Martin Weiser

Affine Invariance Concepts applied to Central Path Methods for Optimal Control

Nonlinear problem

$$F: X \to Y$$
$$F(x) = 0$$

X, Y Banach spaces

F differentible

Newton's Method

$$F'(x^k)\Delta x^k = -F(x^k)$$
$$x^{k+1} = x^k + \Delta x^k$$

Theorem (Kantorovich '52)

Assume $\exists \alpha, \beta, \gamma:$

$\|F'(x) - F'(y)\| \le \alpha \|x - y\|$ Lipschitz constant

$\|F'(x^0)^{-1}\| \le \beta$ bound on inverse

$\|F'(x^0)^{-1} F(x^0)\| \le \gamma$ length of first correction (~distance from solution)

If $\alpha\beta\gamma < \frac{1}{2}$, then $x^k \to x^*$ r–quadratically.

Nonlinear Problem

$$F: X \to Y$$
$$F(x) = 0$$

linear Transformation

$$A: Y \to Y$$
$$B: X \to X, \quad \xi = B^{-1} x$$

Transformed Problem

$$\hat{F}: X \to \hat{Y}$$
$$\hat{F}(\xi) = 0$$

Solution x^*

$$\boxed{\hat{F}(\xi) = A F(B\xi) = A F(x)}$$
$$\hat{F}'(\xi) = A F'(B\xi) B = A F'(x) B$$

Solution $\xi^* = B^{-1} x^*$

Newton's Method

$$F'(x^k) \Delta x^k = -F(x^k)$$
$$x^{k+1} = x^k + \Delta x^k$$

Newton's Method on transformed Problem

$$\hat{F}'(\xi^k) \Delta \xi^k = -\hat{F}(\xi^k)$$
$$\Rightarrow A F'(x^k) B \Delta \xi^k = -A F(x^k)$$
$$\Rightarrow B \Delta \xi^k = \Delta x^k$$

If $B\xi^0 = x^0$, then $B\xi^k = x^k \; \forall k$

Newton's Method is *invariant* under *linear* transformations.

