Interior point method for nonlinear nonconvex optimization

Ladislav Lukšan, Ctirad Matonoha, Jan Vlček

Institute of Computer Science AS CR, Prague

GAMM Workshop
Applied and Numerical Linear Algebra
September 11-12, 2008
Technische Universität Hamburg-Harburg, Germany

Outline

1. Introduction
2. Direction determination I.
3. Indefinitely preconditioned CGM
4. Linear dependence of gradients of active constraints
5. Numerical experiments
6. Direction determination II.
7. Linear dependence of gradients of active constraints

1. Introduction

General nonlinear programming problem

Consider the general nonlinear programming problem

$$\text{(NP)} \qquad x = \arg \min_{x \in \mathcal{R}^n} f(x)$$

subject to

$$c_I(x) \le 0, \quad c_E(x) = 0,$$

where

$$\begin{aligned} c_I(x) &= [c_i(x) : i \in I]^T, & I &= \{1, \dots, m_I\} \\ c_E(x) &= [c_i(x) : i \in E]^T, & E &= \{m_I + 1, \dots, m_I + m_E = m\}. \end{aligned}$$

We assume that the functions

$$f(x) : \mathcal{R}^n \to \mathcal{R}, \quad c_I(x) : \mathcal{R}^n \to \mathcal{R}^{m_I}, \quad c_E(x) : \mathcal{R}^n \to \mathcal{R}^{m_E}$$

are twice continuously differentiable.

KKT conditions for (NP)

The necessary KKT (Karush-Kuhn-Tucker) conditions for the solution of problem (NP) have the following form:

$$\begin{aligned} g(x,u) &= 0, \\ c_I(x) &\le 0, \quad u_I \ge 0, \quad u_I^T c_I(x) = 0, \\ c_E(x) &= 0, \end{aligned}$$

where

$$g(x,u) = \nabla f(x) + A_I(x) u_I + A_E(x) u_E,$$

and

$$A_I(x) = [\nabla c_i(x) : i \in I], \quad A_E(x) = [\nabla c_i(x) : i \in E].$$

Here

$$u_I = [u_i(x) : i \in I]^T, \quad u_E = [u_i(x) : i \in E]^T$$

are vectors of Lagrange multipliers.

The idea of interior point methods

We introduce of a slack vector

$$s_I = [s_i(x) : i \in I]^T$$

and transform original problem (NP) to the sequence of problems with the logarithmic barrier function

$$\text{(IP)} \qquad x = \arg \min_{(x, s_I) \in \mathcal{R}^{n+m_I}} \left(f(x) - \mu e^T \ln(S_I) e \right),$$

subject to

$$c_I(x) + s_I = 0, \quad c_E(x) = 0,$$

where $\mu > 0$ is a barrier parameter, e is the vector with unit elements, and $S_I = \mathrm{diag}(s_i : i \in I)$.

- The logarithmic barrier term is used to ensure the inequality $s_I \geq 0$ implicitly.
- If $\mu = 0$, then the KKT conditions for (IP) coincide with the KKT conditions for (NP). Therefore $\mu \to 0$ is assumed.

KKT conditions for (IP)

The necessary KKT conditions for the solution of problem (IP) have the following form (primal-dual formulation):

$$
\begin{aligned}
g(x,u) &= 0, \\
S_I U_I e - \mu e &= 0, \\
c_I(x) + s_I &= 0, \\
c_E(x) &= 0,
\end{aligned}
\tag{1}
$$

where $U_I = \mathrm{diag}(u_i : i \in I)$. Inequalities $s_I > 0$ and $u_I > 0$ are required in all iterations.

- condition $s_I > 0$ is necessary for the definition of the logarithmic barrier function,
- condition $u_I > 0$ improves the properties of the linear system solved and is necessary for the construction of an efficient preconditioner.

Newton's method

Linearizing the primal-dual equations, we get one step of the Newton method

(2)
$$\begin{bmatrix} G & 0 & A_I & A_E \\ 0 & U_I & S_I & 0 \\ A_I^T & I & 0 & 0 \\ A_E^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s_I \\ \Delta u_I \\ \Delta u_E \end{bmatrix} = - \begin{bmatrix} g \\ S_I U_I e - \mu e \\ c_I + s_I \\ c_E \end{bmatrix},$$

where $g = g(x, u)$ and

$$G = G(x, u) = \nabla^2 f(x) + \sum_{i \in I} u_i \nabla^2 c_i(x) + \sum_{i \in E} u_i \nabla^2 c_i(x).$$

- The Hessian matrix $G(x, u)$ is not usually given analytically, but automatic or numerical differentiation is used instead.
- We assume that the matrix of system (2) is nonsingular.

Description of algorithm

The algorithm for an interior point method can be roughly described in the following form.

