Nonlinear programming without a penalty function or filter

Philippe Toint[1] Nick Gould[2]

[1]Department of Mathematics,University of Namur, Belgium

(philippe.toint@fundp.ac.be)

[2]Computer Laboratory, University of Oxford, UK

8th US-Mexico Workshop on Optimization and its Applications,
Huatulco, January 2007

Outline

Outline

Outline

1. The new method
2. Convergence theory
3. Numerical experiments
4. Conclusions

Outline

The problem

Consider the equality constrained problem

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{such that} \quad c(x) = 0.$$

Motivation:

- large-scale problems
- PDE constrained optimal control applications
- backup for a new filter method (in development)
- personnal curiosity...

The pedigree and context

An inexact SQP algorithm

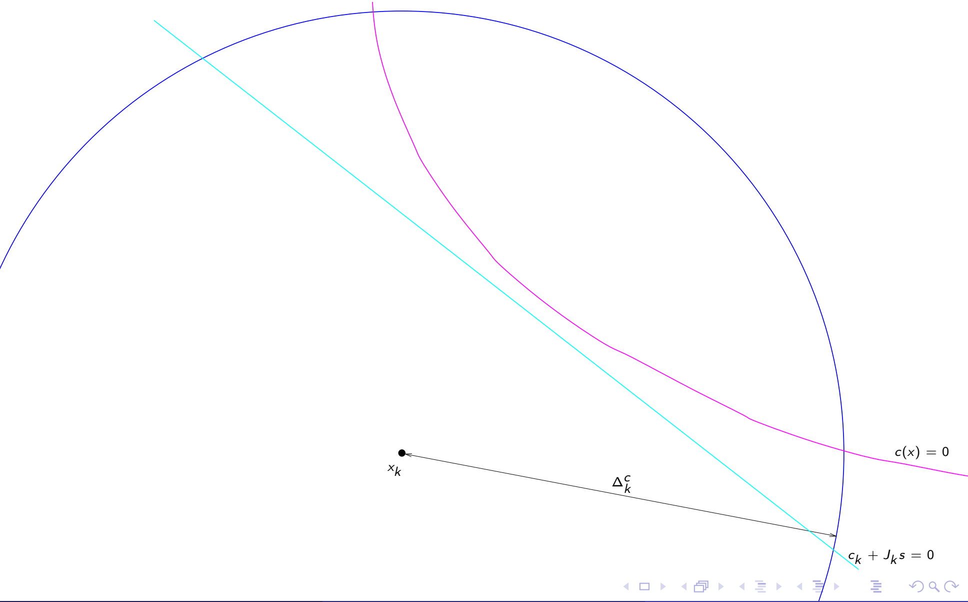
Origin/ inspiration from

- Himmelblau (1971) "flexible tolerance"
- Byrd-Omojokun (1989) SQP technique ...
- ... and much of the SQP literature (e.g. Nocedal-Plantenga, 1998)

