MARCH / MARS 2001

IN THIS ISSUE / DANS CE NUMÉRO

1 Dipendra Prasad and C. S. Yogananda
Bounding the torsion in CM elliptic curves

6 R. Balasubramanian and D. S. Ramana
Additive complements of the squares

12 B. Sury
Values of Euler polynomials

16 Jacek Mrowiec
Remark on approximately Jensen-convex functions

22 Sergio Amat and Sonia Busquier
A remark on inhomogeneous Cauchy problems

28 Tauno Metsänkylä
Catalan's equation with a quadratic exponent

23
No 1

C. R. Math. Rep. Acad. Sci. Canada Vol. **23** (1), 2001 pp. 1–5

BOUNDING THE TORSION IN CM ELLIPTIC CURVES

DIPENDRA PRASAD AND C. S. YOGANANDA

Presented by M. Ram Murty, FRSC

RÉSUMÉ. Nous donnons une borne supérieure pour l'ordre du groupe de torsion d'un courbe elliptique à multiplication complexe defini sur un corps de nombres quelconque. Ce résultat était établi par A. Silverberg en utilisant le théorème principale de CM. Nous présentons une démonstration plus simple qui utilise seulement le théorème de Deuring concernant les nombres premiers supersinguliers et la théorie algébraique des nombres.

1. Introduction. In [6] A. Silverberg, using the main theorem of complex multiplication of Shimura and Taniyama, has obtained a bound for the order of a point of finite order in a CM abelian variety over a number field in terms of only the degree of the number field and the dimension of the abelian variety. As a corollary she obtains the following result for elliptic curves: *Let E be an elliptic curve over a number field K of degree d with CM by an order $\mathcal{O}$ in an imaginary quadratic field k. Suppose $P \in E(K)$ is a point of order N. Then $\varphi(N) \leq \delta\mu d$ where μ is the number of roots of unity in $\mathcal{O}$ and $\delta = 1/2$ or 1 depending on whether k is contained in K or not.* Her results also imply a bound on the full torsion subgroup of CM elliptic curves.

The aim of this paper is to give an estimate on the order of the torsion subgroup of a CM elliptic curve over a number field using only the result of Deuring about supersingular primes, and elementary algebraic number theory. To state our theorem, we need the following notation: if $M = l_1^{j_1} \cdots l_r^{j_r}$ is the prime factorisation of M let $M' = l_1^{(j_1+\delta_1)/2} \cdots l_r^{(j_r+\delta_r)/2}$ where $\delta_i = 0$ if j_i is even and $\delta_i = 1$ if j_i is odd.

THEOREM 1.1. *Let E be an elliptic curve over a number field K of degree d with CM by an order in an imaginary quadratic field k. Then if M is the order of the torsion subgroup of $E(K)$ we have:*

1. *$\varphi(M) \leq 2d$ if $K \cap k = \mathbf{Q}$;*
2. *$\varphi(M') \leq 2d$ if $k \subseteq K$ but $k \neq \mathbf{Q}(i), \mathbf{Q}(\omega)$, ($\omega$ being a cubic root of unity).*

When $k = \mathbf{Q}(i), \mathbf{Q}(\omega)$ our method gives an extra factor of $2^{d(M)+1}$ where $d(M)$ is the number of distinct prime divisors of M.

Received by the editors January 21, 2000.
AMS subject classification: Primary: 11G05; secondary: 11G15, 14K22.

REMARK 1. A. Silverberg has remarked that one can easily give an estimate to the order of the full torsion subgroup of a CM elliptic curve which is the same as our case (i) from her theorem in [9]. Her estimate is better than ours in case (ii). In case (i), this follows as the torsion subgroup on an elliptic curve is always of the form $\mathbf{Z}/a \times \mathbf{Z}/b$ where $a|b$. If the torsion subgroup of $E(K)$ contains $\mathbf{Z}/r \times \mathbf{Z}/r$ for $r > 2$, it follows from [10] that the field of complex multiplication $(= k)$ must be contained in K. Therefore if $K \cap k = \mathbf{Q}$, then the torsion subgroup is either $\mathbf{Z}/N$ or $\mathbf{Z}/N \times \mathbf{Z}/2$, where N is the maximum order of a torsion point, achieving the same bound that we obtain for the order of the torsion subgroup of $E(K)$ from the maximal order of a torsion element.

There is a large amount of literature on the torsion subgroup of elliptic curves over number fields. In [4] Merel has shown that the order of the torsion subgroup of an elliptic curve over a number field K can be bounded in terms of only the degree, d, of K over $\mathbf{Q}$. The bound thus obtained (first by Merel and then improved by Oesterlé) is exponential in d. The initial motivation for this note was to investigate as to what could be the 'right bound' by looking at the CM case when we discovered that Silverberg has already done this. We also refer to the paper of Olson [8] which deals with elliptic curves with complex multiplication.

ACKNOWLEDGEMENT. We are grateful to Prof. J. Oesterlé whose wonderful lectures on Merel's work and his own refinement at the Mehta Research Institute when he was visiting under the Indo-French programme, and at the ICTP, Trieste, stimulated our interest in working out the CM case. We are grateful to Prof. J. Alperin for providing us with the proof of Lemma 2.3 and to Prof. A. Silverberg for Remark 1. This work was done while the second author was visiting Mehta Research Institute whose hospitality he gratefully acknowledges.

2. Preliminary Lemmas.

LEMMA 2.1. *Let k be a quadratic extension of $\mathbf{Q}$ and K an extension of $\mathbf{Q}$ of degree d with $K \cap k = \mathbf{Q}$. Then the set of primes p in $\mathbf{Q}$ which remain inert in k and have the property that there is at least one prime of degree 1 in K above p is of density at least $1/(2d)$.*

PROOF. Let L be a Galois extension of $\mathbf{Q}$ containing K and k, and let $G = \mathrm{Gal}(L/\mathbf{Q})$. Further let H_K and H_k be the subgroups of G corresponding to the subfields K and k, respectively, of L. It is easy to see that the set of prime ideals $\mathfrak{p}$ in L, such that the prime ideal $\mathfrak{p} \cap K$ is of degree 1 are precisely those for which the corresponding Frobenius element σ in G belongs to H_K. The prime $p = \mathfrak{p} \cap \mathbf{Q}$ is inert in k if and only if σ does not belong to H_k. So, the primes p in $\mathbf{Q}$ as desired in the lemma are precisely those for which there is a prime $\mathfrak{p}$ in L above p for which the Frobenius element belongs to $(G \setminus H_k) \cap H_K$. Since the cardinality of $(G \setminus H_k) \cap H_K$ is $|H_K|/2$, it follows from the Cebotarev density theorem that the density of p in $\mathbf{Q}$ as desired in the lemma is at least $1/(2d)$. ∎

LEMMA 2.2. *Let K be a number field of degree d containing an imaginary quadratic field k. Then the set of primes p in $\mathbf{Q}$ which are inert in k and have a prime of degree 2 in K over p is of density at least $1/d$.*

PROOF. Let L be a Galois extension of $\mathbf{Q}$ containing K and $G = \mathrm{Gal}(L/\mathbf{Q})$. Further let H_K and H_k be the subgroups of G corresponding to the subfields K and k, respectively, of L. The set of primes p as desired in the lemma are precisely those for which the corresponding Frobenius substitution σ does not belong to H_k but whose square is in H_K. Since k is imaginary quadratic, the complex conjugation does not belong to H_k. The following lemma combined with the Cebotarev density theorem completes the proof of our lemma. ∎

LEMMA 2.3. *Let G be a finite group, N a subgroup of G of index 2 in G and H a subgroup of N. Suppose that there is an element of order 2, say c, in G which is not in N. Then the set of elements in G which do not belong to N but whose square belongs to H has cardinality at least that of H.*

PROOF (DUE TO J. ALPERIN). We need to count elements $n \cdot c$ with $n \in N$ whose square belongs to H. Clearly HcH is a subset of $NcN = Nc$. We will prove that there are exactly $|H|$ elements in HcH whose square belongs to H which will prove our lemma. To prove this let $A = H \cap cHc^{-1}$, and let $X \subset H$ be a set of left coset representatives of A in H so that every element of H can be written uniquely in the form $x \cdot a$ with $x \in X$ and $a \in A$. From this it is easy to see that an element of HcH can be uniquely written in the form xch with $x \in X$, $h \in H$. Now $(xch)^2 = xchxch$ belongs to H if and only if $chxc$ belongs to H which happens if and only if hx belongs to cHc^{-1}. Since both x and h belongs to H, we find that $(xch)^2$ belongs to H if and only if hx belongs to $A = H \cap cHc^{-1}$. For each x, this means that h belongs to Ax^{-1}. So for each x, there are $|A|$ many choices for h such that $(xch)^2$ belongs to H. Therefore the total number of elements in HcH whose square belongs to H is

$$|A| \cdot \frac{|H|}{|A|} = |H|,$$

proving the lemma. ∎

3. Proof of the main theorem. The proof of our main theorem will be a simple consequence of the lemmas in the previous section, Cebotarev density theorem, and the well known theorem about elliptic curves with complex multiplication that a prime $\mathfrak{p}$ in K which is a prime of good reduction for E over K is a prime of supersingular reduction if and only if $\mathfrak{p} \cap \mathbf{Q} = p$ is inert or ramified in k (see [3]).

CASE 1. $K \cap k = \mathbf{Q}$. We consider the set of rational primes p coprime to M which are inert in k and have a split factor in K, *i.e.*, there is a prime of degree 1, say $\mathfrak{p}$ in K which divides p. Denote by $\mathbf{F}_\mathfrak{p}$ the residue field associated to the prime

ideal $\mathfrak{p}$ of K. Then the torsion subgroup of E over K will inject into $E_{\mathfrak{p}}(\mathbf{F}_{\mathfrak{p}})$ which has cardinality $p+1$ since $\mathfrak{p}$ is a prime of supersingular reduction of E and $\mathbf{F}_{\mathfrak{p}}$ is a finite field with p elements; hence $M|(p+1)$. By Lemma 2.1 the density of such primes p is at least $1/(2d)$ whereas by Cebotarev density theorem, the density of primes p which are congruent to -1 modulo M is $1/\varphi(M)$. Therefore we must have $1/(2d) \leq 1/\varphi(M)$ and so $\varphi(M) \leq 2d$.

CASE 2. $k \subset K$. We consider primes p in $\mathbf{Q}$ such that p is inert in k and has a prime factor $\mathfrak{p}$ of degree 2 in K which is a prime of good reduction of E. The elliptic curve will have supersingular reduction at such primes. The density of such primes is $1/d$ from Lemma 2.2. Since $\mathbf{F}_{\mathfrak{p}}$ is a finite field of order p^2, $|E_{\mathfrak{p}}(\mathbf{F}_{\mathfrak{p}})| = 1 + p^2 + a_p$ where $a_p = 0, \pm p, \pm 2p$. The possibilities $a_p = 0, \pm p$ arises only for $k = \mathbf{Q}(i)$ and $k = \mathbf{Q}(\omega)$ respectively which we have omitted. Therefore we find that $|E_{\mathfrak{p}}(\mathbf{F}_{\mathfrak{p}})|$ is either $(p+1)^2$ or $(p-1)^2$. Hence p is congruent to either 1 or -1 modulo M'. The density of such primes p is $2/\varphi(M')$, and as before we get

$$\frac{1}{d} \leq \frac{2}{\varphi(M')},$$

or $\varphi(M') \leq 2d$ completing the proof of the theorem.