1. Let vectors $x \in \mathcal{R}^n$, $s_I \in \mathcal{R}^{m_I}$, $u_I \in \mathcal{R}^{m_I}$, $u_E \in \mathcal{R}^{m_E}$ such that $s_I > 0$, $u_I > 0$ be given.
2. Let a barrier parameter $\mu > 0$ be given.
3. Determine direction vectors Δx, Δs_I, Δu_I, Δu_E by solving a linear system equivalent to (2).
4. Choose a step-length $0 < \alpha \le \overline{\alpha}$.
5. Set
$$x := x + \alpha \Delta x, \quad s_I := s_I(\alpha, \Delta s_I),$$
$$u_I := u_I(\alpha, \Delta u_I), \quad u_E := u_E + \alpha \Delta u_E,$$
where $s_I(\alpha, \Delta s_I) > 0$ and $u_I(\alpha, \Delta u_I) > 0$ are functions of α depending on Δs_I and Δu_I, which are chosen by a suitable strategy.
6. Determine a new barrier parameter $\mu > 0$.

2. Direction determination I.

Active and inactive constraints

KKT condition (1) implies that $S_I U_I e \approx \mu e$ and if $\mu \to 0$, then either $u_i \to 0$ or $s_i \to 0$ holds for every index $i \in I$. Therefore, we can split the set of inequality constraints to an active and inactive subsets.

Active: $s_i \le \varepsilon_I u_i,\ i \in I$ – denoted by $\hat{.}$, i.e. $\hat{c}_I(x), \hat{s}_I, \hat{u}_I$.
Active constraints are those for which $c_i(x),\ i \in I,$ are close to zero, where $\hat{c}_I \in \mathcal{R}^{\hat{m}_I}$.

Inactive: $s_i > \varepsilon_I u_i,\ i \in I$ – denoted by $\check{.}$, i.e. $\check{c}_I(x), \check{s}_I, \check{u}_I$.
Inactive constraints are those for which $u_i,\ i \in I,$ are close to zero, where $\check{u}_I \in \mathcal{R}^{\check{m}_I}$.

Here $\varepsilon_I > 0$ is a suitable parameter and $\hat{m}_I + \check{m}_I = m_I$.

A general definition of the set of indices of active constraints:

$$\bar{E}(x) = E \cup \{i \in I : c_i(x) = 0\}$$

Elimination of Δs_I

System (2) is nonsymmetric with the dimension $n + m_E + 2m_I$. This system can be symmetrized and reduced by the elimination of the vector Δs_I. One has

$$\Delta s_I = -U_I^{-1} S_I (u_I + \Delta u_I) + \mu U_I^{-1} e$$

so that

$$\begin{bmatrix} G & A_I & A_E \\ A_I^T & -U_I^{-1} S_I & 0 \\ A_E^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u_I \\ \Delta u_E \end{bmatrix} = - \begin{bmatrix} g \\ c_I + \mu U_I^{-1} e \\ c_E \end{bmatrix}. \tag{3}$$

Disadvantage: elements of matrix $U_I^{-1} S_I$ can be unbounded, since $u_i \to 0$ if the i-th inequality constraint is inactive at the solution point.

Elimination of inactive equations

By elimination of inactive equations we obtain

$$\Delta \check{u}_I = \check{S}_I^{-1} \check{U}_I (\check{c}_I + \check{A}_I^T \Delta x) + \mu \check{S}_I^{-1} e$$

so that

$$\begin{bmatrix} \hat{G} & \hat{A}_I & A_E \\ \hat{A}_I^T & -\hat{U}_I^{-1} \hat{S}_I & 0 \\ A_E^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \hat{u}_I \\ \Delta u_E \end{bmatrix} = - \begin{bmatrix} \hat{g} \\ \hat{c}_I + \mu \hat{U}_I^{-1} e \\ c_E \end{bmatrix}, \tag{4}$$

where

$$\hat{G} = G + \check{A}_I \check{S}_I^{-1} \check{U}_I \check{A}_I^T, \tag{5}$$

$$\hat{g} = g + \check{A}_I \check{S}_I^{-1} \check{U}_I \check{c}_I + \mu \check{A}_I \check{S}_I^{-1} e. \tag{6}$$

Boundedness of matrices

Both matrices $\hat{G}$ and $\hat{U}_I^{-1}\hat{S}_I$ are bounded (if G and A are bounded) and if the strict complementarity conditions

$$\lim_{\mu \to 0} (s_i + u_i) > 0, \quad i \in I,$$

hold (recall that $s_i > 0$ and $u_i > 0$), then one has

$$\lim_{\mu \to 0} \hat{U}_I^{-1}\hat{S}_I = 0.$$

Similarly, the matrix $\check{S}_I^{-1}\check{U}_I$ is bounded and if the strict complementarity conditions hold, then

$$\lim_{\mu \to 0} \check{S}_I^{-1}\check{U}_I = 0.$$

Splitting of Δs_I

At the same time, we can split equality for Δs_I into two equalities to obtain

$$\begin{aligned} \Delta \hat{s}_I &= -\hat{U}_I^{-1} \hat{S}_I (\hat{u}_I + \Delta \hat{u}_I) + \mu \hat{U}_I^{-1} e, \\ \Delta \check{s}_I &= -(\check{c}_I + \check{A}_I^T \Delta x + \check{s}_I) \end{aligned}$$

after re-arrangements.

Elimination of inactive constraints is quite a general approach:

- if ε_I is large enough, we obtain original system (3);
- if ε_I is close to zero, all constraints are inactive.

A choice of a suitable ε_I can improve effectiveness of the algorithm and decrease the number of operations in an iterative method.

