21-373, Algebraic Structures, Department of Mathematical Sciences, Carnegie Mellon University
Fall 2011: (Math Studies Section) Monday, Wednesday, Friday, 10:30 am, Porter Hall 226B.
Luc Tartar, University Professor of Mathematics, Wean Hall 6212, tartar@cmu.edu

22- Wednesday October 19, 2011.

Definition 22.1: If $P = a_0 + a_1x + \ldots + a_nx^n \in R[x]$, the *derivative of* P, noted P' is $P' = a_1 + 2a_2x + \ldots + n\,a_nx^{n-1} \in R[x]$.

Remark 22.2: One has $(P+Q)' = P' + Q'$, and $(P\,Q)' = P'Q + P\,Q'$ for all $P, Q \in R[x]$: if $P = a_0 + a_1x + \ldots + a_nx^n$ and $Q = b_0 + b_1x + \ldots + b_mx^m$, then for $k \geq 1$ the coefficient of x^k in $P\,Q$ is $\sum_{j=0}^k a_jb_{k-j}$, so that the coefficient of x^{k-1} in $(P\,Q)'$ is $k\big(\sum_{j=0}^k a_jb_{k-j}\big) = \sum_{j=0}^k k\,(a_jb_{k-j})$, but for $0 \leq j \leq k$ one has $k\,(a_jb_{k-j}) = (j\,a_j)\,b_{k-j} + a_j\big((k-j)\,b_{k-j}\big)$,[1] and $\sum_{j=0}^k (j\,a_j)\,b_{k-j}$ is the coefficient of x^{k-1} in $P'Q$, while $\sum_{j=0}^k a_j\big((k-j)\,b_{k-j}\big)$ is the coefficient of x^{k-1} in $P\,Q'$.

If R is commutative, or simply if P and P' commute, one deduces by induction on ℓ that $(P^\ell)' = \ell\,P^{\ell-1}P'$ for $\ell \geq 2$: the preceding case with $Q = P$ gives $(P^2)' = P'P + P\,P'$, which is $2P\,P'$ since P and P' commute; then for $\ell > 2$ one uses $Q = P^{\ell-1}$, so that by the induction hypothesis one has $Q' = (\ell-1)\,P^{\ell-2}P'$, hence $(P^\ell)' = (P\,Q)' = P'Q + P\,Q' = P'P^{\ell-1} + P\,(\ell-1)P^{\ell-2}P'$, which is $\ell\,P^{\ell-1}P'$ since P and P' commute.

If P is a constant, i.e. $P = a_0$, then $P' = 0$, but in some rings it may happen that a non-constant polynomial has a zero derivative: for example, if R is an integral domain with characteristic p (necessarily a prime), then $P' = 0$ means $j\,a_j = 0$ for all $j \geq 0$, but since for $a_j \neq 0$ it implies that j is a multiple of the characteristic p, one deduces that $P' = 0$ if and only if P is a polynomial in x^p, i.e. it has the form $\sum_{\ell=0}^m b_\ell x^{\ell\,p}$.

Lemma 22.3: If R is a commutative unital ring, then α is a multiple root of $P \in R[x]$ if and only if $P(\alpha) = 0$ and $P'(\alpha) = 0$.
Proof: If α is a root of multiplicity $k \geq 2$, one has $P = (x-\alpha)^kQ$ (with $Q(\alpha) \neq 0$), so that $P' = k\,(x-\alpha)^{k-1}Q + (x-\alpha)^kQ'$, hence $P(\alpha) = 0$ and $P'(\alpha) = 0$. Conversely, if $P(\alpha) = 0$ one has $P = (x-\alpha)\,Q_1$, so that $P' = Q_1 + (x-\alpha)\,Q_1'$, hence $P'(\alpha) = Q_1(\alpha)$; if $P(\alpha) = 0$ and $P'(\alpha) = 0$, one deduces that $Q_1(\alpha) = 0$, so that $Q_1 = (x-\alpha)\,Q_2$, hence $P = (x-\alpha)^2Q_2$, i.e. α is a multiple root of P (of multiplicity $k \geq 2$).

Remark 22.4: If R is a commutative unital ring and α is a root of multiplicity $k \geq 2$, then $P = (x-\alpha)^kQ$ with $Q(\alpha) \neq 0$, so that $P' = k\,(x-\alpha)^{k-1}Q + (x-\alpha)^kQ' = (x-\alpha)^{k-1}Q_1$ with $Q_1 = k\,Q + (x-\alpha)Q'$, hence α is a root of multiplicity at least $k-1$ of P'. Since $Q_1(\alpha) = k\,Q(\alpha)$, it may happen that $k\,Q(\alpha) = 0$ although $Q(\alpha) \neq 0$: if R is an integral domain, it means that R has a finite characteristic, which must be a prime p, and k is a multiple of p.

