PENALTY FUNCTION, REGULARIZATION, DUALITY, AND NEWTON'S METHOD IN LINEAR OPTIMIZATION

Alexander Golikov, Yuri Evtushenko, Igor Kaporin

Dorodnicyn Computing Center
FRC CSC RAS

«Теория и численные методы решения обратных и некорректных задач»
Новосибирск, Академгородок, 12-22 апреля 2021 года

SYSTEM OF LINEAR EQUATIONS WITH NON-NEGATIVE SOLUTION

$$Ax = b, \quad x \geq 0_n$$

$$A \in R^{m \times n}, \quad b \in R^m, \quad m < n$$

$$\min_{x \in X} 0_n^{\mathrm{T}} x, \quad X = \{x \in R^n : Ax = b, \quad x \geq 0_n\}$$

$$\max_{u \in U} b^T u, \quad U = \{u \in R^m : A^T u \leq 0_n\}$$

$$\min_{x \in R^n, y \in R^m} e_m^{\mathrm{T}} y$$

$$Ax + y = b, \, x \geq 0_n, \, y \geq 0_m$$

$$\max_{u \in U} \{b^T u + \frac{1}{2\varepsilon} \| (A^T u)_+ \|^2\}$$

$x^* = \frac{1}{\varepsilon}(A^T u(\varepsilon))_+$ is the exact minimum 2-norm solution to

$$Ax = b, \, x \geq 0_n \quad \text{for all} \quad 0 < \varepsilon \leq \varepsilon^*$$

Farkas' Theorem

Original system (I)

$$Ax = b, \;\; x \geq 0_n$$

Alternative system (II)

$$A^{\mathrm{T}}u \leq 0_n, \;\; b^{\mathrm{T}}u = \rho > 0$$

$$\min_{u \in R^m} \tfrac{1}{2}[(\rho - b^{\mathrm{T}}u)^2 + \|(A^{\mathrm{T}}u)_+\|^2] \rightarrow u^*$$

$$\tilde{x}^* = \frac{(A^{\mathrm{T}}u^*)_+}{\rho - b^{\mathrm{T}}u^*}$$ – normal solution to system (I)

How to Make the Problem Easier

- **Reduce the dimension:**
- **Remove the constraints:**
- **Extract a unique solution:**

- **=> *replace it by the dual***
- **=> *penalize it***
- **=> *regularize it***

$$
\begin{tikzcd}
P_{pr} \arrow[d, leftrightarrow] & P_{p} \arrow[l, "\mathrm{reg}"'] \arrow[d, leftrightarrow] & P_{0} \arrow[l, "\mathrm{pen}"'] \arrow[r, "\mathrm{reg}"] \arrow[d, leftrightarrow] & P_{r} \arrow[r, "\mathrm{pen}"] \arrow[d, leftrightarrow] & P_{rp} \arrow[d, leftrightarrow] \\
D_{rp} & D_{r} \arrow[l, "\mathrm{pen}"] & D_{0} \arrow[l, "\mathrm{reg}"] \arrow[r, "\mathrm{pen}"'] & D_{p} \arrow[r, "\mathrm{reg}"'] & D_{pr}
\end{tikzcd}
$$

P_0: $$\min_{x \in X} 0_n^{\mathrm{T}} x,$$

D_0: $$\max_{u \in U} b^{\mathrm{T}} u,$$

$$X = \{x \in R^n : Ax = b,\ x \geq 0_n\}$$

$$U = \{u \in R^m : A^{\mathrm{T}} u \leq 0_n\}$$

$$\begin{tikzcd}
P_{pr} \arrow[d, leftrightarrow] & P_p \arrow[l, "\text{reg}"'] \arrow[d, leftrightarrow] & P_0 \arrow[l, "\text{pen}"'] \arrow[r, "\text{reg}"] \arrow[d, leftrightarrow] & P_r \arrow[r, "\text{pen}"] \arrow[d, leftrightarrow] & P_{rp} \arrow[d, leftrightarrow] \\
D_{rp} & D_r \arrow[l, "\text{pen}"] & D_0 \arrow[l, "\text{reg}"] \arrow[r, "\text{pen}"'] & D_p \arrow[r, "\text{reg}"'] & D_{pr}
\end{tikzcd}$$