REMARK 2. The technique of using supersingular primes can be used in some situations to get better bounds for the order of the torsion subgroup when K, the field of definition of E, is a non-normal extension. For instance, if L is a Galois extension of $\mathbf{Q}$ with Galois group $\mathrm{GL}_2(\mathbf{F}_q)$ which is disjoint from the field of CM, k, and K is the fixed field of the diagonal torus, then $[K : \mathbf{Q}] = q(q+1)$. The set of primes p in $\mathbf{Q}$ with the property that there is a prime $\mathfrak{p}$ in K of degree 1 above p corresponds to those Frobenius substitutions in $\mathrm{GL}_2(\mathbf{F}_q)$ which have a conjugate which is diagonalisable over $\mathbf{F}_q$. The set of elements in $\mathrm{GL}_2(\mathbf{F}_q)$ which are diagonalisable over $\mathbf{F}_q$ can be easily seen to be of cardinality $[(q-2)(q-1)q(q+1)]/2 + (q-1)$. Therefore the set of primes p which are inert in k and have a prime in K of degree 1 above p is roughly of density $1/4$. By the arguments in Theorem 1.1 this implies that the torsion in $E(K)$ is bounded by M with $\varphi(M) \leq 4$, implying that the set of possible values of M is $\{1, 2, 3, 4, 5, 6, 8\}$, instead of the much larger bound depending on q coming from Theorem 1.1.

4. A conjecture about torsion. In this section we make a few general remarks and state a conjecture on the bound for the order of a torsion point of an elliptic curve defined over a number field.

Let X be a curve over $\mathbf{Q}$ of genus ≥ 2. Let d be the gonality of X, *i.e.*, d is the minimal integer among the degrees of maps from X to $\mathbf{P}^1$ and assume that d is realised for a map π defined over $\mathbf{Q}$. It is clear that any element x of $X(\overline{\mathbf{Q}})$ with $\pi(x) \in \mathbf{P}^1(\mathbf{Q})$ is defined over a number field of degree $\leq d$ and therefore there are infinitely many points in $X(\overline{\mathbf{Q}})$ defined over number fields of degree $\leq d$. Conversely, it has been proved by Debarre and Klassen in [2] that if X is

a smooth *plane* curve then d is the maximal integer with the property that the number of points in $X(\overline{\mathbf{Q}})$ defined over a number field of degree $< d$ is finite. We would like to believe that a suitably modified version of their theorem is valid also for modular curves $X_0(N)$ and $X_1(N)$. More precisely, we believe that these modular curves have *no* points, except for cusps, defined over an extension of $\mathbf{Q}$ of degree $\leq Bd$, B a constant independent of N.

It has been proved in [1] that the gonality of $X_0(N)$ (resp. $X_1(N)$) is at least a constant times the degree of the standard map of $X_0(N)$ to $\mathbf{P}^1$; in fact, $d_0 \geq (7/800)N$ (a similar bound but quadratic in N for $X_1(N)$). Earlier heuristics therefore lead us to the following.

CONJECTURE. Let E be an elliptic curve defined over a number field K with a torsion point of order N in $E(K)$. Then there is a constant C independent of E and K such that $\varphi(N) \leq C[K : \mathbf{Q}]$.

REFERENCES

1. D. Abramovich, *A linear lower bound on the gonality of modular curves.* Internat. Math. Res. Notices **20**(1996), 1005–1011.
2. O. Debarre and M. J. Klassen, *Points of low degree on smooth plane curves.* J. Reine Angew. Math. **446**(1994), 81–87.
3. S. Lang, *Elliptic Functions.* Addison-Wesley, 1973.
4. L. Merel, *Bornes pour la torsion des courbes elliptiques sur le corps de nombres.* Invent. Math. **124**(1994), 437–449.
5. L. Olson, *Points of finite order on elliptic curves with complex multiplication.* Manuscripta Math. **14**(1974), 195–205.
6. A. Silverberg, *Points of finite order on Abelian Varieties.* Contemp. Math. **133**(1992), 175–193.
7. ———, *Fields of definition for homomorphisms of abelian varieties.* J. Pure Appl. Algebra **77**(1992), 253–262.

Mehta Research Institute
Chhatnag Road
Jhusi, Allahabad-211019
India
email: dprasad@mri.ernet.in

Olympiad Cell (NBHM)
Dept. of Mathematics
Indian Institute of Science
Bangalore 560012
India
email: yoga@math.iisc.ernet.in

C. R. Math. Rep. Acad. Sci. Canada Vol. 23 (1), 2001 pp. 6–11

ADDITIVE COMPLEMENTS OF THE SQUARES

R. BALASUBRAMANIAN AND D. S. RAMANA

Presented by M. Ram Murty, FRSC

ABSTRACT. When $N \geq 1$ is an integer, let $B(N)$ denote a subset of $\{0, 1, \ldots, N\}$, of smallest cardinality, satisfying the following condition: every integer n satisfying $1 \leq n \leq N$ is expressible as $n = b + k^2$, for a b in $B(N)$ and an integer k. Let $b(N)$ be the cardinality of a $B(N)$ and let α denote $\liminf_{N\to\infty} b(N)/\sqrt{N}$. We prove that if, for a δ in the interval $(0, 1)$ and all large integers N, the interval $[0, \delta N]$ contains a $B(N)$, then $\alpha \geq 2/(\frac{\sqrt{1-\delta}}{1+\sqrt{\delta}} + \sin^{-1}\sqrt{\delta})$.

RÉSUMÉ. Pour un entier $N \geq 1$, soit $B(N)$ une partie de $\{0, 1, \ldots, N\}$ de plus petit cardinal possible vérifiant : tout entier n satisfaisant $1 \leq n \leq N$ s'exprime $n = b + k^2$, pour b dans $B(N)$ et un entier k. Soit $b(N)$ la cardinalité de $B(N)$ et soit α égal à $\liminf_{N\to\infty} b(N)/\sqrt{N}$. Nous démontrons que si, un δ dans l'interval $(0, 1)$ et tout N assez grand, l'interval $[0, \delta N]$ contient un ensemble $B(N)$, alors $\alpha \geq 2/(\frac{\sqrt{1-\delta}}{1+\sqrt{\delta}} + \sin^{-1}\sqrt{\delta})$.

1. Introduction. Let $N \geq 1$ be an integer. By a *minimal additive complement of the squares up to* N we mean a subset B of the integers $\{0, 1, \ldots, N\}$, of *smallest cardinality*, such that every integer n, $1 \leq n \leq N$, can be written as $n = b + k^2$, for some b in B and some integer k. For each integer $N \geq 1$, we let $b(N)$ denote the cardinality of a minimal additive complement of the squares up to N and finally, we write α to denote the liminf of the sequences $b(N)/\sqrt{N}$, as N tends to infinity.

From the trivial inequality $b(N)\sqrt{N} \geq N$, it follows that $\alpha \geq 1$. Prompted by a question of P. Erdős, a number of authors have worked on improving this lower bound for α. The best known result at present being (independently) due to L. Habsieger and J. Cilleruelo, who proved $\alpha \geq 4/\pi$ (see [1] for a more complete history).

Heuristical reasoning suggests that the elements of minimal additive complement must themselves be small. Therefore we may ask the following question. Can the lower bound $4/\pi$ for α be improved under the assumption that there exists a δ in $(0, 1)$ such that for all large integers N there is a minimal additive complement of the squares contained in the interval $[0, \delta N]$? In this note

Received by the editors February 28, 2000.
AMS subject classification: 11B13, 11B99.
Key words and phrases: additive complements.

we answer this question by proving the following theorem. This theorem is an improvement over a result obtained (by a different method) by Zhai ([2]).

THEOREM 1. *If, for a δ in the interval $(0,1)$ and all large integers N, there is a minimal additive complement of the squares up to N contained in the interval $[0,\delta N]$, then one has the following inequality.*

$$\alpha \geq \frac{2}{\frac{\sqrt{1-\delta}}{(1+\sqrt{\delta})} + \sin^{-1}(\sqrt{\delta})} \tag{1}$$

The right hand side of (1) is a continuous function of δ taking the value 2 when δ is 0 and the value $4/\pi$ when δ is 1. In particular, therefore, it follows from Theorem 1 and the easily verified inequality $2 \geq \alpha$ (see [2]), that if, for all large N, there exists a minimal additive complement of the squares up to N all of whose elements are $o(N)$, then $\alpha = 2$.

2. The basic lemma. Our method rests on the following lemma whose proof is immediate from the definition of additive complements.

LEMMA 1. *Let B be an additive complement of the squares up to N. Then the following inequality holds for any function $f(t)$ which is ≥ 0 for all $t \geq 0$.*

$$\sum_{b\in B} \sum_{0\leq k\leq \sqrt{N-b}} f(b+k^2) \geq \sum_{1\leq n\leq N} f(n). \tag{2}$$

For an integer $m \geq 1$ and t in $[0,1]$, we define $\phi_m(t)$ by the following relation

$$\phi_m(t) = \int_0^{1-t} \frac{(u+t)^m}{2\sqrt{u}}\, du \tag{3}$$

For an integer $m \geq 1$ and t in $[0,1]$, we write $g_m(t)$ to denote $-\phi_m'(t)$. With these notations we apply Lemma 1 to the functions $f_m(t) = t^m$, for integers $m \geq 1$, to deduce the following corollary.

COROLLARY. *Let B be an additive complement of the squares up to N. Let $B(N)$ be the cardinality of the set B. For any t in $[0,1]$, let $\beta(Nt)$ denote the number of b in B that are $< Nt$. For any integer $m \geq 1$, the following inequality holds.*

$$\frac{B(N)}{N} + \int_0^1 \frac{\beta(Nt)}{\sqrt{N}} g_m(t)\, dt \geq \frac{1}{m+1} \tag{4}$$

PROOF. For the functions $f_m(t)$ the right hand side of (2) is $\geq N^{m+1}/m+1$. By the monotonicity of $f_m(t)$, the summation over k on the left hand side of (2) is majorised by $N^m + \int_0^{\sqrt{N-b}} f_m(b+x^2)\, dx$. Inserting these remarks into (2) and writing the summation over b as an integral with respect to the measure $\beta(Nt)$, we obtain (5).

$$B(N)N^m + \int_0^1 \int_0^{\sqrt{N-Nt}} f_m(Nt+x^2)\, dx\, d\beta(Nt) \geq \frac{N^{m+1}}{m+1} \tag{5}$$

Making a change of variable $x^2 = Nu$ in the integral with respect to x on the left hand side of (5) and dividing throughout by N^{m+1} we obtain the following inequality.

$$\frac{B(N)}{N} + \frac{1}{\sqrt{N}} \int_0^1 \phi_m(t)\, d\beta(Nt) \geq \frac{1}{m+1} \tag{6}$$

The corollary now follows from (6) upon integrating by parts with respect to t and noting that $\beta(0) = 0$ and that $\phi_m(1) = 0$ for any integer $m \geq 1$.