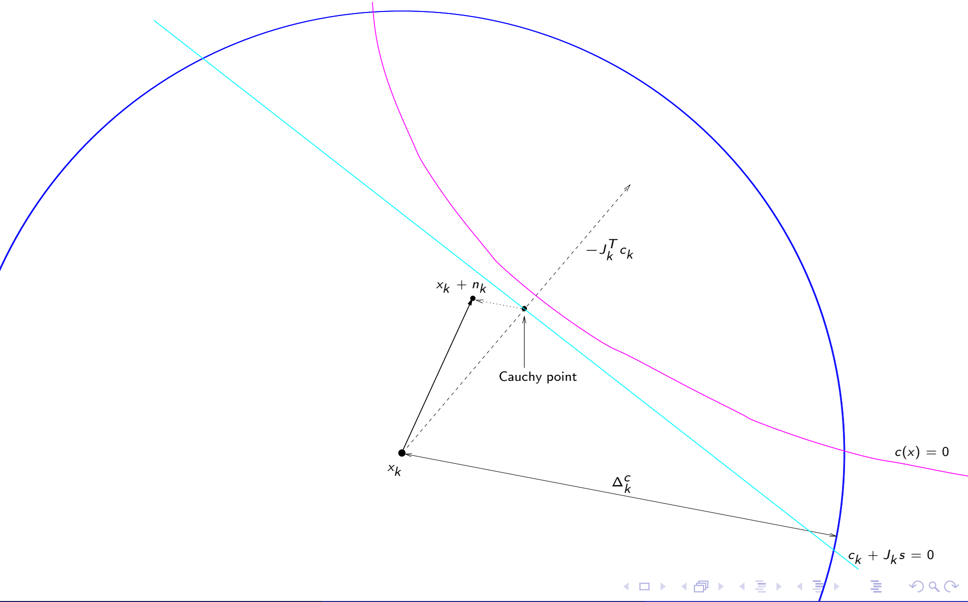
Similar concerns addressed by

- Byrd-Curtis-Nocedal (2006)
- (a posteriori) Zoppke-Donaldson (1995) "tolerance tube" method

At a typical iterate. . .



The normal step (pic)



The normal step (equs)

Constraint violation:

$$\theta(x) = \tfrac{1}{2}\|c(x)\|^2$$

Compute a Gauss-Newton step n_k for the problem $\min_n \theta(x_k + n)$ inside the constraints' trust region:

$$n_k \in \{n \in \mathbf{R}^n \mid \|n\| \le \Delta_k^c\}$$

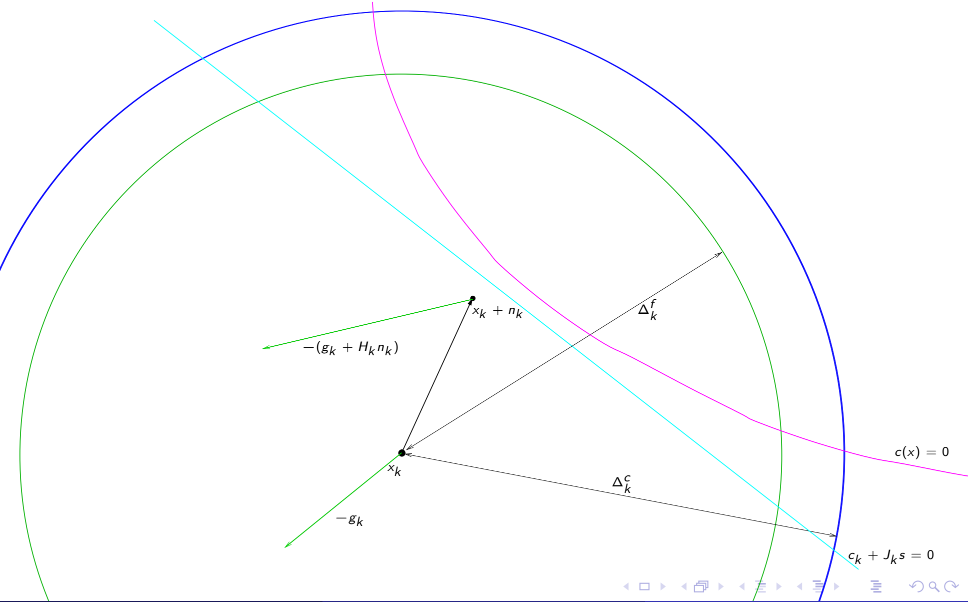
and also

$$\|n_k\| \le \kappa_{\rm n} \|c_k\|.$$

Relevant Cauchy condition:

$$\tfrac{1}{2}\|c_k\|^2 - \tfrac{1}{2}\|c_k + J_k n_k\|^2 \ge \kappa_{\rm nC} \|J_k^T c_k\| \min\left[\frac{\|J_k^T c_k\|}{1 + \|J_k^T J_k\|}, \Delta_k^c\right]$$

Objective's model and gradients (pic)



Objective's model and gradients (equs)

Whenever

$$\|n_k\| \leq \kappa_{\rm B} \Delta_k$$

decrease the objective's model:

$$m_k(x_k + n_k + t) = f_k + \langle g_k + H_k n_k, t\rangle + \tfrac{1}{2}\langle t, H_k t\rangle$$