3. Indefinitely preconditioned CGM

Indefinite system

To simplify the notation, we rewrite system (4) containing only active constraints in the form

$$K\bar{d} = \begin{bmatrix} \hat{G} & \hat{A} \\ \hat{A}^T & -\hat{M} \end{bmatrix} \begin{bmatrix} d \\ \hat{d} \end{bmatrix} = \begin{bmatrix} b \\ \hat{b} \end{bmatrix} = \bar{b}, \tag{7}$$

where $\hat{A} = [\hat{A}_I, A_E]$ and $\hat{M} = \mathrm{diag}(\hat{M}_I, 0)$. Here $\hat{M}_I = \hat{U}_I^{-1}\hat{S}_I$ is a positive definite diagonal matrix. We assume that matrix K is nonsingular, which implies that A_E has a full column rank (gradients of active constraints are linearly independent).

System (7) is symmetric and indefinite of order $n + \hat{m} = n + \hat{m}_I + m_E$. It can be solved

- either directly by using the sparse Bunch-Parlett decomposition
- or iteratively by using Krylov-subspace methods for symmetric indefinite systems.

The preconditioner

We use a nonsingular preconditioner

$$C = \begin{bmatrix} \hat{D} & \hat{A} \\ \hat{A}^T & -\hat{M} \end{bmatrix},$$

where $\hat{D}$ is a positive definite diagonal matrix derived from the diagonal of $\hat{G}$. We restrict to the situation when matrix $\hat{G} - \hat{D}$ is non-singular (a usual situation and the worst case in some sense). One has

$$KC^{-1} = \begin{bmatrix} I + (\hat{G} - \hat{D})\hat{P} & (\hat{G} - \hat{D})\hat{Q} \\ 0 & I \end{bmatrix},$$

where

$$\hat{P} = \hat{D}^{-1} - \hat{D}^{-1}\hat{A}(\hat{A}^T\hat{D}^{-1}\hat{A} + \hat{M})^{-1}\hat{A}^T\hat{D}^{-1},$$

$$\hat{Q} = \hat{D}^{-1}\hat{A}(\hat{A}^T\hat{D}^{-1}\hat{A} + \hat{M})^{-1}.$$

Theorem 1.

Consider preconditioner C applied to system

$$K\bar{d} = \bar{b}$$

and assume that $\hat{G} - \hat{D}$ is nonsingular. Then matrix KC^{-1} has at least $\hat{m}_I + 2m_E$ unit eigenvalues but at most $\hat{m}_I + m_E$ linearly independent eigenvectors corresponding to these eigenvalues exist.

The other eigenvalues of matrix KC^{-1} are exactly eigenvalues of matrix

$$Z_E^T \tilde{G} Z_E (Z_E^T \tilde{D} Z_E)^{-1},$$

where $[Z_E, A_E]$ is a nonsingular square matrix,

$$Z_E^T A_E = 0, \quad Z_E^T Z_E = I$$

and where

$$\tilde{G} = \hat{G} + \hat{A}_I \hat{M}_I^{-1} \hat{A}_I^T, \quad \tilde{D} = \hat{D} + \hat{A}_I \hat{M}_I^{-1} \hat{A}_I^T.$$

If $Z_E^T \tilde{G} Z_E$ is positive definite, then all eigenvalues are positive.

Theorem 2.

Consider preconditioner C applied to system

$$K\bar{d} = \bar{b}$$

and assume that $\hat{G} - \hat{D}$ is nonsingular. Then the Krylov subspace $\mathcal{K}$ defined by matrix KC^{-1} and vector $\bar{r} \in R^{n+\hat{m}}$, where $\hat{m} = \hat{m}_I + m_E$, has a dimension of at most

$$\min(n+1, n-m_E+2).$$

Consequence: using a Krylov-subspace method we obtain a solution of system $K\bar{d} = \bar{b}$ after $\min(n+1, n-m_E+2)$ iterations at most.

Algorithm PCG

d – given, $\quad \hat{d} := 0,$

$r := b - \hat{G}d - \hat{A}\hat{d}, \quad \hat{r} := \hat{b} - \hat{A}^T d + \hat{M}\hat{d},$

$\beta := 0,$

while $\|r\| > \omega\|b\|$ or $\|\hat{r}\| > \omega \min(\|\hat{b}\|, \|\hat{c}\|)$ **do**

$\hat{t} := (\hat{A}^T \hat{D}^{-1} \hat{A} + \hat{M})^{-1} (\hat{A}^T \hat{D}^{-1} r - \hat{r}),$

$t := \hat{D}^{-1}(r - \hat{A}\hat{t}),$

$\gamma := r^T t + \hat{r}^T \hat{t}, \quad \beta := \beta\gamma,$

$p := t + \beta p, \quad \hat{p} := \hat{t} + \beta\hat{p},$

$q := \hat{G}p + \hat{A}\hat{p}, \quad \hat{q} := \hat{A}^T p - \hat{M}\hat{p},$

$\alpha := p^T q + \hat{p}^T \hat{q}, \quad \alpha := \gamma/\alpha,$

$d := d + \alpha p, \quad \hat{d} := \hat{d} + \alpha\hat{p},$

$r := r - \alpha q, \quad \hat{r} := \hat{r} - \alpha\hat{q},$

$\beta := 1/\gamma$

end while.