In Section 4, we will prove Theorem 1 by exploiting the above corollary with the aid of certain properties of the functions $g_m(t)$ and $\phi_m(t)$ described in Section 3.

3. Some properties of $g_m(t)$ and $\phi_m(t)$.

PROPOSITION 1.

(i) For any integer $m \geq 1$, $g_m(t)$ has a unique zero x_m in the interval $(0,1)$. Further, $g_m(t)$ is < 0 on $[0, x_m)$ and is > 0 on $(x_m, 1]$.

(ii) The sequence x_m tends to 1 as m tends to infinity.

PROOF. Let m be an integer ≥ 1. Differentiating the expression (3) defining $\phi_m(t)$ we note that $g_m(t)$ is given by the following relation.

$$g_m(t) = \frac{1}{2\sqrt{1-t}} - \int_0^{1-t} \frac{m(u+t)^{m-1}}{2\sqrt{u}}\, du \tag{7}$$

From (7) it is immediate that $g_m(t) \to \infty$ as $t \to 1$ and that $g_m(0) = -1/(2m-1)$. Therefore $g_m(t)$ has at least one zero in $[0,1]$.

Now set $t = \cos^2\theta$ (with θ in the interval $[0, \pi/2]$) and define, for each integer $m \geq 1$, $h_m(\theta) = g_m(\cos^2\theta)\sec^{2m-1}\theta$. It follows from (7), with $u = \cos^2\theta \tan^2\psi$, that $h_m(\theta)$ is given by the following relation.

$$h_m(\theta) = \frac{\sec^{2m-1}\theta}{2\sin\theta} - m\int_0^\theta \sec^{2m}\psi\, d\psi \tag{8}$$

Differentiating the expression (8) for $h_m(\theta)$, we see that $h_m'(\theta) = -\sec^{2m}\theta/2\sin^2\theta$. This implies that $h_m(\theta)$ is strictly decreasing on $[0, \pi/2]$. Therefore, $g_m(t)$ has at most one root in $[0,1]$. These remarks verify (i). (It turns out that $g_m(t)$ itself is *not* monotonic for large values of m. This is why we have considered the functions $h_m(\theta)$ in order to prove (i).)

To verify (ii), we first observe the following inequality. To obtain (9) we note that $1 - t \geq u/(u+t)$, for t in $[0,1)$ and u in $[0, 1-t]$.

$$\begin{aligned} \int_0^{1-t} \frac{m(u+t)^{m-1}}{2\sqrt{u}}\, du &\geq \int_0^{1-t} \frac{m(u+t)^{m-1}}{2\sqrt{u}} \frac{\sqrt{u}}{\sqrt{(u+t)(1-t)}}\, du \\ &= \int_0^{1-t} \frac{m(u+t)^{m-\frac{3}{2}}}{2\sqrt{1-t}}\, du \end{aligned} \tag{9}$$

Combining (9) with the obvious identity

$$\int_0^{1-t} \frac{m(u+t)^{m-\frac{3}{2}}}{2\sqrt{1-t}}\,du = \frac{m(1-t^{\frac{2m-1}{2}})}{(2m-1)\sqrt{1-t}} \tag{10}$$

we obtain (11), for any t in $[0,1)$.

$$\int_0^{1-t} \frac{m(u+t)^{m-1}}{2\sqrt{u}}\,du \geq \frac{m(1-t^{\frac{2m-1}{2}})}{(2m-1)\sqrt{1-t}} \tag{11}$$

Since $g_m(x_m) = 0$, we obtain (12) from (7) and (11) applied with $t = x_m$.

$$\frac{1}{2\sqrt{1-x_m}} \geq \frac{m(1-x_m^{\frac{2m-1}{2}})}{(2m-1)\sqrt{1-x_m}} \tag{12}$$

Rearranging (12), we see that $x_m \geq 1/(2m)^{\frac{2}{2m-1}}$. Since $x_m < 1$, (ii) follows from this last inequality upon letting m tend to infinity.

PROPOSITION 2.

(i) $\lim_{m\to\infty}(m+1)\phi_m(t) = 1/2\sqrt{1-t}$ *for all* t *in* $[0,1)$.

(ii) When δ *is in* $[0,1)$, *the inequality* $|(m+1)\phi_m(t)| \leq 2/\sqrt{1-\delta}$ *holds for all* t *in* $[0,\delta]$ *and all* $m \geq M$, *where* M *is a real number dependent only on* δ.

(iii) When δ *is in* $[0,1)$ *we have the following relation.*

$$\lim_{m\to\infty}\int_0^{\delta}(1-\sqrt{t})(m+1)g_m(t)\,dt = \frac{1}{2} - \frac{1-\sqrt{\delta}}{2\sqrt{1-\delta}} - \int_0^{\delta}\frac{dt}{4\sqrt{t(1-t)}} \tag{13}$$

PROOF. For a t in $[0,1)$, let $\epsilon > 0$ be chosen such that $1 - t > \epsilon$. From definition of $\phi_m(t)$ given by (3) we have the following relation.

$$(m+1)\phi_m(t) = \int_0^{1-t-\epsilon}\frac{(m+1)(u+t)^m}{2\sqrt{u}}\,du + \int_{1-t-\epsilon}^{1-t}\frac{(m+1)(u+t)^m}{2\sqrt{u}}\,du \tag{14}$$

By the dominated convergence theorem, the first integral on the right hand side of (14) tends to 0 as m tends to infinity. The second integral on the right hand side of (14) is bounded above (resp. below) by $\left(1-(1-\epsilon)^{m+1}\right)/2\sqrt{1-t-\epsilon}$ (resp. $\left(1-(1-\epsilon)^{m+1}\right)/2\sqrt{1-t}$) for all $m \geq 1$. Now letting m tend to infinity for the chosen ϵ and finally noting that this ϵ can be chosen to be arbitarly small, we obtain (i).

When δ is in $[0,1)$, and $\epsilon > 0$ satisfies $1-\delta > \epsilon$, (14) and the preceding remarks also imply (15), for any t in $[0,\delta]$.

$$|(m+1)\phi_m(t)| \leq (m+1)(1-\epsilon)^m\int_0^{1-\delta-\epsilon}\frac{du}{2\sqrt{u}} + \frac{1}{2\sqrt{1-\delta-\epsilon}} \tag{15}$$

Choosing ϵ small enough so that $1/\sqrt{1-\delta-\epsilon} \leq 2/\sqrt{1-\delta}$ and choosing a real number M large enough so that, for this ϵ, the first term on the right hand side of (15) is $\leq 1/\sqrt{1-\delta}$ for all $m \geq M$, we obtain (ii) from (15).

To verify (iii), we recall that $g_m(t) = -\phi_m'(t)$ and integrate by parts to obtain (16). Note that $\phi_m(0) = 1/2m+1$.

$$\int_0^\delta (1-\sqrt{t})(m+1)g_m(t)\,dt = \frac{m+1}{2m+1} - (m+1)\phi_m(\delta)(1-\sqrt{\delta}) - \int_0^\delta \frac{(m+1)\phi_m(t)}{2\sqrt{t}}\,dt \tag{16}$$

(iii) now follows from (16) upon letting m tend to infinity and using (i), (ii) and the dominated convergence theorem.

4. Proof of Theorem 1. Choose m large enough so that $x_m > \delta$. This is possible by (ii) of Proposition 1, Section 3.

Let N_k be a sequence on integers such that $b(N_k)/\sqrt{N_k} \to \alpha$ as $k \to \infty$. For sufficiently large integers k, we will apply the corollary to Lemma 1, Section 2 to the integers N_k and *minimal* additive complements B_k of the squares up to N_k contained in the interval $[0, \delta N_k]$. For a t in $[0,1]$ and an integer k, sufficiently large, we write $\beta_k(N_k t)$ to denote the number of b in B_k that are $< N_k t$. From (4) we then have (17).

$$\frac{b(N_k)}{N_k} + \int_0^\delta \frac{\beta_k(N_k t)}{\sqrt{N_k t}}\sqrt{t}g_m(t)\,dt + \int_\delta^1 \frac{\beta_k(N_k t)}{\sqrt{N_k}}g_m(t)\,dt \geq \frac{1}{m+1} \tag{17}$$

The hypothesis on δ implies that $\beta_k(N_k t) = \beta_k(N_k)$ for t in $(\delta, 1]$. Therefore, the second term on the left hand side of (17) is $\beta_k(N_k)/\sqrt{N_k}\int_\delta^1 g_m(t)\,dt$. Taking limsup of both sides as k tends to infinity and noting that $\limsup_{k\to\infty} b(N_k)/N_k$ is 0, $\limsup_{k\to\infty} \beta_k(N_k)/\sqrt{N_k}$ is α, we have (18).

$$\limsup_{k\to\infty} \int_0^\delta \frac{\beta_k(N_k t)}{\sqrt{N_k t}}\sqrt{t}g_m(t)\,dt + \alpha\int_\delta^1 g_m(t)\,dt \geq \frac{1}{m+1} \tag{18}$$

By (i), Proposition 1, Section 3, $g_m(t)$ is < 0 on $[0, x_m)$ and hence on $[0,\delta]$. Applying Fatou's Lemma and recalling the definition of α, we see that the first term on the left hand side of (18) is $\leq \alpha\int_0^{x_m}\sqrt{t}g_m(t)\,dt$. Inserting this upper bound into (18) and rearranging the terms we obtain the following inequality.

$$\alpha\int_0^1 g_m(t)\,dt + \alpha\int_0^\delta (\sqrt{t}-1)g_m(t)\,dt \geq \frac{1}{m+1} \tag{19}$$

The first integral on the left hand side of (19) is easily seen to be $\phi_m(0)$. Theorem 1 now follows by multiplying both sides of (19) by $m+1$ and taking into account the preceding remark and Proposition 2, Section 3 while letting $m \to \infty$. Note that $1-\sqrt{\delta}/\sqrt{1-\delta} = \sqrt{1-\delta}/(1+\sqrt{\delta})$ and that $\int_0^\delta dt/2\sqrt{t(1-t)}$ is $\sin^{-1}\sqrt{\delta}$.

5. Concluding Remarks. The method of this note may be easily adapted to yield the results on the analogous question for higher powers. For the sake of completeness, however, we record below the statement of the generalisation of Theorem 1 of Section 1 for higher powers. Let p be a fixed integer ≥ 2.

When N is an integer ≥ 1, a *minimal additive complement of the p-th powers up to N* is a subset B of the integers $\{0, 1, \ldots, N\}$, of *smallest cardinality* such that every integer n satisfying $1 \leq n \leq N$ is expressible as $n = b + k^p$, for some b in B and some integer k. Let $b_p(N)$ denote the cardinality of a *minimal* additive complement of the p-th powers up to N and let $\alpha(p)$ denote the liminf of the sequence $b(N)/N^{1-\frac{1}{p}}$, as N tends to infinity.

