Linear Programming:

Part (i): Strict Feasibility and Degeneracy (Pg 2)
Part (ii): Exterior Point Path Following Algorithm (Pg 70)

Combinatorics & Optimization, University of Waterloo
Prof. Henry Wolkowicz
hwolkowicz@uwaterloo.ca

Tues. Mar. 28, 2023

LP Part (i): Strict Feasibility and Degeneracy

Henry Wolkowicz

Dept. Comb. and Opt., University of Waterloo, Canada

Tues. Mar. 28, 2023

joint work with: Jiyoung Im, Univ. of Waterloo

Motivation/Main Results

Background

- Currently: simplex and interior point methods are most popular algorithms for solving linear programs, LPs.
- Unlike general conic programs, (finite) LPs do not require strict feasibility for strong duality. Hence strict feasibility is often less emphasized.

We show that lack of strict feasibility:

1. causes numerical difficulties in both simplex and interior point methods.
2. and $\implies$ all basic feasible solutions, BFS, are degenerate

We present

an extension of Phase-I of simplex method for preprocessing for strict feasibility

Background and Notation

Feasible LPs; standard form (with FINITE opt. value)

$$(\mathcal{P}) \quad \text{(finite) } p^* = \begin{array}{ll} \min_x & c^T x \\ \text{s.t.} & Ax = b \in \mathbb{R}^m \\ & x \in \mathbb{R}^n_+ \end{array}$$

assume wlog $\operatorname{rank}(A) = m$;

with feasible set: $\mathcal{F} = \{x \in \mathbb{R}^n : Ax = b, \ x \geq 0\}$

Dual LP

$$(\mathcal{D}) \quad p^* = d^* = \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y \leq c \in \mathbb{R}^n \\ & y \in \mathbb{R}^m \end{array}$$

(equivalently $A^T y + s = c, s \geq 0$ slack)

History: Kantorovich; Dantzig, Karmarkar

Kantorovich '39, USSR, WWII

- transportation models and optimal solutions (algorithm)
- helped NKVD with transportation problems

Dantzig '47, USA, SIMPLEX METHOD

- following duality/game-theory by Von Neumann
- Hotelling: "but the world is nonlinear"
- Von Neumann: "if you have a linear model, you can now solve it"
- SIAM survey 1970's: 70% of ALL world computer time is spent on the simplex method

Karmarkar '84, Interior Point Revolution

- Lustig-Marsten-Shanno OB1 code '90; large went from: $(m = 1e3 \times n = 1e4)$ to $(m = 1e5 \times n = 1e7)$
- to modern day: $(m = 1e6 \times n = 1e10)$

Strict Feasibility, Slater, Mangasarian-Fromovitz CQ

Feasible LPs; standard form (with FINITE opt. value)

$$(\mathcal{P}) \quad \text{(finite) } p^* = \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \in \mathbb{R}^m \\ & x \in \mathbb{R}^n_+ \end{array}$$

there exists $\hat{x}$ with $A\hat{x} = b, \hat{x} > 0$ (MFCQ)

Dual LP

$$(\mathcal{D}) \quad p^* = d^* = \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y \leq c \in \mathbb{R}^n \\ & y \in \mathbb{R}^m \end{array}$$

there exists $\hat{y}$ with $A^T \hat{y} < c$ (Slater CQ)

Stability: MFCQ/Slater $\iff$

stability wrt RHS perturbations

$\iff$ compact set of dual variables

Basic (Feasible/Degenerate) Solutions

Definition (basic (feasible) solution)

- Given: $x \in \mathbb{R}^n, Ax = b$ and $\mathcal{B} \subset \{1, \ldots, n\}$, $|\mathcal{B}| = m$; let $\mathcal{N} = \{1 \ldots n\} \backslash B$.
  Then x is a basic solution if
  $A(:, \mathcal{B})$ is nonsingular and $x_i = 0, \ \forall i \in \mathcal{N}$
- x is a basic feasible solution, BFS, if in addition $x \geq 0$. It is degenerate, if $\exists i \in \mathcal{B}, x_i = 0$

Equivalently, if $Ax = b, x \geq 0$ (feasible):

x is basic if there exists
$\mathcal{N} \subset \{1, \ldots, n\}, |\mathcal{N}| = n - m, x_i = 0, \forall i \in \mathcal{N}$;
and the corresponding matrix of active constraints

$$\begin{bmatrix} A \\ I_{\mathcal{N}} \end{bmatrix} \quad \text{is nonsingular.}$$

It is degenerate if there are redundant active constraints.

Two Kinds of Degeneracy

Definition (Degenerate BFS)

$$x \text{ BFS is } \begin{cases} \text{nondegenerate}, & \text{if } x_i > 0,\ \forall i \in \mathcal{B}, \\ \text{degenerate}, & \text{otherwise} \end{cases}$$

Definition (variable fixed at 0)

Let $i_0 \in \mathcal{I} = \{1, \ldots, n\}$. x_{i_0} is <u>fixed at 0</u> if $x_{i_0} = 0, \forall x \in \mathcal{F}$. Let

$$\mathcal{I}^{=} = \{i \in \mathcal{I} : x_i \text{ is fixed at } 0\},\ \mathcal{I}^{<} = \mathcal{I} \backslash \mathcal{I}^{=}$$

$\bar{x}$ a degenerate BFS with basis $\mathcal{B}$ is of type:

1. if: $i \in \mathcal{B}, \bar{x}_i = 0 \implies i \in \mathcal{I}^{<}$
2. if: there exists $i \in \mathcal{B} \cap \mathcal{I}^{=}$

Below we see that:
if Type 2 exists, then ALL BFS are of Type 2.

Facial Reduction, FR, for LPs that fail Strict Feasibility

Two Steps

- obtain an equivalent problem with strict feasibility;
- recover full-row rank for the constraint matrix (always needed for MFCQ)

Definition (Face of a convex set K)

A convex set $F \subseteq K \subseteq \mathbb{R}^n$ is a face of K, denoted $F \trianglelefteq K$, if

$$y, z \in K, x = \frac{1}{2}(y+z) \in F \implies y, z \in F.$$

The minimal face for F, $\operatorname{face}(F)$, is the intersection of all faces of K containing C.

faces of $\mathbb{R}^n_+$, nonnegative orthant

for fixed indices $\hat{\mathcal{I}} \subseteq \{1, \ldots, n\}$

$$F = \{x \in \mathbb{R}^n_+ : x_i = 0, \forall i \in \hat{\mathcal{I}}\}$$

Facial Reduction; Basics

Theorem (DW: [12, Theorem 3.1.3] Theorem of the Alternative)

For the feasible system $\mathcal{F}$ of the LP, exactly one of the following statements holds:

1. *There exists $x \in \mathbb{R}^n_{++}$ with $Ax = b$, i.e., strict feasibility holds;*
2. *There exists $y \in \mathbb{R}^m$ such that*

$$(*) \quad 0 \neq z := A^T y \in \mathbb{R}^m_+, \quad \text{and} \ \langle b, y \rangle = 0,$$

exposing vector $z \in \mathbb{R}^n_+$

(*) is equivalent to:
exposing vector $0 \neq z \geq 0$ exists for the minimal face containing the feasible set, i.e.,

$$\begin{aligned} x \in \mathcal{F} \quad &\iff \quad Ax = b, x \geq 0 \\ &\implies \quad \langle z, x \rangle = \langle A^T y, x \rangle = \langle y, Ax \rangle = \langle y, b \rangle = 0 \end{aligned}$$

Facial Reduction two steps; Outline

suppose strict feasibility fails; i.e., get exposing vector z

1. Thm of Alternative implies: $\exists 0 \lneq z = A^T y \in \mathbb{R}^m$:

$$\begin{array}{rcl} x \in \mathcal{F} & \implies & 0 \leq \langle x, z \rangle = \langle x, A^T y \rangle = \langle Ax, y \rangle = \langle b, y \rangle = 0 \\ & \implies & 0 = x \circ z \\ & \iff & 0 = x_j z_j = 0, \forall j \\ & & \quad \text{yields complementary unit vectors } e_k \end{array}$$

cardinality of support of z: $s_z = |\{i : z_i > 0\}|$

2. $z = \sum_{j=1}^{s_z} z_{t_j} e_{t_j}$, t_j nondecreasing order

$x = \sum_{j=1}^{n-s_z} x_{s_j} e_{s_j}$, s_j nondecreasing order.

$V = \begin{bmatrix} e_{s_1} & e_{s_2} & \dots & e_{s_{n-s_z}} \end{bmatrix} \in \mathbb{R}^{n \times (n-s_z)}, \qquad Vz = 0.$

3. $\mathcal{F} = \{x \in \mathbb{R}^n_+ : Ax = b\} = \{x = Vv \in \mathbb{R}^n : AVv = b, v \in \mathbb{R}^{n-s_z}_+\}$
4. Recover full row rank: $A \leftarrow P_{\bar{m}} AV, b \leftarrow P_{\bar{m}} b$

Facial Reduction, FR; Two Steps

matrix $V \in \mathbb{R}^{n\times(n-s_z)}$, facial range vector

Every facial reduction step yields at least one redundant constraint, BW: [7],IW: [18, Lemma 2.7],S: [31, Section 3.5].

Lemma (step 2: redundant constraint)

Consider the facially reduced feasible set

$$\mathcal{F}_r = \left\{v : AVv = b, v \in \mathbb{R}^{n-s_z}_+\right\}.$$

Then at least one linear constraint of the LP *is redundant.*

Proof.

Let: $0 \neq z = A^T y \geq 0$ *exposing vector;* V *corresponding facial range vector; Then:*

$$0 = V^T z = V^T A^T y = (AV)^T y = \sum_{i=1}^m y_i((AV)^T)_i$$

Since $0 \neq y \in \mathbb{R}^m$*, the rows of* AV *are linearly dependent.* □

Summary FR

Result of full two step FR: strict feas.; full rank

$$\begin{aligned}\mathcal{F} &= \{x \in \mathbb{R}^n_+ : Ax = b\} \\ &= \{x = Vv \in \mathbb{R}^n : \bar{A}v := (P_{\bar{m}}AV)v = (P_{\bar{m}}b) =: \bar{b}, \\ &\qquad v \in \mathbb{R}^{n-s_z}_+\}\end{aligned}$$

- after substit: $\min (V^T c)^T v$ s.t. $\bar{A}v = \bar{v}$, $v \in \mathbb{R}^{n-s_z}_+$
- $\exists \hat{v} > 0, \bar{A}\hat{v} = \bar{b}$ (MFCQ)
- full rank $\bar{A} = P_{\bar{m}}AV$: $P_{\bar{m}} : \mathbb{R}^m \to \mathbb{R}^{\bar{m}}$, $\bar{m} = \operatorname{rank}(AV) < m$. $P_{\bar{m}}$ is projection that chooses the linearly independent rows of AV.
- BOTH # variables, # constraints are strictly reduced.

This emphasizes the ILL-CONDITIONING of problems where strict feasibility fails, i.e., Implicit singularity is eliminated using FR.

Two-Step Facial Reduction; $Ax = b$, $x \geq 0$

Facial Reduction, FR

a journey to reformulate a problem until strict feasibility is met

Two-Step Facial Reduction; $Ax = b$, $x \geq 0$

Facial Reduction, FR

a journey to reformulate a problem until strict feasibility is met

Solve the auxiliary system:

$$\text{Find } y \in \mathbb{R}^m \text{ s.t. } A^T y \in \mathbb{R}^n_+ \setminus \{0\}, \quad \langle b, y \rangle = 0$$

Set $V = I(:, \operatorname{supp}(A^T y)^c)$
$x \leftarrow Vv$
$\mathcal{F} \leftarrow \{v \geq 0 : (AV)v = b\}$

Two-Step Facial Reduction; $Ax = b$, $x \geq 0$

Facial Reduction, FR

a journey to reformulate a problem until strict feasibility is met

[STEP 1]
Solve the auxiliary system:

$$\text{Find } y \in \mathbb{R}^m \text{ s.t. } A^T y \in \mathbb{R}^n_+ \setminus \{0\}, \quad \langle b, y \rangle = 0$$

Set $V = I(:, \text{supp}(A^T y)^c)$
$x \leftarrow Vv$
$\mathcal{F} \leftarrow \{v \geq 0 : (AV)v = b\}$

[STEP 2]

Any nontrivial FR

$\Downarrow$

discovery of redundant equalities

Use $P_{\bar{m}}$ to discard redundancies

$\mathcal{F} \leftarrow \{v \geq 0 : P_{\bar{m}}AV(v) = P_{\bar{m}}b\}$

Two-Step Facial Reduction; $Ax = b,\ x \geq 0$

Facial Reduction, FR

a journey to reformulate a problem until strict feasibility is met

[STEP 1]
Solve the auxiliary system:

$$\text{Find } y \in \mathbb{R}^m \text{ s.t. } A^T y \in \mathbb{R}^n_+ \setminus \{0\},\quad \langle b, y \rangle = 0$$

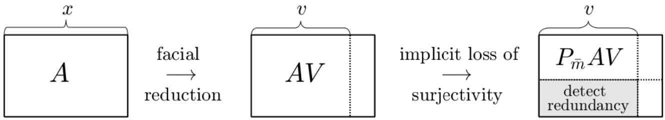
Set $V = I(:, \mathrm{supp}(A^T y)^c)$
$x \leftarrow Vv$
$\mathcal{F} \leftarrow \{v \geq 0 : (AV)v = b\}$

[STEP 2]

Any nontrivial FR
$\Downarrow$
discovery of redundant equalities

Use $P_{\bar{m}}$ to discard redundancies

$\mathcal{F} \leftarrow \{v \geq 0 : P_{\bar{m}}AV(v) = P_{\bar{m}}b\}$



Example

Consider $\mathcal{F}$ with the data

$$A = \begin{bmatrix} 1 & 1 & 3 & 5 & 2 \\ 0 & 1 & 2 & -2 & 2 \end{bmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\text{Set } y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \implies A^T y = \begin{pmatrix} 1 & 0 & 1 & 7 & 0 \end{pmatrix}^T \geq 0 \text{ and } \langle b, y \rangle = 0.$$

$$V = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad x \leftarrow Vv = \begin{pmatrix} 0 \\ v_1 \\ 0 \\ 0 \\ v_2 \end{pmatrix}, \quad Ax = b \leftarrow AVv = b \equiv \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(*) Side note

There are exactly six feasible bases in $\mathcal{F}$; (BFS all degenerate).