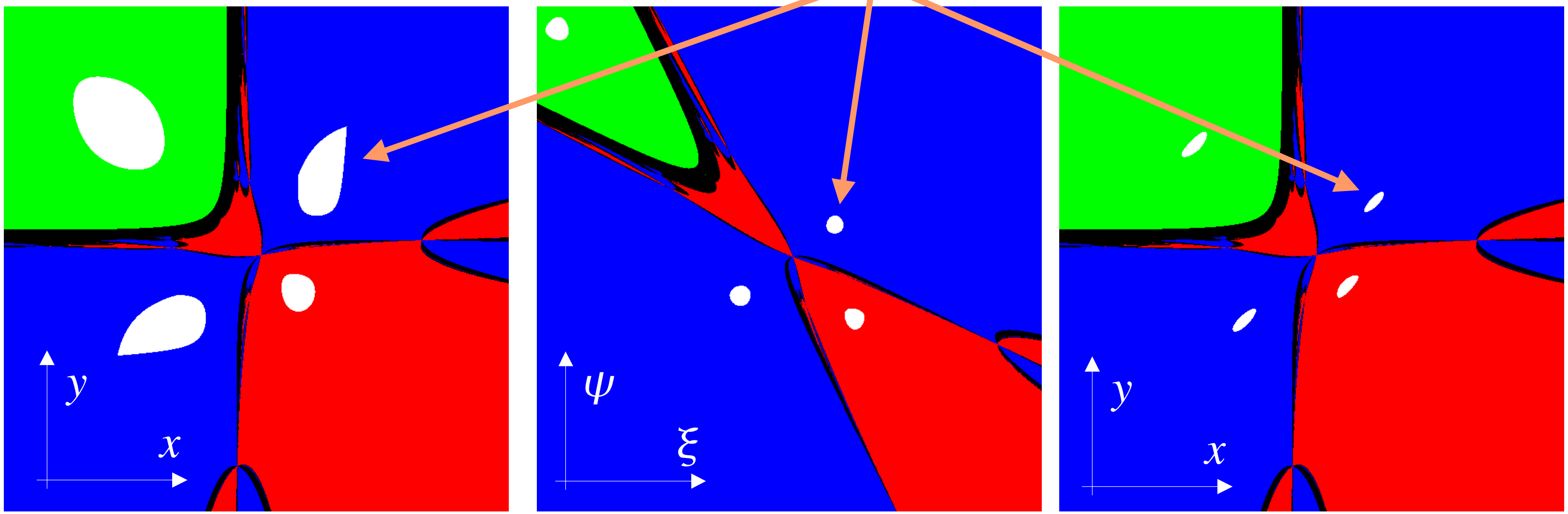
Newton's Method: Quantitative Example

$$F(x,y)=\begin{bmatrix} x^2-y^2-\frac{1}{10} \\ x^2+y^2+x-y-1 \end{bmatrix} \qquad B=\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

$$B\begin{bmatrix} \xi \\ \psi \end{bmatrix}=\begin{bmatrix} x \\ y \end{bmatrix} \qquad \hat{F}(\xi,\psi)=\begin{bmatrix} \frac{3}{4}(\xi^2-\psi^2)-\frac{1}{10} \\ \frac{5}{4}(\xi^2+\psi^2)+2\xi\psi+\frac{1}{2}(\xi-\psi)-1 \end{bmatrix}$$

convergence domains of Newton's method:

convergence domains predicted by theory



Theory and Algorithms should respect and preserve the affine invariance of Newton's method

- domain of convergence
- termination criterion
- convergence check
- damping stepsize
- control of inner iteration in inexact Newton methods

What does $\|x^k - x^*\| \rightarrow 0$ or $\|F(x^k)\| \rightarrow 0$ mean if arbitrary transformations can be applied?

Full affine invariance is impossible (we have to take norms).

Special combinations of transformations A and B can be handled.

affine covariance: $B = I$
(Deuflhard/Heindl 1979)

- invariance under arbitrary image space transformations
- convergence is measured in terms of errors $\|x^k - x^*\|$

affine contravariance: $A = I$
(Hohmann 1994)

- invariance under arbitrary domain space transformations
- convergence is measured in terms of residuals $\|F(x^k)\|$

affine conjugacy: $A = B^*$
(Deuflhard/Weiser 1995)

- invariance under arbitrary domain space transformations in convex minimization problems $F(x) = \nabla f(x)$
- $F'(x)$ is spd, transforms as $B^* F'(x) B$
- convergence is measured in terms of energy norms $\|F'(x^k)^{1/2} \Delta x^k\|$

affine similarity: $A = B^{-1}$
(Deuflhard 2001)

- invariance under simultaneous domain and image space transformations
- convergence is measured in terms of eigen norms $\|T(x^k)^{-1} F(x^k)\|$

Local Affine Invariant Norms

Affine invariant norms often depend on $F'(x)$

- affine covariance $\|F'(x^{?})^{-1}F(x)\|$
- affine conjugacy $\|F'(x^{?})^{-1/2}F(x)\|$

What is $x^{?}$?

$x^{?} = x^{*}$ theoretically convenient, but algorithmically unavailable

$x^{?} = x^{0}$ algorithmically available, but need to store $F'(x^{0})$
(suitable for simplified Newton method)

$x^{?} = x^{k}$ algorithmically convenient, but produces a local norm

Theorem (Ordinary Newton Method)

Assume $\|(F'(x)-F'(y))(x-y)\|_z \le \omega \|F'(x)(x-y)\|_x^2$ for collinear x,y,z

$\omega \|F(x^0)\|_{x^0} < 2$

+ technical assumptions

Lipschitz condition

$F'(x)^{-1}$ bounded is hidden here

Then the residuals converge to zero q–quadratically:

$$\|F(x^{k+1})\|_{x^{k+1}} \le \frac{\omega}{2} \|F(x^k)\|_{x^k}^2$$

$$\min \quad J(x)$$

$$\text{subject to} \quad c(x) = 0$$

assumption:

