A fast algorithm for computing the $p$-curvature

Alin Bostan

joint work with
Xavier Caruso (Univ. Rennes 1) and Éric Schost (Univ. Waterloo)

ISSAC’15, Bath, UK  July 7th 2015
Main objects and goal

- $k = $ a field of prime characteristic $p$, typically $\mathbb{F}_p$
- $k(x)\langle \partial \rangle = $ the non-commutative (right-) Euclidean algebra of linear differential operators $L = a_0 + a_1 \partial + \cdots + a_r \partial^r$, for $a_i \in k(x)$

**Def:** $p$-curvature $A_p(L)$ of $L = $ the matrix in $\mathcal{M}_r(k(x))$ whose $(i, j)$ entry is the coefficient of $x^i$ in $\partial^{p+j} \text{ Rmod } L$ for $0 \leq i, j < r$

**Goal:** design an efficient algorithm for $A_p(L)$

- **Efficiency = complexity estimates with a low exponent w.r.t. $p$**
- **Complexity is measured in terms of arithmetic operations in $k$**
Main objects and goal

- $k = \text{a field of prime characteristic}\ p$, typically $\mathbb{F}_p$
- $k(x)\langle \partial \rangle = \text{the non-commutative (right-) Euclidean algebra of linear differential operators}\ L = a_0 + a_1 \partial + \cdots + a_r \partial^r$, for $a_i \in k(x)$

Def: $p$-curvature $A_p(L)$ of $L = \text{the matrix in } M_r(k(x))$ whose $(i, j)$ entry is the coefficient of $x^i$ in $\partial^{p+j} \text{ Rmod } L$ for $0 \leq i, j < r$

Goal: design an efficient algorithm for $A_p(L)$

- Efficiency = complexity estimates with a low exponent w.r.t. $p$
- Complexity is measured in terms of arithmetic operations in $k$
- Caveat: to simplify matters, assume input $L$ has size $\mathcal{O}(1)$
Example

\[ L = (5x^2 + 4)\partial^2 + (4x^2 + 6x + 5)\partial + 2x + 2 \in \mathbb{F}_7[x]\langle \partial \rangle \]

Euclidean right division in \( \mathbb{F}_7(x)\langle \partial \rangle \):

\[ \partial^7 = (\cdots) L + \frac{(x + 1)(x^2 + x - 1)}{(x + 3)(x - 3)^2} \partial + \frac{4x(x - 1)}{(x + 3)(x - 3)^2} \]

\[ \partial^8 = (\cdots) L + \frac{2(x + 1)(x^2 + x - 1)}{(x + 3)(x - 3)^2} \partial + \frac{x(x - 1)}{(x + 3)(x - 3)^2} \]

\[ \implies A_7(L) = \begin{bmatrix} \frac{4x(x-1)}{(x+3)(x-3)^2} & \frac{x(x-1)}{(x+3)(x-3)^2} \\ \frac{(x+1)(x^2+x-1)}{(x+3)(x-3)^2} & \frac{2(x+1)(x^2+x-1)}{(x+3)(x-3)^2} \end{bmatrix} \]
Basics on differential equations in characteristic $p$

Main differences between characteristic zero and $p$

- (Honda 1981) solutions are simpler in characteristic $p$
  \[
  \dim_{k(x^p)} S_L(k[x]) = \dim_{k(x^p)} S_L(k(x)) = \dim_{k((x^p))} S_L(k[[x]])
  \]

- Cauchy’s theorem does not hold: the common dimension $\dim S_L$ of the solution spaces is generally $< r = \text{ord}(L)$

  Example: $y' = y$ has no solution in $\mathbb{F}_p[[x]]$

Connection between solutions and $p$-curvature

- (Katz & Cartier 1970) $\text{rank}(A_p(L)) = r - \dim(S_L)$

  $\rightarrow$ $p$-curvature measures to what extent $\dim(S_L)$ is close to $r$
Example

\[ L = (5x^2 + 4)\partial^2 + (4x^2 + 6x + 5)\partial + 2x + 2 \in \mathbb{F}_7[x]\langle \partial \rangle \]

- 7-curvature of \( L \)

\[ A_7(L) = \begin{bmatrix} \frac{4x(x-1)}{(x+3)(x-3)^2} & \frac{x(x-1)}{(x+3)(x-3)^2} \\ \frac{(x+1)(x^2+x-1)}{(x+3)(x-3)^2} & \frac{2(x+1)(x^2+x-1)}{(x+3)(x-3)^2} \end{bmatrix} \]

- Katz-Cartier:

\[ 1 = \text{rank}(A_7(L)) = 2 - \dim_{\mathbb{F}_7(x^7)}(S_L) \implies \dim_{\mathbb{F}_7(x^7)}(S_L) = 1 \]

- In fact

\[ \text{Basis}_{\mathbb{F}_7(x^7)}(S_L) = \{1 - 2x^2 - x^3\} \]
A useful tool for theory

**Grothendieck’s conjecture (’70s)** \( \Gamma \in \mathbb{Q}[x]<\partial> \) has a basis of algebraic solutions over \( \mathbb{Q}(x) \) iff \( A_p(\Gamma) = 0 \) for almost all primes \( p \).