Termination conditions

The parameter ω represents precision of the inner iteration. It should satisfy the inequality

$$0 \le \omega \le \overline{\omega} < 1,$$

which is necessary for the global convergence, and also $\omega \to 0$ as $\|\bar{b}\| \to 0$ should hold for assuring the superlinear rate of convergence. Algorithm PCG terminates if

$$\|r\| \le \omega \|b\|, \quad \|\hat{r}\| \le \omega \|\hat{b}\|, \quad \|\hat{r}\| \le \omega \|\hat{c}\|$$

hold simultaneously, where

$$\hat{c} = \begin{bmatrix} \hat{c}_I + \hat{s}_I \\ c_E \end{bmatrix}.$$

Inequality $\|\hat{r}\| \le \omega \|\hat{c}\|$ plays an essential role if ε_I is large. In this case, elements of $\hat{u}_I$ can be small enough, implying a large norm of $\hat{c}_I + \mu \hat{U}_I^{-1} e$ (the first part of vector $\hat{b}$). Thus the resulting equations are badly scaled and the precision $\|\hat{r}\| \le \omega \|\hat{b}\|$ is insufficient.

Use of a Choleski decomposition

The matrix $(\hat{A}^T \hat{D}^{-1} \hat{A} + \hat{M})^{-1}$ used in Algorithm PCG is not computed, but the sparse Choleski decomposition (complete or incomplete) is used instead. Unfortunately, this matrix can be dense when $\hat{A}$ has dense rows. Assume that $\hat{A}^T = [\hat{A}_s^T, \hat{A}_d^T]$ and $\hat{D} = \mathrm{diag}(\hat{D}_s, \hat{D}_d)$, where

$$\hat{M}_s = \hat{A}_s^T \hat{D}_s^{-1} \hat{A}_s + \hat{M}$$

is sparse and $\hat{A}_d$ consists of dense rows. Then

$$(\hat{A}^T \hat{D}^{-1} \hat{A} + \hat{M})^{-1} = (\hat{M}_s + \hat{A}_d^T \hat{D}_d^{-1} \hat{A}_d)^{-1} = \hat{M}_s^{-1} - \hat{M}_s^{-1} \hat{A}_d^T \hat{M}_d^{-1} \hat{A}_d \hat{M}_s^{-1},$$

where

$$\hat{M}_d = \hat{D}_d + \hat{A}_d \hat{M}_s^{-1} \hat{A}_d^T$$

is a (low-dimensional) dense matrix. Again the sparse Choleski decomposition of matrix $\hat{M}_s$ is used instead of its inversion.

Theorem 3.

Consider Algorithm PCG with preconditioner C applied to system $K\bar{d} = \bar{b}$. Assume that the initial $\bar{d}$ is chosen in such a way that $\hat{r} = 0$ at the start of the algorithm. Let matrix $Z_E^T \tilde{G} Z_E$ be positive definite. Then:

1. Vector d^* (the first part of vector $\bar{d}^*$ which solves equation $K\bar{d} = \bar{b}$) is found after $n - m_E$ iterations at most.
2. The algorithm cannot break down before d^* is found.
3. Error $\|d - d^*\|$ converges to zero at least R-linearly with quotient

$$\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1},$$

   where κ is the spectral condition number of $Z_E^T \tilde{G} Z_E (Z_E^T \tilde{D} Z_E)^{-1}$.

4. If $d = d^*$, then also $\hat{d}_I = \hat{d}_I^*$ and d_E^* can be determined by the formula

$$d_E^* = d_E + (A_E^T \tilde{D}^{-1} A_E)^{-1} A_E^T \tilde{D}^{-1} r.$$

4. Linear dependence of gradients of active constraints

Example

Let

$$\begin{aligned} c_1(x) &= -x_1 \le 0 \\ c_2(x) &= -x_2 \le 0 \\ c_3(x) &= x_1^2 + 4x_2^2 - 4 \le 0 \\ c_4(x) &= (x_1 - 2)^2 + x_2^2 - 5 \le 0 \end{aligned}$$

Let $f(x)$ be such that $x^\star = [0, 1]^T \in \mathcal{R}^2$. Therefore $\bar{E}(x^\star) = \{1, 3, 4\}$.
Now

$$\nabla c_1(x^\star) = [-1, 0]^T, \quad \nabla c_2(x^\star) = [2x_1, 8x_2]^T = [0, 8]^T,$$

$$\nabla c_3(x^\star) = [2(x_1 - 2), 2x_2]^T = [-4, 2]^T.$$

These vectors are linearly dependent at the solution point.

Near singularity

In this case, system (7)

$$K\bar{d} = \begin{bmatrix} \hat{G} & \hat{A} \\ \hat{A}^T & -\hat{M} \end{bmatrix} \begin{bmatrix} d \\ \hat{d} \end{bmatrix} = \begin{bmatrix} b \\ \hat{b} \end{bmatrix} = \bar{b},$$

where

$$\hat{A} = [\hat{A}_I, A_E], \quad \hat{M} = \mathrm{diag}(\hat{M}_I, 0), \quad \hat{M}_I = \hat{U}_I^{-1}\hat{S}_I,$$

or matrix

$$\hat{A}^T \hat{D}^{-1} \hat{A} + \hat{M}$$

is singular or near singular, vector $\hat{d}$ obtained from (7) tends to infinity and the interior-point method usually fails.