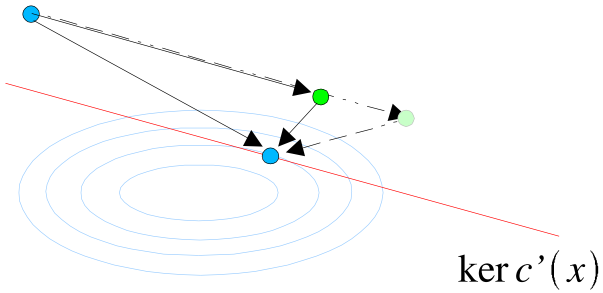
$L'' = J''(x) - c''(x)^* \lambda$ spd on $\ker c'(x)$

in neighborhood of solution

natural splitting: $X = \ker c'(x) \times S$

tangential part
L'' spd
affine conjugate

constraints part
$c'(x)$ invertible
affine contravariant



invariance under transformations of the domain space:
solution is characterized by

- feasibility $c(x^*) = 0$
- optimality $J(x^*) = \min$

how to choose S? (Orthogonality is not an invariant concept.)

Necessary KKT conditions:

$$F(x,\lambda) = \begin{bmatrix} J'(x) - c'(x)^* \lambda \\ -c(x) \end{bmatrix} = 0$$

Define affine invariant norm

$$\|F(x,\lambda)\|^2_{(x,\lambda)} = \langle F'(x,\lambda)\xi, \xi \rangle + \|l\|^2 + \|c(x)\|^2$$

where $\quad F'(x,\lambda)\begin{bmatrix} \xi \\ l \end{bmatrix} = \begin{bmatrix} J'(x) - c'(x)^* \lambda \\ 0 \end{bmatrix}$

second right hand side

Norm is invariant under domain space transformations of type $\begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix}$

Cost functional: $J(u,y) = \int_0^1 f(u(t), y(t))\, dt$

Equality constraints: $c(u,y) = \begin{bmatrix} \dot{y} - \phi(u,y) \\ r(y(0), y(1)) \end{bmatrix}$ differential equation, boundary conditions

Inequality constraints: $g_u(u) \geq 0$ control constraints

$g_y(y) \geq 0$ state constraints

Optimization problem:

$$\min J(x)$$

subject to $c(x) = 0$

$g(x) = 0$

$x = [u, y]$

Indirect method

Pontryagin's maximum principle

known switching structure

multipoint boundary value problem

multiple shooting or collocation discretization

finite–dimensional nonlinear equation

Newton's method

Direct method

multiple shooting or collocation discretization

finite–dimensional nonlinear program

KKT conditions

finite–dimensional nonlinear equation

Newton's method

Indirect method	Direct method
obtains highly accurate solutions	requires no preparatory work or a–priori knowledge
requires a–priori knowledge of switching structure	large convergence domain
requires substantial preparatory work	less accurate solutions
requires good starting point	loss of problem structure

KKT conditions,
regularized by complementarity
formulation

↓

infinite−dimensional nonlinear equation

↓

inexact Newton method
in function space

↓

infinite−dimensional linear equation

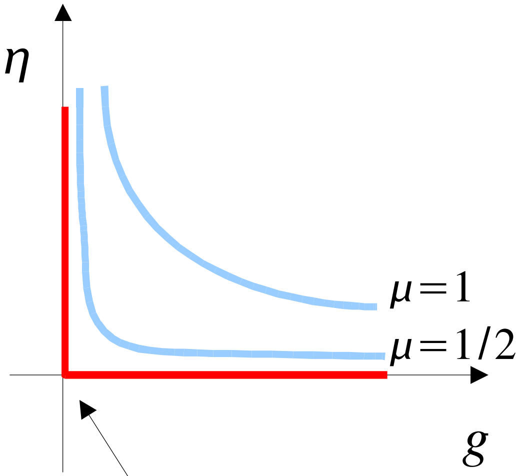
↓

multiple shooting or collocation
discretization with adaptive
mesh refinement

KKT conditions

$$\exists \lambda, \eta:$$

$$\begin{aligned} J'(x) - c'(x)^* \lambda - g'(x)^* \eta &= 0 \\ -c(x) &= 0 \\ \langle \eta, g(x) \rangle &= 0 \\ \eta, g(x) &\geq 0 \end{aligned}$$



nondifferentiable

regularization by complementarity formulation

$$F(x;\mu) = \begin{bmatrix} J'(x) - c'(x)^* \lambda - g'(x)^* \eta \\ -c(x) \\ \psi(\eta, g(x); \mu) \end{bmatrix}$$

with complementarity function

$$\psi(a,b;\mu) = \sqrt{a^2 + b^2 + 2\mu} - a - b$$

Homotopy in duality gap parameter

$$\mu \to 0$$

defines the central path.