THEOREM. *If, for a δ in the interval $(0,1)$ and all large integers N, there is a minimal additive complement of the p-th powers up to N contained in the interval $[0, \delta N]$, then one has the following inequality.*

$$\alpha(p) \geq \frac{p}{\frac{1-\delta^{1-\frac{1}{p}}}{(1-\delta)^{1-\frac{1}{p}}} + (1-\frac{1}{p})\int_0^\delta \frac{dt}{t^{1-\frac{1}{p}}(1-t)^{\frac{1}{p}}}} \tag{20}$$

ACKNOWLEDGEMENTS. We thank Dr. B. Ramakrishnan, M.R.I., Allahabad, for bringing [2] to our attention. We thank the referee for a careful reading of this paper.

REFERENCES

1. L. Habsieger, *On the Additive Completion of Polynomial Sets.* J. Number Theory **51**(1995), 130–135.
2. W. Zhai, *The Additive Completion of k-th Powers.* J. Number Theory **79**(1999), 292–300.

The Institute of Mathematical Sciences
C.I.T. Campus
Taramani
Chennai 600 113
India

C. R. Math. Rep. Acad. Sci. Canada Vol. 23 (1), 2001 pp. 12–15

VALUES OF EULER POLYNOMIALS

B. SURY

Presented by M. Ram Murty, FRSC

ABSTRACT. Recently, G. J. Fox [F] proved a theorem on the values of Euler polynomials at rational numbers analogous to a similar result on Bernoulli polynomials. The result on Bernoulli polynomials was originally discovered by Almkvist and Meurmann [AM]. Another proof was given by the author [S] and the object of this note is to point out that it actually provides a very easy proof of a generalisation of Fox's result.

RÉSUMÉ. Nous démontrons un résultat aux valeurs des polynômes d'Euler aux nombres rationnels. C'est analogue à un résultat connu pour les polynômes de Bernoulli et generalise le nouveau travail de G. J. Fox.

The Bernoulli polynomials $B_n(t)$ are defined by the identity

$$\frac{Xe^{tX}}{e^X-1}=\sum_{n=0}^{\infty}B_n(t)\frac{X^n}{n!}.$$

The Euler polynomials $E_n(t)$ are similarly defined by the identity

$$\frac{2e^{tX}}{e^X+1}=\sum_{n=0}^{\infty}E_n(t)\frac{X^n}{n!}.$$

It will be seen also that the Bernoulli numbers $B_n = B_n(0)$ and the Euler numbers $E_n = E_n(0)$ can be recovered from a Pascal-like triangle. Almkvist and Meurmann [AM] proved that for a rational number r/s, the value $s^n\big(B_n(r/s) - B_n(0)\big) \in \mathbf{Z}$ for every $n \geq 0$. Another simple proof of this was given in [S]. Fox proves that the Euler polynomials satisfy $s^n\big(E_n(r/s) + (-1)^{rs-1}E_n(0)\big) \in \mathbf{Z}$ for every $n \geq 0$. Here, we shall notice that the proof in [S] gives an easy proof of Fox's result and also generalises it slightly. We prove:

THEOREM. *For any arbitrary integer r, $s^nE_n(r/s) \in \mathbf{Z}$ if s is even. If s is odd, then $s^n\big(E_n(r/s) + (-1)^{r-1}E_n(0)\big) \in \mathbf{Z}$.*

PROOF. Write $a_n = s^nE_n(r/s)$. Then, $\sum a_n\frac{X^n}{n!} = \frac{2e^{rX}}{e^{sX}+1}$. On comparing the coefficients on both sides of the identity $\sum a_n\frac{X^n}{n!}(e^{sX}+1) = 2e^{rX}$, one gets $a_0 = 1$, $2(r^n - a_n) = \sum_{k=0}^{n-1}\binom{n}{k}a_ks^{n-k}$ for $n > 0$.

Received by the editors March 7, 2000.
AMS subject classification: 11B68.

If s is even, one can divide out by 2 and this immediately gives by induction that a_n are integers for all $n \geq 0$.

Let s be odd and let us write $b_n = s^n\big(E_n(r/s) + (-1)^{r-1}E_n(0)\big)$. Then, one can rewrite the above identity as $\frac{e^{sX}+1}{e^X+1}\sum \frac{b_n X^n}{n!} = 2\frac{e^{rX}+(-1)^{r-1}}{e^X+1}$, *i.e.*,

$$\sum_{l=0}^{s-1}(-1)^l e^{lX}\sum \frac{b_n X^n}{n!} = 2\frac{e^{rX}+(-1)^{r-1}}{e^X+1} = 2(-1)^{r-1}\Big(\sum_{m=0}^{r-1}(-1)^m e^{mX}\Big).$$

Comparing the coefficients of like powers of X, one obtains $b_0 = 0$ or 2 according as r is even or odd and for $n > 0$,

$$\frac{b_n}{n!} + \sum_{k=0}^{n-1}\frac{b_k p_{n-k}}{k!(n-k)!} = 2(-1)^{r-1}\frac{q_n}{n!}$$

where $p_l = -1+2^l-3^l+\cdots+(s-1)^l$ and $q_l = -1+2^l-3^l+\cdots+(-1)^{r-1}(r-1)^l$. Multiplying by $n!$, we get

$$b_n + \sum_{k=0}^{n-1}\binom{n}{k}b_k p_{n-k} = 2(-1)^{r-1}q_n$$

which again gives immediately by induction that $b_n \in \mathbf{Z}$. This proves the theorem.

REMARKS. It is clear that the $E_n(t)$ and the $B_n(t)$ are related by the relation

$$\sum_{n=0}^{\infty}E_n(t)\frac{X^n}{n!}\sum_{n=0}^{\infty}B_n(t)\frac{X^n}{n!} = \sum_{n=0}^{\infty}B_n(t)\frac{(2X)^n}{n!}.$$

In fact, it may be of interest to point out here that the Euler numbers and the Bernoulli numbers have as ancestors the so-called *up-down* numbers u_n; these can be read off from a Pascal-like triangle.

An *up-down* sequence of length n is a sequence of integers $a_1 < a_2 > a_3 < a_4 > \cdots a_n$ where the set $\{a_1, a_2, \ldots, a_n\}$ is a permutation of the set $\{1, 2, \ldots, n\}$.

For example, $1 < 3 > 2$ and $2 < 3 > 1$ are the only up-down sequences of length 3. The *up-down number* u_n is defined to be the number of up-down sequences of length n. Therefore, $u_3 = 2$ as above. One can also see that $u_4 = 5$, $u_5 = 16$ *etc.*

The relation with the Euler-Bernoulli numbers was discovered by Arnold [A] and is the following. First, note that from the generating functions for the Euler numbers $E_n = E_n(0)$ and the Bernoulli numbers $B_n = B_n(0)$, one gets the fact that $E_n = 0$ if $n \neq 0$ is even and $B_n = 0$ if n is odd. On putting $X = 2iy$ in the generating function for the Euler numbers, one gets

$$\tan(y) = \sum_n E_{2n-1}(-1)^n\frac{(2y)^{2n-1}}{(2n-1)!}.$$

On the other hand, one can break any up-down sequence at its largest or its smallest term to get two sequences and it is a nice and easy exercise to see that the resulting count gives the recursive relation

$$2u_{n+1} = \sum_{l=0}^{n} \binom{n}{l} u_l u_{n-l}.$$

Thus, the generating function $U(t) = \sum_{n=0}^{\infty} u_n \frac{t^n}{n!}$ satisfies the differential equation $2U'(t) = 1 + U(t)^2$. Solving this with the initial condition $U(0) = u_0 = 1$, one has

$$\sum_{n=0}^{\infty} u_n \frac{t^n}{n!} = \frac{1 + \operatorname{Sin} t}{\operatorname{Cos} t}.$$

On putting $-t$ in place of t and subtracting, one gets

$$\tan(t) = \sum_{n} u_{2n-1} \frac{t^{2n-1}}{(2n-1)!}.$$

Therefore,

$$u_{2n-1} = (-1)^n 2^{2n-1} E_{2n-1} = (-1)^{n-1} 2^{2n-1} (2^{2n} - 1) \frac{B_n}{n}.$$

There is a way to read off the up-down numbers (and therefore the Euler and Bernoulli numbers) is to form a triangle akin to the Pascal triangle. The triangle looks as follows:

						1								
					1		0							
				0		1		1						
			2		2		1		0					
		0		2		4		5		5				
	16		16		14		10		5		0			
0		16		32		46		56		61		61		
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

The entries in the triangle are explained as follows. Each entry in a row with an odd (respectively even) number of elements is the sum of all the entries to the left (respectively right) on the row above. The up-down numbers are at the two extreme edges of this triangle—the right side has the numbers u_{even} and the left side has the numbers u_{odd}.

References

[A] V. Arnold, *Snake calculus and the combinatorics of the Euler, Bernoulli and Springer numbers of Coxeter groups.* Russian Math. Surveys 47(1992), 3–40.

[AM] G. Almkvist and A. Meurman, *Values of Bernoulli polynomials and Hurwitz's zeta function at rational points.* C. R. Math. Rep. Acad. Sci. Canada **13**(1991), 104–108.

[F] G. J. Fox, *Euler polynomials at rational numbers.* C. R. Math. Rep. Acad. Sci. Canada **21**(1999), 87–90.

[S] B. Sury, *The values of Bernoulli polynomials at rational numbers.* Bull. London Math. Soc. **25**(1993), 327–329.

Statistics & Mathematics Unit
Indian Statistical Institute
8th Mile Mysore Road
Bangalore 560 059
India
email: sury@isibang.ac.in

C. R. Math. Rep. Acad. Sci. Canada Vol. 23 (1), 2001 pp. 16–21

REMARK ON APPROXIMATELY JENSEN-CONVEX FUNCTIONS

JACEK MROWIEC

Presented by Vlastimil Dlab, FRSC

ABSTRACT. Let X be a real vector space and $D \subset X$ be a convex set. It is shown that every ε-J-convex function $f: D \to \mathbf{R}$ is also $C\varepsilon$-$\mathbb{Q}$-convex, where C is a constant independent of f. It is also proved that every ε-W-convex and J-convex function is W-convex.

RÉSUMÉ. Que X désigne un espace réel des vecteurs et que $D \subset X$ soit un ensemble convexe. Il est montré que chaque ε-J-fonction convexe $f: D \to \mathbf{R}$ est aussi $C\varepsilon$-$\mathbb{Q}$-convexe, où C est une constante indépendante de f. Il est aussi prouvé que chaque ε-W-convexe et J-fonction convexe est W-convexe.

Let X be a real vector space, $D \subset X$ be a convex set and $\varepsilon \geq 0$ be a fixed number. A function $f: D \to \mathbb{R}$ is called:

- ε-*convex* if

$$(*) \qquad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) + \varepsilon$$

  for all $x, y \in D$ and $\lambda \in [0,1]$;
- ε-$\mathbb{Q}$-*convex* if it satisfies $(*)$ for all $x, y \in D$ and $\lambda \in \mathbb{Q} \cap [0,1]$;
- ε-J-*convex* if it satisfies $(*)$ for all $x, y \in D$ and $\lambda = 1/2$, *i.e.*,

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} + \varepsilon, \quad x, y \in D.$$

If f satisfies the above conditions with $\varepsilon = 0$, it is called *convex*, $\mathbb{Q}$-*convex* and J-*convex* (Jensen convex), respectively.