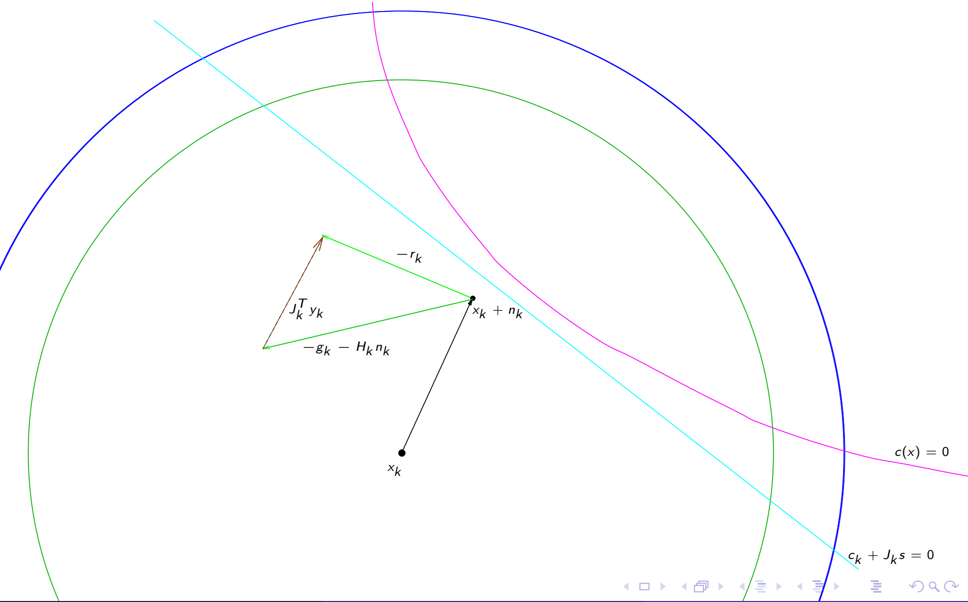
The objective's trust region

$$\{t \in \mathbf{R}^n \mid \|t\| \leq \Delta_k^f\}$$

The composite trust region:

$$s_k \in \mathcal{B}_k = \mathcal{N}_k \cap \mathcal{T}_k = \{s \in \mathbf{R}^n \mid \|s\| \leq \Delta_k = \min[\Delta_k^c, \Delta_k^f]\},$$

The effect of Lagrange multipliers (pic)



The effect of Lagrange multipliers (equs)

When sufficiently inside, i.e.,

$$\|n_k\| \leq \kappa_{\rm B} \Delta_k$$

compute least-squares Lagrange multipliers estimates from

$$\|y_k + [J_k^T]^I (g_k + H_k n_k)\| \leq \omega_1(\|c_k\|)$$

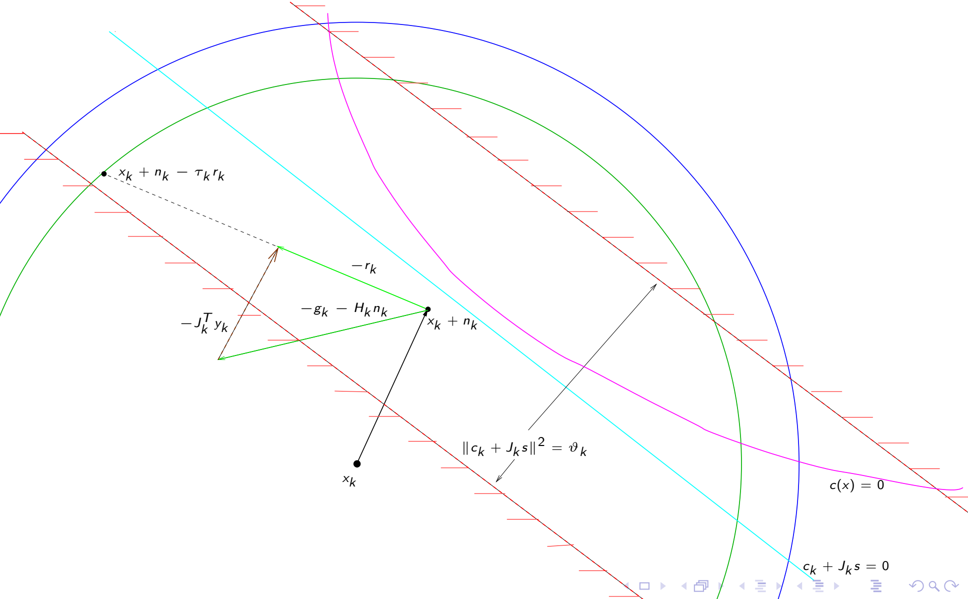
such that the approximate projected gradient

$$r_k \stackrel{\rm def}{=} g_k + H_k n_k + J_k^T y_k$$

satisfies

$$\langle g_k + H_k n_k, r_k \rangle \geq 0 \qquad \text{and} \qquad \|r_k\| \leq \kappa_{\rm nr} \|g_k + H_k n_k\|$$

