A Very Brief Introduction to Differential Algebra

We'll consider differential fields (resp. rings), which is are fields (resp. rings) equipped with one or more additional unary operations (derivations), each of which satisfy two axioms:

$$D(a+b) = Da + Db$$

$$D(ab) = (Da)b + a(Db)$$

I'll often write a derivation as either a subscripted delta (δ_x instead of D), a subscripted variable (f_x instead of Df or $\delta_x f$) or an apostrophe (if only one derivation is being considered, like in a lemma).

Definition An element t of a differential field K is *monomial* over a subfield $k \subset K$ w.r.t a derivation D if t is transcendental over k and $Dt \in k[t]$. Bronstein p. 91.

In simpler terms, the derivation of a monomial element is a polynomial. Most common transcendental field extensions are monomial. For example, if $t = \ln x$, then $t_x = 1/x$, which is a polynomial in $\mathbf{C}(x)[t]$, so $\ln x$ is monomial over rational functions in x with respect to derivation by x, as is $\exp x$, since if $t = \exp x$, then $t_x = t \in \mathbf{C}(x)[t]$. However, $t = \sqrt{x}$ is not monomial, even though $t_x = t/2x \in \mathbf{C}(x)[t]$, since $\sqrt{x}$ is not transcendental (it's algebraic), nor is $t = \exp(\sqrt{x})$ monomial over $\mathbf{C}(x)$ w.r.t. δ_x, since $t_x = \sqrt{x}\exp(\sqrt{x})/(2x) \notin \mathbf{C}(x)[t]$. However, $t = \exp(\sqrt{x})$ is monomial over $\mathbf{C}(x, \sqrt{x})$

Definition Let k be a differential field extended by a single monomial t to form a differential ring $k[t]$. A polynomial $f\ ink[t]$ is *normal* with respect to a derivation D if $\gcd(f, Df) = 1$ and *special* if $\gcd(f, Df) = f$. Bronstein p. 92.

For an irreducible polynomial, these are the only two cases. Reducible polynomials can be factored into a normal and a special component. If Dt is constant (the ordinary case), all irreducible polynomials in $k[t]$ are normal.

Lemma An irreducible normal factor of a polynomial with multiplicity p and non-zero derivative appears in the polynomial's derivative with multiplicity $p-1$. **Proof.** Write the polynomial as $f^p q$, where $\mathrm{GCD(f, q)} = 1$. Its derivative is:

$$pf^{p-1}f'q + f^p q' = f^{p-1}(pf'q + fq')$$

f doesn't divide f' (the factor is *normal*), and both f' and q are not zero, so f doesn't divide either $pf'q$ or $(pf'q + fq')$.

Lemma An irreducible factor of a polynomial with multiplicity p and zero derivative appears in the polynomial's derivative with multiplicity at least p, unless the derivative is zero. Write the polynomial as $f^p q$, where $\mathrm{GCD(f, q)} = 1$. Its derivative is $f^p q'$, and while f doesn't factor q, it might factor q', or q' might be zero.

Lemma An irreducible normal factor of a polynomial with multiplicity p appears in the polynomial's derivative with multiplicity at least $p-1$, unless the derivative is zero.

A Prototype Problem

Let's consider the one-dimensional heat equation:

$$\frac{\delta^2}{\delta x^2}\Psi = \frac{\delta}{\delta t}\Psi$$

We'll use two derivations, δ_x and δ_t, so I'll often write the previous equation as $\delta_x^2\Psi = \delta_t\Psi$ or $\Psi_{xx} = \Psi_t$.

What field do we use? Good question!

Let's start with the rational functions in x and t with complex coefficients: $\mathbf{C}(x,t)$. Are there any solutions to $\delta_x^2\Psi = \delta_t\Psi$ in this differential field?

Everything in $\mathbf{C}(x,t)$ can be written uniquely as a ratio of polynomials in x and t with GCD 1, so let's consider an arbitrary field element $\Psi = n/d$ with $n, d \in \mathbf{C}[x,t]$. Applying our differential operators (I now write $\delta_x n$ as n_x):

$$\delta_x \frac{n}{d} = \frac{n_x d - n d_x}{d^2}$$

$$\delta_x^2 \frac{n}{d} = \frac{n_{xx} d^2 - n d_{xx} d - 2 n_x d_x d + 2 n d_x^2}{d^3}$$

$$\delta_t \frac{n}{d} = \frac{n_t d - n d_t}{d^2}$$

If $\Psi = \frac{n}{d}$ satisfies $\delta_x^2\Psi = \delta_t\Psi$, then we must have:

$$n_{xx} d^2 - n d_{xx} d - 2 n_x d_x d + 2 n d_x^2 = n_t d^2 - n d d_t$$

Rearranging:

$$2 n d_x^2 - n d_{xx} d = n_t d^2 - n d d_t - n_{xx} d^2 + 2 n_x d_x d$$

Consider f, a single irreducible factor of d, appearing with multiplicity $p >= 2$, and $f_x \neq 0$, so δ_x drives its order down by one. f appears in d_x with power $p-1$ and in d_{xx} with power $p-2$, and doesn't appear in n at all, since $\mathrm{GCD(d, n)} = 1$. Thus, the terms on the LHS have f-order exactly $2p-2$, while all of the terms on the RHS have f-order at least $2p-1$ ($2p$, $2p-1$, $2p$, and $2p-1$, plus possible factors in n's derivatives, plus another factor in d_t if f does not involve t):

$$2 d_x^2 - d_{xx} d = m f^{2p-1}$$

Now let's expand $d = f^p q$:

$$d = f^p q \qquad d_x = f^p q_x + p f^{p-1} f_x q$$
$$d_x^2 = f^{2p} q_x^2 + 2p f^{2p-1} f_x q q_x + p^2 f^{2p-2} f_x^2 q^2$$
$$d_{xx} = f^p q_{xx} + 2p f^{p-1} f_x q_x + p f^{p-1} f_{xx} q + (p^2 - p) f^{p-2} f_x^2 q$$

Substituting into $2 d_x^2 - d_{xx} d = m f^{2p-1}$ and cancelling f^{2p-2}:

$$-f^2 q q_{xx} + 2 f^2 q_x^2 + 2p f f_x q q_x - p f f_{xx} q^2 + (p^2 + p) f_x^2 q^2 = m f$$

$$(p^2+p)f_x^2q^2 = mf + f^2qq_{xx} - 2f^2q_x^2 - 2pff_xqq_x + pff_{xx}q^2$$

f factors the RHS, but $\mathrm{GCD(f,q)} = 1$ and $\mathrm{GCD(f,f_x)} = 1$ by construction, so the LHS must be zero, $p^2+p=0$, implying that p is 0 or -1, both of which contradict the assumption that $p \geq 2$. So d can have no irreducible factors that involve x and appear with power greater than unity.

