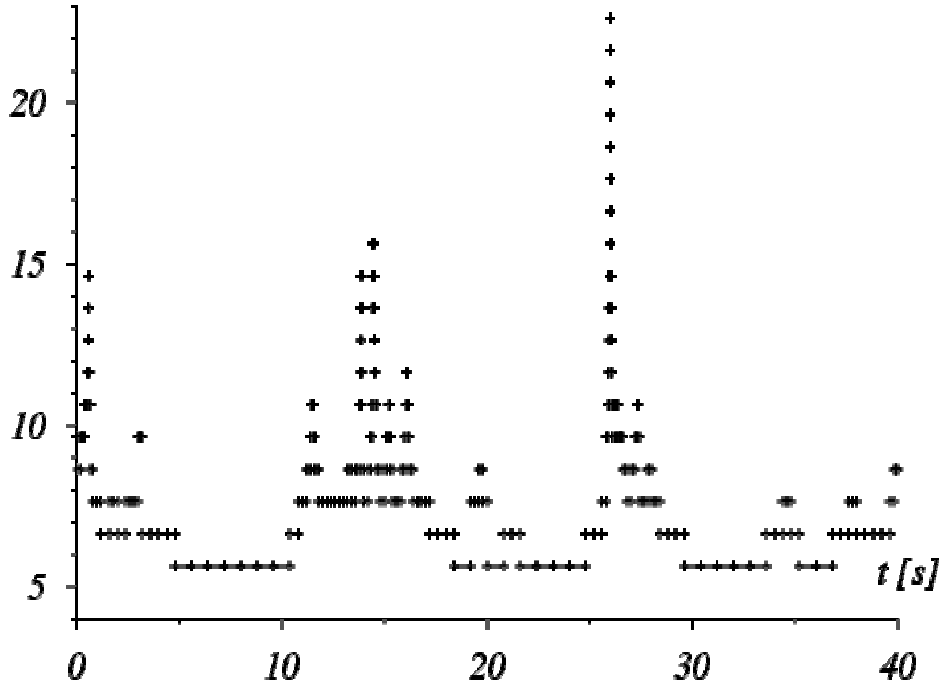


Martin Weiser

Peter Deuflhard

An Adaptive Function Space Interior Point Method for Optimal Control Problems

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

Equality constraints: $c(u,y)=\begin{bmatrix} \dot{y}-\phi(u,y) \\ r(y(0),y(1)) \end{bmatrix}$ BVP for ODEs

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

$g_y(y)\geq 0$ state constraints

Optimization problem:

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

$x=[u,y]$

Indirect methods:
- highly accurate solutions
- small convergence domain
- require immense amount of preparatory work (manpower)

Derive necessary conditions separately for:
- control constrained subarcs
- singular control subarcs
- state constrained subarcs of different order
- touch points
- interior point conditions

Guess switching structure
Update switching structure if incorrect

Direct methods:
- less accurate solutions
- larger convergence domain
- almost no preparatory work

Aim:
- reduce (wo)manpower required
- obtain highly accurate solutions nevertheless
- prospect of extension to PDEs