- $\mathcal{B} \in \{\{1,2\},\{2,3\},\{2,4\}\}$ is $x = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T$;
- $\mathcal{B} \in \{\{1,5\},\{3,5\},\{4,5\}\}$ is $x = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}^T$.

Example

Consider $\mathcal{F}$ with the data

$$A = \begin{bmatrix} 1 & 1 & 3 & 5 & 2 \\ 0 & 1 & 2 & -2 & 2 \end{bmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\text{Set } y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \implies A^T y = \begin{pmatrix} 1 & 0 & 1 & 7 & 0 \end{pmatrix}^T \geq 0 \text{ and } \langle b, y \rangle = 0.$$

$$V = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad x \leftarrow Vv = \begin{pmatrix} 0 \\ v_1 \\ 0 \\ 0 \\ v_2 \end{pmatrix}, \quad Ax = b \leftarrow AVv = b \equiv \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(*) Side note

There are exactly six feasible bases in $\mathcal{F}$; (BFS all degenerate).

- $\mathcal{B} \in \{\{1,2\},\{2,3\},\{2,4\}\}$ is $x = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T$;
- $\mathcal{B} \in \{\{1,5\},\{3,5\},\{4,5\}\}$ is $x = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}^T$.

Detect Redundancy

Lemma (AV is rank deficient)

Consider the facially reduced feasible set

$$\mathcal{F}_r = \left\{ v : AVv = b, v \in \mathbb{R}_+^{n-s_z} \right\}.$$

Then at least one linear equality of $AVv = b$ is redundant.

(proof) Let $z = A^T y$ be the exposing vector, V be a facial range vector induced by z. Then

$$0 = V^T z = V^T A^T y = (AV)^T y.$$

Found a nontrival row combination of AV, i.e., detected redundancy

Definition (implicit problem singularity)

The implicit problem singularity (**ips**) = The number of implicit redundant equalities of $\mathcal{F}$

Detect Redundancy

Lemma (AV is rank deficient)

Consider the facially reduced feasible set

$$\mathcal{F}_r = \left\{ v : AVv = b, v \in \mathbb{R}_+^{n-s_z} \right\}.$$

Then at least one linear equality of $AVv = b$ is redundant.

(proof) Let $z = A^T y$ be the exposing vector, V be a facial range vector induced by z. Then

$$0 = V^T z = V^T A^T y = (AV)^T y.$$

Found a nontrival row combination of AV, i.e., detected redundancy

Definition (implicit problem singularity)

The implicit problem singularity (**ips**) = The number of implicit redundant equalities of $\mathcal{F}$

Detect Redundancy

Lemma (AV is rank deficient)

Consider the facially reduced feasible set

$$\mathcal{F}_r = \left\{ v : AVv = b, v \in \mathbb{R}_+^{n-s_z} \right\}.$$

Then at least one linear equality of $AVv = b$ is redundant.

(proof) Let $z = A^T y$ be the exposing vector, V be a facial range vector induced by z. Then

$$0 = V^T z = V^T A^T y = (AV)^T y.$$

Found a nontrival row combination of AV, i.e., detected redundancy

Definition (implicit problem singularity)

The implicit problem singularity (**ips**) = The number of implicit redundant equalities of $\mathcal{F}$

Singularity Degree $sd(\mathcal{F})$, Sturm '20 [32]

Definition ($d = sd(\mathcal{F}) = \min |FR\ steps|$)

Definition (Hölder regularity)

the pair of closed, convex subsets A, B is γ-*Hölder regular* if $\forall U$ compact, $\exists c > 0$ with:

$$\operatorname{dist}(x, A \cap B) \leq c \cdot \left(\operatorname{dist}^\gamma(x, A) + \operatorname{dist}^\gamma(x, B) \right) \qquad \text{for all } x \in U.$$

Sturm [32] error bound Theorem for SDP, $\mathcal{F} = \mathcal{L} \cap \mathbb{S}^n_+$

$(\mathcal{L}, \mathbb{S}^n_+)$ is $\frac{1}{2^d}$-Hölder regular. ($\mathcal{L}$ linear manifold)

- for LPs, FR in *one* iteration using maximal exposing vector, i.e., $d = \mathbf{sd}(\mathcal{F}) \leq 1$
- FR for LPs does not alter sparsity pattern of A. (only involves discarding columns of A; rows of A, b)

A Theoretical Result on degenerate BFS $\leftrightarrow$ MFCQ

Theorem

[a] *Suppose that strict feasibility of $\mathcal{F}$ fails. Then every basic feasible solution, BFS, $x \in \mathcal{F}$ with basis $\mathcal{B}$ has $\mathcal{B} \cap \mathcal{I}^= \neq \emptyset$ and thus is degenerate.*

[a] Contrapositive found in Bertsimas-Tsitsiklis book [4, Exer. 2.19].

Proof.

- $\mathcal{F} = \{x \in \mathbb{R}^n \;:\; AVv = b,\; v \in \mathbb{R}_+^{n-s_z}\}$, facial range vctr V
- wlog $V = \begin{bmatrix} I_r \\ 0 \end{bmatrix}$ and $r = n - s_z$;
- recall by redundant constraint lemma: $\operatorname{rank} AV < m$
- implies $\operatorname{rank} A(:, \{1, \ldots, r\}) < m$
- BFS implies $\operatorname{rank} A(:, \mathcal{B}) = m$; implies $\exists i \in \mathcal{B}, i > r$
- implies $\exists i \in \mathcal{B} \cap \mathcal{I}^=, x_i = 0$ (degeneracy) $\square$

Corollary, Stability, Converse

Corollary (contrapositive motivates phase I part 2)

If there exists a nondegenerate basic feasible solution, then there exists a strictly feasible point in $\mathcal{F}$.

Stability from above corollary

Recall: strict feasibility (and full rank, MFCQ) is equivalent to stability wrt RHS perturbations.

Example (converse fails; all BFS degenerate $\not\Longrightarrow$ MFCQ fails)

$A = \begin{bmatrix} 1 & 0 & 2 & 0 & -2 \\ 1 & -3 & 2 & 1 & -2 \end{bmatrix}; b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad 0 < x = \frac{1}{10}\begin{pmatrix} 1 & 1 & 5.5 & 3 & 1 \end{pmatrix}^T$
4 deg. feas. bases: $\mathcal{B} = \{\{1,2\},\{1,4\} : x = (1,0,0,0,0)^T$
$\mathcal{B} = \{2,3\},\{3,4\} : x = (0,0,1/2,0,0)^T$
(Also, the linear assignment problem is highly degenerate but has a strictly feasible point (average).)

