Notes on the Linear Complementarity Problem

Jad Nohra

1 Notes

Note 1. A short history of the LCP is presented in the preface of [1]. It all started between 1940 and 1960, in relation to operations research, alongside Linear Programming (LP), where Dantzig proposed the Simplex algorithm in 1947. Linear and quadratic programs originally played a major role in its early development, where the LCP was conceived as a unifying formulation for both, and for the bimatrix game problem. In these problems the matrix M in the associated LCP problem has special properties.

Note 2. [1, p1]. In the restrictions of LCP

$$z \geq 0 \tag{1}$$

$$w = q + Mz \geq 0 \tag{2}$$

$$z^T(q + Mz) = 0\ , \text{ equiv. } \ z^T w = 0\ , \text{ equiv. } \ z^T(q + Mz) \leq 0 \tag{3}$$

Note that the $\leq$ formulation in (3) is explained in [1, 1.7.2]. Is it so that the condition $z \geq 0$ in the basic formulation of LCP is simply due to the most common LCP problems? It might be so that it helps restrict the possible solutions and that, given a z that contains negative elements, one can usually find a solution that is non-negative. Additionally, it is certain that the condition supports determinacy since both of z^T and $(q + Mz)$ are always nonnegative in (3).

Note 3. At least for game physics solvers, it feels that condition (1) above can always be done because of the symmetry of applied impulses. In terms of interpretation, (2) is violated when one of its elements is negative, this is when a physical constraint is violated. Hence, when (2) is violated, the problem is unsolved. Note that subtlety that only one situation violates (3), the situation where both an impulse is applied, and the constraint evaluation is positive (and not zero). The subtlety being that if an impulse is applied and the constraint evaluation is exactly zero, all is fine. In terms of interpretation, imagine a contact constraint: No impulse should be applied when there is no contact, when penetration occurs there should be an impulse (otherwise (2) will be violated), and even if the contact is at the limit (perfect contact with zero penetration), it makes sense to apply an impulse to keep it so. In general, there surely are ways to find algorithms for all possible variations of the conditions knowing an algorithm for the standard one. What is essential for the problem to be called LCP is the existence of the three constraints, signs and inequality strictness is not an issue.

Note 4. [1, p3]. Two special classes of LCP(q, M) occur when q $\geq$ 0 and even more specially q = 0. For the first, z = 0 is always a trivial solution. For the second, called homogeneous, we additionally have that if z is a solution, z is too, given ≥ 0. Just like in the case of matrix inversion, the existence of nonzero solution to the homogenous problem is of great theoretical and algorithmic importance.

Note 5. [1.p1]. In R1, condition (3) becomes $zq + Mz^2$, which seems to hint to a QP problem.

Note 6. The chapter [1.Chap2] is a full review chapter which includes all the prerequisites of full LCP understanding. This includes material from the topics: Real Analysis, Matrix Analysis, Pivotal Algebra, Matrix Factorization, Iterative Methods for Equations, Convex Polyhedra, Linear Inequalities, Quadratic Programming Theory, Degree and Dimension.

Note 7. [1, p45] . Since the Cauchy-Schwarz inequality becomes an equality iff. the vectors are linearly dependent, it could be used as a definition for linear independence.

Note 8. [1, p45] . In the elliptic norm formula, the matrix must be symmetric positive definite (SPD). We have to work on a classification and understanding of such types of matrices on a level higher than their definitions, if possible.

Note 9. [1, p48] . Convexity is a generalization of multiplicative betweenness in one dimension.

Note 10. [1, p38] . In the context of finite two person games, zero-sum games are called matrix games and solved by the simplex method, while non-zero-sum games are called bimatrix games; these games are related to the Nash equilibrium point, which itself is related to Brouwer's fixed point theorem. While Brouwer's proof of existence was nonconstructive, Howson (1963) and Lemke (1964) provided an efficient constructive procedure for obtaining a Nash equilibrium point of a bimatrix game. While not the first such procedure (Vorobev, Kuhn, Mangasarian), it was the simplest and most elegant.

Note 11. [1, p39] . Unsurprisingly, finding the convex hull of a set of points is a (special) LCP problem.

Note 12. [1, p40] . On complementary cones: The interpretation of the LCP in terms of complementary cones can be traced to Samelson, Thrall and Wesler (1958), a paper that has great importance for other reasons as well. . . . . The topic was significantly enlarged by Murty (1972).

Note 13. [1, p41] . Finally a mention of the MLCP, especially in the context of a singular matrix A: The mixed LCP (1.5.1) is a special case of a complementarity problem over the cone $R^n \times R^{m+}$. In the case where the matrix A is singular, the conversion of the problem into a standard LCP is not an entirely straightforward matter.

Note 14. [1, p58] . The relation between quadratic functions and convexity can be established through using the positive definiteness of the Hessian.

Note 15. [1, p64] . There is an important connection between the largest and smallest eigenvalues of a matrix A and the quadratic form $x^T Ax$.

Note 16. [1, p64] . There is a tight relationship between the spectral radius of a complex square matrix and any matrix norm: (a)$\rho(A) \leq \|A\|$ for any matrix norm; (b) for every $\epsilon > 0$), there exists at least one matrix norm (induced by some vector norm) such that $\|A\| \leq \rho(A) + \epsilon$.

Note 17. [1, p65] . Convergent matrices are important for the analysis of iterative methods: complex square matrix A is said to be convergent if $lim_{k \to \infty} A^k = 0$; i.e., if all the entries of A^k approach zero as $k \to \infty$. Convergent matrices play an important role in the convergence analysis of iterative methods, see Section 2.5. These matrices can be characterized in terms of their spectral radii. And finally, we have: A complex square matrix A is convergent if and only if $\rho(A) < 1$. If A is convergent, then $(I - A)$ is nonsingular, and $(I - A)^{-1} = \sum_{k \to 0}^{\infty} = A^k$.

Note 18. [1, p69] . The tableau is a version of our additive axis.

Note 19. [1, p70] . It seems that a good reduction to thinking about pivoting is by relating it to Real Algebra, generalized to Matrix Algebra (sandwiching method). Concepts that did not need a name now do to the multiplicity of determination, per example basic and non-basic variables. In that sense, a simple pivot is roughly the multiply determined version moving of a variable from one side to the other in a 'linear' equation in singly determined setting (R^1). The singularity of zero in R^1 is mirrored by the condition $m_{rs} \neq 0$.

Note 20. [1, p75] . Similarly, the Schur complement is only needed in the multiply determined setting, separating a submatrix, in a very rough way, reminding of a remainder; a bit more precisely, the nonsingularity of a square matrix is related to an (always) nonsingular principal submatrix, by the nonsingulatiry of the Schur complement of the latter Indeed, confirming the remainder analogy, we read: Equation (17) $<(M/M_{11}) = ((M/A_{11})/(M_{11}/A_{11}))>$ is called the quotient formula for the Schur complement. One feature to notice is how it formally resembles the algebraic rule used for simplifying a complex fraction <A doubly

determined problem $>$. The Schur complement can be used to determine whether a symmetric matrix is positive definite, through the signs of leading principle minors, practically subject to round-off problems.

Note 21. [1, p79] . On lexicographic ordering we read: An unfortunate fact of life is that the usual linear ordering of the real number system does not hold in integer dimensions. For instance, we can say that (1,0) $\geq$ (0,0) and (0,1) $\geq$ (0,0) but there is no such relationship between (1,0) and (0,1). The notion of lexicographic ordering of vectors remedies this problem in a particular way. We feel that this ordering is a generated concept, and that the best way to internalize is by its use. Despite that, we can still establish the same rough analogy as above: a multiply determined extension of ordering in R1, where each vector component becomes analogous to a digit in a 'number' of fixed digit count. Finally, we read [p.132]: We do know that the concept was used by Dantzig, Orden and Wolfe (1955) as a way to avoid cycling in the simplex method $<\ldots>$ There is a close connection between "lexicography" and "epsilon perturbation" of the constant (right-hand side) vector.