If α is a root of multiplicity $k \geq 3$, then $P(\alpha) = 0$ and the successive derivatives of P up to order $k-1$ are 0 at α. If R is an integral domain of characteristic p, the converse holds if $k \leq p$, and the proof is by induction on k: since $P(\alpha) = P'(\alpha) = 0$ implies $P = (x-\alpha)^2Q$, it is true for $k = 2$; assume that $k \geq 3$ (so that $p \geq 3$) and that it has been proved up to $k-1$, so that $P = (x-\alpha)^{k-1}Q$, and then the derivative of order $k-1$ has a term in $(k-1)!Q$ and all other terms have $x-\alpha$ as a factor, so that the $(k-1)$th derivative of P evaluated at α is $(k-1)!Q(\alpha)$, and since $(k-1)!$ is not a multiple of p and the $(k-1)$th derivative of P evaluated at α is 0 by hypothesis, one deduces that $Q(\alpha) = 0$, so that $Q = (x-\alpha)\,Q_1$ and $P = (x-\alpha)^kQ_1$.

One has almost used Leibniz's formula giving the kth derivative of a product,[2] that if one denotes $P^{(j)}$ the jth derivative of P, so that $P^{(1)}$ means P' and $P^{(0)}$ means P, then Leibniz's formula is that $(P\,Q)^{(k)} = \sum_{j=0}^k \binom{k}{j} P^{(j)}Q^{(k-j)}$, and it was proved for $k = 1$, and the proof is by induction on k, and it follows easily by using the properties of binomial coefficients.

[1] Although R may not be commutative, for $a, b \in R$ and $\ell \in \mathbb{Z}$, one has $\ell\,(a\,b) = (\ell\,a)\,b = a\,(\ell\,b)$: for $\ell > 0$, it is about adding ℓ copies of $a\,b$, and the formula follows from distributivity; for $\ell < 0$, it is a consequence of $-(a\,b) = (-a)\,b = a\,(-b)$, which is about having $0 = (a\,b) + (-a)\,b = (a\,b) + a\,(-b)$, which again follows from distributivity.

[2] Gottfried Wilhelm von Leibniz, German mathematician, 1646–1716. He worked in Frankfurt, in Mainz, Germany, in Paris, France, and in Hanover, Germany, but never in an academic position.

Remark 22.5: If R is a commutative unital ring, one can prove Taylor's expansion for polynomials: the usual formula taught in analysis is $P(x+h) = P(x) + P'(x)\,h + \frac{P''(x)\,h^2}{2!} + \ldots$, but for a polynomial the sum is finite, since $P^{(n+1)} = 0$ if P has degree n; since a term $\frac{P^{(j)}(x)\,h^j}{j!}$ appears, which may not make sense in some rings because one cannot always divide elements of R by $j!$, one should pay attention to the notation. If $P = x^k$ then $P^{(j)} = k \cdots (k+1-j)\,x^{k-j}$ if $j \le k$ and 0 if $j > k$, so that $\frac{P^{(j)}(x)\,h^j}{j!} = \binom{k}{j} x^{k-j} h^j$ and since $\binom{k}{j}$ is an integer, one never divides an element of R by an integer. Then the proof is obtained by writing the binomial formula for $(x+h)^k$, which one multiplies by a_k before summing in k.

In particular, if $P \in \mathbb{Z}[x]$, then one has observed that $\frac{P^{(j)}}{j!} \in \mathbb{Z}[x]$, so that if $a, h \in \mathbb{Z}$ one has $P(a+h) = P(a) + P'(a)\,h + \sum_{j=2}^{deg(P)} c_j h^j$, with $c_j \in \mathbb{Z}$ for $j = 2, \ldots, deg(P)$. In the following application, if p is a prime and h is a multiple of p^m (with $m \ge 1$), then $P(a+h) = P(a) + P'(a)\,h \pmod{p^{2m}}$.