Empirics for FR Preprocessing

We want to avoid implicit singularity

- improve conditioning, number of iterations

interior point methods

- Condition number of normal equation system
- stopping criteria

$$\text{KKT} = \left(\frac{\|Ax^* - b\|}{1 + \|b\|}, \frac{\|A^T y^* + s^* - c\|}{1 + \|c\|}, \frac{\langle x^*, s^* \rangle}{n} \right).$$

simplex methods (NETLIB data set)

- percentage of degenerate iterations

Optimality Conditions at current $(x > 0, y, s > 0), \mu > 0$

$X = \mathrm{Diag}\,(x),\; S = \mathrm{Diag}\,(s).$

$$\begin{array}{rcll} A^T\Delta y + \Delta s - c & = & 0 & \text{dual feasibility} \\ A\Delta x - b & = & 0 & \text{primal feasibility} \\ S\Delta x + X\Delta s & = & \mu e & \text{complementary slackness} \end{array}$$

After block elimination, solve normal equations for Δy

- Use Δs in eqn 1 to eliminate Δs in eqn 3.
- Solve for Δx in eqn 3 and eliminate it in eqn 2.
- We get the normal equations

$$AS^{-1}XA^T\Delta y = RHS.$$

- Backsolve for $\Delta x, \Delta s$ to get the Newton direction.

Numerical Experiments with Interior Point Methods

condition numbers of normal matrix; x^*, s^* near optimal

$$\kappa\left(AD^*A^T\right), \quad \text{where } D^* = \mathrm{Diag}\,(x^*)\mathrm{Diag}\,(s^*)^{-1} \tag{1}$$

three families of instances

1. $(\mathcal{P}_{(A,b,c)})$ do not have strictly feasible points;
2. $(\bar{\mathcal{P}}_{(A,\bar{b},c)})$ have strictly feasible points;
3. $(\mathcal{P}_{(A_{FR},b_{FR},c_{FR})})$ facially reduced instances of $(\mathcal{P}_{(A,b,c)})$.

Condition Numbers of Normal Matrix Near Optimum

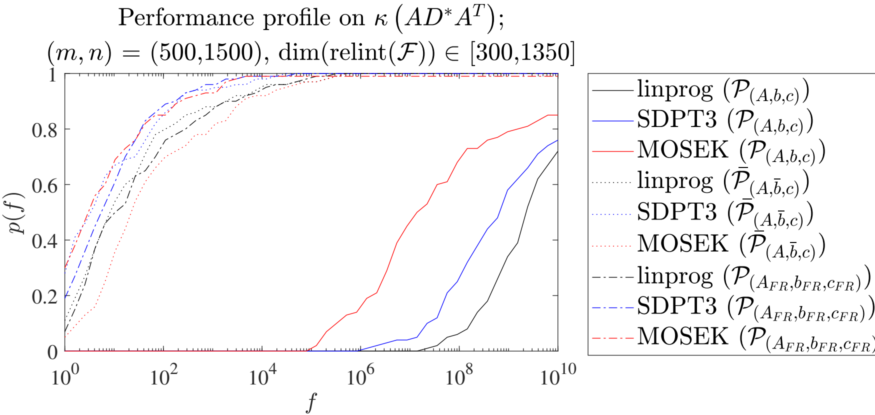


Figure: Performance profile on $\kappa\left(ADA^T\right)$ with(out) strict feasibility near optimum; various solvers

Empirics on Stopping Criteria

test the average performance of 10 instances of size $(n, m, r) = (3000, 500, 2000)$

$$\text{KKT} = \left(\frac{\|Ax^* - b\|}{1 + \|b\|}, \frac{\|A^T y^* + s^* - c\|}{1 + \|c\|}, \frac{\langle x^*, s^* \rangle}{n} \right)$$

		Non-Facially Reduced System	Facially Reduced System
linprog	KKT	(9.58e-16, 1.80e-12, 5.17e-09)	(5.78e-16, 1.51e-15, 5.57e-08)
	iter	23.30	17.60
	time	1.10	0.76
SDPT3	KKT	(1.51e-10, 1.49e-12, 4.67e-03)	(8.54e-12, 3.75e-16, 4.19e-06)
	iter	25.40	19.80
	time	0.82	0.53
MOSEK	KKT	(8.40e-09, 7.54e-16, -5.16e-06)	(5.16e-09, 3.81e-16, -2.03e-08)
	iter	35.90	10.10
	time	0.58	0.31

Table: Average of KKT conditions, iterations and time of (non)-facially reduced problems

Numerical Experiments with (Dual) Simplex Method

Empirics on the Number of Degenerate Iterations

- MOSEK (values in the table) reports percentage of degenerate iterations i.e„ 'DEGITER(%)' is ratio of degenerate iterations. (smaller value is better).
- $r = |\text{supp}(s)|$; smaller value $(r/n)\%$ means entries of s are identically 0; 100% means strict feasibility holds.
- note significant decrease in 'DEGITER(%)'.

		$(r/n)\%$				
		60%	70%	80%	90%	100%
(n,m)	(1000, 250)	36.62	10.18	0.01	0.02	0.00
	(2000, 500)	39.72	18.28	0.07	0.15	0.01
	(3000, 750)	25.99	10.66	0.32	0.75	0.02
	(4000, 1000)	29.78	18.25	0.25	0.53	0.02

Table: Average of ratio of degenerate iterations DEGITER(%)

Phase I(b): Towards Strict Feasibility

- $\bar{x}, \mathcal{B}$ degenerate BFS/basis; Wlog basic variables located first $\bar{x}$ as are degenerate variables. Solve (using basis from phase I simplex method)

$$p_1^* = \max\{x_1 \ : \ Ax = b, \ x \geq 0\}.$$

1. Suppose that $p_1^* > 0$. Then, the the variable x_1 is not an identically 0 variable, i.e., $1 \notin \mathcal{I}_0$.
2. Suppose that $p_1^* = 0$. Then, the variable x_1 is an identically 0 variable, i.e., $1 \in \mathcal{I}_0$. Let $\mathcal{B}^*$ be an optimal basis. Then we have an exposing vector

$$y^* = A(:, \mathcal{B}^*)^T e_1, \ \langle b, y^* \rangle = 0 \ \text{ and } A^T y^* \geq e_1.$$

- Add up certificates: $y^\circ = \sum_j y^j$ to get exposing vector

$$A^T y^\circ = \sum_j A^T y^j \geq 0, A^T y^\circ \neq 0, \langle b, y^\circ \rangle = \sum_j \langle b, y^j \rangle = 0.$$

Conclusion

- loss of strict feasibility has many applications recent survey Drusvyatskiy-W. [12].
- though not needed theoretically in LP, loss of MFCQ results in stability/numerical issues.
- In the paper we introduced new concept: Implicit Singularity Degree, maximum number of FR steps, and presented an algorithm, phase I (b), that regularizes an LP, for strict feasibility holding.