Note 22. [1, p88] . We already know that iterative approximation is often the extraction of a (easily) computable dynamic equilibrium algorithm from a static (mathematical) equilibrium. Iterative solution of a system of linear equations by matrix splitting is a perfect example. Starting from the static equilibrium Ax=b. we formally transform the expression (with the goal in mind) into the equivalent: Ax=b, (B+C)x=b, Bx=b-Cx, x = B^{-1}(b-Cx), and finally extract dynamic equilibrium with $\mathrm{x}^{i+1} = \mathrm{B}^{-1}(\mathrm{b}\text{-}\mathrm{Cx}^{i})$. Of course, this is but a heuristic and the theory about why, how, and when (not always) this works, is the topic of [1-2.2.5]. Indeed, for the nonsingular case, matrix splitting converges iff. $\rho(B^{-1}C) < 1$ [p98]. Additional restrictions like symmetry and positive definiteness enlarge the domain of convergence. The singular case has a different convergence theorem [p.90].

Note 23. [1, p45] . Matrix splitting includes both the Jacobi and the Gauss-Seidel methods, the difference being the choice of the splitting. Successive overrelaxation (SOR) is a generalization of Gauss-Seidel, whose convergence is given by the Ostrowski-Reich theorem [p.90].

Note 24. [1, p91] . Much of what is needed from real analysis is to study the convergence of iterative methods for nonlinear equations of the form f(x)=0 in R^n; namely the Newton method, the damped Newton descent method (which is relevant for solving LCP problems [p.93]), and the contraction principle for computing fixed points. Specifically, Lipschitz continuity guarantees a quadratic rate of convergence for the Newton method, noting that the convergence is only guaranteed locally, meaning that the initial estimate must be 'sufficiently close' to the solution. Eliminating this restriction is the goal of the damped version. Again by a similar same equilibrium tactic as above, we have that the Newton method uses: $x^{i+1} = x^i - \nabla f(xi)^{-1} f(xi)$. This is nicely understood in terms of approximation as a general method. In this case, we reduce the computation of zeros to the most computationally basic, as we know 'linear', operations. In this case, starting with a full power series expansion of a function, with a specific starting point chosen for approximating iteration, we have $f(x) = f(x^i) + \nabla f(x^i)(x - x^i) + (1/2) \ldots$ and 'ignoring' all higher-than-linear terms, we are left with $f(x) = f(x^i) + \nabla f(x^i) - (x - x^i)$, hoping that the iteration will recover their contribution. Contrast this to finding the zeros of a third degree polynomial algebraically.

Note 25. [1, p91] . Indeed, our general but fuzzy idea of the equilibrium, that is, of a dynamical system, that is, of an 'iterative' system, is not far from the truth. The more precise concept, formally unifying both tactics above is that of fixed points. Note that the change of form from $f(x) = 0$ to $g(x) = x$ is simply the transformation into a form that enables iteration and maybe equilibrium. We read: This is a computational scheme for finding the (unique) fixed point of a function of a special type. More specifically, let $g : R^n \to R^n$ be a mapping whose fixed point is being sought. One may generate a sequence $\{x^v\}$ by choosing an arbitrary starting point x0 and by letting $x^{v+1} = g(x^v)$, v=01,2,... This is known as a fixed point iteration. $>$ ... $<$ $g(x) = B^{-1}b - B^{-1}Cx$. Similarly $>$ ... $<$ $g(x) = x - \nabla f(x)^{-1} f(x)$. Rigorously speaking, the latter mapping g is defined only at those points x at which the Jacobian matrix $\nabla f(x)$ is nonsingular.

Note 26. A more precise explanation of the philosophy of iterative approximation may be found in [13, p283]

Note 27. [1, p92] . A contraction mapping (always uniformly continuous), is Lipschitz continuous with modulus less than 1. By the related contraction principle, we find the domain in which sequences converge to fixed points. (Note that this relates us to ZAI, p199). It is here where the spectral radius of the Jacobian matrix relates to the convergence of Newton's method, with $\rho(\nabla g(x))$ being the local Lipschitz modulus (Lipschitz constant).

Note 28. [1, p96] . The author points out the fascinating aspect of Convex Polyhedra that they can be described both externally, as intersections of halfspaces, and internally, as intersection of convex hulls (containing a set of points).

Note 29. [1, p98] . These two theorem from the Convex Polyhedra section are interesting and unexpected at first sight: If the system Ax=b has a nonnegative solution, then the system has a nonnegative basic ¡using only linearly independent vectors¿ solution. The intimate connection between nonnegative basic solutions and extreme points is summarized by the following result. Let x' in X = {x : Ax=b, x≥0}. Then x' is an extreme point of X if and only if x' is a basic solution of Ax=b. In short, there is an intimate solution between nonnegative solutions, basic solutions, and extremes of the convex nonnegative solution sets. All this in turn, is related to the relation between matrices, cones, and extreme rays of cones, as defined in [1.16].

Note 30. [1, p99] . A simple example of a relative interior and that [1.99] In general, a convex set $\in R^n$ need not have a nonempty (topological) interior in R^n; nevertheless, it is a known fact that a relative interior of a nonempty convex set must be nonempty, is the following: A segment embedded in R^3 is convex, has no open neighborhood of any of its points in R^3, but does relatively to its affine hull, a translation of R^1.

Note 31. [1, p99] . The definition used for a cone [p16] allows for an elegant condition for a cone to be convex, based simply on closure under addition: Let C be a cone. Then C is convex if and only if C = C+C = {z: z = x+y, x,y ∈ C }. As a result to construct a conical (and convex) hull of a set of points (vectors), it suffices to construct the cone of all rays through the points, using nonnegative multipliers (formally taking $Pos(A)$ of a set A), and then closing these cones under addition making them convex. At this stage, noting that all linear subspaces are in fact cones, it is very important to think of cones as very important algebraic and geometric creatures, closures of vectors under spans, possible restricted to only positive scalars. Indeed, closure is used to explain cones and convex cones in [7, p230]: If a set if closed under multiplication by non-negative scalars, it is called a cone. This is in analogy with the familiar cones of elementary geometry with vertex at the origin which contain with any point not at the vertex all points on the same half-line from the vertex through the point. If the cone is also closed under addition, it is called a convex cone. It is easily seen that a convex cone is a convex set.
Good and complementary expositions of cones in this context are given in [7, p229-237], [6, p519-531], [9, p42-69].

Note 32. In [9, p42-69], Gale provides dependable pictures for most of the relevant convex cone theory. We retain the following observations.

1. Although it is not the case, the duality of the problem 'nonnegative solutions of linear equations' should be explained this way: For every vector (coordinates) in $Pos(A)$ there is one or many vector (coordinates, taking the matrix theoretic perspective) that are mapped to it, and vice versa, there are no vectors (coordinates) that are mapped by A to vectors outside $Pos(A)$. It

is therefore impossible to have any solutions if b is not in $Pos(A)$. 'Not in $Pos(A)$' is a problem of separability, and in vector spaces, this can be done (separability). It turns out, and this requires some effort of geometrical intuition, or, more formal skills, that the problem of finding a separating plane between b and $Pos(A)$, is a rather elegant problem, one of linear inequalities, called the 'dual problem'. The geometrical intuition is the existence of a hyperplane that splits the space in two, such that b is on one side, and all of a_i are on the other. This mean a hyperspace whose normal, chosen to be on the side of a_i, has a negative inner product with b (geometrically an obtuse angle), and a nonnegative product with each of a_i. Here we note the use of the transpose (we know this is related to the inverse, and the dual space) or A's column vectors, to be able to talk about the inner product, in contrast to taking the vectors themselves, to be able to talk about $Pos(A)$.

2. The exact point at which coordinates enter (and coordinate freedom is lost) into the theory is at p52. There, the total ordering of real numbers is used to single out positive numbers, which are used to create the cone $Pos(x)$ for some vector x, explaining that cones are a generalization of subspaces. What is not mentioned is that the subspace spanned by a vector (or multiple) is invariant under the linear combination scalars when taken from vector coordinates, while the cone is. Per example, suppose $ax + by$ is the span of vectors x and y, with a and b forming the coordinates of a vector. As long as the coordinates are unconstrained, the choice of basis for that vector is irrelevant, and the subspace is the same. If however, there are constraints on the scalars, then, coordinates enter the game, and the choice of basis affects the result (cone per example).
   This becomes clear with such subtle examples as: *The set of all nonnegative vectors P of R^m is a convex cone.* , where the subtlety is to notice the explicit R^m.
3. All the warnings of the previous points taken into account, it is clear how cones are the highest level mental creature to talk about systems of linear inequalities (and their duals). Per example we read:

   3. For any vector b in R^m the set of vectors of the form $\lambda b, \lambda \geqq 0$ is a convex cone, called a halfline and denoted by (b). Thus

   $$(b) = \{x|\ x = \lambda b, \quad \lambda \geqq 0\}$$

   4. The set of all solutions of the inequality $xb \leqq 0$ is a convex cone, called halfspace and denoted by $(b)^*$. Formally, we have

   $$(b)^* = \{x|\ xb \leqq 0\}$$

5. As a generalization of 4, the set of all solutions of the homogeneous inequality $xA \leqq 0$, for any matrix A, forms a convex cone.