Limiting infeasibility (pic)



Limiting infeasibility (equs)

Avoid jeopardizing the gain in feasibility due to the normal step

$$\|c_k + J_k(n_k + t_k)\|^2 \le \kappa_{\rm nt}\|c_k\|^2 + (1-\kappa_{\rm nt})\|c_k + J_k n_k\|^2 = \vartheta_k$$

for $\kappa_{\rm nt} \in (0,1)$, but also

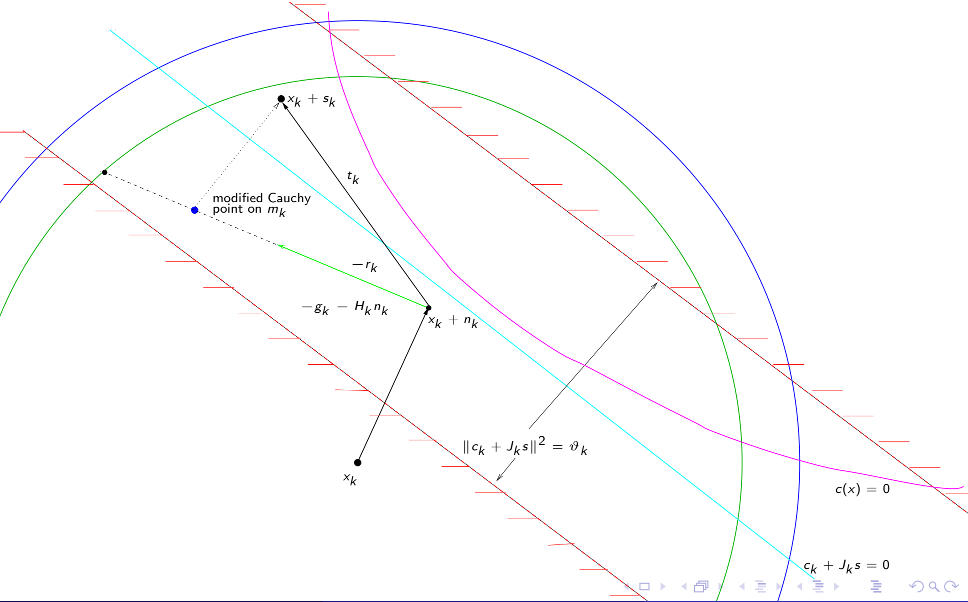
$$\|c_k + J_k(n_k - \tau_k r_k)\|^2 \le \vartheta_k$$

where

$$\tau_k = \frac{-\beta_k + \sqrt{\beta_k^2 + \Delta_k^2 - \|n_k\|^2}}{\|r_k\|} \quad \text{with} \quad \beta_k \stackrel{\rm def}{=} \frac{\langle n_k, r_k\rangle}{\|r_k\|}$$

is the maximum stepsize such that $x_k + n_k - \tau r_k$ remains in the composite trust region

The complete step (pic)



The complete step (equs)

(Maybe) compute a tangential step t_k that reduces m_k beyond the modified Cauchy point:

$$m_k(x_k+n_k)-m_k(x_k+n_k+t_k) \geq \kappa_{\rm tC1}\pi_k \min\left[\frac{\pi_k}{1+\|H_k\|}, \tau_k\|r_k\|\right]$$

where

$$\pi_k \stackrel{\rm def}{=} \frac{\langle g_k + H_k n_k, r_k\rangle}{\|r_k\|}$$