Regularized Nonsmooth Newton Algorithms for Best Approximation with Applications

Prof. Henry Wolkowicz
hwolkowicz@uwaterloo.ca

Tues. Mar. 28, 10:00-11:20 EST, 2023

joint work with: Yair Censor (Univ. of Haifa);
Walaa Moursi and Tyler Weames (Univ. of Waterloo)

Motivation/Main Results

Main Problem/Best Approximation

Given $v \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedral set,
find the nearest point to v from the set P

Nonsmooth Algorithms

• Application of Moreau Decomposition/elegant equation
• present regularized nonsmooth method; singular Jacobian
• compare computational performance to classical projection methods (e.g., HLWB projection method)

Applications

solving large scale linear programs; triangles from branch and bound methods; generalized constrained linear least squares.

Notation

best approximation problem to polyhedral set $P \subset \mathbb{R}^n$

find the nearest point $x^* \in P$ to a given point $v \in \mathbb{R}^n$

uniquely attained optimum (projection of v onto P)

$$\text{optimum: } x^*(v) = \operatorname{argmin}_{x \in P} \frac{1}{2}\|x - v\|^2$$

optimal value: $p^*(v) = \frac{1}{2}\|x^*(v) - v\|^2$

Nonsmooth Newton Method

We apply a
(regularized/scaled) nonsmooth Newton method to a special form of the optimality conditions based on a
Moreau decomposition.

- The special Moreau decomposition for the optimality conditions comes from work in infinite dimensional Hilbert space e.g., [9, 10, 23, 8], where the projection is actually differentiable, and typically P is the intersection of a cone and a linear manifold of finite co-dimension (finite # constraints).
- parametrized quadratic problem to solve finite dimensional linear programs [30] applied in our work here below. (In this finite dimensional case differentiability was lost.)
- infinite dimensional applications appear in the theory of *partially finite programs* in [5, 6] Further references in [29, 19, 2].

Semismoothness

- differentiability is lost in finite dimensional; this led to application of semismoothness [24, 26, 25].
- More recently: applications for nearest Euclidean distance matrices and nearest doubly stochastic in [1, 17].
- The optimum $x^*(v)$ is often called the *projection onto the polyhedral set* and is known to be unique. Differentiability properties are nontrivial as discussed in e.g., [16]. A characterization of differentiability in terms of normal cones is given in [13]. Further results and connections to semismoothness is in e.g., [16, 15]. A survey presentation is at [28].

Projection onto a Polyhedral Set

$$
\text{(P)} \qquad
\begin{array}{rll}
x^*(v) := & \operatorname{argmin}_x & \frac{1}{2}\|x - v\|^2 \\
 & \text{s.t.} & Ax = b \in \mathbb{R}^m \\
 & & x \in \mathbb{R}^n_+,
\end{array}
$$

$$
\text{optimal value: } p^*(v) \quad = \quad \frac{1}{2}\|x^*(v) - v\|^2,
$$

Assumptions: A full row rank; feasible set nonempty

Optimality Conditions

Theorem ($F : \mathbb{R}^m \to \mathbb{R}^m$; find root y^*; Newton)

The optimum $x^*(v)$ exists and is unique. Let

$$(*) \quad \boxed{F(y) := A(v + A^T y)_+ - b}, \quad f(y) := \tfrac{1}{2}\|F(y)\|^2$$

Then $F(y) = 0$ has a root y^*, $F(y^*) = 0 \iff y \in \operatorname{argmin} f(y^*)$

$$x^*(v) = (v + A^T y^*)_+, \text{ for any root } F(y^*) = 0.$$

Moreover, strong duality holds and the dual problem is

$$\begin{aligned} p^*(v) &= d^*(v) \\ &:= \max_{z \geq 0, y} \phi(y, z) \quad (= \min_x L(x, y, z)) \\ &:= -\tfrac{1}{2}\left\|z - A^T y\right\|^2 + y^T(Av - b) - z^T v. \end{aligned}$$

AND

At each iteration, we get a provable/calculable lower bound

$$\max_{z \geq 0, y} \phi(y, z) = -\frac{1}{2}\left\|z - A^T y\right\|^2 + y^T(Av - b) - z^T v$$

Proof of Optimality Conditions

Proof.

$L(x,y,z) = \frac{1}{2}\|x-v\|^2 + y^T(b-Ax) - z^T x;$
$\nabla_x L(x,y,z) = x - v - A^T y - z;$
stationarity: $0 = \nabla_x L(x,y,z) \implies x = (v + A^T y) + z$
$\implies L(x,y,z) = -\frac{1}{2}\|z + A^T y\|^2 + y^T(b - Av) - z^T v.$

KKT optimality conditions

$\frac{\partial}{\partial x}L(x,y,z)$	$=$	$x - v - A^T y - z$	$=$	0	(dual feasibility)
$\frac{\partial}{\partial y}L(x,y,z)$	$=$	$Ax - b$	$=$	0	(primal feasibility)
$\frac{\partial}{\partial z}L(x,y,z)$	$\cong$	$x \in (\mathbb{R}^n_+ - z)^+$			(compl. slackness, $z^T x = 0$ or $z \circ x = 0$)

$\square$

Proof continued...

(cont... Solve opt. cond.

$$\begin{bmatrix} x - v - A^T y - z \\ Ax - b \\ z^T x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad x, z \in \mathbb{R}^n_+, y \in \mathbb{R}^m.$$

Moreau Decomposition:

$v + A^T y = x - z \quad = x + (-z), x^T z = 0$

$x = (v + A^T y)_+; \; z = -(v + A^T y)_-$

$$F : \mathbb{R}^m \to \mathbb{R}^m; \quad \boxed{F(y) = A(v + A^T y)_+ - b = 0, \;\; y \in \mathbb{R}^m}$$

□

Apply Newton at current y_c; Newton direction Δy

$$F'(y_c)\Delta y = -F(y_c); \qquad y_p = y_c + \Delta y$$

Compare Interior Point Methods

Block Elimination on Perturbed KKT Conditions

$$\begin{bmatrix} r_d \\ r_p \\ r_c \end{bmatrix} := \begin{bmatrix} x - v - A^T y - z \\ Ax - b \\ Zx - \mu e \end{bmatrix}, \quad x, z \in \mathbb{R}^n_+, y \in \mathbb{R}^m.$$

$$F'_\mu \Delta s = \begin{bmatrix} \underline{\Delta x} - A^T \Delta y - \Delta z \\ A\Delta x - b \\ X\Delta z + Z\Delta x \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_c \end{bmatrix}, \quad x, z \in \mathbb{R}^n_+, y \in \mathbb{R}^m.$$

Normal Equations Reduction to Δy

Currently, normal equations are not considered efficient. But the Newton equation was a percursor and appears to be efficient?

$$F : \mathbb{R}^m \to \mathbb{R}^m; \qquad \boxed{F(y) = A(v + A^T y)_+ - b = 0, \;\; y \in \mathbb{R}^m}$$

$$F'(y_c)\Delta y = -F(y_c); \qquad y_p = y_c + \Delta y$$

Nonlinear Least Squares, Generalized Jacobians

minimize squared residual $f(y) = \frac{1}{2}\|F(y)\|^2$

differentiable case $\{i : (v + A^T y)_i = 0\} = \emptyset$:
$\nabla f(y) = (F'(y))^* F(y)$

Definition ((local) Lipschitz Continuity)

Let $\Omega \subseteq \mathbb{R}^n$. A function $F : \Omega \to \mathbb{R}^n$ is *Lipschitz continuous* on Ω if there exists $K > 0$ such that

$$|F(y) - F(z)| \leq K\|y - z\|, \forall y, z \in \Omega.$$

F is *locally Lipschitz continuous* on Ω if for each $x \in \Omega$ there exists a neighbourhood U of x such that F is Lipschitz continuous on U.