Explicitly:
If C is a convex cone, the dual cone C^* is defined by

$$C^* = \{y|\ yx \leqq 0 \quad for\ all\ x \in C\}$$

> ... < *The dual of the cone C is seen to consist of all vectors making anonacute angle with all vectors of the cone C.*
Note that in the above, $\lambda > 0$ would not work, and that a *halfspace* is not a finite cone. Finally, It is important not to confuse addition of cones as defined, with their union.

4. Not only is the cone a generalization of a subspace, the duals also fit. The dual of a subspace seen as a cone (which then must include 'opposing' vectors), is the intersection of two halfspaces, giving a hyperplane. This can be seen with the example of the cone of a vector, its dual, the subspace (its span), and its (orthogonal) dual.
5. The idea that the double dual of a cone must not necessarily coincide with the cone itself, is explained without a counter-example. In fact, within our context, there is no counter-example (Lakatos). To find one, we need to go beyond. An example is provided in `http://www1.se.cuhk.edu.hk/~zhang/Courses/Seg5660/Lecture2.pdf`. *The above standard cones are so regular that their corresponding dual cones coincide with the original cones themselves (the primal cones). In general, they are not the same. As an example, consider the following Lp cone > ... < In convex analysis, the dual object for a convex function f(x) is known as the conjugate conjugate of f(x) > ... < The conjugate function is also known as the Legendre-Fenchel transformation. > ... < If f(x) is strictly convex and differentiable, then $f^*(s)$ is also strictly convex and differentiable. The famous bi-conjugate theorem states that $f^{**} = cl\ f$. There is certainly a relationship between the two dual objects: the dual cone and the conjugate function. > ... < In linear programming, it is known that whenever the primal and the dual problems are both feasible then complementary solutions must exist. This important property, however, is lost, for a general conic optimization model .*
6. `http://www1.se.cuhk.edu.hk/~zhang/Courses/Seg5660/Lecture1.pdf` gives a good peak into the larger picture of convex optimization, throwing new light on cones. Similarly, [13] gives a glance at the full theory. Per example we read:
The modern theory of optimization in normed linear space is largely centered about the interrelations between a space and its corresponding dual–the space

consisting of all continuous linear functionals on the original space. In this chapter we consider the general construction of dual spaces, give some examples, and develop the most important theorem in this book– the Hahn-Banach theorem.

In the remainder of the book we witness the interplay between a normed space and its dual in a number of disctinct situations. Dual space plays a role analogous to the inner product in Hilbert space; by suitable interpretation we can develop results extending the projection theorem solution of minimum norm problems to arbitrary normed linear spaces. Dual space provides the setting for an optimization problem "dual" to a given problem in the original (primal) space in the sense that if one of these problems is a minimization problem, the other is a maximization problem. The two problems are equivalent in the sense that the optimal values of objective functions are equal and solution of either problem leads to solution of the other. Dual space is also essential for the development of the concept of a gradient, which is basic for the variational analysis of optimization problems. And finally, dual spaces provide the setting for Lagrange mutlipliers, fundamental for a study of constrained optimization problems.

We read between the lines of this excellent paragraph and note:

(a) Despite all warnings, it is true that dual spaces really get their importance with finite dimensional spaces.

(b) It is true that the 'dual of a problem' is directly related with 'dual of a space' in the sense that the vector spaces in which the vectors reside, are duals.

Additionally, we read the sobre:

Our approach in this chapter is largely geometric. To make precise mathematical statements however, it is necessary to translate these geometric concepts into concrete algebraic relations.

7. Perhaps the most striking illustration of duality in optimization is given in [13, p9]: *Many of these duality principles are based on the geometric relation illustrated in Figure 1.3. The shortest distance from a point to a convex set is equal to the maximum of the distances from the point to a hyperplane separating the point from the convex set. Thus, the original minimization over vectors can be converted to a maximization over hyperplanes.*
8. We note explicitely that the existence of a norm functional for some space, with the constraint of 'linearity', gives a dual space. Now it happens that the dual for a finite vector space is of the same dimension, and we note here that the specific situation arises in most generality exactly in Hilbert spaces, where the inner product is equivalent to orthgonality *orthogonality; a concept with is not generally available in normed space but which is available in*

Hilbert space. . The path leads form vector spaces, to normed spaces [p22], to Hilbert spaces [p46], where we read:

> *Every student of high school geometry learns that the shortest distance from a point to a line is given by the perpendicular from the point to the line. This highly intuitive <never was for me> result is easily generalized to the problem of finding the shortest distance from a point to a plane; furthermore one might reasonably conjecture that in n-dimensional Euclidean space the shortest vector from a point to a subspace is orthogonal to the subspace. This is, in fact, a special case of one of the most powerful and important optimization principles–the projection theorem.*
> *The key concept in this observation is that of orthogonality; a concept with is not generally available in normed space but which is available in Hilbert space. A Hilbert space is simply a special form of normed space having an inner product defined which is analogous to the dot product of two vectors in analytic geometry. Two vectors are then defined as orthogonal if their inner product is zero.*
> *Hilbert spaces, equipped with their inner products, possess a wealth of structural properties generalizing many of our geometrical insights for two and three dimensions. Correspondingly, these structural properties imply a wealth of analytical results applicable to problems formulated in Hilbert space. The concepts of orthonormal bases, Fourrier series, and least-squares minimization all have natural setting in Hilbert space.*

Having understood 'structure' with the aid of Abstract Algebra, we can finally appreciate the axioms of a pre-Hilbert space ($(x|y)$ being the scalar inner product):

(a) $(x|y) = \overline{(x|y)}$
(b) $(x + y|z) = (x|y) + (y|z)$
(c) $(\lambda x|y) = \lambda(x|y)$
(d) $(x|x) =\leq 0$ and $(x|x) = 0$ iff. $x = \mathbf{0}$.

Finally, since a pre-Hilbert space is normed linear, the concepts of convergence, closure, completeness, etc. apply, and we have: *A complete pre-Hilbert space is called a Hilbert space.*
In other words, in the list of spaces, Hilbert space is the most general space in which we can start to talk about a dual space whose structure describes orthogonality (which is stronger than linear independence) in the primal space. This is useful for optimzation thanks to the projection-theorem. One could say that Hilbert spaces are the most general ones where the experience based two and three dimentional intuition still works reliably.

9. naturality has a precise meaning in cat theory, as explained in `http://www.math.umn.edu/~garrett/m/algebra/notes/25.pdf`

Note 33. On Duality.