(Maurer '81)

Nonlinear constraints and complementarity function are continuously differentiable only if

$$u, y, \lambda, \eta \in L_\infty$$

But state constraints Lagrange multipliers often contain Dirac functionals!
$\longrightarrow$ convergence in L_∞ at most for purely control constrained problems

The sufficient second order condition must be formulated in the weaker L_2 – norm:

$$\langle \partial_x^2 L(v^*)\xi, \xi \rangle \geq \alpha \|\xi\|_2^2 \quad \text{for} \quad \xi \in \ker c'(x^*) \cap \{\xi \,|\, \langle \lambda^*, \mathrm{g}'(\mathrm{x}^*)\xi \rangle \geq \beta \|\xi\|_2\}$$

where $L(v) = J(x) - \langle \lambda, c(x) \rangle - \langle \eta, g(x) \rangle$ is the Lagrange functional

Let $v = [x, \lambda, \eta]$ denote all variables.

Theorem (Existence of the central path on open inverval $\mu \in (0, \mu_0]$)

Assume

- f, ϕ, r, g are twice continuously differentiable
- the linearized state equation is solvable:

$$\forall \delta : \exists \delta y : \quad \partial_y c(x) \delta y = \delta \quad , \quad \| \delta y \| \leq \text{const} \| \delta \|$$

- the strengthened Legendre–Clebsch condition holds:

with $$H = f''(x) - \phi''(x)^T \lambda - g''(x)^T \eta + g'(x)^T \partial_\eta \psi(g, \eta)^{-1} \partial_g \psi(g, \eta) g'(x)$$

holds $\xi(t)^T H(t) \xi(t) \geq \text{const} |\xi(t)|^2$ for $\xi \in \ker c'(x)$

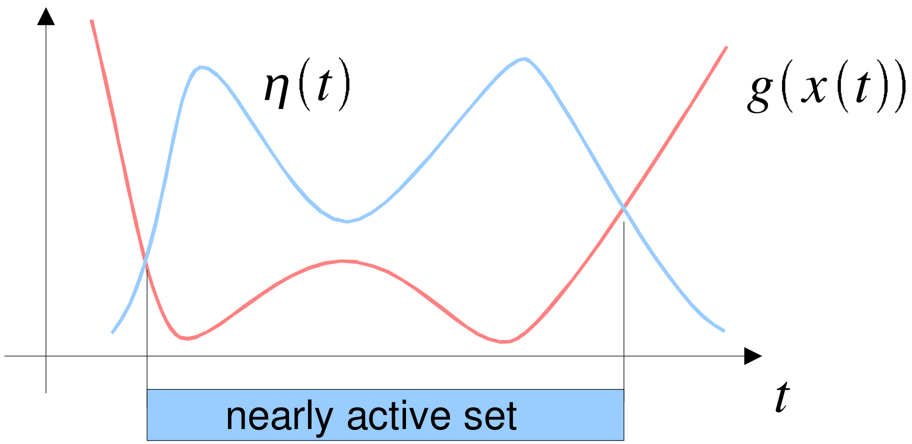
- there exists a starting point v_0 with $F(v_0; \mu_0) = 0$

Then there is a path $v(\mu)$ with $F(v(\mu); \mu) = 0$ for $\mu \in (0, \mu_0]$

Nearly Active Constraints

Nearly active set $\Omega_\rho(x) = \{t \in [0,1] : g(x(t)) \leq \rho\, \eta(t)\}$

$$\Omega_\rho^c = [0,1] \setminus \Omega_\rho$$



splitting of inequality constraints $g(x) = \begin{bmatrix} g_\rho(x) \\ g_\rho^c(x) \end{bmatrix} \qquad \eta = \begin{bmatrix} \eta_\rho \\ \eta_\rho^c \end{bmatrix}$