We still have to consider the case of a square-free irreducible factor involving x. In this case, we can rearrange like this:

$$2nd_x^2 = n_td^2 - ndd_t - n_{xx}d^2 + nd_{xx}d + 2n_xd_xd$$

Note that d divides the RHS, and a purely square-free factor of d involving x would not appear in d_x or n, so the LHS must be zero, which can only occur if either $n=0$ or if $d_x = 0$, which contradicts the assumption that d's factor involves x.

To summarize, we've concluded that either $n = 0$ or that $d_x = 0$, and thus be d must be a polynomial in $\mathbf{C}[t]$. In this case, the derivatives d_x and d_{xx} vanish from the original equation and it becomes:

$$n_{xx}d^2 = n_td^2 - ndd_t$$

Cancelling and rearranging:

$$nd_t = n_td - n_{xx}d$$

d can't divide its own derivative, and $\mathrm{GCD(n,d)}$ is 1 by hypothesis, so this equation can only hold if d_t is zero, so d must be a constant (LHS), and $n_t = n_{xx}$ (RHS). Analysing this equation using the techniques later in this paper [1] shows that $n_t = n_{xx}$ admits an infinity of solutions in $\mathbf{C}[x,t]$, with an infinite basis set $\{B_0, B_1, \ldots\}$:

$$B_{2i} = \sum_{j=0}^{i} \frac{(2i)!}{(2i-2j)!j!}x^{2i-2j}t^j \qquad B_{2i+1} = \sum_{j=0}^{i} \frac{(2i+1)!}{(2i-2j+1)!j!}x^{2i-2j+1}t^j$$

$$\{1, x, x^2+2t, x^3+6xt, x^4+12x^2t+12t^2, x^5+20x^3t+60xt^2, \ldots\}$$

So the only solutions to $\delta_x^2\Psi = \delta_t\Psi$ in $\mathbf{C}(x,t)$ are linear combinations of these polynomials.

Incidentally, what happens if we apply the irreducible factor analysis to the numerator equation $n_{xx} = n_t$ with $n = f^pq$? We find that p must be one, and an additional differential condition, namely, that $f_{xx}q + 2f_xq_x - f_tq$ be a multiple of f. Notice that $B_5 = x^5+20x^3t+60xt^2 = x(x^2+(10+2\sqrt{10})t)(x^2+(10-2\sqrt{10})t)$, [2] and that any selection of one of these factors as f (and the product of the remaining two as q) satisfies the differential condition. While interesting as the logical conclusion of the irreducible factor solution technique, this information doesn't seem to lead directly to a solution. Open question: would an elimination analysis with differential Groebner bases turn up additional relations on the irreducible factors?

Can we find a different field that contains more solutions to this PDE?

An extension field is the obvious choice.

It's a basic theorem in differential algebra that differentials extend in a single unique way into an algebraic extension (Bronstein Theorem 3.2.3), while transcendental extensions require only that the differential be specified for the primitive element that extends the field (Bronstein Theorem 3.2.2).

Let's extend $\mathbf{C}(x,t)$ by r (an algebraic) and z (a transcendental). I'll set $r^2 = t$, so its differentials are:

$$\delta_x r = 0 \qquad \delta_t r = \frac{1}{2r}$$

[1] see *Solving Polynomial Differential Equations*

[2] `expand(x*(x**2+(10 + 2 *sqrt(10))*t)*(x**2+(10 - 2*sqrt( 10))*t));`

Since z is transcendental, I can pick (almost) anything for its differentials. For reasons that will become apparent later, let's try

$$\delta_x z = -\frac{x}{2t} z \qquad \delta_t z = \frac{x^2}{4t^2} z$$

Now, is there a solution to $\delta_x^2 \Psi = \delta_t \Psi$ in $\mathbf{C}(x, t, r, z)$? You bet! Let's try $\Psi = z/r$.

$$\delta_x \Psi = -\frac{xz}{2rt}$$
$$\delta_x^2 \Psi = -\frac{2zt - x^2 z}{4rt^2}$$
$$\delta_t \Psi = \frac{x^2 z - 2zt}{4rt^2} = \delta_x^2 \Psi$$

So, what's the point? That I can concoct some screwball extension that solves a PDE? Actually, z is just an exponential that solves $\frac{\delta z}{\delta y} = z$ for $y = -x^2/(4t)$, and those weird differentials are simply the result of applying the chain rule to compute $\frac{\delta z}{\delta x}$ and $\frac{\delta z}{\delta t}$. Analytically, we would write this solution as:

$$\Psi(x, t) = \frac{e^{-\frac{x^2}{4t}}}{\sqrt{t}}$$

Later, I'll use differential algebra to show that this is the only solution (times an arbitrary constant) in $\mathbf{C}(x, t, r, z) \backslash \mathbf{C}(x, t)$.

The two main extension types I'm interested in are algebraic extensions (since they've been so heavily studied), and "holonomic" extensions of the form $\frac{\delta z}{\delta r} = f$ for r and f arbitrary field elements. This models first order semilinear ODEs.

Then we can start asking questions like, given a finite number of algebraic and holonomic extensions, can we find solutions to a given PDE? This corresponds to asking whether we can solve a PDE by breaking it down into algebraic functions and ODEs, so would subsume separation of variables as a special case.

How can we deal with boundary conditions? Well, once we've found all solutions to a PDE in a particular extension field, we can then restrict them to the boundary (say, $t = 1$, in the example above), and ask whether they form a basis for whatever function space (typically L^2) our boundary condition exists in. Since exponentials (as we saw above) are a simple case of the holonomic extension, Fourier analysis would be subsumed as a special case.

It would be nice to have theorems telling us what conditions are needed to get a basis set in some extension field, and of course how to setup that extension and compute the basis elements.

Is there any hope of proving such theorems? Well, since we're working with algebra, we've got the machinery of algebraic geometry available (not something you typically expect in PDE theory). Reducing modulo p, for example, is definitely in the mix, though this suggestion does has the flavour of an Army captain musing about tactical nukes while contemplating an enemy bunker.

