Issue 2 JULY–DECEMBER 2000 ISSN 1083-0464

SOUTHWEST JOURNAL OF PURE AND APPLIED MATHEMATICS

CAMERON UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCES
LAWTON, OK 73505, U.S.A

Southwest Journal
of
Pure and Applied Mathematics

Editorial Office

SWJPAM
Cameron University
Department of Mathematical Sciences
Lawton, OK 73505, U.S.A.

Publication Program

2 issues per year
July, December

Information for Authors

Scope. SWJPAM is an electronic journal devoted to all aspects of Pure and Applied mathematics, and related topics. Authoritative expository and survey articles on subjects of special interest are also welcomed.

Since SWJPAM aims to serve not only the specialists, but also the user of mathematics (promoting, in such a way, a strong interaction between them), particular emphasis is placed on applications. Thus, SWJPAM publishes papers which, although not directly contributing to mathematics, contain, however, significant results (from any field) obtained by making an essential and well stressed use of the methods and results of mathematics.

Submission of Manuscripts. Since SWJPAM is a completely electronic journal, all manuscripts must be submitted in either AMSTEX, LaTEX, or TEXformat. Papers for publication should be submitted to ftp://rattler.cameron.edu/SWJPAM_incoming by anonymous ftp. The articles must be placed in the 'SWJPAM_incoming' directory. The submission should include the title of the paper, the author's name and affiliation, and a short abstract which summarizes the paper; a list of keywords and a classification according to the Mathematical Reviews' Subject Classification (1991). Authors are informed by e-mail of referee communication.

Key Words. Key words should be comprehensive and subject specific. It is not necessary to list the subject area of the journal's coverage as a key word. Four key words should be sufficient to cover the major subjects off a given paper, although more can always be supplied if the author deems it necessary.

 (continued on inside back cover)

SOUTHWEST JOURNAL
OF
PURE AND APPLIED MATHEMATICS
ISSUE 2, DECEMBER–JULY 2000

CONTENTS

ISSN 1083-0464

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2000, pp. 1 – 16
Submitted: June 22, 2000. Published: December 15, 2000

A Bargaining Procedure with Offer-Dependent Breakdown Probabilities

F. Szidarovszky and L. Shen

Abstract

This paper introduces a new solution concept for two-person bargaining. First a sequential bargaining process is proposed where break-down probabilities depend on the current offers. After showing that this process terminates in finitely many steps in a large class of bargaining problems, a one-stage solution concept is introduced which is the optimal offer for both players under given relative power indices. Three numerical examples illustrate the concept and solution methodology. The generalization of the solution for the n-person case is outlined in the last section.

A.M.S. Subject Classification Codes: 90
Keywords: *bargaining, dynamic games, solution concept.*

1 Introduction

This paper introduces a new solution concept and methodology for resolving

Ferenc Szidarovszky and Ling Shen Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona 85721, U.S.A.

two-person conflicts. In the literature, there are two major approaches for conflict resolution. The first kind of methods is based on certain conditions which the solution must satisfy. These conditions are given as axioms. In most cases, the existence and uniqueness of a solution satisfying these axioms are proved. Nash's bargaining model has been one of the simplest yet most fruitful paradigms in such concept (Nash 1950). Following the idea of Nash, several authors have suggested modification of his set of axioms. For example, Harsanyi and Selten(1972) have dropped the symmetry assumption, and have proved the existence of a unique solution which maximizes the weighted geometric average of the gains. Kalai and Smorodinsky(1975) have criticized the axiom requiring the independence of irrelevant alternatives. They have shown that in many cases Nash's solution is unfair. They have suggested that this axiom be replaced by individual monotonicity. The resulting solution is the same as a direction based method known from multi-objective programming. This concept has been further extended by Anbarci(1995) as the reference function solution. Many additional solution concepts have been introduced during the last decade. Among the most commonly known concepts we mention here the egalatarian solution of Kalai(1977), the equal sacrifice solution of Chun(1988), the equal area solution of Anbarci(1993), and the super-additive solution of Perles and Maschler(1981). The second type of methods is based on ficticious negotiation processes. Either the equilibrium of the resulting dynamic game is accepted as the solution, or the convergence of the process is proved and the limit is accepted as the solution. Nash(1953) claimed that the equilibrium set coincides with the set of Pareto solutions if bargaining is considered as a two-person noncooperative game. Bargaining can be also considered as a single player decision problem, when the strategy selection of the other player is considered to be random. Each player then maximizes his/her expected profit. If uniform distribution is assumed, then the resulting solution is equivalent to the Nash-product maximizing alternative. In dynamic negotiations, the main problem is to determine the order in which concessions are made. If the principle of Zeuthen(1930) is assumed, then the limit of the dynamic process coincides with the Nash solution. The alternating offer method (Rubinstein, 1982) is one of the most popular bargaining models. It is based on a sequence of alternating offers, and the process is complicated by the introduction of break down probabilities. Under general conditions, it is proved that there is a unique subgame perfect equilibrium. Howard(1992) has introduced an extended version of the alternating offer method, and in the model of Anbarci(1995) the payoffs of the players depend not only on their offers for themselves but also on how generous their offers are.

In this paper, a new solution concept is introduced and examined, which has certain elements from both major concepts. First a dynamic process is introduced, and then the one-step solution concept is discussed, which is an extension of the one-shot bargaining solution of Anbar and Kalai(1978). It is assumed that the solution is Pareto optimal, and it maximizes the expected profit for both players. Instead of assuming that the alternative selection of the competitor is random, offer dependent break down probabilities are introduced. The resulting nonsymmetric solution is based on the relative power indices of the players.

The paper develops as follows. In section 2, the two player concept is explored. A dynamic bargaining process is examined first, and then the existence of a unique solution is verified. In addition, an iteration scheme is proposed to find the solution. In section 3, some examples are given to illustrate the concept and the solution algorithm. Conclusions are drawn in section 4.

2 Dynamic Process and Solution Concept in the Two-Person Case

A standard two-person bargaining game S is considered, which is illustrated in Fig. 1. It is assumed that S is compact and comprehensive in R^2. We may assume without loss of generality (if necessary, after a suitable positive affine transformation of utility scales) that the disagreement point is the origin. Let the Pareto frontier be given by the graph of a twice continuously differentiable, concave and strictly decreasing function $x_2 = g(x_1)$. The concavity assumption on function g will be relaxed later. The domain of g is assumed to be a finite closed interval, and there is an $x_1 > 0$ such that $g(x_1)$ is defined and $g(x_1) > 0$.

The state of the negotiation process at each time period is a pair $((x_1, x_2), (y_1, y_2))$ of points on the Pareto frontier, such that $x_1 > y_1$ and $y_2 > x_2$. Here points (x_1, x_2) and (y_1, y_2) are considered as the positions (or current offers) of player 1 and player 2, repectively. It is also naturally assumed that $y_1 \geq 0$

and $x_2 \geq 0$, that is, no player offers less to the other player than the disagreement payoff value.

As the negotiation preceeds, the state changes when any of the players makes a concession. If player 1 makes the concession, then he/she moves to another point (z_1, z_2), which decreases his/her payoff and at the same time, increases the payoff of player 2. Player 2 will either accept the offer with probability $P_2(z_2)$ or break up the bargaining process with probability $1 - P_2(z_2)$, where $P_2 : [x_2, y_2] \mapsto [0, 1]$ is a real valued function. In the case of acceptance the players receive the agreed payoffs, and in the case of breakdown, both players get zero payoffs. Suppose that $P_2(x_2) = 0$ and $P_2(y_2) = 1$, which models the concept that no concession from player 1 will result in certain breakdown and full concession will be certainly accepted by player 2. For the sake of simplicity, assume that

$$P_2(z_2) = (\frac{z_2 - x_2}{y_2 - x_2})^{p_2} \tag{1}$$

where $p_2 > 0$ is a given constant.

Under this assumption, player 1 will select the point that maximizes his/her expected payoff:

$$\begin{array}{l} \text{maximize } z_1(\frac{z_2-x_2}{y_2-x_2})^{p_2} + 0(1 - (\frac{z_2-x_2}{y_2-x_2})^{p_2}) \\ \text{subject to } x_1 \geq z_1 \geq y_1. \end{array} \tag{2}$$

Since $y_2 - x_2$ is fixed and (z_1, z_2) is on the Pareto frontier, this optimization problem is equivalent to maximizing function

$$z_1(g(z_1) - x_2)^{p_2}. \tag{3}$$

Note that the objective value is zero at $z_1 = x_1$, and positive at $z_1 = y_1$. Hence, at the optimum, $y_1 \leq z_1 < x_1$. The optimal solution z_1 is either y_1 or a stationary point, which is a solution of equation

$$(z_1(g(z_1) - x_2)^{p_2})' = 0. \tag{4}$$

Since at the optimum, $z_1 \neq x_1$ and $g(z_1) - x_2 \neq 0$, this equation can be rewritten as

$$h(z_1) = z_1 p_2 g'(z_1) + g(z_1) - x_2 = 0 \qquad (x_1 > z_1 \geq y_1). \tag{5}$$

It is easy to see that the sign of h equals the sign of the derivative of the objective function (3). Note first that

$$h(x_1) = x_1 p_2 g'(x_1) + g(x_1) - x_2 \leq 0, \tag{6}$$

since $g'(x_1) \leq 0$, and $g(x_1) = x_2$ at the Pareto frontier. Furthermore,

$$h'(z_1) = p_2 g'(z_1) + z_1 p_2 g''(z_1) + g'(z_1) < 0, \tag{7}$$

since $g' < 0$ and $g'' \leq 0$. Therefore, h strictly decreases. In the case of $h(y_1) > 0$, there is a unique stationary point z_1 in the interval (y_1, x_1). Otherwise, $z_1 = y_1$ is the best choice. In this case player 1 accepts the offer of player 2 and the negotiation terminates.

In proving the unique optimal choice of player 1, the strict monotonicity of function h is very important. This property is directly assured by assuming that g is concave. We will next show how this condition can be relaxed and the strict monotonicity of h still holds. From equation (7), we conclude that $h'(z_1) < 0$ for all $p_2 > 0$ if and only if

$$z_1 g''(z_1) + g'(z_1) \leq 0. \tag{8}$$

This relation is equivalent to the condition that $z_1 g'(z_1)$ decreases in z_1 for $z_1 \in [y_1, x_1]$. This is a much weaker condition than the assumption that g is concave. In Example 3 we will demonstrate a case when condition (8) holds but g is convex.

Interchanging the role of the two players, the same concept is applied to player 2. Function g is replaced by its inverse function, which will be denoted by G. Obviously G is also twice continuously differentiable, strictly decreasing and concave if the same properties hold for g. Similarly to the case of player 1, the necessary and sufficient condition for the monotonicity of the equation giving the unique optimal choice for player 2 with all $p_1 > 0$ is that $x_2 G'(x_2)$ decreases in x_2 (or increases) in x_1. Notice that

$$z_1 = G(z_2). \tag{9}$$

By differentiating,

$$G'(z_2) g'(z_1) = 1, \tag{10}$$

that is

$$z_2 G'(z_2) = \frac{g(z_1)}{g'(z_1)}. \tag{11}$$

It is easy to see that this function is increasing if and only if

$$g''(z_1) \leq \frac{g'(z_1)^2}{g(z_1)}. \tag{12}$$

This is also a weaker condition than assuming the concavity of G. Combining the conditions for the two players leads to the following observation. The necessary and sufficient condition for having monotone equations for both players for all $p_1, p_2 > 0$ is the following:

$$g''(z_1) \leq min\{\frac{g'(z_1)^2}{g(z_1)}, -\frac{g'(z_1)}{z_1}\}. \tag{13}$$

The right hand side is positive for all z_1, so this condition allows even small positive values of g''. That is, g might be even slightly convex.

Based on the above described expected payoff maximizing rule the bargaining process can be defined in several different ways. One possible approach is the following. Assume that player 1 makes the move in each odd(even) step, and player 2 makes offer in each even(odd) step. A pair of such steps is considered as a round. It is also assumed that at each step, the offer making player selects his/her optimal alternative. Notice that at each step the players optimize their expected payoffs occuring in that step only, and they do not consider payoffs of the later steps. This approach is realistic in cases when there is a large uncertainty in future payoff values or even in continuing the negotiations. In such cases each player wants to get immediately as much payoff as possible. We mention that in dynamic economic modeling this assumption is commonly accepted (see, for example, Okuguchi and Szidarovszky, 1990). We will next prove that this process is always finite.

Theorem 1. Assume that g is twice continuously differentible, $g'(z_1) < 0$ and relation (13) holds for $z_1 \in [y_1^0, x_1^0]$. Then the above alternating offer procedure is always finite.

Proof. Suppose the process is infinite. Let $(x_1^{(1)}, x_2^{(1)})$, $(x_1^{(2)}, x_2^{(2)}), \cdots$ denote the sequential positions of player 1. At round k, when player 1 makes the next offer, the maximizer of his/her expected payoff must be an interior point between $(x_1^{(k)}, x_2^{(k)})$ and $(y_1^{(k)}, y_2^{(k)})$, such that $h(x_1^{(k+1)}) = 0$. After some algebra and the use of the mean value theorem of derivatives, we have

$$x_1^{(k+1)} = \frac{g'(c_k)(x_1^{(k)} - x_1^{(k+1)})}{p_2 g'(x_1^{(k+1)})}, \tag{14}$$

where $x_1^{(k+1)} \leq c_k \leq x_1^{(k)}$.

Since g' is continuous, it is bounded, say, $0 < c_1 \leq |g'(x)| \leq c_2$, which implies that

$$x_1^{(k+1)} \leq \frac{c_2}{p_2 c_1}(x_1^{(k)} - x_1^{(k+1)}). \tag{15}$$

Denote $c_2/(p_2 c_1)$ by K, then

$$x_1^{(k+1)} \leq \frac{K}{K+1} x_1^{(k)}. \tag{16}$$

This implies that $x_1^{(k+1)}$ tends to zero. This is an obvious contradiction, since for all $k \geq 2$, $x_1^{(k)} \geq y_1^{(1)} > 0$. Thus the proof is complete.

Another bargaining process can be defined in the following way. At the initial step each player finds his/her expected payoff maximizing offer $(x_1^{(1)}, x_2^{(1)})$ and $(y_1^{(1)}, y_2^{(1)})$. If $x_1^{(1)} \leq y_1^{(1)}$, then bargaining terminates and each player receives his/her demanded payoff (or any point of the Pareto frontier between $(x_1^{(1)}, x_2^{(1)})$ and $(y_1^{(1)}, y_2^{(1)})$ can be accepted as the final outcome). Otherwise negotiation continues in the same manner as in step 1, starting from these new offers. It is easy to see that under the conditions of Theorem 1 this process is also finite.

The bargaining processes outlined above require the knowledge of the power indices of both players, which is usually hard to quantify. A one-step solution concept is next introduced, when only the ratio $\alpha = p_2/p_1$ has to be known by the participants. This ratio α can be considered as the relative power of player 2 over player 1. A point (z_1^*, z_2^*) is considered as the solution of the bargaining problem, when it is the optimal choice for both players, starting from the initial offers with some parameters p_1 and p_2 such that $p_2/p_1 = \alpha$. We will next prove that such point exists and is unique.

Theorem 2. Under the conditions of the previous theorem, for all $\alpha > 0$, there is a unique point (z_1^*, z_2^*) on the Pareto frontier between the initial offers (x_1, x_2) and (y_1, y_2), which is the optimal choice for both players with some power indices p_1 and p_2 such that $p_2/p_1 = \alpha$.

Proof. A point (z_1, z_2) is optimal for both players if and only if it is the solution of equations

$$z_1 p_1 g'(z_1) + g(z_1) - x_2 = 0 \tag{17}$$

and

$$z_2 p_2 G'(z_2) + G(z_2) - y_1 = 0. \tag{18}$$

Relations $p_2/p_1 = \alpha$, $z_2 = g(z_1)$, and $z_2 G'(z_2) = g(z_1)/g'(z_1)$ imply that these equations are equivalent to a single equation for the single unknown z_1:

$$\frac{x_2 - g(z_1)}{z_1 g'(z_1)} - \alpha \frac{y_1 - z_1}{\frac{g(z_1)}{g'(z_1)}} = 0. \tag{19}$$

One can easily verify that the left hand side is strictly decreasing, it is positive at y_1 and negative at x_1. Consequently, there is a unique solution.

The monotonicity of equation (19) allows us to use special iteration methods (such as bisection, or secant method) to find the solution (Szidarovszky and Yakowitz,1978).

Using the monotonicity of function h, one may easily verify the following monotonicity properties:

1. For fixed values of α and x_2, z_1 strictly increases in y_1. Similarly, for fixed α and y_1, z_1 strictly decreases in x_2. In other words, better initial position of any player will result in his/her higher payoff at the solution.

2. For fixed values of x_2 and y_1, z_1 strictly decreases in α. It represents the fact that as α increases, the solution moves closer to the position of player 2.

3 Numerical Examples

In this section, three examples are presented. The first one shows a linear case, the second example presents a strictly concave case, and in the third example, a slightly convex g is examined.

Example 1

Select $g(x) = -ax + b$, where $a > 0, b > 0$. Suppose $a = 0.5$ and $b = 1$ are given. The initial positions of player 1 and player 2 are $(1.6, 0.2)$ and

$(0.6, 0.7)$, respectively(see Fig. 2). Assume that the power index of player 1 is 0.4 and that of player 2 is 0.6.

If the two players make concessions alternatively, then the process is finite and the solution depends on the order in which concessions are made. As shown in Tables 1 and 2, the payoff of each player is smaller in the case when he/she makes the first offer assuming that the process does not terminate with disagreement and at each step the offering player selects a point where his/her expected payoff is maximized.

Consider next the one-step solution with $\alpha = 1.5$. Equation(19) has a unique solution $(0.9243, 0.5378)$. Notice that this solution is located between the previous two solutions. Therefore, it can be considered as a compromise solution when no "first concession" is made.

Example 2

Select $g(x) = \sqrt{1-x^2}$ where $0 \leq x \leq 1$ (see Figure 3). Suppose the initial positions are given as $(0.8, 0.6)$, $(0.6, 0.8)$, and the power index is 20 for both players. Tables 3 and 4 give the solutions when the alternating offer strategy is used. Obviously the problem is symmetric. The one-step solution is $(\sqrt{2}/2, \sqrt{2}/2)$ which is also symmetric in the two players.

We will next repeat the computations when the power index of player 2 is doubled. The solution in the alternating process remains the same. By solving equation(19) again, the one-step solution is $(0.68155, 0.7318)$. It is easy to see that the payoff of the second player has been increased with this higher power index.

Example 3

Assume now that $g(x) = (2-x)^2 - 1$ for $0 \leq x \leq 1$ and the initial positions of player 1 and player 2 are (0.1,2.61) and (0.9,0.21), respectively (see Figure 4). Tables 5 and 6 give the solutions when the alternating offer process is used.

Solving equation (19) with $\alpha = 0.5$ leads to the one-step solution (0.46165,1,3665). Notice that function g is strictly convex but it satisfies all conditions of Theorem 2.

Table 1: Result of Example 1 when player 1 makes the first concession

round(k)	$(x_1^{(k)}, x_2^{(k)})$	$(y_1^{(k)}, y_2^{(k)})$
0	$(1.6, 0.2)$	$(0.4, 0.8)$
1	$(1, 0.5)$	$(6/7, 4/7)$
2	$(6/7, 4/7)$	———

Table 2: Result of Example 1 when player 2 makes the first concession

round(k)	$(y_1^{(k)}, y_2^{(k)})$	$(x_1^{(k)}, x_2^{(k)})$
0	$(0.4, 0.8)$	$(1.2, 0.2)$
1	$(6/7, 4/7)$	$(1, 0.5)$
2	$(1, 0.5)$	———

Table 3: Result of Example 2 when player 1 makes the first concession

round(k)	$(x_1^{(k)}, x_2^{(k)})$	$(y_1^{(k)}, y_2^{(k)})$
0	$(0.8, 0.6)$	$(0.6, 0.8)$
1	$(0.6, 0.8)$	———

Table 4: Result of Example 2 when player 2 makes the first concession

round(k)	$(y_1^{(k)}, y_2^{(k)})$	$(x_1^{(k)}, x_2^{(k)})$
0	$(0.6, 0.8)$	$(0.8, 0.6)$
1	$(0.8, 0.6)$	———

Table 5: Result of Example 3 when player 1 makes the first concession

round(k)	$(x_1^{(k)}, x_2^{(k)})$	$(y_1^{(k)}, y_2^{(k)})$
0	$(0.9, 0.21)$	$(0.1, 2.61)$
1	$(0.575, 1.030)$	$(0.358, 1.696)$
2	$(0.375, 1.640)$	$(0.375, 1.640)$

Table 6: Result of Example 3 when player 2 makes the first concession

round(k)	$(y_1^{(k)}, y_2^{(k)})$	$(x_1^{(k)}, x_2^{(k)})$
0	$(0.1, 2.61)$	$(0.9, 0.21)$
1	$(0.358, 1.696)$	$(0.575, 1.030)$
2	$(0.549, 1.107)$	$(0.549, 1.107)$

4 Discussions and Conclusions

In this paper new solution concepts were introduced to solve 2-person conflict situations. A dynamic process was first examined when the players make concessions alternatively based on offer-dependent break-down probabilities. We have also examined a variant of this bargaining process when the players make simultaneous concessions. We proved that these processes are always finite. A one-step solution concept was next introduced, in which the relative power index of player 2 over player 1 is assumed to be given. The solution is the best alternative selection for both players with some power indices such that their ratio is given. The existence of the unique solution was verified in the paper.

Three examples illustrated the methodology. We selected a linear, a concave and a convex function in order to show all possible cases.

The computation of the solution is based on solving a special nonlinear equation. The equation is monotonic, and therefore routine methods are available for obtaining the solutions. The N-person extension of the solution concepts and methods will be the subject of a future paper.

References

[1] Anbar, D. and E. Kalai (1978), A One-shot Bargaining Problem. *International J. of Game Theory*, Vol. 7, pp. 13-15.

[2] Anbarci, N. (1993), Noncooperative Foundations of the Area Monotonic Solution. *Quarterly J. of Econ.*, Vol. 108, pp. 245-258.

[3] Anbarci, N. (1995), Reference Functions and Balanced Concessions in Bargaining. *Canadian J. of Economics*, Vol. 28, pp. 675-682.

[4] Chun, Y. (1988), The Equal-Loss Principle for Bargaining Problems. *Econ. Letters*, Vol. 26, pp. 103-106.

[5] Harsanyi, J. C. and R. Selten (1972), A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information. *Management Science*, Vol. 18, Part II, pp. 80-106.

[6] Howard, J. V. (1992), A Social Choice Rule and Its Implementation in Perfect Equilibrium. *J. of Econ. Theory*, Vol. 56, pp. 142-159.

[7] Kalai, E. and M. Smorodinsky (1975), Other Solutions to Nash's Bargaining Problem. *Econometrica*, Vol. 43, pp. 513-518.

[8] Nash, J. F. (1950), The Bargaining Problem. *Econometrica*, Vol. 28, pp. 155-162.

[9] Nash, J. F. (1953), Two-Person Cooperative Games. *Econometrica*, Vol. 21, pp. 129-140.

[10] Okuguchi, K. and F. Szidarovszky (1990) *The Theory of Oligopoly with Multi-Product Firms.* Springer-Verlag, Berlin/Heidelberg/New York.

[11] Perles, M. A. and M. Maschler (1981), A Super-Additive Solution for Nash's Bargaining Game. *International J. of Game Theory*, Vol. 10, pp. 163-193.

[12] Roth, A. E. (1979) *Axiomatic Models of Bargaining.* Springer-Verlag, Berlin/ Heidelberg/ New York.

[13] Rubinstein, A. (1981), Perfect Equilibrium in a Bargaining Model. *Econometrica*, Vol. 50, pp. 97-110.

[14] Szidarovszky, F. and S. Yakowitz (1978) *Principles and Procedures of Numerical Analysis.* Plenum, New York/London.

[15] Zeuthen, F. (1930), *Problems of Monopoly and Economic Welfare.* Routledge, London.

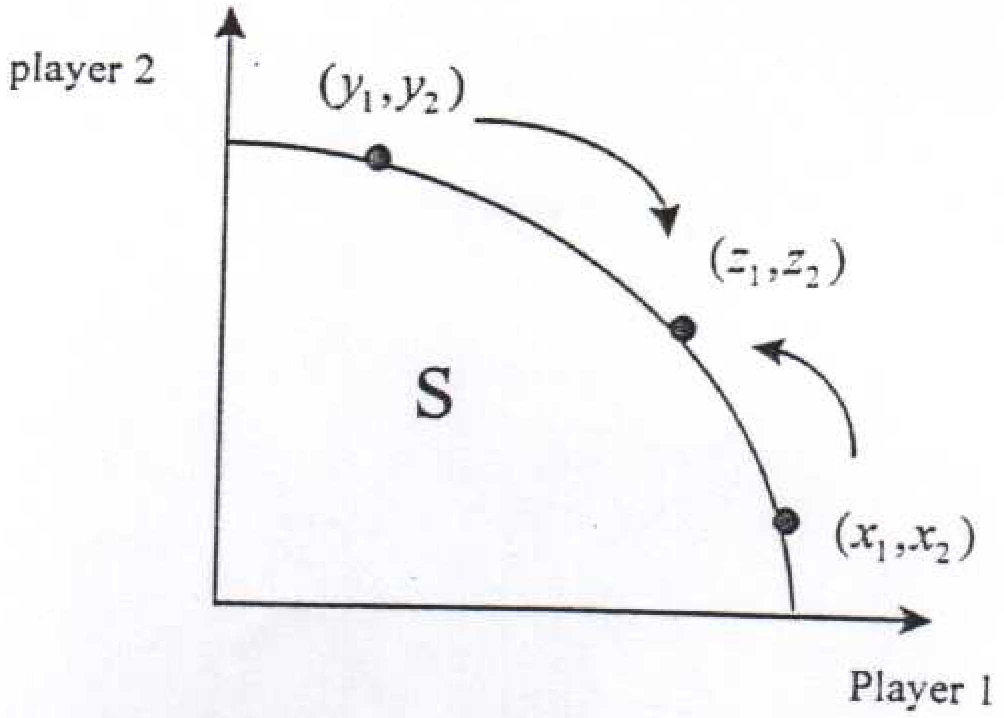


Figure 1: The general 2-person bargaining game

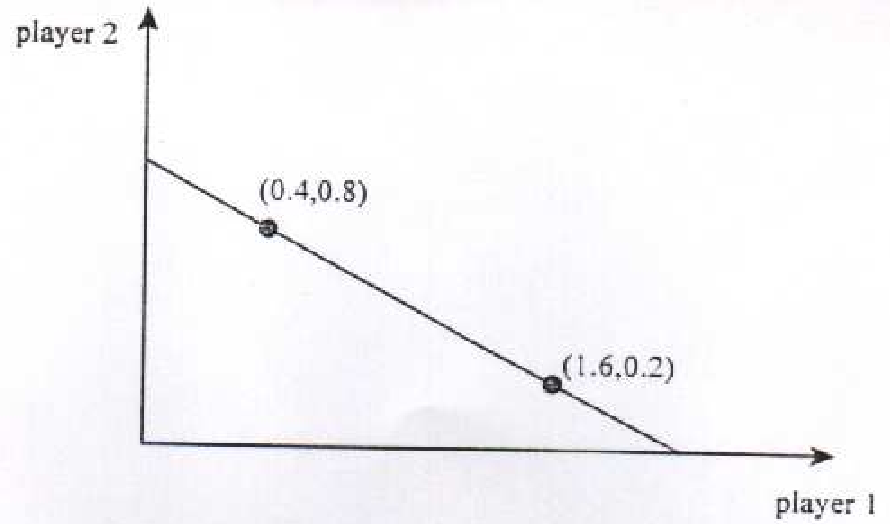


Figure 2: Pareto frontier and initial positions in Example 1

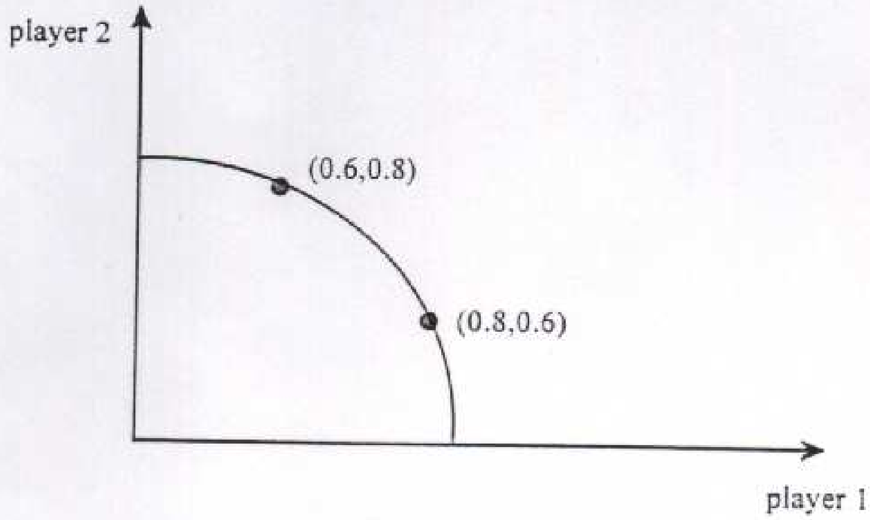


Figure 3: Pareto frontier and initial positions in Example 2

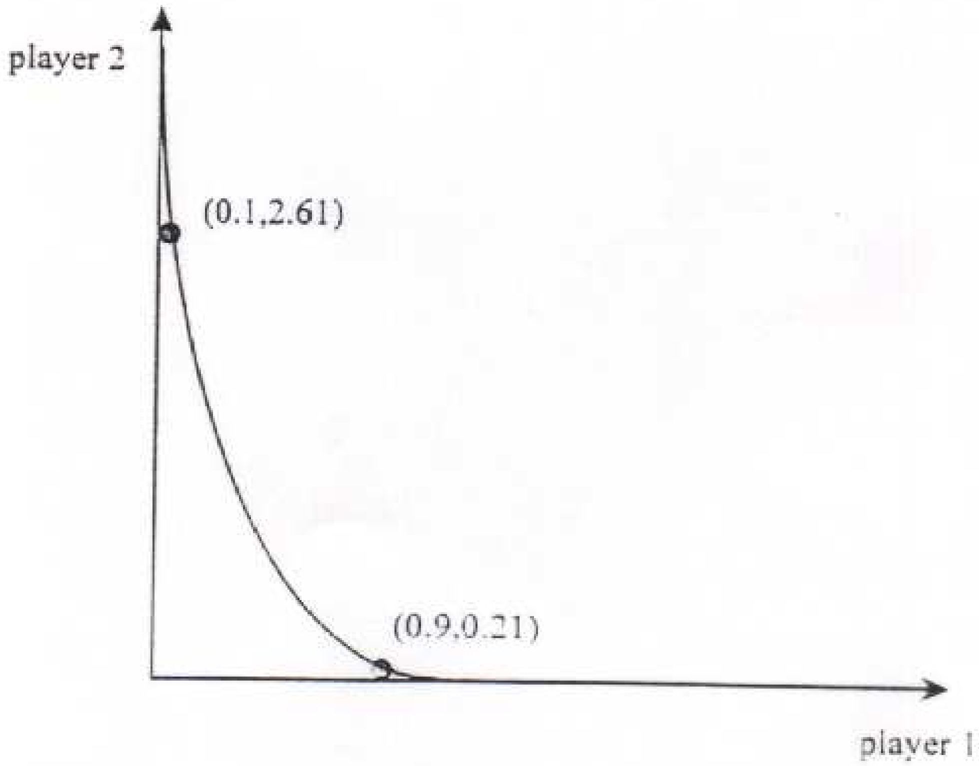


Figure 4: Pareto frontier and initial positions in Example 3

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2000, pp. 17–27.
Submitted: July 3, 2000. Published: December 15, 2000.

ON THE ASYMPTOTIC BEHAVIOR OF DYNAMIC RENT-SEEKING GAMES

Carl Chiarella and Ference Szidarovszky

Abstract. Dynamic rent-seeking games with nonlinear cost functions are analyzed. The local asymptotical stability of the solution is first examined. We show that in the absence of a dominant agent all eigenvalues of the Jacobian are real. Conditions are given for the local asymptotical stability as well as for the local instability of the equilibrium. In the presence of a dominant agent complex eigenvalues are possible. Simple stability conditions are presented for cases when all eigenvalues are real, and the possibility of limit cycles is analyzed in the case of complex eigenvalues.

1991 A.M.S. Subject Classification Codes: 90.

Key Words and Phrases: Asymptotic behavior, Rent-Seeking games, Nash Equilibrium.

1. INTRODUCTION

During the last two decades an increasing attention has been given to the analysis of rent-seeking games. The introductory paper of Tullock [7] has initiated a series of studies on the subject. Decreasing return in rent-seeking technology as well as increasing returns were analyzed in Pérez-Castrillo and Verdier [3], who also examined the possibility of the agent's reaction function to be upward sloping. A systematic approach to establish the existence of an equilibrium has also been offered. Okuguchi [2] has pointed out that rent-seeking games are mathematically equivalent to profit maximizing Cournot oligopolies with hyperbolic price functions. He also proved the existence of a pure symmetric Nash equilibrium. These results have been extended to the nonsymmetric case by Szidarovszky and Okuguchi [4].

The dynamic extension of rent-seeking games has been examined by Szidarovszky and Okuguchi [5] and by Xu and Szidarovszky [8], however the stability of the

C. Chiarella, School of Finance and Economics, University of Technology, Sydney, P.O. Box 123, Broadway, NSW 2007, Australia

F. Szidarovszky, Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona 85721, U.S.A.

E-mail: szidar@sie.arizona.edu

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

equilibrium was proved under certain convexity and monotonicity conditions, which are not always satisfied.

In this paper we will present a systematic approach to examine the local asymptotical stability of the equilibrium under much more general conditions than in the previously cited studies, and in the case of local instability the possibility of limit cycles will be investigated.

2. THE MATHEMATICAL MODEL

Assume that n agents compete for a rent, which will earn a unit profit for the firm which actually will win the rent. Let x_i denote the effort agent i spends in order to win the rent, and let $c_i(x_i)$ be its cost. The probability of winning the rent for agent i is $\frac{x_i}{\sum_{j=1}^{n} x_j}$, therefore its expected profit is given as

$$\varphi_i(x_1, \dots, x_n) = \frac{x_i}{\sum_{j=1}^{n} x_j} - c_i(x_i). \tag{1}$$

Let $X_i = \mathcal{R}_+$ be the set of feasible strategies of agent i, then an n-person noncooperative game $\Gamma = \{n, X_1, \dots, X_n, \varphi_1, \dots, \varphi_n\}$ is formally defined, and is called a *rent-seeking game.*

For each agent i, denote $Q_i = \sum_{j \neq i} x_j$, then for each $Q_i \geq 0$, its best response can be obtained as

$$x_i(Q_i) = \arg\max_{x_i \geq 0} \left\{ \frac{x_i}{x_i + Q_i} - c_i(x_i) \right\}. \tag{2}$$

Assume that

(i) for all i, c_i is twice continuously differentiable.

Assume in addition that $x_i(Q_i) > 0$, then the first and second order conditions can be written as

$$\frac{Q_i}{(x_i(Q_i) + Q_i)^2} - c_i'(x_i(Q_i)) = 0 \tag{3}$$

and

$$\frac{-2Q_i}{(x_i(Q_i) + Q_i)^3} - c_i''(x_i(Q_i)) < 0. \tag{4}$$

Notice that relation (4) also guarantees that the left-hand side of equation (3) is strictly monotonic, therefore $x_i(Q_i)$ is unique.

Let now $(\overline{x}_1, \dots, \overline{x}_n)$ be a positive equilibrium, that satisfies both conditions (3) and (4). That is, with the notation $\overline{Q}_i = \sum_{l \neq i} \overline{x}_l$ and $\overline{s} = \sum_{j=1}^{n} \overline{x}_j$,

$$\overline{Q}_i = c_i'(\overline{x}_i)\overline{s}^2 \tag{5}$$

and

$$c_i''(\overline{x}_i) > \frac{-2\overline{Q}_i}{\overline{s}^3}. \tag{6}$$

Select any arbitrary positive numbers $\overline{x}_1, \dots, \overline{x}_n$, and let $\overline{s} = \sum_{i=1}^{n} \overline{x}_i$ and for all i, $\overline{Q}_i = \sum_{l \neq i} \overline{x}_i$. In addition, select any sequence $\overline{c}_1'', \dots, \overline{c}_n''$ of real numbers that satisfy inequality (6). Then there is a rent-seeking game in which $(\overline{x}_1, \dots, \overline{x}_n)$ is an equilibrium and for all i, $c_i''(\overline{x}_i) = \overline{c}_i''$, since for each i there is a strictly increasing cost function with given first and second derivatives at a single point.

Assume next that the time scale is continuous, and at each time period $t \geq 0$, each agent adjusts its output proportionally to its marginal profit. The resulting dynamic model has the form:

$$\dot{x}_i = k_i \left(\frac{Q_i}{(x_i + Q_i)^2} - c_i'(x_i) \right) \tag{7}$$

where $k_i > 0$ is given for each $i = 1, 2, \dots, n$. Notice that the vector $(\overline{x}_1, \dots, \overline{x}_n)$ is an equilibrium of this dynamic system as the consequence of the first order conditions (5). Assume that the initial vector $\underline{x}(0)$ is selected from a neighborhood of this equilibrium. In the next section the local stability and the dynamic behavior of the solution $\underline{x}(t)$ will be investigated.

3. STABILITY ANALYSIS

In order to analyze the asymptotical behavior of the solution of system (7), the Jacobian has to be first determined. Simple differentiation shows that at the equilibrium

$$\begin{aligned}
\underline{J} &= \frac{1}{\overline{s}^3} \\
&\begin{pmatrix} k_1(-2\overline{Q}_1 - \overline{s}^3 c_1''(\overline{x}_1)) & k_1(\overline{x}_1 - \overline{Q}_1) & \cdots & k_1(\overline{x}_1 - \overline{Q}_1) \\ k_2(\overline{x}_2 - \overline{Q}_2) & k_2(-2\overline{Q}_2 - \overline{s}^3 c_2''(\overline{x}_2) & \cdots & k_2(\overline{x}_2 - \overline{Q}_2) \\ \vdots & \vdots & \ddots & \vdots \\ k_n(\overline{x}_n - \overline{Q}_n) & k_n(\overline{x}_n - \overline{Q}_n) & \cdots & k_n(-2\overline{Q}_n - \overline{s}^3 c_n''(\overline{x}_n)) \end{pmatrix} \\
&= \frac{1}{\overline{s}^3}(\underline{D} + \underline{a}\underline{1}^t)
\end{aligned} \tag{8}$$

with

$$\begin{aligned}
\underline{D} &= \operatorname{diag}(k_1(-\overline{s} - \overline{s}^3 c_1''(\overline{x}_1)), \dots, k_n(-\overline{s} - \overline{s}^3 c_n''(\overline{x}_n))), \\
\underline{a} &= (k_1(\overline{x}_1 - \overline{Q}_1), \dots, k_n(\overline{x}_n - \overline{Q}_n))^T, \quad \text{and} \quad \underline{1}^t = (1, \dots, 1).
\end{aligned} \tag{9}$$

Since the signs of the real parts of the eigenvalues determine the asymptotical behavior of the solution, we might ignore the positive factor $\frac{1}{\overline{s}^3}$. The characteristic polynomial of $\underline{D} + \underline{a}\underline{1}^T$ can be written as

$$\begin{aligned}
\det(\underline{D} - \lambda \underline{I} + \underline{a}\underline{1}^T) &= \det(\underline{D} - \lambda \underline{I}) \det(\underline{I} + (\underline{D} - \lambda \underline{I})^{-1} \underline{a}\underline{1}^T) \\
&= \prod_{i=1}^{n} (k_i(-\overline{s} - \overline{s}^3 c_i''(\overline{x}_i)) - \lambda) \left[1 + \sum_{i=1}^{n} \frac{k_i(\overline{x}_i - \overline{Q}_i)}{k_i(-\overline{s} - \overline{s}^3 c_i''(\overline{x}_i)) - \lambda} \right].
\end{aligned} \tag{10}$$

In order to simplify the notation let $d_i = k_i(-\bar{s} - \bar{s}^3 c_i''(\bar{x}_i))$ and $e_i = k_i(\bar{x}_i - \overline{Q}_i)$ for $i = 1, 2, \ldots, n$, and assume that the different d_i values are $d_1 < d_2 < \cdots < d_r$. Let I_l denote the set of agents with the same d_l value $(l = 1, 2, \ldots, r)$, and let m_l denote the number of agents in set I_l. Then the eigenvalues are the solutions of the equation

$$\prod_{l=1}^{r} (d_l - \lambda)^{m_l} \left[1 + \sum_{l=1}^{r} \frac{\gamma_l}{d_l - \lambda} \right] = 0, \tag{11}$$

where $\gamma_l = \sum_{i \in I_l} e_i$. The eigenvalues can be obtained in the following way. If $m_l > 1$, then $\lambda = d_l$ is an eigenvalue with multiplicity $m_l - 1$, and if $m_l = 1$, then the corresponding factor cancels, so in this case $\lambda = d_l$ is not an eigenvalue ($l = 1, 2, \ldots, r$). In this way we have $m_1 + \cdots + m_r - r = n - r$ eigenvalues (including with multiplicities). The other eigenvalues are the solutions of the nonlinear equation

$$1 + \sum_{l=1}^{r} \frac{\gamma_l}{d_l - \lambda} = 0. \tag{12}$$

Since this is equivalent to a polynomial equation of degree r, it has r real (or complex) roots.

The previous derivation implies the following.

Theorem 1. (i) *Assume that with some l, $m_l \geq 2$ and $d_l > 0$. Then the equilibrium is locally unstable.*

(ii) *Suppose that for all l such that $m_l \geq 2$, $d_l < 0$, then let $\lambda_1, \ldots, \lambda_r$ denote the roots of equation*

$$1 + \sum_{l=1}^{r} \frac{\gamma_l}{d_l - \lambda} = 0. \tag{13}$$

If for at least one i, $Re\,\lambda_i > 0$, then the equilibrium is locally unstable. If for all i, $Re\,\lambda_i < 0$, then the equilibrium is locally asymptotically stable.

We note that in the case when for all roots, $\mathrm{Re}\,\lambda_i \leq 0$ and for at least one root $\mathrm{Re}\,\lambda_i = 0$, no conclusion can be given based on only the eigenvalues, since the system is nonlinear.

In order to apply the above theorem, the locations of the roots of equation (13) should be examined. We will consider two major cases.

Case A. Assume that for all l,

$$\gamma_l \leq 0. \tag{14}$$

Notice that if for some l the value of γ_l is zero, then the corresponding term is missing from the left-hand side of equation (13) without changing its structure. Therefore in the following discussion we may assume that inequality (14) is strict for all l. Introduce

$$g(\lambda) = \sum_{l=1}^{r} \frac{\gamma_l}{d_l - \lambda}. \tag{15}$$

It is easy to see that

$$\lim_{\lambda\to d_l+0} g(\lambda)=\infty,\quad \lim_{\lambda\to d_l-0} g(\lambda)=-\infty,$$
$$\lim_{\lambda\to\pm\infty} g(\lambda)=0,\quad \text{and}\quad g'(\lambda)<0 \tag{16}$$

for all λ, except the poles. The graph of $g(\lambda)$ is illustrated in Figure 1, where we assume that the agents are numbered so that $d_1 < d_2 < \cdots < d_r$.

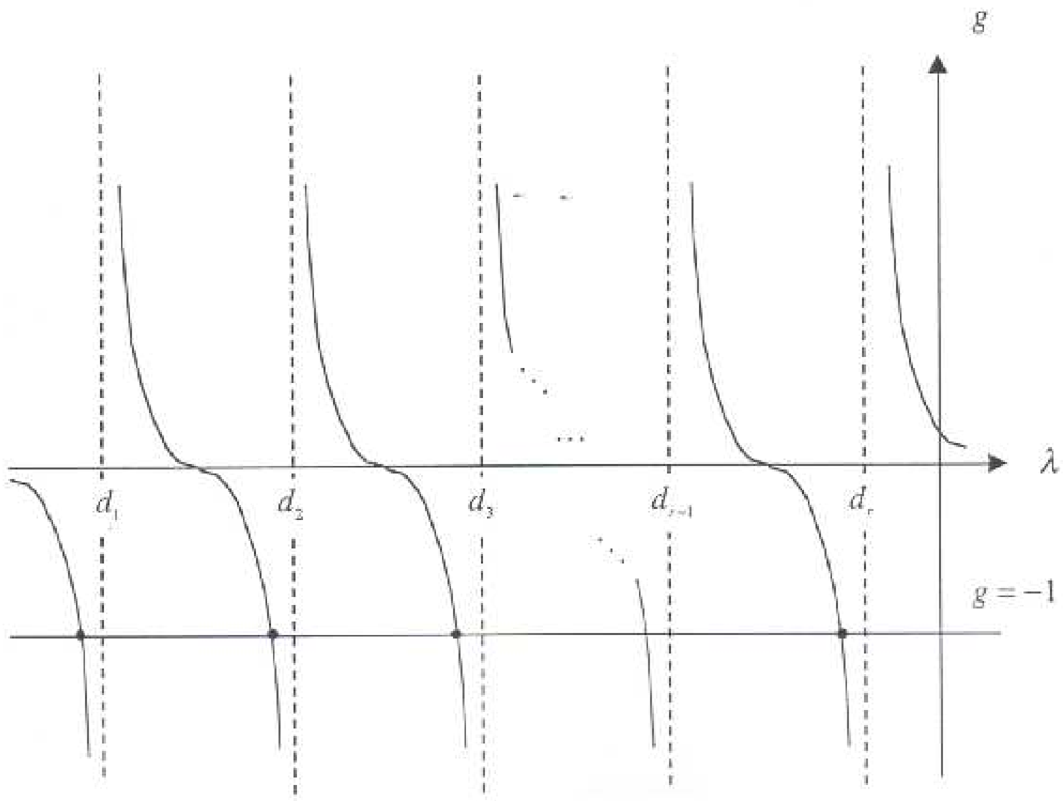


Figure 1. The shape of $g(\lambda)$ in Case A.

Since equation (13) can be written as $g(\lambda) = -1$, we see that all roots are real: one before d_1, and one between each d_i and d_{i+1} $(i = 1, 2, \ldots, r-1)$.

Here we have the following result.

Theorem 2. *Consider the case in which $\gamma_l \le 0$ for all l.*

(i) *If for all l, $d_l \le 0$, then all roots are negative implying the local asymptotical stability of the equilibrium.*

(ii) *Assume that $d_r > 0$ and all $d_i < 0$ $(i = 1, 2, \ldots, r-1)$. If $g(0) < -1$, then the equilibrium is locally asymptotical stable and if $g(0) > -1$, then the equilibrium is locally unstable.*

(iii) *If $d_r > 0$ and $d_{r-1} \ge 0$, then the equilibrium is locally unstable regardless of the signs of the other d_i values $(i = 1, 2, \ldots, r-2)$.*

We note again that in the case when the largest root is zero, no conclusion can be given, since the system is nonlinear.

Case B. For some l^*,

$$\gamma_{l^*} > 0. \tag{17}$$

We first show that this relation can hold for only one value of l^*. Notice first that there is at most one dominant agent p such that $\overline{x}_p > \overline{Q}_p$. If set I_t does not contain this agent, then

$$\gamma_t = \sum_{i \in I_t} k_i(\overline{x}_i - \overline{Q}_i) \leq 0,$$

therefore agent p must belong to set I_{l^*}. We show next that $d_{l^*} < 0$ necessarily for the set satisfying inequality (17). The second order conditions (6) imply that for the dominant agent p,

$$-\overline{s} - \overline{s}^3 c_p''(\overline{x}_p) < -\overline{s} + 2\overline{Q}_p = \overline{Q}_p - \overline{x}_p,$$

and by multiplying this inequality by k_p we have

$$d_{l^*} = k_p(-\overline{s} - \overline{s}^3 c_p''(\overline{x}_p)) < k_p(\overline{Q}_p - \overline{x}_p) = -e_p.$$

Since for all other agents $i \in I_{l^*}$, $-e_i \geq 0$, we see that

$$d_{l^*} < -\sum_{i \in I_{l^*}} e_i = -\gamma_{l^*} < 0. \tag{18}$$

For the sake of simplicity assume that $l^* = r$. We keep the assumption that $d_1 < d_2 < \cdots < d_{r-1}$ and note that d_r is not necessarily the largest among the d_l values. In this case function g satisfies the following properties:

$$\begin{aligned} &\lim_{\lambda \to d_l + 0} g(\lambda) = \infty, \ \lim_{\lambda \to d_l - 0} g(\lambda) = -\infty \quad (l = 1, 2, \ldots, r-1), \\ &\lim_{\lambda \to d_r + 0} g(\lambda) = -\infty, \ \lim_{\lambda \to d_r - 0} g(\lambda) = \infty, \end{aligned} \tag{19}$$

and

$$\lim_{\lambda \to \pm\infty} g(\lambda) = 0.$$

Notice that no condition can be given on the sign of the derivative $g'(\lambda)$ in this case. The shape of function g is illustrated in Figure 2, where d_r is either the largest, or the smallest among the d_r values, or it is in the middle.

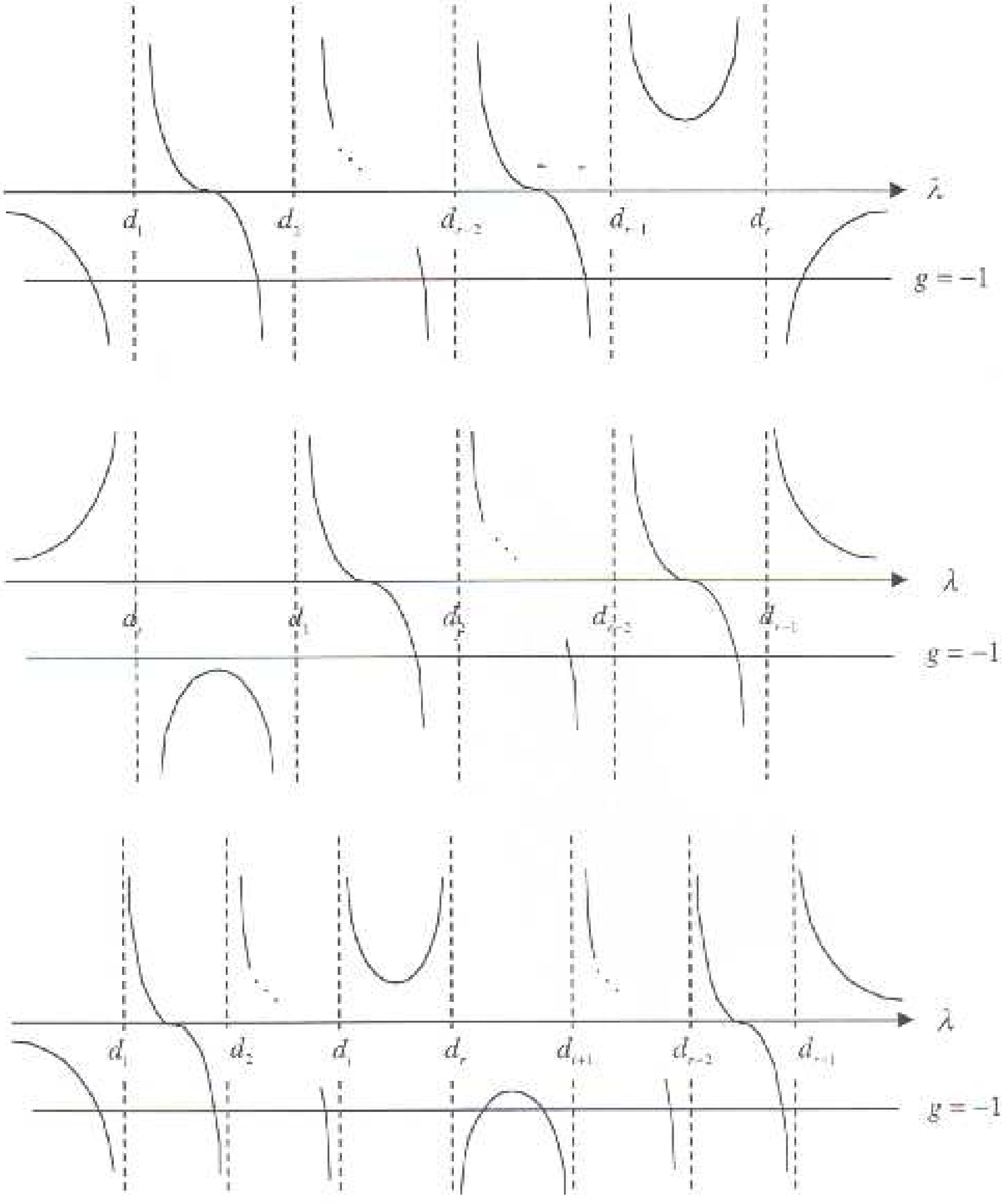


Figure 2. The shape of $g(\lambda)$ in Case B.

The above discussion implies the following

Theorem 3. *Consider the case in which $\gamma_{l^*} > 0$ for some l^*.*

(i) *Assume first that $d_r > d_l$ $(l = 1, 2, \dots, r-1)$. If $d_r \geq 0$, then the equilibrium is unstable. If $d_r < 0$, then in the case of $g(0) > -1$ the equilibrium is locally asymptotically stable, and if $g(0) < -1$, the equilibrium is locally unstable.*

(ii) *Assume now that $d_r < d_l$ $(l = 1, 2, \dots, r-1)$ and there are real roots between d_r and d_1. If $d_{r-2} \geq 0$, then the equilibrium is unstable. If $d_{r-2} < 0 < d_{r-1}$, then in the case of $g(0) < -1$ the equilibrium is locally asymptotically stable, and if $g(0) > -1$, the equilibrium is locally unstable. If $d_{r-1} \leq 0$, then the equilibrium is locally asypmtotically stable.*

(iii) *Assume next that $d_{r-2} < d_r < d_{r-1}$, and there are real roots between d_{r-2} and d_{r-1}. If $d_{r-2} \geq 0$, then the equilibrium is locally unstable. If $d_{r-2} < 0 < d_{r-1}$, then in the case when the roots between d_{r-2} and d_{r-1} are negative, the equilibrium is locally asymptotically stable, and if at least one root between d_{r-2} and d_{r-1} is positive, the equilibrium is locally unstable. If $d_{r-1} \leq 0$, then the equilibrium is locally asymptotically stable.*

(iv) *Assume finally that $d_{i-1} < d_r < d_i$ with some $i \leq r-2$, and there are real roots between d_{i-1} and d_i. If $d_{r-2} \geq 0$, then the equilibrium is locally unstable. If $d_{r-2} < 0 < d_{r-1}$ then in the case of $g(0) < -1$ the equilibrium is locally asymptotically stable, and if $g(0) > -1$, the equilibrium is locally unstable. If $d_{r-1} \leq 0$, then the equilibrium is locally asymptotically stable.*

Notice that the theorem does not apply for cases when the largest real root is zero or there are no roots in the intervals indicated. In every case there are always $(r-2)$ real roots, therefore either all roots are real, or there is exactly one complex conjugate pair of roots. The next example shows that both possibilities might occur.

Example 1. Consider the special case when $r = 2$, $k_1 = k_2 = 1$, $d_2 = -1$, $d_1 = 0$. Let $2\overline{x}_2 - \overline{s} = \gamma > 0$ be a parameter. From equation (18) we must have $\gamma < 1$. Furthermore

$$2\overline{s}_1 - m_1\overline{s} = 2(\overline{s} - \overline{x}_2) - m_1\overline{s} = -(m_1 - 2)\overline{s} - 2\overline{x}_2 = -(m_1 - 1)\overline{s} - \gamma.$$

By introducing the notation $\alpha = (m_1 - 1)\overline{s}$ we see that α can take on any nonnegative value. Then equation (13) has the form

$$\frac{-\alpha - \gamma}{0 - \lambda} + \frac{\gamma}{-1 - \lambda} = -1,$$

which simplies as

$$\lambda^2 + \lambda(1 + \alpha) + (\alpha + \gamma) = 0.$$

The discriminant is

$$D = (1 + \alpha)^2 - 4(\alpha + \gamma) = (1 - \alpha)^2 - 4\gamma$$

showing that the cases of both real and complex roots are possible.

1. The Possibility of Limit Cycles.

In the case of locally unstable equilibrium we can demonstrate the existence of limit cycle motion by applying the Hopf bifurcation theorem (see, for example, Guckenheimer and Holmes, 1983). We need to determine the region in the parameter space where pure complex eigenvalues exist. Letting $\lambda = i\alpha$, equation (13) has the form

$$\sum_{l=1}^{r-1} \frac{\gamma_l}{d_l - i\alpha} + \frac{\gamma_r}{d_r - i\alpha} = -1, \tag{20}$$

where $\gamma_l < 0$ $(i = 1, 2, \dots, r-1)$, and $\gamma_r > 0$.

By equating the real and imaginary parts we have

$$\sum_{l=1}^{r-1} \frac{\gamma_l d_l}{d_l^2 + \alpha^2} + \frac{\gamma_r d_r}{d_r^2 + \alpha^2} = -1$$

and

$$\sum_{l=1}^{r-1} \frac{\gamma_l i\alpha}{d_l^2 + \alpha^2} + \frac{\gamma_r i\alpha}{d_r^2 + \alpha^2} = 0. \tag{21}$$

Since we assume that $\alpha \neq 0$, the second equation implies that

$$\gamma_r = -(d_r^2 + \alpha^2) \sum_{l=1}^{r-1} \frac{\gamma_l}{d_l^2 + \alpha^2} \tag{22}$$

and combining this relation with the first equation of (21) leads to a simple equation for α^2:

$$\sum_{l=1}^{r-1} \frac{\gamma_l (d_l - d_r)}{d_l^2 + \alpha^2} = -1. \tag{23}$$

Notice that this is a polynomial equation of degree $r-1$ for α^2, which can be solved by routine methods (see for example, Szidarovszky and Yakowitz [6]). Notice that in case (i) of Theorem 3, the left-hand side is always positive, so no real root exists. This observation coincides with the corresponding results when we say that all roots of equation (13) are real, so no pure complex root exists. The existence of real roots of equation (23) depends on the signs and the orders of magnitudes of the differences $d_l - d_r$. The following example illustrates the possibility of real roots and hence the possibility of limit cycles.