Assume $\partial_x^2 L(v)$ spd on $\ker c'(x) \cap \ker g_\rho{}'(x)$ ← similar to sufficient second order condition

Theorem (Existence of the central path on closed inverval $\mu \in [0, \mu_0]$)

Assume
- only control constraints
- the operator $C = \begin{bmatrix} c'(x) \\ g_\rho'(x) \end{bmatrix}$ of nearly active constraints

  satisfies the inf–sup–condition
- the modified Hessian $H = \partial_x^2 L + (g_\rho^c{}')^* (\partial_\eta \psi_\rho^c)^{-1} \partial_g \psi_\rho^c g_\rho^c{}'$

  is positive definite on $\ker C$
- the central path exists on $(0, \mu_0]$

Then the path can be continuously extended (w.r.t. L_2–norm) to include a KKT point $v(0)$

Affine invariant norm:

$$\|F(v;\mu)\|^2_{(x,\mu)} = \|c(x)\|^2 + \|\psi(g(x),\eta;\mu)\|^2 + \langle L''(x)\xi,\xi\rangle + \|l\|^2 + \|e\|^2$$

L_2–norm part!

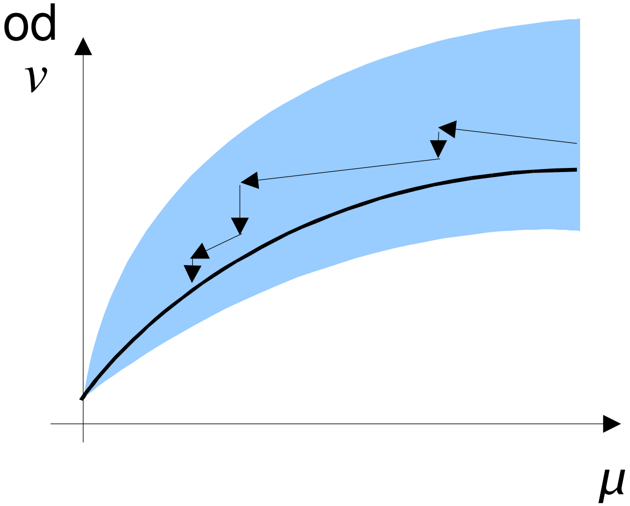
where

$$F'(v;\mu)\begin{bmatrix}\xi\\ l\\ e\end{bmatrix} = \begin{bmatrix}J'(x) - c'(x)^*\lambda - g'(x)^*\eta\\ 0\\ 0\end{bmatrix}$$

assumes $L''(x)$ spd on $F'(v;\mu)^{-1}(X^*\times\{0\}\times\{0\}) \approx \ker c'(x) \cap \ker g_\rho{}'(x)$

For L_∞ –norm replace $\langle L''(x)\xi,\xi\rangle$ by $\sup\limits_{t\in[0,1]} |\langle L''(x)(t)\xi(t),\xi(t)\rangle|$

Employ a Newton type predictor–corrector method in function space



Reduction of discretization error expensive

→ inexact Newton corrector
inexact tangential predictor

→ corrector does not reach path
predictor is not tangential to path

→ new stepsize selection scheme

inexact Newton corrector

$$F'(v^k)\,\delta v^k = -F(v^k) + r^k$$
$$v^{k+1} = v^k + \delta v^k$$

inexact tangential predictor

$$F'(v;\mu)\,p = -\partial_\mu F(v;\mu) + r$$

Theorem (Inexact Newton Method)

Lipschitz condition

Assume $\|(F'(v)-F'(w))(v-w)\|_z \le \omega \|F'(v)(v-w)\|_v^2$ for collinear v,w,z

