Interior-Point Methods as Inexact Newton Methods

Silvia Bonettini
Università di Modena e Reggio Emilia
Italy

- Valeria Ruggiero – Università di Ferrara
- Emanuele Galligani – Università di Modena e Reggio Emilia
- Carla Durazzi – Università di Ferrara

A Nonlinear Programming Problem

min $f(x)$

s.t.

$g_1(x) = 0$

$g_2(x) \geq 0$

$\longleftarrow \quad f \quad : \mathbb{R}^n \to \mathbb{R}$

$\longleftarrow \quad g_1 \quad : \mathbb{R}^n \to \mathbb{R}^{neq}$

$\longleftarrow \quad g_2 \quad : \mathbb{R}^n \to \mathbb{R}^m$

$$\Updownarrow$$

min $f(x)$

s.t.

$g_1(x) = 0$

$g_2(x) - s = 0$

$s \geq 0$

$\longleftarrow \quad s \in \mathbb{R}^m$ Slack variables

Introduce Lagrange multipliers for equality and inequality constraints

$$\lambda \in \mathbb{R}^{neq}, \qquad w, z \in \mathbb{R}^m$$

Lagrangian function: $L(x, \lambda, w, s, z) = f(x) - \lambda^t g_1(x) - w^t(g_2(x) - s) - z^t s$

Karush-Kuhn-Tucker Optimality Conditions

$$\begin{array}{lllll} L_x = & \nabla f(x) - \lambda^t \nabla g_1(x) - w^t g_2(x) & = & 0 & n \\ L_\lambda = & -g_1(x) & = & 0 & neq \\ L_w = & -g_2(x) + s & = & 0 & m \\ & SWe_m & = & 0 & m \\ & s, w \geq 0 & & & \end{array}$$

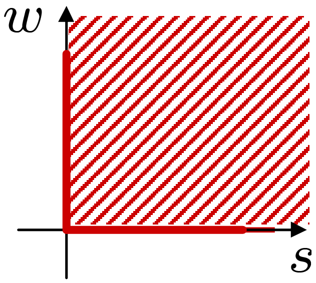
$$L_s = w - z = 0 \rightarrow w = z$$

Notations: $S = diag(s)$ $W = diag(w)$ $e_m = (1, ..., 1)^t \in \mathbb{R}^m$

Complementarity conditions: $\boxed{s_i w_i = 0 \quad i = 1, \dots, m}$

Define a new variable $v \in \mathbb{R}^{n+neq+2m}$ $v = (x^t, \lambda^t, w^t, s^t)^t$

$$\boxed{\begin{array}{c} H(v) = \begin{pmatrix} H_1(v) \\ SWe_m \end{pmatrix} = 0 \\ s, w \geq 0 \end{array}}$$



Standard Assumptions

1. EXISTENCE

   $\exists v_* \text{ s.t. } H(v_*) = 0 \text{ and } s_*, w_* \geq 0$

2. SMOOTHNESS

   The hessian matrices $\nabla^2 f, \nabla^2 (g_1)_i, \nabla^2 (g_2)_i$ exist and are Lipschitz continuous in a neigborhood of v_*

3. REGULARITY

   The gradients of the equality contraints and of the active inequality constraints, in the solution are linearly independent.

4. STRICT COMPLEMENTARITY

   $(w_*)_i + (s_*)_i > 0 \quad i = 1, \ldots, m$

5. SECOND ORDER SUFFICIENCY

   The matrix $\nabla^2 L_{xx}(v_*)$ is positive definite on the space

$$N(x_*) = \{y \in \mathbb{R}^n : \nabla (g_1)_i(x_*)y = 0, \quad \nabla (g_2)_j(x_*)y = 0, \text{ where } i = 1, \ldots, neq \text{ and } (g_2)_j \text{ is active in } x_*\}$$

