Starting from the fact that the 1974 version of the NNLS algorithm of Lawson and Hanson does indeed constitute by itself a new constructive proof of Farkas’ lemma (in primal and dual form), we apply it to solve the restricted feasibility sub-problems that occur in two primal-dual active-set linear programming algorithms, thus establishing the relationship with a recently proposed (dated in October, 2002) primal-dual active-set scheme. As a result, the proposed algorithms are impervious to degeneracy, both in Phase I and in Phase II, and finite in exact arithmetic without having to resort to any anti-cycling pivoting rule, and hence they are by themselves constructive proofs of the strong duality theorem. Finally, we describe the sparse implementation of both algorithms in terms of certain sparse Cholesky factorization.

Keywords: Linear programming; Non-negative least squares; Primal-dual active-set method; Farkas’ lemma; Degeneracy
1 CONSTRUCTIVE PROOF OF FARKAS’ LEMMA IMPERVIOUS TO DEGENERACY

Let us consider the usual unsymmetric primal-dual pair of linear programs using a non-standard notation (we have deliberately exchanged the usual roles of $b$ and $c$, $x$ and $y$, $n$ and $m$, and $(P)$ and $(D)$, as in e.g. [16, §2]):

$$(P) \quad \text{min } \ell(x) \equiv c^T x, \ x \in \mathbb{R}^n$$

subject to $A^T x \geq b$.

$$(D) \quad \text{max } \mathcal{L}(y) \equiv b^T y, \ y \in \mathbb{R}^m$$

subject to $Ay = c, \ y \geq 0$.

where $A \in \mathbb{R}^{n \times m}$ with $m \geq n$ and rank$(A) = n$. We denote with $\mathcal{F}$ and $\mathcal{G}$ the feasible region of $(P)$ and $(D)$, respectively.

In his elementary proof of Farkas’ lemma,

or $\exists y \in \mathbb{R}^m : Ay = c, \ y \geq 0$,  
or else $\exists d \in \mathbb{R}^n : c^T d < 0, \ A^T d \geq 0$,

Dax (cf. [8] and also [7]) links the dual feasibility problem

$$\text{find } y \in \mathbb{R}^m : Ay = c, \ y \geq 0$$

with the NNLS convex problem

$$\min \{ \| Ay - c \|_2 : y \geq 0 \}.$$  

He gave a particular method to deal with this problem, in which he had to manage the possibility of degeneracy. Indeed, as was pointed out in [3], “this possibility seems to be an intrinsic part of the problem”. Our proposal is to apply the NNLS algorithm of Lawson and Hanson [13, §23.3] directly to this problem, because in exact arithmetic the algorithm is finite and so we do not have to resort to anti-
cycling rules to prove Farkas’ lemma. Note that problem (2) usually appears in the literature in strictly convex form, but a careful view of the proof of the finiteness in [13, §23.3] does not rule out the convex case. We obtain a dual feasible point (in particular, a point $y \doteq [y_B; O]$ such that $A_By_B = c$, $y_B > O$ and $A_B$ has full column rank, so $y$ is a basic feasible solution of (1)), or else we obtain a descent direction $d$ of $\mathcal{F}$. Note that the number of columns in $A_B$ is less than or equal to $n$.

Gale’s theorem is equivalent to the dual form of Farkas’ lemma (e.g., [20]),

$$\text{or } \exists x \in \mathbb{R}^n : A^T x \geq b, \quad \text{or else } \exists \delta \in \mathbb{R}^m : b^T \delta > 0, A\delta = O, \delta \geq O.$$ 

Our favourite constructive proof consists on linking the primal feasibility problem

$$\text{find } x \in \mathbb{R}^n : A^T x \geq b \quad (3)$$

with the least distance problem

$$\min \{ \|x\|_2 : A^T x \geq b \}. \quad (4)$$

We then proceed with Cline’s method as in [13, §23.4], solving

$$\min \left\{ \left\| \begin{bmatrix} b^T \\ \delta - \begin{bmatrix} 1 \\ O \end{bmatrix} \end{bmatrix} \right\|_2 : \delta \geq O \right\} \quad (5)$$

via NNLS instead of (4), so we also escape from degeneracy. We obtain an $x \in \mathcal{F}$ (i.e., a solution of (3)), or else we obtain an ascent direction $\delta$ of $\mathcal{G}$. 