Example 2. Consider now the case of $r = 3$, $\gamma_1 = -a$, $\gamma_2 = -b$ (where a and b are positive parameters), $\gamma_3 = \gamma > 0$ is the bifurcation variable, and $d_1 = -2$, $d_2 = 1$, $d_3 = -1$. Then equation (20) has the form

$$\frac{-a}{-2-\lambda} + \frac{-b}{1-\lambda} + \frac{\gamma}{-1-\lambda} = -1. \tag{24}$$

If $\lambda = i\alpha$, then

$$\frac{-a}{-2-i\alpha} + \frac{-b}{1-i\alpha} + \frac{\gamma}{-1-i\alpha} = -1.$$

Equating the real and imaginary parts,

$$\frac{2a}{4+\alpha^2} + \frac{-b}{1+\alpha^2} + \frac{-\gamma}{1+\alpha^2} = -1$$

and

$$\frac{-ai\alpha}{4+\alpha^2} + \frac{-bi\alpha}{1+\alpha^2} + \frac{\gamma i\alpha}{1+\alpha^2} = 0.$$

Notice that we are interested in the case of $\alpha \neq 0$. So solving both equations for $\frac{\gamma}{1+\alpha^2}$, we get

$$\frac{\gamma}{1+\alpha^2} = \frac{2a}{4+\alpha^2} - \frac{b}{1+\alpha^2} + 1 = \frac{a}{4+\alpha^2} + \frac{b}{1+\alpha^2}. \tag{25}$$

A fourth order equation is obtained for α:

$$\alpha^4 + \alpha^2(5 + a - 2b) + (4 + a - 8b) = 0. \tag{26}$$

Notice that this equation will have positive roots if $4 + a - 8b < 0$. In order to satisfy relation (18) we have to guarantee that $\gamma < 1$, that is,

$$\frac{\gamma}{1+\alpha^2} = \frac{a}{4+\alpha^2} + \frac{b}{1+\alpha^2} < \frac{1}{1+\alpha^2},$$

which can be rewritten as

$$\alpha^2(a + b - 1) < 4 - a - 4b.$$

For example, by selecting $a = 0.15$, $b = 0.75$, this inequality is satisfied for all $\alpha^2 \geq 0$, furthermore from equation (26) we have

$$\alpha^2 \approx 0.451 \quad \text{and} \quad \alpha^2 \approx -4.101,$$

from which we see that the real roots are

$$\alpha_{1,2} \approx \pm 0.672.$$

Substituting these values into equation (25) we see that $\gamma \approx 0.799$. Differentiating equation (24) with respect to γ leads to the derivative

$$\frac{d\lambda}{d\gamma} = \frac{\frac{1}{1+\lambda}}{\frac{-a}{(-2-\lambda)^2} + \frac{-b}{(1-\lambda)^2} + \frac{\gamma}{(1+\lambda)^2}},$$

which has the approximating value $(0.468 + 0.719i)$ at $\lambda = i\alpha$ with the above values of α and γ. Hence the real part is nonzero showing the existence of a limit cycle.

CONCLUSIONS

In this paper nonlinear rent-seeking games were analyzed where the cost functions were assumed to be nonlinear. A dynamic adjustment process in which each agent adjusts its output proportionally to its marginal profit was examined. The local asymptotical stability of the equilibrium was first examined. In the absence of a dominant agent it turns out that all eigenvalues are real. Conditions for the local asymptotical stability and also for the local instability of the equilibrium were then derived. In the presence of a dominant agent complex eigenvalues are also possible. Simple conditions were presented for the stability and also for the instability of the equilibrium in the case of real eigenvalues. If complex eigenvalues are possible, then we have shown the possibility of limit cycles via the use of the Hopf bifurcation theorem.

REFERENCES

1. Guckenheimer, J. and Holmes, P.: *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.
2. Okuguchi, K.: Decreasing Return and Existence of Nash Equilibrium in Rent-Seeking Games, *Mimeo*, Department of Economics, Nanzan University, Nagoya, Japan, 1995.
3. Pérez-Castrillo, J.D. and Verdier, T.: A General Analysis of Rent-Seeking Games, *Public Choice* **73** (1992), 335–350.
4. Szidarovszky, F. and Okuguchi, K.: On the Existence and Uniqueness of Pure Nash Equilibrium in Rent-Seeking Games, *Games and Econ. Behavior* **18** (1997), 135–140.
5. Szidarovszky, F. and Okuguchi, K.: On a Dynamic Rent-Seeking Game, *Mimeo*, Department of Economics, Nanzan University, Nagoya, Japan, 1997.
6. Szidarovszky, F. and Yakowitz, S.: *Principles and Procedures of Numerical Analysis*, Plenum Press, New York, 1978.
7. Tullock, G.: Efficient Rent-Seeking. In: *Toward a Theory of Rent-Seeking Society* (J.M. Buchanan, R.D. Tollison and G. Tullock, eds.), Texas A & M Press, College Station, 1980.
8. Xu, L. and Szidarovszky, F.: The Stability of Dynamic Rent-Seeking Games, *International Game Theory Review* **1**, No. 1 (1999), 87–102.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2000, pp. 28–43.
Submitted: July 13, 2000. Published: December 15, 2000.

TENDER SELECTION PROCEDURE IN THE POWER GENERATING INDUSTRY

Sándor Molnár, Ferenc Szidarovszky and Tibor Takács

Abstract. The paper presents a method evaluating tenders to bring on-line new power generating capacities. However, this methodology can be applied to all kind of multiobjective evaluation problems.

1991 A.M.S. Subject Classification Codes: 90.

Key Words and Phrases: Probability, density function, distribution.

In 1997, the Hungarian Power Companies (HPC) issued two tenders for installing new capacities in the period of 2002–2006: one for big capacities (200 MW<), and one for smaller ones. HPC is responsible for the electricity supply of the country. It buys power from plants

and sells it to distributors. Bid evaluations in Hungary are governed by the Electric Power Act passed in 1994. This act basically requires the HPC to ensure that any future power acquisitions abide by the principle of least cost. Of course, there is not a single cost indicator according to which bids could unambiguously be ranked. Costs can be divided into fixed and variable costs (expected energy fee and capacity charge), and also these two types of costs can be further broken down. The planned sites of the investments influence the costs, too, since new plants have to be connected into the national power transmission grid. In the evaluation process, other criteria have to be taken into consideration, too. One of these criteria is the environmental impacts of the production. In case of power plants, the amount of air pollutants have to be taken into account. In the past decades, Hungary joined most of the international conventions and agreements regulating the emission levels of different air pollutants. We mention, for example the UN Framework Convention of Climate Change regularing primarily the CO_2 emissions of anthropogenic origin. Since one third of the overall anthropogenic CO_2 emission comes from power plants, the efficiency and input fuel structure have to be taken into consideration, too. The

S. Molnár, Systems and Control Laboratory, Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary.

F. Szidarovszky, Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona 85721, U.S.A.

Tibor Takács, Systemexpert Consulting Ltd., Hungary.

level of other pollutants (SO_2, NO_X, CO, particulate) are also among the evaluation criteria.

The methodology described below is a decision supporting tool for policymakers. A lot of factors influence the electrical sector. The two main decisive factors are the development of demand on electrical energy, and the price escalation of fuels. Therefore, methodology has to take into account different paths of these variables. In the evaluation process we propose to consider three variants for both factors. The demand will be determined basically by the economic development in the next 10–20 years. After a deep recession in the late 80's and early 90's, the Hungarian economy has begun to recover, and 4–5% GDP growth is predicted in the next years. Certainly, an overall increase of efficiency of electricity consumption is expected as well, which will compensate in part the demand increase. Of course, the forecase of fuel prices is also very uncertain even for short-term.

1. INTRODUCTION

The evaluation procedure for tenders for the establishment of new electric power generating capacities is a very complex task. The presence of several interest groups (share holders), many evaluation criteria, and the dynamism of the need for electric power makes this problem challenging. This particular problem has been raised in Hungary, where the privatization of all fields of the economy is one of the most important trends in adjusting the Hungarian economy to the Western European community.

The selection of the most appropriate tenders and partners cannot be considered as a simple multiobjective programming problem because of the presence of multiple decision makers. Since each interest group's decision is based on several evaluation criteria, and the criteria are conflicting to each other, the problem cannot be approached by classical game theoretical methodology either. Therefore a hybrid methodology had to be developed in order to supply the final decision makers with a satisfactory solution.

This paper, after a brief methodology review, will describe the basics of the methodology and their details that was suggested by the authors of this paper to the leadership of the Hungarian Power Companies as well as to the Ministry of Economic Affairs of Hungary.

The presence of multiple evaluation criteria suggests the use of certain concepts and methodology of multicriteria decision making. The fundamentals of this topic are given for example in Szidarovszky et al. [1]. Depending on the preference informations of the decision makers different methods are available. To name a few, sequential optimization, the ε-constraint method, weighting, distance-, and direction-based methods can be used. Multicriteria problems are sometimes solved by using multiattribute utility theory (Keeney and Raiffa [3]), however not all preference orders can be characterized by single-valued utility functions. For example, the lexicographic order is not equivalent to any order generated by a real-valued function (Fishburn [4]). Decision problems with more than one decision makers are usually modeled by using the methodology of noncooperative- or cooperative games, including group decision making and conflict resolution. The game theoretical aspects are summarized for example in Selten [5], or in Myerson [6]. If the theoretical groups are considered as the decision makers, then since each of them has several objective functions, game theoretical methods cannot be used directly.

Recently Szidarovszky and Eskandari [7] have introduced a general framework to examine games with multiple payoffs. Future demands and economic conditions cannot be predicted exactly, therefore the uncertainty of these factors has to be incorporated into the methodology. In the decision science literature stochastization or fuzzification is used to handle such uncertainty. Szidarovszky et al. [1] contains the basics of both approaches.

The diversity of the methodology in each of the above-mentioned topics offers a large variety of particular methods to be adapted and combined to the tender evaluation process. In the next sections the major steps and the suggested methodology will be outlined.

2. STEPS OF THE PROCESS

For each interest group the evaluation process should be done separately, and it was suggested to follow the following pattern.

Step 1. A list of possible economic scenarios is constructed which are not controlled by the electric power suppliers and have significant effects on the outcome of each tender (e.g., fuel price, economic indices, regulation variants, etc.). Let $k = 1, 2, \dots, N$ denote these scenarios. For different interest groups the selected scenarios may be different.

One possibility is given below.

In the analysis three assumptions will be considered on fuel prices: low, medium, and high increase. Also, three different rates of demand expansions (low, medium, high) will be considered. The nine possible combinations of these assumptions will give the basic assumptions:

1: low fuel price increase and low demand expansions ($F_L \cap D_L$)
2: low fuel price increase and medium demand expansions ($F_L \cap D_M$)
3: low fuel price increase and high demand expansions ($F_L \cap D_H$)
4: medium fuel price increase and low demand expansions ($F_M \cap D_L$)
5: medium fuel price increase and medium demand expansions ($F_M \cap D_M$)
6: medium fuel price increase and high demand expansions ($F_M \cap D_H$)
7: high fuel price increase and low demand expansions ($F_H \cap D_L$)
8: high fuel price increase and medium demand expansions ($F_H \cap D_M$)
9: high fuel price increase and high demand expansions ($F_H \cap D_H$)

These scenarios will be then assumed in evaluating the tenders for estabilishing new power generating capacities. If we do not consider unchanged, or decreasing fuel prices and/or demands, then it will be assumed that the above nine scenarios cover all possibilities. Therefore we propose to select $N = 9$.

Step 2. The occurrence probability p_k for each scenario is then estimated.

The estimation of the individual probabilities

$$p(F_i) \quad \text{and} \quad p(D_j)$$

for $i, j = L, M, H$ are possible by forecasting economic indices that mostly effect these quantities.

Let $x_1, x_2, \dots, x_n$ denote the indices and assume that fuel price and demand expansions can be obtained in some function relations

$$fp = f(x_1, \dots, x_n) \tag{2.1}$$

$$de = g(x_1, \dots, x_n). \tag{2.2}$$

First of all, these functions have to be estimated. Using historical data, nonlinear (or as special case linear) least squares method is suggested. Let

$$x_1^{(t)}, \ldots, x_n^{(t)}, fp^{(t)}, de^{(t)}$$

denote historical data set with $t = 1, 2, \ldots, T$. The function types can be guessed by graphing the $(x_1^{(t)}, fp^{(t)}), \ldots, (x_n^{(t)}, fp^{(t)})$ as well as the $(x_1^{(t)}, de^{(t)}), \ldots, (x_n^{(t)}, de^{(t)})$ points. These one-dimensional graphs represent "fp" and "de" as functions of the individual economic indices. In the computer analysis several function types will be examined and the functions with best least squares fit will be accepted. Additive and multiplicative expressions of the single-dimensional $x_i \to fp$ and $x_i \to de$ functions will be first examined. If no satisfactory fit can be established, then multivariable polynomial forms will be considered. The details of linear and nonlinear least squares analysis can be found in all elementary textbooks.

After functions f and g are determined, the economic indices have to be estimated. The result of the forecasting procedure is a probability distribution for each index. Depending on the forecasting methodology, either a discrete or a continuous distribution is provided. Let $\tilde{x}_1, \ldots, \tilde{x}_n$ denote these random variables. Based on their distributions, the probability distributions of "fp" and "de" can be obtained in the following way.

If the distributions of $\tilde{x}_i$ are discrete, then the process is as follows. Let x_{i1}, x_{i2}, $\ldots$, x_{ir_i} denote the discrete values and $p_{i1}, p_{i2}, \ldots, p_{ir_i}$ the occurring probabilities. Then we accept the values

$$fp(j_1, j_2, \ldots, j_n) = f(x_{1j_1}, \ldots, x_{nj_n}) \tag{2.3}$$

and

$$de(j_1, j_2, \ldots, j_n) = g(x_{1j_1}, \ldots, x_{nj_n}) \tag{2.4}$$

with occurring probabilities $p_{1j_1} \cdot p_{2j_2} \cdot \ldots \cdot p_{nj_n}$. Notice that here we used the independence of the economic indices. This necessary independence has to be taken care by the careful selection of the indices. One possible way of doing it is the use of factor analysis. The details of this methodology can also be found in most advanced statistical textbooks, so the details are omitted.

Let next $[a_L, b_L]$, $[a_M, b_M]$, $[a_H, b_H]$ denote the low, medium, and high fuel-price intervals, and let $[A_L, B_L]$, $[A_M, B_M]$, $[A_H, B_H]$ be the low, medium, and high demand expansion intervals. For $\alpha = L, M, H$ and $\beta = L, M, H$, the probability of α fuel-price with β demand expansion can be computed as

$$P(\alpha, \beta) = \sum_{j_1} \sum_{j_2} \cdots \sum_{j_n} p_{1j_1} p_{2j_2} \ldots p_{nj_n}, \tag{2.5}$$

where summation is made for all $(j_1, \ldots, j_n)$ such that

$$fp(j_1, \ldots, j_n) \in [a_\alpha, b_\alpha]$$

and

$$de(j_1, \ldots, j_n) \in [A_\beta, B_\beta].$$

In the case of continuous disributions on the economic indices the distribution function of "fp" and "de" is very difficult to get, since an n-dimensional integral has to be computed. Therefore we suggest stochastic simulation in the following way. Simulate random values of $x_1, \ldots, x_n$, and with each n-tuple, compute the corresponding values of "fp" and "de" by using the already computed f and g functions. If this simulation is repeated N^*-times, then $N^*(fp, de)$ pairs are obtained. With a simple counter in the program it is easy to see how many times the (fp, de) pair belongs to category (α, β) for $\alpha, \beta = L, M, H$. Let $N^*(\alpha, \beta)$ denote these counted values. Then the requested probabilities can be obtained by the relative frequencies:

$$P(\alpha, \beta) \approx \frac{N^*(\alpha, \beta)}{N^*}. \tag{2.6}$$

The final outcome of this step is the sequence of the probability values $P(\alpha, \beta)$ representing the occurring probabilities for each scenario. These probability values are then ordered as $P_1, P_2, \ldots, P_9$ depending which (α, β) pair corresponds to any given scenario.

Step 3. The list of evaluation criteria is next constructed. They may differ from an interest group to another. The goodness of the tenders are measured by these criteria.

The construction of the list of criteria will be done in three major stages.

In the first stage we have to think over the main reasons why the tenders are needed. In other words, we have to construct a list of goals and conditions the tender should satisfy. These goals and conditions are not necessarily quantifiable. For example goal such as "satisfying energy need" or condition to be "environment friendly" should be formulated only verbally. This list simply should serve as a systematic frame in the beginning of the process.

In the second stage we have to divide each goal and condition into verbal subcategories. For example, the "environment friendly" condition might be divided into groups such as air-quality, ground-water quality, noise, etc.

The third stage formulates the measurable criteria, by lising those quantifiable measures that can be used for concrete evaluation of the subcategories. For example, in "air-quality" we might use the CO_2, SO_2, NO_X, etc. content. The corresponding % numbers can be finally used for evaluation. A similar approach is then suggested for each goal and condition.

In practice, the following procedure seems reasonable. Representatives of the different stake-holders (or interest groups) should have the first meeting to determine the list required in the first stage. For each goal and condition we have to identify the group being mainly interested in. After this primary list is completed, then in the second stage (maybe in a later time) the different interest groups should hold separate meetings to construct and get agreements on the further subdivision of the goals and conditions. After an agreement has been reached, the third stage can start. We advise to have at least a couple of days before the third stage starts. The reasons for this break is to give a chance to the participants to think over the results of the previous stages, to have second thoughts on these lists. If any changes or additions are suggested, they have to be incorporated into the final lists. The third stage will need separate meetings for each interest group. They have to go through the lists of the second stage systematically and for each item they have to suggest measures to be used to evaluate the goodness of each tender with respect to these measures. After each interest group has finalized its list, the meetings have to be finished. After two–three days it is advised that the coordinator makes contact (for example in the way of conferences calls) to the leaders of each group in order to check if any new idea or suggestion came up. If there is no suggestion, then the final list of objectives should be considered as the final outcome of the Step 3, otherwise the list should be modified, and the modified list has to be considered final. Let S denote the number of objectives in the final list. Finally we note that a very similar procedure has been successfully applied for finding the best forestry treatment in a Northern Arizona region (see, Eskandari et al. [8]).

Step 4. In constructing the payoff matrices for each (in our case, nine) scenarios, we have to consider two possibilities.

If an objective is completely quantifyable and certain, then the actual data should be put into the payoff matrix. In the case of an uncertain data, expected values should be used in the table. A common method to do so (in absence of a sequence of data in cases of uncertainty) is the following. Ask experts about a most probable value, and the smallest and largest possibilities for that value. Let $x_{\min}$, $x_{\max}$ and x_{prob} denote these values. Then a triangular distribution can be assumed as shown in Figure 1.

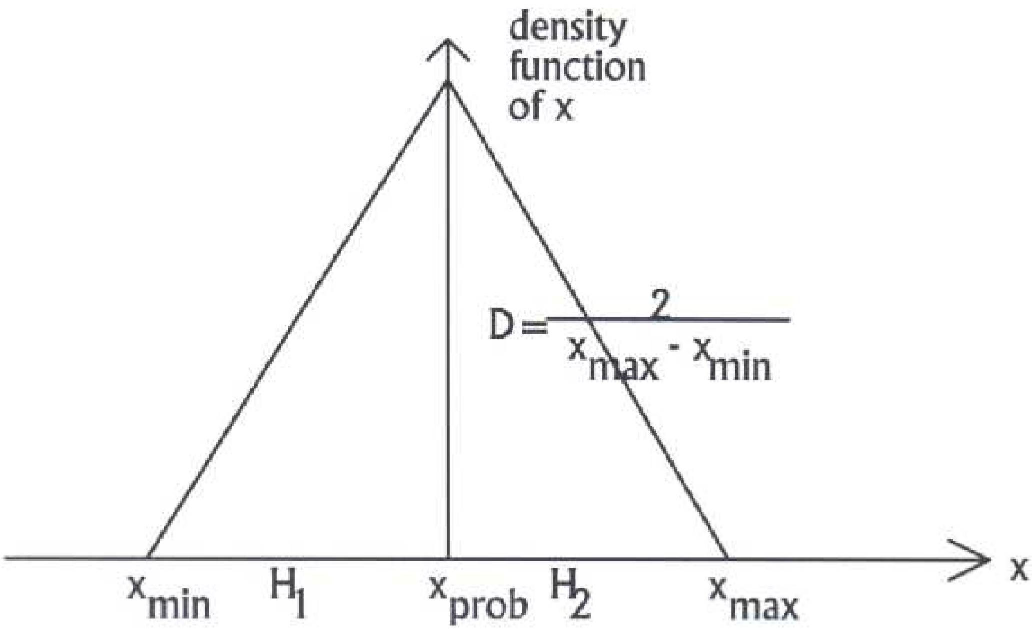


Figure 1. Triangular distribution for uncertain data

The value of the density function at x_{prob} is unknown. Let D denote this value. Then it can be obtained by using the fact that the integral of the density function has to be one. Therefore, using the area of triangles,

$$\frac{(x_{\max}-x_{\min})D}{2}=1$$

implying that

$$D=\frac{2}{x_{\max}-x_{\min}}. \tag{2.7}$$

So the density function can be written as

$$f(x)=\begin{cases}(x-x_{\min})\frac{D}{H_1} & \text{if } x_{\min}\le x\le x_{\text{prob}}\\ (x_{\max}-x)\frac{D}{H_2} & \text{if } x_{\text{prob}}\le x\le x_{\max}\\ 0 & \text{otherwise}\end{cases} \tag{2.8}$$

where $H_1 = x_{\text{prob}} - x_{\min}$ and $H_2 = x_{\max} - x_{\text{prob}}$. The expected value can be obtained by simple integration:

$$\begin{aligned}\overline{X} &= \int_{x_{\min}}^{x_{\text{prob}}}(x^2-x\cdot x_{\min})\frac{D}{H_1}dx+\int_{x_{\text{prob}}}^{x_{\max}}(x\cdot x_{\max}-x^2)\frac{D}{H_2}dx \\ &=\frac{D}{H_1}\left[\frac{x^3}{3}-\frac{x^2}{2}x_{\min}\right]_{x_{\min}}^{x_{\text{prob}}}+\frac{D}{H_2}\left[\frac{x^2}{2}x_{\max}-\frac{x^3}{3}\right]_{x_{\text{prob}}}^{x_{\max}} \\ &=\frac{D}{H_1}\left[\frac{x_{\text{prob}}^3-x_{\min}^3}{3}-x_{\min}\frac{x_{\text{prob}}^2-x_{\min}^2}{2}\right] \\ &\quad+\frac{D}{H_2}\left[x_{\max}\frac{x_{\max}^2-x_{\text{prob}}^2}{2}-\frac{x_{\max}^3-x_{\text{prob}}^3}{3}\right] \\ &=\frac{D}{H_1}\left[\frac{1}{3}x_{\text{prob}}^3+\frac{1}{6}x_{\min}^3-\frac{1}{2}x_{\min}x_{\text{prob}}^2\right] \\ &\quad+\frac{D}{H_2}\left[\frac{1}{6}x_{\max}^3-\frac{1}{2}x_{\max}x_{\text{prob}}^2+\frac{1}{3}x_{\text{prob}}^3\right]\end{aligned} \tag{2.9}$$

which is very easy to compute. We wish to mention here that $\overline{X}$ equals x_{prob} if and only if

$$x_{\text{prob}}=\frac{x_{\min}+x_{\max}}{2}, \tag{2.10}$$

that is, when the most probable value is the mean of the smallest and largest values. So, in this special case the above calculation can be omitted. If for any of the payoff-matrix elements there is a sequence of (e.g., historical) data, then their average value can be accepted as expectation. In the case of repeated values they have to be considered in the average as many times as they appear. If $x_1, x_2, \ldots, x_s$ denote the values with multiplicities $m_1, m_2, \ldots, m_s$, then $\overline{X}$ is the weighted average:

$$\overline{X}=\frac{m_1x_1+m_2x_2+\cdots+m_sx_s}{m_1+m_2+\cdots+m_s}. \tag{2.11}$$

In the case of unquantifyable objectives, subjective measures should be selected. For example, issues concerning beauty or aesthetics cannot be quantified. In this case a subjective $[0, 100]$ scale can be chosen, where 0 corresponds to the possible worst, and 100 corresponds to the possible best outcome. The members of the interest group associated to this objective should make the judgments. It is suggested that after a short discussion, each individual gives his/her assessments, which will be averaged later, and the averaged values will be used in the payoff table.

The above process has to be repeated for each matrix element in order to construct the entire payoff matrix. We strongly recommend that the meeting must continue until the entire table is finalized.

Each of the nine scenarios requires the construction of a payoff matrix. If time permits, several scenarios might be considered in one meeting, however the meetings have to be terminated only at times when a complete payoff matrix is constructed.

The outcome of this step is 9 payoff-matrices for each interest group. We note here certain matrix elements are not affected by the scenarios. These elements do not need to be recalculated. They have to be assessed only once and these values be repeated in each case. For the further analysis we have to assume that each objective function is to be maximized. Therefore all objective values of minimizing objectives should be multiplied by (-1). Therefore, in the final payoff-matrices larger objective value always should reflect better performance.

Step 5. The utility function associated to each objective will transform the objective values into an appropriate scale that can be easily interpreted by the decision makers. The assessment of the utility functions is suggested to be performed by the interest groups (stake holders) which are associated to the particular objective functions. The procedure to be outlined below has to be repeated for each objective.

The first step is to identify a possible worst objective value. Even in cases when an objective function is unbounded, a practical finite minimal value can be identified. Let this value be denoted by $g_{\min}$. Similarly, a possible maximal value has to be then identified, which can be denoted by $g_{\max}$. If U is the utility function associated to this objective, then define

$$U(g_{\min}) = 0 \quad \text{and} \quad U(g_{\max}) = 1. \tag{2.12}$$

To construct the shape of the function in between these extreme values, a bisection-type procedure is suggested. Let

$$g_{\text{mid}} = \frac{g_{\min} + g_{\max}}{2} \tag{2.13}$$

denote the midpoint of the initial interval $[g_{\min}, g_{\max}]$, and ask the interest group to estimate the value $U(g_{\text{mid}})$. In the case of disagreement, the average value has to be selected. Now, we have two subintervals. Repeat the procedure next to each subintervals: estimate the utility function value at both midpoints. Then we will have 4 subintervals. After the same procedure is repeated again, we will have 8 subintervals. So, after k repeats, the number of subintervals becomes 2^k.

The utility function is now given in a finite set of points, and in the second stage, a continuous function has to be constructed. In the literature (see for example, Szidarovszky et al. [1]) three basic types of utility functions are discussed representing risk-seeking, risk-averse, and risk-neutral viewpoints. These types are illustrated in Figures 2, 3, and 4.

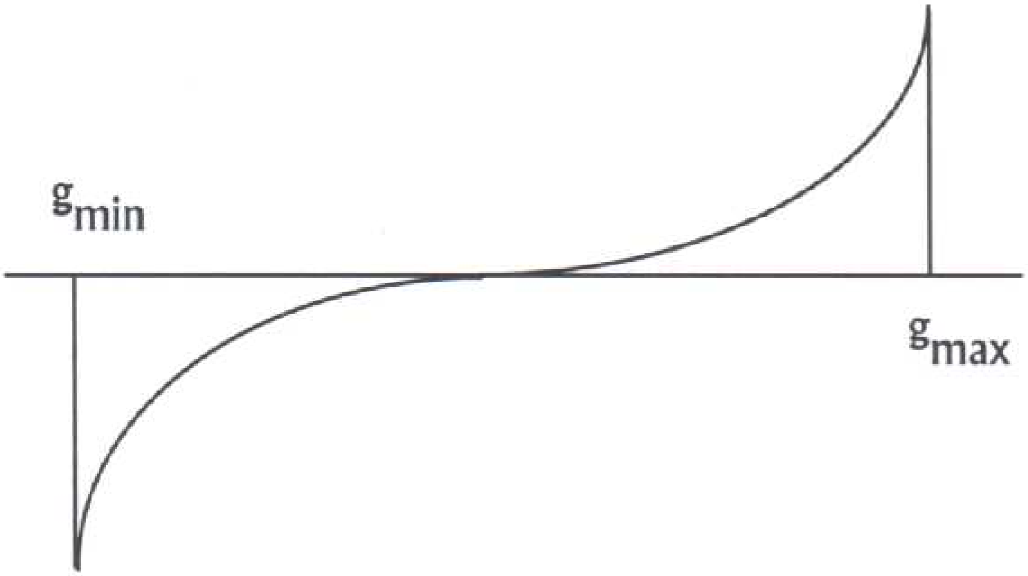


Figure 2. Risk-seeking utility function

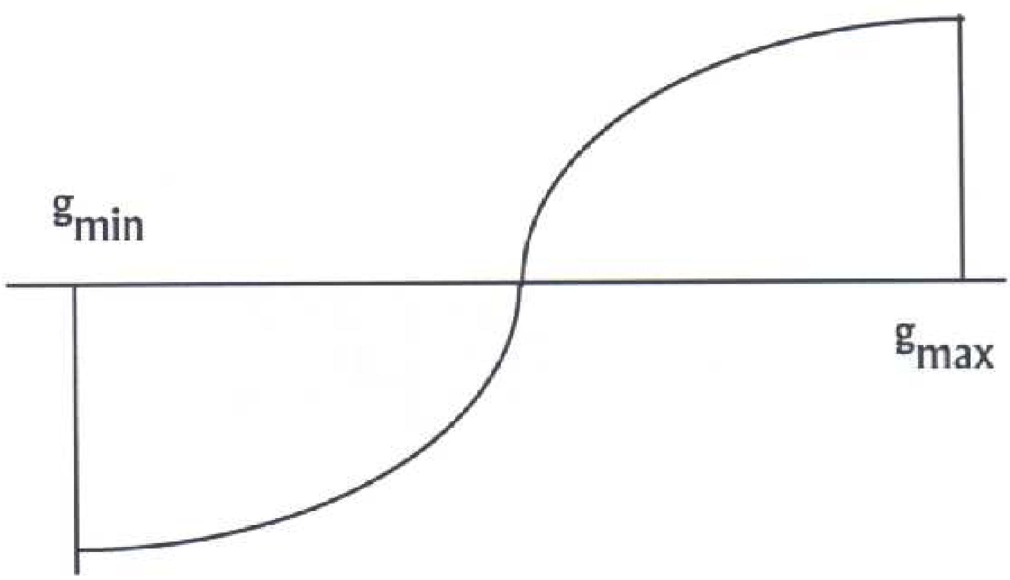


Figure 3. Risk-averse utility function

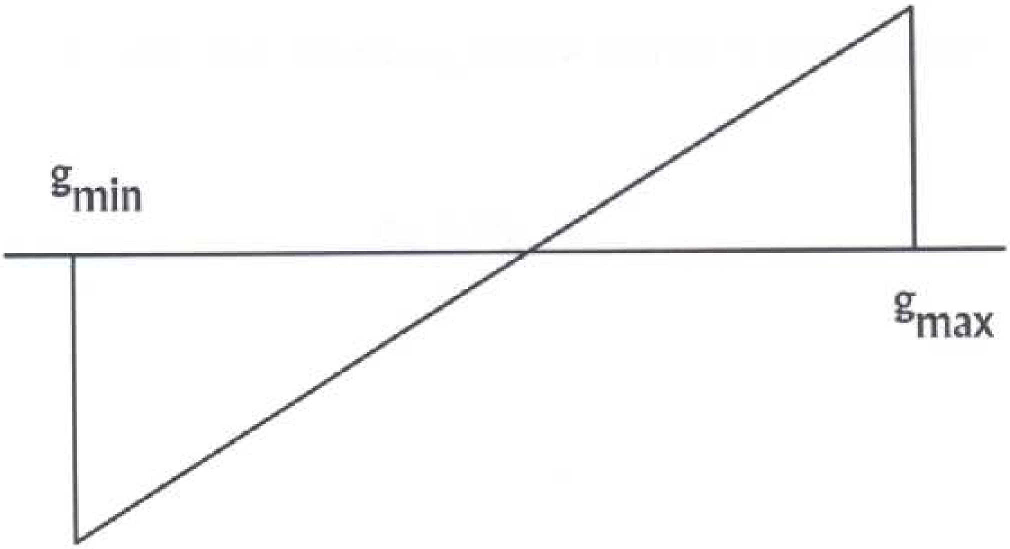


Figure 4. Risk-neutral utility function

In practical applications, however these types are often combined. Therefore we do not propose to fit any of these types to the discrete data. The continuous graph connecting the discrete points have to be the most smooth interpolator connecting the neighbor points. It is well known from the numerical analysis literature that the smoothest interpolator is the natural cubic spline (see for example, Yakowitz and Szidarovszky [9]). The construction of the natural cubic spline can be summarized as follows. Assume, that the independent variable values are $g_1 < g_2 < \cdots < g_n$ and the corresponding function values are $U_1, U_2, \ldots, U_n$. Define first

$$\left.\begin{aligned} h_k &= g_{k+1} - g_k, \\ s_k &= \tfrac{U_{k-1} - U_k}{h_k} \end{aligned}\right\} \quad (k = 1, 2, \ldots, n-1), \tag{2.14}$$

Select $c_1 = c_n = 0$, and for the unknowns $c_2, \ldots, c_{n-1}$ solve the following tridiagonal system of linear equations

$$c_{k-1}h_{k-1} + 2c_k(h_{k-1} + h_k) + c_{k-1}h_k = 3(s_k - s_{k-1}) \tag{2.15}$$

where $k = 2, 3, \ldots, n-1$. After the values c_k are determined, then define

$$a_k = U_k \tag{2.16}$$

$$b_k = s_k - \frac{h_k}{3}(2c_k + c_{k-1}) \tag{2.17}$$

and

$$d_k = \frac{c_{k+1} - c_k}{3h_k} \quad (k = 1, 2, \ldots, n-1). \tag{2.18}$$

And finally, the natural spline in the interval $[g_k, g_{k+1}]$ is given as follows:

$$a_k + b_k \cdot (g - g_k) + c_k \cdot (g - g_k)^2 + d_k \cdot (g - g_k)^3. \tag{2.19}$$

After the natural spline segments are computed for all subintervals of all objective functions, the continuous utility functions are completely constructed. After the procedure is completed, the resulted utility functions (in graphical form) are suggested to be sent back to the interest groups for final examination in order to have second thoughts on their assessed point values. If any changes are suggested by them, then either the resulting utility functions are slightly modified on the graphs, or the entire computation has to be repeated. We strongly believe that the possibility has to be given to the interest groups to see the consequences of their assessments, and to give them an additional chance to make necessary corrections if needed.

The final outcome of this step is a continuous utility function for each interest group and each objective function. The utility functions are given in graphical representations as well as in natural cubic spline forms.

Step 6. In this step the multiattribute utility functions will be constructed. Let $U_1, U_2, \ldots, U_s$ denote the individual utility functions associated to the different objective functions. We expect that these utility functions satisfy the mutually utility independence condition, that is, the utility function of any objective does not depend on particular values of the other utilities (for precise definition see, for example, Szidarovszky et al. [1]). In this case the multiattribute utility function is either additive:

$$U = k_1 U_1 + k_2 U_2 + \cdots + k_s U_s \tag{2.20}$$

or multiplicative:

$$kU + 1 = (kk_1 U_1 + 1)(kk_2 U_2 + 1) \cdots (kk_s U_s + 1) \tag{2.21}$$

where $k_1, k_2, \ldots k_s$, and k are unknown constants. The main problem is to determine these unknown values. For the proof of the above forms, see for example, Keeney and Raiffa [3].

We will present here two approaches, which are the most popular ones in practical application: least squares estimation, and a systematic assessment procedure.

Assume first that the multiattribute utility function is linear. In the least squares approach we will ask the members of the interest groups to assess the multiattribute utility function values at a set of given values of $U_1, U_2, \ldots, U_s$. We suggest to have 5 values: 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 for each individual utility function. Altogether we might have 5^s points. If it is too much, we can select 3 points only: 0, $\frac{1}{2}$, and 1, and consider only 3^s points. For each point, say $U_1^{(j)}, \ldots, U_s^{(j)}$, let the assessed multiattribute utility value be $U^{(j)}$, for $j = 1, 2, \ldots, 3^s$ or 5^s. The unknown k_i values can be then obtained by using standard linear least squares method for relation (2.20).

In the systematic approach we have to ask the stake-holders to assess the multiattribute utility function values at S special points:

$$U_1 = 0, \dots, U_{i-1} = 0,\ U_i = 1,\ U_{i+1} = 0, \dots, U_s = 0 \tag{2.22}$$

for $i = 1, 2, \dots, S$. The multiattribute utility value will correspond to the case, when the ith utility is at maximum level, but all others are at their minima. If $U^{(i)}$ is the assessed multiattribute utility value, then simple substitution shows that

$$k_i = U^{(j)}. \tag{2.23}$$

Consider next the multiplicative form. With keeping the above notation, the least squares method minimizes the least squares fit

$$\sum_j \left[kU^{(j)} + 1 - (kk_1U_1^{(j)} + 1)(kk_2U_2^{(j)} + 1) \cdots (kk_sU_s^{(j)} + 1)\right]^2. \tag{2.24}$$

This unconstrained optimization problem can be solved by standard methods such as gradient-type algorithms. For details, see for example, Hadley [10]. We mention here that several professional software packages, like MATLAB, IMSL, etc. might be used to perform the actual optimization. In the systematic approach substitute points (2.22) into equation (2.21) to see that for $i = 1, 2, \dots, S$,

$$kU^{(j)} + 1 = kk_i + 1, \tag{2.25}$$

since for $j \neq i$, $kk_jU_j + 1 = kk_j \cdot 0 + 1 = 1$, and $kk_iU_i + 1 = kk_i \cdot 1 + 1 = kk_i + 1$. From equation (2.25) we conclude that

$$k_i = U^{(i)} \tag{2.26}$$

similarly to the linear case. After the values k_i are recovered, substitute $U_1 = U_2 = \cdots = U_s = 1$ into the multiplicative formula (2.21) and normalize the multiattribute utility function to 1 to have an additional equation

$$k + 1 = (kk_1 + 1)(kk_2 + 1) \cdots (kk_s + 1) \tag{2.27}$$

for the remaining unknown k. Notice that $k = 0$ is always a solution which reduces the multiplicative form to linear. If $k \neq 0$, then equation (2.27) has to be solved for k. Simple algebra shows that after dividing by k, it can be rewritten as

$$k^{s-1}k_1k_2 \cdots k_s + \cdots + k(k_1k_2 + k_1k_3 + \cdots + k_{s-1}k_s) + (k_1 + k_2 + \cdots + k_s - 1) = 0$$

which is a polynomial equation for k. In the case of multiple solutions, the best solution can be obtained by an additional point selection in the following way. Select for example, $U_1 = U_2 = \cdots = U_s = \frac{1}{2}$ and let the assessed associated multiattribute utility function value be $U^{1/2}$. Then the measure to minimize is the following:

$$\left|kU^{1/2} + 1 - \left(\frac{kk_1}{2} + 1\right)\left(\frac{kk_2}{2} + 1\right) \cdots \left(\frac{kk_s}{2} + 1\right)\right|, \tag{2.28}$$

or by selecting a set of points, then least squares measure (2.24) can be used to select the best value of k from the solution set.

The final outcome of this step is the particular form (2.20) or (2.21) of the multiattribute utility function with the particular values of the coefficient, $k_1, k_2, \dots, k_s$ (or in addition, the value of) k.

Step 7. For each scenario and each tender the corresponding multiattribute utility function values are already computed. Let now consider a given tender. Let $g_1, \dots, g_s$ denote the computed objective function values (as the results of Step 4). Using the individual utility functions (as the results of Step 5) we may easily calculate the sequence of utility values

$$U_1^* = U_1(g_1),\ U_2^* = U_2(g_2), \dots, U_s^* = U_s(g_s),$$

and then using the result of Step 6, the corresponding multiattribute utility function value U^* can be determined by using either equation (2.20) or equation (2.21) depending on which type gave better fit in Step 6.

The outcome of this computation is shown in Table 1, where U_{ij}^* gives the multiattribute utility value of tender i $(i = 1, 2, \dots, p)$ under scenario j $(j = 1, 2, \dots, 9)$. Finally, for each tender i, we can compute the expected multiattribute utility values as

$$U_i^* = \sum_{j=1}^{9} P_j U_{ij}^*,$$

where the P_j probability values were obtained as the results of Step 2.

Tender	Scenario				
	1	2	3	$\cdots$	g
1	U_{11}^*	U_{12}^*	U_{13}^*	$\cdots$	U_{19}^*
2	U_{21}^*	U_{22}^*	U_{23}^*	$\cdots$	U_{29}^*
3	U_{31}^*	U_{32}^*	U_{33}^*	$\cdots$	U_{39}^*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p	U_{p1}^*	U_{p2}^*	U_{3p}^*	$\cdots$	U_{p9}^*

Table 1. Results of Step 7

Repeating the previous steps for each interested group, a final payoff matrix is obtained, in which the rows correspond to the interest groups, the columns to the tenders (decision alternatives), and the (k, l) element U_{kl} of this table gives the final evaluation of tender l by interest group k. This table (given in Table 2) can be considered as the payoff table of a discrete group decision problem, in which the rows correspond to the different members and the columns represent the decision alternatives.

Group	Tenders			
	1	2	$\cdots$	p
1	U_{11}	U_{12}	$\cdots$	U_{1p}^*
2	U_{21}	U_{22}	$\cdots$	U_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	U_{K1}	U_{K2}	$\cdots$	U_{Kp}

Table 2. Group decision payoff table

The final selection procedure will be outlined in the next section.

FINAL SELECTION PROCEDURE

Notice that the utility function for each group are normalized, so no normalization is needed in the final selection procedure. The final selection is suggested to be based on the following simple algorithm.

The importance of the different interest groups is assumed to be assessed by the final decision makers. Let $\alpha_1, \alpha_2, \dots, \alpha_K$ denote these factors, where $\alpha_i \geq 0$ for all i, and $\alpha_1 + \alpha_2 + \cdots + \alpha_K = 1$. Then for each tender l, compute the overall evaluation measures

$$U_l^F = \sum_{k=1}^{K} \alpha_k U_{kl}^2, \tag{3.1}$$

and select the tender l which has the largest U_l^F value.

Notice that any other multiobjective procedure, or any social choice of conflict resolution method can be selected for the final decision making. Equation (3.1) maximizes the I_2 distance from the origin, so this method is a special distance-based procedure. The large variety of multiobjective programming methods can be found in for exmaple, Szidarovszky et al. [1], conflict resolution concepts are examined, for example in Myerson [6], and social choice procedures are analyzed, for example in Taylor [11].

REFERENCES

1. Szidarovszky, F., Gershon, M., and Duckstein, L.: *Techniques for Multiobjective Decision Making in Systems Management*, Elsevier, Amsterdam, 1986.
2. Szidarovszky, F. and Molnár, S.: *Game Theory with Engineering Applications* (in Hungarian), Müszaki Könyvkiadó, Budapest, 1986.
3. Kenney, R.L. and Raiffa, H.: *Decisions with Multiple Objectives: Preference and Value Trade-Off*, Wiley, New York, 1976.
4. Fishburn, P.: *Utility Theory for Decision Making*, Wiley, New York, 1970.
5. Selten, R. (ed.): *Game Equilibrium Models*, Vols. 1–4, Springer-Verlag, Berlin, Heidelberg, New York, 1991.
6. Myerson, R.B.: *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA, 1991.
7. Szidarovszky, F. and Eskandari, A.: A Note on Pareto Equilibria, *Pure Math. and Appl.* **7**, No. 1–2 (1996), 191–195.
8. Eskandari, A., Ffolliott, P., and Szidarovszky, F.: Uncertainty in Multicriterion Watershed Management Problems, *Technology* **332A** (1995), 199–207.
9. Yakowitz, S. and Szidarovszky, F.: *Introduction to Numerical Computations*, (2nd editor, 1989), Macmillan, New York, 1986.
10. Hadley, G.: *Nonlinear and Dynamic Programming*, Addison-Wesley, Reading, MA, 1964.
11. Taylor, A.D.: *Mathematics and Politics*, Springer-Verlag, Berlin, Heidelberg, New York, 1995.
12. Molnár, S., Staub, K., Molnár, M., Conzelmann, G., Koritarov, V., Guziel, K., and Veselka, T.: *Final Audit Report, New Generating Capacity Tenders 97/1 and 97/2*. Prepared for the Hungarian Power Companies Ltd.,

published in Decision and Information Sciences Division, Argonne National Laboratory, USA, 1999.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2000, pp. 44–51.
Submitted: July 23, 2000. Published: December 15, 2000.

NEGOTIATIONS UNDER UNCERTAINTY

Ferenc Szidarovszky and Abdollah Eskandari

Abstract. A one-step solution for two-person conflicts is examined when the break-down probabilities depend on the goodness of the offers and the relative power indices of the players. In the one-step solution each player maximizes his/her expected payoff. The uncertainty in the power index is characterized by its CDF and fuzzy membership function. Based on the monotonicity of solution simple algorithms are suggested to obtain the CDF and membership function of the solution in closed forms. A simple example illustrates the methodology.

1991 A.M.S. Subject Classification Codes: 90.

Key Words and Phrases: Negotiation, uncertainty, fuzzy function, break down probability.

1. INTRODUCTION

Before 1950, only few negotiation models had been developed and most were based on economic or accounting theory. Since 1950, negotiation and bargaining became an independent and important research area. For example, Nash's papers laid the foundation for the modern theoretical works in negotiation and bargaining ([9], [10], [11]). His first paper

stated that there is more than one way that two individuals in a bargaining situation could collaborate for mutual benefits. His second paper argued that non-cooperative game theory should be used to study the bargaining and cooperation in games. His third paper suggested the idea to relate the axiomatic solutions to the equilibria of strategic models, which implies that in order to understand the outcomes of bargaining, it is appropriate to model the bargaining as a noncooperative game and solve for its equilibria. This procedure is known as the *Nash program.*

There were several bargaining or negotiation models developed later. The non-symmetric Nash solution was introduced in [5], and solutions satisfying individual

---

F. Szidarovszky, A. Eskandari, Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona 85721, U.S.A.
E-mail: szidar@sie.arizona.edu

monotonicity have been introduced in [8]. This solution has been further generalized in [3] as the reference function solution. Among the mostly known solution concepts we mention the egalitarian solution [7], the superadditive solution [12], the equal sacrifice solution [4], and the equal area solution [2]. Nash has shown that the equilibrium set coincides with the set of Pareto solutions if the problem is considered as a two-person noncooperative game. One might consider bargaining as a single player decision problem, when the strategy selection of the other player is considered random. If uniform distribution is assumed and each player maximizes his/her expected payoff, then the optimal selection is equivalent to the Nash solution. Similar equivalence holds for applying the principle of Zeuthen [17] in determining the order in which concessions are made. One of the most popular bargaining methods is the alternating offer process [13], which has been extended in [6]. In the model of [3] the payoffs of the players depend not only on their offers to themselves but also on how generous their offers are.

The bargaining process to be examined in this paper is a combination of the ideas presented in [13] and [3], when the break-down probabilities depend on how good the offers are, and also on the relative power of the participants. In [16] the uncertainty of the Pareto frontier has been examined. In addition, time-variant negotiations have been discussed. In this paper the uncertainty in the relative power index is our main concern. Randomization as well as fuzzification will be used to predict the effect of this uncertainty in the solution of the conflict.

2. THE BARGAINING SOLUTION

A two-person conflict is identified mathematically with a pair $(S, \mathbf{d})$, where $S \subseteq R^2$ is a convex, closed, comprehensive feasible set, and $\mathbf{d} \in S$ is the disagreement payoff vector. Any point $(x_1, x_2) \in S$ can be considered as a simultaneous payoff pair for the two players. Without loss of generality we may assume that $\mathbf{d} = \mathbf{0}$, otherwise we may introduce the new variables $x_i^1 = x_i - d_i$ $(i = 1, 2)$. It is also assumed that there exists a concave, strictly decreasing, continuously differentiable function $g: [0, A] \mapsto R$ such that $g(A) = 0$. Since no person accepts an offer worse than the disagreement payoff, we may assume that $S = \{(x_1, x_2) \mid 0 \leq x_1 \leq A, 0 \leq x_2 \leq g(x_1)\}$. The Pareto frontier is therefore the graph of function g. Let (x_1^0, x_2^0) and (y_1^0, y_2^0) denote the initial positions of the two players, where $A \geq x_1^0 > y_1^0 \geq 0$, $x_2^0 = g(x_1^0)$, and $y_2^0 = g(y_1^0)$. If player 1 presents an offer (z_1, z_2), then we assume that it will be accepted by player 2 with probability $[(z_2 - x_2^0)/(y_2^0 - x_2^0)]^{P_2}$, where P_2 is the power of player 2. Notice that the selection $z_2 = x_2^0$ leads to certain rejection, and offer $z_2 = y_2^0$ will result in certain acceptance. Between these extreme values the acceptance probability increases if a better offer is presented. The optimal offer of player 1 can be obtained by maximizing his/her expected payoff $E_1(z_1) = z_1 \cdot [(g(z_1) - x_2^0)/(y_2^0 - x_2^0)]^{P_2}$. Similarly, if player 2 presents an offer (z_1, z_2), then it will be accepted with probability $[(z_1 - y_1^0)/(x_1^0 - y_1^0)]^{P_1}$, where P_1 is the power of player 1. Therefore the optimal offer of player 2 is obtained by maximizing his/her expected payoff $E_2(z_1) = g(z_1) \cdot [(z_1 - y_1^0)/(x_1^0 - y_1^0)]^{P_1}$. It can be proven (see [15]) that under the above assumptions both players have a unique payoff maximizing offers with arbitrary values of $P_1, P_2 > 0$. Finally, let α denote the relative power of player 2 above player 1. The solution of the conflict is a point (z_1, z_2) of the Pareto frontier that is optimal for both players with some values of P_1 and P_2 such that $\alpha = P_2/P_1$. It can be proved (see, for example [14]) that the solution is the unique solution of the single variable monotonic equation

$$\frac{x_2^0 - g(z_1)}{z_1 g'(z_1)} - \alpha \frac{y_1^0 - z_1}{\frac{g(z_1)}{g'(z_1)}} = 0 \tag{1}$$

inside the interval $[y_1^0, x_1^0]$. It is easy to see that in the special case when $x_2^0 = y_1^0 = 0$ this equation can be further simplified as

$$g(z_1) + \alpha z_1 g'(z_1) = 0. \tag{2}$$

Notice that equation (2) shows that in this special case the bargaining solution is equivalent to a nonsymmetric Nash-solution ([5]).

For the sake of simplicity let $h(z_1, \alpha)$ denote the left-hand side of equation (1). Under the assumptions given in the beginning of this section, h is strictly decreasing in z_1 with any fixed $\alpha > 0$, and is also strictly decreasing in α with arbitrary fixed $z_1 \in (y_1^0, x_1^0]$. Therefore there is a pair of unique functions

$$z_1 = h_1(\alpha) \quad \text{and} \quad \alpha = h_2(z_1),$$

where h_1 and h_2 are inverses of each other. Function h_1 can be computed only numerically; for each value of α, the unique root of equation (1) has to be determined, however function h_2 can be obtained in a closed form:

$$\alpha = h_2(z_1) = \frac{[x_2^0 - g(z_1)]g(z_1)}{z_1[g'(z_1)]^2(y_1^0 - z_1)}. \tag{3}$$

Next we show that both functions h_1 and h_2 are strictly decreasing. Consider first function h_1 and assume that with some $\alpha < \alpha'$, $z_1 = h_1(\alpha) \le z_1' = h_1(\alpha')$. Then $0 = h(z_1, \alpha) > h(z_1, \alpha') \ge h(z_1', \alpha') = 0$, which is an obvious contradiction. Since h_2 is the inverse of h_1, it is also strictly decreasing.

The above discussion is next illustrated by a simple example.

Example 1. Assume that two players want to share a pie, that is, $g(x_1) = 1 - x_1$ with $A = 1$. Assume that the initial positions are $(y_1^0, y_2^0) = (0, 1)$ and $(x_1^0, x_2^0) = (1, 0)$. Then equation (1) has the special form

$$\frac{-1 + z_1}{-z_1} - \alpha \frac{-z_1}{\frac{1 - z_1}{-1}} = 0$$

which can be simplied as

$$(1 - z_1)^2 = \alpha z_1^2.$$

Therefore

$$z_1 = \frac{1}{1 + \sqrt{\alpha}} \quad \text{and} \quad \alpha = \frac{(1 - z_1)^2}{z_1^2}.$$

3. STOCHASTIC ANALYSIS

Assume now that α is a random variable with a continuous cumulative distribution function F. Then $z_1 = h_1(\alpha)$ is also a random variable, and its cumulative distribution function can be obtained as follows:

$$\begin{aligned} G(Z) &= P(z_1 \le Z) = P(h_1(\alpha) \le Z) = P(\alpha \ge h_2(Z)) \\ &= 1 - F(h_2(Z)). \end{aligned} \tag{4}$$

Hence $G(Z)$ is obtained in a closed form giving the cumulative distribution function of the solution z_1 of the conflict. The expectation and variance of z_1 can then be determined in a routine method.

Example 2. Consider again the conflict of Example 1 and assume that the probability density function of α is triangular, as shown in Figure 1. Then simple integration shows that the cumulative distribution function is as follows:

$$F(\alpha) = \begin{cases} 0, & \text{if } \alpha \leq 1 \\ \frac{(\alpha-1)^2}{2}, & \text{if } 1 < \alpha \leq 2 \\ 1 - \frac{(3-\alpha)^2}{2}, & \text{if } 2 < \alpha \leq 3 \\ 1, & \text{if } \alpha > 3 \end{cases}$$

This function is illustrated in Figure 2. In our case $h_2(Z) = \frac{(1-Z)^2}{Z^2}$, therefore equation (4) implies that:

$$G(Z) = \begin{cases} 0, & \text{if } Z \leq \frac{\sqrt{3}-1}{2} \\ \frac{(2Z^2+2Z-1)^2}{2Z^4}, & \text{if } \frac{\sqrt{3}-1}{2} \leq Z < \sqrt{2}-1 \\ 1 - \frac{(1-2Z)^2}{2Z^4}, & \text{if } \sqrt{2}-1 \leq Z < \frac{1}{2} \\ 1, & \text{if } Z \geq \frac{1}{2} \end{cases}$$

This is a continuous cumulative distribution function showing that the Z values from the open interval $\left(\frac{\sqrt{3}-1}{2}, \frac{1}{2}\right)$ might occur with positive probabilities. It is also interesting to note that $G(\sqrt{2}-1) = \frac{1}{2}$. The graph of G is shown in Figure 3.

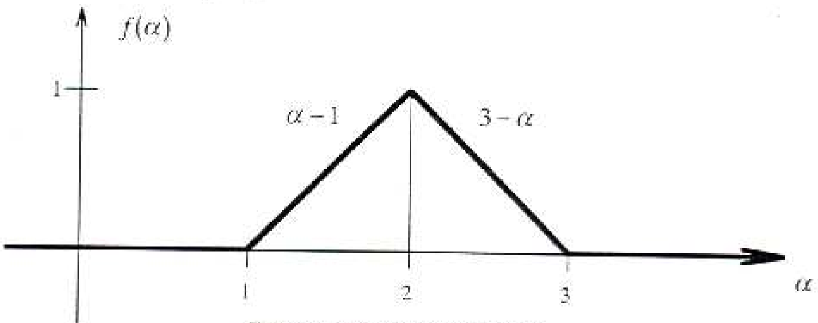


Figure 1. Density Function of α

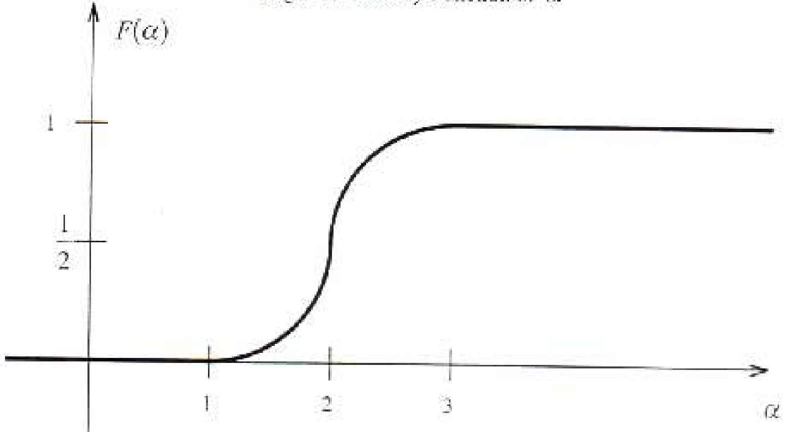


Figure 2. Cumulative Distribution Function of α

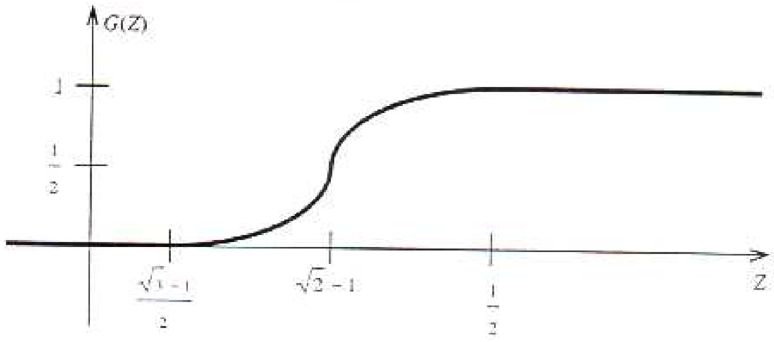


Figure 3. Cumulative Distribution Function of the Solution

4. FUZZY ANALYSIS

In the approach of fuzzification the membership function of parameter α has to be specified as the characterization of its uncertainty. In most applications a triangular or trapezoidal membership function is selected. The membership function of the solution can be then determined in the following way. For each membership level h, let $\alpha_{\min}(h)$ and $\alpha_{\max}(h)$ denote the minimal and maximal α values with membership values at least h. Since Z strictly decreases in α, $h_1(\alpha_{\max}(h))$ and $h_1(\alpha_{\min}(h))$ give the corresponding minimal and maximal Z values for the same membership level h.

Example 3. Consider again the conflict of the previous examples and assume now that α has a triangular membership function in Figure 4.

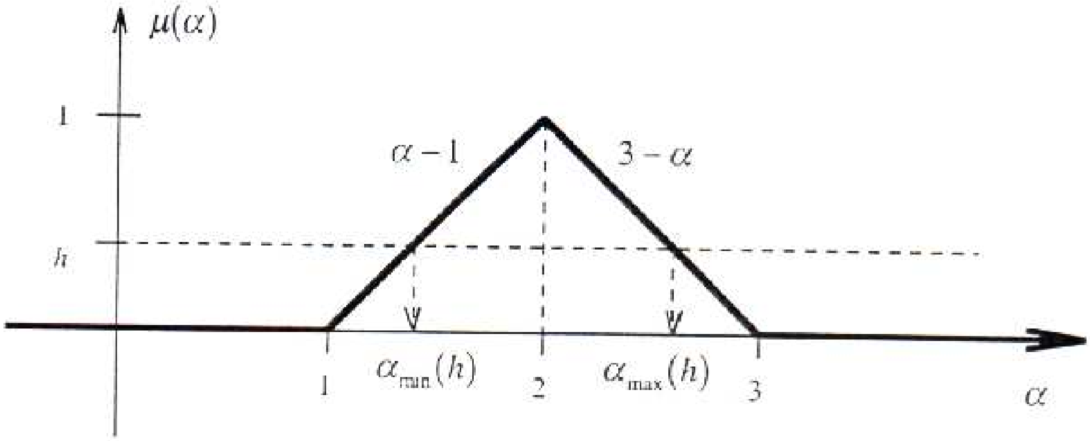


Figure 4. Membership Function $\mu(\alpha)$

For each membership level $h \in (0, 1)$ we have

$$\alpha_{\min}(h) = 1 + h \quad \text{and} \quad \alpha_{\max}(h) = 3 - h$$

therefore

$$Z_{\min}(h) = h_1(\alpha_{\max}(h)) = \frac{1}{1 + \sqrt{3 - h}}$$

and

$$Z_{\max}(h) = h_1(\alpha_{\min}(h)) = \frac{1}{1 + \sqrt{1 + h}}.$$

The resulted membership function $\mu(Z)$ is shown in Figure 5. This membership function is the fuzzy characteristic of the uncertainty of the result of the conflict.