Of course, every ε-convex function is ε-$\mathbb{Q}$-convex and every ε-$\mathbb{Q}$-convex function is ε-J-convex. It is also well known that every J-convex function is $\mathbb{Q}$-convex, but need not be convex (*cf. e.g.* [4, p. 125]). We will show that similar result holds for ε-J-convex functions.

In the proof of our main result (Theorem 1) we will use the following Euler theorem (*cf.* [9, p. 261]).

Received by the editors March 20, 2000.
AMS subject classification: 39B72, 26A51.

THEOREM (EULER). *Let $\varphi(n)$ denote the number of all natural numbers not exceeding* $\mathbf{n}$ *and relatively prime to* $\mathbf{n}$. *If natural numbers* $\mathbf{a}$ *and* $\mathbf{n}$ *are relatively prime, then $(a^{\varphi(n)}-1)$ is divisible by* $\mathbf{n}$.

COROLLARY 1. *For any odd* $\mathbf{n}>1$ *there exists a natural* $\mathbf{d}<\mathbf{n}$ *($1\le d\le n-1$, $d=\varphi(n)$), such that (2^d-1) is divisible by* $\mathbf{n}$.

LEMMA 1 (*cf.* [7, LEMMA 1]). *If a function $f\colon D\to\mathbb{R}$ is ε-J-convex, then the following inequality holds:*

$$(1)\qquad \bigwedge_{x,y\in D}\bigwedge_{m\in\mathbb{N}} f\Big(\frac{p}{2^m}x+\Big(1-\frac{p}{2^m}\Big)y\Big)\le\frac{p}{2^m}f(x)+\Big(1-\frac{p}{2^m}\Big)f(y)+C_0(m)\varepsilon,$$

where $C_0(m):=2(1-2^{-m})<2$, $1\le p<2^m$.

LEMMA 2. *If a function $f\colon D\to\mathbb{R}$ is ε-J-convex, then:*

$$(2)\qquad \bigwedge_{x,y\in D}\bigwedge_{n\in\mathbb{N}} f\left(\frac{1}{n}x+\frac{n-1}{n}y\right)\le\frac{1}{n}f(x)+\frac{n-1}{n}f(y)+C_1(n)\varepsilon,$$

where $C_1(n):=\frac{n-1}{2^m-1}$ and $m:=[\log_2 n]$. Moreover:

$$(3)\qquad 2(1-2^{-m})\le C_1(n)\le 4(1-2^{-m}).$$

PROOF (INDUCTION). Of course, inequality (2) holds for $n=2$ (and for $n=1$). Write

$$a:=\frac{1}{n+1}x+\frac{n}{n+1}y\quad\text{and}\quad b:=\frac{n}{n+1}x+\frac{1}{n+1}y.$$

Then the following equations hold:

$$a=\frac{1}{n}b+\frac{n-1}{n}y\quad\text{and}\quad b=\frac{n-1}{n}x+\frac{1}{n}a.$$

For every $n\in\mathbb{N}$ there exists a natural number $m=m(n)$ such that:

$$(1.1)\qquad 2^m\le n\le 2^{m+1}-1.$$

For induction, assume that (2) holds for some $n\in\mathbb{N}$. So we have:

$$(1.2)\qquad f(b)\le\frac{n-1}{n}f(x)+\frac{1}{n}f(a)+C_1(n)\varepsilon$$

and

$$f(a)\le\frac{1}{n}f(b)+\frac{n-1}{n}f(y)+C_1(n)\varepsilon.$$

By (1.2) we have:

$$f(a)\le\frac{1}{n}\left(\frac{n-1}{n}f(x)+\frac{1}{n}f(a)+C_1(n)\varepsilon\right)+\frac{n-1}{n}f(y)+C_1(n)\varepsilon,$$

and then we get:

$$f(a) \le \frac{1}{n+1} f(x) + \frac{n}{n+1} f(y) + C_1(n+1)\varepsilon,$$

where $C_1(n+1) = C_0(m+1)$ if $n+1 = 2^{m+1}$ (from (1)) and $C_1(n+1) = \frac{n}{n-1} C_1(n)$ if $n+1 \neq 2^{m+1}$.

Assume that $n+1 < 2^{m+1}$. Then $m = [\log_2 n] = [\log_2(n+1)]$ and

$$\begin{aligned} C_1(n+1) &= \frac{n}{n-1} \cdot \frac{n-1}{n-2} \cdot \ldots \cdot \frac{2^m+1}{2^m} \cdot \frac{2^m}{2^m-1} \cdot C_1(2^m) \\ &= \frac{n}{2^m-1} \cdot C_0(m) = \frac{n}{2^m-1} \cdot \frac{2^m-1}{2^{m-1}} = \frac{n}{2^{m-1}}. \end{aligned}$$

Inequality (3) results from (1.1) and the formula on $C_1(n)$.

THEOREM 1. *If a function $f: D \to \mathbb{R}$ is ε-J-convex, then it is $C_2\varepsilon$-$\mathbb{Q}$-convex, i.e.,*

$$\bigwedge_{x,y \in D} \bigwedge_{\substack{n,p \in \mathbb{N} \\ p \le n}} f\left(\frac{p}{n}x + \frac{n-p}{n}y\right) \le \frac{p}{n} f(x) + \frac{n-p}{n} f(y) + C_2(n,p)\varepsilon, \tag{4}$$

where $C_2(n,p) < 10\frac{4}{9}$.

PROOF. Let us fix $x, y \in D$. We need only consider 2 cases.

CASE 1°. n is even.

If $n = 2^m$, then (4) follows from Lemma 1. Assume therefore, that $n \neq 2^m$ and fix a $p < n$. Define

$$c := \frac{n-1}{n}x + \frac{1}{n}y \quad \text{and} \quad a := \frac{p}{n}x + \frac{n-p}{n}y. \tag{2.1}$$

We conclude from Corollary 1 that there exist natural k, q such that $(n-1)k = 2^q - 1$, where $q = \varphi(n-1)$. Now let $b := \frac{nk-k+1}{nk}x + \frac{k-1}{nk}y$. Combining these equalities we obtain

$$a = \frac{kp}{(n-1)k+1} b + \frac{(n-1)k+1-kp}{(n-1)k+1} y = \frac{kp}{2^q} b + \frac{2^q - kp}{2^q} y \tag{2.2}$$

and

$$b = \frac{1}{k}x + \frac{k-1}{k}c. \tag{2.3}$$

From Lemma 2, (2.1) and (2.3) we get

$$f(c) \le \frac{n-1}{n} f(x) + \frac{1}{n} f(y) + C_1(n)\varepsilon$$

and

$$f(b) \leq \frac{1}{k}f(x) + \frac{k-1}{k}\left(\frac{n-1}{n}f(x) + \frac{1}{n}f(y) + C_1(n)\varepsilon\right) + C_1(n)\varepsilon.$$

An easy computation shows that

$$f(b) \leq \frac{nk-k+1}{nk}f(x) + \frac{k-1}{nk}f(y) + \left(C_1(k) + \frac{k-1}{k}C_1(n)\right)\varepsilon. \tag{2.4}$$

By (2.2) and (1):

$$f(a) \leq \frac{kp}{(n-1)k+1}f(b) + \frac{(n-1)k+1-kp}{(n-1)k+1}f(y) + C_0(q)\varepsilon.$$

Next by (2.4):

$$f(a) \leq \frac{p}{n}f(x) + \frac{n-p}{n}f(y) + C_2'(n,p)\varepsilon, \tag{2.5}$$

where

$$C_2'(n,p) = C_0(q) + \frac{kp}{2^q}\left(C_1(k) + \frac{k-1}{k}C_1(n)\right).$$

Since $C_1(n) < 4$ and $C_0(m) < 2$ for any natural n, m, we conclude that $C_2'(n,p) < 2 + 8 \cdot \frac{kp}{2^q}$.

By symmetry in x and y we may assume that $p \leq n/2$. Then

$$C_2'(n,p) < 2 + 4 \cdot \frac{nk}{(n-1)k+1} < 2 + 4 \cdot \frac{n}{n-1}.$$

The sequence $(\frac{n}{n-1})$ is decreasing, for $n = 2, 4, 8$ by (1) and for $n = 6$ by (2) there is better estimation of constant C_2', so we put $n = 10$. Therefore $C_2'(n,p) < 6\frac{4}{9}$.

CASE 2°. n is odd.

Let us denote a by formula $a := \frac{p}{n}x + \frac{n-p}{n}y$ and

$$y_1 := \frac{1}{n}x + \frac{n-1}{n}y. \tag{2.6}$$

Then $a = \frac{p-1}{n-1}x + \frac{n-p}{n-1}y_1$. Since $n-1$ is even, we put in (2.5) $n-1$, $p-1$, y_1 in the place of n, p, y, respectively, and we get

$$f(a) \leq \frac{p-1}{n-1}f(x) + \frac{n-p}{n-1}f(y_1) + C_2'(n-1,p-1)\varepsilon. \tag{2.7}$$

By Lemma 2 and (2.6) we have

$$f(y_1) \leq \frac{1}{n}f(x) + \frac{n-1}{n}f(y) + C_1(n)\varepsilon. \tag{2.8}$$

Therefore from (2.7) and (2.8):

$$f(a) \leq \frac{p-1}{n-1}f(x) + \frac{n-p}{n-1}\left(\frac{1}{n}f(x) + \frac{n-1}{n}f(y) + C_1(n)\varepsilon\right) + C_2'(n-1,p-1)\varepsilon.$$

Hence

$$f(a) \leq \frac{p}{n}f(x) + \frac{n-p}{n}f(y) + C_2''(n,p)\varepsilon,$$

where

$$C_2''(n,p) = C_2'(n-1,p-1) + \frac{n-p}{n-1}C_1(n). \tag{2.9}$$

From (2.9) and (3) we obtain:

$$C_2''(n,p) \leq C_2'(n-1,p-1) + C_1(n) < 10\tfrac{4}{9}.$$

Now it suffices to admit $C_2 := C_2'$ when n is even, $C_2 := C_2''$ when n is odd.

REMARK 1. In 1952 D. H. Hyers and S. Ulam (*cf.* [3]) proved that if $f\colon D \to \mathbb{R}$, where D is a convex subset of $\mathbb{R}^n$, is ε-convex than there exists a convex function g such that $|f(x) - g(x)| \leq k_n\varepsilon$, $x \in D$, where k_n is a constant depending only on the dimension of the domain (*cf.* also [1], [5]). In [1] Cholewa showed that analogous result for ε-J-convex functions is not true (*cf.* also [2]). However the domain of the function in this counterexample is a $\mathbb{Q}$-convex subset of $\mathbb{R}^n$. For convex domain in $\mathbb{R}^n$ this question is still open. The above Theorem 1 reduces the problem of the stability of J-convex functions to the problem of the stability of $\mathbb{Q}$-convex functions.

Now, let us recall that a function $f\colon D \to \mathbb{R}$ is said to be *ε-W-convex* (*cf.* [6], [7], [8]) if

$$\bigwedge_{x,y\in D}\ \bigwedge_{t\in[0,1]} f(tx + (1-t)y) + f((1-t)x + ty) \leq f(x) + f(y) + 2\varepsilon, \tag{**}$$

where $\varepsilon > 0$.