P_0: $\min\limits_{x \in X} 0_n^{\mathrm{T}} x,$

$$X = \{x \in R^n : Ax = b,\ x \geq 0_n\}$$

D_0: $\max\limits_{u \in U} b^{\mathrm{T}} u,$

$$U = \{u \in R^m : A^{\mathrm{T}} u \leq 0_n\}$$

P_r: $\min\limits_{x \in X}(0_n^{\mathrm{T}} x + \tfrac{1}{2} \| x \|^2)$

$$X = \{x \in R^n : Ax = b, \quad x \geq 0_n\}$$

D_p: $\max\limits_{u \in R^m} (b^{\mathrm{T}} u - \tfrac{1}{2} \| (A^{\mathrm{T}} u)_+ \|^2)$

$$x^* = (A^{\mathrm{T}} u^*)_+$$

P_{rp}: $\min\limits_{x \in R_+^n}(\tfrac{1}{2} \| x \|^2 + \tfrac{1}{2\varepsilon} \| b - Ax \|^2)$

$$u(\varepsilon) = \tfrac{1}{\varepsilon}(b - Ax(\varepsilon))$$

D_{pr}: $\max\limits_{u \in R^m}(b^{\mathrm{T}} u - \tfrac{1}{2} \| (A^{\mathrm{T}} u)_+ \|^2 - \tfrac{\varepsilon}{2} \| u \|^2)$

$$x(\varepsilon) = (A^{\mathrm{T}} u(\varepsilon))_+$$

$$
\begin{tikzcd}
\mathrm{P}_{\mathrm{pr}} \arrow[d, leftrightarrow] & \mathrm{P}_{\mathrm{p}} \arrow[l, "\text{reg}"'] \arrow[d, leftrightarrow] & \mathrm{P}_0 \arrow[l, "\text{pen}"'] \arrow[r, "\text{reg}"] \arrow[d, leftrightarrow] & \mathrm{P}_{\mathrm{r}} \arrow[r, "\text{pen}"] \arrow[d, leftrightarrow] & \mathrm{P}_{\mathrm{rp}} \arrow[d, leftrightarrow] \\
\mathrm{D}_{\mathrm{rp}} & \mathrm{D}_{\mathrm{r}} \arrow[l, "\text{pen}"] & \mathrm{D}_0 \arrow[l, "\text{reg}"] \arrow[r, "\text{pen}"'] & \mathrm{D}_{\mathrm{p}} \arrow[r, "\text{reg}"'] & \mathrm{D}_{\mathrm{pr}}
\end{tikzcd}
$$

P_0: $$\min_{x\in X} 0_n^{\mathrm{T}} x,$$

$$X = \{x \in R^n : Ax = b,\ x \geq 0_n\}$$

D_0: $$\max_{u\in U} b^{\mathrm{T}} u,$$

$$U = \{u \in R^m : A^{\mathrm{T}} u \leq 0_n\}$$

P_{p}: $$\min_{x\in R_+^n} \tfrac{1}{2} \| b - Ax \|^2$$

$$u^* = b - Ax^*$$

D_{r}: $$\max_{u\in U} (b^{\mathrm{T}} u - \tfrac{1}{2} \| u \|^2),$$

$$U = \{u \in R^m,\ A^{\mathrm{T}} u \leq 0_n\}$$

P_{pr}: $$\min_{x\in R_+^n} (\tfrac{1}{2} \| b - Ax \|^2 + \tfrac{\varepsilon}{2} \| x \|^2)$$

$$u(\varepsilon) = b - Ax(\varepsilon)$$

D_{rp}: $$\max_{u\in R^m} (b^{\mathrm{T}} u - \tfrac{1}{2} \| u \|^2 - \tfrac{1}{2\varepsilon} \| (A^{\mathrm{T}} u)_+ \|^2$$

$$x(\varepsilon) = \tfrac{1}{\varepsilon} (A^{\mathrm{T}} u(\varepsilon))_+$$

Primal problems

P_0: $$\min_{x\in X} 0_n^{\mathrm{T}} x,$$
$$X = \{x \in R^n : Ax \le b\}$$

P_p: $$\min_{x\in R^n} \tfrac{1}{2} \| (Ax - b)_+ \|^2$$
$$u^* = (Ax^* - b)_+$$

P_{pr}: $$\min_{x\in R^n} (\tfrac{1}{2} \| (Ax - b)_+ \|^2 + \tfrac{\varepsilon}{2} \| x \|^2)$$
$$u(\varepsilon) = (Ax(\varepsilon) - b)_+$$