$\|r\|_w \le (1+\gamma_v \|F'(v)(w-v)\|_v)\|r\|_v$ for all r,v,w

$$\boxed{\frac{\omega}{2}(1+\delta_k)^2\|F(v^k)\|_{v^k} + (1+(1+\delta_k)\gamma_v\|F(v^k)\|_{v^k})\,\delta_k \le \Theta < 1}$$

where $\delta_k = \dfrac{\|r^k\|_{v^k}}{\|F(v^k)\|_{v^k}}$ for all k

accuracy requirement for inner iteration

Then the residuals converge q–linearly to zero:

$$\boxed{\|F(v^{k+1})\|_{v^{k+1}} \le \Theta \|F(v^k)\|_{v^k}}$$

and $\|F(v^{k+1})\|_{v^k} \le \left(\delta_k + \dfrac{\omega}{2}(1+\delta_k)^2\|F(v^k)\|_{v^k}\right)\|F(v^k)\|_{v^k}$

Termination criterion for inner iteration
requires unknown constants ω, γ

→ derive efficiently computable estimates to be inserted in accuracy matching condition

$$[\omega]_k = \frac{2}{(1+\delta_k)^2 \|F(v^k)\|_{v^k}} \left(\frac{\|F(v^{k+1})\|_{v^k}}{\|F(v^k)\|_{v^k}} - \delta_k \right) \leq \omega$$

$$[\gamma_x]_k = \frac{\left| \|F(v^{k+1})\|_{v^{k+1}} - \|F(v^{k+1})\|_{v^k} \right|}{\|F(v^k)\|_{v^k}} \leq \gamma$$

Theorem (inexact tangential predictor)

Assume $\|F(v_2;\tau_2)\|_{(v_2,\tau_2)} \le (1+\gamma_\tau(\tau_2-\tau_1))\|F(v_2;\tau_2)\|_{(v_1,\tau_1)}$

$$\|F(v_2;\tau_2)\|_{(v_1,\tau_1)} \le \|F(v_1;\tau_1)\|_{(v_1,\tau_1)} + \|r\|_{(v_1,\tau_1)}(\tau_2-\tau_1) + \beta(\tau_2-\tau_1)^2$$

for $v_1, v_2, \tau_1, \tau_2, r$ such that

$$F'(v_1;\tau_1)(v_2-v_1) = -(\tau_2-\tau_1)(\partial_\mu F(v_1;\tau_1)+r)$$

Then the inexact Newton corrector converges for all stepsizes $\Delta\tau = \tau_2-\tau_1$ satisfying

$$(1+\gamma_\tau\Delta\tau)\left(\|F(v_1;\tau_1)\|_{(v_1,\tau_1)} + \|r\|_{(v_1,\tau_1)}\Delta\tau + \beta\Delta\tau^2\right) < \frac{2}{\omega}$$

Stepsize selection criterion needs unknown constants γ_τ, β

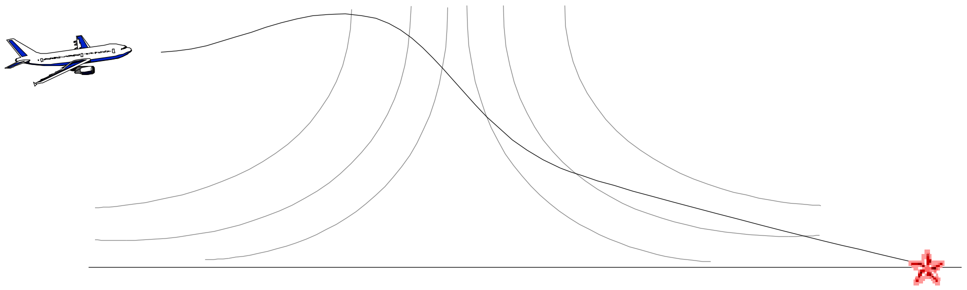
→ define efficiently computable estimates to be inserted in stepsize formula

$$[\gamma_\tau] = \frac{1}{\Delta\tau}\left|\frac{\|F(v_2;\tau_2)\|_{(v_2,\tau_2)}}{\|F(v_2;\tau_2)\|_{(v_1,\tau_1)}} - 1\right| \leq \gamma_\tau$$

$$[\beta] = \frac{1}{\Delta\tau^2}\left|\|F(v_2;\tau_2)\|_{(v_1,\tau_1)} - \|F(v_1;\tau_1)\|_{(v_1,\tau_1)}\right| - \frac{\|r\|_{(v_1,\tau_1)}}{\Delta\tau} \leq \beta$$