Functions satisfying (**) with $\varepsilon = 0$ are called *W-convex* (Wright-convex). The stability problem for W-convex functions is also open. However, we have the following result:

THEOREM 2. *Let X be a real vector space and D be a convex subset of X. If the function $f\colon D \to \mathbb{R}$ is J-convex and ε-W-convex, then it is W-convex.*

PROOF. Let $t \in [0, \frac{1}{2}]$. Then:

$$\begin{aligned} &f\big(tx + (1-t)y\big) + f\big(ty + (1-t)x\big) \\ &\quad = f\Big(\tfrac{1}{2}\big(2tx + (1-2t)y\big) + \tfrac{1}{2}y\Big) + f\Big(\tfrac{1}{2}\big(2ty + (1-2t)x\big) + \tfrac{1}{2}x\Big) \\ &\quad \le \tfrac{1}{2}f\big(2tx + (1-2t)y\big) + \tfrac{1}{2}f(y) + \tfrac{1}{2}f\big(2ty + (1-2t)x\big) + \tfrac{1}{2}f(x) \\ &\quad = \tfrac{1}{2}\big[f\big(2tx + (1-2t)y\big) + f\big(2ty + (1-2t)x\big)\big] + \tfrac{1}{2}[f(x) + f(y)] \\ &\quad \le \tfrac{1}{2}[f(x) + f(y) + 2\varepsilon] + \tfrac{1}{2}[f(x) + f(y)] = f(x) + f(y) + 2\varepsilon_1, \end{aligned}$$

where $\varepsilon_1 = \frac{1}{2}\varepsilon$.

By symmetry in x and y, the above extends to all $t \in [0,1]$. Therefore f is ε_1-W-convex. Iterating this scheme we get that f is ε_n-W-convex for $n = 2, 3, \ldots$, where $\varepsilon_n = \frac{1}{2}\varepsilon_{n-1} = \frac{1}{2^n}\varepsilon$. Since $\varepsilon_n \to 0$ as $n \to \infty$, we obtain the conclusion that f is W-convex.

REFERENCES

1. P. W. Cholewa, *Remarks on the stability of functional equations.* Aequationes Math. **27**(1984), 76–86.
2. R. Ger, *Almost approximately convex functions.* Math. Slovaca (2) **38**(1988), 61–77.
3. D. H. Hyers and S. M. Ulam, *Approximately convex functions.* Proc. Amer. Math. Soc. **3**(1952), 821–828.
4. M. Kuczma, *An introduction to the theory of functional equations and inequalities.* Polish Scientific Publishers and Silesian University, Warszawa-Kraków-Katowice, 1985.
5. M. Laczkovich, *The local stability of convexity, affinity and of the Jensen equation.* Aequationes Math. **58**(1999), 135–142.
6. C. T. Ng, *Functions generating Schur-convex sums.* General Inequalities 5 (Oberwolfach, 1986), Internat. Schriftenreihe Numer. Math. **80**, Birkhäuser, Basel-Boston, MA, 1987, 433–438.
7. K. Nikodem and C. T. Ng, *On approximately convex functions.* Proc. Amer. Math. Soc. **118**(1993), 103–108.
8. A. W. Roberts and D. E. Varberg, *Convex functions.* Academic Press, New York and London, 1973.
9. W. Sierpiński, *Elementary theory of numbers.* PWN—Polish Scientific Publishers, Warszawa, 1987.

Department of Mathematics
Technical University of Łódź
Branch in Bielsko-Biała
Willowa 2
PL 43-309 Bielsko-Biała
Poland

C. R. Math. Rep. Acad. Sci. Canada Vol. 23 (1), 2001 pp. 22–27

A REMARK ON INHOMOGENEOUS CAUCHY PROBLEMS

SERGIO AMAT AND SONIA BUSQUIER

Presented by Vlastimil Dlab, FRSC

ABSTRACT. In this paper, a local existence and uniqueness theorem is presented for some inhomogeneous Cauchy problems. It is shown that we can choose the initial condition in the whole space, and not in a subset, like in the classical theorems.

RÉSUMÉ. Dans cet article, nous présentons un théorème d'existence et d'unicité local pour un problème de Cauchy non-homogène. Nous démontrons que la condition initiale peut être choisie dans tout l'espace et qu'aucune restriction à un sous-espace n'est nécessaire comme c'est le cas dans la théorie classique.

1. Inhomogeneous Cauchy problems. We consider the following Cauchy problem:

$$u_t = Au(t) + f\big(t, u(t)\big), \quad t_0 < t < t_1 \tag{1.1}$$

$$u(t_0) = u_0 \tag{1.2}$$

where $-A$ is a w-sectorial operator ($0 < w < \frac{\pi}{2}$) and $0 \in \rho(A)$.

$f: W \subset [0, \infty[\times X \to X$, X a Banach space, such that $\forall (t,u) \in W$ there exist $V \subset W$ (W is open), constants $C(t,u,V) > 0$ and $0 \leq \gamma \leq 1$, verifying

$$\|f(s_1, y_1) - f(s_2, y_2)\| \leq C(|s_1 - s_2|^\gamma + \|y_1 - y_2\|), \tag{1.3}$$

$(s_1, y_1), (s_2, y_2) \in W$.

DEFINITION 1.1. Let $A: D(A) \subset X \to X$ be a closed linear operator on a Banach space X and $0 < w \leq \pi$. A is w-sectorial if

(i) $\sigma(A) \subset S_w = \{z \in \mathbb{C} : |\arg z| < w\}$.

(ii) There exists a constant $M > 0$ such that, for $z \notin S_w$

$$\|(z - A)^{-1} x\| \leq \frac{M}{|z|} \|x\|, \quad \forall x \in X.$$

DEFINITION 1.2. A function $u(t): [t_0, t_1[\to X$ is a solution of (1.1) and (1.2) if it verifies

Received by the editors August 7, 2000.
AMS subject classification: 47D05, 35F10, 35F25, 35G25.
Key words and phrases: inhomogeneous Cauchy problems, existence, uniqueness.

1) $u(t) \in C([t_0, t_1[\,; X) \cap C^1(]t_0, t_1[\,; X)$ and $u(t_0) = u_0$.
2) $u(t) \in D(A)$ and $\big(t, u(t)\big) \in W \ \forall t_0 < t < t_1$.
3) $u_t = Au(t) + f\big(t, u(t)\big) \ \forall t_0 < t < t_1$.

DEFINITION 1.3. A semigroup $U(t)$, $0 \le t < \infty$, of bounded linear operators on X is a strongly continuous semigroup of bounded linear operators if

$$\lim_{t \downarrow 0} U(t)x = x, \qquad \forall x \in X. \tag{1.4}$$

A strongly continuous semigroup of bounded linear operators on X will be called a semigroup of class C_0.

THEOREM 1.4. *For all $(t_0, u_0) \in W$, the initial value problem* (1.1) *and* (1.2) *has a unique local solution.*

PROOF.

STEP 1. Let $(t_0, u_0) \in W$ be fixed. We choose $\epsilon > 0$ and $\delta > 0$ such that

$$V = \{(t, x) \in [0, +\infty[\times X : t_0 \le t \le t_0 + \epsilon, \|x - x_0\| \le \delta\} \tag{1.5}$$

and

$$\|f(t, u) - f(s, v)\| \le C(|t - s|^\gamma + \|u - v\|) \tag{1.6}$$

$\forall (t, u), (s, v) \in W$ with $C = C(t_0, u_0, V)$.

Let $U(t)$ be the semigroup with generator $-A$. Then there exists $M_\alpha > 0$ $(0 \le \alpha \le 1)$ such that

$$\|A^\alpha U(t)\| \le M_\alpha t^{-\alpha}, \quad t > 0. \tag{1.7}$$

Let t_1 such that

$$0 < t_1 - t_0 < \min\left\{\epsilon, \left(\frac{\delta}{4} \frac{1}{M_{\frac{1}{2}}(B + \delta C)\|A^{-\frac{1}{2}}\|}\right)^2\right\} \tag{1.8}$$

where $B = \max_{t_0 \le t \le t_0 + \epsilon} \|f(t, u_0)\|$.

$U(t)$ is strongly continuous, then we can choose the t_1 such that

$$\|U(t - t_0)u_0 - u_0\| < \frac{\delta}{2}, \quad (t_0 < t < t_1). \tag{1.9}$$

STEP 2. Let $Y = C([t_0, t_1[\,; X)$ with the norm $|||y||| = \max_{t_0 \le t \le t_1} \|y(t)\|$. We define $\Phi: Y \to Y$ as follows:

$$\Phi(y)(t) = U(t - t_0)u_0 + \int_{t_0}^{t} U(t - s) f\big(s, y(s)\big)\, ds,$$

where $U(t - t_0)u_0 \in C([t_0, +\infty[\,; X) \cap C^\infty(]t_0, +\infty[\,; X)$.

Moreover

$$\begin{aligned}\int_{t_0}^{t+h} U(t+h-s)f\big(s,y(s)\big)\,ds - \int_{t_0}^{t} U(t-s)f\big(s,y(s)\big)\,ds \\ = \int_{t_0}^{t} \big(U(t+h-s)-U(t-s)\big)f\big(s,y(s)\big)\,ds \\ + \int_{t_0}^{t+h} U(t+h-s)f\big(s,y(s)\big)\,ds.\end{aligned} \tag{1.10}$$

But, y is continuous and for (1.5) and (1.6) there exists $N>0$ such that

$$\big\|f\big(t,y(t)\big)\big\| \le N, \quad t\in[t_0,t_1]. \tag{1.11}$$

It is well known that $\forall\beta$ $(0<\beta<1)$ there exists $M_\beta>0$ such that

$$\big\|\big(U(h)-I\big)U(t-s)\big\| \le M_\beta h^\beta \|A^\beta U(t-s)\|. \tag{1.12}$$

Then applying (1.7) and (1.12) we have

$$\big\|\big(U(h)-I\big)U(t-s)\big\| \le M_\beta h^\beta C_\beta (t-s)^{-\beta}, \quad h>0. \tag{1.13}$$

Thus,

$$\begin{aligned}\left\|\int_{t_0}^{t}\big(U(t+h-s)-U(t-s)\big)f\big(s,y(s)\big)\,ds\right\| \\ = \left\|\int_{t_0}^{t}\big(U(h)-I\big)U(t-s)f\big(s,y(s)\big)\,ds\right\| \\ \le M_\beta h^\beta C_\beta N\int_{t_0}^{t}(t-s)^{-\beta}\,ds \\ \le \left(\frac{M_\beta C_\beta N}{1-\beta}\right)(t_1-t_0)^{1-\beta}h^\beta, \quad 0<\beta<1,\end{aligned} \tag{1.14}$$

and finally, for (1.7),

$$\begin{aligned}\left\|\int_{t}^{t+h} U(t+h-s)f\big(s,y(s)\big)\,ds\right\| &\le N\int_{t}^{t+h}\|U(t+h-s)\|\,ds \\ &= N\int_{t}^{t+h}\|A^{-\beta}A^\beta U(t+h-s)\|\,ds \\ &\le \left(\frac{NC_\beta}{1-\beta}\right)h^{1-\beta}\|A^{-\beta}\|.\end{aligned} \tag{1.15}$$

Taking $\beta=\frac{1}{2}$, we have $\Phi\big(y(t)\big)\in C([t_0,t_1];X)\cap C^{\frac{1}{2}}([t_0,t_1];X)$.