If successful, this program should result in a solution technique for PDEs that would generalize both Fourier analysis and separation of variables, so I think this is quite promising!

Solving $\Psi_{xx} = \Psi_t$ in $\mathbf{C}(x, t, z)$

What about the intermediate field $\mathbf{C}(x, t, z)$? Remember

$$z_x = -\frac{x}{2t}z \qquad z_t = \frac{x^2}{4t^2}z$$

Lemma For all $f \in \mathbf{C}[x, t, z], f \notin \mathbf{C}[x, t] \implies f_x \neq 0$. Consider $f = \sum f_i z^i$. Then:

$$f_x = \sum \left(f_{ix} - \frac{ix}{2t} f_i \right) z^i$$

If $f_x = 0$, then each of these terms must be 0. The x-degree of f_{ix} is one less than the x-degree of f_i, but the x-degree of xf_i is one greater than that, so either i or f_i is zero. So all of the f_i are zero except f_0, but then f would be in $\mathbf{C}[x, t]$. QED.

Returning to the original problems, we're once again led to consider the equation:

$$n_{xx}d^2 - nd_{xx}d - 2n_x d_x d + 2nd_x^2 = n_t d^2 - ndd_t$$

This time, we'll split d into its normal and special components, with respect to δ_x:

$$d = d_s d_n$$

This time, we consider an irreducible normal factor f of d_n with non-zero x-derivative, and use the same logic as before to conclude that f_x is zero, so d_{nx} is zero. By the previous lemma, d_n must be in $\mathbf{C}[t]$.

Next we attack d_s, using Bronstein's Theorem 5.1.2 (p. 130) which says that d_s has to be a power of z.

$$d = z^a f \qquad f \in \mathbf{C}[t]$$

$$d_x = az^a f \left(-\frac{x}{2t} \right)$$

$$d_{xx} = a^2 z^a f \left(-\frac{x}{2t} \right)^2 - az^a f \left(\frac{1}{2t} \right)$$

$$d_t = z^a \left(a\frac{x^2}{4t^2} f + f_t \right)$$

Our polynomial equation (after cancelling f and z^{2a}, and clearing the denominator) becomes:

$$(a^2 + a)x^2 fn + 4axftn_x + 4ft^2 n_{xx} - 4ft^2 n_t + 2aftn + 4f_t t^2 n = 0$$

This implies that $4f_t t^2 n$ must be a multiple of f, and since f is irreducible and normal, and $\gcd(f, n) = 1$, this can only occur if $f = t$.

So now let's consider a denominator of the form $z^p t^a f$, where f has no t factor:

$$(p^2 + p)x^2 fn + 4pxtfn_x + 4t^2 fn_{xx} - 4t^2 fn_t + 4t^2 f_t n + (2p + 4a)tfn = 0$$

Again the minimum power of f appears multiplied by $4f_t t^2 n$, which is a contradiction (since there's no t factor in f), implying that f_t is zero.

So we've reduced the denominator to the form z^a. Expanding the polynomial equation we get:

$$(a^2+a)x^2n + 4axtn_x + 4t^2n_{xx} - 4t^2n_t + 2atn = 0$$

Now we look at the powers of t and conclude that $(a^2+a)x^2n$ must be a multiple of t, implying that the numerator must include a t factor. Setting the numerator to be $t^b n$, and leaving the denominator as z^a, we obtain:

$$(a^2+a)x^2t^bn + 4axt^{b+1}n_x + 4t^{b+2}n_{xx} - 4t^{b+2}n_t + (2a-4b)t^{b+1}n = 0$$

So, $(a^2+a)x^2n$ must be zero or a multiple of t, which implies that a^2+a must be zero, so a must be 0 or -1. Since it can't be -1, it must be zero, and the denominator is therefore trivial.

Now let's look at the numerator, assume that it has the form $\sum n_i z^i$, where the n_i are polynomials in $\mathbf{C}[x,t]$. Expanding out $n_{xx} - n_t = 0$, we obtain:

$$\sum_i \left[(i^2-i)x^2n_i - 4ixtn_{ix} + 4t^2n_{ixx} - 4t^2n_{it} - 2itn_i\right] z^i = 0$$

Looking at the powers of t, we conclude that $(i^2-i)x^2n_i$ must be a multiple of t, so we now try a numerator of the form $\sum n_i t^a z^i$, and again conclude that $(i^2-i)x^2n_i$ must be a multiple of t, which implies that $i^2-i=0$ and thus i is either zero or one.

Our numerator equation for $i=1$ assumes the form:

$$-2n_x x - n + 2tn_{xx} = 2tn_t$$

Assuming that n has the form $\sum a_b t^b$, where the a_b are polynomials in $\mathbf{C}[x]$, we expand this into:

$$-2\sum xn_{bx}t^b - \sum a_b t^b + 2\sum a_{b-1xx}t^b = 2\sum b a_b t^b$$

The $b=0$ term becomes $-2xa_{0x} - a_0 = 0$. Since a_0 is a polynomial in $\mathbf{C}[x]$, this equation implies that each of its coefficients would have to be at least twice itself, which is impossible, so $a_0 = 0$. Likewise, $a_0 = 0$ implies that a_1's equation is $-2xa_{1x} - a_1 = 2a_1$, which is likewise impossible. By induction, $a_b = 0$ implies that $a_{b+1} = 0$, so all of the a_b are zero and there is no z ($i=1$) term in the numerator.

The numerator thus reduces to the $i=0$ term, for which $n_{xx} = n_t$ and we have the same solution set as before.

Solving $\Psi_{xx} = \Psi_t$ in $\mathbf{C}(x, t, r, z)$

We should be able to find more solutions in $\mathbf{C}(x, t, r, z)$, where $r^2 = t$.

Let's study this field as the fraction field of the ring $\mathbf{C}(x, t, r)[z]$. This enables us to cleanly analyse z as a monomial extension. Any polynomial in z will have a non-zero x-derivative (previous lemma), and we've already dealt with the denominator cases for normal factors with $f_x \neq 0$, so we once again consider a special denominator of the form z^a, and a numerator of the form $t^b n + t^c n_r r$ where both n and n_r are in $\mathbf{C}(x, t)$. This leads to a pair of equations:

$$(a^2 + a)x^2 t^b n + 4axt^{b+1} n_x + 4t^{b+2} n_{xx} - 4t^{b+2} n_t + (2a - 4b)t^{b+1} n = 0$$

$$(a^2 + a)x^2 t^c n_r + 4axt^{c+1} n_{rx} + 4t^{c+2} n_{rxx} - 4t^{c+2} n_{rt} + (2a - 4c - 2)t^{c+1} n_r = 0$$

Since t can not factor either n or n_r, $a^2 + a$ must be zero, again leading to the conclusion that z does not appear in the denominator.