KKT conditions

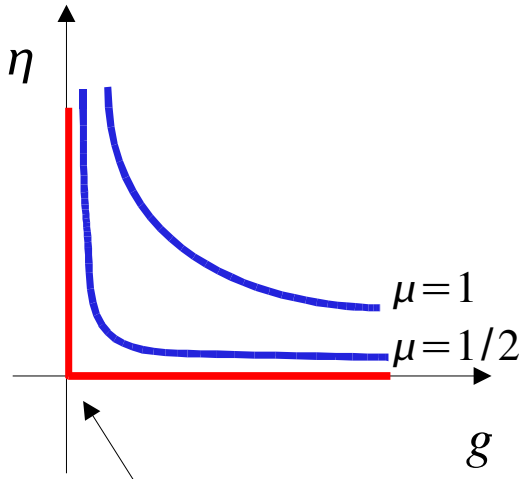
$$\exists \lambda, \eta:$$

$$J'(x) - c'(x)^* \lambda - g'(x)^* \eta = 0$$

$$-c(x) = 0$$

$$\langle \eta, g(x) \rangle = 0$$

$$\eta, g(x) \geq 0$$



nondifferentiable

regularization by complementarity formulation

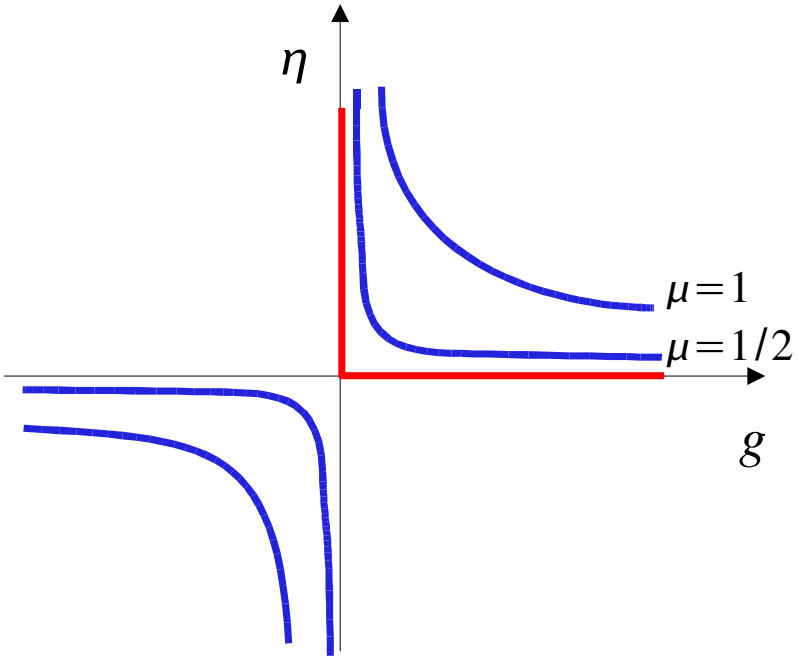
$$F(x;\mu) = \begin{bmatrix} J'(x) - c'(x)^* \lambda - g'(x)^* \eta \\ -c(x) \\ \psi(\eta, g(x); \mu) \end{bmatrix}$$

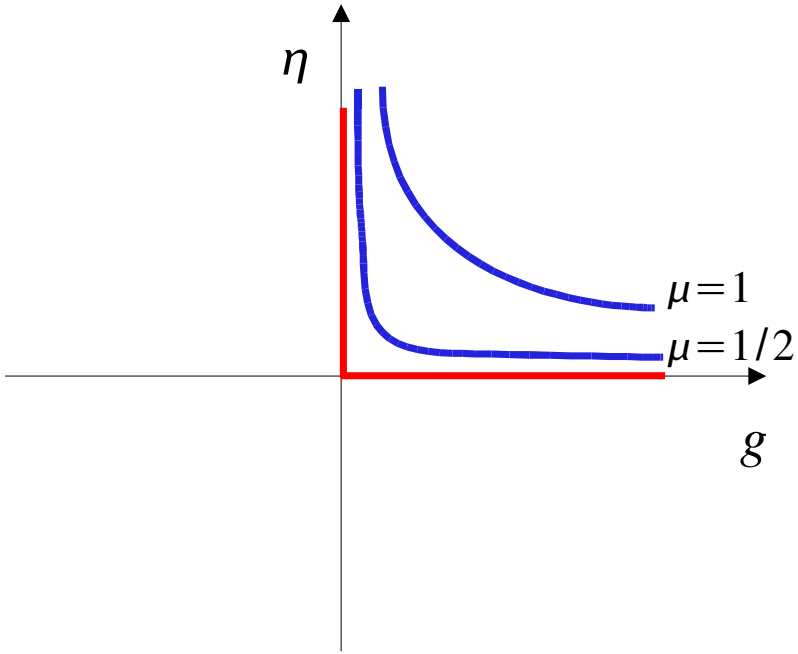
Homotopy in duality gap parameter

$$\mu \to 0$$

defines the central path.

Feasible vs. Infeasible Methods





classical interior point formulation

$$\psi(a,b;\mu)=ab-\mu$$

feasible iterates required

typical for linear problems

complementarity formulation

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

infeasible iterates permitted

advantages for nonlinear problems

The Two Norm Discrepancy

(Maurer '81)

- Nonlinearity in constraints and complementarity function

  Newton iteration requires continuous differentiability

$$c(u,y)=0$$

$$c\colon L_\infty \times W^{1,\infty} \to L_\infty$$

- Sufficient second order condition (coercivity) requires Hilbert space structure:

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

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

The Central Path: Theory I

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 sufficiently smooth (C^2)
- linearized state equation is solvable
- strengthened Legendre–Clebsch condition (pointwise):

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

with $\quad H=\partial_x^2 L(x)+(g')^*\,\partial_\eta\psi^{-1}\,\partial_g\psi\,g'$

- existence of starting point v_0 with $F(v_0;\mu_0)=0$

Then there exists a path $v(\mu)$ with $F(v(\mu);\mu)=0$ for $\mu\in]0,y_0]$

Factorize $\begin{bmatrix} \partial_x^2 L & C^* & G^* \\ C & & \\ G & & -D \end{bmatrix}$ by elimination via $D^{-1} = (g')^* \, \partial_\eta \psi^{-1} \, \partial_g \psi \, g'$