(also used as dual feasibility measure)
Define

$$s_k = n_k + t_k \quad \text{and} \quad \delta_k^c = \tfrac{1}{2}\|c_k\|^2 - \tfrac{1}{2}\|c_k + J_k s_k\|^2$$

f-steps and c-steps

Say that s_k is a f-step whenever

- $$\|n_k\| \leq \kappa_{\rm B}\Delta_k \quad \text{and} \quad \pi_k > \omega_2(\|c_k\|) \qquad (t_k \text{ computed})$$
- $$\delta_k^f \stackrel{\rm def}{=} m_k(x_k) - m_k(x_k + s_k) \geq \kappa_\delta[m_k(x_k + n_k) - m_k(x_k + s_k)]$$
- $$\theta(x_k^+) \leq \theta_k^{\max} \quad \text{for some} \quad \theta_k^{\max} > \theta(x_k)$$

($\theta_k^{\max}$ is the maximal infeasibity tolerated at iteration k)

Otherwise, s_k is a c-step

Accepting f-steps

Accept a f-step if

$$\rho_k^f \stackrel{\text{def}}{=} \frac{f(x_k) - f(x_k^+)}{\delta_k^f} \geq \eta_1$$

and update the f-radius by

$$\Delta_{k+1}^f \in \begin{cases} [\Delta_k^f, \infty) & \text{if } \rho_k^f \geq \eta_2, \\ [\gamma_2 \Delta_k^f, \Delta_k^f] & \text{if } \rho_k^f \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k^f, \gamma_2 \Delta_k^f] & \text{if } \rho_k^f < \eta_1, \end{cases}$$

Also possibly increase the c-radius by

$$\Delta_{k+1}^c \in [\Delta_k^c, +\infty) \quad \text{if} \quad \theta(x_k^+) \leq \eta_3 \theta_k^{\max} \quad \text{and} \quad \rho_k^f \geq \eta_1$$

Accepting c-steps

Accept a f-step if

$$\rho_k^c \stackrel{\text{def}}{=} \frac{\theta(x_k) - \theta(x_k^+)}{\delta_k^c} \geq \eta_1$$

and update the c-radius by

$$\Delta_{k+1}^c \in \begin{cases} [\Delta_k^c, \infty) & \text{if } \rho_k^c \geq \eta_2, \\ [\gamma_2 \Delta_k^c, \Delta_k^c] & \text{if } \rho_k^c \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k^c, \gamma_2 \Delta_k^c] & \text{if } \rho_k^c < \eta_1 \end{cases}$$

Also decrease the maximal infeasibility whenever $\rho_k^c \geq \eta_1$ by

$$\theta_{k+1}^{\max} = \max\left[\kappa_{\text{tx1}} \theta_k^{\max}, \theta(x_k^+) + \kappa_{\text{tx2}}(\theta(x_k) - \theta(x_k^+))\right]$$

Global convergence to first-order points

Assumptions:

- f and c are smooth and have bounded derivatives at x_k
- f is bounded below on $\{x \in \mathbb{R}^n \mid \theta(x) \le \theta_0^{\max}\}$
- the Jacobian is full-rank on the feasible set

Main result:

Either there exists a subsequence indexed by $\mathcal{Z}$ such that

$$\lim_{k\to\infty, k\in\mathcal{Z}} \|J_k^T c_k\| = 0 \quad \text{with} \quad \liminf_{k\to\infty, k\in\mathcal{Z}} \|c_k\| > 0,$$

or

$$\lim_{k\to\infty} \|c_k\| = 0 \quad \text{and} \quad \liminf_{k\to\infty} \|g_k + J_k^T y_k\| = 0.$$

Sketch of the proof: technical lemmas

- $$\theta(x_j) < \theta_k^{\max} \quad (j \geq k)$$

- $$\delta_k^c \geq \kappa_{\rm nC2} \|J_k^T c_k\| \min\left[\frac{\|J_k^T c_k\|}{1+\|J_k^T J_k\|}, \Delta_k^c\right]$$

- $$\text{first-order criticality} \quad \Leftrightarrow \|c_k\| \rightarrow 0 \text{ and } \pi_k \rightarrow 0$$

- $$\pi_k \geq \epsilon_f \text{ for all } f\text{-iterations } \Rightarrow \Delta_k^f \geq \epsilon_{\mathcal{F}}$$
  $$\|J_k^T c_k\| \geq \epsilon_\theta \text{ for all } c\text{-iterations} \Rightarrow \Delta_k^c \geq \epsilon_{\mathcal{C}}$$

- $$\|c_k\| \rightarrow 0 \quad \Rightarrow \quad \text{very successful } c\text{-iterations}$$

Sketch of the proof: the works

- f-iterations only $\Rightarrow \|c_k\| \to 0$ and $\pi_k \to 0$

(f monotone + modified Cauchy condition + infeasibility dominated)

- c-iterations $\Rightarrow \|c_k\| \to 0$ and $\Delta_k^c \geq \epsilon$

(shrinking $\theta_k^{\max}$!!)