P_r: $$\min_{x\in X} (0_n^{\mathrm{T}} x + \tfrac{1}{2} \| x \|^2),$$
$$X = \{x \in R^n : Ax \le b\}$$

P_{rp}: $$\min_{x\in R^n} (\tfrac{1}{2} \| x \|^2 + \tfrac{1}{2\varepsilon} \| (Ax - b)_+ \|^2)$$
$$u(\varepsilon) = \tfrac{1}{\varepsilon} (Ax(\varepsilon) - b)_+$$

Dual problems

D_0: $$\max_{u\in U} b^{\mathrm{T}} u,$$
$$U = \{u \in R^m : A^{\mathrm{T}} u = 0_n, u \ge 0_m\}$$

D_r: $$\max_{u\in U} (b^{\mathrm{T}} u - \tfrac{1}{2} \| u \|^2),$$
$$U = \{u \in R^m, A^{\mathrm{T}} u = 0_n, u \ge 0_m\}$$

D_{rp}: $$\max_{u\in R_+^m} (b^{\mathrm{T}} u - \tfrac{1}{2} \| u \|^2 - \tfrac{1}{2\varepsilon} \| A^{\mathrm{T}} u \|^2$$
$$x(\varepsilon) = \tfrac{1}{\varepsilon} A^{\mathrm{T}} u(\varepsilon)$$

D_p: $$\max_{u\in R_+^m} (b^{\mathrm{T}} u - \tfrac{1}{2} \| A^{\mathrm{T}} u \|^2)$$
$$x^* = A^{\mathrm{T}} u^*$$

D_{pr}: $$\max_{u\in R_+^m} (b^{\mathrm{T}} u - \tfrac{1}{2} \| A^{\mathrm{T}} u \|^2 - \tfrac{\varepsilon}{2} \| u \|^2)$$
$$x(\varepsilon) = A^{\mathrm{T}} u(\varepsilon)$$

Primal problems

P_0: $\min_{x \in X} 0_n^{\mathrm{T}} x,$

$$X = \{x \in R^n : Ax = b, \quad x \geq 0_n\}$$

Dual problems

D_0: $\max_{u \in U} b^{\mathrm{T}} u,$

$$U = \{u \in R^m : A^{\mathrm{T}} u \leq 0_n\}$$

P_r: $\min_{x \in X} \left(0_n^T x + \tfrac{1}{2} \| x - \hat{x} \|^2 \right),$

$$X = \{x \in R^n : \; Ax = b, \;\; x \geq 0_n\}$$

D_p: $\max_{u \in R^m} \left(b^T u - \tfrac{1}{2} \| (\hat{x} + A^T u^*)_+ \|^2 + \tfrac{1}{2} \| \hat{x} \|^2 \right)$

$$x^* = (\hat{x} + A^T u^*)_+$$

P_{rp}: $\min_{x \in R_+^n} \left(\tfrac{1}{2} \| x - \hat{x} \|^2 + \tfrac{1}{2\varepsilon} \| b - Ax \|^2 \right)$

$$u(\varepsilon) = \tfrac{1}{\varepsilon}(b - Ax(\varepsilon))$$

D_{pr}: $\max_{u \in R^m} \left(b^T u - \tfrac{1}{2} \| (\hat{x} + A^T u)_+ \|^2 - \tfrac{\varepsilon}{2} \| u \|^2 + \tfrac{1}{2} \| \hat{x} \|^2 \right)$