Windshear Situation

Descending column of air leads to subsequential encounter of
- headwind
- downburst
- tailwind



Low altitude windshear is a dangerous situation due to
- loss of altitude in the downburst region
- loss of lift in the tailwind region

Flight crew's typical reaction: decrease engine power and/or flight path angle to compensate additional lift in headwind region

Fatal crash possible if settings remain unchanged during downburst and tailwind

At least 30 aircraft accidents attributed to windshear
(New Orleans '82, Denver '85)

If initial altitude is sufficiently high, abort landing is safer than penetration landing

(Miele et al. '87)

$$\max \min h$$

subject to

$$|u| \le u_{\max}$$
$$\alpha \le \alpha_{\max}$$

$$\dot{x} = V \cos\gamma + W_x$$
$$\dot{h} = V \sin\gamma + W_h$$
$$\dot{V} = \frac{T}{m}\cos(\alpha+\delta) - \frac{D}{m} - g\sin\gamma - \dot{W}_x \cos\gamma - \dot{W}_h \sin\gamma$$
$$\dot{\gamma} = \frac{1}{V}\left(\frac{T}{m}\sin(\alpha+\delta) + \frac{L}{m} - g\cos\gamma + \dot{W}_x \sin\gamma - \dot{W}_h \cos\gamma\right)$$
$$\dot{\alpha} = u$$

x	horizontal distance	u	angle of attack rate (control)
h	altitude	$W_x(x,h)$	horizontal wind velocity
V	velocity	$W_h(x,h)$	vertical wind velocity
γ	path inclination angle	$T(h,V,t)$	thrust
m	aircraft mass	$D(h,V,\alpha)$	drag
g	acceleration of gravity	$L(h,V,\alpha)$	lift
α	angle of attack		

Initial position given, terminal position with specified path inclination.

Abort Landing in Presence of Windshear: Results

Extremely difficult problem due to

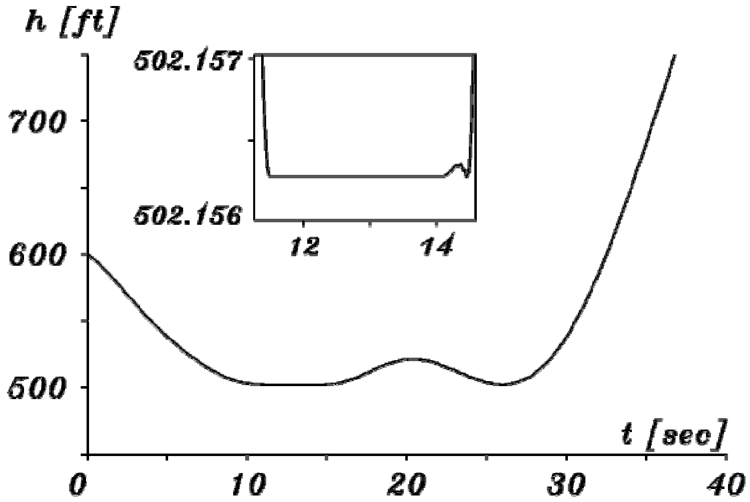
- high nonlinearity
- third order state constraint
- complicated switching structure
- multiple touch points
- state constrained subarcs
- control constrained subarcs
- singular subarcs

Indirect method requires an immense amount of preparatory analytical work.

Direct function space method obtains competitive solution with virtually no preparation.