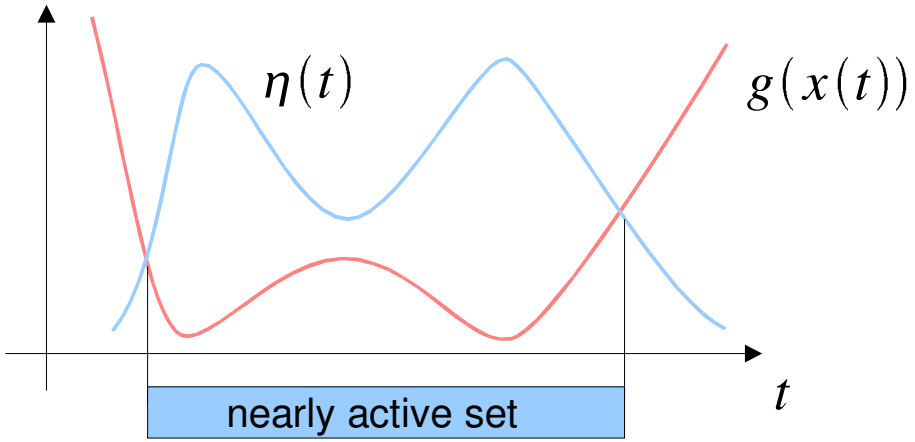
Drawback: D becomes ill–conditioned for $\mu \to 0$

Proof technique useless for quantitative results

Nearly Active Constraints

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

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



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

Assume $\partial_x^2 L(v)$ spd on $\ker c'(x) \cap \ker g_\rho'(x)$

similar to sufficient second order condition

The Central Path: Theory II

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

Assume

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

  satisfies the inf–sup–condition

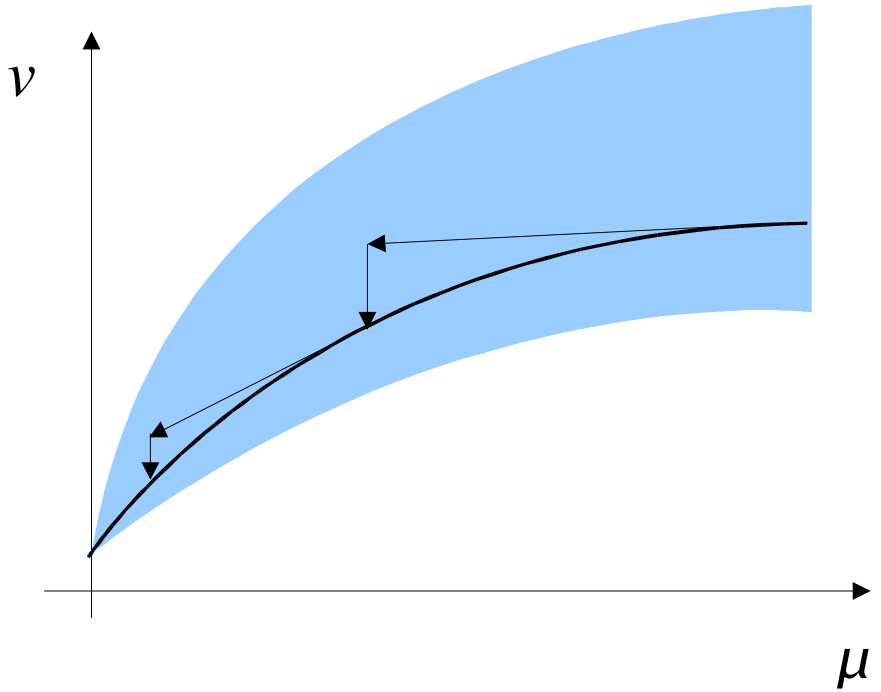
  constraints qualification
- the modified Hessian $H = \partial_x^2 L + (g_\rho^{c\,\prime})^* (\partial_\eta \psi_\rho^c)^{-1} \partial_g \psi_\rho^c g_\rho^{c\,\prime}$

  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 $v(0)$

The Central Path: Continuation

Employ a Newton type predictor–corrector method



Adaptive stepsize control (Deuflhard '78)