$$x(\varepsilon) = (\hat{x} + A^T u(\varepsilon))_+$$

P_p: $\min_{x \in R_+^n} \tfrac{1}{2} \| b - Ax \|^2$

$$u^* = b - Ax^*$$

D_1: $\max_{u \in U} (b^{\mathrm{T}} u - \tfrac{1}{2} \| u \|^2),$

$$U = \{u \in R^m, A^{\mathrm{T}} u \leq 0_n\}$$

P_{pr}: $\min_{x \in R_+^n} \left(\tfrac{1}{2} \| b - Ax \|^2 + \tfrac{\varepsilon}{2} \| x - \hat{x} \|^2 \right)$

$$u(\varepsilon) = b - Ax(\varepsilon)$$

D_{rp}: $\max_{u \in R^m} \left(b^T u - \tfrac{1}{2} \| u \|^2 - \tfrac{\varepsilon}{2} \| (\hat{x} + \tfrac{1}{\varepsilon} A^T u)_+ \|^2 + \tfrac{\varepsilon}{2} \| \hat{x} \|\right.$

$$\hat{x}(\varepsilon) = (\hat{x} + \varepsilon^{-1} A^T u(\varepsilon))_+$$

Projection of $\hat{x}$ onto the set of nonnegative solutions of linear equation system

$$\min_{x \in X} \frac{1}{2} \| x - \hat{x} \|^2, \qquad X = \{x \in R^n : \ Ax = b, \quad x \geq 0_n\} \qquad (\mathrm{P_r})$$

$$\max_{u \in R^m} \left(b^T u - \frac{1}{2} \| (\hat{x} + A^T u)_+ \|^2 + \frac{1}{2} \| \hat{x} \|^2 \right) \qquad (\mathrm{D_p})$$

Proposition 1. The unique solution of the Problem $(\mathrm{P_r})$ is expressed via the solution of the dual problem $(\mathrm{D_p})$ as

$$x^* = (\hat{x} + A^T u^*)_+$$

Newton Method

$$\max_{u \in R^m} S(u), \qquad S(u) = b^T u - \frac{1}{2} \| (\hat{x} + A^T u)_+ \|^2$$

$$\text{gradient:} \qquad S_u(u) = b^T - A(\hat{x} + A^T u)_+$$

$$\text{generalized Hessian:} \quad S_{uu}(u) = -AD(u)A^T,$$

$$\text{where} \quad D(u) = \mathrm{Diag}(sign((\hat{x} + A^T u)_+))$$

$$u^{k+1} = u^k - \tau^k \Delta u$$

$$H^k \Delta u = -S_u(u^k)$$

$$H^k = -AD(u^k)A^{\mathrm{T}} - \delta I_m$$

$$\tau^k = \arg\max_{\tau} S(u^k - \tau \Delta u)$$

The Dual for the Regularized & Penalized Problem (P_{rp})

$$\min_{x \in R^n_+} F(x), \quad F(x) = \frac{1}{2}\| x - \hat{x} \|^2 + \frac{1}{2\varepsilon}\| b - Ax \|^2 \qquad (P_{rp})$$

$$\max_{u \in R^m} W(u), \quad W(u) = b^T u - \frac{1}{2}\| (\hat{x} + A^T u)_+ \|^2 - \frac{\varepsilon}{2}\| u \|^2 \quad (D_{pr})$$

Proposition 2. For any $\varepsilon \geq 0$ the unique solution $x(\varepsilon)$ of Problem (P_{rp}) can be determined via the unique solution $u(\varepsilon)$ of Problem (D_{pr}) as

$$x(\varepsilon) = (\hat{x} + A^T u(\varepsilon))_+$$

and $x(\varepsilon) \to \hat{x}^*$ if $\varepsilon \to 0$.

gradient: $\quad W_u(u) = b^T - A(\hat{x} + A^T u)_+ - \varepsilon u$

generalized Hessian: $\quad W_{uu}(u) = -AD(u)A^T - \varepsilon I_m, \quad \Rightarrow \quad \delta = 0$

where $\quad D(u) = \mathrm{Diag}(sign((\hat{x} + A^T u)_+))$

$$\max_{u\in R^m} (b^T u - \frac{1}{2}\|(\hat{x} + A^T u)_+\|^2) \qquad \max_{u\in R^m} (b^T u - \frac{1}{2}\|(\hat{x} + A^T u)_+\|^2 - \frac{\varepsilon}{2}\|u\|^2)$$

