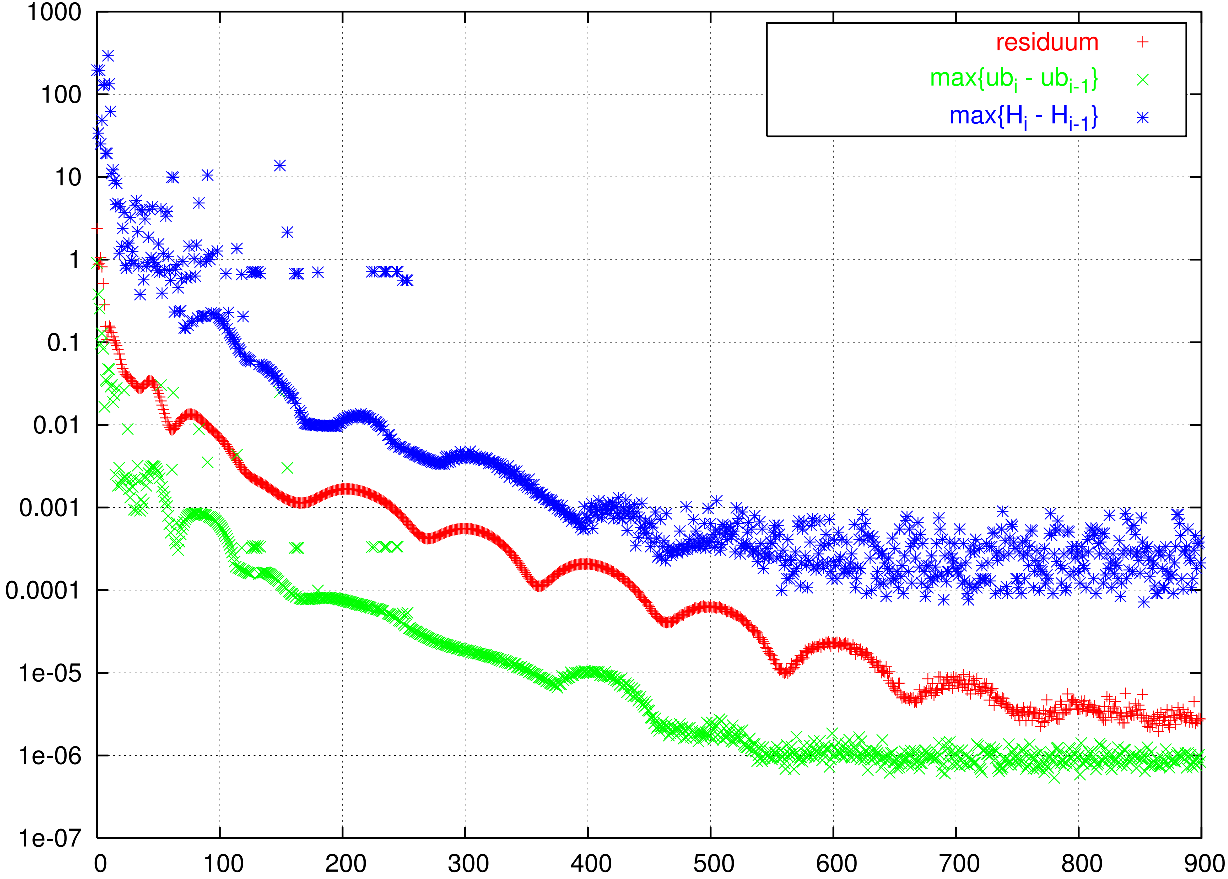
$$bu = 0; bf = 1$$

$$\text{bstep}(bu, u, by1 \ldots by9, y1 \ldots y9, bf, f)$$

$$\text{use}(f, bu)$$

Same structure for second order adjoints $\Longrightarrow$ preconditioners.


mean line
wing
discritisation
flow
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
0.4
0.6
0.8
1
1.2
1.4
1.6


1000
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
0
100
200
300
400
500
600
700
800
900
residuum
max{ub_i - ub_i-1}
max{H_i - H_i-1}

# Summary, Conclusion and Tasks

- Sensitivities 'easily' obtainable from fixed point solver
- Storage requirement does not grow with iteration number
- Derivatives converge also linearly but lag a little behind
- Reduced Hessian no good preconditioner for single step piggy-backing

- Optimal substep sequence and preconditioning depends on spectrum($G_y$)
- Preconditioning cost reducable to $\approx$ cost of simulation step ????
- Avoid second derivatives alltogether or at least in iteration !!!!