1. It is important to differentiate between the different dualities involved (Dual of a vector space V^*, of a linear programming problem P^*, of a cone C^*), before relating them. It is equally important to realize that, within a vector space, 'picking' can only be done using coordinates, and hence relating elements between dual spaces, can only be done after singling out a specific correspondence between a space and its dual, while in general there are many such correspondences, some of which might be special. An example of confusion is that of the dual of a vector space by its linear functional space, and the dual of a subspace (or of a cone more generally). The dual of a subspace usually means its orthogonal complement. This, applied to the whole space, would give the zero vector as the dual, whereas the dual of the space in terms of functionals is a space of equal dimension, and not of dimension zero.
2. [1, p99] . Dual cones are the first mention of duality so far, where the dual cone S* defined by $S^* = \{y : y^T x \geq 0, \text{ for all } x \in S)$. This might be motivated by the following fact: a dual cone is always closed and convex is simple: A dual cone is the intersection of (possibly infinitely many) closed halfspaces [p31]. It seems that the 'closedness' property will lead to an important argument that is yet to come, could it be that there is a nice relation between closedness and approximation and convergence, bringing us back to topology and a finite number of open covers? Actually, this may be the first instance where we see the recurrently mentioned important property of compact sets behaving somehow like finite sets. The fact that dual cones are finite also seems to be important, and the better way of explaining finite cones is found in [7]: If C is a convex cones and there exists a finite set of vectors $\{a_1, .., a_n\}$ in C such that every vector in C can be represented as a linear combination of the ai with non-negative coefficients, a non-negative linear combination, we call $\{a_1, .., a_n\}$ the generators of C and call C a finite cone. The cone generated by a single non-zero vector is called a half-line. A dependable picture of a finite cone is formed by considering that half-lines formed by each of the generators as constituting an edge of a pointed cones as in Fig. 4. By considering a solid circular cone in R^3 it should be clear that there are convex cones that are not finite. A finite cone is the convex full of a finite number of half lines.
3. A better explanation is given for duality in [7], relating it to the simple example of inner products, already in the preface [7, ix] we read:
   The concept of duality receives considerably expanded treatment in this second edition. Because of the aesthetic beauty of duality, it has long been a favorite topic in abstract mathematics. I am convinced that a thorough understanding of this concept also should be a standard tool of the applied mathematicians and others who wish to apply mathematics. Several sections of the chapter concerning applications indicate how duality can be

used. For example, in Section 3 of Chapter V, the inner product can be used to avoid introducing the concept of duality. This procedure is often followed in elementary treatments of a variety of subjects because it permits doing some things with a minimum of mathematical preparation. However, the cost of loss of clarity is a heavy price to pay to avoid linear functionals. Using the inner product to represent linear functionals in the vector space overlays two different structures on the same space. This confuses concepts that are similar but essentially different. The lack of understanding which usually accompanies this shortcut makes facing a new context an unsure undertaking. I think that the use of the inner product to allow the cheap and early introduction of some manipulative techniques should be avoided. It is far better to face the necessity of introducing linear functionals at the earliest opportunity.

We conclude from this that linear functionals and duality should be studied from [7].

4. A tip from Howard Karloff's [Linear Programming] leads to [3] which, unlike many other failed attempts that we found, motivates duality in our context quite constructively, and also links the non-negativity definition to the natural conditions of non-negativity of resource quantities in economic problems. About duality we read:
Every maximization LP problem in the standard form gives rise to a minimzation LP problem called the dual problem. The two problems are linked in an interesting way. Every feasible solution to one yields a bound on the optimal value of the other. In fact, if one of the two problems has an optimal solution, then so does the other, and the two optimal values coincide. This fact, known as the Duality Theorem, is the subject of the present chapter. We shall also note that, in managerial applications, the variables featured in the dual problem can be intereprested in a very useful way. Motivation: Finding Upper Bounds On The Optimal Value. We shall begin this chapter with the following LP problem:

$$\begin{aligned} \text{maximize } 4x_1 + x_2 + 5x_3 + 3x_4 \\ \text{subject to } x_1 - x_2 - x_3 + 3x_4 \leq 1 \\ 5x_1 + x_2 + 3x_3 + 8x_4 \leq 55 \\ -x_1 + 2x_2 + 3x_3 - 5x_4 \leq 3 \\ x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Rather than solving it, we shall try to get a quick estimate of the optimal value z^* of its objective function. To get a reasonably good lower bound on z^*, we

need only come up with a reasonable good feasible solution. For example, the bound $z^* \geq 5$ *comes from considering the feasible solution (0,0,1,0). The feasible solution (2,1,1,*$\frac{1}{3}$*) shows that* $z^* \geq 15$*. Better yet, the feasible soltion (3,0,2,0) yields* $z^* \geq 22$*. Needless to say, such guesswork is vastly inferior to the systematics attack by the simplex method: even if we were lucky enough to hit on the optimal solution, our guess would provide no proof that the solution is indeed optimal.*
We shall not pursue this line any further: the subject of this chapter stems from a similar quest for upper bounds on z^**. For example, a glance at the data suggests that* $z^* \leq \frac{275}{3}$*. Indeed, multiplying the second constraint by* $\frac{5}{3}$ *we obtain the inequality* $\frac{25}{3}x_1 + frac53x_2 + \ldots \leq \frac{275}{3}$*.Hence every feasible solution* $(x_1, x_2, \ldots)$ *satisfies the inequality*

$$4x_1 + x_2 + \ldots \leq \frac{25}{3}x_1 + \frac{5}{3}x_2 + \ldots \leq \frac{275}{3}.$$

In particular, this inequality holds for the optimal solution and so $z^* \leq \frac{275}{3}$*. With a little inspiration, we can improve this bound considerably. For instance, the sum of the second and third constraint reads* $4x_1 + 3x_2 + \ldots \leq 58$*. Therefore,* $z^* \leq 58$*. Rather than searching for further improvements in a haphazard way, we shall now describe the strategy in precise and general terms. We construct linear combinations of the constraints. That is, we multiply the first constraint by some number* y_1*, the second by* y_2*, the third by* y_3 *and then we add them up. >...<. The resulting inequiality reads*

$$(y_1+5y_2-y_3)x_1+(-y_1+y_2+2y_3)x_2+(-y_1+3y_2+3y_3)x_3+(3y_1+8y_2-5y_3)x_4 \leq y_1+55y_2+3y_3$$

Of course, each of these three multipliers y_i *must be nonnegative: otherwise ther corresponding inequality would reverse its direction. Next, we want to use the left-hand side of (5.1) as an upper bound on* $z = 4x_1+x_2+5x_3+3x_4$*.> ... < we want*

$$\begin{aligned} y_1 + 5y_2 - y_3 &\leq 4 \\ -y_1 + y_2 + 2y_3 &\geq 1 \\ -y_1 + 3y_2 + 3y_3 &\geq 5 \\ 3y_1 + 8y_2 - 5y_3 &\geq 3. \end{aligned}$$

>...<Thus, we are led to the following <dual >LP problem.

$$\begin{aligned} \text{minimize } & y_1 + 55y_2 + 3y_3 \\ \text{subject to } & y_1 + 5y_2 - y_3 \geq 4 \\ & \dots \end{aligned}$$

Upper bound motivation aside, duality has both practical and theoretical advantages. *In certain cases, we may find it advantageous to apply the simplex method to the dual of the problem that we are really interested in. (Of course, the optimal solution of the primal problem can then be read directly off the final directory for the dual.) For example, if m=99 and n=9, then the dictionaries will have 100 rows in the primal problem but only 10 rows in the dual. Since the typical number of simplex iterations is proportional to the number of rows in a dictionary and relatively insensitive to the number of variables, we shall most likely be better off solving the dual problem.*
From a theoretical point of view, duality is important because it points out an elegant and succint way of proving optimality of solutions of LP problems: as we have observed, an optimal solution of the dual problem provides a "certificate of optimality" for an optimal solution of the primal problem, and vice versa. Furthermore, the duality theorem asserts that for every optimal solution there is a certificate of optimality.
Duality also leads to Complementary Slackness [p63], which makes a necessary and important appearance even before talking about LCP per se, noting that degeneracy complicates things here as well [p65]: *Of course, this straightforward strategy for verifying optimality of allegedly optimal solutions is applicable only if the system of equations (5.22) has a unique solution.*

5. We find in [6, p507] A good summary with proofs of the relevant 'Theorems of Alternatives', these show duality in a clear light. On the highest level, duality is simply the situation where the complement of the solution set to a problem, can also be described finitely and elegantly, as a solution to a dual problem that is completely determined by the primal one. Logically, there is no need for the primal to be more convenient to solve, and additionally, the sandwiching effect allows for for more ways to check the 'boundary', the 'optimal solution'. The presented theorems are:
6. The unwarranted starting difficulty with dual spaces is caused by the fact that they seem very natural and obvious at first. Reading in detail Nering's treatment [7, p129-p134], we find that the following 'subtleties' are the cause of difficulties:
   (a) There is no a priori specific dual for each vector

$Ax = b$	$\pi A = 0$
	$\pi b = 1$
$Ax = b$	$\pi A \leqq 0$
$x \geqq 0$	$\pi b > 0$
$Ax > 0$	$\pi A + \mu B + \gamma C = 0$
$Bx \geqq 0$	$\pi \geq 0$
$Cx = 0$	$\mu \geqq 0$

(b) The dual base that is constructed is a particular one, used to prove the 'trivial' fact that the dual space (not to be tought of as the space of duals) is of the same dimension as the original space. Given the same dimension, of course there is an isomorophism, per example using that base.