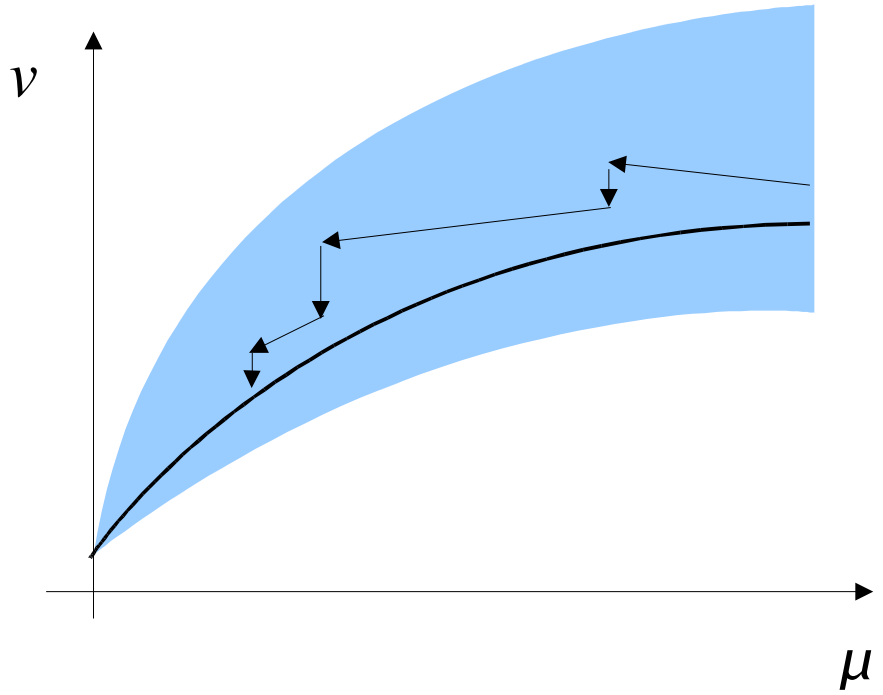
assumes:

- corrector reaches the path
- predictor is exactly tangential

The Central Path: Inexact Continuation

Reduction of discretization error (adaptive mesh refinement) expensive:

- corrector does not reach the path
- predictor not exactly tangential



Stepsize selection scheme needs to be extended

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

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

Lipschitz condition

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

Lipschitz continuity of local norm

$$\delta_k := \frac{\|r^k\|_{v^k}}{\|F(v^k)\|_{v^k}} \quad \text{sufficiently small such that}$$

$$\Theta(\delta_k;\omega,\gamma_v) \le \bar{\Theta} < 1$$

accuracy requirement for inner iteration

where $$\Theta(\delta_k;\omega,\gamma_v) := \frac{\omega}{2}(1+\delta_k)^2\|F(v^k)\|_{v^k} + (1+(1+\delta_k)\gamma_v\|F(v^k)\|_{v^k})\delta_k$$

Then

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

Accuracy requirement uses unknown constants ω, γ_v

Idea (Deuflhard '78): use computable estimates instead

$$[\omega]_k = 2\frac{\|F(v^{k+1}) - r^k\|_{v^k}}{\|F'(v^k)\delta v^k\|_{v^k}^2} \le \omega$$

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

For reliable estimates use problem–adapted affine invariant norm.
(Deuflhard '79, Volkwein/Weiser '00)

The Central Path: Inexact Tangential Predictor

Theorem

$$\tau = -\log(\mu) \qquad \Delta\tau = \tau_2 - \tau_1$$

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

curvature of central path

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

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

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

Lipschitz continuity of local norm

Then the inexact Newton corrector converges for all stepsizes $\Delta\tau$ 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

Again: use computable estimates

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

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

Stepsize Reduction

Nested stepsize reduction scheme:

1. compute steplength $\Delta\tau([\omega],[\gamma_v],[\beta],[\gamma_\tau],\rho)$ safety factor $\rho<1$

2. check $\|F(v_{p;}\tau_p)\|_v \le \rho^2 TOL_p$

   update $[\beta]$

   check failed: go to 1

3. check $\|F(v_p;\tau_p)\|_{v_p} \le (1+[\gamma_\tau]\Delta\tau\|F_v(v;\tau)p\|_v)\rho\, TOL_p$

   update $[\gamma_\tau]$

   check failed: go to 1

4. check corrector convergence

   update $[\omega],[\gamma_v]$ iteration terminates

   check failed: go to 1

How many corrector steps?

maximize $Q := \dfrac{\Delta\tau}{\text{work}}$ (information gain per work)

depends on
- predictor tolerance TOL_p
- corrector tolerance TOL_c
- relative accuracy δ_k

simplistic model yields:

$$TOL_p \approx \frac{1}{\omega}$$ (half of corrector's convergence radius)

$$TOL_c \approx 0.6\, TOL_p$$ (one corrector step)