The numerator analysis, based on $\sum n_i t^a z^i$ is unchanged, leading again to the conclusion that i is either zero or one. The two resulting equations are:

$$\begin{aligned} n_{xx} &= n_t && (i = 0) \\ -2n_x x - n + 2tn_{xx} &= 2tn_t && (i = 1) \end{aligned}$$

This time our solution space is $\mathbf{C}(x, t, r)$, however. Assuming n has the form $n + n_r r$ (remember that no fraction field is needed for algebraic extensions), we obtain four equations to be solved in $\mathbf{C}(x, t)$:

$$\begin{aligned} n_{xx} &= n_t && (i = 0) \\ n_r &= 2tn_{rxx} - 2tn_{rt} && (i = 0; r) \\ -2xn_x + 2tn_{xx} - 2tn_t - n &= 0 && (i = 1) \\ -2xn_{rx} + 2tn_{rxx} - 2tn_{rt} - 2n_r &= 0 && (i = 1; r) \end{aligned}$$

These four rational equations convert to four polynomial equations:

$$\begin{gathered} -ndd_{xx} + ndd_t + 2nd_x^2 - 2n_x dd_x + n_{xx} d^2 - n_t d^2 = 0 \\ -2tndd_{xx} + 2tndd_t + 4tnd_x^2 - 4tn_x dd_x + 2tn_{xx} d^2 - 2tn_t d^2 - nd^2 = 0 \\ 2xndd_x - 2xn_x d^2 - 2tndd_{xx} + 2tndd_t + 4tnd_x^2 - 4tn_x dd_x + 2tn_{xx} d^2 - 2tn_t d^2 - nd^2 = 0 \\ 2xndd_x - 2xn_x d^2 - 2tndd_{xx} + 2tndd_t + 4tnd_x^2 - 4tn_x dd_x + 2tn_{xx} d^2 - 2tn_t d^2 - 2nd^2 = 0 \end{gathered}$$

The first one we've analysed already; its solutions are all in $\mathbf{C}(x, t)$.

Analyzing the remaining three using the techniques already discussed (irreducible factors) reveal that their denominator's x-derivatives are zero. Setting d_x and d_{xx} to zero, we obtain:

$$\begin{gathered} 2tnd_t + 2tn_{xx} d - 2tn_t d - nd = 0 \\ -2xn_x d + 2tnd_t + 2tn_{xx} d - 2tn_t d - nd = 0 \\ -2xn_x d + 2tnd_t + 2tn_{xx} d - 2tn_t d - 2nd = 0 \end{gathered}$$

Considering the next one, we see that $2tnd_t$ must be a multiple of d, so we're led to consider t as a factor of d, i.e, $d = t^a q$:

$$2ntq_t + 2tn_{xx}q - 2tn_tq + (2a-1)nq = 0$$

which requires either $2ntq_t$ to be a multiple of q (impossible), or $a = \frac{1}{2}$, so the second equation has no solution.

The next-to-last equation also requires either $2ntq_t$ to be a multiple of q or $a = \frac{1}{2}$, so it's also impossible.

The last equation becomes:

$$-2xn_xq + 2ntq_t + 2tn_{xx}q - 2tn_tq + (2a-2)nq = 0$$

At first, this appears to require $2ntq_t$ to be a multiple of q, but there is another possibility! If n and q are both constant, then the equation reduces to:

$$(2a-2)nq = 0$$

which can be satisfied if $a = 1$, which means $d = tq$.

In short, we've found another solution: zr/t, as expected from analysis. This is the only solution in $\mathbf{C}(x,t,r,z)\backslash\mathbf{C}(x,t)$.

Solving $\Psi_{xx} = \Psi_t$ with a free exponential

In previous sections, I took $z = \exp(-\frac{x^2}{4t})$, but this requires too cleaver a guess to be useful as a general technique. What happens if we relax this relationship, assume only that z is an exponential of some element in $\mathbf{C}(x,t)$, and look for solutions in $\mathbf{C}(x,t,z)$?

We begin our analysis as before by concluding that there can be no normal irreducible factors in the solution's denominator, so next let's consider a solution of the form $\frac{n}{z^p}$, with an exponential of the form $z = \exp(\frac{n_e}{f^a d_e})$. Expanding out our equation, we find that the only term not involving f is $p^2a^2f_x^2n_e^2d_e^2n$. f is coprime to n_e and d_e, and let's add the assumption that it's also coprime to n. If p and a are both non-zero, we must have f_x zero, so adding that assumption produces an equation of the form:

$$[\cdots]f^{2a+1} + [\cdots]f^{a+1} + [\cdots]f^a + [\cdots]f = 0$$

Here, for the first time, we find exponents that are *incomparable*, since we can't separate the final two terms without additional assumptions on the value of a. Assuming that a is at least two, the f terms are:

$$p^2n_e^2d_{ex}^2n - 2p^2n_en_{ex}d_ed_{ex}n + p^2n_{ex}^2d_e^2n = p^2(n_ed_{ex} - n_{ex}d_e)^2n = p^2(n_ed_e)_x^2n$$

which must be a multiple of f (or zero), so we integrate and conclude that $n_ed_e = f^bq + c$, where $c_x = 0$. This makes $p^2(n_ed_e)_x^2n$ a multiple of f^{2b-2}. If we now assume that $2b-1 > a$, then the f^a term dominates and $paf_tn_ed_e^3n$ must be a multiple of f, which is impossible. So $a >= 2$ and $2b+1 > a$ (along with the coprimality assumptions) is an impossible combination, and we can begin to build a *case tree* of possibilities.

What about $a >= 2$ and $2b-1 = a$? Then we've insured that the f and f^a terms have equal powers of f, so we can move on and conclude that the two terms together must also be a multiple of f (to cancel the f^{a+1} term):

$$p(n_ed_e)_x^2 - af_tn_ed_e^3 = rf$$

We might think that r and f must be coprime, since f^a must be multiplied by exactly f to match f^{a+1}, but this neglects the possibility that the f^{a+1} coefficient might itself be a multiple of f, in fact has to be to match f^{2a+1}.