(c) That special base is a coordinate picking base, of course, mapping each original base vector to its Kroenecker delta vector. This of course picks a dual for each vector, the dual being the coordinates given a chosen original base. If the chosen original base is identity, the the dual of a vector is the vector itself. A short examination shows that the dual bases is nothing but the rows of the inverse matrix of the matrix formed by the original basis. Coordinate picking has in essence nothing to do with the dual space, but it can be fruitfully used a simple device to find a basis for the dual space, picking coordinates using linear functionals.

(d) Recursicely, the dual of the dual is also isomorphic, hence the double dual is isomorphic to the original.

(e) There should be no attempt to interpret geometrically dual vectors, at least not in the trivial way, thinking of geometrical projections in euclidean space. Nonetheless, a geometric interpretation is possible using projective spaces, passing through level sets [10]. In that context, the dual of a vector is that 'vector' (functional) that has the same coordinates, each in their respective base. So the dual of the 'point' with coordinates $\binom{a}{b}$, is the linear map with coordinates $\binom{a}{b}$ with action $\binom{x}{y} \xrightarrow{f} (a\phi_1 + b\phi_2)\binom{x}{y}$ with (ϕ_1, ϕ_2) being the basis of the dual space and $\binom{x}{y}$ being a vector and not its coordinates. Indeed we read: *The duality between V and V^* reminds us of the duality inside projective spaces between k-dimensional subpaces or k-blades and (n-k-1)-dimensional ones, e.g. between points and planes in 3-space. Is this a coincidence, or is there a connection between these two dualities?*

and *Conversely, given a projective space P, one can -by singleing out a hyperplane* ∞ *and a point- construct a vector space* $V = P\backslash\infty$ *and its dual* $V^* = P\backslash O$. *(To be precise, one should consider P as the set of all points, lines etc., including the empty set and the whole space. V would be the set of points that are not in* ∞, V^* *would be the set of hyperplanes that do not contain O.) See again my On the Fundamentals of Geometry.*
The projective space duality mentioned, is summarized in the note: *The equation of a line in the projective plane may be given as sx + ty + uz = 0 where s, t and u are constants. Each triple (s, t, u) determines a line, the line determined is unchanged if it is multiplied by a nonzero scalar, and at least one of s, t and u must be non-zero. So the triple (s, t, u) may be taken to be homogeneous coordinates of a line in the projective plane, that is line coordinates as opposed to point coordinates. If in sx + ty + uz = 0 the letters s, t and u are taken as variables and x, y and z are taken as constants then equation becomes an equation of a set of lines in the space of all lines in the plane. Geometrically it represents the set of lines that pass though the point (x, y, z) and may be interpreted as the equation of the point in line-coordinates. In the same way, planes in 3-space may be given sets of four homogeneous coordinates, and so on for higher dimensions.*
and expounded in countless references, per example `http://math.berkeley.edu/~monks/papers/DualityV3.pdf` which gives the 'Fano projective plane F_7' as a basic example. de Boer stresses the tight connection between vector spaces and projective geometry, and goes to say that *from a geometrical point of view, it is natural to consider a vector space as a subset of a projective space.* Indeed, he explains using projective geometry the addition of covectors, their scalar multiplication, the action of a covector on a vector, and free covectors. Finally, he mentions the many (and dangerous) cases of confusion that arise from being blind to dual spaces. Per example, he states: *New, vectors are closely related to infinitesimal calculus and differential equations. It is at the same time lucky and tragic that the tangent space of Euclidean space is isomorphic to the space itself: lucky, because it simplifies computations a lot, tragic, because it obscures the difference in quality between points and vectors.*. His free book [11] is devoted to expounding this view.

(f) As obvious as it may sound, there should be no confusion between the basis vectors of the constructed dual basis, and the fact that the linear functionals are row matrices and not square ones.

(g) A motivating path to tensor diagrams (and geometric algebra) for the sake of trivializations of (multi)linear algebra is the presentation [12]. Within the slides, we find the following trivializers:

   i. The idea that a relatively simple and elegant system emerges when abstract away the details in (multi-)linear algebra, to which one might be intuitively resistant, is in fact an important technique of general scope.
   ii. A diagram on change of basis.
   iii. Covariance and Contravariance find their natural meaning: A covariant 'thing' transforms like the basis, while a contravariant 'thing' does so inversly. Per example, column vectors are contravariant, viz. given such a fixed vectors, and given a change of basis, the coordinates of the vector change contravariantly. Symbolically with C being the basis change: $[e'_1, e'_2] = [e_1, e_2]C$, $\binom{x'_1}{x'_2} = c^{-1}\binom{x_1}{x_2}$, while the coordinates of a linear function (covector) are covariant. Taking this further, a linear map is covariant in 'input' and contravariant in 'output' $A' = D^{-1}AC$.
   iv. For multilinear functions, duality happens when there is no fixed input/output role, a transformation of vectors transforms covectors in the other direction.
   v. Tensor diagrams are extremely interesting.
   vi. Projective relations (equality up to scale) reduce to standard homogeneous systems.

7. `http://math.stackexchange.com/questions/598258/determinant-of-transpose/636198#636198`

Note 34. There is a tight relation between Lagrange multipliers, the Simplex Method, Saddle points and duality in LP, explained in [8, p59]: *A natural question to ask is why not use the classical method of Lagrange multipliers to solve the linear programming problem. It we were to do so we would be requited to find optimal multipliers (or prices Pi), which, if they exist, must satisfy a "dual" system with the property that > ... <. The latter leads to 2n possible cases of either $c_j = 0$ or $x_j = 0$. If is here that this classical approach breaks down, for it is not practical to consider all 2^n possible cases for large n. In a certain sense, however, the Simplex Method can be viewed as a systematic way to eliminate most of these cases and to consider only a few. Indeed, it immediately restricts the number of cases by considering only those with n-m of the xj=0 at one time and such that the coefficients matrix of the remaining m variables is nonsingular; moreover the unique value of these variables is positive (under nondegeneracy). > ... <* .
The relation to 'Either Or' theorems is also stressed: *there are several important duality-type results, known as "Either Or" theorems of the alternatives, that pre-*

dated the linear programming era: > ... < The earliest known result on feasibility is one concerning homogeneous systems, Gordan [1873].
Additionally, the Duality Gap is mentioned in [p45], providing a good dependable picture.

2 Background

Apart from "absolute basics", we extract a list of needed concepts from the "Background" chapter of [1].
From Real Analysis: convex sets, connected sets, Lipschitz continuity, Fréchet differentiability, strong F-derivative, inverse function theorem, Brouwer's fixed point theorem, convex functions.

From Matrix Analysis: principal submatrix rearrangement, matrix reducibility, matrix norm (vector induced), positive definite and semi-definite matrices, diagonally dominant matrix, nonnegative and positive matrices and their relation to the Perron-Frobenius theorem. Note that for Matrix Analysis, it seems that the eigenvalues can provide a way to reduce of all the needed concepts and their inter-relations to a small and intuitive set of ideas.

From Pivotal Algebra: the basic concepts of pivotal transform (and what it preserves), the geometrical interpretation in term of complementary cones, the Schur complement which is of great importance during pivoting but is a more general concept, and lexicographic ordering.

From Matrix Factorization: the basics (principal minors, LU factorization, permutation matrices, Cholesky factorization, Gaussian elimination, matrix factorization update).

From Iterative Methods for Equations: Specific applications of what is needed from Real Analysis, including the convergence properties of descent methods, like the damped Newton method.

From Convex Polyhedra (related to the geometry of linear inequalities): The basics (affine hull, convex hull, k-simplex, extreme point, basic and nonnegative solution of a linear system, relative interior) along with the strongly interrelated group: halfspace, polytope, convex set, closed set, cone, convex cone, finite cone, dual cone, closed cone, polyhedral cone.

From Linear Inequalities ...

From Quadratic Programming Theory ...

From Degree and Dimension ...