Generalized Jacobian

Rademacher's Theorem [27, 14]

$F : \Omega \to \mathbb{R}^n$ locally Lipschitz on Ω implies that it is Frechét differentiable almost everywhere on Ω.

Definition (Clarke [11] Generalized Jacobian)

Suppose that $F : \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz. Let D_F be the set of points such that F is differentiable. Let $F'(y)$ be the usual Jacobian matrix at $y \in D_F$. The *generalized Jacobian of F at y*, $\partial F(y)$ is

$$\partial F(y) = \operatorname{conv}\left\{ \lim_{\substack{y_i \to y \\ y_i \in D_F}} F'(y_i) \right\}.$$

In addition, $\partial F(y)$ is nonsingular if every $V \in \partial F(y)$ is nonsingular.

Case: Differentiable and $F'(y)$ invertible

Newton Direction; Newton Equation

$$(F'(y))^*(F'(y))\Delta y = -(F'(y))^* F(y) \iff F'(y)\Delta y = -F(y).$$

$$\Delta y = -\left((F'(y))^*(F'(y))\right)^{-1}(F'(y))^* F(y) = -(F'(y))^\dagger F(y)$$

directional derivative: $\Delta y^T \nabla f(y) = \dots$

$$-\left[(F'(y))^* F(y)\right]^T \boxed{\left((F'(y))^*(F'(y))\right)^{-1}} \left[(F'(y))^* F(y)\right]$$
$$< 0$$

We now see that we maintain a descent direction.

Lemma (for handling singularity in $(F'(y))^*(F'(y))$)

LM *direction is always a descent direction.*

Proof.

$(J \cong F'(y))$

$$(J^* J + \lambda I)\Delta y = -J^* F.$$

$$\Delta y = -\left(J^T J + \lambda I\right)^{-1}(J^T F).$$

Therefore, the directional derivative is

$$\begin{aligned}\Delta y^T \nabla f(y) &= -\left(\left(J^T J + \lambda I\right)^{-1}(J^T F)\right)^T (J^T F) \\ &= -(J^T F)^T \left(\left(J^T J + \lambda I\right)^{-1}\right)(J^T F) \\ &< 0.\end{aligned}$$

□

Max. Rank Generalized Jacobian

Cols chosen $\cong$ pos. variables of w

$$Aw_+ = A(\mathcal{P}_{\mathcal{N}} w) = (A\mathcal{P}_{\mathcal{N}})w_+ = \sum_{w_i>0} A(:,i)w_i$$

Index Set of Columns

Note: $v + A^T y \geq 0 \implies F'(\Delta y) = AIA^T \Delta y = AA^T \Delta y$

$$\mathcal{U}(y) := \left\{ u \in \mathbb{R}^n \mid u_i \in \left\{ \begin{array}{cl} 1 & \text{if } (v + A^T y)_i > 0 \\ [0,1] & \text{if } (v + A^T y)_i = 0 \\ 0 & \text{if } (v + A^T y)_i < 0 \end{array} \right. \right\}$$

generalized Jacobian at y; after convex hull

$\partial F(y) = \{A \operatorname{Diag}(u) A^T \mid u \in \mathcal{U}(y)\}$
(max-rank: choose $u_i = 1$ when possible)

Semismooth Newton Method solving $F(y) = 0$

Solve $(V_k + \lambda I) d_{Newton} = -F(y^k)$, with
$V_k \in \partial F(y^k), \lambda > 0, c \in (0,1)$

$y^{k+1} = y^k + d_{Newton};\ \left(\text{or avging } y^{k+1} = (1-c)y^k + +c d_{Newton}\right)$

Max-rank Jacobian

$$\begin{aligned} AMA^T &:= A\mathrm{Diag}\,(u)A^T \\ &= \textstyle\sum_{i\in\mathcal{I}_+} A_{:i}A_{:i}^T + \sum_{i\in\mathcal{I}_0} \alpha_i A_{:i}A_{:i}^T,\ \alpha_i \in [0,1], \forall i \in \mathcal{I}_0 \end{aligned}$$

maximum (resp. minimum) rank for AMA:
$\alpha_i = 1, \forall i \in \mathcal{I}_0$ ($\alpha_i = 0, \forall i \in \mathcal{I}_0$, resp.)

Vertices and Polar Cones

Choosing the optima for the tests; (nondegenerate) vertex

In our tests we can decide on the characteristics of the optimal solution using the properties of (degenerate) vertices.
Recall: x optimal iff $x - v \in \mathcal{F}(x)^+$

Lemma (vertex and polar cone)

$y \in \mathbb{R}^m, x(y) = (v + A^T y)_+ \in \mathcal{F}$. Then:
$x(y)$ vertex $\iff A_{\mathcal{I}_+}$ nonsingular
$\iff$ corresp. gen. Jac. nonsingular.
$x = x(y) \in \mathcal{F} \implies$
$\mathcal{F}(x)^+ = \{w : w = A^T u + z, u \in \mathbb{R}^m, z \in \mathbb{R}^n_+, x^T z = 0\}$

Proof of Lemma

Proof.

wlog $A = [A_{\mathcal{I}_+} \; A_{\mathcal{I}_0}]$ implies active set is $\begin{bmatrix} A_{\mathcal{I}_+} & A_{\mathcal{I}_0} \\ 0 & I \end{bmatrix} x = \begin{pmatrix} b \\ 0 \end{pmatrix}$;
This has unique solution $x(y)$ iff $A_{\mathcal{I}_+}$ is nonsingular.
gradient of objective satisfies

$$x - v = A^T y + \sum_{j \in \mathcal{I}_0} z_j e_j.$$

Optimality conditions yield polar cone at a vertex. □

degeneracy of optimal solutions

Let $x \in \operatorname{bdry} \mathcal{F}$;
x is optimal iff $x - v \in \mathcal{F}(x)^+$, i.e., we can choose v with
$v = x - A^T u + z$, $z \geq 0$, $z^T x = 0$.
and
$x^*(v)$ is differentiable at $v \iff (x^*(v) - v) \in \operatorname{ri}(\mathcal{F} - x^*(v))^+$

Best Approx.; Nonsmooth Algor.