$m\times n$	d	it	$\|Ax-b\|_\infty$	t, sec	it	$\|Ax-b\|_\infty$	t, sec
250×10^4	1	6	$2{,}8\times10^{-11}$	2,8	6	$2{,}6\times10^{-11}$	2,7
500×10^4	1	6	$5{,}6\times10^{-11}$	8,7	6	$3{,}7\times10^{-11}$	8,9
1000×10^4	1	7	$6{,}0\times10^{-11}$	35,5	7	$4{,}1\times10^{-11}$	35,1
2000×10^4	1	8	$4{,}4\times10^{-11}$	166,3	8	$3{,}7\times10^{-11}$	156,1
3000×10^4	1	9	$6{,}2\times10^{-11}$	446,1	9	$6{,}5\times10^{-11}$	451,6
4000×10^4	1	10	$7{,}4\times10^{-11}$	822,1	10	$7{,}0\times10^{-11}$	825,1
1000×10^5	0,5	6	$1{,}8\times10^{-10}$	167,6	6	$1{,}4\times10^{-10}$	164,9
2000×10^5	0,5	6	$5{,}8\times10^{-10}$	671,7	6	$5{,}0\times10^{-10}$	658,9
3000×10^5	0,5	6	$4{,}5\times10^{-10}$	1479,0	6	$5{,}3\times10^{-10}$	1414,0
4000×10^5	0,5	6	$5{,}2\times10^{-10}$	2508,5	6	$5{,}9\times10^{-10}$	2470,3
500×10^6	0,1	6	$1{,}5\times10^{-10}$	57,4	6	$2{,}2\times10^{-10}$	59,3
1000×10^6	0,1	6	$1{,}5\times10^{-09}$	180,5	6	$1{,}6\times10^{-09}$	180,1
2000×10^6	0,1	6	$1{,}2\times10^{-09}$	685,8	6	$1{,}1\times10^{-09}$	687,6

Newton Method 2

$$\max_{u \in R^m} S(u), \qquad S(u) = b^T u - \frac{1}{2} \|(\hat{x} + A^T u)_+\|^2$$

gradient: $\quad S_u(u) = b^T - A(\hat{x} + A^T u)_+$

generalized Hessian: $\quad S_{uu}(u) = -AD(u)A^T,$

where $\quad D(u) = \text{Diag}(sign((\hat{x} + A^T u)_+))$

$$u^{k+1} = u^k - \tau^k \Delta u$$

$$H^k \Delta u = -S_u(u^k)$$

where $H^k = -AD(u^k)A^T - \delta \text{Diag}(AA^T)$, (instead of $\ldots - \delta I_m$)