3 Historic Overview

3.1 From *The Linear Complementarity problem (Cottle, Pang, Stone)*

Note 35. The name of the problem we are studying in this book has undergone several changes. It has been called the "composite problem," the "fundamental problem," and the "complementary pivot problem." In 1965, the current name "linear complementarity problem" was proposed by Cottle..

Note 36. Perhaps the earliest existence result for the LCP appears in the paper by Smelson, Thrall and Wesler (1958). Motivated by an engineering application, their work was concerned with a geometric problem which can be expressed as a CP... their results yield theorem 3.3.7

Note 37. There is no known method for efficiently testing a matrix for membership in one of these classes. Aganagie and Cottle (1978) >...¡ describes an inefficient test for membership in Q... This and finite characterization for Q0-matrices can be found in Mutry (1988)...

Note 38. Laypunov (1947) introduced the notion of a (negative) stable matrix (with complex entries) in his study of solution stability of differential equations >...¡ The last conclusion in our theorem 3.3.9 is a special case of this famous result >...¡

Note 39. The H-matrices are closely connected with the diagonally dominant matrices which, of course, have a long history.

Note 40. The invariance (of P) under principal pivoting was shown by Tucker in [TUCKER, A.W. (1963)].

Note 41. Reading these sections, it becomes very clear that the classification of matrices in mathematical programming is fundamental, and that there are many classes. It almost seems like fixing the LCP algorithm and ignoring the matrix class is simply wrong, and that the right point of view is a per matrix class one, of course with inter-relations. A brief search points to Cottle's recent article [4], from the abstract of which we quote: *There are more than 50 matrix classes discussed in the literature of the Linear Complementarity Problem. This guide is offered as a compendium of notations, definitions, names, source information, and commentary on these many matrix classes. Also included are discussions of certain properties possessed by some (but not all) of the matrix classes considered in this guide. These properties—fullness, completeness, reflectiveness, and sign-change invariance—are the subject of another table featuring matrix classes that have one or more of them. Still another feature of this work is a matrix class inclusion map depicting relationships among the matrix classes listed herein.* Dantzig also mentions the thesis 'Stone, R.E.: Geometric Aspects of the Linear Complementarity Problem. Ph.D. thesis, Department of Operations Research, Stanford University (1981)' where a very comprehensive (until 1981) compilation of matrix classes including diagrams and relations is presented.

Note 42. The richness of the subject [classes K and Z] is evident in a remarkable theorem of Fiedler and Ptak stating 13 equivalent conditions under which a Z-matrix is a K-matrix. This feat was surpassed by Berman and Plemmons (1979) who listed 50 such conditions … Even the Berman-Plemmons collection is not exhaustive; …

Note 43. Minus the name, the class of hidden Z-matrices was introduced in a paper by Mangasarian (1976a) which had to do with the idea of solving LCPs as linear programs. … The least-element theory of the LCP with a hidden Z-matrix was developed in an effort to explain the phenomenon observer by Mangasarian in the aforementioned papers.

Note 44. The invariance of P under principal pivoting, was also shown by Tucker (1963). The invariance theorem for positive definite and positive semi-definite matrices is from an unpublished paper of Tucker and Wolfe, it is cited in Cottle (1964a)…

Note 45. Algorithm 4.4.1 and its streamlined version are what is usually meant by the term "Lemke's method". Lemke, himself, called it "Scheme I."… For references on the effect of the covering vector d on the performance of Lemke's method see Mylander (1971, 1974), Todd (1986), and Krueger (1985).

Note 46. The treatment of the traffic equilibrium problem by complementarity and variational inequality methods originated from a paper by Smith (1979)… Many paradoxes arise in the study of the traffic equilibrium problem. …. [p.306] The most famous among these is the one due to D. Braess which demonstrates that in a congested transportation network, it is possible for the elimination of a path to decrease the cost of all paths with positive flows. This is somewhat counter-intuitive because one would expect … at least one (and hence every, …) used path in the network would suffect an increase in cost.

Note 47. The vast amount of literature on the LCP of the Z-type is evidenced by... This abundance of work is attributable to the numerous applications that this special class of LCP's possesses, and to the fact that the Z-property easily lends itself to some fruitful analysis.

Note 48. The latter study < of Theorem 4.8.7 > was inspired by a geometric question >... < which asked whether the set of n-step vectors would form the interior of a simplical cone. Morris and Lawrence provided an affirmative answer >... < The ingenuity of their proof lies in its use of the matrices Mi whose significance in LCP theory had not been emphasized before.

Note 49. Degeneracy resolution in the LCP has been studied by Chang (1979). >... < Chang obtained results on the sizes of the smallest LCP's in which cycling can occur with different pivoting methods. (Chang's findings do not quite agree with those of Kostreva (1979) who used different ground rules). A double least-index rule for resolving degeneracies in linear programming was pioneered by Bland (1977) and later extended to quadratic programming by Chang and Cottle (1980). >... < Chang (1979) extended the concept of a least-index to Lemke's method. In each case, however, the technique was restricted to instances in which the matrix M of the LCP was either positive semi-definite or a P-matrix. >... < As of this writing, a way to do the analogous thing relative to Lemke's method has not been found...

Note 50. The examples of exponential worst case behavior of Algorithms >... < are due to Murty (1978a)...

Note 51. Several pivoting methods for the LCP have not been discussed in this chapter. >... < Some of these algorithms address the heroic task of finding all solutions to an LCP or an even more general problem.

Note 52. The early study of iterative methods for solving the LCP was mainly concerned with the symmetric problem and its applications to a non-negatively constrained convex quadratic program. Hilderth (1954, 1957) developed a projected Gauss-Seidel relaxation method for solving a strictly convex quadratic program with only inequality constraints; his method actually solved the dual problem which was equivalent to an LCP (although it was not recognized at the time).>... <

Note 53. On the application to the contact problem: *The finite-dimensional contact problem was sketchily formulated as an LCP by Fridman and Chernina (1967) who proposed an iterative scheme (PGS) for solving it.>... <*

Note 54. On the convergence study for iterative schemes: *Mangasarian (1977) proposed a very general iterative scheme for solving the symmetric LCP and established its convergence under some fairly broad assumptions. This paper can be credited as being the first one in which a systematic study of the convergence of iterative methods for the LCP was carried out; it has provided the impetus for much subsequent research on this subject, among which is the work of Aganagic (1978a) and Ahn (1981) who investigated the convergence of iterative methods for solving the asymmetric LCP. On matrix splitting: Inspired by the aforementioned papers of Mangasarian, Aganagic, and Ahn, Pang (1982) introduced a matrix-splitting algorithm as a unification of many of the iterative methods for solving the LCP. Incidentally, Mangasarian's scheme is general enough to include the splitting algorithm as a special case.* >... < *The advantage of the splitting framework is its simplicity and ease of analysis.*

Note 55. Aganagic (1978a) initiated the use of the norm contraction approach in the convergence analysis of the simple iteration...

Note 56. Ahn (1981) was the first one to employ a vector contraction approach to analyze the convergence of Mangasarian's 1977 scheme for an asymmetric LCP.

Note 57. On the relaxation parameter in SOR methods: *The family of SOR methods is among the most effective for solving large, sparse LCPs. The efficiency of these methods is crucially dependent on the choice of the relaxation parameter w, see (5.2.1). There is little theory concerning the choice of an optimal parameter value. Generally speaking, if one can identify the positive variables of the limit solution in a finite number of iterations, then the iterative method essentially reduces to that for solving a system of linear equations. (Exercises 5.11.2 and 5.11.3 are relevant to this consideration.) In this case, it becomes possible to borrow from the theory of linear equations to help identify a good value for w. See the text by Hageman and Young (1981) for more discussion on the latter subject.*

Note 58. On large scale linear programs and method mixing: *The SOR methods, in conjunction with a type of proximal-point scheme (see 5.12.16), have been used extensively for solving large-scale linear programs. Discussions of these applications can be found in the work of Mangasarian* >... ¡

Note 59. On parallelization: *Implementation of the resulting algorithms in a parallel computation environment is discussed in Mangasarian and De Leone (1987, 1988b)* >... ¡

Note 60. On a long unresolved question, that of sequential convergence: *The topic treated in Section 5.4 was an open research question for a long time. Inspired by Mangasarian's 1977 paper, many of the early theoretical studies of the iterative methods for solving the symmetric LCP were mainly concerned with the notion of subsequential convergence of the iterates. There was generally a lack of sequential convergence results except in the positive definite case—Cryer (1971b) and a few special instances—Pang (1986a). Then, a breakthrough occurred with a paper by Luo and Tseng (1991) in which they proved Theorem 5.4.6, thus settling an outstanding question. This result was independently established by De Pierro and Iusem (1993) whose proof we follow, as it is somewhat easier to comprehend than that of Luo and Tseng.*

Note 61. Gana (1982) and Venkateswaran (1993) discuss the regularization idea applied to an LCP with a Po –matrix >. . . ¡ In Gana (1982) and Kostreva (1989), the authors claim the convergence of the entire sequence of iterates; however, their "proofs" are based on an invalid argument. . . .