- c-iterations $\Rightarrow \pi_k \to 0$

($\Delta_k \geq \epsilon$ + infinite subsequences of c-iterations impossible)

May include Lagrange multipliers in model for tangential step!

Numerical experiments: context

Results are for the following choices:

- normal step by TCG on $\|c_k + J_k n\|_2$ within $\|n\|_2 \leq \Delta_k^c$
- multipliers by factorization if J_k is full rank or CG otherwise
- tangential step by GLTR without preconditioner
- Attempts are made to identify (and then remove) dependent linearized constraints

Code still very experimental!!!

CUTEr equality constrained problems 1

Name	n	m	iter	ngeval	*f*	*c*	time
ALLINITC	2	1	9	8	-1.00e+00	2.15e-11	0.03
BT1	2	1	15	7	-1.00e-00	1.09e-11	0.05
BT2	3	1	35	20	3.26e-02	1.78e-15	0.11
BT3	5	3	6	6	4.09e+00	0.00e+00	0.03
BT4	3	2	8	7	-4.55e+01	0.00e+00	0.03
BT5	3	2	15	9	9.62e+02	5.33e-15	0.05
BT6	5	2	22	11	2.77e-01	3.26e-13	0.09
BT7	5	3	40	22	3.06e+02	7.02e-12	0.13
BT8	5	2	10	10	1.00e+00	3.81e-06	0.05
BT9	4	2	28	17	-1.00e+00	2.50e-08	0.08
BT10	2	2	7	7	-1.00e+00	4.18e-09	0.21
BT11	5	3	9	9	8.25e-01	1.05e-12	0.12
BT12	5	3	7	6	6.19e+00	1.29e-08	0.03
BYRDSPHR	3	2	13	9	-4.68e+00	8.52e-09	0.03
DIXCHLNG	10	5	31	16	2.47e+03	2.22e-16	0.13
EIGENA2	110	55	4	4	0.00e+00	0.00e+00	0.06
EIGENACO	110	55	4	4	0.00e+00	0.00e+00	0.07
EIGENB2	110	55	3	3	1.80e+01	0.00e+00	0.06
EIGENBCO	110	55	3	3	9.00e+00	0.00e+00	0.07
EIGENC2	462	231	14	9	1.37e-18	1.04e-11	31.81
EIGENCCO	462	231	46	30	1.03e-25	2.22e-16	171.12

CUTEr equality constrained problems 2

Name	n	m	iter	ngeval	f	c	time
ELEC	150	50	287	132	1.06e+03	1.27e-09	4.75
FCCU	19	8	5	5	1.11e+01	3.55e-15	0.09
GRIDNETE	60	36	7	7	3.96e+01	4.44e-16	0.21
GRIDNETH	264	144	7	7	5.71e+01	4.44e-16	1.52
HS6	2	1	13	11	4.34e-24	7.77e-15	0.04
HS7	2	1	10	8	-1.73e+00	2.35e-07	0.04
HS8	2	2	8	6	-1.00e+00	6.53e-12	0.02
HS9	2	1	4	3	-5.00e-01	0.00e+00	0.01
HS26	3	1	24	19	6.36e-11	4.48e-06	0.08
HS27	3	1	17	17	4.00e-02	1.04e-32	0.06
HS28	3	1	3	3	1.23e-32	0.00e+00	0.01
HS39	4	2	28	17	-1.00e+00	2.50e-08	0.07
HS40	4	3	6	6	-2.50e-01	1.11e-16	0.02
HS42	4	2	4	4	1.39e+01	5.21e-09	0.03
HS46	5	2	16	16	9.67e-10	7.57e-06	0.07
HS47	5	3	16	16	5.13e-10	6.35e-07	0.06
HS47	5	3	16	16	5.13e-10	6.35e-07	0.06
HS48	5	2	3	3	0.00e+00	0.00e+00	0.01
HS49	5	2	16	16	6.96e-09	8.88e-16	0.06
HS50	5	3	9	9	4.93e-32	8.88e-16	0.03
HS51	2	1	6	6	0.00e+00	0.00e+00	0.02