$$\delta_k = 0.1$$

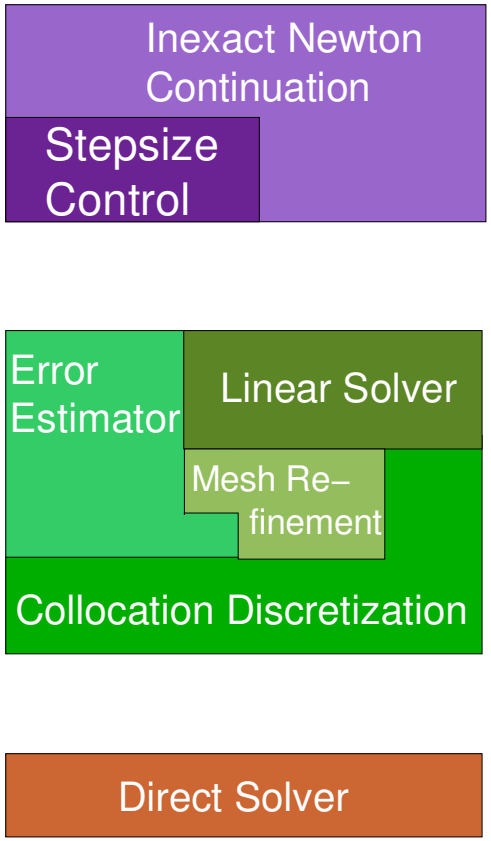
Algorithm Prototype

- inexact Newton continuation in function space
- adaptive mesh refinement via Newton corrector:

  relative discretization error $\delta_k := \dfrac{\|r^k\|_{v^k}}{\|F(v^k)\|_v^k}$

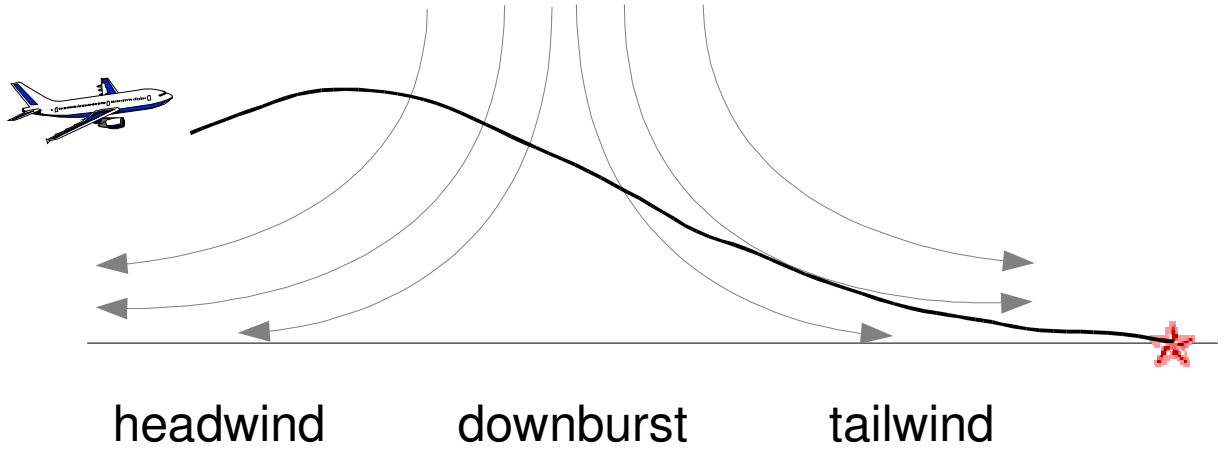
  small enough such that $\Theta(\delta_k; [\omega], [\gamma_v]) \le \bar{\Theta} < 1$

- higher order collocation ($p=2$ up to 6)
- direct solver (band/sparse)



Low Altitude Windshear Situation

- loss of altitude
- loss of lift



Flight crew's reaction often fatal

At least 30 aircraft accidents attributed to windshear
(New Orleans '82, Denver '85, Zürich '01?)

Abort Landing in Presence of Windshear: Model

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

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.

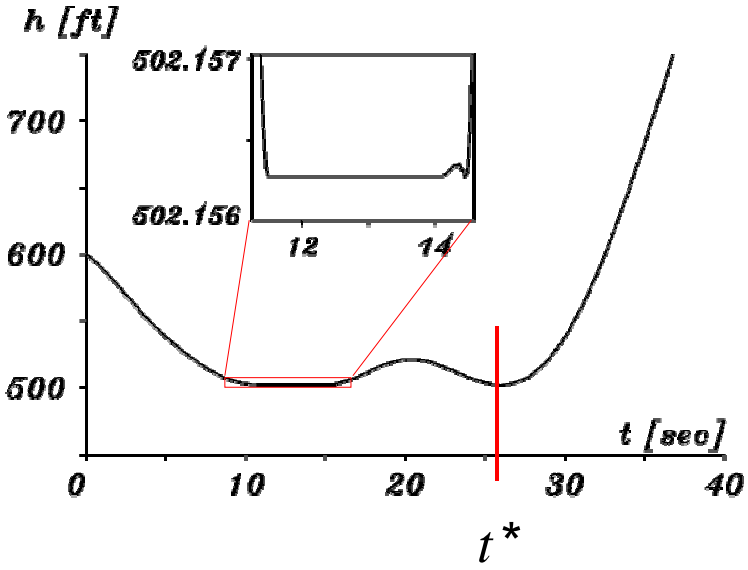
Direct function space method obtains competitive solution with virtually no preparation.