Note 62. On the damped-Newton method and the NLCP: *The damped-Newton method (Algorithm 5.8.5) is a specialization of some Newton-type methods for solving certain B-differentiable systems of nonsmooth equations proposed by Pang (1990a). Our presentation of this algorithm is based on Harker and Pang (1990b) in which some computational results are reported. A refinement of the algorithm that involves solving LCPs of smaller sizes can be derived from the method described in Pang (1991b). The theoretical advantage of the refined algorithm is that a quadratic rate of convergence can be established under the assumption of sequential convergence Harker and Xiao (1990) and Xiao (1990) discuss extensively the application of the nonsmooth Newton methods for solving the nonlinear complementarity problem. The computational results they report provide evidence of the practical efficiency of these methods for solving realistic applied problems.*

Note 63. The all-change algorithm, as the resulting method was called in the reference, differs from Algorithm 5.8.5 in that it contains no linesearch routine. In the absence o this important step, the algorithm becomes a kind of heuristic procedure for solving the LCP. Indeed, Kostreva offered no theoretical justification for his all-change algorithm.

Note 64. On the interior-point method. *The interior-point method (Algorithm 5.9.3) originates from an algorithm introduced by Karmarkar (1984) for solving linear programs. Due to its remarkable practical efficiency and dramatic departure from the traditional simplex method, there is an abundance of research on the latter algorithm. The volumes edited by Megiddo (1989b) and Gay, Kojima and Tapia (1991) contain excellent collections of papers in this area. The extension* *of Karmarkar's algorithm to the LCP has been the subject of many papers. >. . . ¡ The last ¡ Noma and Yoshise (1991).¿ of these is an extensive survey paper which presents a unified approach to the entire subject and contains a long bibliography.*

Note 65. On numerical continuation (of the interior-point method) methods: *The reader is referred to the book by Allgower and Georg (1990) for a general introduction to the subject of numerical continuation methods for solving systems of nonlinear equations.*

Note 66. On homotopy: *This concept dates back to at least the nineteenth century. Homotopies are a powerful tool both analytically and computationally. They can be used to prove theorems and, indeed, much topological theory is based on them. They can also be used to find numerical solutions to differential equations, integral equations, and nonlinear systems of equations. >. . . ¡ One may view Lemke's method as being akin to the homotopy concept (see the beginning of Section 6.3). In addition, homotopy algorithms exist for the general complementarity problem, fixed point problems, and other LCP related problems. The reader is referred to 2.11.1 for a brief historical account and some references concerning these homotopy (fixed-point) methods. >. . . ¡ For a different ¡then the degree-theoretic one¿ topological approach to the LCP, see Naiman and Stone (1998) where homology theory is used to obtain a characterization of the class Q which leads to a test for membership. This test has better time complexity, but worse space complexity, than the one attributed to Gale (see Note 3.13.4). Unfortunately, both tests are inefficient.*

Note 67. On the lack of formal studies of residues and error bounds: *some of the earliest iterative methods for solving the LCP appeared in the nineteen fifties. Nevertheless, in the field of mathematical programming, there are rather few formal studies of residues and error bounds. In the context of the LCP, the papers by Mangasarian and Shiau (1986), Mathias and Pang (1990), Mangasarian (1990a, 1992), and Luo and Tseng (1992b, 1992c) seem to be the only available references. The two related papers, Pang (1986b, 1987), discuss some error-bound results for the nonlinear complementarity problem and the variational inequality problem. >. . . ¡ based on the earlier results of Mangasarian (1981b) which introduced the notion of a condition number for systems of linear inequalities (see Lemma 5.10.13). Hoffman (1952) is believed to be the first person to have derived such error-bound results for linear inequalities.*

Note 68. On error bounds and sensitivity theory: *The subject of local error bounds is closely related to that of sensitivity theory. The reader is referred to Section 7.7 for notes and references on the latter subject.*

Note 69. On complementary cones and path-following: *Subsequently ¡Murty¿ , complementary cones, their facets, and path-following arguments have become standard tools in the study of the LCP. >. . . ¡ When path-following arguments are employed to study the LCP, it is common to require that certain "degenerate" points not be on the path which is used. At these points it is difficult to keep track of which complementary cones the path is entering and/or leaving.*

Note 70. On "Lemke's ghost", a graph-theoretic convergence argument*: the underlying idea behind the proof of Proposition 6.2.18 is a graph-theoretic argument. This is also the reasoning which underlies the proof of convergence of Lemke's algorithm (Theorem 4.4.4). In fact, this argument can be found in the original papers describing Lemke's algorithm (Lemke and Howson (1964) and Lemke (1965)). One sometimes encounters this reasoning referred to as "Lemke's ghost" argument alluding to a memorable interpretation due to B.C. Eaves (see Section 2.2.6 of Murty (1988)). Cottle and Dantzig (1970) explicitly describe the graph theory behind this argument. They then extend the argument to cover the vertical generalization of the LCP (see (1.5.3)). For a treatment of Lemke's algorithm in terms of graph theory and combinatorial topology see Shapley (1974).*

Note 71. By the definitions given in Saigal (1972b), the common facet of a full cone and a degenerate cone would be considered a proper facet. This means that some results given in Saigal (1972b) require certain corrections, as is documented in Stone (1981) and Saigal and Stone (1985)

Note 72. The basic results concerning the behavior of Lemke's method as it relates to complementary cones and their facets is first taken up in Saigal (1972b)

Note 73. On relations to spherical geometry: *Several works make use of spherical geometry in studying the LCP. Instead of working with a complementary cone directly, one works with the intersection of the complementary cone and the unit sphere. (If the complementary cone is nondegenerate, this intersection will be a nondegenerate spherical simplex.)*

Note 74. On the first works on sensitivity analysis in linear and nonlinear programming, and its relation to the classical implicit function theorem: *The systematic study of sensitivity analysis in nonlinear programming starts with the seminal work of Fiacco and McCormick (1968) in which the classical implicit function theorem was used to obtain the first perturbation properties of parametric nonlinear programs. This subject has since developed into a very fruitful area of research within mathematical programming. A large body of literature is available, and several texts have been written, with Bank, Guddat, Klatte, Kummer and Tammer (1983) and Fiacco (1983) being two of the more recent additions. The* *former text contains some discussion of the sensitivity issues pertaining specifically to the LCP. >... ¡ The most significant departure of Robinson's research from the classic results of Fiacco and McCormick is that the assumption of strict complementary slackness is removed. This is a major contribution because with this assumption in place, sensitivity analysis of nonlinear programming (including complementarity problems) can typically be carried out by a straightforward application of the implicit function theorem,*

Note 75. On the relation of sensitivity analysis to Lipschitzian properties and to the Schur complement: *although much of Robinson's work is not directly cast in the framework of the LCP, many of the sensitivity results presented in this chapter are derived as special cases of his discoveries. In particular, Theorem 7.2.1 is a specialization of an upper Lipschitzian property of general polyhedral multifunctions obtained in Robinson (1981). >... < throughout the study of the sensitivity and stability of a solution of the LCP, the Schur complement N defined by the expression (7.2.4) has played a major role.*

Note 76. On the recency of convergence results for splitting methods: *Several of the convergence results of the splitting methods described in Section 7.2 appeared in the literature only recently. The implication (a) r (b) in Theorem 7.2.10 was obtained by Iusem (1990a) and Tseng and Luo (1990).*

Note 77. Using Robinson's notion of a strongly regular solution, Josephy (1979a, 1979b) was the first person to establish the convergence of Newton's method and the quasi-Newton methods for solving a generalized equation.