Finally, we consider the f^{a+1} term and conclude that it also has to be a multiple of f, in fact exactly a multiple of f^a in order to cancel the f^{2a+1}:

$$2n_ed_ed_{ex}n_x + n_ed_ed_{exx}n - n_ed_ed_{et}n - 2n_ed_{ex}^2n - 2n_{ex}d_e^2n_x + 2n_{ex}d_ed_{ex}n - n_{exx}d_e^2n + n_{et}d_e^2n = sf^a$$

Now things are starting to get really complicated! We want to put everything in a computer algebra system that can handle differential algebra. Specifically, we want to solve systems of differential/polynomial equations with simple polynomials in their exponents. I call polynomials with polynomial exponents a $K[Z[p]]$ ring; there's a latter section in this paper about them. The next section is about handling systems of poly-differential equations.

The Rosenfeld-Groebner algorithm

When working with systems of polynomial equations, the best current tool by far is Buchberger's algorithm to compute a Groebner basis of an algebraic ideal. For systems of poly-differential equations, we'd like to compute a differential Groebner basis of a differential ideal, but this is more difficult. There are simple examples of differential ideals (like $[y'^2]$) that admit no finite differential basis at all. However, a theorem of Raudenbush (and Ritt) states that radical ideals admit finite bases.

Are radical ideals adequate for our purposes? A radical ideal is one with no "missing roots". If f exists in the $K[Z[p]]$ ring and f^p is in the ideal, for some integer p, then f must itself be in the ideal. A radical ideal can't contain f^p without also containing f.

Recasting this slightly, remember that we're looking for solutions built up out of fairly simple pieces – nothing worse than semilinear ordinary differential equations. A more immediate goal is solution by integrals, exponentials, and algebraics. Starting with rational functions (ratios of polynomials) in the variables required to state the problem, we extend by a finite number of integrals and exponentials, at each step using rational functions from the previous step as coefficients.

$$\mathbf{C}(x,t) \subset \mathbf{C}(x,t,z=e^{-\frac{x^2}{4t}}) \subset \mathbf{C}(x,t,z=e^{-\frac{x^2}{4t}}, r=\sqrt{t})$$

All of these functions are sufficiently well behaved (PROOF NEEDED) that we can add an additional inference to our work. If f^p is zero, where f is some function in our solution space, then f is also zero. This is in addition to the axioms implicit in our use of ideals, namely that if f and g are both zero, then $f+g$, mf, and mg, where m is any other function whatsoever, are all zero. It looks like working in a radical ideal shouldn't be a problem.

On the other hand, can we also safely assume that fg being the zero function implies that either f or g is the zero function? If so, we could restrict our attention to *integral domains*. However, when we take the domain of the function into consideration, we see that f could be non-zero only for x values greater than zero (say), while g is non-zero only for x values less than zero. So their product could be identically zero although neither multiplicand is. Further investigation is required to see if cases like this appear in actual practice.

Returning to the question of radical ideals, it seems that we can safely introduce this assumption, and therefore work in the smallest radical differential ideal generated by a system of poly-differential equations. For such purposes, the Rosenfeld-Groebner algorithm seems state of the art.

Rosenfeld-Groebner, introduced in [Bo95], takes a system of generators as input and computes a finite number of Groebner bases that each describe a prime differential ideal, and the intersection of these ideals forms the radical differential ideal generated by the original system. The operation of the algorithm can be modified by altering the ranking on the derivative terms, in a manner similar to Buchberger's. In particular, elimination rankings are possible, to isolate any ordinary differential equations in one of the variables, for instance.

However, the manner in which Rosenfeld-Groebner presents its results defies immediate simplication. Each prime differential ideal is described by a characteristic set and a Groebner basis, such that reducing the differential polynomial modulo the set produces a reduced differential polynomial that is in the ideal of the Groebner basis iff the original polynomial is in the prime differential ideal. Different prime differential ideals can be described using different characteristic sets, so a simple intersection of Groebner bases won't suffice.

Solving Polynomial Differential Equations

Often we're confronted with an equation like $n_{xx} = n_t$, where n is constrained to be a polynomial. A fair amount is known about such equations, but there are still big gaps.

An important result is the Abramov-Petkovšek theorem, which proves the impossibility of algorithmically solving PDEs in a polynomial ring as a straightforward consequence of the undecidability of Hilbert's tenth problem. The proof is based on the close correlation between solving polynomial differential equations and solving integer Diophantine equations. Abramov and Petkovšek demonstrated a simple construction that allows any polynomial equation to be converted into a PDE with the property than any polynomial solution to the PDE can be trivially lifted to a solution of the original polynomial. Thus, any algorithm that could solve arbitrary PDEs in a polynomial ring could be used to solve arbitrary Diophantine equations, which is impossible (the MRDP theorem).

However, their proof technique suggests to me that the degree of the resulting Diophantine equations should be bounded by the order of the PDE. PROOF NEEDED. If true, then this is good news since Schödinger's equation is second order and Cohen's GTM 239 suggests that second degree Diophantine equations may be completely solvable. See section 6.3 (p. 341) in GTM 239.

The form of the Diophantine equations can be further restricted. Since there are no cross derivatives of the form $\delta^2/\delta x \delta y$, there should be no mixed monomials in the Diophantine equations, so they can be separated into a linear component (solvable using matrix reduction into Hermite normal form) and a sum of squares. If the squares all have the same sign, which is likely since the second derivatives in Schrödinger's equation are invariant under exchange of variables, then the equation is of elliptic type, meaning that it describes an n-dimensional ellipsoid in real space $\mathbf{R}^n$, so its integer solutions are bounded and thus computable. Different values from the linear component, however, might lead to an infinite number of elliptic quadratics with unbounded total size, and it also seems likely that a single PDE could generate an infinite number of Diophantine equations.

Systems of partial differential equations have been extensively studied in the form of D-modules, and algorithms have been developed to find their polynomial and rational solutions in the case of finite holonomic rank. Most of our equations are of infinite rank, however. The known D-module algorithms for rational solutions depend on finding an algebraic variety (the singular locus), which is basically the Zariski closure of the solution's singularities and thus provides crucial information about what factors can be present in the denominator. This seems difficult to generalize into transcendental extension fields, since the relationship between x and e^x, for example, is not algebraic. Futuremore, Tsai's Lemma 2.1.5 states that in the infinite rank case, the singular locus is trivial (it's the entire space).