Perturbation of $\hat{M}$

Motivated by Tikhonov regularization, we use a perturbation of $\hat{M}$ to eliminate singularity (or near singularity) of matrix $\hat{A}^T\hat{D}^{-1}\hat{A}+\hat{M}$. Therefore, we solve equation

(8) $$K(\varepsilon)\bar{d}(\varepsilon)=\begin{bmatrix}\hat{G} & \hat{A}\\ \hat{A}^T & -(\hat{M}+\varepsilon\hat{E})\end{bmatrix}\begin{bmatrix}d(\varepsilon)\\ \hat{d}(\varepsilon)\end{bmatrix}=\begin{bmatrix}b\\ \hat{b}\end{bmatrix}=\bar{b},$$

and use a preconditioner

$$C(\varepsilon)=\begin{bmatrix}\hat{D} & \hat{A}\\ \hat{A}^T & -(\hat{M}+\varepsilon\hat{E})\end{bmatrix},$$

where $\hat{E}$ is a positive semidefinite diagonal matrix (e.g. $\hat{E}=I$) and $\varepsilon>0$.

Theorem 4.

Consider perturbed system (8) with non-singular $\hat{G}$. Then

$$\frac{1}{2}\frac{\partial(\hat{d}^T(\varepsilon)\hat{E}\hat{d}(\varepsilon))}{\partial\varepsilon} = -\hat{d}^T(\varepsilon)\hat{E}(\hat{A}^T\hat{G}^{-1}\hat{A} + \hat{M} + \varepsilon\hat{E})^{-1}\hat{E}\hat{d}(\varepsilon).$$

If there is a number $\bar{\varepsilon} \geq 0$ such that

$$\hat{A}^T\hat{G}^{-1}\hat{A} + \hat{M} + \varepsilon\hat{E} \succ 0 \quad \forall\varepsilon \geq \bar{\varepsilon},$$

then the above expression is negative (if $\hat{d}(\varepsilon) \neq 0$) $\forall\varepsilon \geq \bar{\varepsilon}$ and

$$\hat{d}^T(\varepsilon)\hat{E}\hat{d}(\varepsilon) \to 0 \quad \text{if} \quad \varepsilon \to \infty.$$

Properties of regularization

- Singularity (or near singularity) of matrix $\hat{A}^T\hat{D}^{-1}\hat{A}+\hat{M}$ is usually detected during the Choleski decomposition. We choose ε and if the Gill-Murray modification of the Choleski decomposition is used, then a suitable matrix $\hat{E}$ is obtained as a by-product.
- The regularization described above deteriorates properties of preconditioner $C(\varepsilon)$. If $\hat{E}=\mathrm{diag}(\hat{E}_I, E_E)$, where E_E is non-singular, then the situation is the same as in case all constraints are inequalities. Thus, the Krylov subspace has a dimension of at most $n+1$ and using Krylov-subspace method we obtain the solution after $n+1$ iterations at most.

5. Numerical experiments

Test problems

To see the effect of regularization, numerical experiments were performed by using a set of test problems obtained as modifications of 18 test problems for equality constrained minimization which can be downloaded from

`http://www.cs.cas.cz/luksan/test.html`

We have used inequalities

$$x \le 0 \quad \text{and} \quad c(x) \le 0$$

and all problems have dimension $n = 1000$.

No regularization, $\varepsilon = 0$

P	NIT	NFV	NFG	F	C	G
1	21	21	126	999.000	0.0D+00	0.2D-15
2	22	22	308	24325.8	0.0D+00	0.3D-08
3	19	19	114	0.465461E-04	0.0D+00	0.2D-06
4	24	31	144	688.717	0.0D+00	0.7D-06
5	13	13	130	0.612895E-11	0.0D+00	0.3D-06
6	33	33	462	0.508804E-11	0.0D+00	0.2D-07
7	21	21	147	-13.8978	0.0D+00	0.2D-09
8	174	298	1218	82510.7	0.4D-15	0.1D-06
9	44	47	308	100.389	0.6D-09	0.1D-07
10	20	25	120	352.426	0.0D+00	0.1D-08
11	12	12	72	996.000	0.0D+00	0.7D-08
12	16	16	112	0.249966E-07	0.0D+00	0.3D-06
13	40	163	320	0.238126E+25	0.6D+00	0.1D+07
14	23	28	161	996.000	0.0D+00	0.4D-10
15	37	37	222	0.941987E-09	0.0D+00	0.2D-07
16	17	17	85	1494.00	0.1D-12	0.3D-06
17	23	24	115	4482.00	0.1D-11	0.7D-08
18	18	18	90	1494.00	0.6D-09	0.5D-06
Σ	577	845	4254	TIME=6.89		