Algorithm 1 Best Approx. of v in P; Exact Newton

Require: $v \in \mathbb{R}^n, y_0 \in \mathbb{R}^m$, $(A \in \mathbb{R}^{m\times n}, \operatorname{rank}(A) = m)$, $\varepsilon > 0$, maxiter

Output. Primal-dual opt: $x_{k+1}, (y_{k+1}, z_{k+1})$

Initialization. $k \leftarrow 0$, $x_0 \leftarrow (v + A^T y_0)_+$, $z_0 \leftarrow (x_0 - (v + A^T y_0))_+$,
$F_0 = Ax_0 - b$, stopcrit $\leftarrow \|F_0\|/(1 + \|b\|)$

while ((stopcrit $> \varepsilon$) & ($k \le$ maxiter)) **do**

  $\lambda = \min(1e^{-3},$ stopcrit$)$

  $\bar{V} = (V_k + \lambda I_m)$

  solve pos. def. $\bar{V} d = -F_k$ for Newton direction d

  updates

  $y_{k+1} \leftarrow y_k + d$

  $x_{k+1} \leftarrow (v + A^T y_{k+1})_+$

  $z_{k+1} \leftarrow (x_{k+1} - (v + A^T y_k))_+$

  $F_{k+1} \leftarrow Ax_{k+1} - b$ (residual)

  stopcrit $\leftarrow \|F_{k+1}\|/(1 + \|b\|)$

  $k \leftarrow k + 1$

end while

Halpern-Lions-Wittmann-Bauschke [3]

Halpern-Lions-Wittmann-Bauschke [3] (HLWB)

Algorithm 2 Extended HLWB algorithm

Require: $v \in \mathbb{R}^n$, $(A \in \mathbb{R}^{m \times n}, \operatorname{rank}(A) = m)$, $\varepsilon > 0$, maxiter $\in \mathcal{N}$.

Output. x_{k+1}

Initialization. $k \leftarrow 0$, $msweeps \leftarrow 0$ $x_0 \leftarrow max(v, 0)$, $y_0 \leftarrow x_0$, $i_0 = 1$
  stopcrit $\leftarrow \|Ay_0 - b\|/(1 + \|b\|)$ $(= \|F_0\|/(1 + \|b\|))$

while ((stopcrit $> \varepsilon$) & ($k \leq$ maxiter)) **do**
  if $1 \leq i(k) \leq m$ **then**
    $y_k = x_k + \frac{b_{i_k} - \langle a_{i_k}, x^k \rangle}{\|a_{i_k}\|^2} a_{i_k}$
  else
    $y_k = \max(0, x_k)$
  end if
  updates
  $\sigma_k = \frac{1}{k+1}$ (change to $\sigma_k = \frac{1}{msweeps+1}$??)
  $x^{k+1} \leftarrow \sigma_k v + (1 - \sigma_k) y^k$
  stopcrit $\leftarrow \|Ay_0 - b\|/(1 + \|b\|)$
  $k \leftarrow k + 1$
  if $k \mod (m+1) == 0$ **then**
    $msweeps = msweeps + 1$
  end if
  $i_k = k(\text{mod } m) + 1$
end while

Numerical Tests varying sizes m, n

Table: Varying $m = 100, 600, 1100, 1600$

Specifications			Time (s)					Rel. Resid		
	n	% density	Exact	Inexact	HLWB	LSQ	QPPAL	Exact	Inexact	HLWB
0	3000	8.1e-01	2.13e-03	1.98e-02	1.89e+01	3.22e+00	8.04e-01	2.55e-16	2.41e-15	2.29e-04
0	3000	8.1e-01	8.35e-02	3.03e-01	1.94e+02	4.28e+00	1.27e+00	5.10e-16	5.10e-18	2.19e-04
00	3000	8.1e-01	7.02e-01	1.29e+00	4.16e+02	6.18e+00	2.53e+00	5.20e-16	8.71e-16	2.08e-04
00	3000	8.1e-01	1.40e+00	3.59e+00	6.57e+02	7.65e+00	5.13e+00	9.84e-18	1.11e-15	2.27e-04

Table: Varying n, $m = 200$

Specifications			Time (s)					Rel. Resids.		
	n	% density	Exact	Inexact	HLWB	LSQ	QPPAL	Exact	Inexact	HLWB
0	3000	8.1e-01	3.12e-03	3.69e-02	4.45e+01	3.50e+00	8.66e-01	8.64e-18	7.39e-17	2.56e-04
0	3500	8.1e-01	3.08e-03	4.05e-02	5.17e+01	4.93e+00	1.00e+00	9.07e-18	1.26e-17	2.78e-04
0	4000	8.1e-01	3.24e-03	3.70e-02	5.82e+01	7.31e+00	1.09e+00	1.46e-16	8.91e-16	2.80e-04
0	4500	8.1e-01	3.99e-03	4.17e-02	6.58e+01	1.01e+01	1.18e+00	1.80e-15	2.05e-16	3.13e-04
0	5000	8.1e-01	3.93e-03	3.42e-02	7.30e+01	1.45e+01	1.26e+00	4.09e-17	1.80e-15	3.16e-04

Numerical Tests varying density

Table: Varying problem density, $m = 300$

	Specifications		Time (s)					Rel. Resids.		
	n	% density	Exact	Inexact	HLWB	LSQ	QPPAL	Exact	Inexact	HLWB
0	1000	1.0e+00	5.65e-03	5.69e-02	1.67e+01	3.02e-01	5.32e-01	7.48e-16	7.27e-16	1.54e-04
0	1000	6.0e+00	4.80e-02	2.52e-01	4.58e+01	3.15e-01	1.22e+00	3.44e-17	1.18e-16	1.51e-04
0	1000	1.1e+01	6.18e-02	2.49e-01	5.41e+01	3.07e-01	2.10e+00	5.65e-17	1.54e-17	1.44e-04
0	1000	1.6e+01	7.79e-02	2.60e-01	5.34e+01	3.03e-01	2.11e+01	6.92e-17	7.98e-17	1.61e-04

Solving (maximization) Linear Programs

primal (maximization) LP in standard form

$$
\text{(PLP)} \qquad p^*_{LP} := \begin{array}{ll} \max & c^T x \\ \text{s.t.} & Ax = b \in \mathbb{R}^m \\ & x \in \mathbb{R}^n_+ . \end{array}
$$

dual LP

$$
\text{(DLP)} \qquad d^*_{LP} := \begin{array}{ll} \min & b^T y \\ \text{s.t.} & A^T y - z = c \in \mathbb{R}^n \\ & z \in \mathbb{R}^n_+ . \end{array} \qquad (2)
$$

Assumptions

A full row rank;

$p^*_{\mathrm{LP}} \in \mathbb{R}$ (so $p^*_{\mathrm{LP}} = d^*_{\mathrm{LP}} \in \mathbb{R}$ and both attained)

Geometric Algorithm

solution can be found from the limit as $R \uparrow \infty$ of the projection of the vector $v_R = Rc \in \mathbb{R}^n$ onto the feasible set.