3
2 THE PROPOSED PRIMAL-DUAL ACTIVE-SET ALGORITHM

The well-known primal-dual active-set method of [5] can be applied to our primal-dual pair if we have a primal feasible point \( x \) available. The solution procedure proceeds by solving a sequence of *restricted dual feasibility subproblems*, namely

\[
\text{find } y \in \mathbb{R}^m : A_A y_A = c, y_A \geq O, \tag{6}
\]

where \( A_A \) is constructed with the columns of \( A \) corresponding to active primal constraints in \( x \). Note that \( A_A \) can have more than \( n \) columns when \( x \) is a degenerate primal vertex. We do not reproduce the details here: the interested reader can consult [14, pp. 263–265].

The original proposal was to solve the restricted feasibility subproblem (6) by constructing a Phase I problem

\[
\min \{ e^T t : A_A y_A + t = c, y_A \geq O, t \geq O \}
\]

and solving it by the simplex method with \( t \in \mathbb{R}^n \) as the initial artificial set of basic variables. The main point is that we have to resort to anti-cycling rules as the inefficient Bland’s one to ensure the convergence of the overall procedure. Several authors (e.g., [11], [12], [10]) have proposed to solve the Phase I subproblem using a criss-cross variant, whose efficiency is an open question [11, §7].

Once again, Dax [7] was the first one to link the restricted feasibility subproblem (6) with the NNLS convex subproblem

\[
\min \{ \| A_A y_A - c \|_2 : y_A \geq O \}. \tag{7}
\]
He gave a particular method to deal with this subproblem, in which he had to man-
age the possibility of degeneracy; this method is known as Dax’s steepest-descent
primal-dual active-set algorithm since [9].

Our proposal is to resort to our favourite proof of Farkas’ lemma and then to apply
the NNLS algorithm directly to subproblem (7), obtaining an optimal $y_A \geq 0$
or else a descent direction $d$ such that $e^T d < 0$ and $A^T A d \geq 0$; such a direction
maintains primal feasibility and strictly decreases the primal objective function $\ell$.
This is the algorithm proposed in [1], but Dantzig et al. [6] apply NNLS to (2) nine
years ago and claim that several NNLS variants “can be applied to solve the Phase
II linear program efficiently by means of a sequence of improving approximations
to the optimal value of the objective function”. Note that in [6] no explicit Farkas’
connection was given.

But what about the computation of the initial primal feasible $x$? Dax [7] proposes to
apply a similar algorithm with a modified piecewise linear objective function with
the sum of primal infeasibilities, and [1] refer to [17, §5] where the classical Beale’s
device (also in [5]) modifying the dual feasible region is described. Note that both
proposed Phase I’s modify the original problem in a somewhat anti-natural sense.

Our proposal is to apply our constructive proof of the dual form of Farkas’ lemma
to find the initial $x \in F$. As a result, we can detect if (D) is unbounded above
during Phase I with the dual form of Farkas’ lemma. If not, we solve a sequence of
restricted dual feasibility subproblems with the primal form of Farkas’ lemma, us-
using the 1974 version of NNLS all the time to do not have to worry about degeneracy
in exact arithmetic.

This proposal is by itself a constructive proof of the strong duality theorem (e.g.,
[20]), also known as the existence-duality theorem. Indeed, it is related to the im-
We think that the first proposed criss-cross-based implicit proof of Farkas’ lemma was done in [11] (using Bland’s least index rule), although they made it explicit later in [12]. A new criss-cross-based proof of Farkas’ lemma (using two different anti-cycling rules) was given recently in [10], but their proof of the strong duality theorem does not rely upon the hungarian primal-dual connection.

3 THE PROPOSED DUAL-PRIMAL ACTIVE-SET ALGORITHM

But we can also see the problem from the dual perspective (and hence obtain an alternative proof of the strong duality theorem), namely to apply our constructive proof of the primal form of Farkas’ lemma to find an initial \( y \in G \). As a result, we can detect if \( (P) \) is unbounded below during Phase I with the primal form of Farkas’ lemma. If not, we solve a sequence of restricted primal feasibility subproblems using the dual form of Farkas’ lemma.

To the best of our knowledge, the first author that proposed to solve a sequence of restricted primal feasibility subproblems was Murty [14, pp. 261–262], and so we gave this method the name Murty’s steepest-ascent dual-primal active-set algorithm. Although he pointed out in his book that it was described in [5], they indeed only included in 1956 the description of the above section for a linear program in standard form. The solution procedure has to compute first a \( y \in G \) and then proceeds by solving a sequence of restricted primal feasibility subproblems, namely

\[
\text{find } x \in \mathbb{R}^n : A_A^T x = b_A, A_N^T x \geq b_N,
\]

where \( A_A \) is constructed with the columns of \( A \) corresponding to positive compo-
nents of $y$ (and $A_N$ with those to zero components). Note that $A_N$ can have more than $m - n$ columns when $y$ is a degenerate dual vertex.