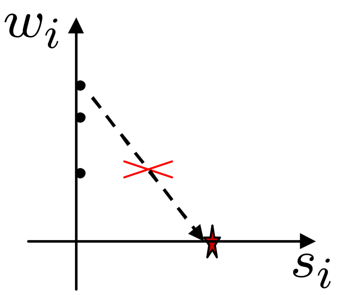
NONLINEAR PROGRAMMING PROBLEM

NONLINEAR SYSTEM WITH BOUNDS ON SOME VARIABLES

Solving the system with Newton's method…

$$H'(v_k)\Delta v_k = -H(v_k)$$

$$S_k \Delta w_k + W_k \Delta s_k = -S_k W_k e_m$$

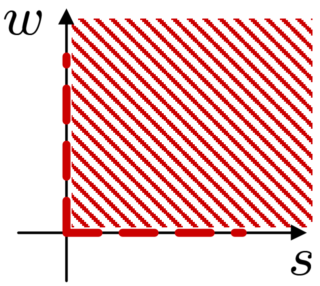


If $(s_k)_i = 0$ [or $(w_k)_i = 0$], then $(s_{k+j})_i = 0$ [or $(w_{k+j})_i = 0$]

$j = 1, 2, \dots$

Perturbe the complementarity conditions:

$$\boxed{\begin{array}{c} H(v) = \begin{pmatrix} H_1(v) \\ SWe_m \end{pmatrix} = \begin{pmatrix} 0 \\ \rho e_m \end{pmatrix} = \rho \tilde{e} \\ s, w > 0 \end{array}}$$



$\rho > 0$ perturbation parameter

$\tilde{e} = (0_{n+neq+m}, e_m^t)^t$

Framework of Interior-Point Methods

- Choose an initial guess v_0 s.t. $(s_0, w_0) > 0$
- Choose the perturbation parameter

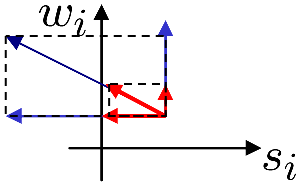
$$\rho_k = \sigma_k \mu_k \text{ s.t. } \quad \sigma_k \in [0,1)$$

$$\mu_k = \frac{s_k^t w_k}{m}$$

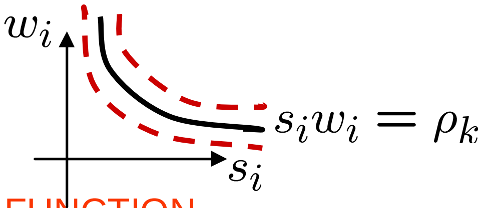
- Solve the perturbed Newton equation

$$H'(v_k)\Delta v_k = -H(v_k) + \rho_k \tilde{e}$$

- “Move” along the direction computed at the previous step: choose the damping parameter α_k such that the new iterate satisfies
  - FEASIBILITY



  - CENTRALITY CONDITIONS



  - SUFFICIENT DECREASE OF A MERIT FUNCTION

$$\Phi(v) = \frac{1}{2}\|H(v)\|_2^2$$

(Armijo, Eisenstat-Walker)

- Update the iterate $v_{k+1} = v_k + \alpha_k \Delta v_k$

Hypotheses for the convergence

$\Omega(\varepsilon) = \{v \in \mathbb{R}^n : \quad \varepsilon \leq \Phi(v) \leq \Phi(v_0)$ s.t. centrality coditions hold$\}$

- $f(x), g_1(x), g_2(x) \in C^2(\Omega(0))$ and $H_1{}'(v)$ is Lipschitz continuous in $\Omega(0)$;

- $\{\nabla(g_1)_i\}_{i=1,\ldots,neq}$ are linearly independent in $\Omega(0)$;

- $H'(v_k)$ is nonsingular $\forall k \geq 0$;

- The sequences $\{x_k\}, \{\lambda_k\}, \{w_k\}$ are bounded in $\Omega(\varepsilon), \forall \varepsilon > 0$.

[Durazzi, Ruggiero, JOTA, *2004*]

SOLVE “EXACTLY” THE PERTURBED NEWTON EQUATION

$$H'(v_k)\Delta v_k + H(v_k) = \rho_k \tilde{e}$$

SOLVE THE NEWTON EQUATION WITH RESIDUAL

$$\boxed{r_k} = H'(v_k)\Delta v_k + H(v_k) = \boxed{\rho_k \tilde{e}}$$

$$\|r_k\| > 0$$

Inexact Newton Method for the Solution of a Nonlinear System

Generate a sequence $\{v_k\}$ s.t.

$$\|H'(v_k)\Delta v_k + H(v_k)\| \leq \eta_k \|H(v_k)\|$$

Residual condition

$\eta_k \in [0, 1)$ forcing term

$$\|H(v_k + \Delta v_k)\| \leq (1 - \beta(1 - \eta_k))\|H(v_k)\|$$

Norm condition

$\beta \in (0, 1)$

$$v_{k+1} = v_k + \Delta v_k$$

Updating rule

[Eisenstat, Walker, *SIAM J. Optimization, 1994*]

CONVERGENCE:

If v_* is a limit point of the sequence $\{v_k\}$ s.t. $H'(v_*)$ is nonsingular, then $H(v_*) = 0$ and $\{v_k\}$ converges to v_*.