Regularization with $\varepsilon = 10^{-14}$

P	NIT	NFV	NFG	F	C	G
1	21	21	126	999.000	0.0D+00	0.2D-15
2	22	22	308	24325.8	0.0D+00	0.3D-08
3	19	19	114	0.465461E-04	0.0D+00	0.2D-06
4	24	31	144	688.717	0.0D+00	0.7D-06
5	13	13	130	0.612895E-11	0.0D+00	0.3D-06
6	33	33	462	0.508804E-11	0.0D+00	0.2D-07
7	21	21	147	-13.8978	0.0D+00	0.2D-09
8	174	298	1218	82510.7	0.4D-15	0.1D-06
9	44	47	308	100.389	0.6D-09	0.1D-07
10	20	25	120	352.426	0.0D+00	0.1D-08
11	12	12	72	996.000	0.0D+00	0.7D-08
12	16	16	112	0.249966E-07	0.0D+00	0.3D-06
13	37	159	296	0.430686E+15	0.3D+01	0.7D+05
14	23	28	161	996.000	0.0D+00	0.4D-10
15	37	37	222	0.941987E-09	0.0D+00	0.2D-07
16	17	17	85	1494.00	0.1D-12	0.3D-06
17	23	24	115	4482.00	0.1D-11	0.7D-08
18	18	18	90	1494.00	0.6D-09	0.5D-06
Σ	574	841	4230	TIME=6.86		

Regularization with $\varepsilon = 10^{-10}$

P	NIT	NFV	NFG	F	C	G
1	21	21	126	999.000	0.0D+00	0.2D-15
2	22	22	308	24325.8	0.0D+00	0.3D-08
3	19	19	114	0.465461E-04	0.0D+00	0.2D-06
4	24	31	144	688.717	0.0D+00	0.7D-06
5	13	13	130	0.612895E-11	0.0D+00	0.3D-06
6	33	33	462	0.508804E-11	0.0D+00	0.2D-07
7	21	21	147	-13.8978	0.0D+00	0.2D-09
8	174	298	1218	82510.7	0.4D-15	0.1D-06
9	44	47	308	100.389	0.6D-09	0.1D-07
10	20	25	120	352.426	0.0D+00	0.1D-08
11	12	12	72	996.000	0.0D+00	0.7D-08
12	16	16	112	0.249966E-07	0.0D+00	0.3D-06
13	1893	2000	15152	0.749888E+13	0.3D+01	0.1D-06
14	23	28	161	996.000	0.0D+00	0.4D-10
15	37	37	222	0.941987E-09	0.0D+00	0.2D-07
16	18	18	90	1494.00	0.3D-13	0.1D-06
17	23	24	115	4482.00	0.2D-11	0.6D-07
18	17	17	85	1494.00	0.3D-08	0.3D-06
Σ	2430	2682	19086	TIME=14.84		

Regularization with $\varepsilon = 10^{-6}$

P	NIT	NFV	NFG	F	C	G
1	21	21	126	999.000	0.0D+00	0.2D-15
2	22	22	308	24325.8	0.0D+00	0.3D-08
3	19	19	114	0.465461E-04	0.0D+00	0.2D-06
4	24	31	144	688.717	0.0D+00	0.7D-06
5	13	13	130	0.612895E-11	0.0D+00	0.3D-06
6	33	33	462	0.508804E-11	0.0D+00	0.2D-07
7	21	21	147	-13.8978	0.0D+00	0.2D-09
8	1875	2000	13132	82510.7	0.2D-05	0.2D-08
9	44	47	308	100.389	0.6D-09	0.1D-07
10	20	25	120	352.426	0.0D+00	0.1D-08
11	12	12	72	996.000	0.0D+00	0.7D-08
12	16	16	112	0.249966E-07	0.0D+00	0.3D-06
13	1941	2000	15536	0.174126E+10	0.3D+01	0.8D-07
14	23	28	161	996.000	0.0D+00	0.4D-10
15	37	37	222	0.941987E-09	0.0D+00	0.2D-07
16	18	18	90	1494.00	0.3D-14	0.1D-06
17	19	19	95	4482.00	0.2D-09	0.7D-09
18	17	17	85	1494.00	0.3D-08	0.7D-09
Σ	4175	4379	31364	TIME=23.73		

Regularization with $\varepsilon = 10^{-2}$

P	NIT	NFV	NFG	F	C	G
1	26	27	156	999.000	0.4D-09	0.2D-06
2	22	22	308	24325.8	0.0D+00	0.3D-08
3	19	19	114	0.465461E-04	0.0D+00	0.2D-06
4	28	30	168	688.717	0.0D+00	0.9D-06
5	13	13	130	0.612895E-11	0.0D+00	0.3D-06
6	238	368	3332	0.156388E-11	0.0D+00	0.2D-06
7	22	22	154	-13.8978	0.0D+00	0.2D-08
8	490	2008	3437	82452.2	0.2D-01	0.8D-01
9	44	47	308	100.389	0.6D-09	0.1D-07
10	26	31	156	352.426	0.0D+00	0.2D-09
11	12	12	72	996.000	0.0D+00	0.7D-08
12	16	16	112	0.249966E-07	0.0D+00	0.3D-06
13	1993	2000	15952	102605.	0.3D+01	0.6D-07
14	44	52	308	996.000	0.0D+00	0.9D-06
15	82	92	492	0.807095E-08	0.0D+00	0.8D-06
16	18	18	90	1494.00	0.3D-14	0.1D-06
17	18	18	90	4482.00	0.4D-09	0.3D-08
18	15	15	75	1494.00	0.4D-07	0.2D-08
Σ	3126	4810	25454	TIME=46.65		