and $\tau^k \in \{1,\ 1/2,\ 1/4,\ \ldots\}$ is chosen using Armijo rule

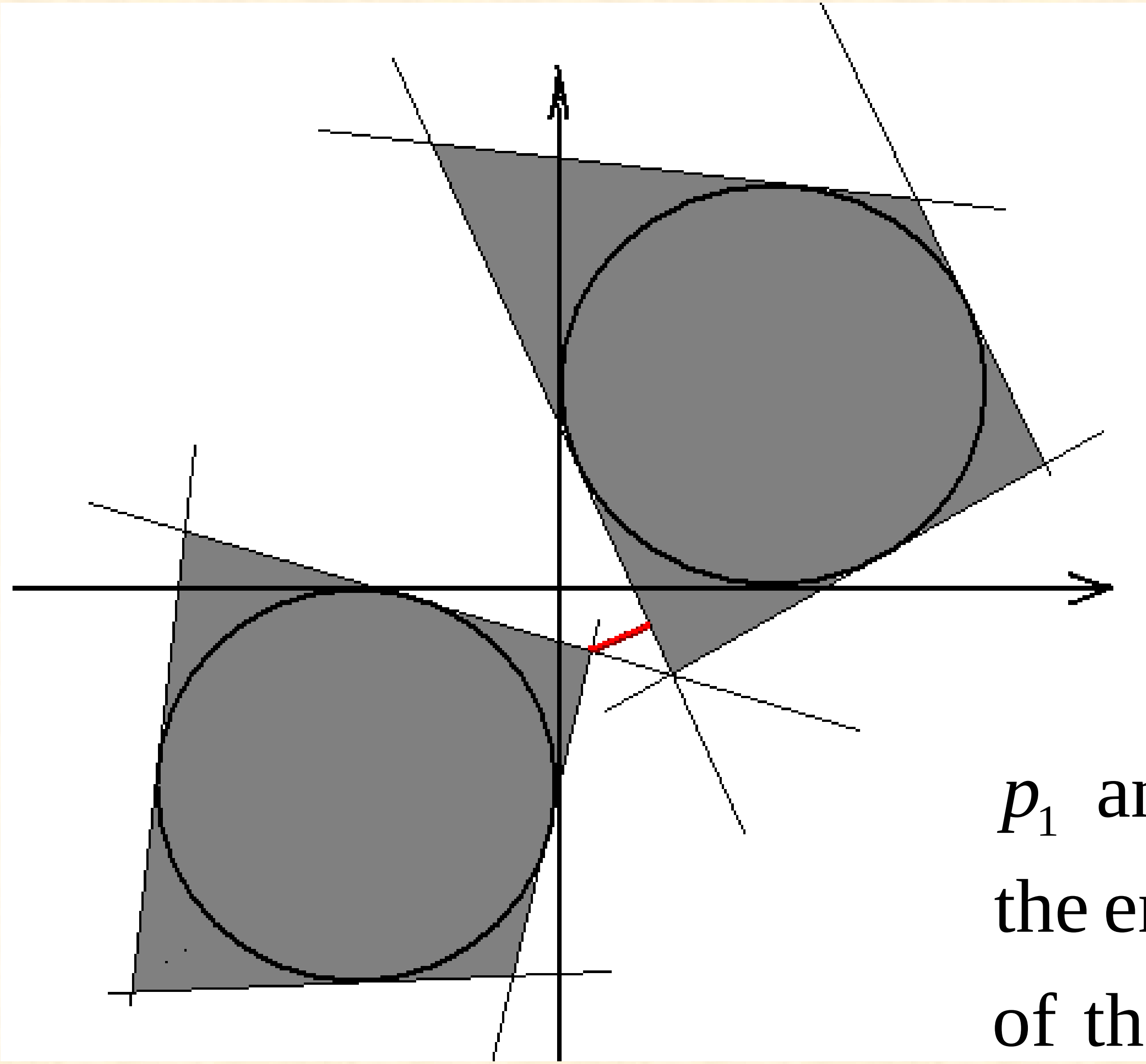
Свойства матриц A для 11 задач из библиотеки NetLib

name	m	n	d	$\|\|x^*\|\|$	$\min_i (AA^T)_{ii}$	$\max_i (AA^T)_{ii}$
Afiro	27	51	0.07	634.029569	1.184900	44.95628
Addlittle	56	138	0.05	430.764399	1.000000	10654.00
agg3	516	758	0.01	765883.022	1.000000	179783.8
25fv47	821	1876	0.007	3310.45652	0.000000	88184.03
pds_02	2953	7716	0.00007	60697.180	1.000000	91.00000
cre_ a	3516	7248	0.00007	1162.32987	0.000000	27476.84
80bau3b	2262	12061	0.0008	4129.96530	1.000000	321739.6
ken_13	28362	42659	0.00008	25363.3224	1.000000	170.0000
maros r7	3136	9408	0.005	141313.207	3.051759	3.371325
cre_b	9648	77137	0.003	624.270129	0.000000	27476.84
osa_14	2337	54797	0.002	119582.321	18.00000	845289.9