IMPLEMENTATION:

Residual condition: apply an iterative method to the Newton equation using the residual condition

$$\|r_k\| \leq \eta_k \|H(v_k)\|$$

as stopping rule ⟶ find a direction.

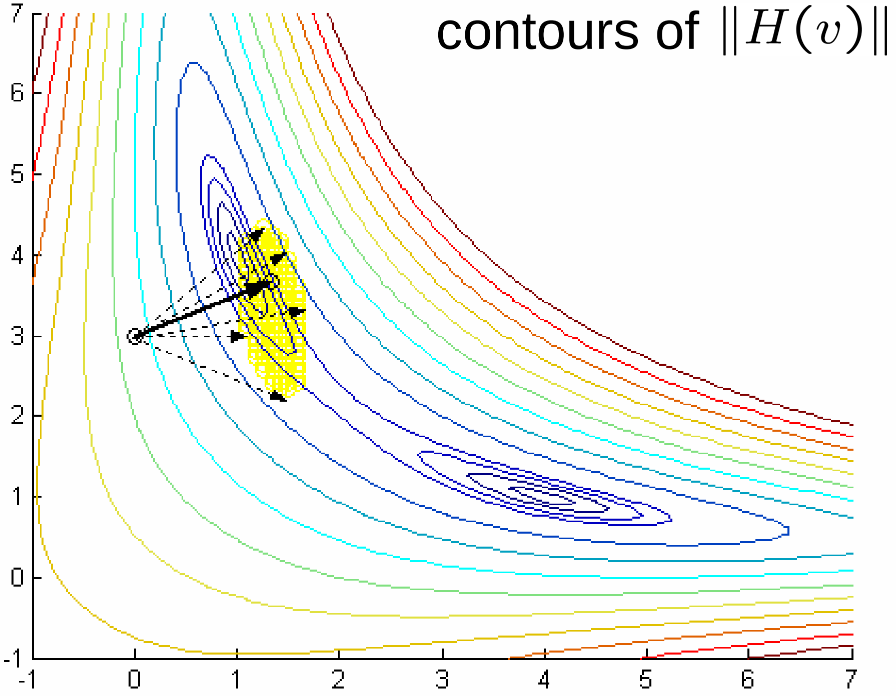
Norm condition: reduce the step along the direction that satisfy the residual condition with a damping parameter α_k until

$$\|H(v_k + \alpha_k \Delta v_k)\| \leq (1 - \alpha_k \beta (1 - \eta_k))\|H(v_k)\|$$

is satisfied ⟶ backtracking.

Example

$$H(v) = \begin{pmatrix} v_1 + v_2 - 5 \\ v_1 v_2 - 4 \end{pmatrix}$$



contours of $\|H(v)\|$

IP Methods as Inexact Newton Methods

- Solve exactly the perturbed Newton equation

$$r_k = \begin{pmatrix} 0 \\ \sigma_k \mu_k e_m \end{pmatrix} \longrightarrow \qquad \|r_k\| = \sigma_k \mu_k \sqrt{m}$$

Condition on the parameter μ_k

$$\boxed{\mu_k \le \frac{\|H(v_k)\|}{\sqrt{m}}}$$

➡

the residual condition is satisfied with forcing term σ_k

- Solve approximately the perturbed Newton equation

$$r_k = \begin{pmatrix} \bar{r}_k \\ \sigma_k \mu_k e_m \end{pmatrix} \longrightarrow \qquad \|r_k\|^2 = \|\bar{r}_k\|^2 + \sigma_k^2 \mu_k^2 m$$

Conditions on $\bar{r}_k$ and on the parameter μ_k

$$\boxed{\begin{array}{c} \|\bar{r}_k\| \le \delta_k \|H(v_k)\| \\ \delta_k < 1, \quad \delta_k + \sigma_k < 1 \\ \mu_k \le \dfrac{\|H(v_k)\|}{\sqrt{m}} \end{array}}$$