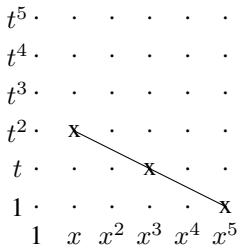
How can we analyse the infinite rank case? Let's consider a simple case: the operator $\delta_x^2 - \delta_t$ acting on $\mathbf{C}[x,t]$. Any polynomial in this ring can be expressed in the form $\sum c_{m,n} x^m t^n$, so applying the operator we obtain:

$$\sum c_{m,n} m(m-1) x^{m-2} t^n - \sum c_{m,n} n x^m t^{n-1} = 0$$

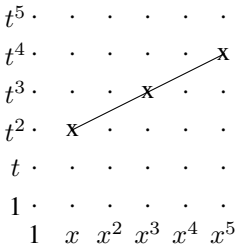
Changing variables, we can collapse the sums together:

$$\sum c_{m+2,n}(m+2)(m+1) x^m t^n - \sum c_{m,n+1}(n+1) x^m t^n = 0$$
$$\sum \left[(m+2)(m+1)c_{m+2,n} - (n+1)c_{m,n+1}\right] x^m t^n = 0$$

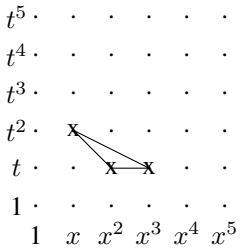
This equation specifies relationships between the various $c_{m,n}$ coefficients. For example, the xt term specifies a relationship between $c_{3,1}$ and $c_{1,2}$, specifically, $3c_{3,1} = c_{1,2}$. The x^3 term specifies a relationship between $c_{5,0}$ and $c_{3,1}$, $20c_{5,0} = c_{3,1}$. If we chart all of the $c_{m,n}$ coefficients, we can graph these relationships as follows:

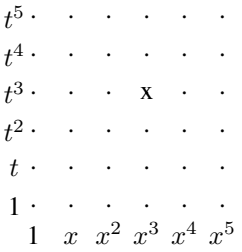


A dimension can be associated with the differential operator, corresponding to the dimensions spanned in the graph. The example shows a one-dimensional operator, spanning a line. Such an operator generates an infinite family of *potential* solutions, each corresponding to a line of the appropriate slope in the chart. Two subcases arise. For a line of negative slope, as in the illustrated example, boundary conditions exist on both ends of the line, as the coefficients must become zero as the line passes out of the first quadrant. These conditions must be satisfied for the potential solution to be an actual solution. In the other case, non-negative slope, only one boundary condition exists on each line, and a recursion relationship gives rise to an infinite sequence of coefficients on each line. An example is the operator $\delta_x^2 + t$, which graphs like this:



For a two-dimensional operator, such as $\delta_x^2 + \delta_x - \delta_t$, all of the coefficients are linked together and can not be separated. For a zero-dimensional operator, such as $x\delta_x + t\delta_t$, each coefficient is independent of all others, and a Diophantine equation determines if each coefficient is an admissible solution. Abramov and Petkovšek used zero-dimensional operators to tie this problem to Hilbert's tenth problem. These two examples are illustrated as follows:





Polynomial rings with more indeterminates give rise to more possibilities. Operators can have dimension from zero up to the number of indeterminates.

The operator $x^a \delta_x^b$ acts on $c_{m,n} x^m t^n$ to produce $m_{(b)} c_{m,n} x^{m-b+a} t^n$, which transforms to $(m+b-a)_{(b)} c_{m+b-a,n} x^m t^n$, where $m_{(b)}$ is the Pochhammer falling factorial. Applying this operation on each term in a differential operator transforms the operator into a recursion relationship on the coefficients.

For example, the operator $\delta_x^2 - \delta_t$ transforms as:

$$\delta_x^2 - \delta_t \Longrightarrow (m+2)(m+1)c_{m+2,n} - (n+1)c_{m,n+1}$$

Treating m as the horizontal coordinate and n as the vertical, we see from this expression that the operator is one-dimensional and has slope $-\frac{1}{2}$, since it contains the points $(m+2, n)$ and $(m, n+1)$.

Having identified the dimension of the operator, it is now profitable to change coordinates into a system that separates dimensions within the operator's hyperplane and dimensions orthogonal to it. Thus, for an m-dimensional operator in an n-dimensional polynomial ring, we seek to find $n - m$ dimensions to identify m-dimensional hyperplanes, and m dimensions within the hyperplanes.

In our example, we can characterize the lines by introducing p as the m-intercept, and thus the highest x power appearing in the line's solution polynomial. We also introduce a as a coordinate along the line, with $a = 0$ corresponding to the point on the m-axis, and higher values of a moving upwards and left along the line.

Setting $m = p - 2a$ and $n = a - 1$, we can transform the coefficients and their recursion relationship as follows, calling them d's in the new numbering system:

$$c_{m,n+1} = c_{p-2a,a} = d_{p,a} \qquad c_{m+2,n} = c_{p-2(a-1),a-1} = d_{p,a-1}$$

$$(m+2)(m+1)c_{m+2,n} - (n+1)c_{m,n+1} = 0 \Longrightarrow (p-2a+2)(p-2a+1)d_{p,a-1} = ad_{p,a}$$

The new numbering clearly relates coefficients along a single line, as evidenced by the p subscript remaining unchanged through the recursion relationship.

Now, in order to have a polynomial and not an infinite sum, we must have clear starting and ending points, beyond which the coefficients are all zero. A starting point requires $d_{p,a}$ to be non-zero, while $d_{p,a-1}$ is zero, so a must be zero. Likewise, an ending point requires $d_{p,a}$ to be zero, while $d_{p,a-1}$ is non-zero, so $(p-2a+2)(p-2a+1)$ must be zero. Since p and a are both integers, this requires (in general) solving Diophantine equations, which comes as no surprise by now!

These specific Diophantine equations are easy, though. $a = 0$ is the only possible starting point, and the only possible end points are $a = \frac{p}{2} + 1$, if p is even, or $a = \frac{p+1}{2}$, if p is odd. We also check to ensure that the resulting m, n coordinates fall in the first quadrant (they do), to prevent any negative powers from appearing in our "polynomials".