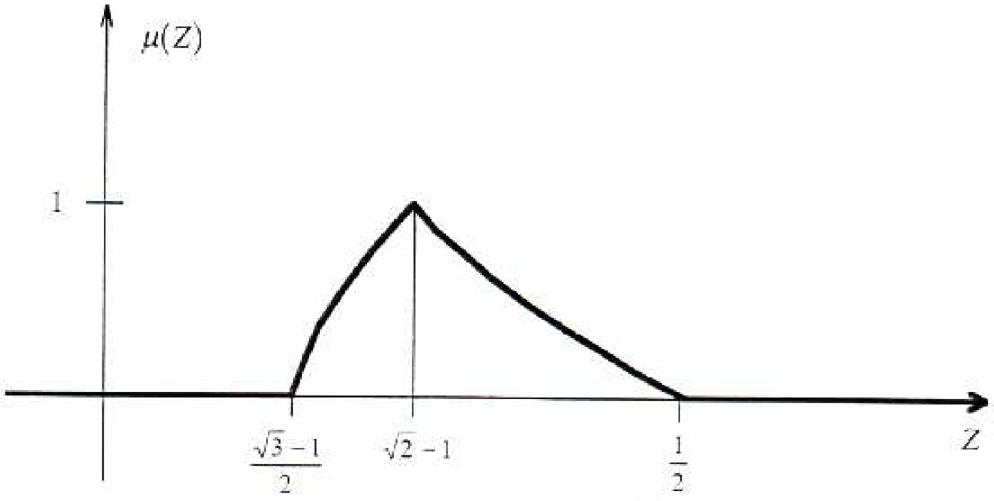


Figure 5. Membership Function $\mu(Z)$

5. CONCLUSIONS

The uncertainty in the relative power index was modelled by randomization and fuzzification. In the first case the cumulative distribution function of the uncertain parameter is assumed to be known, and the same for the solution of the conflict is then determined in a closed form expression. The resulted distribution function illustrates the uncertainty of the solution, and allows us to compute such characteristics as expectation and variance. In the case of fuzzification it is assumed that the membership function of the uncertain parameter is given, and the membership funciton of the solution is then determined by using the transformation model. The resulted membership function illustrates the uncertainty of the solution, and by using defuzzification (such as center of mass) a crisp solution is obtained which takes the uncertainty of the model parameter.

REFERENCES

1 Anbar, D. and Kalai, E.: A One-Shot Bargaining Problem, *International J. of Game Theory* **7** (1978), 13–15.
2 Anbarci, N.: Noncooperative Foundations of the Area Monotonic Solution, *Quarterly J. of Econ.* **108** (1993), 245–258.
3 Anbarci, N.: Reference Functions and Balanced Concessions in Bargaining, *Canadian J. of Economics* **28** (1995), 675–682.
4 Chun, Y.: The Equal-Loss Principle for Bargaining Problems, *Econ. Letters* **26** (1998), 103–106.
5 Harsanyi, J.C. and Selten, R.: A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information, *Management Science* **18**, Part II (1972), 80–106.

6 Howard, J.V.: A Social Choice Rule and Its Implementation in Perfect Equilibrium, *J. of Econ. Theory* **56** (1992), 142–159.
7 Kalai, E.: Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons, *Econometrica* **45** (1977), 1623–1630.
8 Kalai, E. and Smorodinsky, M.: Other Solutions to Nash's Bargaining Problem, *Econometrica*, **43** (1975), 513–518.
9 Nash, J.F.: The Bargaining Problem, *Econometrica* **18** (1950), 155–162.
10 Nash, J.F.: Non-cooperative Games, *Ann. Math.* **54** (1951), 286–295.
11 Nash, J.F.: Two-person Cooperative Games, *Econometrica* **28** (1953), 128–140.
12 Perles, M.A. and Maschler, M.: A Super-additive Solution for Nash's Bargaining Game, *International J. of Game Theory* **10** (1981), 163–193.
13 Rubinstein, A.: Perfect Equilibrium in a Bargaining Model, *Econometrica* **50** (1982), 97–110.
14 Szidarovszky, F.: A Relation Between Stochastic Bargaining with Offer-dependent Break-down Probabilities and the Non-symmetric Nash Solution, Proceedings of the 27th Annual Meeting of the Western Decision Science Institute, April 7–11, 1998, Reno, Nevada, pp. 485–587.
15 Szidarovszky, F.: Bargaining with Offer-dependent Break-down Probabilities, *Appl. Math. and Comp.* **90** (1998), 117–127.
16 Szidarovszky, F.: Negotiations with Changing Conditions and Payoffs, Proceedings of the 28th Annual Meeting of the Western Decision Science Institute, April 6–10, 1999 (accepted).
17 Zeuthen, F.: *Problems of Monopoly and Economic Welfare*, Routledge, London, 1930.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 1 December 2000, pp. 52 – 59
Submitted: August 7, 2000. Published: December 15, 2000

On Solutions of Linear Difference Equations with Infinite Delay

Mustapha Lakrib

Abstract

The difference equation with infinite delay $(DE)_\varepsilon$ $x_{n+1} = Ax_n + \varepsilon \sum_{k=-\infty}^{n} R_{n-k}x_k$ *is considered. Sufficient conditions for the existence of a unique solution are proved. An* $O(\varepsilon)$*-approximation of this solution for* $-N \leq n \leq N$, $N > 0$, *by the solution of the equation* $(DE)_0$ *is obtained.*

(1991) A.M.S Subject Classification Codes. 45D, 34C.
Key Words and Phrases. Difference equation, Infinite delay, Approximation

In [1], Ryabov considers a problem concerning the existence of so-called two-sided solutions of linear integrodifferential equations of the form

$$\frac{dx}{dt} = Ax + \varepsilon \int_{-\infty}^{t} R(t-s)x(s)ds, \tag{1}$$

where ε is a positive small parameter, $t \in \mathbb{R}$, $x \in \mathbb{R}^m$, A is a constant $m \times m$ matrix and R is a matrix of continuous functions satisfying the inequality

$$\|R(t)\| \leq c\frac{e^{-\gamma t}}{t^{1-\alpha}},$$

Institut d'Informatique, Université Djillali Liabès, Sidi Bel Abbès, Algérie

where c, γ, α are positive constants and $0 < \alpha < 1$. It is proven there that if $\min\{\mathcal{R}e(\lambda_i) : 1 \leq i \leq m\} > -\gamma$, where $\lambda_1, \ldots, \lambda_m$ are eigenvalues of A, then there is an $\varepsilon_0 > 0$ such that for any $\varepsilon \in (0, \varepsilon_0]$ and any $x_0 \in \mathbb{R}^m$ the equation (1) has a unique solution x_ε (so-called two-sided solution) defined on the entire axis t for every given initial condition $x_\varepsilon(0) = x_0$ and satisfying $\lim_{\varepsilon\to 0} \|x_\varepsilon - x\|_L = 0$ for any $L > 0$, where $\|x_\varepsilon - x\|_L = \sup\{\|x_\varepsilon(t) - x(t)\| : -L \leq t \leq L\}$, $x(t) = e^{At}x_0$.

The equation (1) has the discretization

$$x_{n+1} = Ax_n + \varepsilon(R_0 x_n + R_1 x_{n-1} + \cdots + R_k x_{n-k} + \cdots), \tag{2}$$

which is a difference equation with infinite delay. The purpose of this paper is to study (2) in a Banach space E, with $A, R_k \in \mathcal{L}(E)$, the space of continuous linear mappings from E into E ($k = 0, 1, \ldots$), A is invertible and $A^{-1} \in \mathcal{L}(E)$. Under sufficient conditions, we shall prove the existence of solutions of (2) determined uniquely by the initial value and defined for all integers.

Theorem. *Suppose that the following conditions are satisfied*

$$\|A^{-1}\| < 1, \qquad \|R_0\| \leq 1, \qquad \|R_n\| \leq \frac{\gamma^{-n}}{n^{1-\alpha}}, \quad n = 1, 2, \ldots, \tag{3}$$

where γ, α are constants, $e \leq \gamma$, $0 < \alpha < 1$.

Then there exist a positive constant ε_0 and a unique operator $B \in \mathcal{L}(E)$, such that, for any $\varepsilon \in (0, \varepsilon_0]$ and any $x_0 \in E$, the sequence $x = \{x_n(\varepsilon)\}_{n=-\infty}^{+\infty} = \{(A + \varepsilon B)^n x_0\}_{n=-\infty}^{+\infty}$ is the unique solution of (2) satisfying the condition $x_0(\varepsilon) = x_0$ with $x \in F := \{y = \{y_n\}_{n=-\infty}^{\infty} : y_n \in E,\ \sup\{|y_n| : -\infty < n \leq 0\} < \infty\}$, where $|.|$ is the norm on E. Moreover, for any positive integer N, there exists a positive constant C (depending on ε_0, x_0 and N) such that $\sup\{|x_n(\varepsilon) - A^n x_0| : -N \leq n \leq N\} \leq C\varepsilon$.

Proof. Let us seek the solution of (2) in the form $x = \{x_n(\varepsilon)\}_{n=-\infty}^{+\infty} = \{(A+\varepsilon B)^n x_0\}_{n=-\infty}^{+\infty}$, where $B \in \mathcal{L}(E)$ is such that $AB = BA$ and $A + \varepsilon B$ is invertible. The matrix B satisfy the equation

$$B = P_\varepsilon(B) := R_0 + R_1(A+\varepsilon B)^{-1} + \cdots + R_k(A+\varepsilon B)^{-k} + \cdots. \tag{4}$$

Consider the mapping $P_\varepsilon : \mathcal{D} \to \mathcal{L}(E)$, defined by (4), where $\mathcal{D} = \{B \in \mathcal{L}(E) : AB = BA \text{ and } \|B\| \leq 1\}$, $\|B\| := \sup\{|Bx| : |x| \leq 1\}$. If $B_1, B_2 \in \mathcal{D}$

then

$$\begin{aligned}\|P_\varepsilon(B_1)-P_\varepsilon(B_2)\| &= \|R_1[(A+\varepsilon B_1)^{-1}-(A+\varepsilon B_2)^{-1}]+R_2[(A+\varepsilon B_1)^{-2}\\ &\quad -(A+\varepsilon B_2)^{-2}]+\cdots+R_k[(A+\varepsilon B_1)^{-k}-(A+\varepsilon B_2)^{-k}]+\cdots\|.\end{aligned}$$

Since $(A+\varepsilon B_i)=A(I+\varepsilon A^{-1}B_i)$ and $AB_i=B_iA$, $i=1,2$ where I is the unit matrix, we have

$$\begin{aligned}\|P_\varepsilon(B_1)-P_\varepsilon(B_2)\| \le \|R_1\|\|A^{-1}\|\|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|\\ +\|R_2\|\|A^{-1}\|^2\|(I+\varepsilon A^{-1}B_1)^{-2}-(I+\varepsilon A^{-1}B_2)^{-2}\|+\cdots\\ +\|R_k\|\|A^{-1}\|^k\|(I+\varepsilon A^{-1}B_1)^{-k}-(I+\varepsilon A^{-1}B_2)^{-k}\|+\cdots.\end{aligned} \tag{5}$$

We have the following estimate

$$\begin{aligned}&\|(I+\varepsilon A^{-1}B_1)^{-k}-(I+\varepsilon A^{-1}B_2)^{-k}\| = \|[(I+\varepsilon A^{-1}B_1)^{-1}]^k-[(I+\varepsilon A^{-1}B_2)^{-1}]^k\|\\ &\quad\le \|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|\|[(I+\varepsilon A^{-1}B_1)^{-1}]^{k-1}\\ &\qquad +[(I+\varepsilon A^{-1}B_1)^{-1}]^{k-2}(I+\varepsilon A^{-1}B_2)^{-1}+\cdots+[(I+\varepsilon A^{-1}B_2)^{-1}]^{k-1}\|\\ &\quad\le \|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|\{\|(I+\varepsilon A^{-1}B_1)^{-1}\|^{k-1}\\ &\qquad +\|(I+\varepsilon A^{-1}B_1)^{-1}\|^{k-2}\|(I+\varepsilon A^{-1}B_2)^{-1}\|+\cdots+\|(I+\varepsilon A^{-1}B_2)^{-1}\|^{k-1}\}.\end{aligned}$$

As for $B_1, B_2\in\mathcal{D}$ and $\varepsilon\in(0,\varepsilon_1]$, where $\varepsilon_1<1/\|A^{-1}\|$, we have

$$\begin{aligned}\|(I+\varepsilon A^{-1}B_i)^{-1}\| &= \|I-\varepsilon(A^{-1}B_i)+\varepsilon^2(A^{-1}B_i)^2-\cdots\|\\ &\le 1+\varepsilon\|A^{-1}\|+\varepsilon^2\|A^{-1}\|^2+\cdots=\frac{1}{1-\varepsilon\|A^{-1}\|},\end{aligned}$$

we obtain the inegality

$$\begin{aligned}&\|(I+\varepsilon A^{-1}B_1)^{-k}-(I+\varepsilon A^{-1}B_2)^{-k}\|\\ &\qquad\le \|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|\frac{k}{(1-\varepsilon\|A^{-1}\|)^{k-1}}.\end{aligned} \tag{6}$$

Thus, for all $B_1, B_2\in\mathcal{D}$ and $\varepsilon\in(0,\varepsilon_1]$, from (5) we obtain the estimate

$$\begin{aligned}\|P_\varepsilon(B_1)-P_\varepsilon(B_2)\| &\le \{\|R_1\|\|A^{-1}\|+\|R_2\|\|A^{-1}\|^2\frac{2}{1-\varepsilon\|A^{-1}\|}+\cdots\\ &\quad +\|R_k\|\|A^{-1}\|^k\frac{k}{(1-\varepsilon\|A^{-1}\|)^{k-1}}+\cdots\}\|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|\\ &\le S_1\|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\|,\end{aligned} \tag{7}$$

where $S_1 = (1-\varepsilon\|A^{-1}\|)\sum_{k=1}^{\infty} k\|R_k\|\frac{\|A^{-1}\|^k}{(1-\varepsilon\|A^{-1}\|)^k}$. As $\|A^{-1}\| < \gamma$ then $\gamma^{-1} < 1/\|A^{-1}\|$. Choose ε_2 sufficiently small such that $\gamma^{-1}+\varepsilon < 1/\|A^{-1}\|$ for $\varepsilon \in (0,\varepsilon_2]$. We obtain

$$\frac{\gamma^{-1}\|A^{-1}\|}{1-\varepsilon\|A^{-1}\|} < 1 \quad \text{for } \varepsilon \in (0,\varepsilon_2]. \tag{8}$$

The conditions (3) and (8) insure the convergence of the series S_1.

The following estimate is needed

$$\begin{gathered}\|(I+\varepsilon A^{-1}B_1)^{-1}-(I+\varepsilon A^{-1}B_2)^{-1}\| \\ = \|(I-\varepsilon(A^{-1}B_1)+\varepsilon^2(A^{-1}B_1)^2-\cdots)-(I-\varepsilon(A^{-1}B_2)+\varepsilon^2(A^{-1}B_2)^2-\cdots)\| \\ \le \varepsilon\|A^{-1}B_1-A^{-1}B_2\|+\varepsilon^2\|(A^{-1}B_1)^2-(A^{-1}B_2)^2\|+\cdots \\ +\varepsilon^k\|(A^{-1}B_1)^k-(A^{-1}B_2)^k\|+\cdots.\end{gathered} \tag{9}$$

The mean value theorem yields

$$\begin{gathered}\|(A^{-1}B_1)^k-(A^{-1}B_2)^k\| \le \|A^{-1}\|^k\|B_1^k-B_2^k\| \\ \le \|A^{-1}\|^k \sup\{\|k[(1-\theta)B_1+\theta B_2]^{k-1}\| : 0\le\theta\le 1\}\|B_1-B_2\| \\ \le k\|A^{-1}\|^k\|B_1-B_2\|.\end{gathered} \tag{10}$$

From (7), (9) and (10) it follows that for all $B_1, B_2 \in \mathcal{D}$ and $\varepsilon \in (0,\min\{\varepsilon_1,\varepsilon_2\}]$, we have

$$\|P_\varepsilon(B_1)-P_\varepsilon(B_2)\| \le \varepsilon\|A^{-1}\|S_1S_2\|B_1-B_2\|, \tag{11}$$

where $S_2 = \sum_{k=0}^{\infty}(k+1)\varepsilon^k\|A^{-1}\|^k$ is a convergent series. Therefore the inequality (11) implies that if $\varepsilon \in (0,\varepsilon_3]$, where $\varepsilon_3 < \frac{1}{\|A^{-1}\|S_1S_2}$, then the mapping P_ε is a contraction on $\mathcal{D}$ and thus it has a unique fixed point in $\mathcal{D}$. So if we set $x_n(\varepsilon) = (A+\varepsilon B)^n x_0$, where B is the unique fixed point of P_ε, the sequence $x = \{x_n(\varepsilon)\}_{n=-\infty}^{\infty}$ is a solution of (2) satisfying the condition $x_0(\varepsilon) = x_0$. As $\|A^{-1}\| < 1$, choose ε_4 sufficiently small, such that $\|(A+\varepsilon B)^{-p}\| \le \|(A+\varepsilon B)^{-1}\|^p < 1$ for all positive integer p and $\varepsilon \in (0,\varepsilon_4]$. This implies that $\sup\{|x_n(\varepsilon)| : -\infty < n \le 0\} \le |x_0| < \infty$, i.e. $x = \{x_n(\varepsilon)\}_{n=-\infty}^{\infty} \in F$.

We shall prove the uniqueness of the solution in F of (2).

Consider the space

$$\tilde{F} := \{x = \{x_n\}_{n=-\infty}^{0} : x_n \in E,\ \sup\{|x_n| : -\infty < n \le 0\} < \infty\}$$

which is a Banach space with the norm $\|x\| := \sup\{|x_n| : -\infty < n \le 0\}$.

Let $x_n = A^n y_n$, $-\infty < n \le 0$. Then by substituting into (2), we obtain

$$y_{n+1} = y_n + \varepsilon A^{-(n+1)}(R_0 A^n y_n + R_1 A^{n-1} y_{n-1} + \cdots + R_k A^{n-k} y_{n-k} + \cdots).$$

Observe that

$$\begin{array}{rcl} y_0 - y_{-1} & = & \varepsilon(R_0 A^{-1} y_{-1} + R_1 A^{-2} y_{-2} + \cdots + R_k A^{-1-k} y_{-1-k} + \cdots), \\ y_{-1} - y_{-2} & = & \varepsilon A(R_0 A^{-2} y_{-2} + R_1 A^{-3} y_{-3} + \cdots + R_k A^{-2-k} y_{-2-k} + \cdots), \\ \vdots & & \\ y_{-(p-1)} - y_{-p} & = & \varepsilon A^{p-1}(R_0 A^{-p} y_{-p} + R_1 A^{-p-1} y_{-p-1} + \cdots + R_k A^{-p-k} y_{-p-k} + \cdots). \end{array}$$

By adding these equations, we get

$$\begin{array}{rcl} y_0 - y_{-p} & = & (y_0 - y_{-1}) + (y_{-1} - y_{-2}) + \cdots + (y_{-(p-1)} - y_{-p}) \\ & = & \varepsilon A^{p-1}(R_0 A^{-p} y_{-p} + R_1 A^{-p-1} y_{-p-1} + \cdots + R_k A^{-p-k} y_{-p-k} + \cdots) \\ & & + \cdots + \varepsilon A(R_0 A^{-2} y_{-2} + R_1 A^{-3} y_{-3} + \cdots + R_k A^{-2-k} y_{-2-k} + \cdots) \\ & & + \varepsilon(R_0 A^{-1} y_{-1} + R_1 A^{-2} y_{-2} + \cdots + R_k A^{-1-k} y_{-1-k} + \cdots). \end{array}$$

A return to the initial variable gives

$$\begin{array}{rcl} A^{-p} x_0 - x_{-p} & = & \varepsilon A^{-1}(R_0 x_{-p} + R_1 x_{-p-1} + \cdots + R_k x_{-p-k} + \cdots) \\ & & + \cdots + \varepsilon A^{-p+1}(R_0 x_{-2} + R_1 x_{-3} + \cdots + R_k x_{-2-k} + \cdots) \\ & & + \varepsilon A^{-p}(R_0 x_{-1} + R_1 x_{-2} + \cdots + R_k x_{-1-k} + \cdots). \end{array}$$

Define the mapping

$$Q_\varepsilon : \tilde{F} \to \{x = \{x_n\}_{n=-\infty}^{0}, x_n \in E\},$$

by

$$\begin{array}{rcl} (Q_\varepsilon x)_{-p} & = & \varepsilon A^{-1}(R_0 x_{-p} + R_1 x_{-p-1} + \cdots + R_k x_{-p-k} + \cdots) + \cdots \\ & & + \varepsilon A^{-p}(R_0 x_{-1} + R_1 x_{-2} + \cdots + R_k x_{-1-k} + \cdots), \end{array}$$

for all positive integer p.

If $x = \{x_n\}_{n=-\infty}^{0} \in \tilde{F}$, we have for all positive integer p the estimate

$$\begin{aligned}
|(Q_\varepsilon x)_{-p}| &\leq \varepsilon\|x\|\{\|A^{-1}\|(\|R_0\| + \|R_1\| + \cdots + \|R_k\| + \cdots) + \cdots \\
&\qquad +\|A^{-1}\|^p(\|R_0\| + \|R_1\| + \cdots + \|R_k\| + \cdots)\} \\
&\leq \frac{\varepsilon\|x\|}{1-\|A^{-1}\|}(1+\gamma^{-1} + \cdots + \frac{\gamma^{-k}}{k^{1-\alpha}} + \cdots) \\
&\leq \frac{\varepsilon\|x\|}{1-\|A^{-1}\|}\int_0^\infty e^{-t}t^{\alpha-1}dt = \frac{\varepsilon\|x\|}{1-\|A^{-1}\|}\Gamma(\alpha),
\end{aligned}$$

where $\Gamma(\alpha)$ is Euler's Gamma-function. Thus, for $x \in \tilde{F}$, we have $Q_\varepsilon x \in \tilde{F}$. Since Q_ε is linear, it satisfies

$$\|Q_\varepsilon x_1 - Q_\varepsilon x_2\| = \|Q_\varepsilon(x_1 - x_2)\| \leq \frac{\varepsilon\Gamma(\alpha)}{1-\|A^{-1}\|}\|x_1 - x_2\|$$

for all $x_1, x_2 \in \tilde{F}$. This implies that if $\varepsilon \in (0, \varepsilon_5]$, where $\varepsilon_5 < \dfrac{1-\|A^{-1}\|}{\Gamma(\alpha)}$, then the mapping Q_ε is a contraction on $\tilde{F}$ and thus has a unique fixed point $x^* \in \tilde{F}$. Since Q_ε is linear, we conclude that $x^* = 0$.

Let $x = \{x_n\}_{n=-\infty}^{\infty}$, $y = \{y_n\}_{n=-\infty}^{\infty} \in F$ be two solutions of (2), satisfying the condition $y_0 = x_0$ and let $x^* = \{y_n - x_n\}_{n=-\infty}^{0}$. The sequence x^* is in $\tilde{F}$ and is the fixed point of Q_ε and therefore $x^* = 0$, i.e. $y_n = x_n$, for all $n \leq 0$. This prove the uniqueness of the solution of (2) for all $n \leq 0$.

On the other hand, from (2) and for $n > 0$ we have

$$|x_n - y_n| = \varepsilon|\sum_{i=0}^{n-1} A^{n-i-1}[R_0(x_i-y_i)+R_1(x_{i-1}-y_{i-1})+\cdots+R_k(x_{i-k}-y_{i-k})+\cdots]|.$$

As $y_k = x_k$ for all integer $k \leq 0$, the above equality gives

$$\begin{aligned}
|x_n - y_n| &\leq \varepsilon\{\|A\|^{n-1}|x_0 - y_0| + \|A\|^{n-2}[|x_1 - y_1| + \gamma^{-1}|x_0 - y_0|] \\
&\qquad + \cdots + [|x_{n-1} - y_{n-1}| + \gamma^{-1}|x_{n-2} - y_{n-2}| + \cdots + \frac{\gamma^{-(n-2)}}{(n-2)^{1-\alpha}}|x_0 - y_0|]\}.
\end{aligned}$$

From this inequality, it follows that

$$|x_1 - y_1| \leq \varepsilon|x_0 - y_0| = 0,$$

$$|x_2 - y_2| \leq \varepsilon\{\|A\|[|x_0 - y_0| + |x_1 - y_1|]\} = 0,$$

etc. Therefore, by induction we have that $y_n = x_n$ for all $n > 0$.

Consequently, the solution in F of (2) is uniquely defined by the sequence $x = \{x_n(\varepsilon)\}_{n=-\infty}^{+\infty} = \{(A+\varepsilon B)^n x_0\}_{n=-\infty}^{+\infty}$.

Finally, we shall prove that for any positive integer N, there exists a positive constant C, depending on ε_0, x_0 and N, where $\varepsilon_0 = \min\{\varepsilon_i : 1 \leq i \leq 5\}$, such that $\sup\{|x_n(\varepsilon) - A^n x_0| : -N \leq n \leq N\} \leq C\varepsilon$.

Let n be an integer such that $0 \leq n < N$. By use of the binomial formula, we deduce the following estimate

$$\begin{aligned}
|x_n(\varepsilon) - A^n x_0| &\leq |[(A+\varepsilon B)^n - A^n]x_0| \leq \|(A+\varepsilon B)^n - A^n\||x_0| \\
&\leq \varepsilon |x_0| \{C_n^{n-1}\|A\|^{n-1}\|B\| + \varepsilon C_n^{n-2}\|A\|^{n-2}\|B\|^2 + \cdots + \varepsilon^{n-1}\} \\
&\leq c_1 \varepsilon,
\end{aligned}$$

where

$$c_1 = c_1(\varepsilon_0, x_0, N) = |x_0|\{C_N^{N-1}\|A\|^{N-1} + \varepsilon_0 C_N^{N-2}\|A\|^{N-2} + \cdots + \varepsilon_0^{N-1}\},$$

and

$$C_n^p = \frac{n!}{p!(n-p)!} \quad \text{for } 0 \leq p \leq n \leq N.$$

For $-N \leq n < 0$, we set $n = -k$. Then we have

$$\begin{aligned}
|x_n(\varepsilon) - A^n x_0| &\leq |[(A+\varepsilon B)^n - A^n]x_0| \leq \|(A+\varepsilon B)^n - A^n\||x_0| \\
&= |x_0|\|(A+\varepsilon B)^{-k} - A^{-k}\| = |x_0|\|[(A+\varepsilon B)^{-1}]^k - [A^{-1}]^k\| \\
&\leq |x_0|\|(A+\varepsilon B)^{-1} - A^{-1}\|\{[(A+\varepsilon B)^{-1}]^{k-1} \\
&\qquad + [(A+\varepsilon B)^{-1}]^{k-2}A^{-1} + \cdots + [A^{-1}]^{k-1}\} \\
&\leq |x_0|\|A^{-1}\|\|(I+\varepsilon A^{-1}B)^{-1} - I\|\{\|(A+\varepsilon B)^{-1}\|^{k-1} \\
&\qquad + \|A^{-1}\|\|(A+\varepsilon B)^{-1}\|^{k-2} + \cdots + \|A^{-1}\|^{k-1}\}.
\end{aligned} \tag{12}$$

As we have seen above, $\|(A+\varepsilon B)^{-1}\|^k < 1$, for all positive integer k and $\varepsilon \in (0, \varepsilon_0] \subset (0, \varepsilon_4]$. On the other hand, we have

$$\begin{aligned}
\|(I+\varepsilon A^{-1}B)^{-1} - I\| &= \|(I - (\varepsilon A^{-1}B) + (\varepsilon A^{-1}B)^2 - \cdots) - I\| \\
&\leq \varepsilon\|A^{-1}\|\{1 + \varepsilon\|A^{-1}\| + \varepsilon^2\|A^{-1}\|^2 + \cdots\} \\
&\leq \varepsilon \frac{\|A^{-1}\|}{1 - \varepsilon\|A^{-1}\|}.
\end{aligned}$$

Therefore, from (12) we obtain

$$\begin{array}{rcl} |x_n(\varepsilon) - A^n x_0| & \leq & \varepsilon |x_0| \dfrac{\|A^{-1}\|^2}{1-\varepsilon\|A^{-1}\|}\{1 + \|A^{-1}\| + \cdots + \|A^{-1}\|^{k-1}\} \\ & \leq & \varepsilon |x_0| \dfrac{\|A^{-1}\|^2}{1-\varepsilon\|A^{-1}\|} \cdot \dfrac{1}{1-\|A^{-1}\|} \\ & \leq & c_2\varepsilon, \end{array}$$

where

$$c_2 = c_2(\varepsilon_0, x_0) = |x_0| \frac{\|A^{-1}\|^2}{1-\varepsilon_0\|A^{-1}\|} \cdot \frac{1}{1-\|A^{-1}\|}.$$

Hence, the following inequality holds for $C = \min\{c_1, c_2\}$

$$\sup\{|x_n(\varepsilon) - A^n x_0| : -N \leq n \leq N\} \leq C\varepsilon.$$

This completes the proof of the theorem.

References

1. Ryabov, Yu.A.: The existence of two-sided solutions of linear integrodifferential equations of Volterra type with delay, Čas. Pěstov. Mat. 111, 2 (1986), 26-33. (In Russian.)

Typeset by LaTeX 2.09

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2000, pp. 60 – 87
Submitted: August 11, 2000. Published: December 15, 2000

Vector spaces over function fields Vector spaces over analytic function fields being associated to ordinary differential equations

J. E. Palomar Tarancón

Abstract

It is a well-known fact, that the general solution of a linear ordinary differential equation belongs to a function vector space over the real (complex) number field generated by a finite exponential function set. This circumstance allows us to handle the large-time behavior and stability problems easily. However there is a wide class of non-linear differential equations whose solutions belong to some function space being also generated by an arbitrary exponential function set; so that the stability problems can be handled as in the linear case. To discern whether a given autonomous system belongs to such an equation class, sufficient conditions have been stated in a previous article. Now, we generalize those conditions to a wider class of non-linear autonomous systems.

A.M.S. Subject Classification Codes: 46A19, 46N20, 34A25.
Keywords and Phrases: Autonomous system, analytic function field, topological vector space, integral transform, Θ-linear equation.

These generalized conditions imply the existence of an analytic function field K(X) for each smooth vector X, and a vector space V over the function field K(X) for the corresponding equation dx(t))/dt=X(x(t)), such that each solution of this equation belongs to the vector space V. Thus, a vector space V over an analytic function field K(X) is associated to every member of a wide class of autonomous systems of ordinary differential equations. The algebraic and topological structures for V and K(X) are also investigated.

Finally, by means of a useful integral transform we build the general solution for those autonomous systems satisfying our generalized conditions.

Department of Mathematics, Inst. Jaume I, C/. Europa, Castellón), Spain
E-mail: jepalomar@argen.net

1. Introduction

The large time behavior and stability problems in ordinary differential equations are not difficult to handle in linear autonomous systems, because the solution of a linear equation belongs to a vector space over the real (complex) number field generated by an exponential function set, say the function set $\left\{e^{\lambda_j} \mid j \in A\right\}$, where A stands for any finite index set. However, there is a wide class of non-linear equations each solution of which belongs to a function vector space being also generated by an exponential function set, hence they can be handled as if they were linear. In this case, we will say the non-linear system to be equivalent to some linear equation. Sufficient conditions for equivalence have been stated in [6], but the involved vector spaces are built over the real number field. As we shall show, in $\mathcal{C}^k$-differentiable manifolds whose dimension is larger than one, the results obtained in [6] can be generalized to a wide family of non-linear autonomous systems. To this end, the structure of vector space over function fields is necessary. It is not difficult to see, that for every smooth manifold $(M_n, \mathcal{A}_k)$ and each chart $(U, \varphi) \in \mathcal{A}_k$ one can associate some function field $\Theta[\mathbf{X}, U]$ to each smooth vector $\mathbf{X}$ as follows.

Let $\mathcal{C}^k(U, \mathbb{C})$ stand for the set of all $\mathcal{C}^k$-maps from U into $\mathbb{C}$. It is easy to see, that the collection of all members of $\mathcal{C}^1(U, \mathbb{C})$ being solutions of the equation

$$\mathbf{X}\phi = 0 \tag{1.1}$$

is a commutative ring, because if ϕ_1 and ϕ_2 are solutions of (1.1), then so are both $\phi_1 + \phi_2$ and $\phi_1\phi_2$. Likewise, if for every p lying in some open subset A of $\mathrm{dom}\,(\phi)$ it is true that $\phi(p) \neq 0$, besides, ϕ satisfies (1.1), then $\dfrac{1}{\phi|_A}$ is again a solution of (1.1); therefore the solutions of (1.1) form a field. However, the condition $\phi(p) \neq 0$ reduces the domain of the inverse of ϕ to the set $A = \{p \in \mathrm{dom}\,(\phi) \mid \phi(p) \neq 0\}$, but the ordinary sum and product of two functions require to have the same domain. To avoid this inconvenient we state the following conditions.

- **a):** The topology $\mathcal{T}$ for M_n is separated.
- **b):** For every $p \in U$ the vector $\mathbf{X}$ satisfies the relation $\mathbf{X}(p) \neq 0$, besides, if $(X^1, X^2, \cdots X^n)$ are the coordinates of $\mathbf{X}$ with respect to the chart (U, φ), then for every $j \leq n$ the function $X^j : \varphi[U] \to \mathbb{R}$ is analytic.
- **c):** The domain of every member of $\Theta[\mathbf{X}, U]$ belongs to the subset $\mathcal{D}(U)$ of $\wp(U)$ defined by

$$\mathcal{D}(U) = \left\{V \subseteq U \mid \overline{\overset{\circ}{V}} = U\right\} \tag{1.2}$$

accordingly if $\phi \in \Theta[\mathbf{X}, U]$, then the interior of $\mathrm{dom}\,(\phi)$ is dense in U.

Notation 1.1. *From now on, denote by $\mathcal{C}^k(\mathcal{D}(U),\mathbb{C})$ the set consisting of every $\mathcal{C}^k$-map $\phi : V \to \mathbb{C}$ whose domain V belongs to $\mathcal{D}(U)$.*

Lemma 1.2. *The family $\mathcal{D}(U)$ satisfies the following properties.*

a): *$\mathcal{D}(U)$ contains with every finite collection of sets their intersection, and with every arbitrary collection their union.*
b): *U belongs to $\mathcal{D}(U)$.*
c): *$W \subseteq V \subseteq U$ and $W \subseteq \mathcal{D}(U)$ imply $V \subseteq \mathcal{D}(U)$.*

In section §2.1 we define two composition laws $\dot{+}$ and $\star$ in $\mathcal{C}^k(\mathcal{D}(U),\mathbb{C})$, closely related with the ordinary sum and product, such that in order to apply each of these laws between two maps ϕ_1 and ϕ_2 the domains of ϕ_1 and ϕ_2 need not be the same. To this end, denote by $()^{\smile} : \mathcal{C}^k(\mathcal{D}(U),\mathbb{C}) \to \mathcal{C}^k(\mathcal{D}(U),\mathbb{C})$ the map carrying each function f into its extension by continuity. Now, define the composition $f \dot{+} g$ by $(f|_{\mathrm{dom}(f)\cap\mathrm{dom}(g)} + g|_{\mathrm{dom}(f)\cap\mathrm{dom}(g)})^{\smile}$, and likewise define $f \star g$ by $(f|_{\mathrm{dom}(f)\cap\mathrm{dom}(g)} \cdot g|_{\mathrm{dom}(f)\cap\mathrm{dom}(g)})^{\smile}$. Obviously, to apply both composition laws $\dot{+}$ and $\star$ between f and g the sets $\mathrm{dom}\,(f)$ and $\mathrm{dom}\,(g)$ need not be the same, besides, by virtue of the preceding lemma the intersection $\mathrm{dom}\,(f) \cap \mathrm{dom}\,(g)$ is non-void and it is dense in U. Analogously, define the inverse $f^\flat$ of f by $f^\flat = \left(\dfrac{1}{f|A}\right)^{\smile}$ being $A = \{p \in \mathrm{dom}\,(f) \mid f(p) \neq 0\}$.

By means of $\dot{+}$, $\star$ and $^\flat$ one can build the algebraic field structure for some subset $\Theta[\mathbf{X}, U]$ of $\mathcal{C}^k(\mathcal{D}(U),\mathbb{C})$ the members of which are analytic solutions of the equation (1.1). In Theorem 2.12, we prove that if the coefficients of an algebraic equation belong to $\Theta[\mathbf{X}, U]$, besides, they are rational functions of a fixed $g \in \Theta[\mathbf{X}, U]$, then its solutions are also members of $\Theta[\mathbf{X}, U]$.

In section §2.2, we define compatible topological structures for function fields. To this end, the ordinary functional families cannot be handled in the usual way. This is why we build these compatible topologies by means of partial seminorms. Likewise, in section §3 we briefly describe both algebraic and topological structures for those vector spaces over analytic function fields used in the proofs for the results in the next section.

Section §4 begins with a generalization of the equivalence criterion stated in [6], and consisting of two conditions, namely, Condition 1 and Condition 2, by means of which one can discern whether a given non-linear autonomous system is equivalent to some linear one, that is to say, by means of these conditions one can discern whether the general solution of a given autonomous system belongs to a vector space over a function field being generated by an arbitrary exponential function set $\mathcal{B}$. Both conditions can be summarized as follows:

If $(X^1, X^2 \cdots X^n)$ are the coordinates of the vector $\mathbf{X}$ with respect to the chart (U, φ), then for each $k \leq n$ there are an integration path C and two functions F^k and Φ^k satisfying the relations

$$\sum_{m=1}^{n} \frac{\partial}{\partial x^m} \Phi^k = h^k \in \Theta[\mathbf{X}, U] \tag{1.3}$$

$$\forall p \in U : \qquad X^k(p) = F^k\big(p, \Phi^k(p)\big) \tag{1.4}$$

besides, there are two functions $\mathcal{F}^k, g^k \in \Theta[\mathbf{X}, U]$ such that for some positive number K and every positive integer r it is true that

$$F^k(p, s) = \int_C \mathcal{F}^k(p, z) \star e^{g^k(p,z)s} dz \tag{1.5}$$

$$\left| \int_C \Big(g^k(p, z) \star h^k(p)\Big)^{(r-1)} \star \mathcal{F}^k(p, z) \star e^{g^k(p,z)s} dz \right|^{\smallsmile} \leq Kr! \tag{1.6}$$

In Theorem 4.3 we show that the general solution for each autonomous system

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{X}\left(\mathbf{x}(t)\right) \tag{1.7}$$

whose associated vector $\mathbf{X}$ satisfies the above conditions can be obtained by means of the integral transform

$$\begin{aligned} x^k_{p_0}(\gamma(t)) = x^k_0 \star 1_U \dotplus \int_C \mathcal{F}^k(p_0, z) \star \Big(g^k(p_0, z) \star h^k(p_0)\Big)^{\flat} \star \\ e^{g^k(p_0,z)\star\Phi^k(p_0)} \star \Big(e^{g^k(p_0,z)\star h^k(p_0)(t-t_0)} \dot{-} 1_U\Big) dz \end{aligned} \tag{1.8}$$

where $(x^1_{p_0}, x^2_{p_0} \cdots x^k_{p_0}) = \mathbf{x}_{p_0}$ stands for the solution of (1.7) satisfying the initial condition $\mathbf{x}_{p_0}(t_0) = \varphi(p_0)$ and 1_U denotes the map with domain U being constantly equal to 1.

Obviously, the former transform implies that the solution coordinate $x^k_{p_0}(\gamma(t))$ belongs to a vector space over the analytic function field $\Theta[\mathbf{X}, U]$ being generated by the non-countable exponential function set

$$\mathcal{B} = \left\{ e^{g^k(p_0,z)\star\Phi^k(p_0) \dotplus g^k(p_0,z)\star h^k(p_0)(t-t_0)} \mid z \in C \right\} \tag{1.9}$$

Thus, one can associate a vector space $\mathcal{V}$ over the analytic function field $\Theta[\mathbf{X}, U]$ to the ordinary differential equation (1.7). This fact justifies the title of this article.

To handle stability problems, the integral transform (1.8) is very useful, because if every value of z satisfying $\Re\big(g^k(p_0, z) \star h^k(p_0)\big) > 0$ can be neglected, then each non-periodic part of the solution vanishes when $t \mapsto \infty$; for instance, if $\mathcal{F}^k(p_0, z) \star \big(g^k(p_0, z) \star h^k(p_0)\big)^{\flat} = 0$ whenever $\Re\big(g^k(p_0, z) \star h^k(p_0)\big) > 0$.

In Corollary 4.6 we state a certain relation between those autonomous systems satisfying the above conditions and an associated partial differential equation. In the same section, we introduce the concept of Θ-linear equation, that is to say, systems being linear over the associated function field, which need not be linear over the real (complex) number field. In Theorem 4.12 we show that Θ-linear equations satisfy our generalized equivalence criterion, accordingly they can be solved by means of the integral transform (4.9).

2. Analytic function fields

2.1. **The algebraic structure.** In the set $\mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$ consisting of every continuous map $f : V \to \mathbb{C}$, whose domain V belongs to $\mathcal{D}(U)$, define an ordering $\preceq$ as follows.

Definition 2.1. *For every pair $\phi_1, \phi_2 \in \mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$, $\phi_1 \preceq \phi_2$ if and only if the following conditions hold:*

c1): $\operatorname{dom}(\phi_1) \subseteq \operatorname{dom}(\phi_2)$ *and* $\forall p \in \operatorname{dom}(\phi_1)$: $\phi_1(p) = \phi_2(p)$.

c2): *The map ϕ_2 is the extension by continuity of ϕ_1; i.e., for each p_0 in* $\operatorname{dom}(\phi_2)$ *there is a sequence $\{p_n \mid n \in \mathbb{N}\}$ in* $\operatorname{dom}(\phi_1)$*, converging to p_0, such that* $\lim\limits_{p_n \to p_0} \phi_1(p_n) = \phi_2(p_0)$.

It is worth pointing out, that since both sets $\operatorname{dom}(\phi_1)$ *and* $\operatorname{dom}(\phi_2)$ *belong to $\mathcal{D}(U)$, then for every $p_0 \in U$ there is a sequence in* $\operatorname{dom}(\phi_1)$ *converging to p_0, for the standard topology for $\mathbb{R}^n$ satisfies the first axiom of countability, and U is homeomorphic to some open subset of $\mathbb{R}^n$, besides,* $\operatorname{dom}(\phi_1)$ *is dense in U.*

By means of the ordering $\preceq$, and taking into account the well-known Zorn lemma, one can define a mapping

$$( )^{\breve{}} : \mathcal{C}^0(\mathcal{D}(U),\mathbb{C}) \to \mathcal{C}^0(\mathcal{D}(U),\mathbb{C}) \tag{2.1}$$

carrying each $\phi \in \mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$ into the maximal element $\breve{\phi}$, such that $\phi \preceq \breve{\phi}$. In other words, $\breve{}$ carries each map ϕ into its extension by continuity.

Notation 2.2. *From now on, denote by 0_U the map with domain U being constantly equal to 0; and for every map $\phi \in \mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$ denote by* $\operatorname{zero}(\phi)$ *the subset of U defined by* $\operatorname{zero}(\phi) = \{p \in \operatorname{dom}(\phi) \mid \phi(p) = 0\}$.

Definition 2.3. *Henceforth, we are concerned with those analytic manifolds whose underlying topology $\mathcal{T}$ is separated. Given such a manifold $(M_n, \mathcal{A})$ and a chart (U, φ), let $\mathcal{C}^\omega(\mathcal{D}(U),\mathbb{C}) \subset \mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$ stand for the subset consisting of each map $\phi \in \mathcal{C}^0(\mathcal{D}(U),\mathbb{C})$ such that the function $\phi \circ \varphi^{-1}$ is analytic; i.e., letting $A = \varphi[\operatorname{dom}(\phi)] \subseteq \mathbb{R}^n$, then both real functions $\Re(\phi\circ\varphi^{-1}) : A \to \mathbb{R}$ and $\Im(\phi\circ\varphi^{-1}) : A \to \mathbb{R}$ are analytic. Given a $\mathcal{C}^\omega$-vector $\mathbf{X}$ satisfying the relation $\mathbf{X}(p) \neq 0$ for every $p \in U$, let $\Theta[\mathbf{X}, U]$ stand for the subset of $\mathcal{C}^\omega(\mathcal{D}(U),\mathbb{C})$ consisting of every function being maximal with respect to $\preceq$, and satisfying the following condition: ϕ is an analytic solution of the equation*

$$\mathbf{X}\phi = 0 \tag{2.2}$$

explicitly

$$\sum_{m=1}^{n} X^m \frac{\partial}{\partial x^m}\, \widehat{\phi}(x^1, x^2 \cdots x^n) = 0 \tag{2.3}$$

where $\widehat{\phi} : \varphi[\mathrm{dom}\,(\phi)] \to \mathbb{C}$ *is the mapping defined by* $\widehat{\phi} = \phi \circ \varphi^{-1}$, *and* $(x^1, x^2 \cdots x^n)$ *are the coordinates of the homeomorphism* φ *of the considered chart* (U, φ). *It is worth pointing out, that because it is assumed* $\mathbf{X}$ *to be a* $\mathcal{C}^\omega$*-vector, then the existence of analytic solutions of (2.3) is guaranteed* [1].

Obviously, letting $\mathbf{X}\breve{\ }\phi = (\mathbf{X}\phi)\breve{\ }$ *the equation (2.2) can be written as follows,*

$$\mathbf{X}\breve{\ }\phi = 0_U \tag{2.4}$$

In the function set $\mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$ define the composition laws $\dot{+}$ and $\star$ by means of the following statements: $\forall \phi_1, \phi_2 \in \mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$,

1): $\phi_1 \dot{+} \phi_2 = (\phi_1\big|_W + \phi_2\big|_W)\breve{\ }$.
2): $\phi_1 \star \phi_2 = (\phi_1\big|_W \cdot \phi_2\big|_W)\breve{\ }$.

where $W = \mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2)$. Since for every $\mathcal{C}^\omega$-vector $\mathbf{X}$ the set $\Theta[\mathbf{X}, U]$ belongs to $\mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$, then both laws $\dot{+}$ and $\star$ are also defined for $\Theta[\mathbf{X}, U]$.

Now, let $()^\flat : (\Theta[\mathbf{X}, U] - \{0_U\}) \to (\Theta[\mathbf{X}, U] - \{0_U\})$ stand for the map carrying each $\phi \in (\Theta[\mathbf{X}, U] - \{0_U\})$ into $\phi^\flat = \left(\dfrac{1}{\phi|_V}\right)\breve{\ }$, where

$$V = \{p \in \mathrm{dom}\,(\phi) \mid \phi(p) \neq 0\} = \mathrm{dom}\,(\phi) - \mathrm{zero}\,(\phi)$$

It is worth pointing out, that according to the definition of the mapping $^\flat$, for every $\phi \in \Theta[\mathbf{X}, U]$ the image under $^\flat$ belongs to $\Theta[\mathbf{X}, U]$, but Definition 2.3 requires that $\mathrm{dom}\left(\phi^\flat\right) \in \mathcal{D}(U)$, and this relation cannot be satisfied whenever the interior of $\mathrm{zero}\,(\phi)$ does not vanish. As the Lemma 2.7 shows, the required analyticity for the members of $\Theta[\mathbf{X}, U]$ implies $(\mathrm{zero}\,(\phi))^\circ = \emptyset$.

Remark 2.4. *For one-dimensional differentiable manifolds the system 2.3 consists of the only equation* $X^1 \frac{\partial}{\partial x^1}\widehat{\phi}(x^1) = 0$, *hence* $\widehat{\phi}$ *is a constant map and* $\Theta[\mathbf{X}, U]$ *can be identified with the complex number field.*

Lemma 2.5. *For every couple* $\phi_1, \phi_2 \in \Theta[\mathbf{X}, U]$ *the following is true,*

a): $\mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2) \subseteq \mathrm{dom}\,(\phi \dot{+} \phi_2)$.
b): $\mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2) \subseteq \mathrm{dom}\,(\phi \star \phi_2)$.

besides, for each $\phi \in \Theta[\mathbf{X}, U]$ *the following relation holds:*

$$\mathrm{dom}\,(\phi) - \mathrm{zero}\,(\phi) \subseteq \mathrm{dom}\left(\phi^\flat\right)$$

Proof. It follows immediately from the definitions of the laws $\dot{+}$, $\star$ and $^\flat$. □

Lemma 2.6. *For every vector* $\mathbf{X}$ *each constant map* $\phi : U \to \mathbb{C}$ *belongs to* $\Theta[\mathbf{X}, U]$, *besides, if a map* $\phi : V \subseteq U \to$ *lying in* $\Theta[\mathbf{X}, U]$ *is constantly equal to* $c \in \mathbb{C}$, *then* $V = U$.

Proof. On the one hand, every constant map satisfies equation (2.4), besides, if the domain is U, then it satisfies Definition 2.3. On the other hand, let $\phi : V \in \mathcal{D}(U) \to \mathbb{C}$ be a map being constantly equal to $c \in \mathbb{C}$. Since V is dense in U, then for each $p_0 \in U$ there is a sequence $(p_n)_{n\in\mathbb{N}}$ in V which converges to p_0; so that

$$\lim_{n\to\infty} \phi(p_n) = c \tag{2.5}$$

and because ϕ is maximal with respect to $\preceq$, then $\phi(p_0) = c$, therefore $p_0 \in \mathrm{dom}\,(\phi)$ and $\mathrm{dom}\,(\phi) = U$. $\square$

Lemma 2.7. *For every member ϕ of $\Theta[\mathbf{X}, U]$ the following statements hold.*

a): *If there is an open subset V of U such that $\forall p \in V$, $\phi(p) = 0$, then $\phi = 0_U$; consequently, if $\phi \neq 0_U$, then $(\mathrm{zero}\,(\phi))^\circ = \emptyset$.*
b): *If $\phi \neq 0_U$, then $U - \mathrm{zero}\,(\phi) \in \mathcal{D}(U)$.*
c): *If $\phi \neq 0_U$, then $A = \mathrm{dom}\,(\phi) \cap (U - \mathrm{zero}\,(\phi)) \in \mathcal{D}(U)$.*

Proof. **a**) It follows immediately from the required analyticity of each member of $\Theta[\mathbf{X}, U]$.

b) On the one hand, because the topology $\mathcal{T}$ of the considered analytic manifold $(M_n, \mathcal{A})$ is assumed to be separable, then $\mathrm{zero}\,(\phi)$ is closed and $U - \mathrm{zero}\,(\phi)$ is open in the relative topology $\mathcal{T}\big|_U$. On the other hand, if $E = U - \mathrm{zero}\,(\phi)$ is not dense in U, then there is $p \in \mathrm{zero}\,(\phi)$ and an open neighborhood B_p of p such that $B_p \cap E = \emptyset$, therefore $B_p \subseteq \mathrm{zero}\,(\phi)$, and taking into account the preceding statement it follows that $\phi = 0_U$, which contradicts the assumption $\phi \neq 0_U$, hence E must be dense in U and because it is open, then $E \in \mathcal{D}(U)$.

c) By Definition 2.3 $\mathrm{dom}\,(\phi)$ belongs to $\mathcal{D}(U)$, and by the preceding statement we know that $U - \mathrm{zero}\,(\phi)$ is also a member of $\mathcal{D}(U)$, therefore taking into account Lemma 1.2, $A \in \mathcal{D}(U)$. $\square$

Remark 2.8. *From now on, for every $c \in \mathbb{C}$ denote by c_U the map with domain U being constantly equal to c. Likewise, denote by $\mathbb{C}_U$ the subset $\{c_U \mid c \in \mathbb{C}\}$ of $\Theta[\mathbf{X}, U]$.*

Corollary 2.9. *If a map $\phi \in \Theta[\mathbf{X}, U]$ is constantly equal to $c \in \mathbb{C}$ over an open subset V of U, then ϕ is equal to c at every point of U, therefore $\mathrm{dom}\,(\phi) = U$.*

Proof. Let $-c_U$ stand for the map with domain U being constantly equal to $-c$. Thus, $\psi = \phi \dotplus (-c_U) = 0$ at every point of V. Now, since V is open and $V \subseteq \mathrm{zero}\,(\psi)$, then $U - \mathrm{zero}\,(\psi)$ is not dense in U, therefore from Lemma 2.7 it follows, that $\psi = 0_U$ and $\phi = c_U$. $\square$

Lemma 2.10. *The set $\Theta[\mathbf{X}, U]$ together with the laws $\dotplus$, $\star$ and $^\flat$ form a field.*

Proof. Let ϕ_1 and ϕ_2 be two members of $\Theta[\mathbf{X}, U]$. On the one hand, by Lemma 1.2 one can obtain easily, that $W = \mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2)$ belongs to $\mathcal{D}(U)$, hence taking into account Lemma 2.5, we have that $\mathrm{dom}\,(\phi_1 \dotplus \phi_2) \supseteq$

W and $\operatorname{dom}(\phi_1 \star \phi_2) \supseteq W$ are also members of $\mathcal{D}(U)$. On the other hand, by Definition 2.3

$$(\mathbf{X}(\phi_1 \dotplus \phi_2))^{\smallsmile} = \mathbf{X}^{\smallsmile}\phi_1 \dotplus \mathbf{X}^{\smallsmile}\phi_2 = 0_U \tag{2.6}$$

$$(\mathbf{X}(\phi_1 \star \phi_2))^{\smallsmile} = (\mathbf{X}^{\smallsmile}\phi_1) \star \phi_2 \dotplus \phi_1 \star \mathbf{X}^{\smallsmile}\phi_2 = 0_U \tag{2.7}$$

therefore both maps $\phi_1 \dotplus \phi_2$ and $\phi_1 \star \phi_2$ satisfy Definition 2.3. Taking into account Lemma 2.6, for each $\phi \in \Theta[\mathbf{X}, U]$ it is true that $\phi \dotplus (-\phi) = 0_U$. The remaining properties defining the ring structure are immediate.

Now, for every $\phi \in \Theta[\mathbf{X}, U]$, $\operatorname{dom}(\phi^\flat)$ contains the set $\operatorname{dom}(\phi) - \operatorname{zero}(\phi)$, hence by virtue of Lemma 2.7 we have that $\operatorname{dom}(\phi^\flat)$ belongs to $\mathcal{D}(U)$, because $\operatorname{dom}(\phi^\flat) \supseteq \operatorname{dom}(\phi) - \operatorname{zero}(\phi) = \operatorname{dom}(\phi) \cap (U - \operatorname{zero}(\phi))$. Denoting by 1_U the map with domain U being constantly equal to 1, by definition, for each $p \in \operatorname{dom}(\phi) \cap \operatorname{dom}(\phi^\flat)$, $\phi(p)\phi^\flat(p) = 1$, therefore taking into account Lemma 2.6, $\phi \star \phi^\flat = 1_U$, and $\phi^\flat$ is the inverse of ϕ. Finally, for every $p \in (\operatorname{dom}(\phi) - \operatorname{zero}(\phi))$

$$\frac{-1}{\phi^2(p)}\mathbf{X}_p\phi(p) = \mathbf{X}_p\phi^\flat(p) = 0 \tag{2.8}$$

hence

$$\mathbf{X}^{\smallsmile}\phi^\flat = 0_U \tag{2.9}$$

therefore $\phi^\flat$ satisfies Definition 2.3 and $\phi^\flat \in \Theta[\mathbf{X}, U]$. $\square$

Definition 2.11. *Notice, that by virtue of Lemma 2.6, for each chart* (U, φ) *and every smooth vector* $\mathbf{X}$, *the set* $\mathbb{C}_U = \{c_U : U \to \mathbb{C} \mid c \in \mathbb{C}\}$ *consisting of every constant map belongs to* $\Theta[\mathbf{X}, U]$, *besides,* $\mathbb{C}_U$ *equipped with the relative laws* $\dotplus$, $\star$ *and* $^\flat$, *is a field being isomorphic to the complex number field. Thus, the analytic function field* $\left(\Theta[\mathbf{X}, U], \dotplus, \star, ^\flat\right)$ *can be regarded as a vector space over* $\left(\mathbb{C}_U, \dotplus, \star, ^\flat\right)$, *which we shall say to be the underlying vector space of* $\left(\Theta[\mathbf{X}, U], \dotplus, \star, ^\flat\right)$.

Theorem 2.12. *If each of the coefficients* $a_0, a_1 \cdots a_m$ *of the algebraic equation*

$$P\big(z\big) = a_0 \dotplus a_1 \star z \cdots \dotplus a_m \star (z)^{\smallsmile m} = 0_U \tag{2.10}$$

belongs to the function field $\Theta[\mathbf{X}, U]$, *besides, there are a fixed* $g \in \Theta[\mathbf{X}, U]$ *and for each* $j \leq m$ *a rational function* $h_j : \mathbb{C} \to \mathbb{C}$ *such that* $a_j = h_j \circ g$, *then each solution of (2.10) belongs to* $\Theta[\mathbf{X}, U]$; *where*

$$(z)^{\smallsmile m} = \underset{1}{z} \star \underset{2}{z} \cdots \star \underset{m}{z}$$

Proof. First, assume that $P(z)$ has no multiple root. In this case, taking into account equation (2.3), from (2.10) we obtain that

$$\begin{aligned} \mathbf{X}^{\smallsmile}(a_0 \dotplus a_1 \star z \cdots \dotplus a_m \star (z)^{\smallsmile m}) &= \\ &\Big(a_1 \dotplus 2a_2 \star z \cdots \dotplus m a_m \star (z)^{\smallsmile (m-1)}\Big)\, \mathbf{X}^{\smallsmile} z = 0_U \end{aligned} \tag{2.11}$$

and because we are assuming that $P(z)$ has no multiple root, then it is true that $\left(a_1 \dotplus 2a_2 \star z \cdots \dotplus ma_m \star (z)^{\smile(m-1)}\right) \neq 0_U$, hence from the equation above one can obtain that $\mathbf{X}\breve{z} = 0_U$, accordingly z satisfies equation (2.3).