Note 78. On solution rays and structural mechanics: *Our discussion of solution rays of the LCP is based on Cottle (1974b). This study was inspired by some questions arising from structural mechanics that were raised in a private communication by G. Maier to Cottle in October, 1973.*

Note 79. On stability, boundedness and ellipsoids: *the stability of a copositive-plus LCP is intimately related to the boundedness of the solution set of the problem.>... ¡ Kojima, Mizuno and Yoshise (1990) investigate properties of ellipsoids that contain all solutions of a positive semi-definite LCP with bounded solution set. The latter investigation is motivated by the family of interior-point methods discussed in Section 5.9.*

3.2 From *Nonlinear programming: A historical view (Kuhn)*

Note 80. He explains that the Sylvester problem, published in the one sentence note: *It is required to find the least circle which shall contain a given set of points in the plane (1857), generalized, was later recognized as a quadratic program, more precisely a hybrid program.*

Note 81. In summary, it was not the calculus of variations, programming, optimization, or control theory that motivated Fritz John but rather the direct desire to find a method that would help to prove inequalities as they occur in geometry.

Note 82. It is poetic justice that Fritz John was aided in solving this problem by a heuristic principle often stressed by Richard Courant that in a variational problem where an inequality is a constraint, a solution always behaves as if the inequality were absent, or satisfies strict equality.

Note 83. On slack variables: *Karush's proof is a direct application of a result of Bliss [20] for the equality constrained case, combined with a trick used earlier by Valentine [16] to convert inequalities into equations by introducing squared slack variables..*

Note 84. Tucker has a chance to return to question: What was the relation between linear programming and the Kirchhoff-Maxwell treatment of electrical networks? It was at this point that he recognized the parallel between Maxwell's potentials and Lagrange multipliers and identified the underlying optimization problem of minimizing heat loss.

Note 85. Kuhn produced a one-page working note expressing the duality of linear programming as saddlepoint of the Lagrangian expression.

Note 86. The background of the work of Karush was so different from that of Kuhn and Tucker that one must marvel that the same theorem resulted.

Note 87. Katayama says: Linear programming aroused interest in constraints in the form of inequalities and in the theory of linear inequalities and convex sets. The Kuhn-Tucker study appeared in the middle of this interest with a full recognition of such developments. However, the theory of nonlinear programming when the constraints are all in the form of equalities has been known for a long time – in fact, since Euler and Lagrange. ... Karush's work is apparently under the influence of a similar work in the calculus of variations by Valentine.

Note 88. With malice aforethought and considerable historical hindsight, a nonlinear program will be defined as a problem of the following form:...
Maximize f(x1,...,xn) for "feasible" solutions to g1(x1,...,xn)- b1= y1, ... , gm(x1,...,xn)- bm= ym

1. *If all x are free and all y and b are zero, we get the classical case of equality constrained (nonlinear) optimization treated first by Lagrange.*
2. *If f,g are linear, x,y nonnegative, then we get a linear program in canonical form.*

3. *If y are zero, then we get a linear program in standard form.*
4. *The most general problem of Maximize f(x) subject to x in S is too broad for any but the most superficial results.*
5. *Even in the simple case of a 'triangle in a plane'-like problem, there are several presentations, some of which better behaved than others, per example, below, the first case is better behaved.*
   (a) *Maximize f(x1, x2) with x1+ x2-1 = - y1, -x1- 2x2+1 = - y2, with x1, x2, y1, y2* $\geq$ *0*
   (b) *Maximize f(x1, x2) with (x1+ x2-1)(x1+ 2x2-1)= - y1, with x1, x2, y1* $\geq$ *0*

Note 89. With the example of linear programming before us, the nonlinear program ... is subjected to a natural linearization which yields a set of likely necessary conditions for a local optimum... Of course, these conditions do not hold in full generality without a regularity condition... When it is invoked, the result is a theorem which has been incorrectly attributed to Kuhn and Tucker.

Note 90. On duality: *To motivate the derivation of the necessary conditions for optimality ..., let us place ourselves in the position of mathematical programmers in the late 40's. Von Neumann had given a formulation of the dual for a linear program [4] and Gale, Kuhn, and Tucker had provided rigorous duality theorems and generalizations [5]. The dual of Maximize f(x) = cx for Ax-b=-y is Minimize h(v) = vb with vA-c=u. It seems that this helps solver the original problem by considering the dual program. Kuhn adds: The phenomenon had even been raised to the level of a method... by Courant and Hilbert [6] in the following passage: "The Lagrange multiplier method leads to several transformations which are important both theoretically and practically. By means of these transformations new problems equivalent to a given problem can be so formulated that stationary conditions occur simultaneously in equivalent problems. In this way we are led to transformations of the problems which are important because of their symmetric character. ". Moreover, for a given maximum problem with maximum M, we shall often find an equivalent minimum problem with the same value M as a minimum; this is a useful tool for bounding M from above and below. On the first instances of the recognition of such a pattern Kuhn mentions among others, Fermat early in the 17th century: "Given three points in the plane, find a fourth point such that the sum of its distances to the three given points is a minimum".*

3.3 Unresolved Questions

1. Does not (c) conflict with convergence in [1, p48] ? Can a sequence in a normed space have multiple accumulation points but still converge?

2. [1, p98]Check why it follows from the above theorem that a convex polyhedron can have only finitely many extreme points.
3. More precision on the Lagrangian in relation to the Simplex Method.
4. Elaborate on Chvatal's problem dual motivation (note 4) by being more explicit about in which space the y scalars act, and relate to the geometric closest point dual interpretation (note 7). Additionally, expand on the slack variable, and its role in linking the two spaces, investigating it in its own right.

3.4 Potentially Interesting Unread Papers

1. Amaral, P., J. Judice, and H.d. Sherali. "A Reformulationlinearizationconvexification Algorithm for Optimal Correction of an Inconsistent System of Linear Constraints." Computers & Operations Research 35.5 (2008): 1494-509. Print.`http://www.co.it.pt/~judice/Articles/AmaralSheraliJudice/_CAOR1877.pdf`
2. Cottle, Richard W. "A Field Guide to the Matrix Classes Found in the Literature of the Linear Complementarity Problem." Journal of Global Optimization 46.4 (2010): 571-80. Print.
3. Linear Complementarity Since 1978, Richard W. Cottle.
4. Duality
   (a) Teaching Duality in Linear Programming - the Multiplier Approach `http://www1.imada.sdu.dk/Courses/DM85/lec4b.pdf`
   (b) Two views of Duality: Lagrangians and Geometric `http://www.cs.berkeley.edu/~satishr/cs270/sp12/slides/lecture8/lecture8.outline.pdf`
   (c) Intuition Behind Primal-Dual Interior-point Methods For Linear And Quadratic Programming `http://www.cs.ubc.ca/~pcarbo/lp.pdf`

3.5 People

1. Joaquim João Júdice `http://www.co.it.pt/~judice`

References

1. Cottle, R. and Pang, J. S. and Stone, R. E. The Linear Complementarity problem (1992)
2. P. Amaral, J. Judice, and H. D. Sherali. A reformulation–linearization–convexification algorithm for optimal correction of an inconsistent system of linear constraints. Computers and Operations Research, 35(5):1494–509, (2008/05).
3. V. Chvátal. Linear programming. Freeman and Co., (1983).
4. Richard Cottle. A field guide to the matrix classes found in the literature of the linear complementarity problem. Journal of Global Optimization.

5. Kuhn. Nonlinear programming: A historical view. Nonlinear Programming, SIAM-AMS Proceedings, 9:1–26.
6. Katta G. Murty. Linear Complementarity, Linear and NonlinearProgramming. University of Michigan, (1988).
7. Evar D. Nering. Linear algebra and matrix theory. Wiley, New York, (1970).
8. George B. Dantzig and Mukund N. Thapa. Linear Programming Vol. 2: Theory and Extensions. Springer, (2003).
9. David Gale. The Theory of Linear Economic Models. McGraw-Hill, (1960).
10. Lou de Boer. Vector spaces and projective geometry, (2009).
11. Lou de Boer. On the Fundamentals of Geometry, (2009).
12. Ruiz, A. multilinear algebra for visual geometry, (2013).
13. D Luenberger. Optimization by Vector Space Methods. John Wiley and Sons, New York, (1969).