Remark 22.6: If $P \in \mathbb{Z}[x]$ and $f(N)$ is the number of solutions in $\mathbb{Z}_N$ of $P(x) = 0 \pmod N$, then f is a multiplicative function by the Chinese remainder theorem, so that one must just wonder how many solutions there is modulo p^k for a prime p and an integer $k \ge 1$. If a_1 is a solution of $P(a_1) = 0 \pmod p$ and one has $P'(a_1) \ne 0 \pmod p$, then one can construct a sequence $a_2, \ldots, a_k$ such that $a_j = a_{j-1} \pmod{p^{j-1}}$ for $j = 2, \ldots, k$ and $P(a_k) = 0 \pmod{p^k}$, so that $P'(a_k) = P'(a_1) \ne 0 \pmod p$. For example, one looks for $a_2 = a_1 + b_1 p$, and one uses the Taylor expansion, which gives $P(a_2) = P(a_1) + P'(a_1)\,b_1 p + \ldots$ where the terms not written contain $b_1 p$ to a power ≥ 2, so that $P(a_2) = P(a_1) + P'(a_1)\,b_1 p \pmod{p^2}$; since $P(a_1) = 0 \pmod p$, one has $P(a_1) = c_1 p \pmod{p^2}$ for some $c_1 \in \mathbb{Z}$, so that $P(a_2) = 0 \pmod{p^2}$ is equivalent to $c_1 + P'(a_1)\,b_1 = 0 \pmod p$, which has a unique solution b_1 modulo p, because $P'(a_1)$ has an inverse modulo p.

Essentially, it is the same idea used in a method of NEWTON for solving equations, which is now known as the *Newton–Raphson method*:[3] if f is a differentiable function on $\mathbb{R}$ and $f'(x_0) \ne 0$, a guess for a solution of $f(x) = 0$ is to replace $f(x) = 0$ by $f(x_0) + f'(x_0)\,(x - x_0) = 0$, so that one takes $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$, and the iterative method $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ converges under some condition.[4]

HENSEL must have thought of this analogy when he invented the p-adic numbers $\mathbb{Q}_p$ in 1897,[5] by using a different metric on $\mathbb{Q}$ (hence on $\mathbb{Z}$) than the usual one, so that the sequence a_k constructed converges to an element of $\mathbb{Q}_p$. For example, if $P = x^2 - 2$ and $p = 7$, then $P(3) = 7 = 0 \pmod 7$ and $P'(3) = 6 \ne 0 \pmod 7$, so that the method creates a sequence of integers, which converges in Q_7 to a root of P, but is this root $+\sqrt{2}$ or $-\sqrt{2}$? For example, $1 + 2 + \ldots + 2^n + \ldots$ converges in $\mathbb{Q}_2$, to -1, and it is quite similar to what will be shown later for formal power series that $(1-x)^{-1} = 1 + x + \ldots + x^n + \ldots$, but it then must be explained in what sense one may take $x = 2$ in this formula.

Definition 22.7: A field F is said to be *algebraically closed* if every non-constant polynomial has a root, hence a polynomial $P \in F[x]$ of degree $n \ge 1$ can be written as $a_n\,(x-\alpha_1)\cdots(x-\alpha_n)$ for some $\alpha_1, \ldots, \alpha_n \in F$.

Remark 22.8: It will be shown that $\mathbb{C}$ is algebraically closed, but $\mathbb{R}$ is obviously not since x^2+1 has no root. $P = (x^2+1)\,(x^2+2)$ has no roots, but it can be "reduced", because $P = P_1 P_2$ with $P_1 = x^2+1$ and $P_2 = x^2+2$, so that one will need a notion of irreducibility for polynomials in $R[x]$, but the definitions will actually be given for general rings. Irreducible polynomials of degree ≥ 2 in $\mathbb{R}[x]$ must have degree 2, and $x^2 + A\,x + B$ is irreducible if and only if $A^2 < 4B$, but the situation is different for $\mathbb{Q}[x]$ and for every $m \ge 2$ there is an irreducible polynomial in $\mathbb{Q}[x]$ of degree m.

---

[3] Joseph RAPHSON, English mathematician, c. 1648–1715. The Newton–Raphson method is partly named after him: he published it in 1690, and it is simpler than the method that NEWTON wrote in 1671, but which was only published in 1736.

[4] For example, if $|f'(x)| \ge \frac{1}{2}\,|f'(x_0)|$ and $|f''(x)| \le M$ on $I = [x_0 - a, x_0 + a]$, one deduces that $|x_{n+1} - x_n| \le \frac{2|f(x_n)|}{|f'(x_0)|}$ and $|f(x_{n+1})| \le \frac{M\,|x_{n+1}-x_n|^2}{2} \le \frac{2M\,|f(x_n)|^2}{|f'(x_0)|^2}$ as long as the points stay in I; if $2M\,|f(x_0)| \le \theta\,|f'(x_0)|^2$ with $\theta < 1$, then $|f(x_n)| \le \theta^{2^n-1}|f(x_0)|$ as long as the points stay in I, which is the case if $2|f(x_0)| \le (1-\theta)\,a\,|f'(x_0)|$.

[5] Kurt Wilhelm Sebastian HENSEL, German mathematician, 1861–1941. He worked in Marburg, Germany. Hensel's lemma is named after him.