Conversely, if $P(z)$ has some multiple root, then denoting by $Q(z)$ the greatest common factor of $P(z)$ and its derivative, the equation $\dfrac{P(z)}{Q(z)} = 0_U$ has the same solutions as (2.10), however it has not multiple root and the proof goes as in the preceding case.

It remains to be shown that $\operatorname{dom}(z)$ belongs to $\mathcal{D}(U)$. The equation (2.10) is defined at every point p lying in the set

(2.12)
$$E = \left(\bigcap_{0\le k\le m} \operatorname{dom}\left(a_k(p)\right)\right) \bigcap \left(\bigcap_{1\le k\le m} \left(U - \operatorname{zero}\left(a_k(p)\right)\right)\right) \in \mathcal{D}(U)$$

therefore, each solution $z(p)$ of (2.10) is defined for every $p \in E$. Thus, by continuity, we have that $(P(z^{\smile}))^{\smile} = 0_U$, so that each solution z of 2.10 is maximal, besides, $\operatorname{dom}(z)$ contains E, and taking into account Lemma 1.2, it follows that $\operatorname{dom}(z)$ belongs to $\mathcal{D}(U)$.

Finally, because each of the h_j is assumed to be rational, then it is a known fact that each solution $f(g)$ of the equation

$$(2.13)\quad h_0(g) \dotplus h_1(g) \star f(g) \cdots \dotplus h_m(g) \star (f(g))^{\smile m} = \\ a_0 \dotplus a_1 \star z \cdots \dotplus a_m \star (z)^{\smile m} = 0_U$$

is an analytic function of g, and because g is assumed to be a member of $\Theta[\mathbf{X}, U]$, then the composition $z = f \circ g$ is also an analytic function. $\square$

2.2. **Topological function fields.** Since for each $\mathcal{C}^\omega$-vector $\mathbf{X}$ the field $(\mathbb{C}, \dotplus, \star, ^\flat)$ belongs to each analytic function field $\left(\Theta[\mathbf{X}, U], \dotplus, \star, ^\flat\right)$, then every topology $\mathfrak{T}$ being compatible with the algebraic field structure of $\left(\Theta[\mathbf{X}, U], \dotplus, \star, ^\flat\right)$ is also compatible with its underlying vector space structure described in Definition 2.11. However, because the domains of two members of $\Theta[\mathbf{X}, U]$ need not be the same, then such a topology $\mathfrak{T}$ cannot be built from the ordinary functionals or seminorms in the usual way.

Notation 2.13. *Given any member A of $\mathcal{D}(U)$ denote by $\mathcal{R}(\mathbf{X}, A)$ the subset of $\Theta[\mathbf{X}, U]$ consisting each map $\phi \in \Theta[\mathbf{X}, U]$ whose domain contains A. By Lemma 2.6 we know that every constant map in $\Theta[\mathbf{X}, U]$ belongs to $\mathcal{R}(\mathbf{X}, A)$, for the domain of each constant map is $U \supseteq A$. As a immediate consequence of Definition 2.3 we have that the set $\mathcal{R}(\mathbf{X}, A)$ together with both relative laws $\dotplus$ and $\star$ form a function ring $(\mathcal{R}(\mathbf{X}, A), \dotplus, \star)$. Since $\mathcal{R}(\mathbf{X}, A)$ contains each constant map from U into $\mathbb{C}$, it is not difficult to see that $(\mathcal{R}(\mathbf{X}, A), \dotplus, \star)$ can be regarded as a vector space over the field $(\mathbb{C}, \dotplus, \star, ^\flat)$, that we shall say to be the underlying vector space of $(\mathcal{R}(\mathbf{X}, A), \dotplus, \star)$; i.e. the subspace of the vector space described in Definition 2.11 whose underlying set is $\mathcal{R}(\mathbf{X}, A)$.*

Definition 2.14. *For each* $p \in A$ *the mapping* $q_p : \mathcal{R}(\mathbf{X}, A) \to [0, \infty)$ *defined by*

$$\forall \phi \in \mathcal{R}(\mathbf{X}, A) : \qquad q_p(\phi) = |\phi(p)| \tag{2.14}$$

is an ordinary seminorm with respect to the underlying vector space structure of $\mathcal{R}(\mathbf{X}, A)$*. Thus, the family of seminorms* $\mathcal{S}(A) = \{q_p \mid p \in A\}$ *induces a topology* $\mathfrak{T}_A$ *for* $\mathcal{R}(\mathbf{X}, A)$ *such that* $(\mathcal{R}(\mathbf{X}, A), \mathfrak{T}_A)$ *is a locally convex space. If the mapping* $m_{\mathcal{R}(\mathbf{X},A)} : \mathcal{R}(\mathbf{X}, A) \to \Theta[\mathbf{X}, U]$ *is the canonical inclusion, then the pair* $(m_{\mathcal{R}(\mathbf{X}A)}, q_p)$*, explicitly*

$$\Theta[\mathbf{X}, U] \xleftarrow{m_{\mathcal{R}(\mathbf{X},A)}} \mathcal{R}(\mathbf{X}, A) \xrightarrow{q_p} [0, \infty)$$

can be denoted by partial seminorm in $\Theta[\mathbf{X}, U]$. *Given a partial seminorm family* $\mathfrak{M} = \big\{(m_{R(\mathbf{X},A)}, q_p) \mid A \in \mathcal{D}(U), p \in A\big\}$*, let* $\mathfrak{T}_{\mathfrak{M}}$ *denote the topology for* $\Theta[\mathbf{X}, U]$ *a subbase of which is the set collection*

$$\mathfrak{S} = \{V_{p,\varepsilon}(\phi_0) \mid p \in U, \varepsilon > 0, \phi_0 \in \Theta[\mathbf{X}, U]\}$$

where each of the $V_{p,\varepsilon}(\phi_0)$ *stands for the subset of* $\Theta[\mathbf{X}, U]$ *defined by*

$$\begin{aligned} V_{p,\varepsilon}(\phi_0) = \Big\{\phi \in \Theta[\mathbf{X}, U] \Big| \\ \forall A \in \mathcal{D}(U), \exists p_0 \in \operatorname{dom}(\phi_0) \cap \operatorname{dom}(\phi) \cap A : \\ \|\varphi(p_0) - \varphi(p)\| < \varepsilon \text{ and } q_{p_0}\big(\phi \dot{-} \phi_0\big) < \varepsilon\Big\} \end{aligned} \tag{2.15}$$

where $\varphi : U \to \mathbb{R}^n$ *is the homeomorphism of the considered chart* (U, φ)*, and* $\phi \dot{-} \phi_0$ *denotes the composition* $\phi \dot{+} (-\phi_0)$*. Obviously, the seminorm* q_{p_0} *occurring in (2.15) is defined in the ring* $\mathcal{R}(\mathbf{X}, B)$ *being* B *the intersection* $\operatorname{dom}(\phi_0) \cap \operatorname{dom}(\phi) \cap A$*.*

Remark 2.15. *From the preceding definition it follows immediately, that a net* $S = \{\phi_\delta \mid \delta \in (D, \leq)\}$ *in* $\Theta[\mathbf{X}, U]$ $\mathfrak{T}_{\mathfrak{M}}$*-converges to* ϕ_0 *if and only if the following condition holds: For each* $p \in U$ *and every* $\varepsilon > 0$ *there exists* $\delta_0 \in D$ *such, that for every* $\delta \geq \delta_0$ *and each* $A \in \mathcal{D}(U)$ *there is* $p_\delta \in A \cap \operatorname{dom}(\phi_d) \cap \operatorname{dom}(\phi_0)$ *satisfying the following relations*

$$\begin{cases} \|\varphi(p_\delta) - \varphi(p)\| < \varepsilon \\ q_{p_\delta}(\phi_\delta \dot{-} \phi_0) = \big|\phi_\delta(p_\delta) \dot{-} \phi_0(p_\delta)\big| < \varepsilon \end{cases} \tag{2.16}$$

Theorem 2.16. *The topology* $\mathfrak{T}_{\mathfrak{M}}$ *is compatible with the algebraic structure of* $\big(\Theta[\mathbf{X}, U], \dot{+}, \star, ^\flat\big)$*.*

Proof. To see the continuity of the map

$$(\,) \dot{+} (\,) : \Theta[\mathbf{X}, U] \times \Theta[\mathbf{X}, U] \to \Theta[\mathbf{X}, U]$$

it must be shown that for every pair $\phi_1, \phi_2 \in \Theta[\mathbf{X}, U]$ and each neighborhood $V_{p,\varepsilon}(\phi_1 \dot{+} \phi_2)$ of $\phi_1 \dot{+} \phi_2$ there are a neighborhood $V_{p,\varepsilon_1}(\phi_1)$ of ϕ_1 and a neighborhood $V_{p,\varepsilon_1}(\phi_2)$ of ϕ_2 such that

$$V_{p,\varepsilon_1}(\phi_1) \dot{+} V_{p,\varepsilon_1}(\phi_2) \subset V_{p,\varepsilon}(\phi_1 \dot{+} \phi_2) \tag{2.17}$$

On the one hand, $\psi_1 \in V_{p,\varepsilon_1}(\phi_1)$ implies that for each positive number $\varepsilon_1 < \varepsilon/6$ and every $A_1 \in \mathcal{D}(U)$ there is $p_1 \in A_1 \cap \operatorname{dom}(\phi_1) \cap \operatorname{dom}(\psi_1)$ such that

$$\begin{cases} \|\varphi(p_1) - \varphi(p)\| < \varepsilon_1 < \varepsilon/6 \\ \left|\psi_1(p_1) \dot{-} \phi_1(p_1)\right| < \varepsilon_1 < \varepsilon/6 \end{cases} \tag{2.18}$$

Analogously, the relation $\psi_2 \in V_{p,\varepsilon_1}(\phi_2)$ implies that for every A_2 in $\mathcal{D}(U)$ there is $p_2 \in A_2 \cap \operatorname{dom}(\phi_2) \cap \operatorname{dom}(\psi_2)$ such that

$$\begin{cases} \|\varphi(p_2) - \varphi(p)\| < \varepsilon_1 < \varepsilon/6 \\ \left|\psi_2(p_2) \dot{-} \phi_2(p_2)\right| < \varepsilon_1 < \varepsilon/6 \end{cases} \tag{2.19}$$

On the other hand, setting $A_1 = A \cap \operatorname{dom}(\phi_2) \cap \operatorname{dom}(\psi_2)$ and $A_2 = A \cap \operatorname{dom}(\phi_1) \cap \operatorname{dom}(\psi_1)$ for any $A \in \mathcal{D}(U)$, then, by continuity and taking into account that φ is a homeomorphism, for every $\varepsilon/6 > 0$ there is ε_1 such that the relations

$$\begin{cases} \|\varphi(p_1) - \varphi(p)\| \, \varepsilon_1 < \varepsilon/6 \\ \|\varphi(p_2) - \varphi(p)\| \, \varepsilon_1 < \varepsilon/6 \\ \|\varphi(p_1) - \varphi(p_2)\| \le \|\varphi(p_1) - \varphi(p)\| + \|\varphi(p_2) - \varphi(p)\| < 2\varepsilon_1 \end{cases} \tag{2.20}$$

imply the following

$$\begin{cases} \left|\phi_2(p_1) \dot{-} \phi_2(p_2)\right| < \varepsilon/6 \\ \left|\psi_2(p_1) \dot{-} \psi_2(p_2)\right| < \varepsilon/6 \end{cases} \tag{2.21}$$

hence

$$\begin{aligned} \left|\psi_2(p_1) \dot{-} \phi_2(p_1)\right| \le \left|\psi_2(p_1) \dot{-} \psi_2(p_2)\right| + \\ \left|\psi_2(p_2) \dot{-} \phi_2(p_2)\right| + \left|\phi_2(p_2) \dot{-} \phi_2(p_1)\right| < \varepsilon/2 \end{aligned} \tag{2.22}$$

therefore from (2.18) and (2.22) together with Lemma 2.5 it follows, that for each $A \in \mathcal{D}(U)$ there is p_1 lying in

$$\begin{aligned} A \cap \operatorname{dom}(\phi_1) \cap \operatorname{dom}(\phi_2) \cap \operatorname{dom}(\psi_1) \cap \operatorname{dom}(\psi_2) \subseteq \\ A \cap \operatorname{dom}(\phi_1 \dot{+} \phi_2) \cap \operatorname{dom}(\psi_1 \dot{+} \psi_2) \end{aligned}$$

satisfying the relations,

$$\begin{cases} \|\varphi(p_1) - \varphi(p)\| < \varepsilon_1 < \varepsilon/6 < \varepsilon \\ \left|(\psi_1(p_1) \dot{+} \psi_2(p_1)) \dot{-} (\phi_1(p_1) \dot{+} \phi_2(p_1))\right| < \varepsilon \end{cases} \tag{2.23}$$

accordingly $(\psi_1 \dot{+} \psi_2)$ belongs to $V_{p,\varepsilon}(\phi_1 \dot{+} \phi_2)$ and because both maps ψ_1 and ψ_2 are arbitrary, then the relation (2.17) follows.

Likewise, setting $\varepsilon = \varepsilon_0$ in (2.18) and (2.22) we obtain, that

$$\begin{cases} \|\varphi(p_1) - \varphi(p)\| < \varepsilon_1 < \varepsilon_0/6 \\ \left|\psi_1(p_1) \dot{-} \phi_1(p_1)\right| < \varepsilon_1 < \varepsilon_0/6 \\ \left|\psi_2(p_1) \dot{-} \phi_2(p_1)\right| < \varepsilon_0/2 \end{cases} \tag{2.24}$$

therefore

$$\left|(\psi_1(p_1)\star\psi_2(p_1)) \dot{-} (\phi_1(p_1)\star\phi_2(p_1))\right| < (|\psi_2(p_1)|+|\phi_1(p_1)|)\,\varepsilon_0 \tag{2.25}$$

and from (2.24) one can obtain $|\psi_2(p_1)| < \varepsilon_0/2+|\phi_2(p_1)|$. Using this relation in (2.25) it follows, that

$$\left|(\psi_1(p_1)\star\psi_2(p_1)) \dot{-} (\phi_1(p_1)\star\phi_2(p_1))\right| < (|\phi_2(p_1)|+|\phi_1(p_1)|+\varepsilon_0/2)\,\varepsilon_0/6 \tag{2.26}$$

therefore, setting $(|\phi_2(p_1)|+|\phi_1(p_1)|+\varepsilon_0/2)\,\varepsilon_0/6=\varepsilon$ it follows that $\psi_1\star\psi_2$ belongs to $V_{p,\varepsilon}(\phi_1\star\phi_2)$, hence one can deduce that for each neighborhood $V_{p,\varepsilon}(\phi_1\star\phi_2)$ of $\phi_1\star\phi_2$ there are $V_{p,\varepsilon_1}(\phi_1)$ and $V_{p,\varepsilon_1}(\phi_2)$ satisfying the relations

$$V_{p,\varepsilon_1}(\phi_1)\star V_{p,\varepsilon_1}(\phi_2)\subset V_{p,\varepsilon}(\phi_1\star\phi_2) \tag{2.27}$$

and the mapping $(\,)\star(\,):\Theta[\mathbf{X},U]\times\Theta[\mathbf{X},U]\to\Theta[\mathbf{X},U]$ is continuous.

Finally, consider a net $S=\{\phi_\delta\mid(D,\leq)\}$ converging to ϕ_0 and assume that $\phi_0\neq 0_U$, accordingly for every $p\in U$ and each $\varepsilon>0$ there is $\delta_0\in D$ such that $\forall\delta\geq\delta_0$ and $\forall A\in\mathcal{D}(U)$ there is $p_\delta\in A\cap\operatorname{dom}(\phi_\delta)\cap\operatorname{dom}(\phi_0)$ satisfying the following relations

$$\begin{cases}\|\varphi(p_\delta)-\varphi(p)\|<\varepsilon\\ \left|\phi_\delta(p_\delta)\dot{-}\phi_0(p_\delta)\right|<\varepsilon\end{cases} \tag{2.28}$$

Letting $A=U-\operatorname{zero}(\phi_0)$ by Lemma 2.7 we know that $A\in\mathcal{D}(U)$, besides $p_\delta\in A$ implies that

$$|\phi_0(p_\delta)|>0 \tag{2.29}$$

and from (2.28) one can obtain, that

$$\begin{aligned}\left|\phi_\delta^\flat(p_\delta)\dot{-}\phi_0^\flat(p_\delta)\right| &= \\ &\left|\left((\phi_\delta(p_\delta)\dot{-}\phi_0(p_\delta)\right)\star\phi_\delta^\flat(p_\delta)\star\phi_0^\flat(p_\delta)\right| \leq \\ &\qquad\left|\phi_\delta(p_\delta)\dot{-}\phi_0(p_\delta)\right|\,|\phi_\delta(p_\delta)|^{-1}\,|\phi_0(p_\delta)|^{-1}\end{aligned} \tag{2.30}$$

therefore setting $\varepsilon<|\phi_0(p_\delta)|$, from the relations (2.28) and (2.29) we obtain that $|\phi_\delta(p_\delta)|>|\phi_0(p_\delta)|-\varepsilon>0$. Using this relation in (2.30) it follows that

$$\forall\delta\geq\delta_0:\qquad\left|\phi_\delta^\flat(p_\delta)\dot{-}\phi_0^\flat(p_\delta)\right|<\frac{\varepsilon}{|\phi_0(p_\delta)|\,(|\phi_0(p_\delta)|-\varepsilon)} \tag{2.31}$$

and now it is easy to see, that the net $\{\phi_\delta^\flat\mid\delta\in D\leq\}$ converges to $\phi_0^\flat$. □

Lemma 2.17. *If $S=\{\phi_\delta\mid\delta\in(D,\leq)\}$ is a net in $\Theta[\mathbf{X},U]$ such that there are a function ring $(\mathcal{R}(\mathbf{X},A),\dot{+},\star)$ and a $\delta_0\in D$ satisfying the relation*

$$\forall\delta\geq\delta_0:\qquad\phi_\delta\in\mathcal{R}(\mathbf{X},A) \tag{2.32}$$

then the net S $\mathfrak{T}_{\mathfrak{M}}$-converges to some function ϕ_0 in $\Theta[\mathbf{X},U]$ if and only if S $\mathfrak{T}_A$-converges to ϕ_0.

Proof. First, suppose that S $\mathfrak{T}_A$-converges to ϕ_0. On the one hand, $\forall p \in U$, $\forall \varepsilon > 0$, $\forall B \in \mathcal{D}(U)$, $\forall \delta \in D$, by virtue of Lemma 1.2 together with Definition 2.3 there is $p_0 \in B \cap \mathrm{dom}\,(\phi_\delta) \cap \mathrm{dom}\,(\phi_0) \cap A \subseteq B \cap \mathrm{dom}\,(\phi_\delta) \cap \mathrm{dom}\,(\phi_0)$ satisfying the relation

$$\|\varphi(p_0) - \varphi(p)\| < \varepsilon \tag{2.33}$$

On the other hand, because it is assumed that the net S $\mathfrak{T}_A$-converges to ϕ_0, for every $p_0 \in A$ there is $\delta_1 \geq \delta_0$, such that

$$\forall \delta \geq \delta_1 : \qquad \left|\phi_\delta(p_0) \dot{-} \phi_0(p_0)\right| < \varepsilon \tag{2.34}$$

accordingly setting $p_\delta = p_0$ we have that $\forall B \in \mathcal{D}(U)$, $\forall \delta \geq \delta_1$ there is $p_\delta \in B \cap \mathrm{dom}\,(\phi_\delta) \cap \mathrm{dom}\,(\phi_0)$ satisfying the relations

$$\begin{cases} \|\varphi(p_\delta) - \varphi(p)\| < \varepsilon \\ \left|\phi_\delta(p_\delta) \dot{-} \phi_0(p_\delta)\right| < \varepsilon \end{cases} \tag{2.35}$$

so that S $\mathfrak{T}_{\mathfrak{M}}$-converges to ϕ_0.

Conversely, suppose that the net S $\mathfrak{T}_{\mathfrak{M}}$-converges to ϕ_0; accordingly, because $A \in \mathcal{D}(U)$, then $\forall p \in A \subseteq U$, $\forall \varepsilon_1 > 0$, there is $\delta_1 \geq \delta_0$ such that for each $\delta \geq \delta_1$, there is p_δ lying in $A \cap \mathrm{dom}\,(\phi_\delta) \cap \mathrm{dom}\,(\phi_0)$ satisfying the relations

$$\begin{cases} \|\varphi(p_\delta) - \varphi(p)\| < \varepsilon_1 \\ \left|\phi_\delta(p_\delta) \dot{-} \phi_0(p_\delta)\right| < \varepsilon_1 \end{cases} \tag{2.36}$$

By continuity for every ε_2 there is $\varepsilon_1 < \varepsilon_2$ such that the relation

$$\|\varphi(p_\delta) - \varphi(p)\| < \varepsilon_1 < \varepsilon_2$$

implies

$$\begin{cases} \left|\phi_\delta(p_\delta) \dot{-} \phi_\delta(p)\right| < \varepsilon_2 \\ \left|\phi_0(p_\delta) \dot{-} \phi_0(p)\right| < \varepsilon_2 \end{cases} \tag{2.37}$$

Finally, setting $\varepsilon_2 = \varepsilon/3 > 0$ from (2.36) and (2.37) one can deduce that

$$\begin{aligned} \forall \delta \geq \delta_1 : \qquad & \left|\phi_\delta(p) \dot{-} \phi_0(p)\right| \leq \\ & \left|\phi_\delta(p) \dot{-} \phi_\delta(p_\delta)\right| + \left|\phi_\delta(p_\delta) \dot{-} \phi_0(p_\delta)\right| + \left|\phi_0(p_\delta) \dot{-} \phi_0(p)\right| < \varepsilon \end{aligned} \tag{2.38}$$

accordingly, the net S $\mathfrak{T}_A$-converges to ϕ_0. □

Corollary 2.18. *For every $A \in \mathcal{D}(U)$ the inclusion*

$$m_{\mathcal{R}(\mathbf{X},A)} : (\mathcal{R}(\mathbf{X}, A), \mathfrak{T}_A) \to (\Theta[\mathbf{X}, U], \mathfrak{T}_{\mathfrak{M}})$$

is a continuous map.

Proof. By Lemma 2.17, if a net S $\mathfrak{T}_A$-converges to $\phi \in \mathcal{R}(\mathbf{X}, A)$, then the net $m_{\mathcal{R}(A)} \circ S$ $\mathfrak{T}_{\mathfrak{M}}$-converges to $m_{\mathcal{R}(A)}(\phi) = \phi$. □

Theorem 2.19. $(\Theta[\mathbf{X}, U], \mathfrak{T}_{\mathfrak{M}})$ *is Hausdorff space.*

Proof. Let $S = \{\phi_\delta \mid \delta \in (D, \leq)\}$ be a net which converges to both maps ϕ_1 and ϕ_2, and suppose that there is $p \in \mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2)$ such that

$$\left|\phi_1(p) \dot{-} \phi_2(p)\right| = a > 0 \tag{2.39}$$

By convergence, for each ε_1 there is $\delta_0 \in D$ such that for every $\delta \geq \delta_0$ there exist $p_1, p_2 \in \mathrm{dom}\,(\phi_\delta) \cap \mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2)$ satisfying the relations

$$\begin{cases} \|\varphi(p_1) - \varphi(p)\| < \varepsilon_1/2 \\ \|\varphi(p_2) - \varphi(p)\| < \varepsilon_1/2 \\ \left|\phi_\delta(p_1) \dot{-} \phi_1(p_1)\right| < \varepsilon_1/2 \\ \left|\phi_\delta(p_2) \dot{-} \phi_2(p_2)\right| < \varepsilon_1/2 \end{cases} \tag{2.40}$$

hence

$$\|\varphi(p_1) - \varphi(p_2)\| < \varepsilon_1 \tag{2.41}$$

Now, by continuity, for each $\varepsilon_0 > 0$ there is $\varepsilon_1 < \varepsilon_0$ such that the relations (2.40) and (2.41) imply the following

$$\begin{cases} \left|\phi_1(p_1) \dot{-} \phi_1(p)\right| < \varepsilon_0 \\ \left|\phi_2(p_1) \dot{-} \phi_2(p)\right| < \varepsilon_0 \\ \left|\phi_2(p_1) \dot{-} \phi_2(p_2)\right| < \varepsilon_0 \\ \left|\phi_\delta(p_1) \dot{-} \phi_\delta(p_2)\right| < \varepsilon_0 \end{cases} \tag{2.42}$$

and from (2.40) and (2.42) it follows that, $\forall \varepsilon_0 > 0$:

$$\begin{aligned} \left|\phi_\delta(p_1) \dot{-} \phi_2(p_1)\right| &\leq \left|\phi_\delta(p_1) \dot{-} \phi_\delta(p_2)\right| \\ &\quad + \left|\phi_\delta(p_2) \dot{-} \phi_2(p_2)\right| + \left|\phi_2(p_2) \dot{-} \phi_2(p_1)\right| < 3\varepsilon_0 \end{aligned} \tag{2.43}$$

hence $\forall \varepsilon_0 > 0$:

$$\begin{aligned} \left|\phi_1(p_1) \dot{-} \phi_2(p_2)\right| &\leq \left|\phi_1(p_1) \dot{-} \phi_\delta(p_1)\right| + \\ &\left|\phi_\delta(p_1) \dot{-} \phi_2(p_1)\right| + \left|\phi_2(p_1) \dot{-} \phi_2(p_2)\right| < \varepsilon_1 + 4\varepsilon_0 < 5\varepsilon_0 \end{aligned} \tag{2.44}$$

and analogously, $\forall \varepsilon_0 = \varepsilon/8 > 0$:

$$\begin{aligned} \left|\phi_1(p) \dot{-} \phi_2(p\right| &\leq \left|\phi_1(p) \dot{-} \phi_1(p_1)\right| + \\ \left|\phi_1(p_1) \dot{-} \phi_2(p_2)\right| &+ \left|\phi_2(p_2) \dot{-} \phi_2(p_1)\right| + \left|\phi_2(p_1) \dot{-} \phi_2(p)\right| < 8\varepsilon_0 = \varepsilon \end{aligned} \tag{2.45}$$

which contradicts (2.39), consequently for every $p \in A = \mathrm{dom}\,(\phi_1) \cap \mathrm{dom}\,(\phi_2)$ the equation $\left|\phi_1(p) \dot{-} \phi_2(p)\right| = 0$ holds, accordingly

$$\phi_1\big|_A = \phi_2\big|_A = 0 \tag{2.46}$$

Finally, because A is dense in U it follows that $\left(\phi_1\big|_A\right)^{\smile} = \phi_1 = \left(\phi_2\big|_A\right)^{\smile} = \phi_2$, therefore $(\Theta[\mathbf{X}, U], \mathfrak{T}_{\mathfrak{M}})$ is a Hausdorff space. $\square$

3. Topological vector spaces over $\Theta[\mathbf{X}, U]$

3.1. **Function vector spaces.** The subset $\mathcal{V}$ of $\left(\mathcal{C}^0(\mathcal{D}(U), \mathbb{C}), \preceq\right)$ consisting of every maximal function lying in $\mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$ equipped with the law $\dot{+}$ is a commutative group containing every constant map in $\mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$, because by definition, for every couple $f_1, f_2 \in \mathcal{V}$ the composition $f_1 \dot{+} f_2$ is also a maximal member of $\mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$. Analogously, from the definition of the law $\star$ it follows immediately, that for every analytic function field $\Theta[\mathbf{X}, U] \subseteq \mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$ and each pair $(\phi, f) \in \Theta[\mathbf{X}, U] \times \mathcal{V}$ the composition $\phi \star f$ is also maximal, therefore $\phi \star f \in \mathcal{V}$. Now, it is not difficult to see, that if $(M_n, \mathcal{A})$ is an analytic manifold, then for every $\mathcal{C}^\omega$-vector $\mathbf{X}$ and each chart $(U, \varphi) \in \mathcal{A}$ the subset $\mathcal{V} \subseteq \mathcal{C}^0(\mathcal{D}(U), \mathbb{C})$ consisting of every maximal element is a vector space over the analytic function field $\Theta[\mathbf{X}, U]$.

Enriching the vector space $\mathcal{V}$ with a topology $\mathfrak{T}$, being compatible with its algebraic vector space structure, we can calculate the sums of function series, for instance, the Taylor expansion of an analytic function. To prove some results described in the following section, we are only concerned with subspaces of $\mathcal{V}$, besides, for our purposes we only consider the coarser topology $\mathfrak{T}$ for $\mathcal{V}$ such that every member of its algebraic dual $\mathcal{V}^*$ is a continuous map.

Definition 3.1. *Let $\mathfrak{T}$ stand for the coarser topology for $\mathcal{V}$ under which every map $h : (\mathcal{V}, \mathfrak{T}) \to (\Theta[\mathbf{X}, U], \mathfrak{T}_{\mathfrak{M}})$ lying in the algebraic dual $\mathcal{V}^*$ of $\mathcal{V}$ is a continuous map. It is not difficult to see, that a subbase for $\mathfrak{T}$ is the set collection $\mathfrak{F} = \{W_{p,\varepsilon,h}(f_0) \mid f_0 \in \mathcal{V}, p \in U, \varepsilon > 0, h \in \mathcal{V}^*\}$ defined as follows:*

$$W_{p,\varepsilon,h}(f_0) = \Big\{ f \in \mathcal{V} \Big| \; h(f) \in V_{p,\varepsilon}\left(h(f_0)\right) \Big\} \tag{3.1}$$

where $V_{p,\varepsilon}\left(h(f_0)\right) \in \mathfrak{T}_{\mathfrak{M}}$ is a member of the subbase for $\mathfrak{T}_{\mathfrak{M}}$ defined in (2.15).

Taking into account (2.15) from the preceding relation it follows immediately, that

$$\begin{aligned} W_{p,\varepsilon,h}(f_0) = \Big\{ f \in \mathcal{V} \Big| \\ &\forall A \in \mathcal{D}(U), \exists p_0 \in \operatorname{dom}\left(h(f_0)\right) \cap \operatorname{dom}\left(h(f)\right) \cap A : \\ &\qquad \|\varphi(p_0) - \varphi(p)\| < \varepsilon \text{ and } q_{p_0}\left(h(f) \dot{-} h(f_0)\right) < \varepsilon \Big\} \end{aligned} \tag{3.2}$$

In locally convex spaces the coarser topology for which the algebraic dual coincides with the topological one is compatible with the vector space structure. By virtue of relation (3.1), the compatibility of $\mathfrak{T}$ with the algebraic structure of $\mathcal{V}$ does not depend on the nature of the considered field $\Theta[\mathbf{X}, U]$, as one can see proving that for every pair $f_1, f_2 \in \mathcal{V}$ and every neighborhood $W_{p,\varepsilon,h}(f_1 \dot{+} f_2)$ of $f_1 \dot{+} f_2$ there are a neighborhood $W_{p,\varepsilon_1,h}(f_1)$ of f_1 and a neighborhood $W_{p,\varepsilon_1,h}(f_2)$ of f_2 such that

$$W_{p,\varepsilon_1,h}(f_1) \dot{+} W_{p,\varepsilon_1,h}(f_2) \subset W_{p,\varepsilon,h}(f_1 \dot{+} f_2)$$

Of course, from (3.1) we have that

$$W_{p,\varepsilon,h}(f_1 \dot{+} f_2) = \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon}(h(f_1 \dot{+} f_2))\} \tag{3.3}$$

$$W_{p,\varepsilon_1,h}(f_1) = \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon_1}(h(f_1))\} \tag{3.4}$$

$$W_{p,\varepsilon_1,h}(f_2) = \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon_1}(h(f_2))\} \tag{3.5}$$

therefore,

$$\begin{aligned} W_{p,\varepsilon_1,h}(f_1) \dot{+} W_{p,\varepsilon_1,h}(f_2) &= \\ \{f \dot{+} g \mid h(f) \in V_{p,\varepsilon_1}(h(f_1)), h(g) \in V_{p,\varepsilon_1}(h(f_2))\} &\subseteq \\ \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon_1}(h(f_1)) \dot{+} V_{p,\varepsilon_1}(h(f_2))\}& \end{aligned} \tag{3.6}$$

Now, by Theorem 2.16 there is $\varepsilon_1 > 0$ such that

$$V_{p,\varepsilon_1}(h(f_1)) \dot{+} V_{p,\varepsilon_1}(h(f_2)) \subset V_{p,\varepsilon}(h(f_1) \dot{+} h(f_2))$$

Using this relation in (3.6) we obtain that

$$\begin{aligned} W_{p,\varepsilon_1,h}(f_1) \dot{+} W_{p,\varepsilon_1,h}(f_2) \subset \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon}(h(f_1) \dot{+} h(f_2))\} &= \\ \{f \in \mathcal{V} \mid h(f) \in V_{p,\varepsilon}(h(f_1 \dot{+} f_2))\} = W_{p,\varepsilon,h}(f_1 \dot{+} f_2)& \end{aligned} \tag{3.7}$$

therefore the mapping $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$ carrying each pair (f_1, f_2) into $f_1 \dot{+} f_2$ is continuous. The proof for the continuity of the mapping $\Theta[\mathbf{X}, U] \times \mathcal{V} \to \mathcal{V}$ carrying each pair $(\phi, f) \in \Theta[\mathbf{X}, U] \times \mathcal{V}$ into $\phi \star f$ is similar; so that $\mathfrak{T}$ is compatible with the algebraic vector space structure of $\mathcal{V}$.

4. Application to differential equation theory

4.1. **General equivalence conditions.** The general solution of a linear system of ordinary differential equations belongs to some vector space over the real (complex) number field generated by an exponential function set of the form

$$\mathcal{B} = \left\{e^{\lambda_\alpha (t-t_0)} \mid (t, t_0) \in \mathbb{R}^2, \lambda_\alpha \in \mathbb{C}, \alpha \in A\right\}$$

where A is any finite index set. A system of ordinary differential equations we shall say to be equivalent to some linear equation, provided that its general solution belongs to a vector space generated by a function set like $\mathcal{B}$, but now A can be also a non-countable index set [6]. We generalize such an equivalence concept to vector spaces over function fields. Thus, an equation system we shall say to be equivalent to some linear system, provided that its solution belongs to a vector space over a function field $\Theta[\mathbf{X}, U]$ being generated by a function set of the form

$$\mathcal{B} = \left\{e^{\phi_1 \star \Phi \dot{+} \phi_2 (t-t_0)} \mid (t, t_0) \in \mathbb{R}^2, \phi_1, \phi_2 \in \Theta[\mathbf{X}, U], \mathbf{X}\check{\ }\Phi \in \Theta[\mathbf{X}, U]\right\} \tag{4.1}$$

In the finite dimensional case, each solution satisfies an equation lying in an n-parameter linear system family. To state the equivalence between an autonomous system

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{X}(\mathbf{x}(t)) \tag{4.2}$$

and some linear equation system, i.e., to determine whether the general solution of (4.2) belongs to a vector space over the function field $\Theta[\mathbf{X}, U]$ generated by the function set (4.1), the following conditions are sufficient.

Condition 1. *Let $(M_n, \mathcal{A})$ an analytic manifold and $\mathbf{X}$ a $\mathcal{C}^\omega$-vector such that if its coordinates in a given chart $(U, \varphi) \in \mathcal{A}$ are $(X^1, X^2 \cdots X^n)$, then they satisfy the following properties: For each index $k \leq n$ there exist a subset $K \subseteq \mathbb{C}$, a continuous mapping $F^k : U \times K \to \mathbb{C}$, a member h^k of $\Theta[\mathbf{X}, U]$, and a maximal function $\Phi \in \mathcal{C}^1(\mathcal{D}(U), \mathbb{C})$ such that*

$$\mathbf{X}\check{}\,\Phi^k = h^k \in \Theta[\mathbf{X}, U] \tag{4.3}$$

$$\forall p \in U: \quad X^k(p) = F^k\left(p, \Phi^k(p)\right) \neq 0 \tag{4.4}$$

Condition 2. *For each positive integer $k \leq n$ there are a subset $D \subseteq \mathbb{C}$, a member A of $\mathcal{D}(U)$ satisfying the relation $A \subseteq \operatorname{dom}\left(h^k\right)$, and two maps $\mathcal{F}^k : A \times D \to \mathbb{C}$ and $g^k : A \times D \to \mathbb{C}$ such that for some integration path $(\sigma : I \subseteq \mathbb{R} \to D)$, this being continuous and differentiable except, at most, in a finite point set, the following is true,*

1): *For every fixed $z \in D : \mathcal{F}^k(p, z) \in \Theta[\mathbf{X}, U]$ and $g^k(p, z) \in \Theta[\mathbf{X}, U]$.*

2): *$F^k(p, s) = \displaystyle\int_C \mathcal{F}^k(p, z) \star e^{g^k(p,z)s} dz.$*

3): *$\forall r \in \mathbb{N}, \forall p \in A$ there is a real number $K(p, s, k) > 0$ such that:*

$$\left|\int_C \left(g^k(p, z) \star h^k(p)\right)^{\smile(r-1)} \star\ \mathcal{F}^k(p, z)\ \star\ e^{g^k(p,z)s} dz\right|^{\smile} \leq K(p, s, k) r!.$$

Notice, that by virtue of Lemma 2.5, for each $r \in \mathbb{N}$ the following relation holds: $A \times D \subseteq \operatorname{dom}\left(\left(g^k \star h^k\right)^{\smile r} \star \mathcal{F}^k\right)$.

Remark 4.1. *It is worth pointing out, that in general, it is not difficult to obtain the function $\mathcal{F}^k(p, z)$ by knowing $F^k(p, s)$; for instance, if the integration path C is a Bromwich one and $g(p, z) = z$, then $\mathcal{F}^k(p, z)$ is nothing but the Laplace transform for $F^k(p, s)$. Conversely, if $C = [0, \infty)$, then $F^k(p, s)$ is the Laplace transform for $\mathcal{F}^k(p, z)$ etc.*

Lemma 4.2. *If a vector $\mathbf{X}$ satisfies both Condition 1 and Condition 2, then for each positive integer r the associated linear operator $\mathbf{X}\check{}$ can be iterated r times over each of its coordinates.*

Proof. By virtue of Condition 1 and Condition 2 we obtain that

$$\forall p \in A: \qquad X^k(p) = F^k\left(p, \Phi^k(p)\right) = \int_C \mathcal{F}^k(p, z) \star e^{g^k(p,z) \star \Phi^k(p)}\, dz \tag{4.5}$$

therefore,

$$\mathbf{X}\check{}\,X^k(p) = \int_C \mathbf{X}\check{}\,\mathcal{F}^k(p, z) \star e^{g^k(p,z) \star \Phi^k(p)}\, dz \tag{4.6}$$

Since by assumption, regarding z as a parameter, both map families $\mathcal{F}^k(p,z)$ and $g^k(p,z)$ belong to the function field $\Theta[\mathbf{X},U]$ and $\Phi^k(p)$ satisfies equation 4.3, then from the equation above it follows immediately, that $\forall p \in A$,

$$\mathbf{X}^{\smile} X^k(p) = \int_C \mathcal{F}^k(p,z) \star g^k(p,z) \star h^k(\mathbf{x}) \star e^{g^k(p,z)\star\Phi^k(p)}\, dz \tag{4.7}$$

Now, by induction and taking into account Condition 2, it follows that $\forall r \in \mathbb{N}$:

$$\begin{aligned} \left|(\mathbf{X}^{\smile})^r X^k(p)\right|^{\smile} &= \\ \left|\int_C \mathcal{F}^k(p,z) \star \left(g^k(p,z) \star h^k(p)\right)^{\smile r} \star e^{g^k(p,z)\star\Phi^k(p)} dz\right|^{\smile} &\leq \\ K(p,\Phi^k(p))(r+1)! \end{aligned} \tag{4.8}$$

□

Theorem 4.3. *If a vector field* $\mathbf{X}$ *satisfies both Condition 1 and Condition 2, then for each* $p_0 \in U$ *there is a positive number* H_{p_0} *and an open interval* $I = (t_0 - H_{p_0},\ t_0 + H_{p_0}) \subseteq \mathbb{R}$ *such that the map* $\gamma : I \to M_n$ *satisfying* $\gamma(t_0) = p_0$ *whose* k*-th coordinate in the considered chart* (U,φ) *is defined by the integral transform*

$$\begin{aligned} x^k_{p_0}(\gamma(t)) = x^k_0 \star 1_U \dotplus \int_C \mathcal{F}^k(p_0,z) \star \left(g^k(p_0,z) \star h^k(p_0)\right)^{\flat} \star \\ e^{g^k(p_0,z)\star\Phi^k(p_0)} \star \left(e^{g^k(p_0,z)\star h^k(p_0)(t-t_0)} \dot{-} 1_U\right) dz \end{aligned} \tag{4.9}$$

is a solution for the autonomous system (4.2) satisfying the initial condition

$$\left(x^1_{p_0}(\gamma(t_0)), x^2_{p_0}(\gamma(t_0)) \cdots x^n_{p_0}(\gamma(t_0))\right) = (x^1_0, x^2_0 \cdots x^n_0) = \varphi(p_0)$$

Remark 4.4. *From the equation (4.9) it is easy to see, that letting* $\phi_1 = g^k$ *and* $\phi_2 = g^k \star h^k$*, then* $x^k_{p_0}(\gamma(t))$ *belongs to the subspace of the vector space* $\mathcal{V}$ *over the field* $\Theta[\mathbf{X},U]$ *generated by the function set* $\mathcal{B}$ *defined in (4.1). Thus, both Condition 1 and Condition 2 form a criterion for equivalence between a non-linear autonomous system and some linear equation.*

Proof. Since $\mathbf{X}$ is assumed to be a differentiable vector, then it satisfies the standard conditions for the existence of a solution $\gamma : I \subseteq \mathbb{R} \to M_n$ of the system (4.2), where I is some open subset of $\mathbb{R}$; accordingly if the coordinates of $\varphi(\gamma(t))$ in the considered chart (U,φ) are

$$\left(x^1_{p_0}(\gamma(t)), x^2_{p_0}(\gamma(t)), \cdots, x^n_{p_0}(\gamma(t))\right) = \mathbf{x}_{p_0}(\gamma(t))$$

then for each $k \leq n$,

$$\frac{d}{dt}\, x^k_{p_0}(\gamma(t)) = X^k(x^1_{p_0}(\gamma(t)), x^2_{p_0}(\gamma(t)) \cdots) \tag{4.10}$$

Now, by virtue of the preceding lemma we obtain that $\forall m \leq n$,

$$(4.11)\quad \frac{d^{m+1}}{dt^{m+1}}\, x^k_{p_0}(\gamma(t))\bigg|_{t=t_0} = (\mathbf{X}^{\smile})^m\, X^k(p)\bigg|_{p=p_0} =$$

$$\int_C \mathcal{F}^k(p_0,z) \star \Big(g^k(p_0,z) \star h^k(p_0)\Big)^{\smile m} \star e^{g^k(p_0,z)\star \Phi^k(p_0)}\, dz$$

therefore, the Taylor's expansion for $x^k_{p_0}(\gamma(t))$ is

$$(4.12)\quad x^k_{p_0}(\gamma(t)) = x^k_0 \star 1_U \dot{+}$$

$$\breve{\sum}_{m=1}^{\infty} \int_C \mathcal{F}^k(p_0,z) \star \left(\frac{(t-t_0)^{m+1}}{(m+1)!} \Big(g^k(p_0,z) \star h^k(p_0)\Big)^{\smile m} \right) \star e^{g^k(p_0,z)\star\Phi^k(p_0)} dz$$

where $\breve{\sum}_{m=1}^{\infty} = \left(\lim_{r\to\infty} \sum_{m=1}^{r} \right)^{\smile}$ is assumed to be the limit with respect to the topology $\mathfrak{T}$ in Definition 3.1. If $g^k(p_0,z) \star h^k(p_0) = 0$, then the right-hand of the former equation converges; otherwise, as a consequence of Lemma 2.7, the set $A = \operatorname{dom}\big(g^k(p_0,z) \star h^k(p_0)\big) - \operatorname{zero}\big(g^k(p_0,z) \star h^k(p_0)\big)$ belongs to $\mathcal{D}(U)$. Obviously, for each $p_0 \in A$ the former equation can be written as follows,

$$(4.13)\quad x^k_{p_0}(\gamma(t)) =$$

$$x^k_0 \star 1_U \dot{+} \breve{\sum}_{m=0}^{\infty} \int_C \left(\mathcal{F}^k(p_0,z) \star \frac{\big(g^k(p_0,z) \star h^k(p_0)(t-t_0)\big)^{\smile(m+1)}}{(m+1)!} \star \right.$$

$$\left. \Big(g^k(p_0,z) \star h^k(p_0)\Big)^{\flat} \star e^{g^k(p_0,z)\star\Phi^k(p_0)} \right) dz$$

Finally, using (4.8) it follows, that

$$(4.14)\quad \left| \breve{\sum}_{m=0}^{\infty} \int_C \mathcal{F}^k(p_0,z) \star \frac{\big(g^k(p_0,z) \star ((t-t_0)h^k(p_0))\big)^{\smile(m+1)}}{(m+1)!} \star \right.$$

$$\left. \Big(g^k(p_0,z) \star h^k(p_0)\Big)^{\flat} \star e^{g^k(p_0,z)\star\Phi^k(p_0)}\, dz \right|^{\smile}$$

$$\leq \sum_{m=0}^{\infty} \left| \Big(K\left(p_0, \phi^k(p_0)\right)(t-t_0) \Big)^{m+1} \right|$$

Now, taking into account Lemma 2.5, for each $r \in \mathbb{N}$ it is true that the set $\operatorname{dom}\big(g^k(p_0,z) \star h^k(p_0)\big)$ belongs to $\operatorname{dom}\Big(\big(g^k(p_0,z) \star h^k(p_0)\big)^{\smile r}\Big)$, accordingly the map $\left(g^k(p_0,z)\star h^k(p_0)\right)^{\smallsmile r}$ belongs to the ring $\mathcal{R}(\mathbf{X},A)$ for each $r\in\mathbb{N}$, therefore taking into account Lemma (2.17), we have that the convergence with respect to the topology $\mathfrak{T}_A$ is sufficient; i.e., one can calculate the limits with respect to the weak topology for the underlying vector space of $\mathcal{R}(\mathbf{X},A)$. Thus, letting $H_{p_0}=\min\limits_{k\leq n}\left(K\left(p_0,\phi^k(p_0),k\right)\right)^{-1}$, for every $|t-t_0|<H_{p_0}$ the right-hand of (4.12) converges and from (4.13) one can obtain (4.9) easily, for the sequence

$$\left(\breve{\sum}_{m=0}^{n}\frac{\left(g^k(p_0,z)\star h^k(p_0)(t-t_0)\right)^{m+1}}{(m+1)!}\right)_{n\in\mathbb{N}}$$

converges to $e^{g^k(p_0,z)\star h^k(p_0)(t-t_0)}\dot{-}1_U$. ☐

Remark 4.5. *The former theorem is nothing but a generalization to vector spaces over analytic function fields of a similar result proved in* [6] *for function spaces over the real (complex) number field.*

Corollary 4.6. *If the vector* $\mathbf{X}$ *satisfies both Condition 1 and Condition 2, being* $\mathbf{x}(x_0^1,x_0^2\cdots x_0^n,t)$ *the general solution of (4.2) for the initial values* $x_0^1,x_0^2\cdots x_0^n$*, then each coordinate* $x^k(x_0^1,x_0^2\cdots,t)$ *of* $\mathbf{x}(x_0^1,x_0^2\cdots x_0^n,t)$*, regarded as a function from* $\mathbb{R}^n\times\mathbb{R}$ *into* $\mathbb{C}$*, satisfies the linear partial differential equation*

$$\sum_{m=1}^{n}X^m\frac{\partial}{\partial x_0^m}x^k(x_0^1,x_0^2\cdots,t)-\frac{\partial}{\partial t}x^k(x_0^1,x_0^2\cdots,t)=0 \tag{4.15}$$

Proof. By straightforward computations and taking into account both Condition 1 and Condition 2, from (4.9) and (4.5) we obtain that

$$(4.16)\quad \sum_{m=1}^{n} X^m \frac{\partial}{\partial x_0^m} x^k(x_0^1, x_0^2 \cdots, t) =$$

$$X^k + \int_C \widehat{\mathcal{F}}^k(x_0^1, x_0^2 \cdots, z) \star \left(\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{h}^k(x_0^1, x_0^2 \cdots)\right)^{\flat} \star$$

$$e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{\Phi}^k(x_0^1, x_0^2 \cdots)} \star \left(\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{h}^k(x_0^1, x_0^2 \cdots)\right) \star$$

$$\left(e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{h}^k(x_0^1, x_0^2 \cdots)(t-t_0)} \dot{-} 1_U\right) dz =$$

$$X^k + \int_C \widehat{\mathcal{F}}^k(x_0^1, x_0^2 \cdots, z) \star e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{\Phi}^k(x_0^1, x_0^2 \cdots)} \star$$

$$e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{h}^k(x_0^1, x_0^2 \cdots)(t-t_0)}\, dz -$$

$$\int_C \widehat{\mathcal{F}}^k(x_0^1, x_0^2 \cdots, z) \star e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{\Phi}^k(x_0^1, x_0^2 \cdots)}\, dz$$

therefore, using equation (4.5) in the above one, we obtain that

$$(4.17)\quad \sum_{m=1}^{n} X^m \frac{\partial}{\partial x_0^m} x^k(x_0^1, x_0^2 \cdots, t) =$$

$$\int_C \widehat{\mathcal{F}}^k(x_0^1, x_0^2 \cdots, z) \star e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{\Phi}^k(x_0^1, x_0^2 \cdots)} \star$$

$$e^{\widehat{g}^k(x_0^1, x_0^2 \cdots, z) \star \widehat{h}^k(x_0^1, x_0^2 \cdots)(t-t_0)}\, dz = \frac{\partial}{\partial t} x^k(x_0^1, x_0^2 \cdots, t)$$

where we denote by $\widehat{\mathcal{F}} : \operatorname{img}(\varphi) \to \mathbb{C}$, $\widehat{g} : \operatorname{img}(\varphi) \to \mathbb{C}$, $\widehat{h} : \operatorname{img}(\varphi) \to \mathbb{C}$ and $\widehat{\Phi} : \operatorname{img}(\varphi) \to \mathbb{C}$ the maps defined by

$$(4.18)\quad \widehat{\mathcal{F}} = \mathcal{F} \circ \varphi^{-1}$$

$$(4.19)\quad \widehat{g} = g \circ \varphi^{-1}$$

$$(4.20)\quad \widehat{h} = h \circ \varphi^{-1}$$

$$(4.21)\quad \widehat{\Phi} = \Phi \circ \varphi^{-1}$$

□

Remark 4.7. *It worth noticing, that the solutions of (4.15) contains a function field* $\Theta[\mathbf{X}^*, U \times I]$*; where* $\mathbf{X}^* = \left(X^1 \frac{\partial}{\partial x^1}\big|_p, \cdots X^n \frac{\partial}{\partial x^n}\big|_p, -\frac{\partial}{\partial t}\big|_{t_0}\right)$.

4.2. **The finite dimensional case.** In this section we are concerned with the equivalence problem between non-linear and linear systems whose general solution belongs to a finite dimensional vector space. The following results form a criterion for the existence of such an equivalence.

Theorem 4.8. *Let $(M_n, \mathcal{A})$ be an analytic manifold and $\mathbf{X}$ a $\mathcal{C}^\omega$-vector whose coordinates in a chart $(U, \varphi) \in \mathcal{A}$ are $(X^1, X^2 \cdots X^n)$. If for each $k \leq n$ the linear operator $\mathbf{X}^{\smile}$ can be iterated r times over the k-th coordinate X^k of $\mathbf{X}$, and there is a linear dependence relation*

$$(\mathbf{X}^{\smile})^r \, X^k = \breve{\sum}_{j=0}^{r-1} c_j^k \star (\mathbf{X}^{\smile})^j \, X^k \tag{4.22}$$

whose coefficients $c_0^k, c_1^k \cdots c_{r-1}^k$ belong to the function field $\Theta[\mathbf{X}, U]$, being r the smallest positive integer satisfying (4.22), besides, the polynomial

$$P_k(z) = \big(z)\big)^{\smile r} \dot{-} c_{r-1}^k \big(z\big)^{\smile (r-1)} \dot{-} \cdots \dot{-} c_0^k \tag{4.23}$$

has no multiple root, and each of its roots belongs to $\Theta[\mathbf{X}, U]$, then X^k lies in the finite dimensional vector space over the associated function field $\Theta[\mathbf{X}, U]$ generated by the function set $\mathcal{B} = \left\{ \mathrm{e}^{\Phi_m^k} \mid m \leq r \right\}$, where the Φ_m^k are maximal members of $\mathcal{C}^\omega(\mathcal{D}(U), \mathbb{C})$ satisfying the relation

$$\mathbf{X}^{\smile} \Phi_m^k = \lambda_m^k \in \Theta[\mathbf{X}, U] \tag{4.24}$$

and the $\lambda_1^k \cdots \lambda_r^k$ are the solutions of

$$P\big(z\big) = 0 \tag{4.25}$$

moreover, the set $\mathcal{B}$ is a basis for the generated space.

