Definition and Evaluation of Projected Hessians for Piggyback Optimization

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

AD Workshop 2005, Nice, April 16

Table of Content

- General Scenario for Optimal Design
- Single-step-one-shot = Piggy Back Optimization
- Time-lagged derivative convergence for fixed u
- Spectral analysis of single-step iteration
- Numerical Result on Modified Bratu Example
- Software and Implementation Issues on CFD Code

Optimal Design Scenario

•**Problem:**

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

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

•**Available:**

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

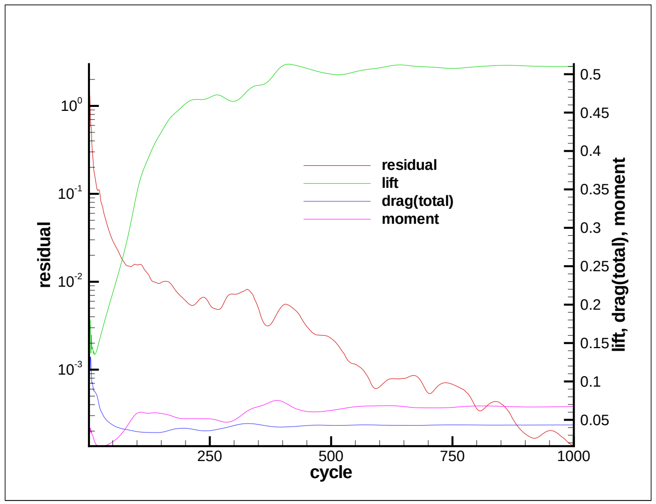
•**Assumption:**

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

•**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. $c(y,u) \equiv G(y,u) - y = 0$.



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

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

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

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

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

$\Longrightarrow$

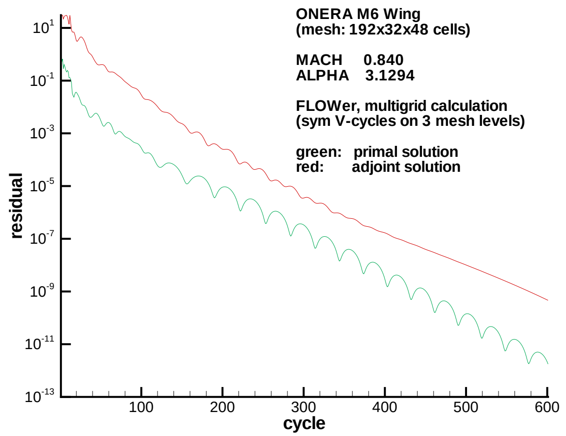
$$\overline{\lim} \sqrt[k]{\|\Delta y_k\|} \quad \leq \quad \varrho \quad \geq \quad \overline{\lim} \sqrt[k]{\|\Delta \bar{y}_k\|}$$

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 ω small enough.



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

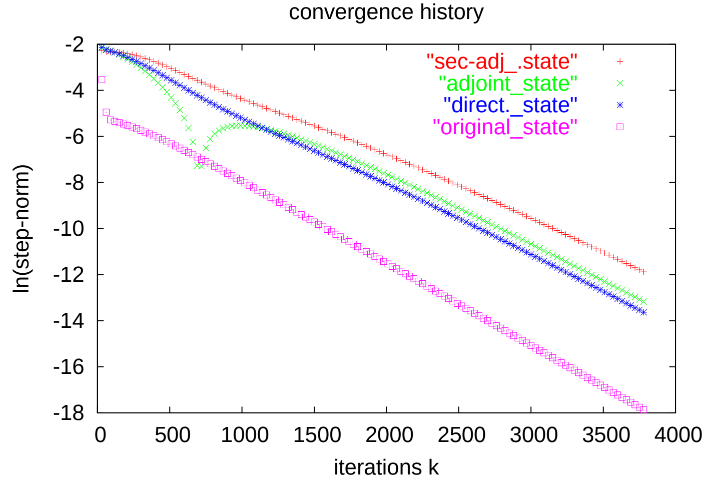
$$\dot{y}_{k+1} = \dot{G}(y_k, \dot{y}_k, \dot{u}) \equiv G_y(y_k, u_k)\dot{y}_k + G_u(x_k, u)\dot{u}$$