Left continuity is similar:

$$\begin{aligned}
&\int_{t_0}^{t-h} U(t-h-s)f\big(s,y(s)\big)\,ds - \int_{t_0}^{t} U(t-s)f\big(s,y(s)\big)\,ds \\
&\quad = \int_{t_0}^{t-h} \big(U(t-h-s)-U(t-s)\big)f\big(s,y(s)\big)\,ds - \int_{t-h}^{t} U(t-s)f\big(s,y(s)\big)\,ds \\
&\quad = \int_{t_0}^{t-h} \big(I-U(h)\big)U(t-h-s)f\big(s,y(s)\big)\,ds - \int_{t-h}^{t} U(t-s)f\big(s,y(s)\big)\,ds
\end{aligned} \tag{1.16}$$

and both terms are bounded as above.

STEP 3. Let $J = \{y \in Y : y(t_0) = u_0, \max_{t_0\le t\le t_1} \|y(t)-u_0\| \le \delta\}$, with $J \neq \emptyset$ $(y(t) = u_0,\ \forall t \in [t_0,t_1])$.

Since J is closed and X is a Banach space, J is a complete metric space.

We are going to prove:

a) $\Phi\colon J \to J$.

b) Φ is contractive.

For a):

a-1) $\Phi(y)(t_0) = u_0$.

a-2) Applying (1.6), (1.7), (1.8) and (1.9), we have

$$\begin{aligned}
&\|\Phi((y)(t)) - u_0\| \\
&\quad = \left\|U(t-t_0)u_0 - u_0 + \int_{t_0}^{t} U(t-s)f\big(s,y(s)\big)\,ds\right\| \\
&\quad \le \|U(t-t_0)u_0 - u_0\| + \left\|\int_{t_0}^{t} A^{-\beta}A^{\beta}U(t-s)\Big(f\big(s,y(s)\big) - f(s,u_0)\Big)\,ds\right\| \\
&\quad\quad + \left\|\int_{t_0}^{t} A^{-\beta}A^{\beta}U(t-s)f(s,u_0)\,ds\right\| \\
&\quad \le \frac{\delta}{2} + \|A^{-\beta}\|M_{\beta}(C_{\delta}+\beta)\int_{t_0}^{t}(t-s)^{-\beta}\,ds \\
&\quad \le \frac{\delta}{2} + \|A^{-\beta}\|M_{\beta}\frac{(C_{\delta}+\beta)}{1-\beta}(t_1-t_0)^{1-\beta} \le \delta; \quad (\beta = \frac{1}{2}).
\end{aligned} \tag{1.17}$$

b)

$$\begin{aligned}
&\|\Phi(y_1(t)) - \Phi(y_2(t))\| \\
&\quad = \left\|\int_{t_0}^{t_1} U(t-s)\big[f\big(s,y_1(s)\big) - f\big(s,y_2(s)\big)\big]\,ds\right\| \\
&\quad = \left\|\int_{t_0}^{t_1} A^{-\frac{1}{2}}A^{\frac{1}{2}}U(t-s)\big[f\big(s,y_1(s)\big) - f\big(s,y_2(s)\big)\big]\,ds\right\| \\
&\quad \le \int_{t_0}^{t_1} \|A^{-\frac{1}{2}}A^{\frac{1}{2}}U(t-s)\|\,\big\|f\big(s,y_1(s)\big) - f\big(s,y_2(s)\big)\big\|\,ds
\end{aligned}$$

$$\begin{aligned} &\leq \|A^{-\frac{1}{2}}\| M_{\frac{1}{2}} C \int_{t_0}^{t} (t-s)^{-\frac{1}{2}} \|y_1(s) - y_2(s)\|\, ds \\ &\leq \|A^{-\frac{1}{2}}\| M_{\frac{1}{2}} C \frac{(t_1-t_0)^{\frac{1}{2}}}{1-\frac{1}{2}} \|y_1(s) - y_2(s)\| \\ &\leq \frac{1}{2} |||y_1 - y_2|||. \end{aligned} \tag{1.18}$$

Then (fixed point theorem), there exists a unique $y \in J$ such that $\Phi(y) = y$, *i.e.*,

$$y(t) = U(t-t_0)u_0 + \int_{t_0}^{t} U(t-s) f\big(s, y(s)\big)\, ds.$$

Moreover $y \in C^{\frac{1}{2}}([t_0, t_1]; X)$ therefore there exists a constant $C_{t_0'} > 0$ such that

$$\|y(t) - y(s)\| \leq C_{t_0'} |t-s|^{\frac{1}{2}}, \quad t, s \in [t_0', t_1].$$

On the other hand, $f\big(t, y(t)\big)$ is locally Hölder continuous:

$$\big\| f\big(t, y(t)\big) - f\big(s, y(s)\big) \big\| \leq C\big(|t-s|^{\gamma} + \|y(t) - y(s)\|\big) \tag{1.19}$$

$$\leq C\big(|t-s|^{\gamma} + CC_{t_0'} |t-s|^{\frac{1}{2}}\big), \quad (0 \leq \gamma \leq 1). \tag{1.20}$$

STEP 4. We consider

$$u_t = Au(t) + f\big(t, y(t)\big), \quad t_0 < t < t_1, \tag{1.21}$$

$$u(t_0) = u_0. \tag{1.22}$$

The solution of this problem is

$$u(t) = U(t-t_0)u_0 + \int_{t_0}^{t} U(t-s) f\big(s, y(s)\big)\, ds.$$

Moreover, $u(t) \in C([t_0, t_1]; X) \cap C^1(]t_0, t_1[; X)$. If $y(t) = u(t)$, then $u(t)$ is the solution of problem (1.1). Since, $\Phi\big(y(t)\big) = y(t)$, then $y(t) = u(t)$.

UNIQUENESS. We suppose that $w(t)$ and $v(t)$ are two solutions of (1.1). Then

$$w(t) = U(t-t_0)u_0 + \int_{t_0}^{t} U(t-s) f\big(s, w(s)\big)\, ds$$

and

$$v(t) = U(t-t_0)u_0 + \int_{t_0}^{t} U(t-s) f\big(s, v(s)\big)\, ds$$

are solutions of the problems

$$u_t = Au(t) + f\big(t, w(t)\big), \quad t_0 < t < t_1, \tag{1.23}$$

$$u(t_0) = u_0, \tag{1.24}$$

and

$$u_t = Au(t) + f\big(t, v(t)\big), \quad t_0 < t < t_1, \tag{1.25}$$

$$u(t_0) = u_0, \tag{1.26}$$

respectively.

Thus, $w(t)$ and $v(t)$ are fixed points of Φ. It follows that $w(t) = v(t)$. ∎

Then we can choose the initial condition in all the space, and not in a subset of it like in the classical theorems [8].

ACKNOWLEDGMENTS. We thank Miguel Sanz Alix for several useful discussions and for providing a good background about this topic, and the referees for many valuable suggestions that made this paper substantially better.

REFERENCES

1. R. Beals, *On the abstract Cauchy problem.* J. Funct. Anal. **10**(1972), 281–299.
2. J. A. Goldstein, *Abstract evolution equations.* Trans. Amer. Math. Soc. **141**(1969), 159–185.
3. ———, *Semi-groups of operators and abstract Cauchy problems.* Tulane Univ. Lecture Notes, 1970.
4. E. Hille, *Functional analysis and semi-groups.* Amer. Math. Soc. Colloq. Publ. **31**, New York, 1948.
5. T. Kato and H. Tanabe, *On the abstract evolution equation.* Osaka Math. J. **14**(1962), 107–133.
6. H. Komatsu, *Fractional powers of operators.* Pacific J. Math. **19**(1966), 285–346.
7. Ju. I. Ljubic, *Conditions for the uniqueness of the solution of Cauchy's abstract problem.* Dokl. Akad. Nauk SSSR **130**(1960), 969–972.
8. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer-Verlag, New York, 1983.
9. E. Sinestrari, *On the solutions of the inhomogeneous evolution equation in Banach space.* Rend. Acc. Naz. Lincei **LXX**, 1981.
10. K. Yosida, *Lectures on semi-group theory and its applications to Cauchy's problem in partial differential equations.* Tata Inst. Fund. Research, 1957.

Departament de Matemàtica Aplicada
Universitat de València
C/Doctor Moliner 50
46100 Burjassot (València)
Spain
email: *sergio.amat@upct.es*
sonia.busquier@uv.es

C. R. Math. Rep. Acad. Sci. Canada Vol. **23** (1), 2001 pp. 28–32

CATALAN'S EQUATION WITH A QUADRATIC EXPONENT

TAUNO METSÄNKYLÄ

Presented by M. Ram Murty, FRSC

RÉSUMÉ. On démontre que l'équation de Catalan $x^p - y^{q^2} = \pm 1$ (p et q premiers distincts impairs) n'a que de solutions sauf si $x \equiv 0$ et $p^{q-1} \equiv 1 \pmod{q^3}$. Ceci est analogue à un résultat récent de Mihăilescu, et la démonstration, basée sur la théorie des corps cyclotomiques, est aussi similaire. On donne de plus un court supplément élémentaire.

1. Introduction. According to Catalan's conjecture, the Diophantine equation

$$x^M - y^N = 1 \quad (xy \neq 0, M > 1, N > 1)$$

has no solution except $x^M = 3^2$, $y^N = 2^3$. For an update survey on the status of this conjecture, see Mignotte's article [6]. In particular, we know that the conjecture is true in case M or N is even, so that M and N can be assumed odd.

The most interesting case, which also would suffice to settle the conjecture in full, is that of the equation $x^p - y^q = 1$, where p and q are different odd primes. On the other hand, Mignotte [5] has proved that if $x^M - y^N = \pm 1$ with $M < N$, then M is a prime and N, if not a prime, is a product of two prime factors, say q_1 and q_2. The aim of the present note is to prove the following result about the case $q_1 = q_2$.

THEOREM. *If* $x^p - y^{q^2} = \pm 1$, *then*

$$x \equiv 0, \quad p^{q-1} \equiv 1 \pmod{q^3}.$$

This result should be compared with a recent theorem by Mihăilescu [7] stating that a solution (x, y) of the equation $x^p - y^q = 1$ is possible only if

$$x \equiv 0, \quad p^{q-1} \equiv 1 \pmod{q^2},$$
$$y \equiv 0, \quad q^{p-1} \equiv 1 \pmod{p^2}.$$

Of course, the latter pair of congruences follows from the former, since the equation may be written as $(-y)^q - (-x)^p = 1$. Previously, Inkeri and others had arrived at these same congruences under various class number conditions.

Received by the editors November 3, 2000.
AMS subject classification: 11D41.

Mihăilescu was able to eliminate these conditions by using the Stickelberger relation in a cyclotomic field. Our proof will be a straightforward adaptation of his argument; in particular, readers not acquainted with [7] can readily reconstruct the proof therein, in fact in a somewhat simplified form.