Proof. On the one hand, by virtue of Theorem 2.12 every solution of 4.23 lies in $\Theta[\mathbf{X}, U]$. On the other hand, letting

$$\forall n < r : \quad f_n^k = \Big(\mathbf{X}^{\smile}\Big)^n X^k \tag{4.26}$$

define the functions $\psi_1^k, \psi_2^k, \cdots \psi_r^k$ by

(4.27)

$$\forall p \in U : \begin{cases} \psi_1^k(p) &= \alpha_1^1(p) \star X^k(p) \dot{+} \alpha_2^1(p) \star f_1^k(p) \cdots \dot{+} \alpha_r^1(p) \star f_{r-1}^k(p) \\ \psi_2^k(p) &= \alpha_1^2(p) \star X^k(p) \dot{+} \alpha_2^2(p) \star f_1^k(p) \cdots \dot{+} \alpha_r^2(p) \star f_{r-1}^k(p) \\ \cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \psi_r^k(p) &= \alpha_1^r(p) \star X^k(p) \dot{+} \alpha_2^r(p) \star f_1^k(p) \cdots \dot{+} \alpha_r^r(p) \star f_{r-1}^k(p) \end{cases}$$

where $\alpha_j^i \in \Theta[\mathbf{X}, U]$ for each $i, j \leq r$; hence, using (4.26) and (4.22) we obtain, that $\forall i \leq r$:

$$\begin{aligned} \mathbf{X}^{\smile} \psi_i^k &= \alpha_1^i \star f_1^k \dot{+} \cdots \dot{+} \alpha_{r-1}^i \star f_{r-1}^k \\ & \alpha_r^i \star c_0^k \star X^k \dot{+} \alpha_r^i \star c_1^k \star f_1^k \dot{+} \cdots \alpha_r^i \star c_{r-1}^k \star f_{r-1}^k \end{aligned} \tag{4.28}$$

therefore, if for some $A \in \mathcal{D}(U))$ we require the α^i_j to satisfy the conditions:

$$(4.29) \qquad \forall i \leq r, \forall p \in A: \quad \begin{cases} c^k_0(p) \star \alpha^i_r(p) = \lambda^k_i(p) \star \alpha^i_1(p) \\ \alpha^i_1(p) \dot{+} c^k_1(p) \star \alpha^i_r(p) = \lambda^k_i(p) \star \alpha^i_2(p) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \alpha^i_{r-1}(p) \dot{+} c^k_{r-1}(p) \star \alpha^i_r(p) = \lambda^k_i(p) \star \alpha^i_r(p) \end{cases}$$

then the maps ψ^k_i satisfy the equation

$$(4.30) \qquad \breve{\mathbf{X}}\psi^k_i = \lambda^k_i \star \psi^k_i$$

The equation (4.29) possesses non-trivial solution provided that

$$(4.31) \qquad \begin{vmatrix} -\lambda^k_i & 0 & 0 & \cdots & 0 & c^k_0 \\ 1 & -\lambda^k_i & 0 & \cdots & 0 & c^k_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 1 & c^k_{r-1} \dot{-} \lambda^k_i \end{vmatrix} = P(\lambda^k_i) = 0$$

accordingly, the λ^k_i are solutions of (4.23), therefore they are also eigenvalues of the endomorphism associated to the matrix

$$(4.32) \qquad \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & c^k_0 \\ 1 & 0 & 0 & \cdots & 0 & c^k_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 1 & c^k_{r-1} \end{pmatrix}$$

Setting $A = \bigcap_{0 \leq n \leq r} \operatorname{dom}(c_n)$, because by hypothesis the solutions of (4.31) belong to $\Theta[\mathbf{X}, U]$, then the corresponding eigenvectors

$$(\alpha^1_1, \alpha^1_2 \cdots), (\alpha^2_1, \alpha^2_2 \cdots) \cdots$$

belong to the vector space $(\Theta[\mathbf{X}, U])^r$, besides, since we are assuming that $P(z)$ has no multiple root, then the eigenvectors are linearly independent; consequently, taking into account (4.27), X^k can be written as a unique linear combination of the $\psi^k_1, \psi^k_2, \cdots \psi^k_r$ with coefficients in $\Theta[\mathbf{X}, U]$. Finally, setting

$$(4.33) \qquad \forall j \leq r: \quad \Phi^k_j = \ln(\psi^k_j)$$

the equation (4.24) are satisfied and X^k can be written as a linear combination of $e^{\Phi^k_1}, e^{\Phi^k_2} \cdots e^{\Phi^k_r}$ with coefficients in $\Theta[\mathbf{X}, U]$. Since the $\psi^k_1, \psi^k_2, \cdots \psi^k_r$ are linearly independent, then $\mathcal{B}$ is a basis. $\square$

Corollary 4.9. *Let*

$$\begin{cases} \dfrac{d}{dt}\,x^1(t) = X^1\left(x^1(t), x^2(t)\cdots\right) \\ \dfrac{d}{dt}\,x^2(t) = X^2\left(x^1(t), x^2(t)\cdots\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{d}{dt}\,x^n(t) = X^n\left(x^1(t), x^2(t)\cdots\right) \end{cases} \tag{4.34}$$

be an autonomous system. If $\mathbf{X}$ *satisfies the hypotheses of the preceding theorem, then it satisfies both Condition 1 and Condition 2, therefore the corresponding equation is equivalent to some linear system, besides, its solutions can be obtained by means of the integral transform (4.9).*

Proof. By the former theorem we know that the roots of $P(z)$ belong to $\Theta[\mathbf{X}, U]$, besides, from (4.24) we obtain that

$$\forall i \le r: \quad \mathbf{X}^{\smile}\Big((\lambda_i^k)^\flat \star \Phi_i^k\Big) = 1_U \tag{4.35}$$

consequently

$$\forall i \le r, \forall m \le r: \qquad \mathbf{X}^{\smile}\Big((\lambda_i^k)^\flat \star \Phi_i^k \dot{-} (\lambda_m^k)^\flat \star \Phi_m^k\Big) = 0_U \tag{4.36}$$

Since each of the coordinates of $\mathbf{X}$ is supposed to be an analytic function, and taking into account Definition 2.3, this equation implies that

$$(\lambda_i^k)^\flat \star \Phi_i^k \dot{-} (\lambda_m^k)^\flat \star \Phi_m^k \in \Theta[\mathbf{X}, U] \tag{4.37}$$

accordingly, setting $(\lambda_1^k)^\flat \star \Phi_1^k = \Phi^k$ there are

$$\vartheta_1^k, \vartheta_2^k \cdots \vartheta_r^k \in \Theta[\mathbf{X}, U]$$

such that

$$\forall m \le r: \quad \Phi_m^k = \lambda_m^k \star \Phi^k \dot{+} \vartheta_m^k \tag{4.38}$$

besides,

$$\mathbf{X}^{\smile}\Phi^k = 1_U \tag{4.39}$$

consequently, X^k can by written as a linear combination of the exponential function set $e^{\lambda_1^k \Phi^k}, e^{\lambda_1^k \Phi^k} \cdots e^{\lambda_r^k \Phi^k}$, for the maps $e^{\vartheta_1^k}, e^{\vartheta_2^k} \cdots$ are also members of $\Theta[\mathbf{X}, U]$. If $a_0^k, a_1^k \cdots a_r^k$ are the corresponding coefficients, we have that

$$X^k = \breve{\sum}_{m=1}^{r} a_m^k \star e^{\lambda_m^k \star \Phi^k} \tag{4.40}$$

Now, letting $\forall k \le n$, $\forall p \in U$:

$$F^k(p, s) = \breve{\sum}_{m=1}^{r} a_m^k(p) e^{\lambda_m^k(p)s} \tag{4.41}$$

$$g(p, z) = z \tag{4.42}$$

$$h^k(p) = 1_U \tag{4.43}$$

and

$$\mathcal{F}^k(p,z) = \frac{1}{2\pi i}\left(\sum_{m=1}^{r} \frac{a_m^k(p)}{z-\lambda_m^k(p)}\right)^{\smile} \tag{4.44}$$

it is not difficult to see, that these functions satisfy both Condition 1 and Condition 2 for a Bromwich integration path. Thus by virtue of theorem 4.3, each coordinate of the general solution of the corresponding autonomous system (4.34) can be written as a linear combination of the functions

$$1_U, e^{\lambda_1^k \star \left(\Phi^k \hat{+} (t-t_0)\right)}, \cdots e^{\lambda_r^k \star \left(\Phi^k \hat{+} (t-t_0)\right)}$$

with coefficients in the function field $\Theta[\mathbf{X}, U]$. □

Example 4.10. *Consider the space $\mathbb{R}^2$ with the standard smooth structure. Consider the vector $\mathbf{X} = \dfrac{x^1}{x^2}\mathbf{e}_1 + \dfrac{x^2}{x^1}\mathbf{e}_2$ and let (φ, U) be the chart whose domain is $U = \left\{(x^1,x^2) \in \mathbb{R}^2 \mid \left|x^1\right| > a, \left|x^2\right| > b\right\}$ and φ is the identity, where a and b are two fixed positive numbers. In this case, $\Theta[\mathbf{X}, U]$ is nothing but a function field $(\mathcal{K}, +, \cdot)$ whose underlying set $\mathcal{K}$ consists of functions of the form $f(\frac{1}{x^1} - \frac{1}{x^2})$, with domain in $\mathcal{D}(U)$, and $f : \mathbb{R} \to \mathbb{C}$ is a map such that both functions $\Re(f)$ and $\Im(f)$ are analytic.*

Now, consider the corresponding non linear autonomous system

$$\begin{cases} \dfrac{d}{dt}\, x^1(t) = \dfrac{x^1(t)}{x^2(t)} \\ \dfrac{d}{dt}\, x^2(t) = \dfrac{x^2(t)}{x^1(t)} \end{cases} \tag{4.45}$$

It is not difficult to see, that the field $\mathbf{X}$ satisfies the conditions of the preceding theorem, for there are the dependence relations

$$\begin{cases} \mathbf{X}^{\smile}X^1 = \left(\dfrac{x^1}{x^2}\dfrac{\partial}{\partial x^1} + \dfrac{x^1}{x^2}\dfrac{\partial}{\partial x^1}\right)\dfrac{x^1}{x^2} = -\left(\dfrac{1}{x^1} - \dfrac{1}{x^2}\right)X^1 \\ \mathbf{X}^{\smile}X^2 = \left(\dfrac{x^1}{x^2}\dfrac{\partial}{\partial x^1} + \dfrac{x^1}{x^2}\dfrac{\partial}{\partial x^1}\right)\dfrac{x^2}{x^1} = \left(\dfrac{1}{x^1} - \dfrac{1}{x^2}\right)X^2 \end{cases} \tag{4.46}$$

In this case, letting $\Phi = \widehat{\Phi} \circ \varphi$ we have that,

$$\widehat{\Phi}^m(x^1,x^2) = (-1)^{m+1}\left(\frac{1}{x^2} - \frac{1}{x^1}\right)^{-1} \ln\left(\frac{x^1}{x^2}\right) \quad (m = 1,2)$$

satisfies the equation

$$\sum_{j=1}^{2} X^j \frac{\partial}{\partial x^j}\, \widehat{\Phi}^m(x^1,x^2) = (-1)^{m+1} \tag{4.47}$$

hence the function $F^m\left(x^1, x^2, \widehat{\Phi}^m(x^1,x^2)\right) = X^m(x^1,x^2)$ defined in Condition 1, with respect to the chart (U,φ), for this example is:

$$F^m\left(x^1,x^2,s\right) = e^{(-1)^m\left(\frac{1}{x^1} - \frac{1}{x^2}\right)s} \quad (m = 1,2) \tag{4.48}$$

accordingly, setting $\mathcal{F}^m(z) = \dfrac{1}{2\pi i(z-(-1)^m)}$, $(m = 1, 2)$ *with the obvious integration path, we have that*

$$F^m\left(x^1, x^2, s\right) = \oint \mathcal{F}^m(z) e^{s\left(\frac{1}{x^1}-\frac{1}{x^2}\right)z} dz \tag{4.49}$$

which satisfies the Condition 2, accordingly (4.45) is equivalent to some 2-parameter linear differential equation family, as it can be deduced from the solution for (4.45),

$$\begin{cases} x^1(t) = -\dfrac{1}{\alpha(x_0^1, x_0^2)} + \dfrac{C(x_0^1, x_0^2)}{\alpha(x_0^1, x_0^2)}\, e^{\alpha(x_0^1, x_0^2)(t-t_0)} \\ x^2(t) = \dfrac{1}{\alpha(x_0^1, x_0^2)} - \dfrac{1}{C(x_0^1, x_0^2)\alpha(x_0^1, x_0^2)}\, e^{-\alpha(x_0^1, x_0^2)(t-t_0)} \end{cases} \tag{4.50}$$

where $\alpha(x_0^1, x_0^2) = \left(\dfrac{1}{x_0^2} - \dfrac{1}{x_0^1}\right)$ *and* $C(x_0^1, x_0^2) = \dfrac{x_0^1}{x_0^2}$, *for instance, the coordinate* $x^1(t)$ *is solution of* $y'(t) - \alpha(c_1, c_2)y(t) - 1 = 0$ *for* $c_1 = x_0^1$ *and* $c_2 = x_0^2$. *It is easy to see, that these functions* $\alpha(x^1, x^2)$ *and* $C(x^1, x^2)$ *satisfy the following equations,*

$$\sum_{j=1}^{2} X^j \frac{\partial}{\partial x_0^j}\, \alpha(x_0^1, x_0^2) = \sum_{j=1}^{2} X^j \frac{\partial}{\partial x_0^j} \left(\frac{1}{x_0^2} - \frac{1}{x_0^1}\right) = 0 \tag{4.51}$$

$$C(x_0^1, x_0^2) = e^{\alpha(x_0^1, x_0^2)\widehat{\Phi}(x_0^1, x_0^2)} = \frac{x_0^1}{x_0^2} \tag{4.52}$$

therefore $\mathbf{X}\check{\Phi}^m(p) = 1_U \in \Theta[\mathbf{X}, U]$, *besides, the coefficient* $\alpha(x_0^1, x_0^2)$ *lies in* $\Theta[\mathbf{X}, U]$. *Indeed, every coordinate in (4.50) belongs to a vector space over the function field* $\Theta[\mathbf{X}, U] = \Theta\left[\dfrac{x^1}{x^2}\mathbf{e}_1 + \dfrac{x^2}{x^1}\mathbf{e}_2, U\right]$.

Definition 4.11. *By means of the function field* $\Theta[\mathbf{X}, U]$ *associated to each smooth vector* $\mathbf{X}$, *one can generalize the concept of linear equation system as follows. Assume* $\left(a_j^i\right)_{i,j\leq n}$ *to be a matrix such that for each couple* $i, j \leq n$ *it is true that*

$$a_j^i \in \Theta\left[\sum_{m=1}^{n}\sum_{k=1}^{n} a_k^m \star x^k \mathbf{e}_m, U\right] \tag{4.53}$$

where $\mathbf{x} = (x^1, x^2 \cdots x^n)$ *are the coordinates of the homeomorphism* φ *of the considered chart* (U, φ). *In this case, the autonomous system*

$$
(4.54) \qquad \begin{cases} \dfrac{d}{dt}x^1(t) = \sum_{m=1}^{n} a_m^1(x^1(t), x^2(t)\cdots) \star x^m(t) \\ \dfrac{d}{dt}x^2(t) = \sum_{m=1}^{n} a_m^2(x^1(t), x^2(t)\cdots) \star x^m(t) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{d}{dt}x^n(t) = \sum_{m=1}^{n} a_m^n(x^1(t), x^2(t)\cdots) \star x^m(t) \end{cases}
$$

that, in general, is not linear over the real (complex) number field, can be regarded as a linear system over the associated function field

$$
\Theta[\mathbf{X}, U] = \Theta\left[\sum_{m=1}^{n}\sum_{k=1}^{n} a_k^m \star x^k \mathbf{e}_m, U\right]
$$

Henceforth, an equation system satisfying this property we shall say to be Θ*-linear.*

Theorem 4.12. *If the characteristic polynomial* $P(z) = c_0 \dotplus c_1 \star z \dotplus \cdots c_n z^{\smile n}$ *of an* $n \times n$*-matrix* $\left(a_j^i\right)_{i,j \le n}$ *defining a* Θ*-linear equation system has no multiple root, and every solution of* $P(z) = 0$ *belongs to* $\Theta[\mathbf{X}, U]$*, then the corresponding vector field whose* i*-th coordinate is*

$$
X^i = \sum_{j=1}^{n} a_j^i \star x^j
$$

satisfies the conditions of theorem 4.8, therefore by virtue of Corollary 4.9 the equation system (4.54) satisfies both Condition 1 and Condition 2, accordingly the corresponding solution can be obtained from the integral transform (4.9).

Proof. Letting $\mathbf{A} = \left(a_j^i\right)_{i,j \le n}$ and $\mathbf{X} = \mathbf{A} \star \mathbf{x}$, then the f_n^k defined in equation (4.26) satisfy the following relation,

$$
(4.55) \qquad f_r^k = \mathbf{A}^{r+1} \star \mathbf{x}
$$

where $\mathbf{x} = (x^1, x^2 \cdots x^n)$ stand for the coordinates of φ. Now, setting

$$
(4.56) \qquad P(\lambda) = \begin{vmatrix} a_1^1 \dot{-} \lambda & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 \dot{-} \lambda & \cdots & a_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ a_1^n & a_2^n & \cdots & a_n^n \dot{-} \lambda \end{vmatrix}^{\smile}
$$

it is a well-known fact that each matrix satisfies its characteristic equation $P(\mathbf{A}) = 0$; therefore, if $c_0, c_1 \cdots c_n$ are the corresponding coefficients, we have that

$$
(4.57) \qquad P(\mathbf{A}) \star \mathbf{A} = c_n \star \mathbf{A}^{n+1} \dotplus c_{n-1} \star \mathbf{A}^n \cdots \dotplus c_0 \star \mathbf{A} = 0_U
$$

and by assumption the corresponding eigenvalues belong to the associated function field

$$\Theta\left[\sum_{m=1}^{n}\sum_{k=1}^{n} a_k^m \star x^k \mathbf{e}_m, U\right]$$

Finally, from the former equation it follows immediately, that

$$c_n \star \mathbf{A}^{n+1} \star \mathbf{x} \dot{+} c_{n-1} \star \mathbf{A}^n \star \mathbf{x} \cdots \dot{+} c_0 \star \mathbf{A} \star \mathbf{x} = 0_U \tag{4.58}$$

and taking into account (4.55), there is a linear dependence relation like (4.22), and the theorem follows. □

Remark 4.13. *Since every function field* $\Theta[\mathbf{X}, U]$ *contains* $(\mathbb{C}_U, \dot{+}, \star, ^\flat)$ *as a subfield, this being isomorphic to the complex number field, then every linear equation system is also a* Θ*-linear one; consequently linear equations satisfy the conditions of the preceding theorem.*

References

[1] Dieudonné, J.: Éléments d' analyse, Cahiers scientifiques, fascicule XXXIV, Gauthier-Villars, Paris, 1971.

[2] Kelley, J.L.: General Topology, Springer-Verlag, New York, 1955.

[3] Kobayashi, S. and Nomizu, K.: Foundations of Differential Geometry, Interscience Publish, 1963.

[4] Köthe, G.: Topological Vector Spaces, Springer-Verlag, New York, 1969.

[5] Lang, S.: Algebra, Addison-Wesley, MA, 1965.

[6] Palomar, J. E.: Sufficient Conditions for Equivalence between non Linear and Linear Autonomous Systems, Dynamics of Cont., Discr. and Imp. Sys. (DCDIS), **6**,1999, pp. 237-249.

[7] K. Yosida, K.: Équations Différentielles et Intégrales, Dounod, Paris, 1971.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2000, pp. 88 – 121
Submitted: August 22, 2000. Published: December 15, 2000

Numerical Bifurcation and Stability Analysis of the Nicolis-Puhl Reaction as an Application of Class of Integro Partial Differential Equation

G.H. Erjaee

Abstract

Numerical bifurcation and stability analysis techniques is applied to the specific class of integro-partial differential equations (IPDE). This class of IPDE can be used to study many physical and chemical reaction systems as a dynamical system. The Nicolis-Puhl(NP) reaction an isothermal two-component, autocatalytic reaction exhibiting bistability, perfectly micromixed continuous-flow stirred tank reactor (CSTR) has been shown to be sensitive to the degree of micromixing under non-premixed feed conditions. These properties make it an intersting model to be study as an application of this class of IPDE. The numerical simulation methods studied here has shown excellent numerical stability and the results are similar to those are found by Fox, Villermaux (1990), Fox, Curtis and Halasi (1990), and Fax (1990) for NP reaction. These results also has been verified through numerical simulations of time dpendent solution model.

AMS Subject Clasification Codes: 34C35, 34A47 and 34A50.
Keywords and Phrases: Dynamical systems, bifureation analysis, continuous flow stirred tank reactor.

Mathematics Department, College of Science, Shiraz University,Shiraz, Iran

1. Introduction

The theory of dynamical systems is a vast and vibrant area. This theory is playing an ever important role in the sciences and engineering, provoking a blurring of boundaries between scientific disciplines and a resurgence of interest in the modern as well as the classical techniques of applied mathematics. Concerning this subject the bifurcation and stability analysis of the chemical reaction systems is an old and much-honored chapter in dynamical systems, one that is experimentaly may insightful get devoid of extensive formalism and study. It is the aim of this work to introduce a class of nonlinear Integro Partial Differential Equations (IPDE) and imploy this system as a model towards a contribution and understanding the dynamical behavior of the chemical reactions such as Nicolis-Puhl (NP) reaction in the Continuous-flow Stirred Tank Reactor (CSTR). The well known CSTR, in chemical engineering, is a stirred pot into which chemicals of a certain recipe flow continuously and are stirred and heated or cooled while undergoing chemical reaction. These reactors often occur in connected multiples which can have ever increasing static and dynamic complexity. The bifurcation behavior of the CSTR has been studies for more than 30 years [13], [14] and yet qualitatively new results seem to continually appear [2], [3], [4].

The two-component reaction scheme proposed by Puhl and Nicolis in 1987 [12]. In the perfect-micromixing limit, this reaction exhibits bistability and oscillatory behavior. The rate constants for the three steps in the NP reaction are all of the same order of magnitude, hence, its numerical investigation does not require the special considerations for stiff equations nor the long computing times needed for this reaction [8]. In addition, since the NP reaction in a perfectly micromixed flow reactor is two-dimensional, any increase in the dynamical dimension observed for finite micromixing rates can be unambiguously linked to micromixing effects [6].

As we will see in next section the simplest models which can described the isotermal reaction scheme of NP in the perfect micromixing CSTR model for a flow reactor with equal concentrations of two species, is in the form of the system consists of two ordinary differential equations. Puhl and Nicolis have used this model [12] for analyzing the effect of micromining on the NP reaction. Although, expansive work has been done in their work, yet in order to invastigate more detail of dynamical behavior of this reaction another model has been imployed. This model which is known as Interaction-by-Exchange-with-the Mean (IEM) model is an IPDE with one-parameter micromixing model derived from the one-point joint probability density function by Pope in 1985 [11]. The alternative derivation of the IEM model equations along with a more detailed comparison of the model to other simple micromixing models can be found elsewhere [17], [9] and [2]. Nevertheless, due to the difficulty of solving IEM system, and further investigation and study of its time-dependent solution, it will be useful to introduce a transformation technique, to transfer this system to a simpler ordinary differential equations (ODE) system [3]. One of the main objectives of this work is to use this ODE system to analyze the behavior of the NP reaction under poorly micromixed conditions. So the remainder of this work has been arranged as follows.

In section 2, the NP reaction is briefly introduced and formulated in two deferent model equations. One is in the case of perfect-micromixing which consists of two ODEs, and another is in the form of IEM model under nonpremixed feed conditions. Among of existence derivations of IEM model outhor has attempted to illustrate an easire method of derivation in [2]. In section 3 the transformed IEM model to a simpler ODE system and the numerical simulation of this system is briefly reviewed [3], [5]. The main contributions of this work appear in section 4 where the results of the numerical bifurcation and stability analysis are presented. In order to test the results found in section 4 we have imployed transformed IEM model to find the time dependent solution using ODE solver LSODES [10]. So time dependent solutions for different types of fixed points are presented in section 5. The closing remark and conclusion is given in section 6.

2. Modeling of the Nicolis-Puhl Reaction

The NP reaction is expressed in terms of the isothermal reaction scheme [12] as follows:

$$2A + B \xrightarrow{k_1} 3A$$

$$A \xrightarrow{k_2} R$$

$$C + A \xrightarrow{k_3} A + B.$$

Here, C is held constant and when $k_2 = k_3$, the last two reaction become

$$C + A \xrightarrow{k_2} B + R.$$

This bistable reaction studied previously by Puhl and Nicolis [12], Villermaux [16] and Fox [7], [6].

Puhl and Nicolis have studied a normal mode analysis for the above reaction scheme by varying k_2 and k_3 independently so that the resultant dynamical system can have Takens-Bogdanov (TB) points in addition to multiple steady stats [12]. Their analysis of the perfect-micromixing case will be usful when analyzing our results in section 4 found from the transformed IEM model.

From the kinetic scheme above, the dimensionless, perfect-micromixing CSTR model for a flow reactor with equal concentrations of A and B in the feed stream can be found as follows:

$$\frac{dC_A}{d\theta} = \tau k_1 C_A^2 C_B - \tau k_2 C_A + 1 - C_A$$
$$\frac{dC_B}{d\theta} = -\tau k_1 C_A^2 C_B + \tau k_3 C_A + 1 - C_B. \tag{1}$$

Here, the mean residence time, τ, has been employed to make the time variable dimensionless. This will be convenient when formulating the IEM model [6] and our transformed IEM model [3]. The parameter σ appearing in Puhl and Nicolis work [12] corresponds to $(\tau k_1)^{-1}$ in Fox studies [6], [5] and [7]. Hence if we let k_1 be unity, then this paramer σ will be equal to τ^{-1} in over study. Based on the bifurcation analysis presented in [12] Puhl and Nicolis have generated bifurcation sets from constant k_2-cross sections in (k_2, k_3, τ^{-1}) parameter space. The resultant bifucation sets incloding Hoph bifurcations and TB points (for $k_2 = \frac{52}{75}$) can be found in [12] and [6], so they are not presented here. But in order to compare our bifurcation results to these found by Puhl-Nicolis and Fox we will use some of the Fox's results using the IEM model. In section 4 we shall follow the lead of these results and fix k_2 and k_3 so that, in the perfect micromixing limit, the TB point can be attained by varing τ^{-1}.

Now we shall restrict our attention to the IEM model formulated for the NP reaction with non premixed feed and equal inlet flow rates for both components A and B. For this case, the model equations can be written in dimensionless form [6] as

$$\frac{\partial C_A^j}{\partial \alpha}(\alpha,\theta)+\frac{\partial C_A^j}{\partial \theta}(\alpha,\theta) = \tau (C_A^j)^2 C_B^j - \tau k_2 C_A^j + b[< C_A(\theta) > - C_B^j],$$

$$\frac{\partial C_B^j}{\partial \alpha}(\alpha,\theta)+\frac{\partial C_B^j}{\partial \theta}(\alpha,\theta) = -\tau (C_A^j)^2 C_B^j + \tau k_3 C_A^j + b[< C_B(\theta) > - C_B^j] \tag{2}$$

for $j = 1, 2$ with

$$C_A^1(0,\theta) = 2, C_A^2(0,\theta) = 0, C_B^1(0,\theta) = 0, C_B^2(0,\theta) = 2$$

where

$$< C_A(\theta) >= \frac{1}{2}\int_0^{+\infty} [C_A^1(\alpha,\theta) + C_A^2(\alpha,\theta)] e^{-\alpha} d\alpha$$

$$< C_B(\theta) >= \frac{1}{2}\int_0^{+\infty} [C_B^1(\alpha,\theta) + C_B^2(\alpha,\theta)] e^{-\alpha} d\alpha. \tag{3}$$

In this IPDE system $C_A^j(\alpha,\theta)$ is the dimensionless concentration of component A in fluid particles with dimensionless age α at dimensionless time θ that entered the reactor from stream j, and $< C_A >$ is the dimensionless, volume-averaged concentration of component A in the reactor. Analogous notation is employed for component B. The inlet concentrations have been chosen so that the dimensionless, flow averaged inlet concentrations of both A and B are unity [6]. The dimensionless micromixing parameter appears as b in equation [13] and, theoretically, may take on values between 0 (complete segregation) and $+\infty$ (prefect-micromixing). All other parameters are the same as those appearing in above Puhl and Nicolis model.

The above IEM model gives a mean field description of the system where chemical reactions occur locally and mass is exchanged globally, equation (2). The mean concentration is found by weighting the local concentrations according to their age in the reactor, equation (3). So the nonlinear integral equation (2), for the mean concentration, mathematically involve non-Markovian or history-dependent behavior and are known, in many cases, to exhibit dynamical behavior absent from the ordinary differential equation, such as model (1), found in the history-independent limit. Moreover, as discussed in [6] it can be shown that if b becomes very large, then $C_A^j \to < C_A >$, $C_B^j \to < C_B >$ for $\alpha > 0$ and the IEM model, in terms of θ alone, is clearly reminiscent of the perfect micromixing model (1). Therefore, using IEM model enable us not only to study the dynamical behavior of NP reaction under perfectly micromixed conditions, but also in the region of incomplete micromixing.

Due to the presence of the integral terms in IEM model it is not possible to find an analytical solution. Nevertheless, extensive work has been done by Fox et. al., [5], [6], [7], [9], [16] and [17], to find the numerical analysis methods for bifurcation and stability analysis of this model. However, their analysis for the unsteady-state IEM model would be considerably less straightforward than the one which we will discuss here. Indeed, the method which we have use here is transformed IEM model as an ordinary differential equations system in infinite dimensional space. The detial technique of this transformation is discussed by author in [3], noting that, slightly modification have been applied here because of its application to the NP reaction (section 3). The assertion of this transformed IEM model, in addition to its simplicity, is availability of the softwares to find the solution and the dynamical behavior of the model in general, and the behavior of time dependent solutions, in particular.

3. IEM Model Transformation for NP Reaction and its Numerical Simulation

Although the general unsteady-state IEM model is given by equations (2) and (3), for the purpose of our transformation technique it is not difficult to see that a straightforward computation leads to the integro-differential equations for the volume-averaged concentrations. To see this, averaging equation (2) over α and j, as implied by equation (3), and employing integration by parts yields

$$\begin{aligned}
\frac{d < C_A >}{d\theta} &= \frac{1}{2}\tau \int_0^{+\infty} [(C_A^1(\alpha,\theta))^2 C_B^1(\alpha,\theta) + (C_A^2(\alpha,\theta))^2 C_B^2(\alpha,\theta)] e^{-\alpha} d\alpha \\
&\quad -\tau k_2 < C_A(\theta) > +1 - < C_A(\theta) >, \\
\frac{d < C_B >}{d\theta} &= -\frac{1}{2}\tau \int_0^{+\infty} [(C_A^1(\alpha,\theta))^2 C_B^1(\alpha,\theta) + (C_A^2(\alpha,\theta))^2 C_B^2(\alpha,\theta)] e^{-\alpha} d\alpha \\
&\quad +\tau k_3 < C_A(\theta) > +1 - < C_B(\theta) > . \qquad (4)
\end{aligned}$$

So the IEM model for NP reaction in component notation can be written as:

$$\begin{aligned}
\frac{\partial C_i^j}{\partial \alpha}(\alpha,\theta) + \frac{\partial C_i^j}{\partial \theta}(\alpha,\theta) &= \lambda\tau (C_1^j)^2 C_2^j - \lambda\tau k_{i+1} C_1^j \\
&\quad + b[< C_i > - C_i^j], \qquad (5)
\end{aligned}$$

$$\begin{aligned}
\frac{d < C_i^j >}{d\theta} &= \lambda\tau \int_0^{+\infty} [(C_1^j)^2 C_2^j - k_{i+1} C_1^j] e^{-\alpha} d\alpha \\
&\quad + C_i^j(0,\theta) - < C_i^j > + b[< C_i > - < C_i^j >] \qquad (6)
\end{aligned}$$

for $i, j = 1, 2 (i \equiv A, B)$ with $C_1^1(0,\theta) = 2, C_1^2(0,\theta) = 0, C_2^1(0,\theta) = 0, C_2^2(0,\theta) = 2$ and $\lambda = \begin{cases} +1 & \text{if} \quad i = 1 \\ -1 & \text{if} \quad i = 2. \end{cases}$

Now suppose $L_n(\alpha)$ be the Laguerre polynomial of degree n and define

$$C_i^j(\alpha, \theta) = \sum_{n=0}^{\infty} F_{i,n}^j(\theta) L_n(\alpha). \tag{7}$$

Then

$$\frac{\partial C_i^j(\alpha, \theta)}{\partial \alpha} = \sum_{n=1}^{\infty} F_{i,n}^j(\theta) L_n'(\alpha). \tag{8}$$

Note that $L_0'(\alpha) = 0$, and

$$\frac{\partial C_i^j(\alpha, \theta)}{\partial \theta} = \sum_{n=0}^{\infty} \frac{dF_{i,n}^j(\theta)}{d\theta} L_n(\alpha). \tag{9}$$

On the other hand, using Rodrigue's formula [1], the orthonormality of the Leguerre polynomials and our definition for $C_i^j(\alpha, \theta)$ (equation 7) we have the following relations:

$$L_n(\alpha) = \frac{e^\alpha}{n!} \frac{d^n}{d\alpha^n}(\alpha^n e^{-\alpha}),$$

$$L_n'(\alpha) = \frac{e^\alpha}{(n-2)!} \frac{d^{n-1}}{d\alpha^{n-1}}(\alpha^{n-2} e^{-\alpha}) - L_{n-1},$$

or

$$L_n'(\alpha) = -L_{n-1} - L_{n-2} - \cdots - L_1 - L_0,$$

so

$$\int_0^\infty L_n(\alpha) L_m'(\alpha) e^{-\alpha} d\alpha = \begin{cases} 0 & \text{if} \quad n \leq m \\ -1 & \text{if} \quad n > m, \end{cases} \tag{10}$$

and

$$\int_0^\infty \sum_{n=1}^{\infty} F_{i,n}^j(\theta) L_n'(\alpha) L_m(\alpha) e^{-\alpha} d\alpha = \lambda \sum_{n=1}^{\infty} F_{i,n}^j(\theta), \tag{11}$$

where $\lambda = \begin{cases} 0 & \text{if } \ n \leq m \\ -1 & \text{if } \ n > m. \end{cases}$ Also, knowing that

$$< C_i(\theta) >= \sum_{j=1}^{2} 1/2 \int_0^\infty C_i^j(\alpha,\theta) e^{-\alpha} d\alpha, \text{ for } \ i = 1, 2,$$

or

$$< C_i(\theta) >= \sum_{j=1}^{2} 1/2 \int_0^\infty \sum_{n=0}^{\infty} F_{i,n}^j(\theta) L_n(\theta) e^{-\alpha} d\alpha,$$

we have

$$< C_i(\theta) >= \sum_{j=1}^{2} 1/2 F_{i,0}^j(\theta) = F_{i,0}^j(\theta). \tag{12}$$

Now, substituting equations (7), (8) and (9) into the equation (5), multiplying both sides of the resultant eqation by $L_m(\alpha)e^{-\alpha}d\alpha$, integrating from 0 to ∞ with respect to α and using the relations (10), (11) and (12) together with equation (6) by some calculation and simplification we obtain our transformed IEM model for NP reaction as follows (for more detail refer to [3]).

$$\frac{dF_{i,k}^j(\theta)}{d\theta} = \lambda\tau \sum_{n=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} F_{1,n}^j(\theta) F_{1,n}^j(\theta) F_{2,m}^j(\theta)$$

$$\int_0^\infty L_n(\alpha) L_n(\alpha) L_m(\alpha) L_k(\alpha) e^{-\alpha} d\alpha - \lambda\tau k_{i+1} F_{1,k}^j(\theta)$$

$$+ C_{0,i}^j - \sum_{n=0}^{k} F_{i,n}^j(\theta) + b[F_{i,0}^j(\theta)\lambda_k - F_{i,k}^j(\theta)]. \tag{13}$$

In above equation (13), $i, j = 1, 2, k = 0, 1, 2, \cdots,\ \ \lambda_k = \begin{cases} 0 & \text{if } \ k \neq 0 \\ 1 & \text{if } \ k = 0, \end{cases}$ $C_{0,1}^1 = 2, C_{0,2}^1 = 0, C_{0,1}^2 = 0$ and $C_{0,2}^2 = 2$.

Equation (13), which we refer to as transformed IEM model for NP reaction, is in the form of ordinary differential equations system and it can be solve much easier than the original IEM model (2) and (3). Indeed, with initial values and special care for computation of non-linear part we can use the continuation soflware BIFPACK [15] and ODE solver LSODES [10] to find bifurcation, stability analysis and time dependent solutions of this transformed system.

So to find the initial values $F^j_{i,n}(0)$, from equation (7) we notice that

$$C^j_i(\alpha, 0) = \sum_{n=0}^{\infty} F^j_{i,n}(0) L_n(\alpha). \tag{14}$$

Multiplying both sides of equation (14) by $L_m(\alpha)e^{-\alpha}$ and taking integral from 0 to ∞, due to the orthonormality of the Laguerre polynomial we get

$$F^j_{i,n}(0) = \int_0^{\infty} C^j_i(\alpha, 0) L_n(\alpha) e^{-\alpha} d\alpha. \tag{15}$$

Now define

$$G^j_{i,n}(\alpha) = \int_0^{\alpha} C^j_i(\alpha, 0) L_n(\alpha) e^{-\alpha} d\alpha. \tag{16}$$

Then

$$\frac{dG^j_{i,n}(\alpha)}{d\alpha} = C^j_i(\alpha, 0) L_n(\alpha) e^{-\alpha}. \tag{17}$$

So $G^j_{i,n}(\alpha) \to F^j_{i,n}(0)$ as $\alpha \to \infty$, (numericaly as $\alpha \to 60$). Therefore, equation (17) together with steady-state of equation (5) yields the intial values for (13).

$$\frac{dC^j_i(\alpha)}{d\alpha} = \lambda\tau(C^j_1)^2 C^j_2 - \lambda\tau k_{i+1} + b[< C_i > - C^j_i]$$

$$\frac{dG^j_{i,n}(\alpha)}{d\alpha} = C^j_i L_n(\alpha) e^{-\alpha}. \tag{18}$$

Here, $i, j = 1, 2$ and $C_i^j(\alpha, 0)$ are the some as before. Note that, as we have said before, since we are going to compare our results to those on [12] and [6], here we have used the same values for $< C_i >, k_2$ and k_3 as in these studies.

Next, during our computation we should compute the integration of non-linear part

$$[\sum_{n=0}^{\infty} F_{1,n}^j(\theta) L_n(\alpha)]^2 \sum_{m=0}^{\infty} F_{2,m}^j(\theta) L_m(\alpha) L_k(\alpha) e^{-\alpha}. \qquad (19)$$

The pseudo-spectral Gauss-Lagurre Integration method [1] gives us the best result. To do this first we define $h_{1,2}(\alpha) = (C_1^j)^2 C_2^j$; from

$$C_i^j(\alpha_p) = \sum_{n=0}^{\infty} F_{i,n}^j(\theta) L_n(\alpha_p),$$

for any $\alpha_p, p = 0, 1, \cdots, P$, we can compute

$$h_{1,2}(\alpha_0), h_{1,2}(\alpha_1), \cdots, h_{1,2}(\alpha_P).$$

Hence, the integration of (19) will be

$$\int_0^{\infty} h_{1,2}(\alpha) L_k(\alpha) e^{-\alpha} d\alpha. \qquad (20)$$

Now, using Gauss-Laguerre Integration method the above integration (20) can be approximated by

$$\sum_{p=1}^{P} A_p h_{1,2}(\alpha_p) L_k(\alpha_p),$$

where

$$A_p = \frac{(P!)^2 \alpha_p}{(P+1)^2 [L_{P+1}(\alpha_p)]^2}.$$

Here, if we choose α_p to be the zero of the Laguerre polynomial of degree P, the integration will be exact for the function $h_{1,2}(\alpha) L_k(\alpha)$ being the polynomial of degree at most 2P-1.

We should notic that the only error in solving the transformed IEM model is the trancation error. Here for NP reaction, we have used 6 terms of the polynomial expansion and the results are very similar to those found by Puhl, Nicolis [12] and Fox [6] for steady-state case.

4. Bifurcation and Stability Analysis

Based on the techniques discussed in the previous section and knowing that

$$C_i^j(\alpha,\theta) \simeq \sum_{n=0}^{N} F_{i,n}^j(0) L_n(\alpha),$$

bifurcation and stability analysis of the NP reaction has been computed using continuation software BIFPACK [15] and ODE solver LSODES [10]. The results of these computations are summarized in Figures 1 and 2. First we compute (b,τ^{-1}) bifurcation sets, since these are the two operating parameters which can be varied in an actual experiment [6]. Indeed the computations were pereformed by fixing k_2 and k_3 so that a TB point existed in the perfect micromixing limit. The possible values for these parameters are given in [12] where the TB point has been parameterized by η, the value of the steady-state concentration of first component ($< C_1 >$) at the TB point. By this parametrization, it is convenient to refor to the parameter pair (k_2, k_3) by its corresponding η-value, denoted by $< \eta >_{TB}$. The values of $< \eta >_{TB}, k_2$ and k_3 used in the first series of computations are the same as Fox's values [6] and are given by (0.3, 0.615, 0.54609), (0.35, 0.65779, 0.58616), (0.375, 0.67634, 0.60459), (0.4, 0.69333, 0.62222), (0.425, 0.709, 0.63921), (0.46, 0.7291, 0.66215), (0.475, 0.73716, 0.67175), (0.5, 0.75, 0.6875), (0.55, 0.77393, 0.71837) and (0.6, 0.79636, 0.74876).

The results of the computations for each of the above-mentioned parameter sets are shown in Figure 1. Hopf and saddle-node bifurcation curves are plotted as dash and solid curves labled H and S, respectively. These two curves were located using BIFPACK with the steady-state system of equations (13) (for $i, j = 1, 2$ and $k = 0, 1, 2, \cdots, 6$). For all values of the parameters used in this section, only the real part of two eigenvalues were changed their signs (saddle-node bifurcation) or only two of them were purely imaginary (Hopf bifurcation). This implies that the dynamical dimension of the transformed IEM model is two for the NP reaction for all values of b.

The existence of periodic solutions to the unsteady-state transformed IEM model for parameter values near the H-curve has been confirmed by solving numerically the time-dependent model equations (13) and the results are shown in next section.

The parameter values used in Figure 1(a)-(c) result in a bifurcation set like the one shown in [12] and [6]. For example as occurs when moving from Figure 1(a) to 1(b), the S-curve above the TB point in Figure 1(a)-(c) becomes smaller and smaller, until the cusp and TB points concide. (Figure 1(d)). The same qualitaive changes can be seen in [12] and [6] for the parameter space (k_3, τ^{-1}). Thus, qualitatively, the parameters b and k_3 have the same effect in the NP reaction: decreasing either parameter while varying τ^{-1} results in an unfolding of the Hopf and saddle-node bifurcations.

It is imporfant to note, however, that for a real reaction the rate constants would be fixed, leaving only manipulatable parameters such as the residence time and the stirring rate as possible bifurcation parameters. Hence, Figure 1 can be viewed as a representation of the observable bifurcation sets in a real reactor

[6]. From this figure, it can be seen that the sensitivity of the NP reaction to micromixing depends strongly on the value of the kinetic rate parameters. In Figure 1(a), the H-curve is relatively insensitive to the value of b in the rang $[100, 1000]$, while the S-curve exhibits a slightly greater dependence. In Figure 1(c), the S-curve exists only near $b = 1000$ and disappears altogether in Figure 1(d). In these figures, the H-curve is seen to move towards larger τ^{-1}-values as b decreases, suggesting that Hopf bifurcations can be induced from a steady state by lowering the stirring rate with the flow rate held constant. In Figure 1(e)-(g) the H-curve moves out of the top of the figure making Hopf bifurcations more and more dependent of the micromixing rate. Finally, in Figure 1(h), the H-curve has disappeared, leaving only the S-curve in the remaining figures. The greatest amount of sensitivity to the micromixing rates occurs, approximatly, in Figure 1(f) where the parameter values correspond to those used in Puhl and Nicolis [12] and Fox [6]. In this figure, the Hopf and saddle-node bifurcation curves both lie above $b = 100$ (thus, in the rangs of validity of the both IEM and fransformed IEM models); hence, for these rate constants, the reaction is very sensitiue to micromixing effects [9].

5. Time Dependent, Periodic Solutions and Chaoc

In addition to the relative simplicity of transormed IEM model to find the solution and bifurcation analysis of NP reaction, it has the interest of finding the time dependent solution of this reaction. Indeed we have employed the same technique outlined in section 3 to investigate time dependent solution to the transformed IEM model for NP reaction. Example solutions for several types of fixed points will be presented here: unstable, periodic after the potential Hopf bifurcation point and even unexpected chaotic point.

We have began time-dependent solution of fixed saddle points in Figure 1. As an example consider the point 1 in Figure 1(a) for values $< \eta >_{TB} = (< C_1 >) = 0.30, \tau^{-1} = 0.078736$ and $b = 750$. We find the result shown in Figure 2(a). As we can see starting from this point as expected the time-dependent solution will change. This is because in finding this bifurcation point, using BIFPACK, in section 4 the corresponding two eighenvalues of this point had different sign of real parts. The similar results to this in Figure 1 have shown that the saddle points found in steady-state case are indeed unstable solutions of the unsteady-state transformed IEM model.

Also of interest is the periodic solution to the transformed IEM model for different parameter b at Hopf bifurcation. In this case if we choose a point along the H-curve in Figure 1 as an initial condition, the periodic solution can be observed by solving the unsteady-state transformed IEM model. Thus, taking point 3 from Figure 1(b) corresponding to $< \eta >_{TB} = (< C_1 >) = 0.35, \tau^{-1} = 0.0981315$ and we find the result presented in Figure 2(c). Similar result can be found for other points along the H-curves in Figure 1. Figure 2(d) is another example of a periodic solution to the transformed IEM model corresponding to the point 4 of Figure 1(b) for $< \eta >_{TB} = (< C_1 >) = 0.35, \tau^{-1} = 0.0935228$ and $b = 200$. The results illustrated in Figures 2(c)-(d) have shown that periodic solutions along the H-curves can be checked by using unsteady-state trunsformed IEM model. Nevertheless, due to the relatively long compoting times required to find the periodic solution by solving the unsteady-state transformed model, only a limited number of parameter values, could actually be investigated by directly solving the time dependent transformed IEM model equations.

Another solution that we have found by sloving the unsteady-state model is unexpectated chaodic solution which we absorved by choosing a few points such as point 5 in Figure 1(a) correspond to $< \eta >_{TB} = (< C_1 >) = 0.30$, $\tau^{-1} = 0.07874$ and $b = 450$. This solution illustrated in Figure 2(e). As we can see starting from this point we never know when we return to the same point. Of course this kind of the solution was to rare during our investigation.

6. Conclusions

The class of IPDE introduced here and elsewhere [2], [3], [6] and [8] can be used to steady the dynamical behavior of many systems [5], [6] and [8]. An unsteady-state version of the IEM model, as the resultant system of this IPDE, however can rarely be solved analytically. So we have derived a polynomial expansion of this model that could be used to investigate the dynamical behavior of the system in general, and the behavior of unsteady-state solutions, in particular.

The numerical simulation procedure, for transformed IEM model, employed in this work represents a direct, albeit time-consuming method for constructing the bifurcation set and time-dependent solutions for imperfectly-micromixed reactions such as NP reaction. Other application of this model can also be developed [5].

As we have seen the steady-state dynamical behavior of the NP reaction can be very sensitive to the degree of micromixing. Using the transformed IEM model, bifurcation sets have been constructed wherein Hopf and saddle-node bifurcation curves in micromixing rate (b) and the mean residence time (τ^{-1}) parameters space. These results are agree with those are found by others [6], [7] and [12].

Acknowledgment. The auther wish to express his gratitude to Shiraz University Research Council for the support of this work under grant number 77-SC-1112-634.

References

1. Abramowitz, M. and Stegun, A.: Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables, Dover Publications. New York, 1965.

2. Erjaee, G.H.: Derivation of the Unsteady-states Interaction-by-Exchange with the Mean Model, Iranian Journal of Science and Technology, 1998, Vol. 22 No. 2.

3. Erjaee, G.H.: Numerical Simulation of a class of IPDE using a Transformation Technique. Bulletin of the Iranian Mathematical Society, 1996, Vol. 22, No. 2.

4. Fathei, A. and Menzinger, M.: Inhomogeneity of a CSTR with Autocatalator: A Classification of Stirring Effects. J. Phys. Chem. 1992, Vol. 96.

5. Fox, R.O., Erjaee, G.H. and Zou, Q.: Bifurcation and Stability Analysis of Micromixing Effects in the Chlorite-Iodide Reaction. Chem. Engng. Sci. 1993.

6. Fox, R.O.: Micromixing Effects in the Nicolis-Puhl Reaction: Numerical Bifurcation and Stability Analysis of the IEM model Chem. Engng. Sci. 1991, Vol. 46, No. 7.

7. Fox, R.O.: Steady-state IEM Model : Singular Perturbation Analysis near Perfect-Micromixing Limit. Chem. Engng. Sci. 1989, Vol. 44.

8. Fox, R.O. and Villermaux, J.: Micromixing Effects in the $ClO_2 + I$ Reaction : Perturbation Analysis and Numerical Simulation of the Unsteady-state IEM model. Chem. Engng. Sci. 1990, Vol. 45.

9. Fox, R.O. and Villermaux, J.: Unsteady-state IEM Model Numerical Simulation and Multiple-scale Perturbation Analysis near Perfect-Micromixing Limit. Chem. Engng. Sci. 1990, Vol. 45.

10. Hindmarsh, A. and Sherman, A.: Livermore Solver for ODE, in Scientific Computing, North-Holland, Amesterdam, Stepleman, R.S. 1988.

11. Pope, S.B.: PDE Methods for Turbulent Reactive Flows. Prog. Energy Combust. Sci. 1985, Vol. 11.

12. Puhl, A. and Nicolis, G.: Micromixing and Multiple Steady-state Transitions in a CSTR, Chem. Engng. Sci. 1987, Vol. 41.

13. Ray, W.H.: Bifurcation Phenomena in Chemically Reacting Systems. Applications of Bifurcation Theory, 285, Academic Press 1977.

14. Ray, W.H. and Jenson, K.F.: Bifurcation Phenomena in Stirred Tanks and Catalytic Reactors. Methematics Research Center and Department Engineering, University of Wisconsin-Medison.

15. Seydel, R.: From Equilibrium to Chaos; Practical Bifurcation and Stability Analysis. Elsevier, New York 1988.

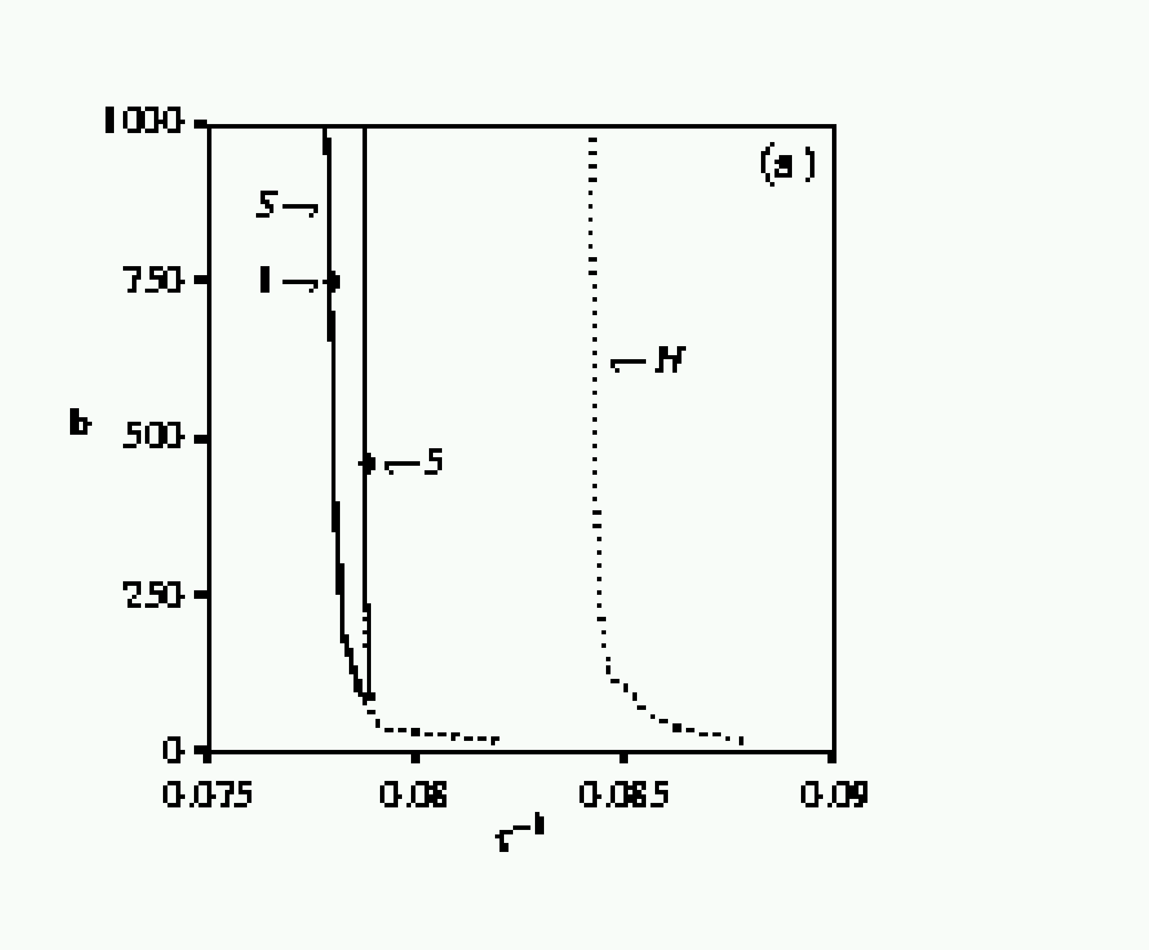
16. Villermaux, J.: Mixing Effects in non-ideal Chemical Reactions in Spatial Inhomogeneities and Transeint Behavior in Chemical Kinetics. Manchester University Press, Manchester, U.K., 1989.

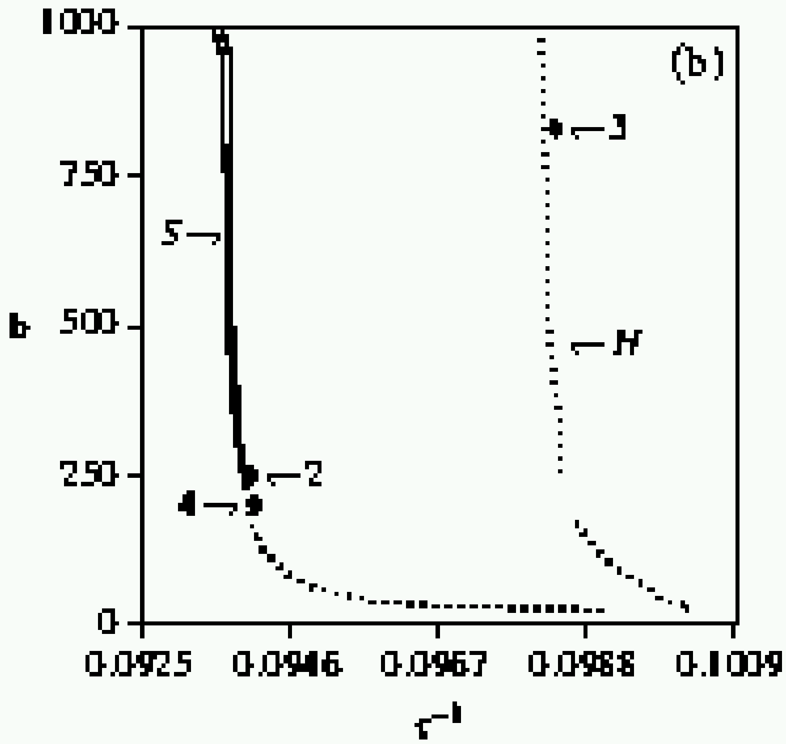
17. Villermaux, J.: Micromixing Phenomena in Stirred Reactors, in Encyclopedia of Fluid Mechanics. Gulf, Houston, TX, 1986.

List of Figures

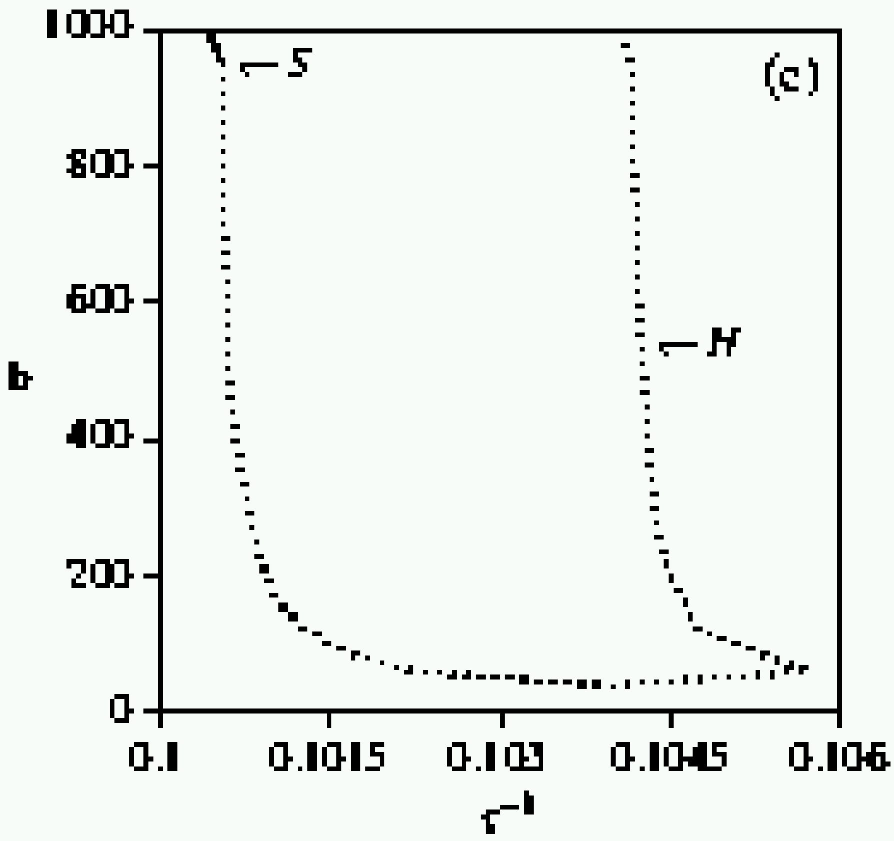
Figure 1. Bifurcation sets in (b, τ^{-1}) parameter space for the NP reaction using transformed IEM model. The kinetic rate constants k_2 and k_3 are given in the text for all figures. 1(a) to 1(j).

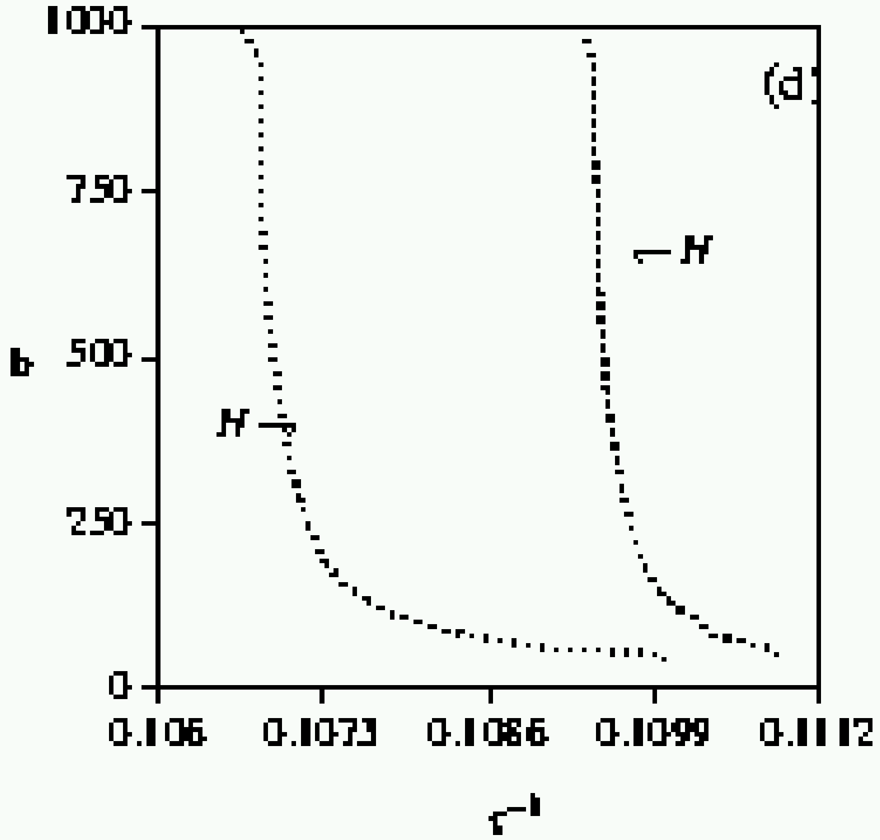
Figure 2. Unsteady-state solution for NP reaction using transformed IEM model: (a), (b) unstable solutions, (c), (d) periodic solutions with relatively small and large periods and (e) chaodic solution.