Lemma ([20, 21, 22, 30])

Let the given LP data be A, b, c with finite optimal value p^*_{LP}. For each $R > 0$ define

$$\begin{array}{rll} x(R) := & \operatorname{argmin}_x & \frac{1}{2}\|x - Rc\|^2 \\ & \text{s.t.} & Ax = b \in \mathbb{R}^m \\ & & x \in \mathbb{R}^n_+. \end{array}$$

Then x^* is the minimum norm solution *of (PLP) if, and only if, there exists $\bar{R} > 0$ such that*

$$R \geq \bar{R} \implies x^* \in \operatorname{argmin}\left\{\frac{1}{2}\|x - Rc\|^2 \,:\, Ax = b,\, x \in \mathbb{R}^n_+\right\}.$$

□

Avoid numerical/roundoff from large numbers

Corollary (scaling $\frac{1}{R}b$)

$A, b, c, R, x(R)$ as in Lemma. Then

$$\frac{1}{R}x(R) = w(R) := \begin{array}{ll} \operatorname{argmin}_w & \frac{1}{2}\|w - c\|^2 \\ s.t. & Aw = \frac{1}{R}b \in \mathbb{R}^m \\ & w \in \mathbb{R}^n_+. \end{array}$$

Proof.

From

$$\|x - Rc\|^2 = R^2 \left\| \frac{1}{R}x - c \right\|^2 = R^2\|w - c\|^2, \; x = Rw,$$

we substitute for x and obtain $A(Rw) = b \iff Aw = \frac{1}{R}b$. The result follows from the observation that argmin does not change after discarding the constant R^2. □

Conclusion

- efficient, robust algorithm for projection of a point onto a polyhedral set.
- One of may applications is to solving linear programs - a type of exterior path following algorithm.

References I

S. Al-Homidan and H. Wolkowicz.
Approximate and exact completion problems for Euclidean distance matrices using semidefinite programming.
Linear Algebra Appl., 406:109–141, 2005.

L.E. Andersson and T. Elfving.
Best constrained approximation in Hilbert space and interpolation by cubic splines subject to obstacles.
SIAM J. Sci. Comput., 16(5):1209–1232, 1995.

H.H. Bauschke.
The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space.
Journal of Mathematical Analysis and Applications, 202:150–159, 1996.

D. Bertsimas and J. Tsitsiklis.
Introduction to Linear Optimization.
Athena Scientific, Belmont, MA, 1997.

J.M. Borwein and A.S. Lewis.
Partially finite convex programming, part I, duality theory.
Math. Program., 57:15–48, 1992.

J.M. Borwein and A.S. Lewis.
Partially finite convex programming, part II, explicit lattice models.
Math. Program., 57:49–84, 1992.

J.M. Borwein and H. Wolkowicz.
Regularizing the abstract convex program.
J. Math. Anal. Appl., 83(2):495–530, 1981.

J.M. Borwein and H. Wolkowicz.
A simple constraint qualification in infinite-dimensional programming.
Math. Programming, 35(1):83–96, 1986.

C.K. Chui, F. Deutsch, and J.D. Ward.
Constrained best approximation in Hilbert space.
Constr. Approx., 6(1):35–64, 1990.

C.K. Chui, F. Deutsch, and J.D. Ward.
Constrained best approximation in Hilbert space. II.
J. Approx. Theory, 71(2):213–238, 1992.

F.H. Clarke.
Optimization and Nonsmooth Analysis.
Canadian Math. Soc. Series of Monographs and Advanced Texts. John Wiley & Sons, 1983.

D. Drusvyatskiy and H. Wolkowicz.
The many faces of degeneracy in conic optimization.
Foundations and Trends® in Optimization, 3(2):77–170, 2017.

F. Facchinei and J.-S. Pang.
Finite-dimensional variational inequalities and complementarity problems, volume 1.
Springer, 2003.

H. Federer.
Geometric measure theory.
Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

References V

M. Goh and F. Meng.
On the semismoothness of projection mappings and maximum eigenvalue functions.
J. Global Optim., 35(4):653–673, 2006.

J.-B. Hiriart-Urruty.
Unsolved Problems: At What Points is the Projection Mapping Differentiable?
Amer. Math. Monthly, 89(7):456–458, 1982.

H. Hu, H. Im, X. Li, and H. Wolkowicz.
A semismooth Newton-type method for the nearest doubly stochastic matrix problem.
Math. Oper. Res., 2023 to appear.
arxiv.org/abs/2107.09631, 35 pages.

- H. Im and H. Wolkowicz.
  A strengthened Barvinok-Pataki bound on SDP rank.
  Oper. Res. Lett., 49(6):837–841, 2021.
- C. Li and X.Q. Jin.
  Nonlinearly constrained best approximation in Hilbert spaces: the strong chip and the basic constraint qualification.
  SIAM J. Optim., 13(1):228–239, 2002.
- O.L. Mangasarian.
  Iterative solution of linear programs.
  SIAM J. Numer. Anal., 18(4):606–614, 1981.

O.L. Mangasarian.
Normal solutions of linear programs.
Number 22, pages 206–216. 1984.
Mathematical programming at Oberwolfach, II
(Oberwolfach, 1983).

O.L. Mangasarian.
A Newton method for linear programming.
J. Optim. Theory Appl., 121(1):1–18, 2004.

C.A. Micchelli, P.W. Smith, J. Swetits, and J.D. Ward.
Constrained l_p approximation.
Journal of Constructive Approximation, 1:93–102, 1985.

R. Mifflin.
Semismooth and semi-convex functions in constrained optimization.
SIAM J. Cont. Optim., 15:959–972, 1977.

References VIII

H. Qi and D. Sun.
A quadratically convergent Newton method for computing the nearest correlation matrix.
SIAM J. Matrix Anal. Appl., 28(2):360–385, 2006.

L. Qi and J. Sun.
A nonsmooth version of Newton's method.
Mathematical programming, 58(1-3):353–367, 1993.

H. Rademacher.
Uber partielle und totale differenzierbarkeit i.
Math. Ann., 89:340–359, 1919.

References IX

E. Sarabi.
A characterization of continuous differentiability of proximal mappings of composite functions.
URL: https://www.math.uwaterloo.ca/~hwolkowi/F22MOMworkshop.d/FslidesSarabi.pdf, 10 2022.
24th Midwest Optimization Meeting, MOM24.

I. Singer.
Best approximation in normed linear spaces by elements of linear subspaces.
Die Grundlehren der mathematischen Wissenschaften, Band 171. Publishing House of the Academy of the Socialist Republic of Romania, Bucharest; Springer-Verlag, New York-Berlin, 1970.
Translated from the Romanian by Radu Georgescu.

P.W. Smith and H. Wolkowicz.
A nonlinear equation for linear programming.
Math. Programming, 34(2):235–238, 1986.

S. Sremac.
Error bounds and singularity degree in semidefinite programming.
PhD thesis, University of Waterloo, 2019.

J.F. Sturm.
Error bounds for linear matrix inequalities.
SIAM J. Optim., 10(4):1228–1248 (electronic), 2000.

Thanks for your attention!

Regularized Nonsmooth Newton Algorithms for Best Approximation with Applications

Combinatorics & Optimization, University of Waterloo

Prof. Henry Wolkowicz
hwolkowicz@uwaterloo.ca

Tues. Mar. 28, 10:00-11:20 EST, 2023

joint work with: Yair Censor (Univ. of Haifa);
Walaa Moursi and Tyler Weames (Univ. of Waterloo)