No pair of odd primes p and q, with $p < q$, is known to satisfy the congruence $p^{q-1} \equiv 1 \pmod{q^3}$. Numerical search has revealed that this congruence is never true for $p < q < 10^6$ [1, p. 1362].

For earlier results about Catalan's equation with composite exponents we refer to Ribenboim's comprehensive monograph [8] and just quote the following:

$$x^p - y^{kq} = \pm 1, k > 1 \implies y \equiv 0, q^{p-1} \equiv 1 \pmod{p^2}.$$

This is an elementary result proved by Hyyrö [2]; his proof can also be found in [8]. Note that the latter congruence (but not the former) also follows from Mihăilescu's theorem.

Hyyrö's result provides an interesting addition to the above theorem. We will conclude the paper by proving that result directly in the special case $k = q$ in a simple (elementary) way, different from Hyyrö's.

2. Beginning the proof of the theorem. It suffices to consider the equation

$$x^p - y^{q^2} = 1. \tag{1}$$

Assume that (x, y) is a solution. First of all, a well-known result tells us that

$$q|x, \quad p|y, \quad \gcd\left(x - 1, \frac{x^p - 1}{x - 1}\right) = p.$$

Let v_p denote the p-adic valuation (with $v_p(p) = 1$). From

$$v_p(x^p - 1) = v_p(y^{q^2}) \geq q^2 > 2$$

it then follows that $v_p(x - 1) > 1$ and $v_p(\frac{x^p-1}{x-1}) = 1$. This leads to the equations

$$\frac{x^p - 1}{x - 1} = pu^{q^2}, \quad x - 1 = p^{q^2-1}a^{q^2}, \quad y = pau, \tag{2}$$

where a and u are coprime integers, $p \nmid u$.

Let ζ be a primitive p-th root of 1. We modify the first equation in (2) by factorizing $x^p - 1$ and p in the cyclotomic field $\mathbb{Q}(\zeta)$, obtaining

$$\prod_{j=1}^{p-1} \frac{x - \zeta^j}{1 - \zeta^j} = u^{q^2}.$$

Since $x \equiv 1 \pmod{p}$, the quotients $\frac{x-\zeta^j}{1-\zeta^j}$ belong to $\mathbb{Z}[\zeta]$, the ring of integers of $\mathbb{Q}(\zeta)$, and the principal ideals they generate are pairwise coprime (*e.g.*, [3, p. 144]). Therefore,

$$\left(\frac{x - \zeta}{1 - \zeta}\right) = \mathcal{A}^{q^2} \tag{3}$$

for some ideal $\mathcal{A}$.

Denote by σ_c the automorphism of $\mathbb{Q}(\zeta)$ defined by $\zeta \mapsto \zeta^c$ $(c = 1, \ldots, p-1)$. By Stickelberger's relation, the element

$$\theta = \sum_{c=1}^{p-1} c\sigma_c^{-1} \in \mathbb{Z}[\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})]$$

annihilates the ideal class group of $\mathbb{Q}(\zeta)$ (see, *e.g.*, [4], where the relation is proved in a special case, sufficient for the present context). Thus we may write

$$\mathcal{A}^\theta = (\gamma), \quad \gamma \in \mathbb{Z}[\zeta].$$

This together with (3) gives the following equation between elements of $\mathbb{Q}(\zeta)$:

$$\frac{(x-\zeta)^\theta}{(1-\zeta)^\theta} = \epsilon\gamma^{q^2}, \tag{4}$$

where ϵ is a unit.

Define $c^{-1} \in \mathbb{Z}$ by the conditions $cc^{-1} \equiv 1 \pmod{p}$, $0 < c^{-1} < p$. We introduce the notation $\lambda = (1-\zeta)^\theta$ and compute

$$\zeta^\theta = \prod_{c=1}^{p-1} (\zeta^{\sigma_c^{-1}})^c = \prod_{c=1}^{p-1} \zeta^{c^{-1}c} = \zeta^{-1},$$

$$\lambda = \zeta^\theta(\zeta^{-1}-1)^\theta = \pm\zeta^{-1}\overline{\lambda}, \tag{5}$$

where the bar denotes complex conjugation.

Since -1 and ζ are q^2-th powers in $\mathbb{Z}[\zeta]$, the equation (4) implies that

$$(1-\zeta^{-1}x)^\theta = \epsilon\lambda\alpha^{q^2}, \quad \alpha \in \mathbb{Z}[\zeta].$$

Now take complex conjugates. By (5), $\overline{\lambda}/\lambda = \pm\zeta$, and by a well-known property of algebraic numbers, $\overline{\epsilon}/\epsilon$ is a root of unity in $\mathbb{Z}[\zeta]$. Hence we get

$$(1-\zeta x)^\theta = \epsilon\lambda\beta^{q^2}, \quad \beta \in \mathbb{Z}[\zeta].$$

These relations give us a key equation containing a difference of two q^2-th powers:

$$(1-\zeta^{-1}x)^\theta - (1-\zeta x)^\theta = \epsilon\lambda(\alpha^{q^2} - \beta^{q^2}).$$

To deal with the left hand side, we write

$$(1-\zeta x)^\theta = \prod_{c=1}^{p-1} (1-\zeta^{c^{-1}}x)^c = 1 - x\sum_{c=1}^{p-1} c\zeta^{c^{-1}} + x^2\xi, \quad \xi \in \mathbb{Z}[\zeta].$$

So our equation becomes, with the notation $\eta = \sum_{c=1}^{p-1} c\zeta^{c^{-1}}$,

$$\epsilon\lambda(\alpha^{q^2} - \beta^{q^2}) = x(\eta - \overline{\eta}) + x^2\omega, \quad \omega \in \mathbb{Z}[\zeta]. \tag{6}$$

3. The proof completed. Let $\mathcal{Q}$ be a prime ideal factor of q in $\mathbb{Q}(\zeta)$. This ideal does not divide (λ), since (λ) is a power of $(1-\zeta)$ which is above p. Hence (6) together with Mihăilescu's result $q^2|x$ gives

$$\alpha^{q^2} \equiv \beta^{q^2} \pmod{\mathcal{Q}^2}.$$

By a general principle, this congruence implies that

$$\alpha^{q^2} \equiv \beta^{q^2} \pmod{\mathcal{Q}^3}. \tag{7}$$

Indeed, first note that $\alpha^{q^f} \equiv \alpha$, $\beta^{q^f} \equiv \beta \pmod{\mathcal{Q}}$, where f denotes the residue field degree of $\mathcal{Q}$. Raise these congruences to the q^2-th power to obtain congruences modulo $\mathcal{Q}^3$, then combine these new congruences with the relation $\alpha^{q^{f+2}} \equiv \beta^{q^{f+2}} \pmod{\mathcal{Q}^3}$.

By (7) and (6),

$$x(\eta - \overline{\eta}) \equiv 0 \pmod{\mathcal{Q}^3}. \tag{8}$$

But

$$\eta - \overline{\eta} = \sum_{c=1}^{p-1} c\zeta^{c^{-1}} - \sum_{c=1}^{p-1} c\zeta^{-c^{-1}} = 2\sum_{c=1}^{p-1} c^{-1}\zeta^c \not\equiv 0 \pmod{q}$$

(note that $\{\zeta, \zeta^2, \ldots, \zeta^{p-1}\}$ is an integral basis of $\mathbb{Q}(\zeta)$). Consequently, we may choose $\mathcal{Q}$ in such a way that $\eta - \overline{\eta} \not\equiv 0 \pmod{\mathcal{Q}}$. With this choice, (8) implies that $x \equiv 0 \pmod{\mathcal{Q}}^3$. Since q is unramified, we finally obtain $x \equiv 0 \pmod{q^3}$.

To prove the latter congruence of the theorem, look at (2) and write

$$x = (p^{q^2-1} - 1)a^{q^2} + a^{q^2} + 1. \tag{9}$$

Since, again by [7], $p^{q-1} \equiv 1 \pmod{q^2}$, we find that $a^{q^2} \equiv -1 \pmod{q^2}$. By the same principle as above, this congruence holds modulo q^3, so that (9) yields $p^{q^2-1} \equiv 1 \pmod{q^3}$. On the other hand, $p^{q(q-1)} \equiv 1 \pmod{q^3}$. These two congruences together imply that $p^{q-1} \equiv 1 \pmod{q^3}$.

4. An elementary supplement. We will show by elementary means that for a possible solution (x, y) of

$$x^p - y^{q^2} = \pm 1$$

one has $y \equiv 0$, $q^{p-1} \equiv 1 \pmod{p^2}$. As mentioned in the introduction, this is a special case of a result in [2]. The following short reasoning provides a direct proof for this particular case.

It suffices to assume that (x, y) is a solution of the equation $x^p - y^{q^2} = 1$. We factorize

$$y^{q^2} + 1 = (y+1) \cdot \frac{y^q + 1}{y+1} \cdot \frac{y^{q^2} + 1}{y^q + 1}$$

and find, iterating a standard argument (*e.g.* Section 2), that the factors on the right hand side are of the following form:

$$y+1 = q^{p-2}a^p, \quad \frac{y^q+1}{y+1} = qu^p, \quad \frac{y^{q^2}+1}{y^q+1} = qv^p \tag{10}$$

with $a, u, v \in \mathbb{Z}$, $q \nmid uv$. Since $p|y$, the last two equations give

$$1 - y \equiv qu^p, \quad 1 \equiv qv^p \pmod{p^2}.$$

Hence $y \equiv q(v^p - u^p) \pmod{p^2}$. This implies that $u^p \equiv v^p \pmod{p}$, thus, $\pmod{p^2}$. Therefore, $y \equiv 0 \pmod{p^2}$.

The first two equations in (10) yield $y^q + 1 = q^{p-1}(au)^p$, so

$$y^q = (q^{p-1} - 1)(au)^p + (au)^p - 1.$$

Again, from $p|y$ we have $(au)^p \equiv 1 \pmod{p}$ and so $\pmod{p^2}$. This gives $q^{p-1} \equiv 1 \pmod{p^2}$.

References

1. R. Ernvall and T. Metsänkylä, *On the p-divisibility of Fermat quotients.* Math. Comp. **66**(1997), 1353–1365.
2. S. Hyyrö, *Über die Gleichung $ax^n - by^n = z$ und das Catalansche Problem.* Ann. Acad. Sci. Fenn. Ser. A I **355**, 1964.
3. K. Inkeri, *On Catalan's conjecture.* J. Number Theory **34**(1990), 142–152.
4. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory.* 2nd edition, Springer-Verlag, New York, 1990.
5. M. Mignotte, *Une remarque sur l'équation de Catalan.* In: Number Theory in Progress (Zakopane, Poland, June 30–July 9, 1997), de Gruyter, Berlin, 1999, 337–340.
6. ———, *Catalan's equation just before 2000.* In: Proc. Turku Symp. on Number Theory (May 31–June 4, 1999), de Gruyter, Berlin, to appear.
7. P. Mihăilescu, *A class number free criterion for Catalan's conjecture.* Manuscript.
8. P. Ribenboim, *Catalan's Conjecture.* Academic Press, New York-London, 1994.

Department of Mathematics
University of Turku
FIN-20014 Turku
Finland
email: taumets@utu.fi