Abort Landing in Presence of Windshear: Results

Altitude above ground

continuum of solutions
for $t > t^*$

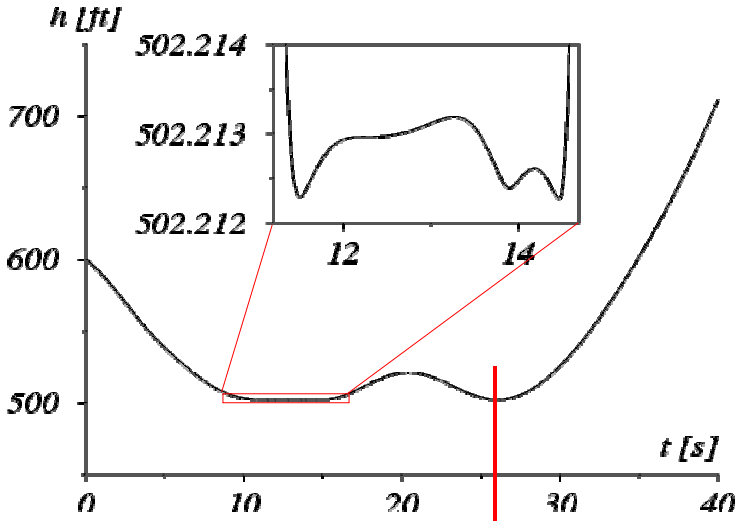
(Bulirsch/Montrone/Pesch '91)



t^*

indirect method
(multiple shooting)

(Weiser/Deuflhard '01)



t^*

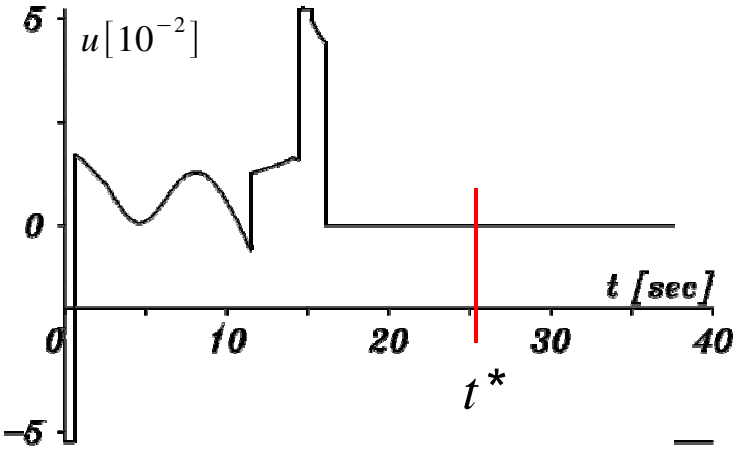
direct function
space method

Abort Landing in Presence of Windshear: Results

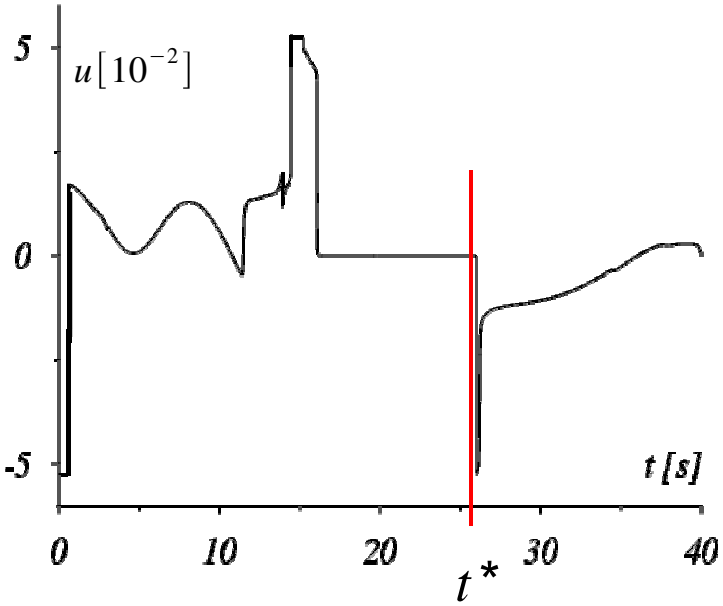
Control
(angle of attack rate)

continuum of solutions
for $t > t^*$

(Bulirsch/Montrone/Pesch '91)



indirect method
(multiple shooting)