Therefore, for each value of p, we can set the value of $d_{p,0}$ arbitrarily, and with further values of $d_{p,a}$ being determined by the recursion relationship:

$$d_{p,a} = \frac{(p-2a+2)(p-2a+1)}{a} d_{p,a-1}$$

Combining the recursions, we obtain:

$$d_{p,a} = \frac{p(p-1)\cdots(p-2a+2)(p-2a+1)}{a!} d_{p,0} = \frac{p!}{(p-2a)!a!} d_{p,0}$$

Earlier in this paper, I separated the even and odd values of p and presented the solutions as:

$$B_{2i} = \sum_{j=0}^{i} \frac{(2i)!}{(2i-2j)!j!} x^{2i-2j} t^j \qquad B_{2i+1} = \sum_{j=0}^{i} \frac{(2i+1)!}{(2i-2j+1)!j!} x^{2i-2j+1} t^j$$

Let's consider another example, the operator $-2x\delta_x - 1 + 2t\delta_x^2 - 2t\delta_t$, which transforms to the recursion:

$$-2mc_{m,n} - c_{m,n} + 2m(m-1)c_{m-2,n+1} - 2nc_{m,n}$$

In this form, we see clearly that this is a one-dimensional operator, even though it has four terms. Setting $m = p - 2a$ and $n = a$, we transform to:

$$(-2p + 4a - 1 - 2a)d_{p,a} + 2(p-2a)(p-2a-1)d_{p,a+1}$$

The problem lies in the first factor, $-2p + 4a - 1 - 2a$, which can never be zero for any integer values of p and a, so our ending condition is never satisified. Thus, this operator has no polynomial solution in $\mathbf{C}[x, t]$.

For two-term recursions like these two examples, zeroing the Diophantine equations provides both necessary and sufficient conditions for the recursion to terminate properly. For higher order one-dimensional recursions, this technique provides necessary conditions that might not be sufficient. Consider a three-term recursion involving d_a, d_{a+1}, and d_{a+2}. For the sequence to start at d_b, d_{a+2}'s coefficient must be zero when $a = b - 2$, since both d_{b-1} and d_{b-2} will be zero. Likewise, for the sequence to end at d_c, d_a's coefficient must be zero when $a = c$, since d_{c+1} and d_{c+2} both must be zero. This does not, however, ensure that the recursion generates zero values for d_{c+1} and d_{c+2}.

Higher dimension operators are more complex, but the same basic ideas apply. Consider the convex hull of the operator. Zeroing the coefficient of each term on the convex hull produces a Diophantine equation that must be satisfied for that point to be in the solution set, while all the other points are outside it. The zeros of the Diophantine equations define boundary hyperplanes that the polynomial solution must lie within.

Theorem. (Hopefully) Given any finite set of points on an n-dimensional integer lattice, an n-dimensional convex polytope whose vertices also lie on the integer lattice, and a distinguished vertex on the convex polytope, an integer translate of the convex polytope

can be found such that the distinguished vertex lies over the only point in the point set that is within the translated convex polytope. PROOF NEEDED.

Furthermore, after changing coordinates as above, we can develop a system of Diophantine equations that may be more restrictive than the individual equations. For example, consider a 2-dimensional operator with three vertices on its convex hull in a 5-dimensional space. Each vertex generates a Diophantine equation of the form $A(m, n, p, a, b)$ where m, n, p identify a two-dimensional plane in the 5-dimensional space, and a, b identify a point on the plane. Since any single solution must be confined to a single plane, we are led to a system of Diophantine equations of the form:

$$A(m, n, p, a, b) = 0$$
$$B(m, n, p, a', b') = 0$$
$$C(m, n, p, a'', b'') = 0$$

A simultaneous solution to this system identifies a single plane, along with three boundary lines on that plane that confine a polynomial solution. Any single solution confines a finite number of points, which can then be tested to see if they can actually be assigned values that satisfy the recursion, requiring nothing more complicated than solving a system of linear equations with constant coefficients. There may, however, be an infinity of solutions to the system of Diophantine equations.

Operating in K[Z[p]]

Often we need to operate in a monoid ring like $K[Z[p]]$, i.e, a polynomial ring whose exponents are polynomials. For example:

$$x^{2p} + 2x^p + 1 = (x^p + 1)^2$$

To order the monomials, an ordering is required on the exponents. I use a "high" ordering where the coefficient of the highest power determines the comparison to zero. [3] Only monoid elements greater than zero are allowed as exponents, so x^{p-50} is in $K[Z[p]]$, but x^{-1} is not.

We can't use Buchberger's algorithm directly, because it may not terminate since the ring is not Noetherian.

$$(x^p) \subset (x^{p-1}) \subset (x^{p-2}) \subset \cdots$$

is an ascending chain of ideals that never stabilises. For the same reason, $K[Z[p]]$ is not a unique factorization domain:

$$x^p = x^{p-10}x^{10} = x^{p-20}x^{20} = \cdots$$

However, $K[Z[p]]$ is a GCD domain. [Gi73] theorem 6.4 states that a semigroup ring is a GCD domain if its coefficient field is a GCD domain and its semigroup is a torsion-free GCD-semigroup with zero. A GCD-semigroup has the property that for any a and b in the semigroup S, there exists a c such that $(a + S) \cap (b + S) = c + S$. Our monoid's strict ordering ensures that this condition is met with $c = \max(a, b)$.

We can compute GCDs (needed for normalisation in the fraction field) using a simple modification to Buchberger's algorithm. Introduce new indeterminates for each combination of lower and upper indeterminates.

$$x^p \to a$$
$$x^q \to b$$
$$x^{(p+q)} \to ab$$

This produces new polynomials with only integer exponents. We can now compute a Gröbner basis, which is how CoCoA computes GCDs, then map the results back into the original ring.

How do we deal with monomials like x^{p-1}? We can't factor it like $x^p x^{-1}$ because x^{-1} isn't in our ring. Instead, we map $x^{p-1} \to a$ and then write x^{p-1} as a and x^p as xa. This implies that we can't construct a single derived ring that can handle all polynomials in $K[Z[p]]$. Instead, we construct a derived ring for any particular GCD calculation.

The derived ring is Noetherian, but is not isomorphic to $K[Z[p]]$. Instead, it is isomorphic to a subring of $K[Z[p]]$ and we can adjust the construction so that the subring encompasses any finite number of elements from $K[Z[p]]$.