➡

the residual condition is satisfied with forcing term $(\delta_k + \sigma_k)$

Inexact Newton Interior-Point Methods

- Choose an initial guess v_0 s.t. $(s_0, w_0) > 0$
- Choose the parameters

$$\rho_k = \sigma_k \mu_k \text{ s.t. } \quad \begin{array}{l} \sigma_k, \delta_k \in [0,1) \\ \sigma_k + \delta_k < 1 \\ \mu_k \in \left[\frac{s_k^t w_k}{m}, \frac{\|H(v_k)\|}{\sqrt{m}} \right] \end{array}$$

- Apply an iterative method to the perturbed Newton equation until

$$\|\bar{r}_k\| \leq \delta_k \|H(v_k)\|$$

obtaining the direction

- Choose the damping parameter α_k such that the new iterate satisfies
  - » FEASIBILITY
  - » CENTRALITY CONDITIONS
  - » SUFFICIENT DECREASE: backtracking until

$$\|H(v_k + \alpha_k \Delta v_k)\| \leq (1 - \beta \alpha_k (\sigma_k - \delta_k)) \|H(v_k)\|$$

- Update the iterate $v_{k+1} = v_k + \alpha_k \Delta v_k$

A Nonmonotone Variant

Let $N \in \mathbb{N}$ a fixed positive integer, and $v_{\ell(k)}$ the element of the sequence $\{v_k\}$ such that

$$\boxed{\|H(x_{\ell(k)})\| = \max_{0\le j\le \min(N,k)} \|H(x_{k-j})\|}$$

Nonmonotone Inexact Newton Method for the solution of a nonlinear system

Generate a sequence $\{v_k\}$ s.t.

$$\|H'(v_k)\Delta v_k + H(v_k)\| \le \eta_k \|H(v_{\ell(k)})\|$$

Nonmonotone residual condition

$\eta_k \in [0,1)$ forcing term

$$\|H(v_k + \Delta v_k)\| \le (1-\beta(1-\eta_k))\|H(v_{\ell(k)})\|$$

Nonmonotone norm condition

$\beta \in (0,1)$

$$v_{k+1} = v_k + \Delta v_k$$

Updating rule

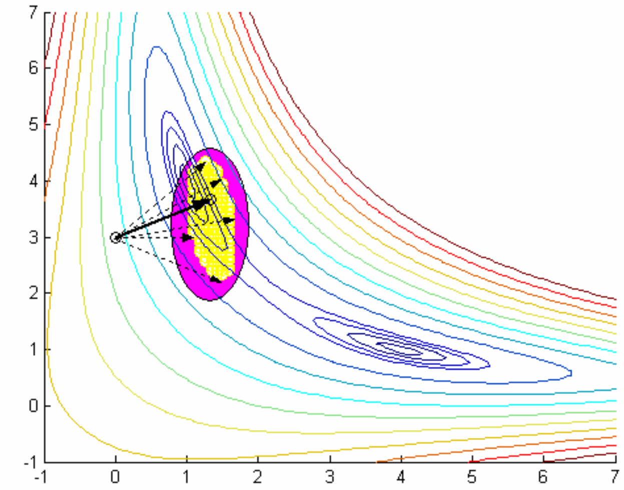
[Bonettini, *Optim. Methods and Software, 2004*]

CONVERGENCE:

If v_* is a limit point of the sequence $\{v_k\}$ such that $H'(v_*)$ is nonsingular and furthermore $\|\Delta v_k\|$ is bounded, then $H(v_*) = 0$ and $\{v_k\}$ converges to v_*.

REMARK:

The consequences of the nonmonotone rules are not only on the steplength, but also on the direction.