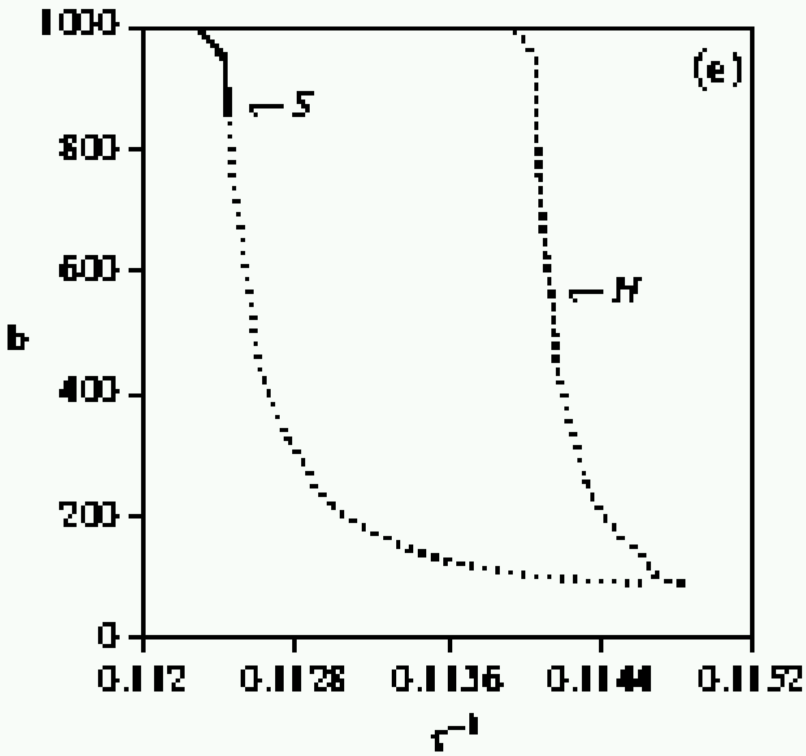

(a)
1000
750
500
250
0
b
0.075
0.08
0.085
0.09
τ⁻¹
5
1
5
H

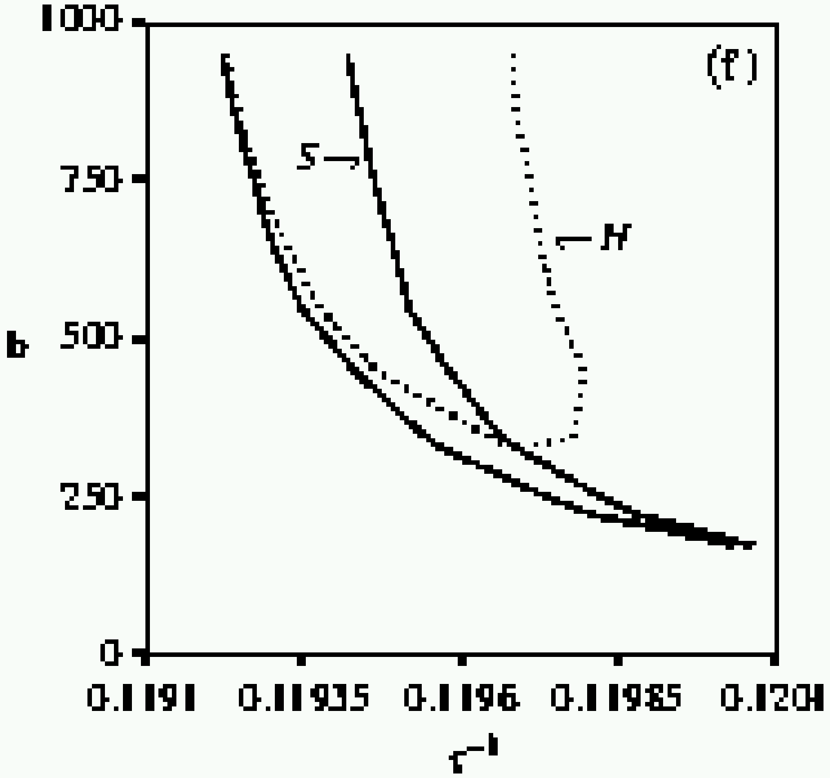


(b)
1000
750
500
250
0
b
0.0925
0.0946
0.0967
0.0988
0.1009
5
2
4
3
H

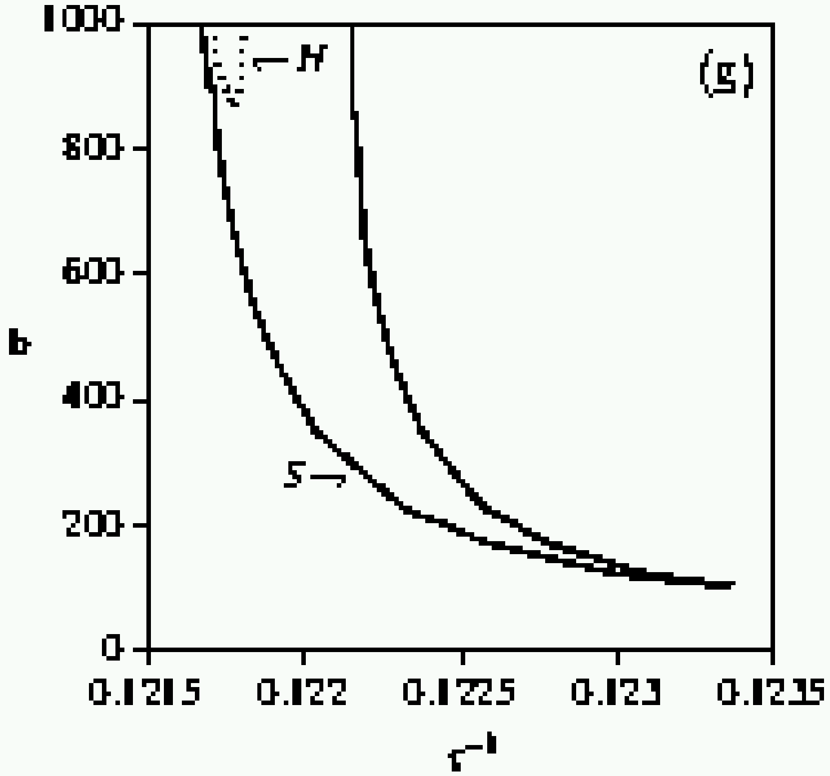

1000
800
600
400
200
0
b
0.1
0.1015
0.103
0.1045
0.106
τ^{-1}
S
H
(c)

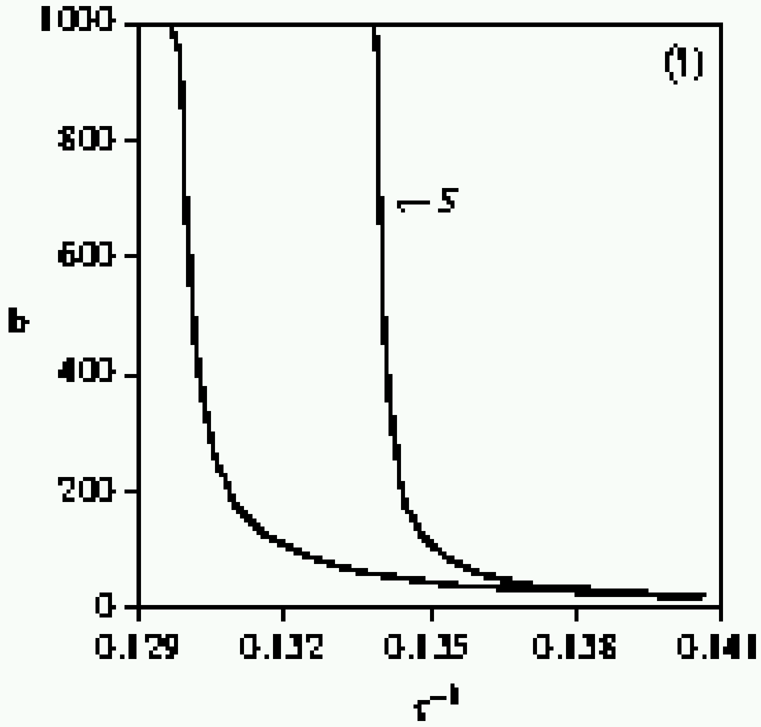

1000
750
500
250
0
b
(d)
H
H
0.106
0.1073
0.1086
0.1099
0.1112
c^{-1}

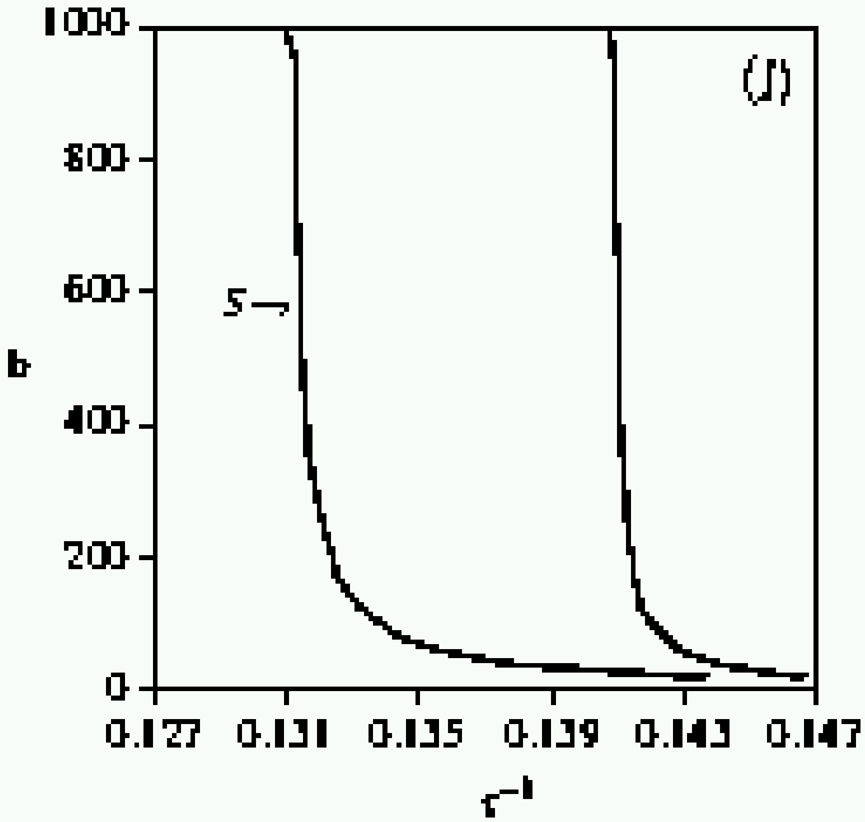

1000
800
600
400
200
0
b
0.112
0.1128
0.1136
0.1144
0.1152
τ^{-1}
(e)
S
H

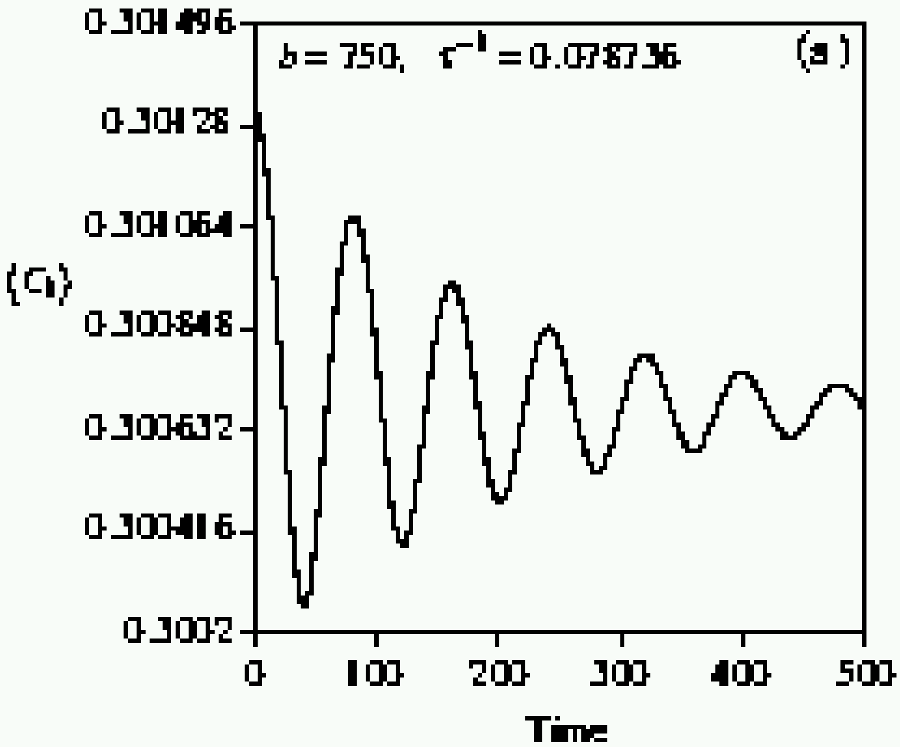


1000
750
500
250
0
b
0.1191
0.11935
0.1196
0.11985
0.1201
r⁻¹
(f)
S
H

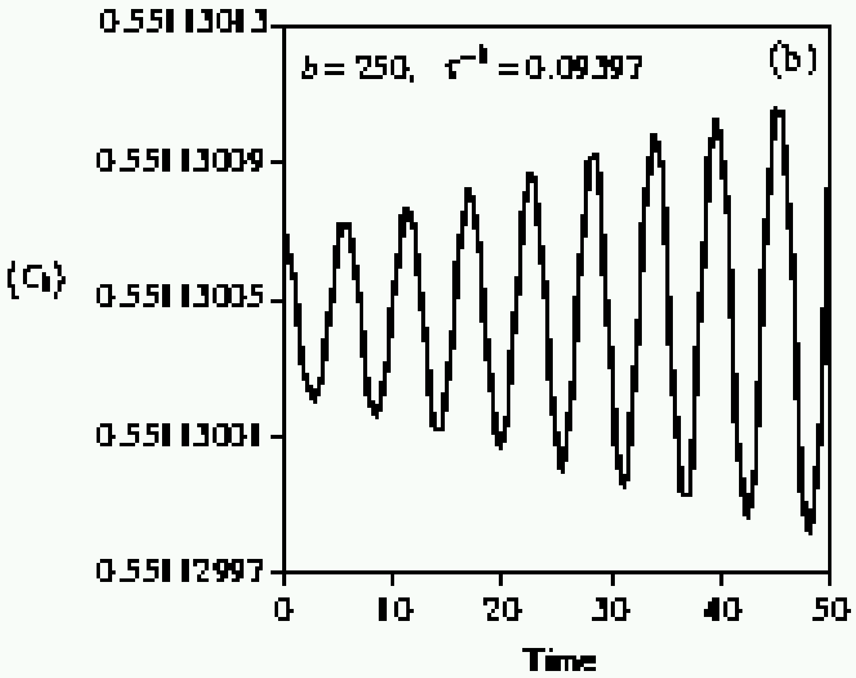

1000
800
600
400
200
0
b
H
S
(g)
0.1215
0.122
0.1225
0.123
0.1235
r⁻¹

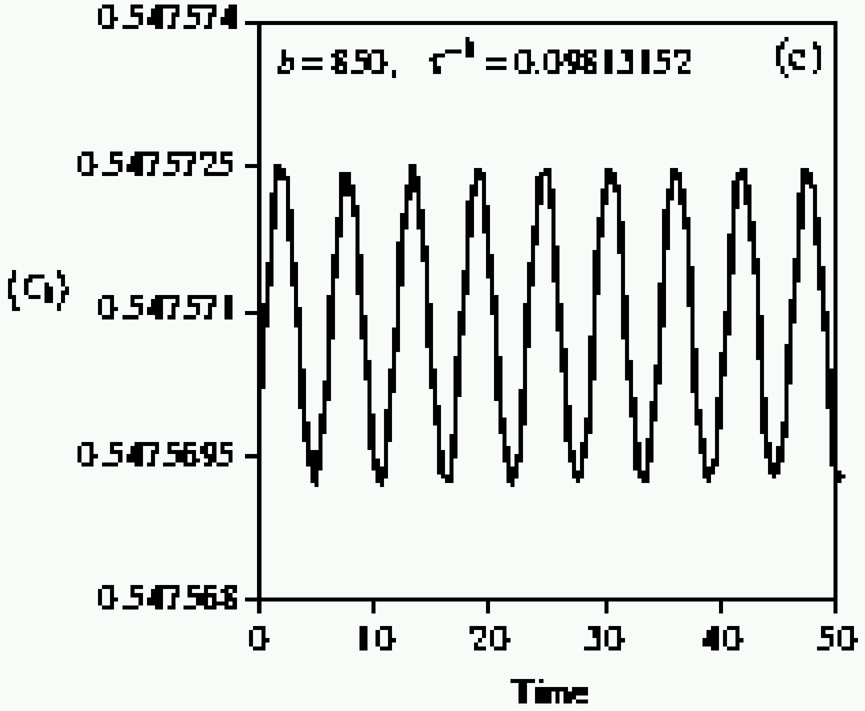

1000
800
600
400
200
0
b
0.129
0.132
0.135
0.138
0.141
τ⁻¹
(1)
5

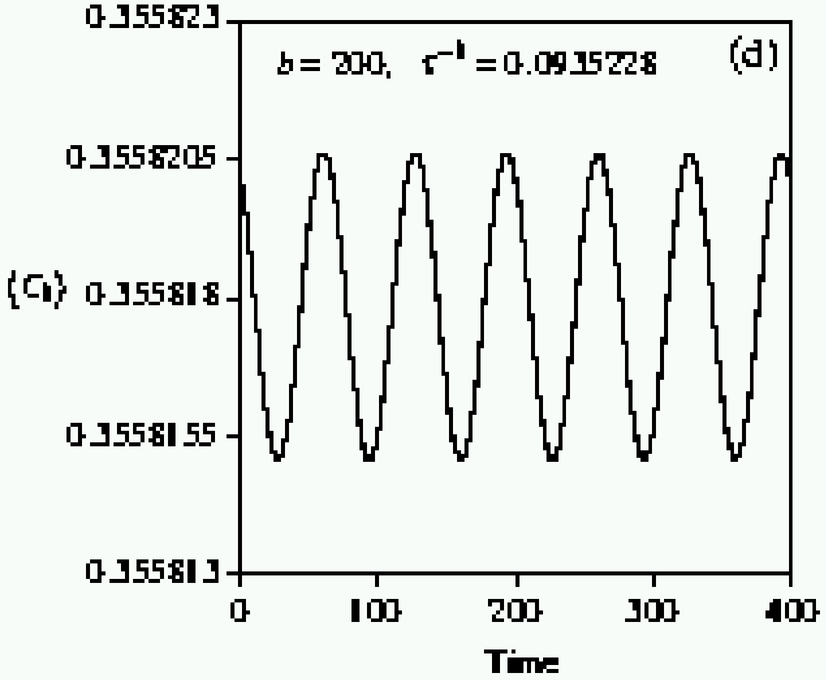

1000
800
600
400
200
0
b
5
(J)
0.127
0.131
0.135
0.139
0.143
0.147
r^{-1}



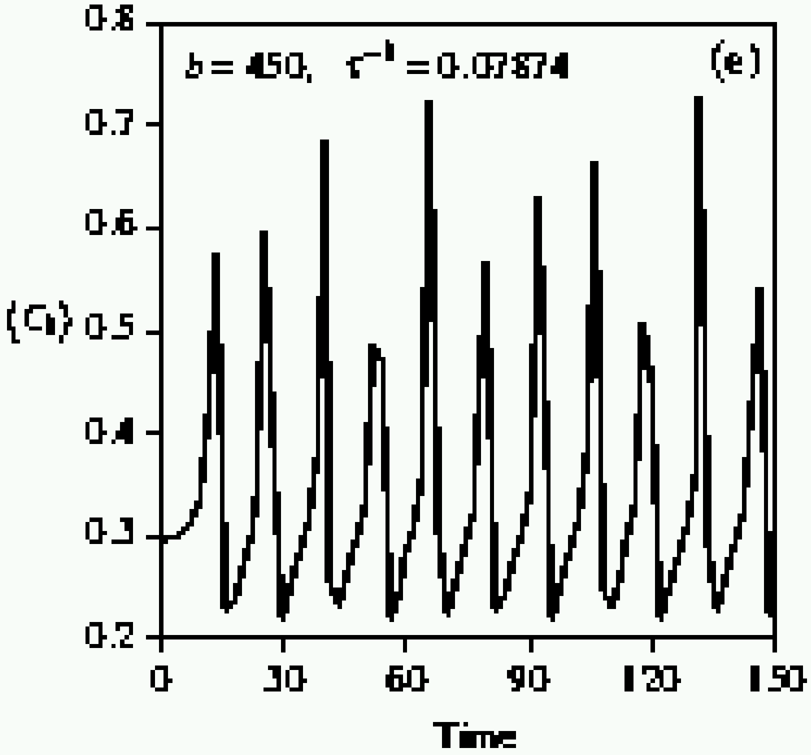
b = 750, τ^-1 = 0.078736
(a)
0.301496
0.30128
0.301064
0.300848
0.300632
0.300416
0.3002
0
100
200
300
400
500
Time


0.55113013
0.55113009
0.55113005
0.55113001
0.55112997
$\langle C_1 \rangle$
$b = 250, \quad \tau^{-1} = 0.09397$
(b)
0
10
20
30
40
50
Time


b = 850, τ^-1 = 0.09813152
(c)
0.547574
0.5475725
0.547571
0.5475695
0.547568
(C1)
0
10
20
30
40
50
Time



0.355823
0.3558205
0.355818
0.3558155
0.355813
$b = 200$, $\tau^{-1} = 0.0935228$
(d)
$\langle C_1 \rangle$
0
100
200
300
400
Time



$b = 450, \quad \tau^{-1} = 0.07874$
(e)
$\langle c_1 \rangle$
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
30
60
90
120
150
Time

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2000, pp. 122 – 151
Submitted: September 3, 2000. Published: December 15, 2000

Finding an Ellipse Tangent to finitely many given Lines

Alan Horwitz

Abstract

Given a finite set Σ of lines in the plane, we discuss necessary and sufficient conditions on when there is an ellipse E, with specified center (h,k) and angle of rotation α, which is tangent to each line in Σ. In all cases we assume that no three of the lines are parallel or have a common intersection point, else no such ellipse could exist. If such an ellipse exists, we say that (h,k) is α admissible, or just admissible if $\alpha = 0$. For two given lines $T_1 : y-k = m_1(x-h)+b_1$ and $T_2 : y-k = m_2(x-h)+b_2$, with $|m_1| \neq |m_2|$, (h,k) is admissible if and only if $\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ and $\frac{b_1^2m_2^2-b_2^2m_1^2}{m_2^2-m_1^2}$ are both positive. Further, the ellipse is unique. We prove similar results when $|m_1| = |m_2|$, in which case the ellipse may not be unique. In certain cases we allow α to vary. We then show that every $(h,k) \notin T_1 \cup T_2$ is α admissible for some α (if $m_1 \neq m_2$). We prove various results for three given lines. In particular, if none of the slopes of the lines are equal in absolute value, and none of the lines are horizontal, then there are cubic polynomials $q(k)$ and $r(h)$ such that the set of admissible centers is precisely the set of points on a hyperbola where q and r are positive. q and r are obtained using the intersection points of the given lines. Finally, for four given lines, we show that there is always some ellipse, rotated of course, tangent to the given lines.

A.M.S. Subject Classification Codes: 51
Key Words and Phrases: Ellipse, rotation, tangent

1. Introduction

Many curve fitting problems involve finding a curve C of a certain type (such as a polynomial of degree n or conic section) which passes through specified points (interpolation), perhaps with specified slopes at those points as well(Hermite interpolation). In this paper we consider a different type of curve fitting: Given a set Σ of finitely many lines and a family $\digamma$ of curves in the plane, find, if possible, a curve $\sigma \in \digamma$ such that each line in Σ is tangent to σ. We emphasize here that the *points of tangency are unknown.* This problem is similar to finding the envelope of a family of lines, except that here we only have *finitely* many lines. To reconstruct a general curve given *all* of its tangents, see [**4**]. The tradeoff is that you then must have some information about σ, the more information the less tangents are needed to reconstruct σIn this paper we assume that $\digamma$ is the family, or some subfamily, of **ellipses** in the plane. Most of our results, however, extend easily to hyperbolas as well. The subfamilies we discuss are ellipses with a specified center or angle of rotation. Many of our results are proven for non-rotated ellipses $\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$, but extend to any fixed angle of rotation. We say that the ellipse has rotation angle α if the major and minor axes have been rotated clockwise through α from the coordinate axes.

It is useful to make the following definition.

DEFINITION 1.1. *Given a finite set of distinct lines $T_1, T_2, ..., T_n$ in the plane, and an angle α, $-\dfrac{\pi}{2}<\alpha<\dfrac{\pi}{2}$, suppose that there is an ellipse with center (h,k) and rotation angle α which is tangent to each of the T_j. Then we say that (h,k) is α* ***admissible****. If $\alpha=0$ we just call (h,k) admissible.*[1]

Of course α admissibility depends on the given lines T_j. In all cases we assume that no three of the lines are parallel or have a common intersection point, else no such ellipse could exist for any α or (h,k).

We discuss the following questions:

(1) Given two distinct lines in the plane, T_1 and T_2, and a point $C=(h,k)\notin T_1\cup T_2$, under what conditions is C admissible. Is the ellipse always unique ?

(2) Is *any* point $(h,k)\notin T_1\cup T_2$ admissible? If not, is *any* point $(h,k)\notin T_1\cup T_2$ α admissible for some α?

(3) For three given lines, what does the set of admissible centers look like for fixed α? What about specifying the eccentricity in advance?

(4) Suppose that we are given four lines, such that no three of the lines are parallel or have a common intersection point. Is there always some ellipse (rotated) tangent to the four lines?

We summarize our results. In Section 2, we give necessary and sufficient conditions for when two lines T_1 and T_2 are tangents to some non-rotated ellipse with center (h,k)(see Theorems 2.4 and 2.6). More precisely, in Theorem 2.4 we assume that T_1 and T_2 are distinct, non–vertical lines with equations $y-k=m_j(x-h)+b_j$, $m_1^2\neq m_2^2$. Then then there is an ellipse $\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$, which has T_1 and T_2 as tangents, if and only if $\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ and $\frac{b_1^2m_2^2-b_2^2m_1^2}{m_2^2-m_1^2}$ are both positive. Further, the ellipse is unique with $a^2=\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ and $b^2=\frac{b_1^2m_2^2-b_2^2m_1^2}{m_2^2-m_1^2}$. In certain cases such an ellipse may *not exist* or be *unique*. For example, if T_1 and T_2 have equations $y=x+2$ and $y=2x-3$, then there is *no* ellipse E (nonrotated) with center $(2,0)$, such that T_1 and T_2 are tangents to E. On the other hand, if T_1 and T_2 have equations $y=2x+3$ and $y=-2x+3$, and $h=k=0$, then T_1 and T_2 are tangent to any ellipse in the family $\mathcal{F}=\{\frac{x^2}{a^2}+\frac{y^2}{b^2}=1: 4a^2+b^2=9\}$.

If one allows the angle of rotation to vary, then for any $(h,k)\notin T_1\cup T_2$ (with $m_1\neq m_2$) is α admissible for some α (see Theorem 2.10). Also, for fixed α, there is always either an ellipse or hyperbola with angle of rotation α tangent to T_1 and T_2 (see Theorem 2.13).

For three given lines, the set of admissible centers either lies on a hyperbola or a straight line. See Theorems 3.1, 3.3, 3.5, and 3.6. In addition, one can choose the center so that the ellipse has any specified eccentricity. In particular, in Theorem 3.1, we assume that none of the slopes of the lines are equal in absolute value, and none of the lines are horizontal. We apply Theorem 2.4 to T_1 and T_2, and to T_1 and T_3. To get one ellipse tangent to all three lines, the a^2 and b^2 obtained in both

[1]One usually rotates *counterclockwise* if $\alpha>0$, but we prefer to use a clockwise rotation.

cases must be equal. This leads to the equation of a hyperbola γ on which the set of admissible centers S must lie. More precisely, $S = \{(h,k) \in \gamma : q_3(k) > 0$ and $r_3(h) > 0\}$, where $q(k)$ and $r(h)$ are cubic polynomials obtained using the intersection points of the T_js.

For *four* given lines, there is always *some* ellipse, rotated of course, tangent to the given lines(see Theorem 4.3). By using an affine map, one only need prove this result for the special case when the four lines have interesection points $(0,0),(1,0),(0,1)$, and (s,t), with $s>0, t>0,\ s \neq 1 \neq t,\ s+t \neq 1$.

2. Two Given Tangents

The theorems in this section are stated for ellipses with major and minor axes parallel to the x and y *axes*. Later we indicate how to extend the results easily using a rotation map. Indeed, the proofs we give all assume that the center C of E is $(0,0)$. If $C=(h,k)$, let $u = x-h, v = y-k$, and let $\bar{E}$ be the ellipse in the $uv-plane$ which is the image of E under the translation map $T : (x,y) \to (u,v)$. Then $\bar{E}$ has center $(0,0)$, and T maps tangents to E onto tangents to $\bar{E}$.

LEMMA 2.1. *Let E be the ellipse $\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$, and suppose that T is a nonvertical tangent to E at (c,d). Write the equation of T in the form*

$$y - k = m(x-h) + B.$$

Then $B \neq 0$ and $d \neq 0$.

PROOF. The equation of T can be written in the form $\frac{c(x-h)}{a^2} + \frac{d(y-k)}{b^2} = 1$. If $B = 0$, then T passes thru (h,k), which is impossible. If $d = 0$, then T is vertical. $\square$

Suppose that T_1 and T_2 are given distinct lines in the $xy-plane$, which are known to be tangent to an ellipse E with center (h,k). If h and k are known, how do you find the equation of E–i.e. what are a and b in $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$? Also, is E unique ? The next two theorems answer this question.

PROPOSTION 2.2. *Suppose that we are given distinct, non-vertical lines T_1 and T_2, which are known to be tangent to some ellipse E with center $C = (h,k)$. Suppose that T_j has equation*

$$y - k = m_j(x-h) + b_j \tag{2.1}$$

$j = 1,2$, with $m_1^2 \neq m_2^2$. Then the equation of E is $\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$, where $a^2 = \dfrac{b_2^2 - b_1^2}{m_2^2 - m_1^2}$ and $b^2 = \dfrac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2 - m_1^2}$.

PROOF. **We assume first that** $C = (0,0)$. Then E has equation $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$. We shall derive formulas for a^2 and b^2. Suppose that T_j is tangent to E at $(x_j, y_j), j = 1,2$. Since $dy/dx = (-b^2/a^2)(x/y)$ at any point of $E(y \neq 0)$, we have

$$m_j = -\frac{b^2}{a^2}\frac{x_j}{y_j} \tag{2.2}$$

Note that T_j not vertical implies that $y_j \neq 0$, by Lemma 2.1. Also

$$y_j = m_j x_j + b_j \tag{2.3}$$

Using (2.2) and substituting (2.3) gives $a^2 m_j(m_j x_j + b_j) + b^2 x_j = 0$, which implies

$$x_j = -\frac{a^2 m_j b_j}{a^2 m_j^2 + b^2} \tag{2.4}$$

By (2.3) we also have

$$y_j = \frac{b^2 b_j}{a^2 m_j^2 + b^2} \tag{2.5}$$

Using (2.4), (2.5), and $\dfrac{x_j^2}{a^2} + \dfrac{y_j^2}{b^2} = 1$ implies that $\dfrac{a^2 m_j^2 b_j^2 + b^2 b_j^2}{(a^2 m_j^2 + b^2)^2} = 1 \Rightarrow b_j^2 = a^2 m_j^2 + b^2$. This gives the following linear system in the unknowns a^2 and b^2, which is nonsingular since $m_1^2 \neq m_2^2$:

$$\begin{array}{rcl} m_1^2 a^2 + b^2 & = & b_1^2 \\ m_2^2 a^2 + b^2 & = & b_2^2 \end{array} \tag{2.6}$$

with unique solution

$$a^2 = \frac{b_2^2 - b_1^2}{m_2^2 - m_1^2},\ b^2 = \frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2 - m_1^2} \tag{2.7}$$

Now suppose that $C = (h,k)$, and suppose that the equation of E is $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$. Let $u = x-h, v = y-k$, and let $\bar{E}$ be the ellipse in the $uv-plane$ which is the image of E under the translation map $(x,y) \to (u,v)$. Then $\bar{E}$ has center $(0,0)$, the equation of $\bar{E}$ is $\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1$, and the lines $\bar{T}_j$ with equation $v = m_j u + b_j$ are tangent to $\bar{E}$ at $(u_j, v_j) = (x_j - h, v_j - k)$. By the case just proven, a^2 and b^2 are given by (2.7). □

Remark 2.3. If $m_1^2 = m_2^2$, then the ellipse E in Theorem 2.2 is *not unique.* This will follow from Theorem 2.6.

Now suppose that two distinct lines T_1 and T_2 are given which are *not* known to be tangent lines to some ellipse, as in Theorem 2.2. Given $C = (h,k)$, are T_1 and T_2 tangent to some ellipse with a *specified* center C ? The conditions on T_1 and T_2 are more restrictive, as the following theorem shows.

Theorem 2.4. *Let h and k be given real numbers, and let T_1 and T_2 be distinct, non–vertical lines with equations $y - k = m_j(x-h) + b_j,\ j = 1,2$. Suppose that $m_1^2 \neq m_2^2$.*

Part 1: If

(A)

$$\frac{b_2^2 - b_1^2}{m_2^2 - m_1^2} > 0$$

and

(B)

$$\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2 - m_1^2} > 0$$

then there is a <u>unique</u> ellipse E, with center (h,k), which has T_1 and T_2 as tangents. Furthermore, E has equation $\dfrac{(x-h)^2}{a^2}+\dfrac{(y-k)^2}{b^2}=1$, where a^2 and b^2 are given by (2.7).

Part 2: (A) and (B) are also necessary conditions for T_1 and T_2 to be tangent to some ellipse with center (h,k).

PROOF. Part 1: It is not hard to show that (A) and (B) imply

$$b_1 \neq 0 \neq b_2$$

We prove the case $h=k=0$, the general case following easily by a translation map, as in the proof of Proposition 2.2. Then T_j has equation $y=m_jx+b_j,\ j=1,2$. Let a^2 and b^2 be defined as in (2.7), and let x_j and y_j be defined by (2.4) and (2.5). By (A) and (B), a^2 and b^2 are well–defined and positive, and x_j and y_j are well–defined. Now $m_jx_j+b_j = \frac{-a^2m_j^2b_j^2+a^2m_j^2b_j+b^2b_j}{a^2m_j^2+b^2} = \frac{b^2b_j}{a^2m_j^2+b^2} = y_j$. Note that by Lemma2.1, $y_j \neq 0$. Also, since the b_j are nonzero, $\frac{x_j}{y_j} = \frac{-a^2m_jb_j}{b_jb^2} = -\frac{a^2m_j}{b^2} \Rightarrow dy/dx\mid_{x=x_j,y=y_j} = \frac{-b^2x_j}{a^2y_j} = m_j$. Letting E be the ellipse with equation $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, we have proven that the T_j are tangent to E(at (x_j,y_j). Now if $\tilde{E}$ is any ellipse with center $(0,0)$, with the T_j tangent to $\tilde{E}$, then the equation of $\tilde{E}$ is $\frac{x^2}{\tilde{a}^2}+\frac{y^2}{\tilde{b}^2}=1$. By Proposition 2.2, $\tilde{a}$ and $\tilde{b}$ are given by (2.7). But then $\tilde{a}=a$ and $\tilde{b}=b$, which implies that $\tilde{E}=E$, and that proves uniqueness.

Part 2: This follows immediately from Proposition 2.2. $\square$

2.0.1. *Examples.* (1) Suppose that the lines $T_1:\ y=x+3$ and $T_2:\ y=2x-3$ are given, and we want an ellipse E with center $(1,2)$, such that T_1 and T_2 are tangents to E. First we must expand about $(1,2)$. $T_1:\ y-2=(x-1)+2$ and $T_2:\ y-2=2(x-1)-3$. Then $m_1=1, b_1=2, m_2=2, b_2=-3$. Since the conditions of Theorem 2.4 are satisfied, E exists and is unique. From $a^2=\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ and $b^2=\frac{b_1^2m_2^2-b_2^2m_1^2}{m_2^2-m_1^2}$ we obtain the equation of $E:\ \frac{(x-1)^2}{\frac{5}{3}}+\frac{(y-2)^2}{\frac{7}{3}}=1$. See Figure 2.1 for a plot of T_1, T_2, and E.

Note that by using (2.4) and (2.5), along with the translation $(x,y)\to(x+1,y+2)$, one obtains the *points of tangency*: $(\frac{1}{6},\frac{19}{6})$ and $(\frac{19}{9},\frac{11}{9})$.

(2) The lines $T_1:\ y=x+2$ and $T_2:\ y=2x-3$ are given, and we want an ellipse E (nonrotated) with center $(2,0)$, such that T_1 and T_2 are tangents to E. Then $m_1=1, b_1=4, m_2=2, b_2=1$, which implies that $\frac{b_2^2-b_1^2}{m_2^2-m_1^2}<0$. By Theorem 2.4, Part 2, there is **no** such ellipse E.

REMARK 2.5. One can also obtain results similar to Proposition 2.2 or Theorem 2.4 if one of the T_j(but not **both**) is *vertical.* If T_1 has equation $x-h=b_1$,then $b_1=a$ or $-a$, which implies that $a^2=b_1^2$. Since $m_2^2a^2+b^2=b_2^2$ still holds, one then solves to get b^2.

THEOREM 2.6. *Let h and k be given real numbers, and let T_1 and T_2 be distinct, non–vertical lines with equations $y-k=m_j(x-h)+b_j,\ j=1,2$.*

Suppose that $m_1^2=m_2^2$.

(i) If $b_1^2=b_2^2\neq 0$, then T_1 and T_2 are tangent to any ellipse $E\in\digamma=$ the family of ellipses $\dfrac{(x-h)^2}{a^2}+\dfrac{(y-k)^2}{b^2}=1$, where a^2 and b^2 are any positive real numbers

satisfying the equation

$$a^2m_1^2 + b^2 = b_1^2 \tag{2.8}$$

(ii) If $b_1^2 \neq b_2^2$ *and/or* $b_1 = 0 = b_2$, *then* T_1 *and* T_2 *are not tangent to any ellipse with center* (h,k).

PROOF. Again, we prove the case $h = k = 0$, the general case following from a translation. Note first that the equation $m_1^2u + v = b_1^2$ always has infinitely many positive solutions (u, v) since the line $m_1^2u + v = b_1^2$ has a positive v intercept and a nonpositive slope. Now let a^2 and b^2 be any positive real numbers satisfying (2.8). To prove (i), let $x_j = -\frac{a^2b_jm_j}{a^2m_j^2+b^2}$ and $y_j = \frac{b^2b_j}{a^2m_j^2+b^2}$. Note that $b_1 \neq 0 \neq b_2$. Consider first the case $m_1 = m_2 \equiv m$. Then $b_2 = -b_1$ since $T_1 \neq T_2$. Also, $x_j = -\frac{a^2b_jm}{b_1^2}$ and $y_j = \frac{b^2b_j}{b_1^2}$, which implies that $dy/dx\ |_{x=x_j, y=y_j} = \frac{-b^2x_j}{a^2y_j} = \frac{-b^2}{a^2}\frac{-a^2b_jm}{b^2b_j} = m$. Finally, $mx_1 + b_1 = \frac{-a^2m^2b_1+b_1^3}{b_1^2} = \frac{b^2}{b_1} = y_1$ and $mx_2 + b_2 = \frac{-a^2m^2b_2+b_1^2b_2}{b_1^2} = \frac{b^2b_2}{b_1^2} = y_2$. Hence T_1 and T_2 are tangents to E(at (x_j, y_j)).

Now suppose that $m_1 = -m_2$. Then $b_2 = b_1$ since $T_1 \neq T_2$, and $x_j = -\frac{a^2b_jm_j}{b_1^2}$ and $y_j = \frac{b^2b_j}{b_1^2}$, which implies that $dy/dx\ |_{x=x_j, y=y_j} = \frac{-b^2x_j}{a^2y_j} = \frac{-b^2}{a^2}\frac{-a^2b_jm_j}{b^2b_j} = m_j$. Also, $mx_1 + b_1 = \frac{-a^2m_1^2b_1+b_1^3}{b_1^2} = \frac{b^2}{b_1} = y_1$ and $mx_2 + b_2 = \frac{-a^2m_2^2b_2+b_1^2b_2}{b_1^2} = \frac{b^2b_2}{b_1^2} = y_2$. Hence T_1 and T_2 are tangents to E(at (x_j, y_j)).

To prove (ii), suppose that T_1 and T_2 are tangent to an ellipse E with center (h,k). Proving the case $h = k = 0$ and proceeding as in the proof of Theorem 2.2, $b_j^2 = a^2m_j^2 + b^2$, $j = 1, 2$. Since $m_1^2 = m_2^2$, $b_1^2 = b_2^2$. Also, by Lemma2.1, $b_1^2 \neq 0 \neq b_2^2$. □

REMARK 2.7. A result similar toTheorem 2.6 can be proved if both T_1 and T_2 are *vertical*, say with equations $x - h = b_j$, $j = 1, 2$. It is easy to show that $b_2 = -b_1$(so that $b_1^2 = b_2^2$) and $b_1 \neq 0 \neq b_2$. Also, $a^2 = 2b_1$, while b can be any positive real number.

REMARK 2.8. One can obtain the expression for b^2 in (2.7) from the expression for a^2 in a natural way as follows: Write the T_j in the form $x - h = \frac{1}{m_j}(y - k) - \frac{b_j}{m_j}$(assuming $m_1 \neq 0 \neq m_2$). This suggests replacing m_j by $\frac{1}{m_j}$ and b_j by $-\frac{b_j}{m_j}$. Then $\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ becomes $\frac{b_1^2m_2^2-b_2^2m_1^2}{m_2^2-m_1^2}$.

2.0.2. *Examples.* (1) Let $h = k = 0$, $T_1 :\ y = 2x + 3$, $T_2 :\ y = -2x + 3$. Then by Theorem 2.6, T_1 and T_2 are tangent to any ellipse $E \in F = \{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 : 4a^2 + b^2 = 9\}$. See Figure 2.2 for a plot of T_1, T_2, and the two ellipses in F, $x^2 + \frac{y^2}{5} = 1$ and $y^2 + \frac{x^2}{2} = 1$

(2) Let $h = k = 0$, $T_1 :\ y = x + 2$, $T_2 :\ y = -x + 1$. Then by Theorem 2.6, T_1 and T_2 are not tangent to any ellipse with center $(0, 0)$.

2.1. Rotated Ellipses. For simplicity of exposition, we have only considered non-rotated ellipses. However, the previous results extend with little effort to ellipses whose axes are *rotated* thru a specified angle α. Rotate the x and y axes clockwise about (h,k) thru an angle α, $-\dfrac{\pi}{2} < \alpha < \dfrac{\pi}{2}$, and let u and v be the new coordinate axes. Then the line $y-k = m(x-h)+b$ is transformed into the line

$$v-k = \left(\frac{m+\tan\alpha}{1-m\tan\alpha}\right)(u-h) + \frac{b\sec\alpha}{1-m\tan\alpha},\quad 1-m\tan\alpha \neq 0$$

Now let T_1 and T_2 be distinct, non–vertical lines with equations $y-k = m_j(x-h)+b_j$, $j=1,2$, $m_1^2 \neq m_2^2$.

We want to apply conditions (A) and (B) of Theorem 2.4 in the *new* coordinates u and v. Thus we replace m_j by $m_{j,\alpha}$ and b_j by $b_{j,\alpha}$, where

$$m_{j,\alpha} = \frac{m_j+\tan\alpha}{1-m_j\tan\alpha}$$

and

$$b_{j,\alpha} = \frac{b_j\sec\alpha}{1-m_j\tan\alpha}$$

This yields the expressions

$$A = \frac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$$

$$B = \frac{b_{1,\alpha}^2 m_{2,\alpha}^2 - b_{2,\alpha}^2 m_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$$

We now state the following generalization of Theorems 2.4 without proof.

THEOREM 2.9. *Let h and k be given real numbers, and let T_1 and T_2 be distinct lines with equations $y-k = m_j(x-h)+b_j$, $j=1,2$. Suppose that $-\dfrac{\pi}{2} < \alpha < \dfrac{\pi}{2}$ and $1-m_1\tan\alpha \neq 0 \neq 1-m_2\tan\alpha$(thus T_1 and T_2 are not vertical when rotated by α)*

Part 1: If $A>0$ and $B>0$, then there is a unique ellipse E, with center (h,k) and rotation angle α, which has T_1 and T_2 as tangents. Furthermore, in the new rotated coordinates u and v, E has equation $\dfrac{(u-h)^2}{a^2} + \dfrac{(v-k)^2}{b^2} = 1$, where

$$a^2 = \frac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2},\ b^2 = \frac{b_{1,\alpha}^2 m_{2,\alpha}^2 - b_{2,\alpha}^2 m_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$$

Part 2: If $m_{2,\alpha}^2 \neq m_{1,\alpha}^2$, then $A>0$ and $B>0$ are also necessary conditions for T_1 and T_2 to be tangent to some ellipse with center (h,k) and rotation angle α.

It is more interesting to allow the rotation angle to *vary.* This leads to the following

Question: Given two distinct lines T_1 and T_2, is any point $(h,k)\notin T_1\cup T_2$ α admissible for some α, $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$? The following theorem answers this question in the affirmative.

THEOREM 2.10. *Let h and k be given real numbers, and let T_1 and T_2 be distinct, non–vertical lines with equations $y-k = m_j(x-h)+b_j$, $j=1,2$. Assume that $m_1\neq m_2$, and that $(h,k)\notin T_1\cup T_2$. Then there is an angle α, $-\dfrac{\pi}{2} < \alpha < \dfrac{\pi}{2}$, such that the ellipse E with certain (h,k) and rotation angle α has T_1 and T_2 as tangents.*

Proof. Note that $(h,k) \notin T_1 \cup T_2 \Rightarrow b_1 \neq 0 \neq b_2$

Case 1: $m_1^2 \neq m_2^2$

We want to choose α so that $A > 0$ and $B > 0$. Note that $\frac{m+\tan\alpha}{1-m\tan\alpha}$ is a strictly increasing function of m. Since $m_1^2 \neq m_2^2$, we can assume without loss of generality that $m_2 \neq 0$, and that $m_2^2 > m_1^2$. Let $\alpha_0 = \arctan(1/m_2)$. Then $\lim_{\alpha\to\alpha_0}\left(\frac{m_2+\tan\alpha}{1-m_2\tan\alpha}\right)^2 = \infty$ and $\lim_{\alpha\to\alpha_0}\left(\frac{b_2\sec\alpha}{1-m_2\tan\alpha}\right)^2 = \infty$ since $b_2 \neq 0$. (Note that $\tan(\alpha_0) = 1/m_2 \Rightarrow m_2 + \tan\alpha_0 \neq 0$). Hence $\lim_{\alpha\to\alpha_0} A = \infty$. Now

$$B = \left(\left(\frac{m_2+\tan\alpha}{1-m_2\tan\alpha}\right)^2 - \left(\frac{m_1+\tan\alpha}{1-m_1\tan\alpha}\right)^2\right)^2 \left(\frac{b_1\sec\alpha}{1-m_1\tan\alpha}\right)^2 - \left(\frac{m_1+\tan\alpha}{1-m_1\tan\alpha}\right)^2$$

$$\text{A}=\left(\left(\frac{m_2+\tan\alpha}{1-m_2\tan\alpha}\right)^2 - \left(\frac{m_1+\tan\alpha}{1-m_1\tan\alpha}\right)^2\right) C,$$

where $C = \left((\frac{m_2+\tan\alpha}{1-m_2\tan\alpha})^2 - (\frac{m_1+\tan\alpha}{1-m_1\tan\alpha})^2\right)(\frac{b_1\sec\alpha}{1-m_1\tan\alpha})^2 - (\frac{m_1+\tan\alpha}{1-m_1\tan\alpha})^2$

$$\left(\left(\frac{b_2\sec\alpha}{1-m_2\tan\alpha}\right)^2 - \left(\frac{b_1\sec\alpha}{1-m_1\tan\alpha}\right)^2\right).$$

Clearly $\frac{m_{2,\alpha}}{b_{2,\alpha}} \to \infty$ as $\alpha \to \alpha_0$ since $-\frac{\pi}{2} < \alpha_0 < \frac{\pi}{2}$ implies that $\sec\alpha_0 \neq 0$. Hence $C \to \infty \Rightarrow B \to \infty$, as $\alpha \to \alpha_0$. Thus for α sufficiently close to α_0, $A > 0$ and $B > 0$.

Case 2: $0 \neq m_2 = -m_1$. Since $b_1 \neq 0 \neq b_2$, by Theorem 2.6, we need to choose α so that $b_{1,\alpha}^2 = b_{2,\alpha}^2$, which holds if and only if $(\frac{b_1\sec\alpha}{1-m_1\tan\alpha})^2 = (\frac{b_1\sec\alpha}{1+m_1\tan\alpha})^2 \Leftrightarrow \frac{b_1^2}{b_2^2} = \frac{1-m_1\tan\alpha}{1+m_1\tan\alpha} \Leftrightarrow \tan\alpha = \frac{1}{m_1}\frac{b_2^2-b_1^2}{b_2^2+b_1^2}$. Solving gives the required α. ☐

2.1.1. *Example.* $T_1 : y = x + 2,\ T_2 : y = 2x + 1$ Then $(0,0)$ is not admissible if one does not allow rotation. However, letting $\alpha = \arctan(1/m_2) = \arctan(.5) \approx .46365$ implies that $A > 0$ and $B > 0$. Hence the ellipse E with center $(0,0)$ and rotation angle α has T_1 and T_2 as tangents.

Remark 2.11. Clearly rotation is no help when $m_1 = m_2$. Given (h,k), by Theorem 2.6, we need to choose α so that $b_{1,\alpha}^2 = b_{2,\alpha}^2$. However, since $m_{1,\alpha} = m_{2,\alpha}$ for any α, $b_{1,\alpha}^2 = b_{2,\alpha}^2$ if and only if $b_1^2 = b_2^2$. In that case a proof similar to that of Theorem 2.6 shows that any angle α will do. If $b_1^2 \neq b_2^2$, no α will work.

2.2. An Aside on Hyperbolas. Using similar techniques, it is not hard to prove a version of Theorem 2.9 for hyperbolas.

Theorem 2.12. *Suppose that we are given distinct lines T_1 and T_2, which are known to be tangent to some hyperbola H with center $C = (h,k)$. Suppose that T_j has equation*

$$y - k = m_j(x-h) + b_j$$

$j = 1,2$. *Finally, assume that $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$ and $1 - m_1\tan\alpha \neq 0 \neq 1 - m_2\tan\alpha$.*

Part 1:

(a) If

$$\frac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} > 0 \text{ and } \frac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} > 0, \tag{2.9}$$

then the equation of H in the rotated coordinates (u,v) is $\dfrac{(u-h)^2}{a^2} - \dfrac{(v-k)^2}{b^2} = 1$, *where* $a^2 = \dfrac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$ *and* $b^2 = \dfrac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$. [2]

(b) If

$$\frac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} < 0 \text{ and } \frac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} < 0, \tag{2.10}$$

then the equation of H in the rotated coordinates (u,v) is $\dfrac{(v-k)^2}{a^2} - \dfrac{(u-h)^2}{b^2} = 1$, *where* $a^2 = -\dfrac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$ *and* $b^2 = -\dfrac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$.

Part 2: If $m_{2,\alpha}^2 \neq m_{1,\alpha}^2$, then it follows that a necessary condition for a non-rotated hyperbola with T_1 and T_2 as tangents, and center (h,k), to exist is that $\dfrac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$ *and* $\dfrac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2}$ *are either both positive or both negative.*

2.2.1. *Example.* The lines $T_1 : \; y = x + 2$ and $T_2 : \; y = 2x - 3$ are given, and we want a hyperbola (nonrotated) H with center $(2,0)$, such that T_1 and T_2 are tangents to H. Then $m_1 = 1$, $b_1 = 4$, $m_2 = 2$, $b_2 = 1$, which implies that $\frac{b_{2,\alpha}^2 - b_{1,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} < 0$ and $\frac{b_{2,\alpha}^2 m_{1,\alpha}^2 - b_{1,\alpha}^2 m_{2,\alpha}^2}{m_{2,\alpha}^2 - m_{1,\alpha}^2} < 0$. By Theorem 2.12, Part 2, the hyperbola $\dfrac{y^2}{21} - \dfrac{(x-2)^2}{5} = 1$ is tangent to T_1 and T_2.

Recall that for the example above there was no *ellipse*$(\alpha = 0)$ with center $(2,0)$ with T_1 and T_2 as tangents. This leads to the question of whether, for fixed center $(h,k) \notin T_1 \cup T_2$ and fixed rotation angle α, there is always either a hyperbola or ellipse which is tangent to T_1 and T_2. The following theorem answers this in the affirmative. For simplicity of exposition, we first state the result for nonrotated conics.

THEOREM 2.13. *(Nonrotated Version) Let h and k be given real numbers, and let T_1 and T_2 be distinct, non-vertical lines with equations $y-k = m_j(x-h)+b_j$, $j = 1,2$. Assume that $(h,k) \notin T_1 \cup T_2$, and that $m_1^2 \neq m_2^2$. Then there is either a non-rotated hyperbola or a non-rotated ellipse, with center (h,k), which is tangent to T_1 and T_2.*

PROOF. Assume, w.l.o.g., that $m_2^2 - m_1^2 > 0$. It is clear that *both* (A) and (B) of Theorem 2.4 and eq(2.9) cannot hold. Similarly, *both* (A) and (B) of Theorem 2.4 and eq(2.10) cannot hold. Hence there cannot be *both* an ellipse and a hyperbola, with center (h,k), which is tangent to T_1 and T_2. Now suppose that there is no ellipse, with center (h,k), which is tangent to T_1 and T_2. By Theorem 2.4, Part 2, either $b_2^2 - b_1^2 < 0$ or $b_1^2 m_2^2 - b_2^2 m_1^2 < 0$. In the first case, $b_2^2 m_1^2 - b_1^2 m_2^2 < b_2^2 m_2^2 - b_1^2 m_2^2 = m_2^2(b_2^2 - b_1^2) < 0$. By Theorem 2.12, Part 1, there is a hyperbola, with center (h,k), which is tangent to T_1 and T_2. In the second case, $b_2^2 m_1^2 - b_1^2 m_2^2 > 0$ and $m_2^2(b_2^2 - b_1^2) > b_2^2 m_1^2 - b_1^2 m_2^2 > 0 \Rightarrow b_2^2 - b_1^2 > 0$. Again, by Theorem 2.12, Part 1, there is a hyperbola, with center (h,k), which is tangent to T_1 and T_2. Finally, suppose that there is no hyperbola, with center (h,k), which is tangent to T_1 and T_2. By Theorem 2.12, Part 2, either $b_2^2 - b_1^2 > 0$ and $b_2^2 m_1^2 - b_1^2 m_2^2 < 0$, or $b_2^2 - b_1^2 < 0$

[2] Note that a^2 is the same as for the ellipse, but b^2 has the opposite sign.

and $b_2^2m_1^2 - b_1^2m_2^2 > 0$. In the first case, $b_1^2m_2^2 - b_2^2m_1^2 > 0$. By Theorem 2.4, Part 1, there is an ellipse, with center (h,k), which is tangent to T_1 and T_2. Since the second case cannot occur, that completes the proof of the theorem. □

The proof of the rotated version of Theorem 2.13 follows in a similar fashion by replacing m_j by $m_{j,\alpha}$ and b_j by $b_{j,\alpha}$. We leave the details to the reader.

THEOREM 2.14. *(Rotated Version) Let h and k be given real numbers, and let T_1 and T_2 be distinct lines with equations $y-k = m_j(x-h)+b_j$, $j=1,2$. Suppose that $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$ and that $1 - m_1\tan\alpha \neq 0 \neq 1 - m_2\tan\alpha$. Finally, assume that $(h,k) \notin T_1 \cup T_2$, and that $m_{1,\alpha}^2 \neq m_{2,\alpha}^2$. Then there is either a hyperbola or an ellipse, with center (h,k) and rotation angle α, which is tangent to T_1 and T_2.*

3. Three Given Tangents

To simplify things, at first we only consider nonrotated ellipses. Later we indicate how the results can be modified for a fixed angle of rotation.

Given three lines $T_j : y = m_jx + c_j$, $j = 1,2,3$, we want to find an ellipse $E : \frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ which is tangent to each of the T_j. If we consider each pair of tangents $\{T_i, T_j\}$, we can find an ellipse $E_{i,j}$, with center $C = (h,k)$, which is tangent to T_i and T_j, using Theorem 2.4. The formula for a^2 and b^2 is given by (2.7), provided the a^2 and b^2 obtained from each pair of tangents are equal. This yields the curve γ below, which is a hyperbola, on which the admissible centers must lie. Several cases must be considered depending upon whether $m_i^2 = m_j^2$ and/or $m_i = m_j$ for some $i \neq j$. Of course it is geometrically obvious that one must also assume that the T_j do not have a common intersection point.

3.1. No two of the T_j have slopes equal in absolute value. We first state our results when none of the T_j is horizontal or vertical–that is, when none of the T_j are parallel to the major or minor axis of the ellipse. In that case we shall show that the set of admissible centers is given precisely by $\{(h,k) \in \gamma : q_3(k) > 0$ and $r_3(h) > 0\}$, where γ is a hyperbola and q_3 and r_3 are polynomials of degree three with all real zeroes. Further, the zeroes of q_3 and r_3 can be obtained from the intersection points of the T_j using γ. Similar, but less concise results, can be proved if one of the T_j is horizontal or vertical.

THEOREM 3.1. *Let T_j, $j = 1,2,3$ be three distinct non-horizontal and non–vertical lines with equations $y = m_jx + c_j$, and assume that $i \neq j \Rightarrow m_i^2 \neq m_j^2$. Assume also that the T_j do not have a common intersection point. Let (x_1,y_1) equal the intersection point of T_2 and T_3, (x_2,y_2) equal the intersection point of T_1 and T_3, and (x_3,y_3) equal the intersection point of T_1 and T_2.*[3] *Let $D = (m_2-m_1)(m_3-m_2)(m_3-m_1)$*

Let γ be the curve with equation

$$Dhk + a_1h + a_2k + a_3 = 0 \tag{3.1}$$

where

$a_1 = m_1c_1\left(m_3^2 - m_2^2\right) - m_2c_2\left(m_3^2 - m_1^2\right) + m_3c_3\left(m_2^2 - m_1^2\right)$,

[3] In this section (x_j, y_j) denotes the intersection point of two tangents, and *not* the point of tangency, as in the previous section.

$a_2 = -c_1\left(m_3^2 - m_2^2\right) + c_2\left(m_3^2 - m_1^2\right) - c_3\left(m_2^2 - m_1^2\right)$,
$a_3 = (c_1^2(m_3^2 - m_2^2) + c_2^2(m_1^2 - m_3^2) + (m_2^2 - m_1^2)c_3^2)/2$

Write equation (3.1) in the form $k = f(h)$ or $h = g(k)$. Let $v_j = f(x_j)$, $w_j = g(y_j)$, $j = 1,2,3$, and let $r_3(h) = \tau(h-w_1)(h-w_2)(h-w_3)$, $q_3(k) = \beta(k-v_1)(k-v_2)(k-v_3)$, where $\tau = -m_1m_2m_3\dfrac{x_2-x_3}{m_3-m_2}$ and $\beta = \dfrac{x_2-x_3}{m_3-m_2}$. Let S denote the set of admissible centers (h,k). Then

$$S = \{(h,k) \in \gamma : q_3(k) > 0 \text{ and } r_3(h) > 0\} \tag{3.2}$$

For each $(h,k) \in S$, if E is the ellipse $\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$ which is tangent to each of the T_j, then E is unique, and a^2 and b^2 are given by (3.8) and (3.9) below, with $b_j = m_jh + c_j - k$.

Finally, given any $0 \le e_0 < 1$, there is an $(h,k) \in S$ such that E has eccentricity e_0.