Conclusion

- G positive definite – additional advantageous preconditioners
  G indefinite or singular – preconditioner C seems to be most robust
- ??? Effective iterative methods for solving linear KKT systems:
  If the system solved is very ill-conditioned ($m_I \gg n$) then the fill-in can be enormously great and the accurate solution can be unsuitable – directions are too large or almost perpendicular to the gradient of the merit function ($\rightarrow$ inexact solution is preferred).
- Regularization reliably eliminates the numerical explosion caused by linear dependence of active constraints and sometimes gives the solution when the standard iterative method fails (future research).
- Active constraints – solved inaccurately by the PKS method;
  Inactive constraints – obtained by direct elimination.
  $\rightarrow$ equations with bounded coefficients, suitable for iterative solvers, dimension of the system solved is usually decreased.
- Suitable ε_I $(10^{-1}, 10^{-2}, 1)$ can improve effectiveness of the algorithm.
- Regularization – better examples with linear dependence of gradients of active constraints would be more convenient.

6. Direction determination II.

Symmetrization

System (2) (linearizing the primal-dual equations) can be symmetrized with elimination of inactive constraints but without elimination of the vector Δs_I.

(9)
$$\begin{bmatrix} \hat{G} & 0 & \hat{A}_I & A_E \\ 0 & I & \hat{D}_I & 0 \\ \hat{A}_I^T & \hat{D}_I & 0 & 0 \\ A_E^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \hat{D}_I^{-1} \Delta \hat{s}_I \\ \Delta \hat{u}_I \\ \Delta u_E \end{bmatrix} = - \begin{bmatrix} \hat{g} \\ \hat{D}_I \hat{g}_s \\ \hat{c}_I + \hat{s}_I \\ c_E \end{bmatrix},$$

where

$$\hat{D}_I = (\hat{S}_I \hat{U}_I^{-1})^{1/2}, \quad \hat{D}_I \hat{g}_s = (\hat{S}_I \hat{U}_I)^{1/2} e - \mu (\hat{S}_I \hat{U}_I)^{-1/2} e,$$

$$\hat{G} = G + \check{A}_I \check{S}_I^{-1} \check{U}_I \check{A}_I^T, \quad \hat{g} = g + \check{A}_I \check{S}_I^{-1} \check{U}_I \check{c}_I + \mu \check{A}_I \check{S}_I^{-1} e$$

(same as above).

Use of trust region methods

General form of (9):

(10) $$\begin{bmatrix} B & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} d \\ d_u \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

If we consider the subproblem:

(11) $$\min \; Q(d) = \frac{1}{2} d^T B d + g^T d \quad \text{subject to} \quad A^T d + h = 0,$$

then the Lagrange function has the form

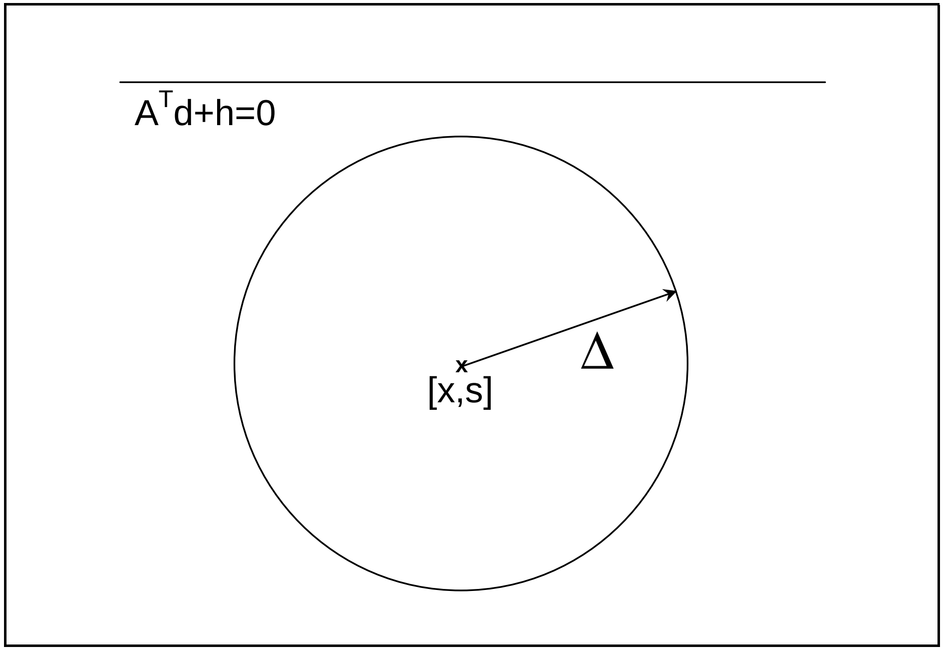
$$\mathcal{L}(d) = Q(d) + d_u^T (A^T d + h)$$

where d_u is a Lagrange multiplier. The optimality conditions has exactly form (10).

We can use a trust region method to (11) with a constraint

$$\|d\| \leq \Delta.$$

Incompatibility of constraints



We will use the idea of Byrd and Omojokun, to make both constraints compatible and secure a sufficient decrease of $Q(d)$:

$$d = d_V + d_H$$

(the sum of vertical and horizontal steps).