Theorem. (Hopefully) Given a ring R, if every pair of elements from R can be embedded in a subring that is a GCD domain, then R is a GCD domain. PROOF NEEDED.

[3] http://math.arizona.edu/~rwilliams/math415A-fall2013/Ordered_Rings_and_Fields.pdf

Representing irreducibility with Gröbner bases

http://mathoverflow.net/questions/217402

We have a ring $R = K[x_1, \ldots, x_n]$ and a ring $S = K[y]$. We want to treat the x_is as polynomials in y, so we're looking for a mapping $f : R \to S$ that sends each x_i to a polynomial in y and satisfies a system of polynomial equations P in the x_i. In other words, we want f to be a ring homomorphism that sends $I(P)$ to 0. f, the map from R to S, is the solution we seek.

Now we want to impose an additional condition: a subset $x_1, \ldots, x_i$ must map to irreducible elements $y_1, \ldots, y_i$ in S. Since $y_1, \ldots, y_i$ are irreducible, they are prime (in S), so we quotient with respect to their ideal I and get a quotient ring S/I that is an *integral domain*. We can also quotient R by the ideal generated by $x_1, \ldots, x_i$ (call it J), and get R/J. f can be similarly restricted, and now we have a homomorphism $\hat{f} : R/J \to S/I$. We can construct a Gröbner basis for R/J by appending the x_i that must be irreducible to the original system P, and reducing $P \cup \{x_1, \ldots, x_i\}$ to a Gröbner basis. This new Gröbner basis gives relationships satisfied by the *equivalence classes* in S/I. "Equal to zero" in this quotient system means "equal to zero or a multiple of an irreducible element" in the original system. However, if the quotient system is inconsistent, then the original system is also inconsistent, at least subject to the restriction that $x_1, \ldots, x_i$ must map to irreducibles.

Can we find additional relationships? Surprisingly, yes! We run this calculation with each irreducible individually. Pick one $x_1, \ldots, x_i$, call it x_j, compute a quotient Gröbner basis for $P \cup \{x_j\}$, take each polynomial in the quotient system's Gröbner basis and test to see if it's in the original system. If so, then it's really equal to zero. Otherwise, it's a multiple of x_j and we can add that polynomial to the original system, equating it a term of the form mx_j, with m a new indeterminate.

The augmented system will have extraneous zeros, at least if we require the irreducible polynomials to be non-zero. We can handle this by computing a primary decomposition and throwing away any primary components that include an irreducible element among their zeros. This is the ideal-theoretic equivalent of factoring a polynomial that must be equal to zero and throwing away factors that we know are non-zero. We can keep repeating these two processes (quotient ring basis and primary decomposition) until our ideal stabilizes.

Example

Consider the equation $af^2 + bf + c = 0$, with f restricted to be irreducible.

Step 1: Form the system $\{af^2 + bf + c, f\}$ and reduce to the Gröbner basis $\{f, c\}$. Of course f is here; our interest is c. Since it isn't in the original ideal, it must be a multiple of f, so we add $c - mf$ our ideal to obtain

$$\{af^2 + bf + c, c - mf\}$$

Step 2: A primary decomposition of this ideal gives two primary ideals, one of which is (f, c). Since f can't be zero, we throw it away and continue with the other primary ideal:

$$(af + b + m, ac + bm + m^2, c - mf)$$

Step 3: Back to the quotient calculation. Now our system is

$$\{af + b + m, ac + bm + m^2, c - mf, f\}$$

and we compute the Gröbner basis $\{f, c, b + m\}$. This implies that $b + m$ must also be a multiple of f, so we add $b + m - nf$ to our ideal, obtaining

$$(af + b + m, ac + bm + m^2, c - mf, b + m - nf)$$

Step 4: Another primary decomposition gives another extraneous ideal $(f, c, b + m)$. Throwing this away, we have

$$(a + n, fn - b - m, fm - c, bm + m^2 - cn)$$

Step 5: A final quotient calculation, with the system

$\{a+n, fn-b-m, fm-c, bm+m^2-cn, f\}$

gives $(f, c, b+m, a+n)$, of which the only new element, $a+n$, reduces to zero.

So we've stabilized on

$$(a+n, fn-b-m, fm-c, bm+m^2-cn)$$

This ideal encodes all of the information I was able to extract in `http://mathoverflow.net/questions/216392`

Bibliography

Bronstein, Symbolic Integration I: Transcendental Functions, Springer 2004.

The single most important reference on the use of differential algebra to solve differential equations. The concept of "special" and "normal" polynomials is from this text.

Abramov, Petkovšek, On Polynomial Solutions of Linear Partial Differential and (q-)Difference Equations. CASC 2012.

Proves the impossibility of algorithmically solving PDEs in a polynomial ring as a straightforward consequence of the undecidability of Hilbert's tenth problem.

Alin Bostan, Thomas Cluzeau, Bruno Salvy. Fast Algorithms for Polynomial Solutions of Linear Differential Equations. 2005.

`http://specfun.inria.fr/bostan/publications/BoClSa05.pdf`

Explains the construction of the indicial polynomial and its use in solving linear ODEs with polynomial coefficients.

Saito, Sturmfels, Takayama, Groebner Deformations of Hypergeometric Equations. Springer Verlag. 1999.

A standard reference work on D-modules. Haven't read it because it's expensive, so I extracted information from stuff like [Ts00].

[Ts00] Harrison Tsai. Algorithms for Algebraic Analysis. Ph.D. thesis, UC Berkeley, 2000.

One of Sturmfels' students. The D-module algorithms developed here and in the previous reference require the module to be of finite holonomic rank, which is generally not the case in the equations of interest to us.

[Bo95] Boulier, et. al, Representation for the radical of a finitely generated differential ideal. ISSAC '95 proceedings.

This is a really good paper, one of my favorites! Not only does it introduce the Rosenfeld-Groebner algorithm for the first time in print, but it's packed with relevant, terse information.

[Ma91] Mansfield, Differential Groebner Bases, Ph.D thesis, University of Sidney, 1991.

[Gi73] Gilmer, Robert, and Tom Parker. Divisibility properties in semigroup rings. The Michigan Mathematical Journal 21.1 (1974): 65-86.

A nice paper in an obscure journal that establishes necessary and sufficient conditions for commutative semigroup rings to be GCD domains, unique factorization domains, and principle ideal domains.