Proof. Note that $x_i = x_j$ for $i \neq j$ if and only if the T_j have a common intersection point. Since none of the T_j are horizontal, it also follows that $y_i = y_j$ for $i \neq j$ if and only if the T_j have a common intersection point. Since we assumed that the T_j do **not** have a common intersection point,

$$x_i \neq x_j \text{ and } y_i \neq y_j \text{ for } i \neq j \tag{3.3}$$

Solving (3.1) for k yields

$$k = f(h) = -\frac{a_1h + a_3}{Dh + a_2} \tag{3.4}$$

and solving (3.1) for h yields

$$h = g(k) = -\frac{a_2k + a_3}{Dk + a_1} \tag{3.5}$$

For any distinct $\{i,j,k\}$ from $\{1,2,3\}$, T_i and T_j intersect at

$$(x_l, y_l) = \left(\frac{c_j - c_i}{m_i - m_j}, \frac{m_ic_j - m_jc_i}{m_i - m_j}\right) \tag{3.6}$$

We shall use the following easily proven identity many times. For any $i,j,l \in \{1,2,3\}$

$$(m_l - m_i)(m_j - m_i)(x_j - x_l) = -c_2m_3 + c_2m_1 + c_1m_3 + c_3m_2 - c_3m_1 - c_1m_2 \tag{3.7}$$

Using (3.7), it is not hard to show that

$Dx_l + a_2 = (m_i^2 - m_j^2)(m_l - m_i)(x_j - x_l)$,
$Dy_l + a_1 = m_l(m_j^2 - m_i^2)(m_l - m_i)(x_j - x_l)$

Since $x_j \neq x_l$ for $j \neq l$ by (3.3), $Dx_l + a_2 \neq 0$ and $Dy_l + a_1 \neq 0$ for any l. Hence by (3.4) and (3.5), $v_j = f(x_j)$ and $w_j = g(y_j)$ are each *finite.*

Now given $C = (h,k)$, write T_j in the form $y - k = m_j(x-h) + b_j$, $j = 1,2,3$. By Theorem 2.4, for each pair of tangents $\{T_i, T_j\}$, we can find an ellipse $E_{i,j}$, with center C which is tangent to T_i and T_j, if and only if

$$\frac{b_j^2 - b_i^2}{m_j^2 - m_i^2} > 0 \text{ and } \frac{b_i^2m_j^2 - b_j^2m_i^2}{m_j^2 - m_i^2} > 0$$

Since we want E to be tangent to all three lines, we need the a^2 and b^2 obtained from each pair of tangents to be **equal**. First, using (2.7) for a^2, we want

$$\frac{b_2^2 - b_1^2}{m_2^2 - m_1^2} = \frac{b_3^2 - b_2^2}{m_3^2 - m_2^2} = \frac{b_3^2 - b_1^2}{m_3^2 - m_1^2} \tag{3.8}$$

Second, using (2.7) for b^2, we want

$$\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2 - m_1^2} = \frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} = \frac{b_2^2 m_3^2 - b_3^2 m_2^2}{m_3^2 - m_2^2} \tag{3.9}$$

For simplicity, let $e_j = b_j^2$ and $f_j = m_j^2$, $j = 1, 2, 3$. Cross multiplying in (3.8) and (3.9) yields a system of four polynomial equations in the variables e_j and f_j :

$s_1 \equiv (e_2 - e_1)(f_3 - f_1) - (e_3 - e_1)(f_2 - f_1) = 0$, $s_2 \equiv (e_2 - e_1)(f_3 - f_2) - (e_3 - e_2)(f_2 - f_1) = 0$, $s_3 \equiv (e_1 f_2 - e_2 f_1)(f_3 - f_1) - (e_1 f_3 - e_3 f_1)(f_2 - f_1) = 0$, $s_4 \equiv (e_1 f_2 - e_2 f_1)(f_3 - f_2) - (e_2 f_3 - e_3 f_2)(f_2 - f_1) = 0$. We found a Grobner Basis(see [**5**]) GB for $\{s_1, s_2, s_3, s_4\}$ using Maple. GB contains the one polynomial

$(e_2 - e_3) f_1 + (e_3 - e_1) f_2 + (e_1 - e_2) f_3$. Hence the solution set of the system of equations in (3.8) and (3.9) is equivalent to the one equation

$$(b_2^2 - b_3^2) m_1^2 + (b_3^2 - b_1^2) m_2^2 + (b_1^2 - b_2^2) m_3^2 = 0 \tag{3.10}$$

Expanding the LHS of (3.10) and using $b_j = m_j h + c_j - k$ yields

$(b_2^2 - b_3^2) m_1^2 + (b_3^2 - b_1^2) m_2^2 + (b_1^2 - b_2^2) m_3^2 = 2m_2^2 c_1 k - 2m_1^2 c_2 k - m_2^2 c_1^2 + m_1^2 c_2^2 - 2m_2^2 m_1 h c_1 + 2m_2^2 m_1 hk + 2m_1^2 m_2 h c_2 - 2m_1^2 m_2 hk - m_1^2 c_3^2 + m_2^2 c_3^2 + m_3^2 c_1^2 - m_3^2 c_2^2 +$
$2m_1^2 c_3 k - 2m_2^2 c_3 k - 2m_3^2 c_1 k + 2m_3^2 c_2 k - 2m_1^2 m_3 h c_3 + 2m_1^2 m_3 hk + 2m_2^2 m_3 h c_3$
$-2m_2^2 m_3 hk + 2m_3^2 m_1 h c_1 - 2m_3^2 m_1 hk - 2m_3^2 m_2 h c_2 + 2m_3^2 m_2 hk$

Dividing thru by 2 and simplifying gives eq.(3.1).

Claim: (3.4) never reduces to $k =$ constant and (3.5) never reduces to $h =$ constant.

PROOF. of claim: To prove the claim, we must show that $a_1 a_2 - a_3 D \neq 0$. Now $a_1 a_2 - a_3 D =$

$-\frac{1}{2}(m_3 + m_2)(m_2 + m_1)(m_3 + m_1)(m_1 c_2 - m_3 c_2 - m_1 c_3 + m_2 c_3 + m_3 c_1 - m_2 c_1)^2 =$

$-\frac{1}{2}(m_3 + m_2)(m_2 + m_1)(m_3 + m_1)\left(\frac{x_3 - x_2}{(m_1 - m_3)(m_1 - m_2)}\right)^2 \neq 0$ by (3.7). Since $m_i \neq -m_j$ for $i \neq j$ and $x_2 \neq x_3$, $a_1 a_2 - a_3 D \neq 0$.

Note that γ is a nontrivial curve since $D \neq 0$, and that part of γ must lie inside the triangle T formed by the T_j since the incenter of T must lie on γ.

If $b_1 = m_1 h + c_1 - k$, $b_2 = m_2 h + c_2 - k$, then by (3.4),

$\frac{b_2^2 - b_1^2}{m_2^2 - m_1^2} = \left(\frac{1}{m_2^2 - m_1^2}\right)((m_2 - m_1)h + c_2 - c_1)((m_1 + m_2)h + c_1 - 2k + c_2)$

$= \frac{p_3(h)}{p_1(h)}$, where p_3 is a monic polynomial of degree ≤ 3 and $p_1(h) = h + \frac{a_2}{D}$, and by (3.5),

$\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2 - m_1^2} =$

$-\left(\frac{1}{m_2^2 - m_1^2}\right)((m_1 - m_2)k + m_2 c_1 - m_1 c_2)((m_1 + m_2)k - 2m_1 m_2 h - m_1 c_2 - m_2 c_1)$

$= \frac{s_3(k)}{s_1(k)}$, where s_3 is a monic polynomial of degree ≤ 3 and $s_1(k) = k + \frac{a_1}{D}$

We can determine the zeroes of p_3 and s_3 as follows:

Using the fact that $\frac{c_j - c_i}{m_i - m_j} = \frac{b_j - b_i}{m_i - m_j} + h$ and $\frac{m_i c_j - m_j c_i}{m_i - m_j} = \frac{m_i b_j - m_j b_i}{m_i - m_j} + k$, we have

$$x_l = \frac{b_j - b_i}{m_i - m_j} + h, \; y_l = \frac{m_i b_j - m_j b_i}{m_i - m_j} + k$$

Hence $b_j - b_i = 0$ if and only if $h = x_l$, and $m_i b_j - m_j b_i = 0$ if and only if $k = y_l$.

This implies that the polynomials $p_3(h)$ and $s_3(k)$ above have all real roots.

p_3 vanishes at the x coordinates of the intersection points of the T_j, and s_3 vanishes at the y coordinates of the intersection points of the T_j. Hence $p_3(h) = (h-x_1)(h-x_2)(h-x_3)$ and $s_3(k) = (k-y_1)(k-y_2)(k-y_3)$.

Now we want to know when both $\frac{b_2^2-b_1^2}{m_2^2-m_1^2}$ and $\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2-m_1^2}$ are positive for $(h,k) \in \gamma$.

First, substituting $h = g(k)$, one can write $\frac{p_3(h)}{p_1(h)} = \frac{(g(k)-x_1)(g(k)-x_2)(g(k)-x_3)}{g(k)+\frac{a_2}{D}}$,

which, using the fact that $v_j = f(x_j)$, vanishes when $k = v_j$, $j = 1,2,3$. Simplifying yields

$\frac{p_3(h)}{p_1(h)} = \frac{2((m_2-m_1)a_2-(c_2-c_1)D)}{(m_2^2-m_1^2)D} \frac{(k-v_1)(k-v_2)(k-v_3)}{\left(k+\frac{a_1}{D}\right)^2} = 2\frac{x_2-x_3}{m_3-m_2} \frac{(k-v_1)(k-v_2)(k-v_3)}{\left(k+\frac{a_1}{D}\right)^2}$,

which is positive precisely when $q_3(k) > 0$. Thus $\frac{b_2^2-b_1^2}{m_2^2-m_1^2} > 0$ precisely when $q_3(k) > 0$. Similarly, substituting $k = f(h)$, one can write

$\frac{s_3(k)}{s_1(k)} = -2m_1 m_2 m_3 \frac{x_2-x_3}{m_3-m_2} \frac{(f(h)-y_1)(f(h)-y_2)(f(h)-y_3)}{\left(h+\frac{a_2}{D}\right)^2}$,

which, using the fact that $w_j = g(y_j)$, vanishes when $h = w_j$, $j = 1,2,3$. Hence $\frac{s_3(k)}{s_1(k)} = -2m_1 m_2 m_3 \frac{x_2-x_3}{m_3-m_2} \frac{(h-w_1)(h-w_2)(h-w_3)}{\left(h+\frac{a_2}{D}\right)^2}$, which is positive precisely when $r_3(h) > 0$. Thus $\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2-m_1^2} > 0$ precisely when $r_3(h) > 0$.

Since $h + \frac{a_2}{D} = 0$ or $k + \frac{a_1}{D} = 0$ cannot yield a point on γ, the set of admissible centers is given by (3.2). Note that $S \neq \emptyset$ since the incenter of T is admissible. One could also prove directly that S is nonempty. The *uniqueness* follows from Theorem 2.2.

Finally, to prove that there is $(h,k) \in S$ such that E has eccentricity e_0: Using (3.6), it follows easily that for any set of distinct $i,j,l \in \{1,2,3\}$,

$r_3(x_i) = \frac{1}{8}(m_j+m_l)^2 \frac{(m_j-m_i)^2(m_l-m_i)^2(x_j-x_l)^4}{(m_j-m_l)^4}$, which implies that

$$r_3(x_i) \neq 0 \text{ for any } i \tag{3.11}$$

and $r_3(-\frac{a_2}{D}) = \frac{1}{8}(m_1+m_2)^2 (x_3-x_2)^4 (m_1+m_3)^2 \frac{(m_3+m_2)^2}{(m_3-m_2)^4}$, which implies that

$$r_3(-\tfrac{a_2}{D}) \neq 0 \tag{3.12}$$

Also, $p_3(g(y_i)) = \frac{1}{8} \frac{(m_i+m_j)(m_i+m_l)(x_j-x_l)^3(m_i-m_j)^2(m_i-m_l)^2}{m_i^3(m_l-m_j)^3}$ and $p_1(g(y_i)) = \frac{1}{2} \frac{(m_i+m_j)(m_i+m_l)(x_j-x_l)}{m_i(m_l-m_j)}$, which implies that

$$p_3(w_i)p_1(w_i) \neq 0 \tag{3.13}$$

Note that from the proof of Theorem 3.1, since $\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2-m_1^2} = \frac{p_3(h)}{p_1(h)}$,

$$S = \{(h, f(h)) : \frac{p_3(h)}{p_1(h)} > 0 \text{ and } r_3(h) > 0\} \tag{3.14}$$

It follows from(2.7) that $\frac{a^2}{b^2} = \left(\frac{b_2^2-b_1^2}{m_2^2-m_1^2}\right) / \left(\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2-m_1^2}\right) = \frac{p_1(h)p_3(h)}{r_3(h)} \equiv R(h)$. Let T be the triangle enclosed by the T_j, and let (h_0, k_0) equal the incenter of

T. Then $R(h_0) = 1, p_1(h_0)p_3(h_0) > 0$, and $r_3(h_0) > 0$. Let $s = \inf\{h < h_0 : p_1(h)p_3(h) > 0 \text{ and } r_3(h) > 0\}$. Since $x_j < h_0$ for some j, $-\infty < s$. First, if $r_3(s) = 0$, then $s = w_j \Rightarrow p_1(s)p_3(s) \neq 0$ by (3.13), and thus $\lim_{h\to s^+} R(h) = \infty$. In that case $\frac{b^2}{a^2}$ varies from 0 to 1 on (s, h_0), and the equation $1 - \frac{b^2}{a^2} = e_0^2$ has a solution (h, k) which is an admissible center by (3.14). Second, if $p_1(s)p_3(s) = 0$, then $s = x_j$ or $s = -\frac{a_2}{D} \Rightarrow r_3(s) \neq 0$ by (3.11) or (3.12), and thus $\lim_{h\to s^+} R(h) = 0$. In that case $\frac{a^2}{b^2}$ varies from 0 to 1 on (s, h_0), and again the equation $1 - \frac{b^2}{a^2} = e_0^2$ has a solution (h, k) which is an admissible center by (3.14).

□

3.1.1. ***Example***. $T_1 : y = x + 1$, $T_2 : y = 2x - 2$, $T_3 : y = -3x$. The intersection points of the $T_j's$ are $(\frac{2}{5}, -\frac{6}{5}), (-\frac{1}{4}, \frac{3}{4}), (3, 4)$, and γ has equation $40hk + 74h - 27 - 42k = 0$. Hence $f(h) = -\frac{74h-27}{40h-42}$ and $g(k) = \frac{42k+27}{40k+74}$. It follows that $p_3(h) = (h-3)\left(h - \frac{2}{5}\right)\left(h + \frac{1}{4}\right)$, $s_3(k) = (k + \frac{6}{5})(k - 4)(k - \frac{3}{4})$, $r_3(h) = \frac{39}{10}\left(h - \frac{9}{16}\right)\left(h + \frac{9}{10}\right)\left(h - \frac{5}{6}\right)$,

$q_3(k) = \frac{13}{10}(k + \frac{5}{2})(k + \frac{7}{8})(k - \frac{1}{10})$, $p_1(h) = h - \frac{21}{20}$, and $s_1(k) = k + \frac{37}{20}$. Hence
$S = \{(h, k) \in \gamma : h > \frac{5}{6} \text{ or } -\frac{9}{10} < h < \frac{9}{16}\} \cap$
$\{(h, k) \in \gamma : k > \frac{1}{10} \text{ or } -\frac{5}{2} < k < -\frac{7}{8}\} \Rightarrow$

$$S = \{(h, -\tfrac{74h-27}{40h-42}) : -\tfrac{9}{10} < h < \tfrac{9}{16} \text{ or } h > 3\}$$

REMARK 3.2. It is interesting to note that there are points $(h, k) \in \gamma$, $(h, k) \notin T_1 \cup T_2 \cup T_3$, such that there is no non-rotated *hyperbola* or ellipse, with center (h, k), tangent to the T_j. For the example above, $h = \frac{9}{16}$, $k = \frac{3}{4}$ is such a point, since $r_3(\frac{9}{16}) = 0$, while the admissible centers must lie on γ for ellipses or hyperbolas.

3.1.2. *One of the Tangents is Horizontal or Vertical.* We now prove a result similar to Theorem 3.1 when one of the given lines, say T_3, is horizontal. Similar results can be proved if one of the lines is vertical by solving for x in terms of y in the equations of the T_j.

THEOREM 3.3. *Let T_j, $j = 1, 2, 3$ be three distinct non–vertical lines with equations $y = m_j x + c_j$, and assume that $0 \neq m_1^2 \neq m_2^2 \neq 0$ and $m_3 = 0$. Assume also that the T_j do not have a common intersection point. Let (x_1, c_3) equal the intersection point of T_2 and T_3, (x_2, c_3) equal the intersection point of T_1 and T_3, and (x_3, y_3) equal the intersection point of T_1 and T_2. Let $D = m_1 m_2 (m_2 - m_1)$ and let γ be the curve with equation $Dhk + a_1 h + a_2 k + a_3 = 0$, where $a_1 = -m_2 m_1 (-m_1 c_2 + m_2 c_1)$,*

$a_2 = (c_1 - c_3)m_2^2 - (c_2 - c_3)m_1^2$, and $a_3 = (-c_1^2 m_2^2 + c_2^2 m_1^2 + (m_2^2 - m_1^2)c_3^2)/2$. Write equation (3.1) in the form $k = f(h)$ or $h = g(k)$. Let $v_j = f(x_j)$, $j = 1, 2, 3$, and let $q_3(k) = \beta(k - v_1)(k - v_2)(k - v_3)$, where $\beta = \dfrac{x_2 - x_3}{m_3 - m_2}$. Let S denote the set of admissible centers (h, k). Then

$$S = \{(h, k) \in \gamma : q_3(k) > 0 \text{ and } h \neq g(c_3)\}$$

For each $(h, k) \in S$, if E is the ellipse $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ which is tangent to each of the T_j, then E is unique, and a^2 and b^2 are given by (3.8) and (3.9), with $b_j = m_j h + c_j - k$.

Finally, given any $0 \leq e_0 < 1$, there is an $(h, k) \in S$ such that E has eccentricity e_0.

PROOF. We sketch the proof, since most of the details follow exactly as in the exactly as in the proof of Theorem 3.1. In particular, letting $m_3 = 0$ in (3.1) gives the curve γ above. $p_3(h)$, $p_1(h)$, $q_3(k)$, and $s_1(k)$ are the same, with $\frac{b_2^2-b_1^2}{m_2^2-m_1^2} = \frac{p_3(h)}{p_1(h)} = \frac{q_3(k)}{(s_1(k))^2}$, $k = f(h)$. However $m_3 = 0$ implies that $y_1 = y_2 = c_3$, and it follows that $\frac{b_1^2 m_2^2 - b_2^2 m_1^2}{m_2^2-m_1^2} = \frac{s_3(k)}{s_1(k)}$, where $s_3 = (k-c_3)^2(k-y_3)$ and $s_1(k) = k + \frac{a_1}{D}$.

If $k = f(h)$, then

$\frac{s_3(k)}{s_1(k)} = \frac{s_3(f(h))}{s_1(f(h))} =$

$\frac{1}{4}\left(m_2c_1 - m_1c_2 - m_2c_3 + m_1c_3\right)^2 \frac{(m_2c_1+2m_2m_1h-m_2c_3+m_1c_2-m_1c_3)^2}{\left(m_2^2m_1h-m_2m_1^2h+c_1m_2^2-c_2m_1^2-c_3m_2^2+c_3m_1^2\right)^2} = \frac{r_2(h)}{\left(h+\frac{a_2}{D}\right)^2}$,

where $r_2(h) = \alpha(h-w)^2$, $\alpha = m_1^2(x_3-x_2)^2$. Since $r_2(h) > 0$ for $h \neq w$, the set of admissible centers is given by S. Finally, to prove that there is $(h,k) \in S$ such that E has eccentricity e_0: First, it is not hard to show that $w \neq x_j$ for any j, $w \neq -\frac{a_2}{D}$, $p_1(w) \neq 0$, and $p_3(w) \neq 0$. The rest of the proof follows exactly as in the proof of Theorem 3.1. □

Example. $T_1 : y = x+1$, $T_2 : y = 2x-2$, $T_3 : y = 0$. The intersection points of the T_j are

$(1,0), (-1,0), (3,4)$, and γ has equation $h(k-4)+3k = 0$. $q_3(k) = 4(k-1)(k-2)(k+2)$ and $r_2(h) = 16h^2$. It follows that $S = \{(h,k) \in \gamma : -2 < k < 1$ or $k > 2,\ k \neq 0\}$. See Figure 3.1 for a plot of γ, the tangents, and the particular ellipse $\frac{(x-3/7)^2}{30/49} + \frac{(y-1/2)^2}{1/4} = 1 (h = 3/7)$ tangent to all three given lines.

Finally, it is interesting to consider the case when T_1 is vertical and T_2 is horizontal.[4]

THEOREM 3.4. *Given the lines $T_1 : x = c_1, T_2 : y = c_2$, and $T_3 : y = m_3x + c_3$, with $m_3 \neq 0$, let S be the set of admissible centers. Then*

$$S = \{(h,k) : 2m_3hk - 2m_3(m_3c_1+c_3)h + 2(c_3-c_2)k = c_3^2 - c_2^2 - m_3^2c_1^2, h \neq c_1, k \neq c_2\}$$

Also, if E is the ellipse $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ which is tangent to each of the T_j, then $a = c_1 - h$ and $b = c_2 - k$. Finally, given any $0 \le e_0 < 1$, there is an $(h,k) \in S$ such that E has eccentricity e_0.

PROOF. Given (h,k), write $T_1 : x-h = b_1, T_2 : y-k = b_2$, and $T_3 : y-k = m_3(x-h)+b_3$. Clearly, $a = b_1$ and $b = b_2$ is both necessary and sufficient for E to be tangent to T_1 and T_2. Applying Theorem 2.4(with $m_2 = 0$) to T_2 and to T_3, respectively, implies the equations $\frac{b_3^2-b_2^2}{m_3^2} = b_1^2$ and $\frac{b_2^2m_3^2}{m_3^2} = b_2^2$, and hence we have the one condition $b_3^2 - b_2^2 = b_1^2m_3^2$. Using $b_1 = c_1 - h, b_2 = c_2 - k$, and $b_3 = m_3h + c_3 - k$ yields the above equation defining S. Note that $h \neq c_1$ and $k \neq c_2$ implies that $b_1 \neq 0 \neq b_2$

We omit the proof that the eccentricity can be specified in advance. □

3.2. Two of the T_j have slopes equal in absolute value.

THEOREM 3.5. *Let T_j, $j = 1,2,3$ be three distinct non–vertical lines with equations $y = m_jx + c_j$, and assume that $m_1^2 = m_2^2 \neq m_3^2$. Assume also that the*

[4] We do not discuss the case when *two* of the given lines are horizontal or vertical, which is trivial.

T_j do not have a common intersection point. Let S denote the set of admissible centers (h,k).

Case 1: $m_2 = -m_1 \neq 0$

Let $h_0 = \dfrac{c_2 - c_1}{2m_1}$, $k_0 = \dfrac{c_1 + c_2}{2}$. Then there exist intervals I_1 and I_2 of the form $(-\infty, s)$ and/or (t, ∞) such that $S = \{(h_0, k) : k \in I_1\} \cup \{(h, k_0) : h \in I_2\}$

Case 2: $m_2 = m_1$

Let (x_1, y_1) and (x_2, y_2) be the intersection points of $\{T_1, T_3\}$ and $\{T_2, T_3\}$, respectively. [5]*Let L be the line with equation $y = m_1 x + \frac{1}{2}c_1 + \frac{1}{2}c_2$.*

(1) If $\dfrac{m_3 - m_1}{m_3 + m_1} > 0$, then assume, without loss of generality, that $x_1 < x_2$.[6]*. If $m_1 \neq 0$, let $u_j = \dfrac{y_j - \frac{1}{2}(c_1 + c_2)}{m_1}$, $j = 1, 2$.*

If $m_1 m_3 > 0$, then $S = \{(h,k) \in L : u_1 < h < x_1$ or $x_2 < h < u_2\}$.

If $m_1 m_3 < 0$, then $S = \{(h,k) \in L : u_2 < h < x_1$ or $x_2 < h < u_1\}$.

If $m_1 = 0$, then $S = \{(h, \frac{1}{2}c_1 + \frac{1}{2}c_2) : h < x_1\} \cup \{(h, \frac{1}{2}c_1 + \frac{1}{2}c_2) : h > x_2\}$.

(2) If $\dfrac{m_3 - m_1}{m_3 + m_1} < 0$, then assume, without loss of generality, that $y_1 < y_2$.[7]*. Let $v_j = m_1 x_j + \frac{1}{2}c_1 + \frac{1}{2}c_2$, $j = 1, 2$.*

If $m_1 m_3 > 0$, then $S = \{(h,k) \in L : v_1 < k < y_1$ or $y_2 < h < v_2\}$.

If $m_1 m_3 < 0$, then $S = \{(h,k) \in L : v_2 < k < y_1$ or $y_2 < h < v_1\}$.

For both case 1 and case 2, for each $(h,k) \in S$, if E is the ellipse $\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$ which is tangent to each of the T_j, then E is unique, and a^2 and b^2 are given by (3.16) and (3.17) below, with $b_j = m_j h + c_j - k$.

Finally, given any $0 \le e_0 < 1$, there is an $(h,k) \in S$ such that E has eccentricity e_0.

PROOF. Given $C = (h,k)$, write T_j in the form $y - k = m_j(x - h) + b_j$, $j = 1, 2, 3$. By Theorem 2.6, in order for an ellipse $E : \frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ to exist which is tangent to T_1 and T_2, we must have $b_1^2 = b_2^2$, and in that case

$$a^2 m_j^2 + b^2 = b_j^2, \; j = 1, 2 \tag{3.15}$$

Again, by Theorem 2.2, since we need the a^2 and b^2 obtained from the pairs of tangents $\{T_1, T_3\}$ and $\{T_2, T_3\}$ to be equal, we want

$$a^2 = \frac{b_3^2 - b_2^2}{m_3^2 - m_2^2} = \frac{b_3^2 - b_1^2}{m_3^2 - m_1^2} \tag{3.16}$$

and

$$b^2 = \frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} = \frac{b_2^2 m_3^2 - b_3^2 m_2^2}{m_3^2 - m_2^2} \tag{3.17}$$

Using (3.16) and (3.17), and substituting into (3.15), with $j = 2$, yields

$$\left(\frac{b_3^2 - b_2^2}{m_3^2 - m_2^2}\right) m_2^2 + \frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} = b_2^2 \tag{3.18}$$

---

[5]This notation differs slightly from that used in Theorem 3.1.

[6]If $x_1 = x_2$, then the T_j have a common intersection point.

[7]If $y_1 = y_2$, then the T_j have a common intersection point.

It is easy to show that (3.16), (3.17), and (3.18), along with $m_1^2 = m_2^2$, is equivalent to the one condition $b_1^2 = b_2^2$.

Note: The condition $(b_2^2 - b_3^2)m_1^2 + (b_3^2 - b_1^2)m_2^2 + (b_1^2 - b_2^2)m_3^2 = 0$ from Section 3.1 still holds, though for Section 3.2, of course, it takes the simpler form $b_1^2 = b_2^2$.

Also, using (3.16) to define a^2 and (3.17) to define b^2 implies that (3.15) holds.

The condition $b_1^2 = b_2^2$ implies $(m_1h + c_1 - k)^2 - (m_2h + c_2 - k)^2$, which implies(using $m_1^2 = m_2^2$) that

$$(2m_2 - 2m_1)hk + (2m_1c_1 - 2m_2c_2)h + (2c_2 - 2c_1)k + c_1^2 - c_2^2 = 0 \tag{3.19}$$

Recall that (x_1, y_1) and (x_2, y_2) are the intersection points of $\{T_1, T_3\}$ and $\{T_2, T_3\}$, respectively. By (3.6),

$$x_1 = \frac{c_3 - c_1}{m_1 - m_3},\ x_2 = \frac{c_3 - c_2}{m_2 - m_3},\ y_1 = \frac{m_1c_3 - m_3c_1}{m_1 - m_3},\ y_2 = \frac{m_2c_3 - m_3c_2}{m_2 - m_3}$$

3.3. *Proof of Case 1:* $m_2 = -m_1 \neq 0$. We shall show that the set of admissible centers is the union of an open, semi–infinite vertical line with an open, semi–infinite horizontal line.

(3.19) reduces to

$-4m_1hk + 2m_1(c_1+c_2)h + 2(c_2-c_1)k + c_1^2 - c_2^2$[8] $= -(2k - c_1 - c_2)(2m_1h - c_2 + c_1) = 0 \Rightarrow k = \frac{c_1+c_2}{2}$ or $h = \frac{c_2-c_1}{2m_1}$. Note that $(\frac{c_2-c_1}{2m_1}, \frac{c_1+c_2}{2})$ is the intersection point of T_1 and T_2 since $m_2 = -m_1$. Hence the admissible centers must lie on $\gamma = \{(h,k) : h = \frac{c_2-c_1}{2m_1}$ or $k = \frac{c_1+c_2}{2}\}$. We must determine the points on γ where $\frac{b_3^2-b_1^2}{m_3^2-m_1^2}$ and $\frac{b_1^2m_3^2-b_3^2m_1^2}{m_3^2-m_1^2}$ are positive.

Recall that $b_1 = m_1h + c_1 - k$, $b_3 = m_3h + c_3 - k$. Then

$\frac{b_3^2-b_1^2}{m_3^2-m_1^2} =$

$((m_3 - m_1)h + c_3 - c_1)((m_1 + m_3)h - 2k + c_3 + c_1)/(m_3^2 - m_1^2) =$

$$(h - x_1)(h - \frac{2k - c_3 - c_1}{m_3 + m_1}) \tag{3.20}$$

and

$\frac{b_1^2m_3^2-b_3^2m_1^2}{m_3^2-m_1^2} = ((m_1 - m_3)k - m_1c_3 + m_3c_1)((m_1 + m_3)k - 2m_3m_1h - m_1c_3 - m_3c_1)/(m_1^2 - m_3^2) =$

$$(k - y_1)(k - \frac{2m_3m_1h + m_1c_3 + m_3c_1}{m_3 + m_1}) \tag{3.21}$$

Recall that $h_0 = \frac{c_2-c_1}{2m_1}$, $k_0 = \frac{c_1+c_2}{2}$

(i) Suppose that $h = h_0$: Then, by (3.20), $\frac{b_3^2-b_1^2}{m_3^2-m_1^2}$ is a linear function of k,

$$L(k) = (h_0 - x_1)(h_0 - \frac{2k - c_3 - c_1}{m_3 + m_1})$$

and it is not hard to show, using (3.20), that $\frac{b_1^2m_3^2-b_3^2m_1^2}{m_3^2-m_1^2}$ is given by

$$P(k) = (k - y_1)(k - y_2)$$

[8] One can also obtain this equation by letting $m_2 = -m_1$ in (3.1).

Now T_1, T_2, T_3 have a common intersection point if and only if $x_1 = h_0$. Also, $L(k)$ is identically zero if and only if $(m_3 - m_1)h + c_3 - c_1 = 0$, which holds if and only if $h = x_1$. Since we assumed that T_1, T_2, T_3 do **not** have a common intersection point, L is not identically the zero function and thus has degree 1.

Now L has one real zero, while P is positive on $(-\infty, y_1) \cup (y_2, \infty)$. Hence $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2}$ and $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2}$ are both positive on an open interval(infinite), I, of k values.

(ii) Suppose that $k = k_0$: Then it is not hard to show, using (3.21), that $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2}$ is given by

$$Q(h) = (h - x_1)(h - x_2)$$

Also, by (3.21), $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2}$ is a linear function of h,

$$R(h) = (k_0 - y_1)(k_0 - \frac{2m_3 m_1 h + m_1 c_3 + m_3 c_1}{m_3 + m_1})$$

R has one real zero(if $m_3 \neq 0$), while Q is positive on $(-\infty, x_1) \cup (x_2, \infty)$. Again, arguing as above, since T_1, T_2, T_3 do **not** have a common intersection point, R is not identically the zero function and thus has degree 1. Hence $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2}$ and $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2}$ are both positive on an open interval(infinite), J, of h values.

If $m_3 = 0$, then $R(h)$ equals the constant function $(c_3 - k_0)^2$, which equals 0 if and only if $c_3 = k_0$. If $c_3 = k_0 = \frac{c_1 + c_2}{2}$, then it is not hard to show that T_1, T_2, T_3 all intersect at the point $(\frac{c_2 - c_1}{2m_1}, \frac{c_1 + c_2}{2})$, which violates the assumption that the T_j do not have a common intersection point. Hence $R(h) > 0$ for all h, and it follows that $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2}$ and $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2}$ are both positive on the union of two open intervals(infinite) of h values. Hence the set of admissible centers has the form $\{(h, k_0) : h \in (-\infty, x_1) \cup (x_2, \infty)\}$.

We also need to choose (h, k) so that $b_1 \neq 0$. Towards that end, it is not hard to show that

$$(h_0 - x_1)(h_0 - \tfrac{2k_0 - c_3 - c_1}{m_3 + m_1}) = \tfrac{(-c_2 m_1 + c_2 m_3 - c_1 m_1 - c_1 m_3 + 2m_1 c_3)^2}{4m_1^2 (m_3 + m_1)(m_3 - m_1)} \tag{3.22}$$

and

$$(k_0 - y_1)\left(k_0 - \tfrac{2m_3 m_1 h_0 + m_1 c_3 + m_3 c_1}{m_3 + m_1}\right) = -\tfrac{(-c_2 m_1 + c_2 m_3 - c_1 m_1 - c_1 m_3 + 2m_1 c_3)^2}{4(m_3 + m_1)(m_3 - m_1)} \tag{3.23}$$

A simple argument shows that if $h_0 - \frac{2k_0 - c_3 - c_1}{m_3 + m_1} = 0$ or $k_0 - \frac{2m_3 m_1 h_0 + m_1 c_3 + m_3 c_1}{m_3 + m_1} = 0$, then T_1, T_2, T_3 have a common intersection point. Hence, using (3.22) and (3.30) proves that $Q(h_0)R(h_0) < 0$ and $L(k_0)P(k_0) < 0$.

But if $h = h_0$, then $b_1 = m_1 h_0 + c_1 - k$ equals 0 only when $k = k_0$, and

if $k = k_0$, then $b_1 = m_1 h + c_1 - k_0$ equals 0 only when $h = h_0$. Hence $\{(h_0, k) : k \in I\}$ and $\{(h, k_0) : h \in J\}$ cannot contain (h_0, k_0).

Summarizing, we have $S = S_1 \cup S_2$, where $S_1 = \{(h_0, k) : k \in I\}$ and $S_2 = \{(h, k_0) : h \in J\}$ if $m_3 \neq 0$, and $S_2 = \{(h, k_0) : h \in (-\infty, x_1) \cup (x_2, \infty)\}$if $m_3 = 0$. The uniqueness in each case follows from Theorem 2.2.

Finally, to prove that there is $(h, k) \in S$ such that E has eccentricity e_0:

(i) Suppose that $h = h_0$: Then, by (3.20), $\frac{b^2}{a^2} = \frac{P(k)}{L(k)}$, where

$L(k) = (h_0 - x_1)(h_0 - \frac{2k-c_3-c_1}{m_3+m_1})$ and $P(k) = (k-y_1)(k-y_2)$. The proof follows exactly as in the proof of Theorem 3.1, if we can show that $L(y_j) \neq 0,\ j = 1, 2$, and $P(k_1) \neq 0$, where $k_1 = \frac{1}{4}\frac{m_3c_2+m_1c_2-m_3c_1+m_1c_1+2m_1c_3}{m_1}$ is the zero of L.

Now $L(y_1) = -\frac{1}{4}\left(-m_3c_2 + m_1c_2 + m_3c_1 + m_1c_1 - 2m_1c_3\right)^2 \frac{m_1-m_3}{m_1^2(m_3+m_1)^3} = \frac{m_3-m_1}{m_1^2(m_3+m_1)}(y_1 - y_3)^2$,

$L(y_2) = -\frac{1}{4}\frac{(-m_3c_2+m_1c_2+m_3c_1+m_1c_1-2m_1c_3)^2}{(m_1-m_3)(m_3+m_1)m_1^2} = \frac{m_3-m_1}{m_1^2(m_3+m_1)}(y_2 - y_3)^2$.

Since $y_1 \neq y_2$ and $y_1 \neq y_3$, $L(y_j) \neq 0,\ j = 1, 2$. Finally,

$P(k_1) = \frac{1}{16}\frac{(-m_3c_2+m_1c_2+m_3c_1+m_1c_1-2m_1c_3)^2}{m_1^2} = \frac{(m_1+m_3)^2}{4m_1^2}(y_1 - y_3)^2 \neq 0$

(ii) Suppose that $k = k_0$: Then, using (3.21), $\frac{b^2}{a^2} = \frac{R(h)}{Q(h)}$, where $Q(h) = (h - x_1)(h-x_2)$ and $R(h) = (k_0 - y_1)(k_0 - \frac{2m_3m_1h+m_1c_3+m_3c_1}{m_3+m_1})$. Again, the proof follows exactly as in the proof of Theorem 3.1, if we can show that $R(x_j) \neq 0,\ j = 1, 2$, and $Q(h_1) \neq 0$, where $h_1 = \frac{1}{4}\frac{-m_3c_1+m_1c_1+m_3c_2+m_1c_2-2m_1c_3}{m_3m_1}$ is the zero of R. Now

$R(x_1) = \frac{1}{4}\left(-m_3c_2 + m_1c_2 + m_3c_1 + m_1c_1 - 2m_1c_3\right)^2 \frac{m_1-m_3}{(m_3+m_1)^3} = \frac{m_1-m_3}{m_3+m_1}(y_1 - y_3)^2 \neq 0$ and

$R(x_2) = \frac{1}{4}\frac{(-m_3c_2+m_1c_2+m_3c_1+m_1c_1-2m_1c_3)^2}{(m_1-m_3)(m_3+m_1)} = \frac{m_1-m_3}{m_3+m_1}(y_2 - y_3)^2 \neq 0$.

Finally, $Q(h_1) = \frac{1}{16}\frac{(-m_3c_2+m_1c_2+m_3c_1+m_1c_1-2m_1c_3)^2}{m_1^2m_3^2} = \frac{(m_1+m_3)^2}{4m_1^2m_3^2}(y_1 - y_3)^2 \neq 0$. $\square$

3.4. Example. $T_1 :\ y = 3x + 2$, $T_2 :\ y = -3x + 1$, $T_3 : y = 2x$. The intersection points of the T_j are $(\frac{1}{5}, \frac{2}{5}), (-2, -4), (-\frac{1}{6}, \frac{3}{2})$. Letting $h = -\frac{1}{6}$, we get $L(k) = \frac{77}{180} - \frac{11}{15}k$ and $P(k) = \frac{1}{5}(k+4)(5k-2)$. Letting $k = \frac{3}{2}$, we get $Q(h) = \frac{1}{5}(h+2)(5h-1)$ and $R(h) = -\frac{66}{5}h + \frac{77}{20}$. It follows that $S = \{(-\frac{1}{6}, k) : k < -4\} \cup \{(h, \frac{3}{2}) : h < -2\}$

3.5. Proof of Case 2: $m_2 = m_1$**.** Note that then $c_1 \neq c_2$ since T_1 and T_2 are distinct. Now (3.19) reduces to $2hm_1(c_1 - c_2) + 2k(c_2 - c_1) + c_1^2 - c_2^2 = 0$[9] $\Rightarrow$

$$2hm_1 - 2k + c_1 + c_2 = 0 \tag{3.24}$$

Recall that L is the line with equation $y = m_1x + \frac{1}{2}c_1 + \frac{1}{2}c_2$. Note that by (3.24), an admissible center must lie on L, and L is parallel to, and lies exactly halfway between, T_1 and T_2.

Solving (3.24) for k yields

(i) $k = m_1h + \frac{1}{2}c_1 + \frac{1}{2}c_2$

That gives

$$\frac{b_3^2-b_1^2}{m_3^2-m_1^2} = \frac{((m_1-m_3)h-c_3+c_2)((m_1-m_3)h-c_3+c_1)}{m_3^2-m_1^2} = \left(\frac{m_3-m_1}{m_3+m_1}\right)(h-x_1)(h-x_2) \equiv Q(h) \tag{3.25}$$

Solving (3.24) for h yields(if $m_1 \neq 0$)

(ii) $h = -\frac{1}{2}\frac{-2k+c_1+c_2}{m_1}$

That gives

[9] One can also obtain this equation by letting $m_2 = m_1$ in (3.1).

(3.26)

$$\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} = \frac{((m_3-m_1)k + c_3 m_1 - m_3 c_1)((m_1-m_3)k + m_3 c_2 - c_3 m_1)}{m_3^2 - m_1^2} = -\left(\frac{m_3-m_1}{m_3+m_1}\right)(k-y_1)(k-y_2) \equiv R(k)$$

Now $x_2 - x_1 = \frac{c_3-c_2}{m_1-m_3} - \frac{c_3-c_1}{m_1-m_3} = \frac{c_2-c_1}{m_3-m_1} \neq 0$. Also, since T_3 passes through (x_1, y_1) and (x_2, y_2), $m_3 = \frac{y_2-y_1}{x_2-x_1}$. There are several cases to consider. We prove the following case,

$$m_3 > 0 \text{ and } \frac{m_3 - m_1}{m_3 + m_1} > 0$$

, the other cases following in a similar fashion. Then $y_2 - y_1 > 0$ and $c_1 < c_2$. The latter implies that $T_1(x) < L(x) < T_2(x)\ \forall x$, and in particular

$$y_1 = T_1(x_1) < L(x_1) \text{ and } L(x_2) < T_2(x_2) = y_2$$

Clearly $Q(h) > 0$ on $(-\infty, x_1) \cup (x_2, \infty)$ and $R(k) > 0$ on (y_1, y_2). Recall that if $m_1 \neq 0$, then $u_j = \frac{y_j - \frac{1}{2}(c_1+c_2)}{m_1}$, $j = 1, 2$, which implies that $L(u_j) = y_j$. We consider three subcases,

(a) $m_1 > 0$: Then $L(u_1) = y_1 < L(x_1) \Rightarrow u_1 < x_1$, and $L(u_2) = y_2 > L(x_2) \Rightarrow u_2 > x_2$

Now $y_1 < k < y_2 \Leftrightarrow y_1 < L(h) < y_2 \Leftrightarrow L(u_1) < L(h) < L(u_2) \Leftrightarrow u_1 < h < u_2$. Hence by (3.25) and (3.26), $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2} > 0$ and $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} > 0$ on $(u_1, x_1) \cup (x_2, u_2)$.

(b) $m_1 < 0$: Then $L(u_1) = y_1 < L(x_1) \Rightarrow u_1 > x_1$, and $L(u_2) = y_2 > L(x_2) \Rightarrow u_2 < x_2$.

Now $y_1 < k < y_2 \Leftrightarrow y_1 < L(h) < y_2 \Leftrightarrow L(u_1) < L(h) < L(u_2) \Leftrightarrow u_2 < h < u_1$. Hence by (3.25) and (3.26), $\frac{b_3^2 - b_1^2}{m_3^2 - m_1^2} > 0$ and $\frac{b_1^2 m_3^2 - b_3^2 m_1^2}{m_3^2 - m_1^2} > 0$ on $(u_2, x_1) \cup (x_2, u_1)$.

(c) $m_1 = 0$: (h, k) now lies on the horizontal line with equation $y = \frac{1}{2}c_1 + \frac{1}{2}c_2$. Since $y_1 = c_1$ and $y_2 = c_2$, $k = \frac{1}{2}c_1 + \frac{1}{2}c_2 \Rightarrow y_1 < k < y_2$. The set of admissible centers now equals

$\{(h, \frac{1}{2}c_1 + \frac{1}{2}c_2) : h < x_1\} \cup \{(h, \frac{1}{2}c_1 + \frac{1}{2}c_2) : h > x_2\}$.

The *uniqueness* follows for each case from Theorem 2.2.

The proof that there is $(h, k) \in S$ such that E has eccentricity e_0 differs somewhat from the proof of Theorem 3.1, since the T_j do not enclose a triangle, and thus there is no incenter. However, if we let

$h_0 = \left(\pm(1/2)\sqrt{(1+m_3^2)/(1+m_1^2)}(c_2 - c_1) - (c_1+c_2)/2 + c_3\right)/(m_1 - m_3)$

and $k_0 = m_1 h_0 + (c_1 + c_2)/2$, then there is a circle with center (h_0, k_0) which is tangent to the T_j. Assume that $m_1 \neq 0$, the case $m_1 = 0$ following in a similar fashion. Then the admissible centers lie on the curve γ with equation $k = f(h)$ or $h = g(k)$, where $f(h) = m_1 h + \frac{1}{2}c_1 + \frac{1}{2}c_2$ and $g(k) = \frac{k - \frac{1}{2}c_1 + \frac{1}{2}c_2}{m_1}$. By (3.25) and (3.26), the set of admissible centers $= \{(h, k) \in \gamma : Q(h) > 0 \text{ and } R(k) > 0\}$, and $\frac{b^2}{a^2} = \frac{R(k)}{Q(h)} = \frac{R(f(h))}{Q(h)} = \frac{R(k)}{Q(g(k))}$ if $(h, k) \in \gamma$. Now

$$Q(g(y_1)) = Q(g(y_2)) = \frac{1}{4}\frac{(c_1 - c_2)^2}{m_1^2} \neq 0 \tag{3.27}$$

and

$$R(f(x_1)) = R(f(x_2)) = \frac{1}{4}(c_1 - c_2)^2 \neq 0 \tag{3.28}$$

Case 1: $m_3 < m_2$

Let $s = \inf\{h < h_0 : R(f(h)) > 0 \text{ and } Q(h) > 0\}$. Since $(h_0 - x_1)(h_0 - x_2) = \frac{1}{4}\left(\sqrt{\left(\frac{1+m_3^2}{1+m_2^2}\right)} - 1\right)(-c_2 + c_1)^2 \frac{\sqrt{\left(\frac{1+m_3^2}{1+m_2^2}\right)}+1}{(m_2-m_3)^2} < 0$, h_0 is between x_1 and x_2, and hence $-\infty < s$. Note that $\frac{R(f(h_0))}{Q(h_0)} = 1$. First, if $Q(s) = 0$, then $s = x_j \Rightarrow R(f(s)) \neq 0$ by (3.28), and thus $\lim_{h \to s^+} \frac{R(f(h))}{Q(h)} = \infty$. In that case $\frac{a^2}{b^2}$ varies from 0 to 1 on (s, h_0), and the equation $1 - \frac{a^2}{b^2} = e_0^2$ has a solution (h, k) which is an admissible center. Second, if $R(f(s)) = 0$, then $f(s) = y_j \Rightarrow s = g(y_j) \Rightarrow Q(s) \neq 0$ by (3.27), and thus $\lim_{h \to s^+} \frac{R(f(h))}{Q(h)} = 0$. In that case $\frac{b^2}{a^2}$ varies from 0 to 1 on (s, h_0), and again the equation $1 - \frac{b^2}{a^2} = e_0^2$ has a solution (h, k) which is an admissible center.

Case 2: $m_3 > m_2$

Let $s = \inf\{k < k_0 : R(k) > 0 \text{ and } Q(g(k)) > 0\}$. Since $(k_0 - y_1)(k_0 - y_2) = \frac{1}{4}(-c_2 + c_1)^2 \frac{m_2^2 - m_3^2}{(1+m_2^2)(m_2-m_3)^2}$, k_0 is between y_1 and y_2, and hence $-\infty < s$. The rest of the proof follows as in case 1.

□

3.5.1. *Example.* $T_1 : y = x + 1$, $T_2 : y = x$, $T_3 : y = -2x$; The intersection points of T_1 and T_3, and of T_2 and T_3, respectively, are $(-1/3, 2/3)$ and $(0, 0)$. Hence $x_1 = -\frac{1}{3}$, $x_2 = 0$, $u_1 = \frac{1}{6}$, $u_2 = -\frac{1}{2}$. The admissible centers lie on the line L with equation $k = h + \frac{1}{2}$. Since $m_1 m_3 > 0$, $S = \{(h, h + \frac{1}{2}) : -\frac{1}{2} < h < -\frac{1}{3} \text{ or } 0 < h < \frac{1}{6}\}$. See Figure 3.2 for a plot of the T_j and the ellipse $\frac{48(x+5/12)^2}{5} + \frac{48(y-1/12)^2}{7} = 1$.

3.6. 3. All Three of the T_j have slopes equal in absolute value.

In this case the set of admissible centers consists of two points. For each such center, there are infinitely many ellipses tangent to the T_j.

THEOREM 3.6. *Let T_j, $j = 1, 2, 3$ be three distinct non–vertical lines with equations $y = m_j x + c_j$, and assume that $m_1^2 = m_2^2 = m_3^2$. Assume also that the T_j do not have a common intersection point, and let (x_1, y_1) and (x_2, y_2) be the intersection points of $\{T_1, T_3\}$ and $\{T_2, T_3\}$, respectively. Let S denote the set of admissible centers (h, k). Then S consists of the two points $(\frac{c_3 - c_2}{2m_1}, \frac{c_1 + c_3}{2}) = (x_2, y_1)$ and $(\frac{c_3 - c_1}{2m_1}, \frac{c_2 + c_3}{2}) = (x_1, y_2)$.*

Let a^2 and b^2 be any positive solutions of $a^2 m_1^2 + b^2 = \frac{1}{4}(c_1 - c_2)^2$. Then, for each $(h, k) \in S$, the ellipse $\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$ is tangent to each of the T_j. Finally, given any $0 \le e_0 < 1$ and any $(h, k) \in S$, one can choose a^2 and b^2 such that E has eccentricity e_0.

PROOF. Suppose that $m_1^2 = m_2^2 = m_3^2$. Given $C = (h, k)$, write T_j in the form $y - k = m_j(x - h) + b_j$, $j = 1, 2, 3$. By Theorem 2.6, in order for an ellipse $E : \frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ to exist which is tangent to T_1, T_2, and T_3, we must have

$$b_1^2 = b_2^2 = b_3^2 \tag{3.29}$$

along with

$$a^2m_j^2 + b^2 = b_j^2,\ j = 1, 2, 3 \tag{3.30}$$

Clearly, if $m_1 = m_2 = m_3$, then there is no ellipse tangent to the T_j. Hence we can assume, without loss of generality, that $m_1 = m_2 = -m_3$.

Arguing as in subsection 2(see (3.19)), $(2m_j - 2m_1)hk + (2m_1c_1 - 2m_jc_j)h + (2c_j - 2c_1)k + c_1^2 - c_j^2 = 0$ for $j = 2, 3$, which yields the two simultaneous equations

$$2hm_1 - 2k + c_1 + c_2 = 0,\ (2k - c_1 - c_3)(2m_1h - c_3 + c_1) = 0$$

with solutions $(h,k) = (\frac{c_3-c_2}{2m_1}, \frac{c_1+c_3}{2}) = (x_2, y_1)$ and $(\frac{c_3-c_1}{2m_1}, \frac{c_2+c_3}{2}) = (x_1, y_2)$. Note that those points are distinct since $c_1 \neq c_2$(else $T_1 = T_2$). Using $b_j = m_jh + c_j - k$, it is not hard to show that

$(h,k) = (\frac{c_3-c_2}{2m_1}, \frac{c_1+c_3}{2})$ if and only if $b_1 = \frac{1}{2}(c_1 - c_2)$, $b_2 = -b_1$, and $b_3 = -b_1$, and

$(h,k) = (\frac{c_3-c_1}{2m_1}, \frac{c_2+c_3}{2})$ if and only if $b_1 = \frac{1}{2}(c_1 - c_2)$, $b_2 = -b_1$, and $b_3 = b_1$.

For each choice of (h,k) above, the b_j are nonzero with $b_1^2 = b_2^2 = b_3^2$. As a^2 varies between 0 and $\frac{b_1^2}{m_1^2}$, one obtains infinitely many positive solutions a^2, b^2 of the equations in (3.30), and $\frac{b^2}{a^2}$ varies from ∞ to 0. Thus one can also find an ellipse with any preassigned eccentricity. $\square$

REMARK 3.7. One can easily prove a *rotated* version of Theorems 3.1, 3.3, 3.5, and 3.6 above. The statement of the results is more complicated, but the outline of the proof is the same. One would have to consider separately cases where one of the T_j is parallel to a coordinate axis in the new coordinate system, as we did with horizontal or vertical lines when $\alpha = 0$. Given α, just replace m_j by $\frac{m_j + \tan\alpha}{1 - m_j\tan\alpha}$ and b_j by $\frac{b_j\sec\alpha}{1 - m_j\tan\alpha}$ throughout. For example, in Theorem 3.1, each of the variables a_j, d, β, and τ depend on α, but the intersection points of the T_j do not, of course, depend on α.

REMARK 3.8. It is easy to show that there are points (h,k), $(h,k) \notin T_1 \cup T_2 \cup T_3$, which are not α admissible for any α. For example, for the given lines $y = x + 1$, $y = 2x - 2$, and $y = 0$, $(0, \frac{1}{2})$ is not α admissible for any α. This leads to the following question. Given T_1, T_2, T_3 with no common intersection point, is any $(h,k) \notin T_1 \cup T_2 \cup T_3$ α admissible for some α, if one allows for *hyperbolas* as well as ellipses ?

REMARK 3.9. For Theorems 3.1, 3.3, 3.5, and 3.6, it is not hard to show that one can specify the *area* in advance instead of the eccentricity.

4. Four Given Tangents

Given four lines in the plane, such that no three of the lines are parallel or have a common intersection point, we want to prove that there is some ellipse E which is tangent to each of the four lines. Of course, simple examples show that this result is only true in general for *rotated* ellipses. Hence, in this section we allow the angle of rotation α to vary. First we need the following results.

LEMMA 4.1. *$f(s,t) = s^2 - 4s + 6st + 4 - 4t + t^2$ is **positive** on $\{(s,t) : s > 0$ and $t > 0\}$.*

PROOF. The implicit curve $C : s^2 - 4s + 6st + 4 - 4t + t^2 = 0$ C has horizontal tangents at $(2,0), (-1,1)$, and vertical tangents at $(0,2), (1,-1)$, and hence is tangent to the s–axis at $(2,0)$ and tangent to the t–axis at $(0,2)$. Since the second derivative is never zero on C, it then follows easily that C never touches

$\{(s,t) : s > 0 \text{ and } t > 0\}$. Since $f(1,1) = 4 > 0$, the lemma follows from the Intermediate Value Theorem.

PROPOSTION 4.2. *Suppose that we are given four lines in the plane, such that no two of the lines are parallel or have a common intersection point. Then the four lines must form the boundary of at least one four sided convex polygon D.*

PROOF. Pick a line L. The other lines intersect L at three distinct points since no three lines have a common intersection point. Let M be the line which intersects L between the other two intersection points. The interior of the triangle T formed by the three lines not equal to M is cut by M. The two regions into which M divides T are each convex, since they are the intersections of convex regions(a triangle and a half plane). One of these regions is D.

THEOREM 4.3. *Let T_1, T_2, T_3, T_4 be four given lines in the plane, such that no three of the T_j are parallel or have a common intersection point. Then there is an ellipse E which is tangent to each of the T_j.*

PROOF. We prove the case when no **two** of the lines are parallel, the proof of the other cases being similar. By Proposition 4.2, the T_j form the boundary of a four sided convex polygon D. Choose any three vertices P_1, P_2, P_3 of D, and let A be a non-singular affine map which sends the P_j to the points $(0,0), (1,0)$, and $(0,1)$. By properties of non-singular affine maps, $A(D)$ must be a convex quadrilateral,with no two sides parallel. Hence A sends the other vertex of D to the point (s,t), where $s > 0, t > 0$, $s \neq 1 \neq t$, and $s + t \neq 1$. Let $L_j,\ j = 1,2,3,4$ denote the lines which make up the boundary of $A(D)$. We shall find an ellipse $E : \frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2}$ tangent to each of the L_j. Then the preimage of E under the map A is an ellipse tangent to each of the T_j. Write $L_1 : y = 0, L_2 : x = 0, L_3 : y = 1 + \left(\frac{t-1}{s}\right) x, L_4 : y = \frac{t}{s-1}(x-1)$

To prove the existence of E, we shall apply Theorem 3.4 to L_1, L_2, and L_3, and to L_1, L_2, and L_4. Let γ_1 denote the curve $2m_3hk - 2m_3(m_3c_1 + c_3)h + 2(c_3 - c_2)k = c_3^2 - c_2^2 - m_3^2c_1^2$, with $m_3 = \frac{t-1}{s}$ and $c_3 = 1$, and let γ_2 denote the same curve, with $m_3 = \frac{t}{s-1}$ and $c_3 = -\frac{t}{s-1}$. The key is showing that γ_1 and γ_2 have at least one intersection point. The equations for γ_1 and γ_2 are

$2\left(\frac{t-1}{s}\right) hk - 2\left(\frac{t-1}{s}\right) h + 2k = 1 \Rightarrow$

$$k = \frac{1}{2} \frac{(2t-2)h + s}{(t-1)h + s} \tag{4.1}$$

and

$2\left(\frac{t}{s-1}\right) hk + 2\left(\frac{t}{s-1}\right)^2 h - 2\left(\frac{t}{s-1}\right) k = \left(\frac{t}{s-1}\right)^2 \Rightarrow$

$$k = -\frac{1}{2} t \frac{2h-1}{(h-1)(s-1)} \tag{4.2}$$

Note that if $(t-1)h + s = 0$, then $2\left(\frac{t-1}{s}\right) hk - 2\left(\frac{t-1}{s}\right) h + 2k = 2 \neq 1$. Hence the denominator in (4.1) cannot be zero. Also, if $h = 1$, then $2\left(\frac{t}{s-1}\right) hk + 2\left(\frac{t}{s-1}\right)^2 h -$

$2\left(\frac{t}{s-1}\right)k = 2\left(\frac{t}{s-1}\right)^2 \neq \left(\frac{t}{s-1}\right)^2$ since $t \neq 0$. Hence the denominator in (4.2) is well defined. Now we set the expressions for k in (4.1) and (4.2) equal to one another. This yields $\frac{1}{2}\frac{(2t-2)h+s}{(t-1)h+s} + \frac{1}{2}t\frac{2h-1}{(h-1)(s-1)} = \frac{1}{2}(s-1+t)\frac{(2t-2)h^2+(s+2-t)h-s}{((t-1)h+s)(h-1)(s-1)} = 0$, which holds if and only if $(2t-2)h^2 + (s+2-t)h - s = 0$. The discriminant is

$(s+2-t)^2 + 4s(-2+2t) = s^2 - 4s + 6st + 4 - 4t + t^2$, which is positive on

$\{(s,t) : s > 0 \text{ and } t > 0\}$ by Lemma 4.1. Hence γ_1 and γ_2 have two points of intersection (h_1, k_1) and (h_2, k_2). Finally, to apply Theorem 3.4, we must show that $h_j \neq c_1$ and $k_j \neq c_2$, where $c_1 = c_2 = 0$. It follows easily that $h = 0$ in γ_1 and/or $k = 0$ in γ_2 implies that $s + t = 1$. Hence there are actually two ellipses tangent to each of the L_j. □

Remark 4.4. Similar methods show that Theorem 4.3 also holds for hyperbolas.

Remark 4.5. Simple examples show that in general one cannot specify the eccentricity in advance in Theorem 4.3.