Vertical step

First, consider the problem

$$\min \; \|A^T d + h\| \quad \text{subject to} \quad \|d\| \leq \delta\Delta$$

for $0 < \delta < 1$ (e.g. $\delta = 0.8$), which is equivalent to

$$\min \; Q_V(d) = \frac{1}{2} d^T A A^T d + h^T A^T d \quad \text{subject to} \quad \|d\| \leq \delta\Delta.$$

A solution is the vertical step d_V and there exists w such that $d_V = Aw$.

The algorithm (the dogleg method) works with matrices $A^T A$ and its inverse so we suppose that A has a full column rank.

Horizontal step

Reformulation of the original problem:

$$\min\ Q_H(d) = \frac{1}{2} d^T B d + g^T d \quad \text{subject to} \quad A^T d = A^T d_V, \quad \|d\| \le \Delta.$$

The constraints are compatible and for a solution d_H it follows that

$$A^T d_H = 0, \quad d_H = Z d_Z, \quad d_V^T d_H = 0$$

for some vector d_Z where the columns of Z form a basis of the null space of A^T. Substitution into Q_H leads to a problem

$$\min\ Q_Z(d) = \frac{1}{2} d^T B_Z d + g_Z^T d \quad \text{subject to} \quad \|Zd\| \le \bar{\Delta},$$

where

$$B_Z = Z^T B Z, \quad g_Z = Z^T (B d_V + g), \quad \bar{\Delta} = \sqrt{\Delta^2 - \|d_V\|^2}.$$

The CGM method is used to solve this problem whose solution is d_Z. Finally, the horizontal step is $d_H = Z d_Z$.

Computation of Lagrange multipliers

- The use of $d_H = Zd_Z$ (instead of d_Z) leads to multiplication by the matrix
$$Z(Z^TZ)^{-1}Z^T = I - A(A^TA)^{-1}A^T$$
so the matrix Z need not be computed.
- Back to (10):
$$\begin{bmatrix} B & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} d \\ d_u \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$
Lagrange multipliers d_u cannot be computed from the CG method. From (10) we have
$$Ad_u = -(g + Bd) \equiv -r$$
where r is a residuum. Thus
$$d_u = -(A^TA)^{-1}A^Tr$$
as a solution of a least squares problem.

7. Linear dependence of gradients of active constraints

Regularized problem

When the gradients of active constraints are linearly dependent, the matrix A doesn't have the full column rank. We will consider the problem

(12) $\min\ Q(d,p) = \frac{1}{2} d^T B d + g^T d + \frac{1}{2} p^T p$ s.t. $A^T d + \delta p + h = 0,$

where δ is some small number. The optimality conditions have the form (d_u is a Lagrange multiplier):

$$\begin{aligned} Bd + g + Ad_u &= 0 \\ p + \delta d_u &= 0 \\ A^T d + \delta p + h &= 0 \end{aligned}$$

which leads to a system

$$\begin{bmatrix} B & A \\ A^T & -\delta^2 I \end{bmatrix} \begin{bmatrix} d \\ d_u \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

Conversion into a standard form

Denote

$$\tilde{B} = \begin{bmatrix} B & \\ & I \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} A \\ \delta I \end{bmatrix}, \quad \tilde{g} = \begin{bmatrix} g \\ 0 \end{bmatrix}, \quad \tilde{d} = \begin{bmatrix} d \\ p \end{bmatrix}$$

Then problem (12) becomes

$$\min \; Q(\tilde{d}) = \frac{1}{2} \tilde{d}^T \tilde{B} \tilde{d} + \tilde{g}^T \tilde{d} \quad \text{s.t.} \quad \tilde{A}^T \tilde{d} + h = 0.$$

The matrix $\tilde{A}$ has the full column rank and we can use a theory based on vertical and horizontal steps.

Conclusion

- Trust region methods are an effective tool for solving optimization problems especially when the objective function is nonconvex or the problem is ill-conditioned.
- They are globally convergent.
- So their principle is used for computation of direction vectors.
- Future research: to use their good properties also for the cases when the gradients of active constraints are linearly dependent (suitable regularized subproblem).

References

1. Lukšan L., Matonoha C., Vlček J.: *Interior point method for non-linear non-convex optimization*, Numerical Linear Algebra with Applications, Vol. 11, No. 5-6, 2004, pp. 431-453.
2. Lukšan L., Matonoha C., Vlček J.: *Interior point method for large-scale nonlinear programming*, Optimization Methods and Software, Vol. 20, No. 4-5, 2005, pp. 569-582.
3. Lukšan L., Vlček J.: *Indefinitely Preconditioned Inexact Newton Method for Large Sparse Equality Constrained Nonlinear Programming Problems*, Numerical Linear Algebra with Applications, Vol. 5, 1998, pp.219-247.
4. Lukšan L., Vlček J.: *Numerical experience with iterative methods for equality constrained non-linear programming problems*, Optimization Methods and Software, Vol. 16, 2001, pp. 257-287.
5. Lukšan L., Vlček J.: *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, TR V767, ICS AS CR, 1998.

Thank you for your attention!