The original proposal was to solve this restricted primal feasibility subproblem by a Phase I for the primal simplex method. But since the dual of (8) is

$$\text{find } \delta \in \mathbb{R}^m : b^T \delta > 0, A\delta = O, \delta_N \geq O,$$

(9)

our proposal is to link it with

$$\min \left\{ \| \begin{bmatrix} b^T \\ A \end{bmatrix} \delta - \begin{bmatrix} 1 \\ O \end{bmatrix} \|_2 : \delta_N \geq O \right\},$$

(10)

obtaining an optimal $x$ or else an ascent direction $\delta$ such that it maintains dual feasibility and strictly increases the dual objective function $L$. Note that we cannot solve this problem with the 1974 version of the NNLS algorithm, because not all the components of $\delta \in \mathbb{R}^m$ are restricted to be non-negative. Nevertheless, Dantzig et al. [6, §3.1] describe the minor changes necessary to allow NNLS to deal with this possibility, and they do not compromise the imperviousity to degeneracy.

4 SPARSE IMPLEMENTATION: THE CHOLESKY CONNECTION

The necessity to obtain least-squares solutions occurring in the intermediate steps of the NNLS algorithm lead us to work with the QR factorization of $A_k$, where $A_k$ is a full column rank matrix with a subset of columns of the fixed matrix $A$. This is the implementation chosen in [7], in [6] and in [1] when they are faced with a dense problem.
When the problem is sparse, iterative relaxation techniques are recommended in [7, p. 321] to solve the restricted dual feasibility subproblem, but in this way we do not take advantage of the existing relationship between $A_k$ and $A_{k+1}$, where $A_{k+1}$ is the resulting full column rank matrix after adding (or deleting) columns of $A$ to (or from) $A_k$. Both in [6] and [1], a sparse NNLS implementation using QR update techniques is asked for.

A suitable sparse approach is to adapt the methodology of Björck [2] and Oreborn [15] to be able to apply the sparse NNLS algorithm with a “short-and-fat” matrix $A$. They proposed an active set algorithm for the sparse least squares problem

$$\minimize \quad \frac{1}{2} \cdot y^T C y + d^T y, \quad y \in \mathbb{R}^m \quad \text{subject to} \quad l \leq y \leq u$$

with $C > 0$. In our problem

$$C = A^T A \quad \land \quad d = -A^T c \quad \land \quad \forall i \in 1:m, \quad l_i = 0 \land u_i = +\infty$$

but $C \geq 0$, so to maintain a sparse QR factorization of $A_k$ we have had to adapt the proposed static technique as in [4], but without forming $C$. Details can be found in [18].

We do not maintain the orthonormal factor in the sparse QR factorization of $A_k$; instead, we take advantage of the fact that its triangular factor is the Cholesky factor of $A_k^T A_k$, and we use corrected seminormal equations to solve the least squares subproblems with the desired accuracy. As was claimed by Saunders [19], major Cholesky would feel proud, but we think that Gyula Farkas would also feel proud!

**Acknowledgements**

The authors thank Michael Saunders at Stanford who, with respect to the techni-
cal report prior to [6], commented us: “Their method is equivalent to the known (and very simple) NNLS algorithm for non-negative least squares. Their real contribution was to apply it to linear programs to find a feasible solution”. We also thank Åke Björck at Linköping and Achiya Dax at Hydrological Service for providing us the references [2] and [7] that we were not aware of, and John Gilbert at Xerox for confirming us that the sparse orthogonal factorization in MATLAB 5 is row-oriented, Givens-based and providing us technical details about the interface.

References


Gyula Farkas (natural scientist). From Wikipedia, the free encyclopedia. Gyula Farkas. Born. (1847-03-28)March 28, 1847 Sârșod, Fejér County. Farkas Gyula, or Julius Farkas (March 28, 1847 – December 27, 1930) was a Hungarian mathematician and physicist. He attended the gymnasium at Győr (Raab), and studied law and physics at Budapest. After teaching in a secondary school at Székesfehérvár (Stuhlweissenburg), Farkas became in succession principal of the normal school at Pápa, privat-docent (1881) of mathematics at the University of Budapest, and professor of physics (1888) at Franz Joseph University of Kolozsvár (Cluj-Napoca, Klausenburg). He worked here up to 1915, when he retired and moved to Budapest.