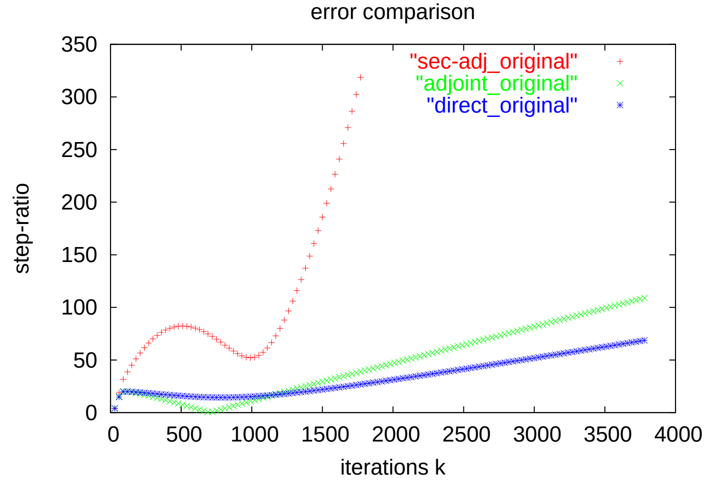
and

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

converge, but lag behind according to

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





•Test Problem:

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

s.t. $\quad \Delta_x y(x) + e^{y(x)} = 0 \quad$ 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.

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 λ the roots of

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

where

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

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

Contractivity in convex case

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

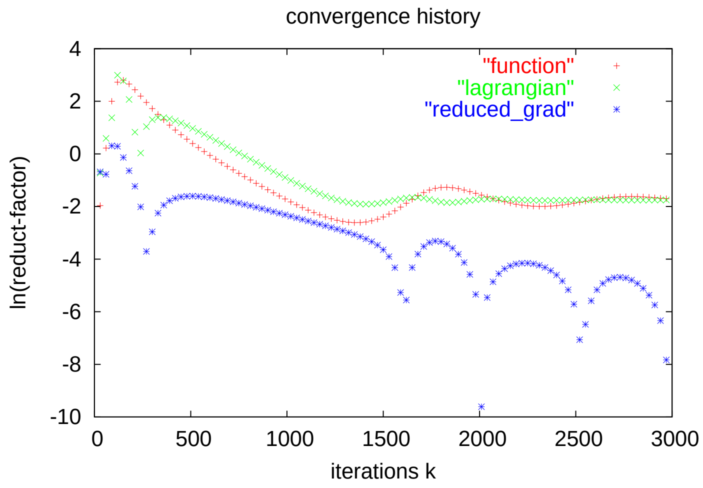
Numerical experience on test problem above:

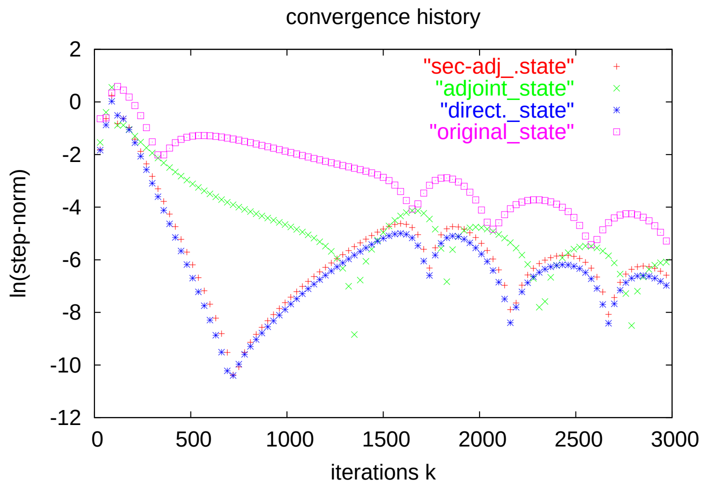
Reduced Hessian $H \equiv H(1) \quad \Longrightarrow \quad$ Immediate Blow-up

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

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

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





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.