NONMONOTONE INEXACT NEWTON METHOD

↓

NONMONOTONE INEXACT NEWTON INTERIOR POINT METHOD

- Choice of the perturbation parameter

$$\mu_k \in \left[\frac{s_k^t w_k}{m}, \frac{\|H(v_{\ell(k)})\|}{\sqrt{m}} \right]$$

- Stopping rule for the inner system

$$\|\bar{r}_k\| \leq \delta_k \|H(v_{\ell(k)})\|$$

- Nonmonotone backtracking rule

$$\|H(v_k + \alpha_k \Delta v_k)\| \leq (1 - \beta \alpha_k (1 - \sigma_k - \delta_k)) \|H(v_{\ell(k)})\|$$

About the Inner Linear System

$$H'(v) = \boxed{\begin{pmatrix} Q & -G_1 & -G_2 & \\ -G_1^t & & & \\ -G_2^t & & & I_m \\ & & S & W \end{pmatrix}}$$

4 × 4 block matrix

$Q = \nabla^2_{xx} L$ hessian of the lagrangian function
$G_1^t = \nabla g_1^t$ jacobian of the equality constraints
$G_2^t = \nabla g_2^t$ " " " inequality "

$$H'(v_k)\Delta v_k = -H(v_k) + \rho_k \tilde{e}$$

$$\boxed{\begin{pmatrix} Q & -G_1 & -G_2 \\ -G_1^t & & \\ -G_2^t & & E \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta w \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}}$$

3 × 3 block matrix

$E = W^{-1} S$

$$\boxed{\begin{pmatrix} A & B \\ B^t & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} r \\ q \end{pmatrix}}$$

2 × 2 block matrix
$A = Q + G_2 S^{-1} W G_2^t$
$B = -G_1$

Solve the 2X2 System (1)

$$\begin{pmatrix} A & B \\ B^t & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} r \\ q \end{pmatrix}$$

Hypothesis: A is positive definite on the space

$$N(B^t) = \{z \in \mathbb{R}^n : B^t z = 0\}$$

⇓

The system has an unique solution, which is also the solution of the minimum problem

$$\min_{B^t y_1 - q = 0} \quad \frac{1}{2} y_1^t A y_1 - r^t y_1$$

Hestenes multiplier's scheme

$$\begin{aligned} y_2^{(j+1)} &= y_2^{(j)} + \chi(B^t y_1^{(j)} - q) \\ (A + \chi B B^t) y_1^{(j+1)} &= -B y_2^{(j)} + r + \chi B q \end{aligned}$$

If χ is sufficiently large ⇒ $(A + \chi B B^t)$ is positive definite

[Bonettini,Galligani,Ruggiero, *Rendiconti di Matematica, 2003*]

Solve the 2X2 System (2)

$$\begin{pmatrix} A & B \\ B^t & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} r \\ q \end{pmatrix}$$

Preconditioned conjugate gradient method

$$M = M_1 M_2$$

$$M_1 = \begin{pmatrix} D & 0 \\ B^t & -B^t D^{-1} B \end{pmatrix} \qquad M_2 = \begin{pmatrix} I_m & D^{-1}B \\ 0 & I_{neq} \end{pmatrix}$$

The conjugate gradient method with preconditioner M converges to a solution of the system in a finite number of steps.

Implementation detail:

Block solution → At each step of the conjugate gradient method, a system whose matrix is the positive definite matrix $B^t D^{-1} B$ has to be solved.

[Durazzi, Ruggiero, *Num. Linear Algebr. Appl., 2003*]

Numerical Experience

- Code in Fortran90 and Matlab function files
    - » Exact solver on the 3X3 system → MA27 (HSL);
    - » Iterative inner solver on the 2X2 system

Hestenes method Preconditioned CG

lipsol.f77 (Ng-Peyton)

- Numerical solution of optimal control problems

Large scale nonlinear programming problems structured and sparse.

[Mittelmann, Maurer, *JCAM,* 2000]

A boundary control problem

[Mittelmann, Maurer, *Optimization Techniques for Solving Elliptic Control Problems With Control and State Constraints,* Comput. Optim and Appl, 2000]

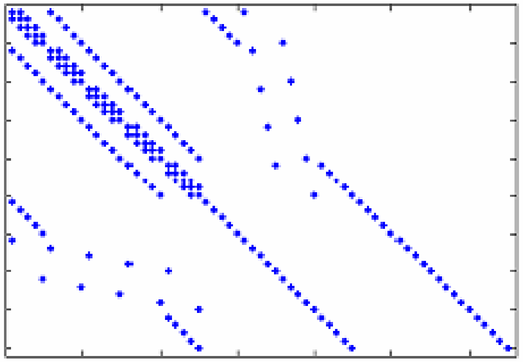
$$F(y,u) = \frac{1}{2}\int_{\Omega} (y(x) - y_d(x))^2 dx + \frac{\alpha}{2}\int_{\Gamma} u(x)^2 dx,$$

in $\Omega = (0,1)\times(0,1)$: $-\Delta y(x) = 0$ $y_d(x) = 2 - 2(x_1(x_1-1) + x_2(x_2-1))$