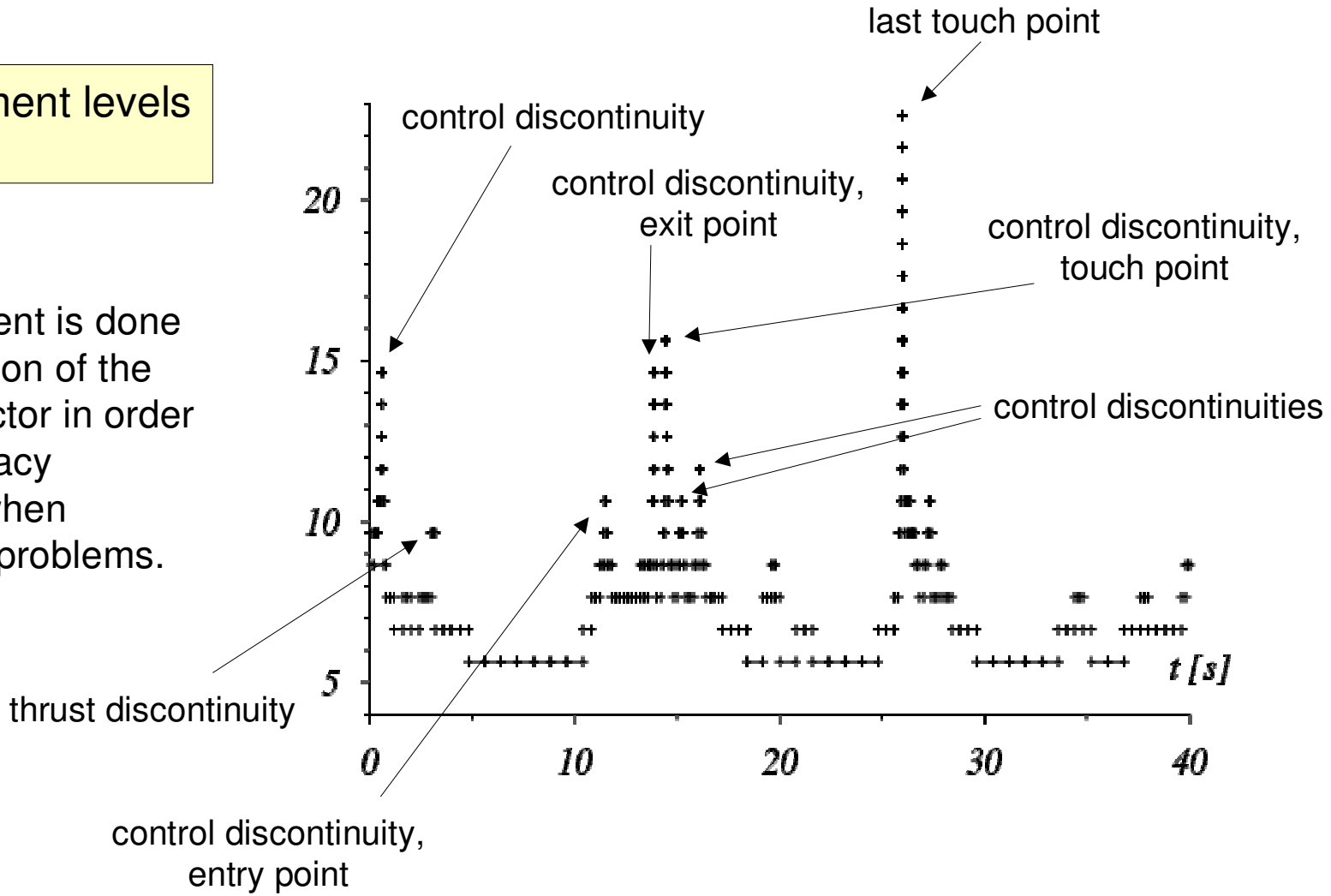


direct function
space method

(Weiser/Deuflhard '01)

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.



Function Space Central Path Methods:

- substantial reduction of (wo)manpower required for solving intricate optimal control problems to high accuracy
- sound theory only for limited situations
- numerical implementation not (yet) tuned

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)

$\exists\, \alpha, \beta, \gamma:$ $\quad \|F'(x) - F'(y)\| \le \alpha \|x - y\|$ 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.

Newton's Method: Affine Invariance

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^*

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

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

$$\min \eta^2+(\xi+\eta)^3-\xi-\eta \quad \text{s.t.} \quad \xi+2\eta=0$$

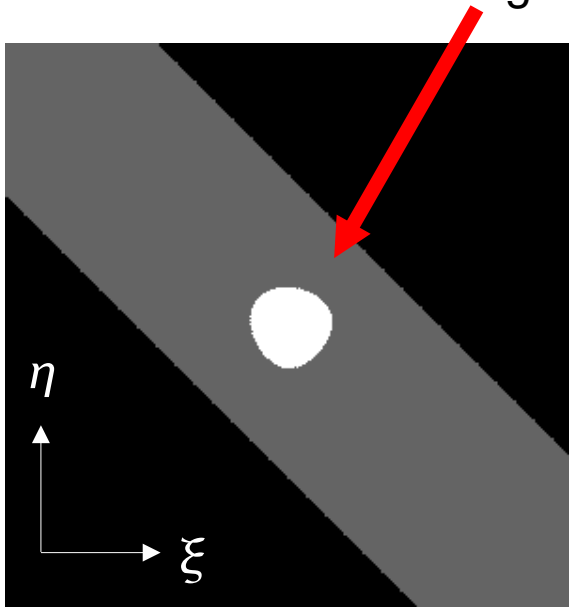
$$\alpha \approx 1.95$$
$$\beta = 12\sqrt{2}$$