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

is rich in *alternating* modes, i.e. eigenvectors with eigenvalues close to -1.
It makes sense that the curvature of the Lagrangian with respect to the alternting 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.

Implementation at the Software Level:

User supplied routine

input: where:

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

output:

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:

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

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:.

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

Coupled basic with first and second adjoint :

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

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

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

Simplified Implementation:

User supplied routine

input: where:

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

output:

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:

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

Coupled basic and adjoint Iteration:

$\text{init}(u);$

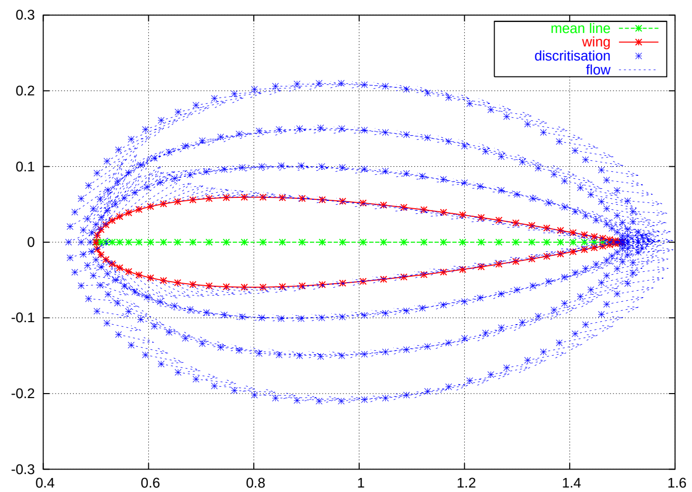
$while(???)do$

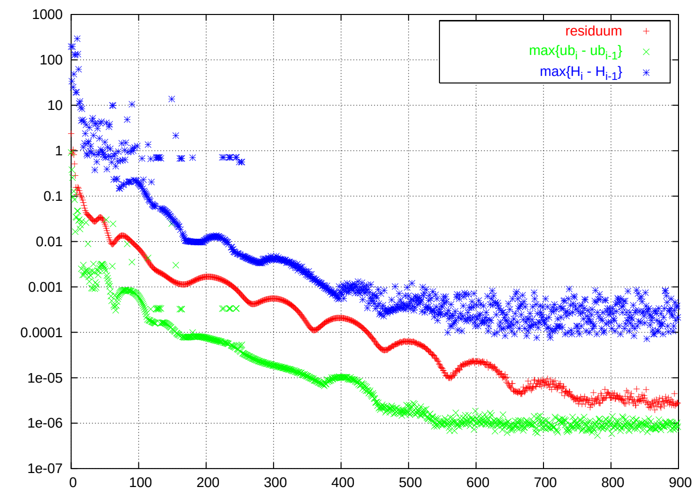
  $bu = 0; bf = 1$

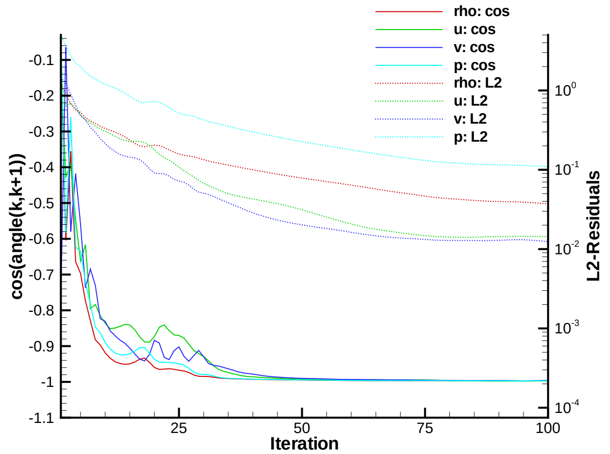
  $\text{bstep}(bu, u, by1 \dots by9, y1 \dots y9, bf, f)$

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

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







Summary and Conclusion

- 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 $\approx$ simulation step ????