CUTEr equality constrained problems 3

Name	n	m	iter	ngeval	f	c	time
HS51	5	3	3	3	0.00e+00	0.00e+00	0.01
HS52	5	3	3	3	5.33e+00	1.39e-17	0.01
HS55SIM	1	1	2	2	0.00e+00	0.00e+00	0.02
HS56MOD	7	4	107	55	-3.46e+00	3.11e-15	0.22
HS56	7	4	107	55	-3.46e+00	3.11e-15	0.21
HS61	3	2	7	6	-1.44e+02	7.20e-11	0.03
HS77	5	2	11	11	2.42e-01	2.52e-14	0.04
HS78	5	3	6	6	-2.92e+00	5.55e-16	0.02
HS79	5	3	6	6	7.88e-02	4.78e-13	0.02
HS100LNP	7	2	15	10	6.81e+02	1.42e-14	0.07
HS111LNP	10	3	17	13	-4.78e+01	4.72e-09	0.08
KOPPEL	12	6	4	3	4.50e+00	2.98e-09	0.03
LCH	150	1	159	63	-4.23e+00	5.98e-12	3.72
LUKVLE1	100	98	13	10	6.23e+00	8.88e-16	0.22
LUKVLE3	100	2	11	11	2.76e+01	1.22e-15	0.15
LUKVLE6	99	49	16	16	6.04e+03	2.86e-16	0.40
LUKVLE7	100	4	15	11	-2.59e+01	1.01e-16	0.21
LUKVLE8	100	98	19	13	1.42e+04	3.71e-16	0.30
LUKVLE9	100	6	67	28	1.12e+01	1.42e-14	0.62
LUKVLE10	100	98	31	17	3.49e+01	6.66e-16	0.50

CUTEr equality constrained problems 4

Name	n	m	iter	ngeval	f	c	time
LUKVLE11	98	64	78	36	4.63e+03	2.66e-14	1.39
LUKVLE13	98	64	26	17	7.90e+02	4.00e-15	0.75
LUKVLE14	98	64	53	29	2.60e+00	2.97e-10	0.77
LUKVLE15	97	72	83	48	4.05e-08	1.11e-15	3.05
LUKVLE16	97	72	23	18	4.72e+01	6.66e-16	0.21
LUKVLE17	97	72	1000	712	3.24e+02	1.65e-18	11.20
LUKVLE18	97	72	1000	659	1.08e+02	1.12e-17	5.96
LUKVLESC	98	64	53	29	2.60e+00	2.97e-10	0.80
MARATOS2	2	1	4	4	-1.00e+00	7.99e-09	0.02
MARATOS	2	1	4	4	-1.00e+00	7.99e-09	0.01
MWRIGHT	5	3	8	8	2.50e+01	1.11e-15	0.03
ORTHRDM2	203	100	12	8	7.78e+00	1.14e-13	0.56
ORTHRDS2	103	50	1000	905	2.40e+01	8.53e-14	3.42
ORTHREGA	133	64	41	24	3.50e+02	4.51e-16	0.55
ORTHREGB	27	6	39	19	9.83e-27	7.51e-13	0.29
ORTHREGC	105	50	15	10	1.98e+00	5.26e-16	0.25
ORTHREGD	103	50	1000	894	7.20e+01	2.84e-13	7.44
ORTHRGDM	23	10	1000	870	2.02e+02	1.14e-13	5.76
ORTHRGDS	155	76	18	13	2.34e+01	8.53e-14	0.27
ROBOT	14	2	11	7	0.00e+00	2.95e-12	0.22
S316-322	2	1	3	3	3.34e+02	5.55e-17	0.01

Conclusions

Encouraging so far!

- finalize the algorithm
- inequality constraints
- theory to be completed
- current code consolidation necessary. . .
- many pending implementation issues
- include a filter mechanism

Thanks for your attention