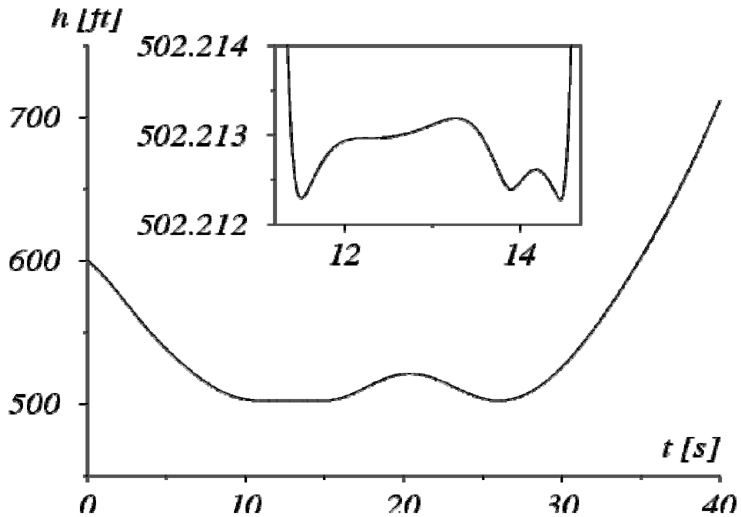
(Bulirsch/Montrone/Pesch '91)
(Weiser/Deuflhard '01)

Altitude above ground vs. time.



indirect method

Solution features a boundary arc and two touch points.



direct function space method

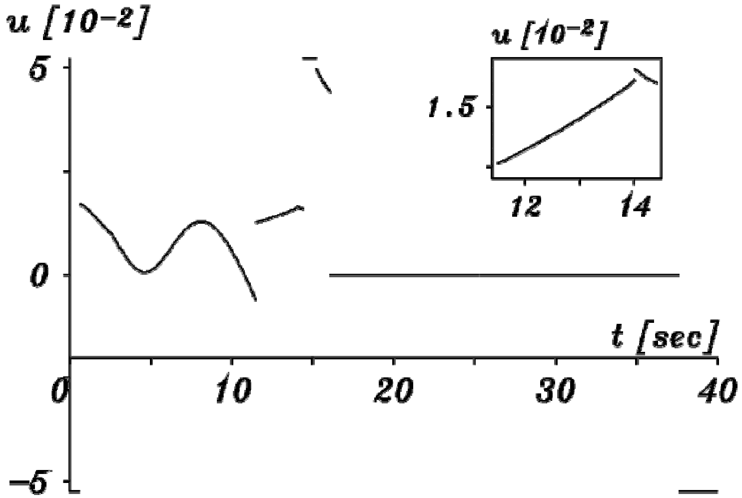
After the last touch point of minimal altitude there is a continuum of optimal solutions.

(Bulirsch/Montrone/Pesch '91)
(Weiser/Deuflhard '01)

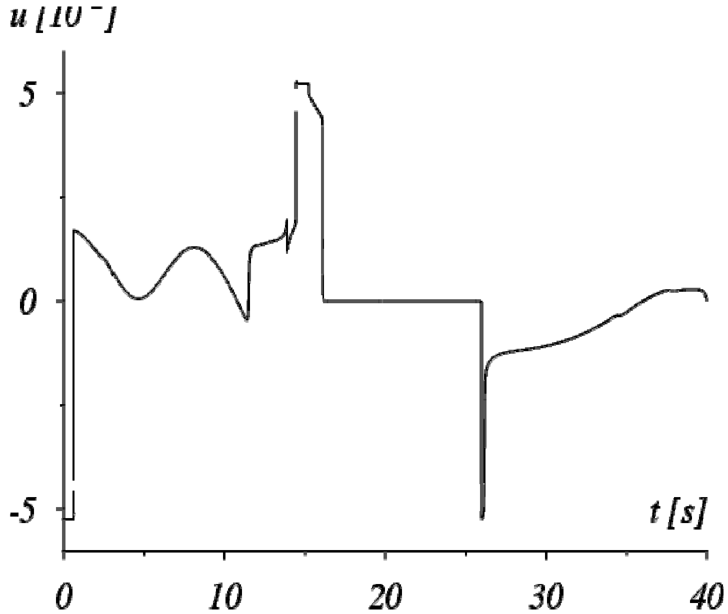
Angle of attack rate (control) vs. time.

Solution features a complicated switching structure and constrained as well as singular subarcs.

After the last touch point of minimal altitude there is a continuum of optimal solutions.



indirect method



direct function space method

Mesh refinement levels versus time.

Mesh refinement is done by inner iteration of the Newton corrector in order to meet accuracy requirement when solving linear problems.