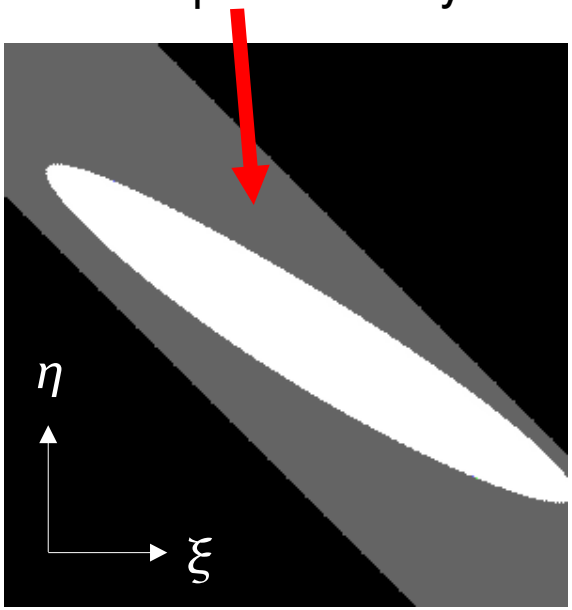
$$\xi = 2\theta - \phi$$
$$\eta = \phi - \theta$$

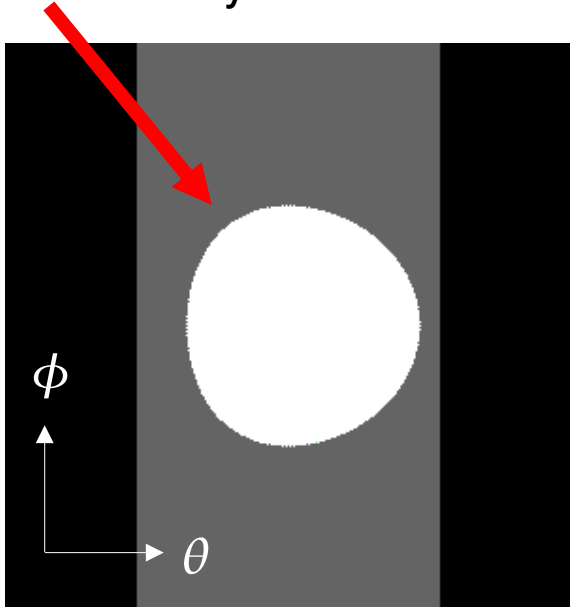
$$\min (\phi-\theta)^2+\theta^3-\theta \quad \text{s.t.} \quad \phi=0$$

$$\alpha \approx 1.86$$
$$\beta = 6$$

convergence domain predicted by Kantorovich theory







Affine Invariant Norms

- Theory and Algorithms should preserve invariance
  - domain of convergence
  - termination criterion
  - convergence check
  - damping
  - accuracy matching

formulate in terms of
local affine invariant norms!

- Full affine invariance impossible

- Invariance w.r.t. restricted transformations possible

min $J(x)$

subject to $c(x)=0$

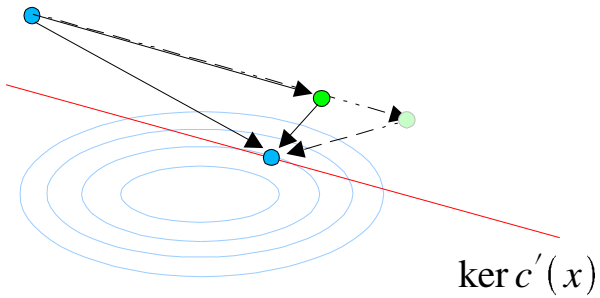
assumption:

$\partial_x^2 L = J''(x) - c''(x)^* \lambda$ spd on $\ker c'(x)$

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

tangential part
$\partial_x^2 L$ spd

constraints part
$c'(x)$ invertible



invariance under transformations of the domain space:

solution is characterized by

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

how to choose S?

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

second right hand side

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

Affine invariant norm:

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

where

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

L_2–norm part!

assumes $\partial_x^2 L(x)$ spd on $F'(v;\mu)^{-1}(X^*\times\{0\}\times\{0\}) \approx \ker c'(x) \cap \ker g'_\rho(x)$

For L_∞ –norm replace $\langle \partial_x^2 L(x)\xi,\xi\rangle$ by $\sup_{t\in[0,1]} |\langle \partial_x^2 L(x)(t)\xi(t),\xi(t)\rangle|$