4.1. Example. $T_1 : y = x+2, T_2 : y = 3x, T_3 : y = -2x-1, T_4 : y = -4x+1$. The T_j enclose the convex quadrilateral D with vertices $\{(-\frac{1}{5}, -\frac{3}{5}), (-1,1), (-\frac{1}{5}, \frac{9}{5}), (\frac{1}{7}, \frac{3}{7})\}$. Let A be the affine transformation with $A(-1,1) = (0,0)$, $A(-\frac{1}{5}, -\frac{3}{5}) = (1,0)$, and $A(-\frac{1}{5}, \frac{9}{5}) = (0,1)$. Then $A(\frac{1}{7}, \frac{3}{7}) = (\frac{5}{7}, \frac{5}{7}) \Rightarrow s = t = \frac{5}{7}$. The two solutions of the equations (4.1) and (4.2) are $(\frac{7}{4} \pm \frac{1}{4}\sqrt{29}, \frac{7}{4} \pm \frac{1}{4}\sqrt{29}) \approx (3.0963, 3.0963)$ and $(.40371, .40371)$. Mapping back to D gives two ellipses tangent to the given T_j with centers $\approx (-.35406, .67703)$ and $(3.9541, -1.477)$. One of the ellipses is inscribed in D(see Figures 4.1 and 4.2). [10]

5. Discussion

Finding general conics tangent to a given set of lines is discussed in books on projective geometry, where the duality between points and lines is used. The dual of finding a conic passing thru given points is finding a conic tangent to given lines. Salmon([**3**]), Liming([**1**], page 242) and Fishback([**2**]) discuss, in particular, how to fit a a conic to *five* given tangents. It is not clear, however, whether duality would more easily yield specific results about fitting less than five tangents to an ellipse, given certain information about the ellipse(such as the center, orientation of the axes, or the eccentricity), or determining when a given set of lines is the tangent to such an ellipse. We have not seen any results give necessary and sufficient conditions on when a finite set of lines is tangent to a certain type of conic, such as the family of ellipses with a specified center we discussed in this paper. Salmon does discuss briefly the locus of centers of a conic tangent to four lines, or to three lines with another condition on the sum of the squares of the axes. However, his results do not imply Theorem 4.3 or any of the Theorems in section 3.

6. Future Research

· Given three lines T_1, T_2, and T_3, let S equal the set of points (h,k) which are α admissible for *some* α. It would be nice to have a simple characterization of S, perhaps using the theorems of Section 3. This would yield a fairly simple algorithm

[10] Using Theorem 3.1, applied to T_1, T_2, T_3, and then T_2, T_3, T_4, it is not hard to show that there is no *non-rotated* ellipse tangent to all four lines!

for determining if a given (h, k) is the center of some ellipse tangent to the T_j. Of course, it is geometrically obvious that S must lie a certain positive distance from each of the given lines.

· The techniques of this paper can be used to obtain similar results for faimilies of curves other than ellipses or hyperbolas. The main difficuty lies in finding explicit solutions of certain nonlinear equations. That difficulty does not arise for, say, the family of curves $y = c_1 e^{c_2 x} + c_3$.

· There are obvious ways to try to extend the results of this paper to higher dimensions–that is, find an ellipsoid tangent to finitely many given planes.

Acknowledgement: We thank Jim Ferry of the Center for Simulation of Advanced Rockets for the proof of Proposition 4.2.

References

[1] Roy A. Liming, Practical Analytic Geometry with Applications to Aircraft, 1944, MacMillan, New York.

[2] W. T. Fishback, Projective and Euclidean Geometry, 2nd ed., 1969, Wiley, New York.

[3] George Salmon, A Treatise on Conic Sections, 1954, Chelsea, New York.

[4] Alan Horwitz, Reconstructing a function from its set of tangent lines, American Mathematical Monthly 96(1989), 807-813.

[5] Cox, Litle, and O'Shea, Ideals, Varieties, and Algorithms, 1992, Springer-Verlag, New York.

Fig. 2.1

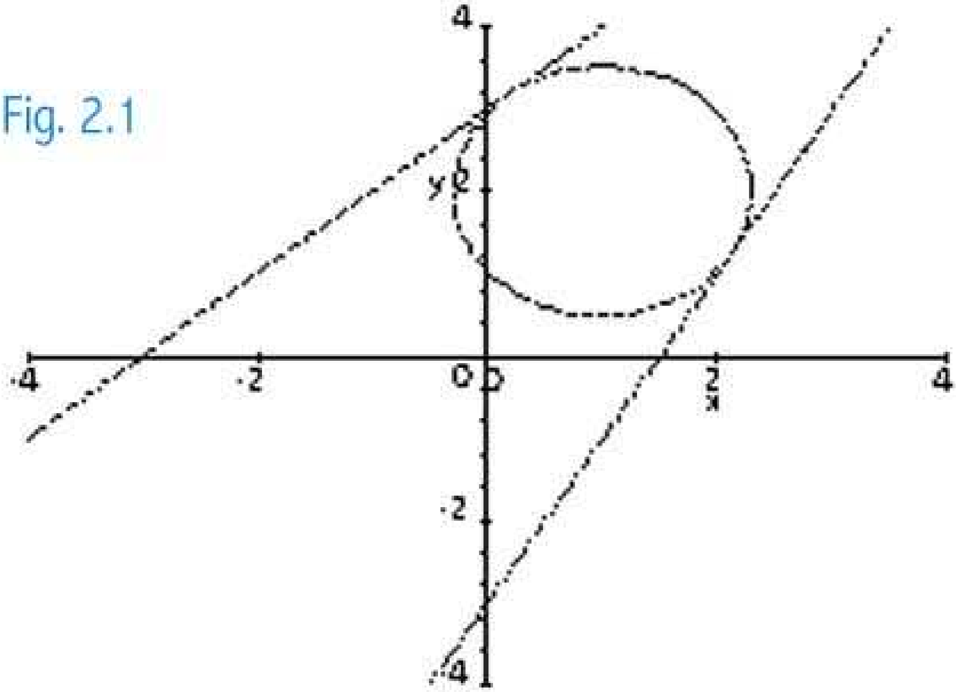


Fig. 2.2

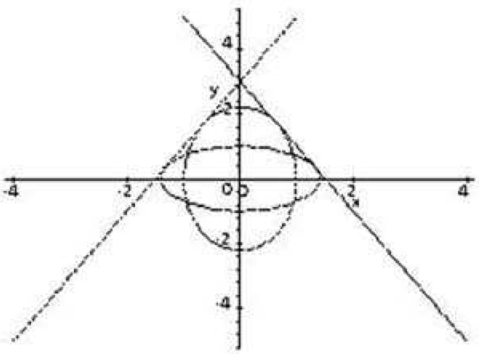


Fig. 3.1

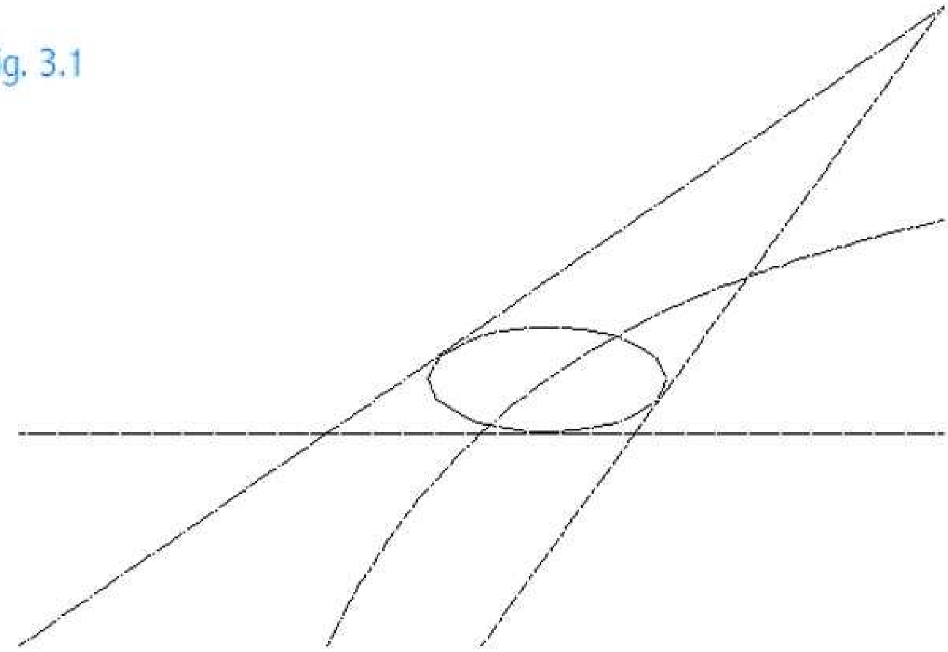
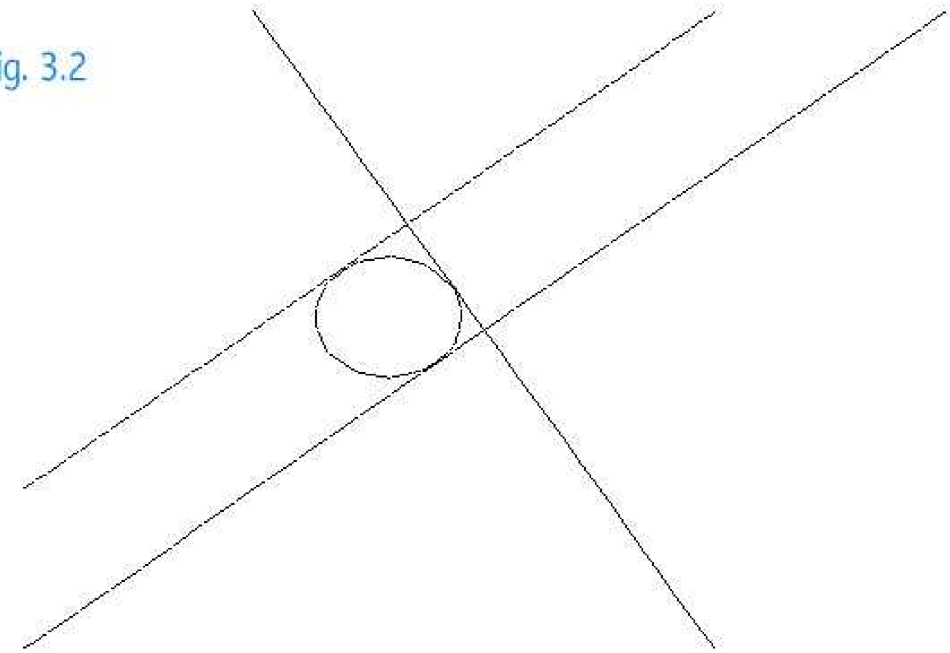


Fig. 3.2



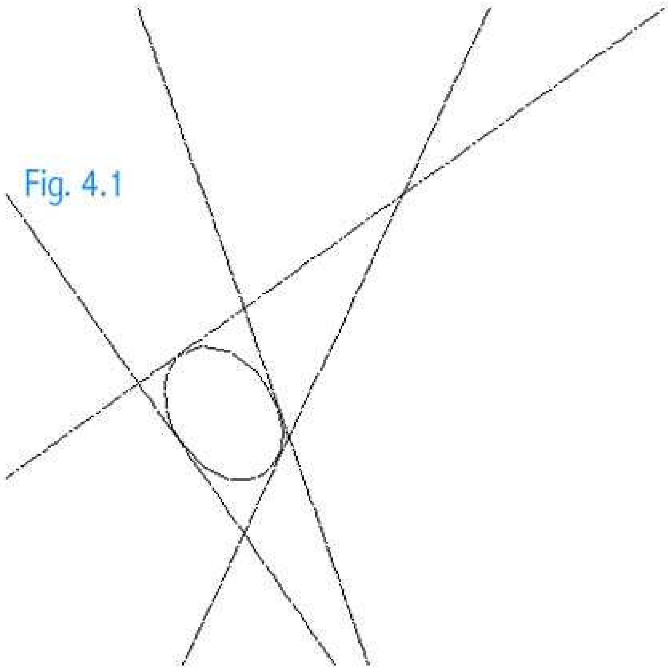
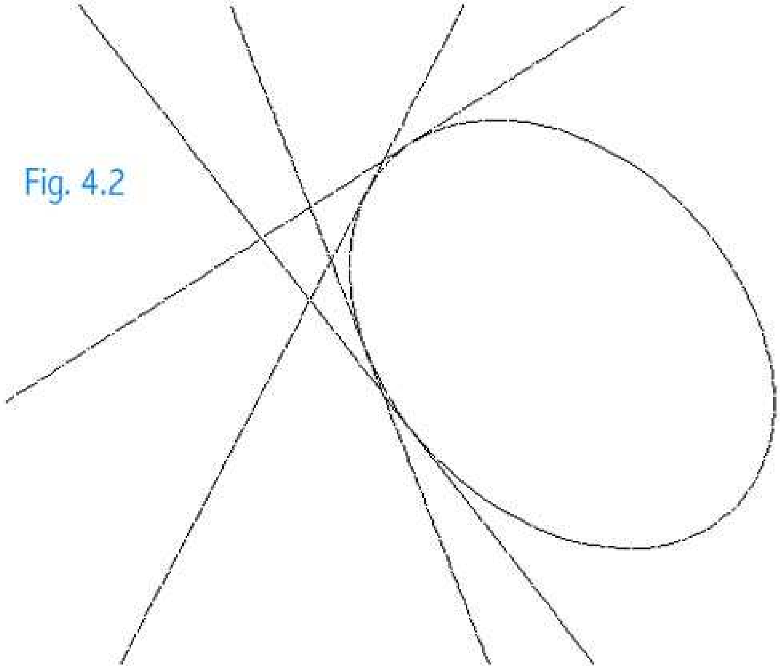


Fig. 4.1

Fig. 4.2



Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2000, pp. 152–159.
Submitted: September 13, 2000. Published: December 15, 2000.

About the Existence of the Thermodynamic Limit for some Deterministic Sequences of the Unit Circle

Stefano Siboni

Abstract We show that in the set $\Omega = \mathbb{R}_+ \times (1, +\infty) \subset \mathbb{R}^2_+$, endowed with the usual Lebesgue measure, for almost all $(h, \lambda) \in \Omega$ the limit $\lim_{n\to+\infty} \frac{1}{n} \ln \left| h\big(\lambda^n - \lambda^{-n}\big) \bmod [-1/2, 1/2) \right|$ exists and is equal to zero. The result is related to a characterization of relaxation to equilibrium in mixing automorphisms of the two-torus. It is nothing but a curiosity, but maybe you will find it nice.

1991 A.M.S. Subject Classification Codes: 28,30,26.

Keywords and Phrases: Thermodynamic Limit.

1. Introduction.

1. Introduction.
In the analysis of relaxation to equilibrium of mixing automorphisms of the two-torus [1−2] one encounters the following problem. Suppose that the one-torus is parametrized by the unit interval $[-1/2, 1/2)$ and for appropriate constants $h > 0$ and $\lambda > 1$ consider the real sequence

$$x_n = h\big(\lambda^n - \lambda^{-n}\big) \bmod [-1/2, 1/2) \qquad \forall\, n \in \mathbb{N}\,. \tag{1}$$

A significant definition of an exponential "relaxation rate" can be given if the so-called "thermodynamic" limit [3]

$$\lim_{n\to+\infty} -\frac{1}{n} \ln |x_n| \tag{2}$$

exists and is equal to zero. Existence of (2) is clearly not obvious, since the x_n's typically wander through the whole interval $[-1/2, 1/2)$ but every so often they visit a small neighborhood of zero, where the logarithm is singular. Actually, not even if one replaces the ordinary limit in (2) with a supremum limit the finiteness of the result is assured.

This short note is devoted to a measure theoretical discussion of the previous problem. One can show that existence to zero of limit (2) occurs almost surely, for almost any choice of the parameters h and λ, with respect to a measure suitably defined.

Dipartimento di Ingegneria dei Materiali Facoltà di Ingegneria, Università di Trento
Trento - Italy
E-mail: Stefano.Siboni@ing.unitn.it

2. Results.

Our goal is to prove the statement below.

Theorem 1 *Consider the set* $\Omega = \mathbb{R}_+ \times (1, +\infty) \subset \mathbb{R}^2_+$ *endowed with the usual Lebesgue measure* μ. *Then, for* μ*-almost all* $(h, \lambda) \in \Omega$ *there holds*

$$\lim_{n\to+\infty} \frac{1}{n} \ln \left| h\left(\lambda^n - \lambda^{-n}\right) \mathrm{mod}\,[-1/2, 1/2) \right| = 0 \; . \tag{3}$$

This result can be easily deduced by means of standard arguments of measure theory once the following main theorem is proved.

Theorem 2 *Let* $h > 0$ *and* $Q \in \mathbb{N}$, $Q > 1$, *some fixed constants. Consider the set* G *of all* $\lambda \in [1, Q]$ *for which a (possibly* λ*-dependent) real sequence* $(a_n)_{n\in\mathbb{N}}$ *and an integer* $n' \in \mathbb{N}$ *exist such that*

$$\begin{aligned} &(a) \quad a_n > 0 \quad \forall\, n > n'; \\ &(b) \quad a_n \le \left| h\left(\lambda^n - \lambda^{-n}\right) \mathrm{mod}\,[-1/2, 1/2) \right| \quad \forall\, n > n'; \\ &(c) \quad \lim_{n\to+\infty} \frac{1}{n} \ln a_n = 0 \end{aligned}$$

Then, if μ *denotes the Lebesgue measure on* $\mathbb{R}$:

(1) *the set* $G \subseteq [1, Q]$ *is actually nonempty;*

(2) G *is* μ*-measurable and its measure holds* $\mu(G) = Q - 1$.

As a consequence, the set $B = [1, Q]\backslash G$ *where conditions* (a), (b), (c) *are not simultaneously satisfied is also* μ*-measurable and of vanishing measure.*

We firstly prove the result by considering values of λ in the interval $[1 + \eta, Q]$, with η small positive number arbitrarily fixed ($\eta < 1/2$). We therefore look for the subset G_η of $\lambda \in [1 + \eta, Q]$ where hypotheses (a), (b), (c) are satisfied by a suitable choice of the sequence $(a_n)_{n\in\mathbb{N}}$ and of the integer $n' \in \mathbb{N}$. The basic idea of the proof is that the μ-measure of G_η turns out to be $Q - 1 - \eta$ even if we confine ourselves to choose the sequence $(a_n)_{n\in\mathbb{N}}$ in the form

$$a_n = \frac{1}{n^2} \quad \forall\, n \in \mathbb{N},$$

which certainly fulfills requirements (a) and (c), and enable us to deal with the *only* condition (b) on λ.

Let us take then $a_n = 1/n^2 \;\forall\, n \in \mathbb{N}$ and an arbitrarily given value of $n \in \mathbb{N}$. Before tackling the real proof, we need some definitions.

Definition 1. We introduce the set $B_n \subseteq [1 + \eta, Q]$

$$B_n = \left\{ \lambda \in [1 + \eta, Q] \,:\, a_n > \left| h\left(\lambda^n - \lambda^{-n}\right) \mathrm{mod}\,[-1/2, 1/2) \right| \right\}$$

i.e. the set of $\lambda \in [1 + \eta, Q]$ where the condition $a_n \le \left| h\left(\lambda^n - \lambda^{-n}\right) \mathrm{mod}\,[-1/2, 1/2) \right|$ is not satisfied for the assigned $n \in \mathbb{N}$.

Notice that B_n is a finite union of intervals because the function $\Phi_n(\lambda) = \lambda^n - \lambda^{-n}$ is strictly increasing in $[+1, +\infty)$ at fixed n. In fact

$$\Phi_n'(\lambda) = n\big(\lambda^{n-1} + \lambda^{-(n+1)}\big) > 0 \quad \forall\, \lambda \in [1, +\infty) .$$

Consequently, B_n is μ-measurable as a finite union of bounded intervals.

Definition 2. We further introduce the set $\hat{B}_{n'} \subseteq [1+\eta, Q]$, $n' \in \mathbb{N}$, given by

$$\hat{B}_{n'} = \Big\{\lambda \in [1+\eta, Q] \,:\, a_n > \big|h\big(\lambda^n - \lambda^{-n}\big) \bmod [-1/2, 1/2)\big| \,,\ n > n'\Big\} = \bigcup_{n>n'} B_n$$

which is obviously μ-measurable as a countable union of μ-measurable sets.

Definition 3. We finally introduce the "bad" set $B_\eta \subseteq [1+\eta, Q]$

$$B_\eta = \bigcap_{n'=1}^{\infty} \hat{B}_{n'} \ ,$$

where condition (b) is not satisfied — with this particular choice of the sequence $(a_n)_{n\in\mathbb{N}}$.

B_η is also a μ-measurable set, as a countable intersection of μ-measurable sets.

An immediate consequence of the previous definitions is that $[1+\eta, Q] \setminus G_\eta = B_\eta$. Our goal is to prove that $\mu(B_\eta) = 0$. To this end, since $\forall\, n' \in \mathbb{N} B_\eta \subseteq \hat{B}_{n'}$ by definition, it is enough to show that

$$\lim_{n' \to +\infty} \mu(\hat{B}_{n'}) = 0 \ .$$

Therefore, we can confine ourselves to consider values of $n' \in \mathbb{N}$ large enough, and owing to Definition 2 we can also assume values of $n \in \mathbb{N}$ greater that n'. More precisely, we impose the following technical requirements on the size of n' and n. We take $n > n' \in \mathbb{N}$ such that:

$$
\begin{aligned}
&(i) \quad a_{n'} = \frac{1}{n'^2} < \eta \quad \Longrightarrow \quad a_n < \eta \quad \forall\, n > n' \\
&(ii) \quad a_n - \frac{h}{(1+\eta)^n} = \frac{1}{n^2} - \frac{h}{(1+\eta)^n} > 0 \quad \forall\, n > n' \\
&(iii) \quad h\big[(1+\eta)^n - (1+\eta)^{-n}\big] > 3/2 \quad \textit{and} \quad h\big[Q^n - Q^{-n}\big] > 5/2 \quad \forall\, n > n' \ .
\end{aligned}
$$

Under the previous conditions we can state the lemmas below.

Lemma 1 *The μ-measure of B_n, n as above, admits the upper bound*

$$\mu(B_n) \le 2\varepsilon_n \left(\frac{1}{h}\right)^{\frac{1}{n}} [h(Q^n - Q^{-n})]^{\frac{1}{n}} \ .$$

where $\varepsilon_n = a_n + \frac{h}{(1+\eta)^n} > 0$.

Proof

We firstly notice that $1 + a_n < 1 + \eta$ by (i); on the other hand, since $\eta < 1/2$ by hypothesis, (i) implies $a_n < 1/2$, so that all the intervals $(p - a_n, p + a_n)$, $p = 2, \dots \lfloor h(Q^n - Q^{-n}) \rfloor + 1$ are disjoint.

By using (iii) and denoted with I_n the integer set $\{2, 3, \dots, \lfloor h(Q^n - Q^{-n}) \rfloor + 1\}$, we deduce

$$
\begin{aligned}
B_n \subseteq \Big\{ \lambda \in [1+\eta, Q] \,:\, h\big(\lambda^n - \lambda^{-n}\big) \in \textstyle\bigcup_{p=2}^{\lfloor h(Q^n - Q^{-n}) \rfloor + 1} (p - a_n, p + a_n) \Big\} &= \\
= \Big\{ \lambda \in [1+\eta, Q] \,:\, p - a_n < h\big(\lambda^n - \lambda^{-n}\big) < p + a_n \;,\, p \in I_n \Big\} &= \qquad (4) \\
= \Big\{ \lambda \in [1+\eta, Q] \,:\, p - \big(a_n - h\lambda^{-n}\big) < h\lambda^n < p + a_n + h\lambda^{-n} \,,\; p \in I_n \Big\} &.
\end{aligned}
$$

Now it is clear that $\forall\, \lambda \in [1 + \eta, Q]$

$$
a_n - \frac{h}{(1+\eta)^n} \le a_n - h\lambda^{-n} < a_n + h\lambda^{-n} \le a_n + \frac{h}{(1+\eta)^n}
$$

and by (ii)

$$
a_n - \frac{h}{(1+\eta)^n} > 0
$$

from which we obtain

$$
0 < a_n - \frac{h}{(1+\eta)^n} \le a_n - h\lambda^{-n} < a_n + h\lambda^{-n} \le a_n + \frac{h}{(1+\eta)^n} \quad \forall\, \lambda \in [1+\eta, Q] \,.
$$

By enlarging each covering interval in (4), we are led then to the inclusion

$$
B_n \subseteq \left\{ \lambda \in [1+\eta, Q] \,:\, p - \left(a_n + \frac{h}{(1+\eta)^n}\right) < h\lambda^n < p + a_n + \frac{h}{(1+\eta)^n} \,,\; p \in I_n \right\}
$$

and recalling the definition of ε_n

$$
\begin{aligned}
B_n \subseteq \left\{ \lambda \in [1+\eta, Q] \,:\, \left(\frac{p - \varepsilon_n}{h}\right)^{\frac{1}{n}} < \lambda < \left(\frac{p + \varepsilon_n}{h}\right)^{\frac{1}{n}} ,\; p \in I_n \right\} &\subseteq \\
\subseteq \bigcup_{p=2}^{\lfloor h(Q^n - Q^{-n}) \rfloor + 1} \left(\left(\frac{p - \varepsilon_n}{h}\right)^{\frac{1}{n}} , \left(\frac{p + \varepsilon_n}{h}\right)^{\frac{1}{n}} \right) &,
\end{aligned}
$$

the final set being μ-measurable as a finite union of intervals. Whence

$$
\mu(B_n) \le \sum_{p=2}^{\lfloor h(Q^n - Q^{-n}) \rfloor + 1} \left(\frac{1}{h}\right)^{\frac{1}{n}} \left[(p + \varepsilon_n)^{\frac{1}{n}} - (p - \varepsilon_n)^{\frac{1}{n}} \right] .
$$

Moreover, $\forall\, p = 2, 3, \dots \lfloor h(Q^n - Q^{-n}) \rfloor + 1$ Lagrange Mean Value Theorem implies the equalities below

$$
(p + \varepsilon_n)^{\frac{1}{n}} - (p - \varepsilon_n)^{\frac{1}{n}} = \frac{1}{n} (p + \xi_p)^{\frac{1}{n} - 1} \, 2\varepsilon_n
$$

for some $\xi_p \in (-\varepsilon_n, \varepsilon_n)$, and since

$$(p+\xi_p)^{\frac{1}{n}-1} \;=\; \frac{1}{(p+\xi_p)^{1-\frac{1}{n}}} \leq \frac{1}{(p-1)^{1-\frac{1}{n}}}$$

we conclude that

$$\mu(B_n) \leq \sum_{p=2}^{\lfloor h(Q^n-Q^{-n})\rfloor+1} \Big(\frac{1}{h}\Big)^{\frac{1}{n}} \, \frac{1}{n} \, \frac{1}{(p-1)^{1-\frac{1}{n}}} \, 2\varepsilon_n \;=\; \frac{2\varepsilon_n}{n} \big(\frac{1}{h}\big)^{\frac{1}{n}} \sum_{p=1}^{\lfloor h(Q^n-Q^{-n})\rfloor} \frac{1}{p^{1-\frac{1}{n}}} \; .$$

As $\big(p^{1-\frac{1}{n}}\big)^{-1}$ is a decreasing function of p, the following upper estimate holds

$$\sum_{p=1}^{\lfloor h(Q^n-Q^{-n})\rfloor} \frac{1}{p^{1-\frac{1}{n}}} \;\leq\; \int\limits_0^{\lfloor h(Q^n-Q^{-n})\rfloor} p^{\frac{1}{n}-1}\, dp \;=\; \Big[np^{\frac{1}{n}} \Big]_0^{\lfloor h(Q^n-Q^{-n})\rfloor} \;=\; n \, \lfloor h(Q^n-Q^{-n}) \rfloor^{\frac{1}{n}}$$

and finally $\mu(B_n) \leq 2\varepsilon_n \left(\frac{1}{h}\right)^{\frac{1}{n}} [h(Q^n-Q^{-n})]^{\frac{1}{n}}$, which completes the proof. □

Lemma 2 *If $n' > 0$ (satisfying (i), (ii) and (iii)) is sufficienlty large, the μ-measure of $\hat{B}_{n'}$ is bounded by*

$$\mu(\hat{B}_{n'}) \leq 2(Q+\varepsilon) \sum_{n=n'+1}^{\infty} \varepsilon_n$$

for some small $\varepsilon > 0$.

Proof
Because of the identity $\hat{B}_{n'} = \bigcup_{n>n'} B_n$ and using Lemma 1, we have the following estimate

$$\mu(\hat{B}_{n'}) \leq \sum_{n=n'+1}^{\infty} \mu(B_n) \leq 2 \sum_{n=n'+1}^{\infty} \varepsilon_n \, \big(\frac{1}{h}\big)^{\frac{1}{n}} [h(Q^n-Q^{-n})]^{\frac{1}{n}} \; .$$

Notice that $\forall\, h > 0, \forall\, Q \in \mathbb{N}, Q > 1$

$$\lim_{n\to+\infty} \Big(\frac{1}{h}\Big)^{\frac{1}{n}} [h(Q^n-Q^{-n})]^{\frac{1}{n}} \;=\; Q \; .$$

so that for some $\varepsilon > 0$, $\varepsilon \ll Q$, and n' sufficienlty large there holds

$$Q-\varepsilon < \Big(\frac{1}{h}\Big)^{\frac{1}{n}} [h(Q^n-Q^{-n})]^{\frac{1}{n}} < Q+\varepsilon \quad \forall\, n \in \mathbb{N} \; , \quad n > n' \; .$$

Whence for n' as above

$$\mu(\hat{B}_{n'}) \leq 2 \sum_{n=n'+1}^{\infty} \varepsilon_n \, (Q+\varepsilon) \;=\; 2(Q+\varepsilon) \sum_{n=n'+1}^{\infty} \varepsilon_n$$

which is finite, owing to

$$\sum_{n=1}^{\infty} \varepsilon_n \;=\; \sum_{n=1}^{\infty}\left(\frac{1}{n^2}+\frac{h}{(1+\eta)^n}\right) \;=\; \frac{\pi^2}{6} \;+\; \frac{h}{\eta}\,. \qquad \square$$

Lemma 3 *The measure of B_η is zero*

$$\mu(B_\eta) \;=\; 0\;.$$

Proof
Since $\forall\, n' \in \mathbb{N}$ we have that $B_\eta \subseteq \hat{B}_{n'}$, in particular this will be true for all $n' \in \mathbb{N}$ large enough to satisfy the requirements of the previous lemmas. Thus

$$\mu(B_\eta) \leq \mu(\hat{B}_{n'}) \leq 2(Q+\varepsilon) \sum_{n=n'+1}^{\infty} \varepsilon_n$$

and therefore

$$\mu(B_\eta) \leq \lim_{n'\to+\infty} 2(Q+\varepsilon) \sum_{n=n'+1}^{\infty} \varepsilon_n$$

where the limit is obviously zero, because of $\sum_{n=1}^{\infty} \varepsilon_n < +\infty$. By the nonnegativity of measure we have the result. $\square$

Proof of theorem 2
As a consequence of Lemma 3, the "good" set $G_\eta = [1+\eta, Q]\backslash B_\eta$ of λ-values in $[1+\eta, Q]$ satisfying condition (b) for the particular choice of $(a_n)_{n\in\mathbb{N}}$, $a_n = 1/n^2$, is of course μ-measurable and with Lebesgue measure

$$\mu(G_\eta) \;=\; Q-1-\eta-\mu(B_\eta) \;=\; Q-1-\eta\;.$$

If we now consider an *arbitrary* choice of the sequence $(a_n)_{n\in\mathbb{N}}$, compatibile again with conditions (a) and (c), the previous set G_η will maybe "grow" by a subset $\tilde{G}_\eta \subseteq [1+\eta, Q]\backslash G_\eta$:

$$G_\eta \quad\longrightarrow\quad G_\eta^o \;=\; G_\eta \cup \tilde{G}_\eta\;.$$

But as $\mu([1+\eta, Q]\backslash G_\eta) = 0$ it follows that $\tilde{G}_\eta$ is also μ-measurable and of vanishing μ-measure. Hence we finally conclude that the *full* set G_η^o, corresponding to arbitrary (a)-(c)-conditioned sequences $(a_n)_{n\in\mathbb{N}}$, is μ-measurable with measure

$$\mu(G_\eta^o) \;=\; Q-1-\eta$$

and that the corresponding *full* set $B_\eta^o[1+\eta, Q]\backslash G_\eta^o$ of λ values where condition (b) is not fulfilled for *any* (a)-(c)-conditioned sequence $(a_n)_{n\in\mathbb{N}}$ is in turn μ-measurable with vanishing μ-measure:

$$\mu(B_\eta^o) \;=\; Q-1-\eta-\mu(G_\eta^o) \;=\; 0\;.$$

So far we have proved that B_η is a set of vanishing measure in any closed interval $[1+\eta, Q]$ with $\eta > 0$. Consider now B, G in $[1, Q]$, that is, according to the previous notation

$$B \;=\; B_0 \qquad \textit{and} \qquad G \;=\; G_0\;.$$

We firstly notice that B and G are both μ-measurable because

$$G = \bigcup_{n'=1}^{\infty} \bigcap_{n>n'} S_n$$

where S_n is the finite union of *subintervals* in $[1, Q]$ (dependent on n), and $B = [1, Q]\backslash G$. Take then

$$B = \Big(B \cap [1, 1+\eta)\Big) \cup \Big(B \cap [1+\eta, Q]\Big)$$

union of disjoint sets, for some fixed $\eta \in (0, 1/2)$. The μ-measurable set $B\cap[1+\eta, Q]$ is the "bad" set in $[1+\eta, Q]$, so that by Lemma 3

$$\mu\big(B \cap [1+\eta, Q]\big) \ = \ 0 \ .$$

As for the μ-measurable set $B \cap [1, 1+\eta)$, we have the identity

$$B\cap[1, 1+\eta) \ = \ B\cap\left(\{1\}\cup\bigcup_{k=1}^{\infty} \left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right)\right) \ = \ B\cap\{1\}\cup\bigcup_{k=1}^{\infty} B\cap\left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right)$$

union of disjoint sets. But the μ-measurable set

$$B \cap \left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right) \ , \quad k \in \mathbb{N}\backslash\{0\} \ ,$$

satisfies

$$B \cap \left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right) \subseteq B \cap \left[1+\frac{\eta}{k+1}, Q\right]$$

and since $1/2 > \dfrac{\eta}{k+1} > 0$ for any given $k \in \mathbb{N}\backslash\{0\}$, we obtain

$$\mu\left(B \cap \left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right)\right) \ = \ 0 \ .$$

On the other hand there trivially holds $\mu\big(B \cap \{1\}\big) \ = \ 0$, so that

$$\mu\big(B \cap [1, 1+\eta)\big) \ = \ \mu(B \cap \{1\}) \ + \ \sum_{k=1}^{\infty} \mu\left(B \cap \left[1+\frac{\eta}{k+1}, 1+\frac{\eta}{k}\right)\right) \ = \ 0 \ .$$

Whence, finally,

$$\mu(B) \ = \ \mu\big(B \cap [1, 1+\eta)\big) \ + \ \mu\big(B \cap [1+\eta, Q)\big) \ = \ 0$$

that is $\mu(B) = 0$ and $\mu(G) = Q - 1$. The proof is complete. $\square$

References

[1] A. Bazzani, S. Siboni, G. Turchetti, S. Vaienti, From Dynamical Systems to Local Diffusion Processes, Preprint of the Centre de Physique Théorique du CNRS - Luminy - Marseille (France) CPT-91/P.2621 November 1991, published in NATO ASI Series Volume *Chaotic Dynamics: Theory and Practice*, T. Bountis editor, Plenum (1991)

[2] S. Siboni, Rilassamento all'equilibrio in un sistema mixing e analisi di un modello di diffusione modulata, Tesi di Dottorato di Ricerca dell'Università degli Studi di Bologna (Italy) (1991)

[3] S. Siboni, G. Turchetti, S. Vaienti, Thermodynamic Limit and Relaxation to Equilibrium in Toral Area-preserving Transformations, *Jour. Stat. Phys.* **75** 1/2, 167 (1994)

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2000, pp. 160 – 171
Submitted: October 22, 2000. Published: December 15, 2000

Common fixed point theorems for commutings k-uniformly Lipschitzian mappings in metric spaces

M.Elamrani, A. Mbarki and B.Mehdaoui

Abstract

In this paper we give a common fixed points existence theorems for any sequence of commuting k-uniformly Lipschitzian mappings (eventualy for $k = 1$ for any sequence of commutings nonexpansive mappings) defined on a bounded and complete metric space (X, d) with uniform normal structure. After that we deduce, by using the T.C.Lim an J.Kulesza result [11], that this result can be generalized for any family of commutings k-uniformly Lipschitzian mappings.

A.M.S. Subject Classification Codes 47H10.
Key Words and Phrases. Complete metric spaces, hyperconvex metric spaces, k-uniformly Lipschitzian mappings, common fixed points.

1 Introduction

In classical theorems concerning the existence of fixed points for families of mappings, such as the Kukatuni theorem [1] and its well know generalization due to Ryll-Nardzewski [2], the mappings of the family are usually assumed

Département de mathématiques et Informatique
Université Mohamed I Oujda, Maroc
E-mail : elamrani, mbarki, mehdaoui@sciences.univ-oujda.ac.ma

to be linear, or at least to be weakly continuous and affine [3]. In the nonlinear theory, stronger geometric structure are utilized. In particuler for a families of nonexpansives mappings, Khamsi proved in [4] that any family of nonexpansives mappings defined on metric space (X, d) with compact and normal convexity structure $\mathcal{F}$, has a common fixed points. In his proof, Khamsi investigated the concept of 1-local retract. In this paper we prove that any sequential family of k-uniformly Lipschitzian mappings defined on bounded metric space with an uniform normal convexity structure $\mathcal{F}$ with constant β, which contains all closed ball of (X, d), has a common fixed point provided that $k^2\beta < 1$. Recall that Khamsi proved in [5] that any nonexpansive mapping defined on a bounded complete metric space with uniform normal structure with constant β has a nonempty fixed point set. For more details on fixed point theory for nonexpansive and k-uniformly Lipschitzian mappings in metric spaces we refere to [8], [9] and [10].

2 Definition and Preliminary

In all the following in this work (X, d) will be a metric space. We use $B(x, r)$ to denote the closed ball centred at $x \in X$ with radius $r > 0$. For a subset A of X we write

$$r_x(A) = \sup_{y \in A} d(x, y) \tag{1}$$

$$r(A) = \inf_{x \in A} r_x(A) \tag{2}$$

$$\delta(A) = \sup_{x \in A} r_x(A) \tag{3}$$

$$cov(A) = \cap_{B \in \mathcal{F}} B \tag{4}$$

where $\mathcal{F}$ is the families of closed ball containing A. A subset A of X is said to be admissible if and only if $A = cov(A)$. In other term A is admissible if it is an intersection of a family of closed ball centred in X.

Definition 1 *Let $\mathcal{F}$ be a nonempty family of subset of X. We say that $\mathcal{F}$ defines a convexity structure on X if and only if it is stable by intersection.*

In this work, we will always assume that $\mathcal{F}$ contains the balls, and therefore we will not recall it. Also we denote by $\mathcal{A}(X)$ the smallest convexity structure on X.

Definition 2 *We say that $\mathcal{F}$ has the property (R) if and only if any decreasing sequence $(X_n)_n$ of nonempty bounded closed subsets of X with $X_n \in \mathcal{F}$ has a nonempty intersection.*

Definition 3 *i) We say that X has uniform normal structure if and only if $r(A) \leq \beta\delta(A)$ for some $0 < \beta < 1$ and every $A \in \mathcal{F}$.*
ii) We say that $\mathcal{F}$ is normal if and only if $r(A) < \delta(A)$ for every $A \in \mathcal{F}$.

Lets recall that a self mapping $T : X \to X$ is said to be k-uniformly Lipschitzian if there exists $k > 0$ such that

$$d(T^i x, T^i y) \leq k d(x, y)$$

for every $i \in \mathbb{N}$ and every x, y in X. An 1-uniformly Lipschitzian maps is also called nonexpansive. For such class of mappings we recall the most important result

Theorem 1 [**5**] *Let (X, d) be a complete bounded metric space. Assume that X has uniform normal structure, then any nonexpansive mapping defined on X has a fixed point.*

In [4], Khamsi gived the definition and a caracterization of 1-local retract subset of a metric space. Now we define

Definition 4 *A subset A is said to be k-local retract if for any family $(B_i)_i$ of closed ball centred in A such that $\cap_{i\in I} B(x_i, r_i) \neq \emptyset$, we have $A \cap_{i\in I} B(x_i, kr_i) \neq \emptyset$.*

It is imediate that uniform normal structure is not hereditary. However for 1-local retract subset we have

Lemma 1 *Let (X, d) be a metric space. Suppose that $\mathcal{A}(X)$ is an uniform normal convexity structure with constant $\beta < 1$. If Y is a 1-local retract subset of X, then $\mathcal{A}(Y)$ is an uniform normal structure with the same constant β.*

The proof is based on the next lemma

Lemma 2 [4] *Let (X, d) be a metric space and A a nonempty bounded subset of X. Then*
1) $cov(A) = \cap_{x \in X} B(x, r_x(A))$.
2) $r_x(A) = r_x(covA)$ for every x in X.
3) $\delta(A) = \delta(covA)$.
4) $r(covA) \leq r(A)$.
5) If (X, d) has the (n, ∞) property and is convex then $\frac{\delta(A)}{2} \leq r(covA) \leq \frac{n-1}{n}\delta(A)$.

Recall that (X, d) is said to have the (n, ∞) property if for any family $(B_i)_{i \in I}$ of closed ball of X such that $\cap_{i \in J} B_i \neq \emptyset$ for any finite subfamily J of I with card(J) less than n, we have $\cap_{i \in I} B_i \neq \emptyset$.
A metric space (X, d) is said to be convex if for all x, y in X and $\alpha \in [0, 1]$ there exists $z \in X$ suh that

$$d(z, x) = \alpha d(x, y) \text{ and } d(z, y) = (1 - \alpha) d(x, y).$$

Proof of lemma 1.
We assume that A is not reduce to a singleton. By 4) of lemma 2, we have $r(covA) \leq r(A)$. Since $A \in \mathcal{A}(Y)$ then $A = \cap_{i \in I} B(x_i, r_i) \cap Y$ with $x_i \in Y$. Hence $covA \subset \cap_{i \in I} B(x_i, r_i)$. Let $z \in cov(A)$ and define $r = r_z(A)$, then $z \in B = \cap_{x \in A} B(x, r) \cap \cap_{i \in I} B(x_i, r_i)$ is in $\mathcal{A}(X)$. Since Y is a 1-local retract of X then $B \cap Y \neq \emptyset$. Let $w \in B \cap Y$, so $w \in A = \cap_{i \in I} B(x_i, r_i) \cap Y$ and $w \in \cap_{x \in A} B(x, r)$. We deduce that $r_w(A) \leq r$. Hence $r(A) \leq r = r_z(A)$.
Since z is arbitrary in $cov(A)$ we obtain from (2) that $r(A) \leq r(cov(A))$. But $cov(A) \in \mathcal{A}(X)$ which is uniform normal, then

$$r(A) \leq r(cov(A)) \leq \beta\delta(cov(A)) = \beta\delta(A)$$

from property 4) of lemma 2. □

3 Fixed Point for k-Uniformly Lipschitzian Mappings

In the next theorem we obtain fixed point theorem for k-uniformly Lipschitzian mapping by utilizing the existence theorem of nonexpansive mapping [4]. In our knowledge This connection was never utilized. Moreover theorem 2 contains the result of theorem 1.

Theorem 2 *Let (X, d) be a complete bounded metric space. Assume that X has an uniform normal structure with constant $\beta < 1$. Then any k-uniformly Lipschitzian mapping $T : X \to X$ has a fixed point if $k^2\beta < 1$.*

Proof.
We need the followings two lemmas

Lemma 3 *Under the sames hypothesis of theorem 2 and for $T : X \to X$ k-uniformly Lipschitzian let*

$$d'(x, y) = \sup_{i=0,1,...} d(T^i x, T^i y).$$

Then
1) (X, d') is a bounded complete metric space.
2) T is d'-nonexpansive, ie

$$d'(Tx, Ty) \leq d'(x, y) \quad \forall x, y \in X.$$

Lemma 4 *Under the sames hypothesis of theorem 2 and for $T : X \to X$ k-uniformly Lipschitzian, the family of all admissibles subset of (X, d') is an uniform normale convexity structure with constant c ($c \leq k^2\beta$).*

Proof of lemma 3.
1-1) d' is a metric on X. Indeed
1-1-a) For every x, y in X we have
$d'(x, y) = 0$ is equivalent to $d(T^i x, T^i y) = 0$ for every $i = 0, 1, 2, ...$
Specifically for $i = 0$, it imply that $d(x, y) = 0$. Since d is a metric on X, then $x = y$.
1-1-b) For every $i = 0, 1, 2, ...$, and every x, y, z in X, we have

$$d(T^i x, T^i y) \leq d(T^i x, T^i z) + d(T^i z, T^i y),$$

since d is a metric on X.
By passing to the supremum on $i \in \mathbb{N}$, we obtain that

$$d'(x, y) \leq d'(x, z) + d'(z, y).$$

1-1-c) It is imediate that $d'(x, y) = d'(y, x)$ for all x, y in X.

1-2) Since T is k-uniformly Lipschitzian on X, and by definition of d', we have the inequality

$$d(x,y) \leq d'(x,y) \leq kd(x,y)$$

for all x, y in X. It follows from this inequality that (X, d') is a bounded complete metric space since (X, d) is it.

2) For every x, y in X we have

$$\begin{aligned} d'(Tx,Ty) &= \sup\{d(T^{i+1}x, T^{i+1}y)/i = 0, 1, 2, ...\} \\ &\leq \sup\{d(T^{i}x, T^{i}y)/i = 0, 1, 2, ...\} \\ &= d'(x,y). \qquad \square \end{aligned}$$

Proof of lemma 4.

Let A be an admissible subset for d', then

$$A = \cap_{x\in X} B'(x, r'_x(A)) \subset cov(A) = \cap_{x\in X} B(x, r_x(A)).$$

On the other hand it follows from the definition of d' that

$$d(z,y) \leq d'(z,y) \leq kd(z,y) \ \ \forall z, y \in X.$$

Hence

$$r'_z(A) \leq kr_z(A) \ \ \forall z \in X. \tag{5}$$

By passing in (5) to the infimum on $z \in \cap_{x\in X} B'(x, r'_x(A))$ we get

$$\inf_{z\in \cap_{x\in X} B'(x, r'_x(A))} r'_z(A) \leq k \inf_{z\in \cap_{x\in X} B'(x, r'_x(A))} r_z(A)$$

Which implies that

$$\begin{aligned} r'(A) = \{\inf r'_z(A)/z \in A = \cap_{x\in X} B'(x, r'_x(A))\} &\leq k \ \ \{\inf r_z(A)/z \in \cap_{x\in X} B'(x, r'_x(A))\} \\ &\leq k \ \inf\{\sup_{x\in A} d(z,x)/d(z,x) \leq \frac{r_x(A)}{k}\} \\ &\leq k \ \ \inf\{k \sup_{x\in A} d(z,x)/d(z,x) \leq r_x(A)\} \end{aligned}$$

since

$$\cap_{x\in X} B(x, \frac{r_x(A)}{k}) \subset \cap_{x\in X} B'(x, r'_x(A)).$$

Therefore

$$r'(A) \leq k^2 r(cov(A)) \leq k^2\beta\delta(cov(A)) = k^2\beta\delta(A) \leq k^2\beta\delta'(A). \qquad \square$$

Proof of theorem 2.
It follows imediatelly from theorem 1, property 2) of lemma 3, and lemma 4.

By theorem 2, we have $Fix(T) \neq \emptyset$ for every k-uniformly Lipschitzian mapping T defined on a bounded complete metric space (X, d) with uniform normal convexity structure $\mathcal{F}$ with constant $\beta < \frac{1}{k^2}$. More than $Fix(T)$ is k-local retract of X, ie for every closed balls $B(x_i, r_i)_{i\in I}$ we have

$$\cap_{i\in I} B(x_i, r_i) \neq \emptyset \text{ implies } \cap_{i\in I} B(x_i, kr_i) \cap Fix(T) \neq \emptyset.$$

Now we are able to show that

Theorem 3 *Let $T_n : X \to X$; $n = 0, 1, 2, ...$ be a family of commutings k-uniformly Lipsitzian mappings. Suppose that X has an uniform normal convexity structure $\mathcal{F}$ with constant $\beta < \frac{1}{k^2}$. Then $\cap_{n\in N} Fix(T_n) \neq \emptyset$ and is k-local retract of X.*

Proof of theorem 3.
The first part of the theorem follows imediatelly from theorem 2. For the second part, let $(B_i)_{i\in I}$ be a family of closed balls centred in $\cap_{n\in N} Fix(T_n)$ such that $(B_i)_{i\in I} \neq \emptyset$.
We have

$$B_d(x_i, r_i) \subset B_{d'}(x_i, kr_i) \subset B_d(x_i, kr_i)$$

Hence

$$\cap_{i\in I} B_{d'}(x_i, kr_i) \neq \emptyset$$

and since $(T_n)_n$ are nonexpansives mappings on (X, d'), it follows from theorem 2 that

$$\cap_{n\in N} Fix(T_n) \cap \cap_{i\in I} B_{d'}(x_i, kr_i) \neq \emptyset$$

which implies that

$$\emptyset \neq \cap_{n\in N} Fix(T_n) \cap \cap_{i\in I} B_{d'}(x_i, kr_i) \neq \emptyset \subset \cap_{n\in N} Fix(T_n) \cap \cap_{i\in I} B_d(x_i, kr_i) \neq \emptyset \qquad \square.$$

The problem whether the conclusion of theorem 3 holds for any commuting family $(T_i)_{i\in I}$ of k-uniformly Lipschitzians mappings ($k > 1$) was still open for several years. However by using the result of T.C.Lim and J.Kulesza [11] in which they shows that the weak compactness and weak countably compactness are equivalent if the metric space have normal structure, we prove that

Theorem 4 *Let (X, d) be a bounded complete metric space with an uniform normal convexity structure ($\beta < 1$). Then any commuting family $T_i : X \to X,\ i \in I$ of k-uniformly Lipschitzians mappings has a common fixed point provided that $k^2\beta < 1$.*

Proof.
Since (X, d') has uniform normal structure with constant c ($c < k^2\beta$), then by the well known theorem of Khamsi [5], $\mathcal{A}(X_{d'})$ is coutably compact.
Hence by the T. C. Lim and J. Kulesza result, it follows that $\mathcal{A}(X_{d'})$ is in fact compact. On the others hand , since each $T_i, i \in I$ is d'-nonexpansive (lemma 4), it follows that the result of theorem 3 is a direct consequence of Khamsi theorem's in which he shows that any commutings family of non-expansive mappings defined on a bounded metric space for which $\mathcal{A}(X_{d'})$ is compact and normal, has a common fixed point. $\square$

Remark that the result of theorem 4, was deduced from T. C. Lim and J. Kulesza theorem's and the uniform convexity of (X, d') (lemma 4); but the problem whether the compacity and normality of (X, d) imply the compacity and normality of (X, d'), still open.

4 Applications

It was proved by Nachbin [12] and Kelley [13] that all Banach spaces which have the $(2, \infty)$ property are those of form $C(E)$ where E is a compact Stonian, for example l_∞ and L_∞. Then by theorem 4 and property 5) of lemma 2, we have

corolary 1 *The unit balls of l_∞, L_∞, and $C(E)$ where E is a compact Stonian have the common fixed point property for every commutings family $T_i : X \to X,\ i \in I$ of k-uniformly Lipschitzians mappings provided that $k < \sqrt{2}$.*

After that, Lindenstrauss [14] proved that l_1 has a $(3, \infty)$ property. Therefore

corolary 2 *The unit ball of l_1 has the common fixed point property for every commutings family $T_i : X \to X$, $i \in I$ of k-uniformly Lipschitzians mappings provided that $k < \sqrt{\frac{3}{2}}$.*

Also, we deduce from theorem 4 and property 5) of lemma 2, the following corolary

corolary 3 *If (X, d) is a Banach space with the (n, ∞) property, and if $k < \sqrt{\frac{n}{n-1}}$, then its unit ball has the common fixed point property for every commutings family $T_i : X \to X$, $i \in I$ of k-uniformly Lipschitzian mappings.*

More recently, Prus [15] proved that all Banach spaces L_p $(1 < p < +\infty)$ have uniform normal structure with constant $\beta = min(2^{\frac{1}{p}}, 2^{\frac{1}{q}})$, where $q = p(p-1)^{-1}$ is the conjugat of p.
Hence, it follows that

corolary 4 *The unit balls of L_p has the common fixed point property for every commutings family $T_i : X \to X$, $i \in I$ of k-uniformly Lipschitzians mappings provided that*
$k < \sqrt{min(2^{\frac{1}{p}}, 2^{\frac{1}{q}})}$.

Now we recall the definition of the most geometrical caracterization of l_∞, L_∞, and $C(E)$ where E is a compact Stonian

Definition 5 *A metric space (X, d) is said to be hyperconvex iff for any family*
$\{B(x_i, r_i), i \in I\}$ of closed ball of (X, d) such that

$$d(x_i, x_j) \leq r_i + r_j$$

for every $i, j \in I$, has a nonempty intersection.

Remark.
1) Every hyperconvex metric space is complete, and if A is an admissible subset of (X, d) then also (A, d) is a hyperconvex metric space (see [8]).
2) Every hyperconvex space is convex. Indeed:
For all x, y in X and for any $\alpha \in [0, 1]$, let u, v in X. We have

$$\alpha[d(x, u) + d(x, v)] + (1 - \alpha)[d(y, u) + d(y, v)] \geq d(u, v).$$

The hyperconvexity of (X,d) imply that

$$\cap_{u\in X} B(u, \alpha d(x,u) + (1-\alpha) d(y,u)) \neq \emptyset.$$

Hence for every x, y in X and for every $\alpha \in [0,1]$, there exists $z \in X$ such that

$$z \in \cap_{u\in X} B(u, \alpha d(x,u) + (1-\alpha) d(y,u)),$$

ie

$$d(u,z) \leq \alpha d(x,u) + (1-\alpha) d(y,u)$$

for all $u \in X$. Therefore

$$d(x,z) = (1-\alpha) d(x,y) \quad \text{and} \quad d(y,z) = \alpha d(x,y). \qquad \square$$

Also by theorem 4 and proprerty 5) of lemma 2, we obtain the following theorem

Theorem 5 *Let (X,d) be a bounded hyperconvex metric space. Then any family of commutings k-uniformly Lipschitzian mappings defined on X has a common fixed point if $k < \sqrt{2}$.*

Proof.
(X,d) is a bounded hyperconvex metric space, then from the above remark, it is complete. Lets prove that (X,d) has the $(2,\infty)$ property. Indeed:
Let $\{B(x_i, r_i), i \in I\}$ be a family of closed balls of (X,d), such that

$$B(x_i, r_i) \cap B(x_j, r_j) \neq \emptyset \;\; \forall i, j \in I (i \neq j).$$

Then we have

$$d(x_i, x_j) \leq d(x_i, x) + d(x_j, x) \leq r_i + r_j$$

where $x \in B(x_i, r_i) \cap B(x_j, r_j)$.
Hyperconvexity of (X,d) imply that $\cap_{i\in I} B(x_i, r_i) \neq \emptyset$. (X,d) is then a convex metric space with the $(2,\infty)$ property. Therefore, by property 5) of lemma 2, $\mathcal{A}(X)$ is an uniform convexity structure with constant $\beta = \frac{1}{2}$. Hence, theorem 4 complete the proof of theorem 5. $\square$

References

[1] Kakutani, S. : Two fixed point theorems concerning bicompact convex sets, Proc. Imp . Acad. Tokyo 14 (1938), 242-245.

[2] Ryll-Nardzewski, C. : On fixed points of semi-groups of endomorphisms of linear spaces, Proc. Fifth Berkeley Symp. On Statistics and Probability II, pt. 1, (1966).

[3] Namioka, I. : A geometric proof of Ryll-Nardzewski's fixed point theorem, Bull. Ame. Math. Soc. 73 (1967) 443-445.

[4] Khamsi, M.A. : A 1-local retract and common fixed point for a commuting family of nonexpansives mappings.

[5] Khamsi, M.A. : On metric spaces with uniform normal structure, Proc. Ame. Math. Soc., volume 106, Number 3, (1989) page 4.

[6] El Amrani, M. : Etude de la propriete du point fixe pour les applications k-uniformement Lipschitziennes, These d'Etat , Universite Libre de Bruxelles, (1994).

[7] Khamsi M.A. and Misane D. , compactness of convexity structures in metric spaces, Math.Japonica 41 n 2 (1995) 321-326.

[8] Goebel K. and Kirk W.A. , topics in metric fixed point theory, Cambridge University Press (1990).

[9] Goebel K. and Kirk W.A. and Thele R.L. , uniformly Lipschitzian families of transformations in Banach spaces, Can.J.Math. n 26 (1974) 1245-1256.

[10] Goebel K. and Kirk W.A. , a fixed point theorems for transformations whose iterate have uniform Lipschitz constant, Studia Math. n 47 (1973) 137-140.

[11] Kulesza J. and Lim T. C. , on weak compactness and countable compactness in fixed point theory, Pro. Math. Ame. Soc., vol 124, n 11 (1996), 3345-3349.

[12] Nachbin L. , a theorem of Hann-Banach type, Trans. A.M.S. n 68 (1950) 28-48.

[13] Kelley J.L. , Banach spaces with the extension property, Trans. A.M.S. n 72 (1952) 323-326.

[14] Lindenstrauss J. , extension of compact operatord memoir of America. Math .Soc. n 48 (1966).

[15] Prus B. , on Bymun's fixed point theorem, Atti semin. mat.fis.univ Modena n 38 (1990) 535-545.

Information for Authors cont.

Copyright. Authors will be asked, upon acceptance of an article, to transfer copyright of the article to the publisher. This will ensure the widest possible dissemination of information under copyright laws. Preprints are available via web page at http://rattler.cameron.edu/SWJPAM. Authorization to copy is granted provided that the SWJPAM copyright symbol is not removed.

Tables. Tables must be numbered consecutively with roman numerals and have titles. They should be prepared in such a manner that no break is necessary.

Figures. All graphs and diagrams should be referred to as 'figure' (not abbreviated) in the text. A detailed caption should be provided for each figure, and figures should be numbered consecutively.

Footnotes. The use of footnotes should be minimized.

Notification. The author will be notified of the posting of his/her work and the time of release for SWJPAM.

References. References should be listed at the end of each paper. They should be numbered consecutively using arabic numbers between square brackets in the text. In the list of references, the numbers should appear without square brackets. A journal reference should comprise of name(s) and initial(s) of author(s), title of article, abbreviated name of journal (following the style of 'Abbreviations of Names of Serials reviewed in Mathematical Reviews'), volume number, year of publication (within parentheses), and full page numbers, e.g.

1. Argyros, I.K. and Szidarovszky, F.: On the Monotone Convergence of Algorithmic Models, Appl. Math. Com. 48, (1992), 167–176.

A book reference should comprise of name(s) and initial(s) of author(s), title of book, publisher, town, year of publication, e.g.

11. Ortega, J.M. and Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970

An edited volume reference should comprise of name(s) and initial(s) of author(s), title of paper, name(s) of editor(s), title of book, publisher, town, year, inclusive page numbers, e.g.

12. Prenter, P.: On Polynomial Operators and Equations, L.B. Rall (ed), Computational Solution of nonlinear operator equations, Johy Wiley Publisher, New York, 1971, pp. 361–398.

Indexing and Abstracting services. Articles published in SWJPAM are selectively indexed or abstracted in: Mathematical Reviews of A.M.S., Zentralblatt für Mathematik, European Mathematical Society (EMIS).