name	solver	time(sec)	$\| Ax - b \|_\infty$	NewtIter
afiro	GNewtEGK	0.001	8.63E−11	17
–”–	cqpMOSEK	0.31	1.13E−13	–
addlittle	GNewtEGK	0.003	6.45E−10	22
–”–	cqpMOSEK	0.35	7.18E−11	–
agg3	GNewtEGK	0.14	3.93E−07	116
–”–	cqpMOSEK	0.40	2.32E−10	–
25fv47	GNewtEGK	0.54	7.15E−10	114
–”–	cqpMOSEK	1.36	1.91E−11	–
pds 02	GNewtEGK	0.32	1.55E−08	75
–”–	cqpMOSEK	0.51	8.20E−06	–

name	solver	time(sec)	$\|\| Ax - b \|\|_\infty$	NewtIter
cre_a	GNewtEGK	3.36	2.64E-09	219
-”-	cqpMOSEK	0.61	2.15E-10	-
80bau3b	GNewtEGK	0.27	3.33E-09	79
-”-	cqpMOSEK	0.80	2.90E-07	-
ken_13	GNewtEGK	1.41	2.70E-08	55
-”-	cqpMOSEK	2.09	1.71E-09	-
maros_r7	GNewtEGK	0.10	1.18E-09	27
-”-	cqpMOSEK	55.20	3.27E-11	-
cre_b	GNewtEGK	9.25	6.66E-10	75
-”-	cqpMOSEK	2.31	1.61E-06	-
osa_4	GNewtEGK	42.59	8.25E-08	767
-”-	cqpMOSEK	4.40	7.82E-05	-

Euclidean Distance Between Two Convex Polyhedrons



p_1 and p_2 are the endpoints of the red segment

Convex Piecewise Quadratic Formulation

$$\text{Polyhedron } \Pi_1 : \quad A_1^T p_1 - c_1 \le 0$$

$$\text{Polyhedron } \Pi_2 : \quad A_2^T p_2 - c_2 \le 0$$

$$\text{Distance} : \quad \| p_1 - p_2 \| \to \min$$

$$A = \begin{bmatrix} A_1^T & 0 \\ 0 & A_2^T \end{bmatrix}, \quad p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad B = \begin{bmatrix} I_d & -I_d \\ -I_d & I_d \end{bmatrix}$$

$$\min_{p \in R^{2d}} \varphi(p); \qquad \varphi(p) = \frac{1}{2} \| (A^T p - c)_+ \|^2 + \frac{\varepsilon}{2} p^T B p$$

Newton Method

gradient: $g(p) = A(A^T p - c)_+ + \varepsilon Bp$

gen. Hessian: $H(p) = AD(p)A^T + \varepsilon B,$

where $D(p) = \text{Diag}(\text{sign}((A^T p - c)_+))$

$$p_{k+1} = p_k - (\delta I_m + H(p_k))^{-1} g(p_k)$$

1. Mangasarian O.L. A Newton method for linear programming // J. Optim. Theory Appl. – 2004. – Vol. 121, – N. 1. – P. 1–18.
2. Golikov A. I., Evtushenko Y. G. A new method for solving systems of linear equalities and inequalities //DOKLADY MATHEMATICS. – MAIK NAUKA, 2001. – Т. 64. – №. 3. – С. 370–373.
3. V. A. Garanzha, A. I. Golikov, Yu. G. Evtushenko, M. Kh. Nguen. Parallel implementation of Newton's method for solving large-scale linear programs // Computational Mathematics and Mathematical Physics, 2009, Volume 49, Issue 8, pp. 1303–1317.
4. Ganin B. V., Golikov A. I., Evtushenko Y. G. Projective-dual method for solving systems of linear equations with nonnegative variables // Computational Mathematics and Mathematical Physics. – 2018. – Т. 58. – №. 2. – С. 159–169.
5. S. Ketabchi, H. Moosaei, M. Parandegan, H. Navidi. Computing Minimum Norm Solution of Linear Systems of Equations by the Generalized Newton Method // Numerical Algebra Control and Optimization. – 2017 – V.7 (2). – P. 113–119.
6. Голиков А. И., Евтушенко Ю. Г., Капорин И. Е. Метод ньютоновского типа для решения систем линейных уравнений и неравенств //Журнал вычислительной математики и математической физики.– 2019. – Т. 59. –– №. 12. – С. 2086 – 2101.

Thank You for Your Attention!