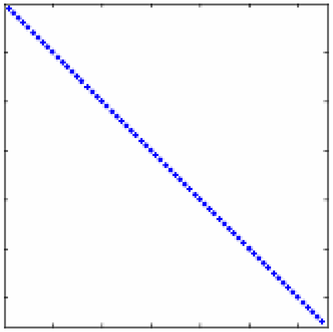
in $\Gamma = \partial\bar{\Omega}$: $\partial_\nu y(x) = u(x) - y(x)^2,$ $3.7 \le u(x) \le 4.5,$ $\alpha = 0.01$

in $\bar{\Omega}$: $y \le 0.271$

G_1



Q



Grid	Direct solver				Iterative solver							
	Monotone		Non monotone		Hestenes method				Preconditioned CG			
					Monotone		Non monotone		Monotone		Non monotone	
50	36	7.99	36	7.99	27(27)	0.92	27(27)	0.9	27(28)	0.8	27(28)	0.8
100	-	-	-	-	35(35)	9.9	33(33)	9.1	35(43)	9.6	35(39)	9.4
200	-	-	-	-	44(45)	89.14	44(44)	87.9	41(45)	82.6	43(55)	86.0

A distributed control problem

[Mittelmann, Maurer, *Optimization Techniques for Solving Elliptic Control Problems With Control and State Constraints,* Comput. Optim and Appl, 2001]

$$\int_\Omega (Mu(x)^2 - Ku(x)y(x))dx$$

in $\Omega = (0,1) \times (0,1)$: $\quad -\Delta y(x) = y(x)(a(x) - u(x) - by(x)) \quad y(x) \le \psi(x),$

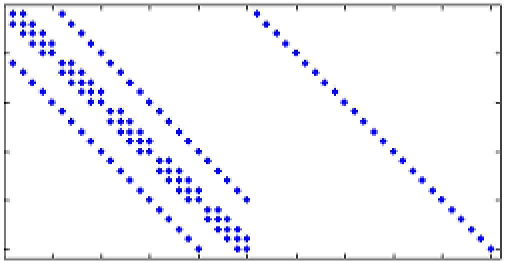
$\qquad u_1 \le u(x) \le u_2,$

in $\Gamma = \partial\bar{\Omega}$: $\quad \partial_\nu y(x) = 0.$

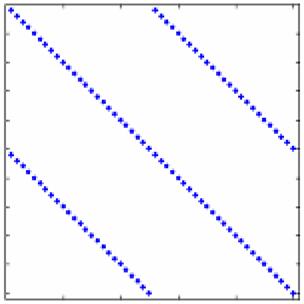
$a(x) = 7 + 4\sin(2\pi x_1 x_2)$

$b = 1, \quad M = 0, \quad K = 1, \quad u_1 = 2, \quad u_2 = 6, \quad \psi(x) = 4.8 \quad .$

G_1



Q



Grid	Direct solver				Iterative solver							
	Monotone		Non monotone		Hestenes method				Preconditioned CG			
					Monotone		Non monotone		Monotone		Non monotone	
50	31	174	31	174	37(37)	1.7	37(37)	1.7	37(49)	1.35	37(44)	1.32
100	-	-	-	-	50(50)	18.4	50(50)	18.4	51(97)	16.1	52(90)	16.0
200	-	-	-	-	62(72)	149.9	62(67)	147.2	61(278)	172.8	62(237)	166.6

Open problems and future work

- Stagnation of the iterates due to:
  - » BACKTRACKING Nonmonotone rules
  - » FEASIBILITY REQUIREMENT (Wachter-Biegler example)
- Singularity of the matrix $H'(v_k)$

unbounded multipliers sequence $\{\lambda_k\}$

  - » REGULARIZATION ?

- Choice of the initial point
- AMPL interface subroutine in “C”

with Gaetano Zanghirati (Univ. Ferrara)

and Federica Tinti (Univ. Padova)