**Def:** A power series \( \sum_{n \geq 0} \frac{a_n}{b_n} x^n \) in \( \mathbb{Q}[[x]] \) is called a **G-series** if it is (a) D-finite; (b) analytic at \( x = 0 \); (c) \( \exists C > 0, \text{lcm}(b_0, \ldots, b_n) \leq C^n \).

**Examples:** algebraic functions; \( \log(1 - x), \, _2F_1\left(\begin{array}{c} \alpha \\ \beta \\ \end{array} \bigg| \frac{x}{\gamma} \right) \); diagonals

**Chudnovsky’s theorem (1985)** The minimal-order operator \( \Gamma \in \mathbb{Q}[x]<\partial> \) annihilating a **G-series** is globally nilpotent, i.e., the \( p \)-curvatures \( A_p(\Gamma) \) are nilpotent for almost all primes \( p \).

**Examples:** algebraic resolvents; \( x(1 - x)\partial^2 + (\gamma - (\alpha + \beta + 1)x)\partial - \alpha\beta x \)
A useful tool for algorithms

$p$-curvature used in computer algebra:

- [van der Put 1995] for factoring operators in $\mathbb{F}_p(x)\langle \partial \rangle$
- [Cluzeau 2003] for decomposing differential systems over $\mathbb{F}_p(x)$
- [Cluzeau & van Hoeij 2004] as filter in modular algorithms for operators in $\mathbb{Q}(x)\langle \partial \rangle$

💡 Improving the complexity of the $p$-curvature computation is an interesting problem in its own right
A useful tool for applications

- in enumerative combinatorics (classification of lattice walks)
- in statistical physics (square lattice Ising model)

**Typical task**: given a power series \( S \in \mathbb{Z}[[x]] \), decide if \( S \) is D-finite

**Differential guessing**: from the first \( N \gg 0 \) terms of \( S \), compute a differential operator \( \Gamma \in \mathbb{Q}[x]\langle \partial \rangle \) that annihilates \( S \mod x^N \)

▷ One way to empirically certify the correction of \( \Gamma \) is to **look at the p-curvature** \( A_p(\Gamma \mod p) \) for a random (large) prime \( p \)
  - if \( A_p(\Gamma \mod p) \) is nilpotent, then \( S \) is very probably D-finite
  - if \( A_p(\Gamma \mod p) \) is zero, then \( S \) is very probably algebraic
A combinatorial application: Gessel’s conjecture

- **Gessel walks**: walks in $\mathbb{N}^2$ using only steps in $S = \{↗, ↖, ←, →\}$
- $g(i, j, n) = \text{number of walks from } (0, 0) \text{ to } (i, j) \text{ with } n \text{ steps in } S$

**Question**: Nature of the generating function

$$G(u, v, x) = \sum_{i,j,n=0}^{\infty} g(i, j, n) u^i v^j x^n \in \mathbb{Q}[u, v, x]$$

**Theorem** (B.-Kauers 2010) \(G(u, v, x)\) is an algebraic function.\(^*\)

→ Effective, computer-driven discovery and proof
→ Key step in discovery: \(p\)-curvature computation of two 11th order (guessed) differential operators for \(G(u, 0, x)\), and \(G(0, v, x)\)

\(^*\)Minimal polynomial \(P(u, v, x, G(u, v, x)) = 0\) has $> 10^{11}$ terms; $\approx 30\text{Gb}$ (!)
Previous work

1. $p$-curvature $A_p(L)$

- [Katz 1982]: algorithm of cost $O(p^2)$, based on recurrence
  
  $A_1 = \text{CompanionMatrix}(L), \quad A_{k+1} = A'_k + A_1 \cdot A_k$

- [B. & Schost 2009]: first subquadratic algorithm $O(p^{1.79})$

- [B. & Schost 2009]: for certain second-order operators $\tilde{O}(p)$

⚠️ Binary powering can not be used to compute $\partial^p$ in $\frac{k(x)\langle \partial \rangle}{k(x)\langle \partial \rangle L}$
Previous work

1. \( p \)-curvature \( A_p(L) \)
   - [Katz 1982]: algorithm of cost \( O(p^2) \), based on recurrence
     \[ A_1 = \text{CompanionMatrix}(L), \quad A_{k+1} = A_k' + A_1 \cdot A_k \]
   - [B. & Schost 2009]: first subquadratic algorithm \( O(p^{1.79}) \)
   - [B. & Schost 2009]: for certain second-order operators \( \tilde{O}(p) \)

   Binary powering can not be used to compute \( \partial^p \) in
   \[ \frac{k(x)\langle \partial \rangle}{k(x)\langle \partial \rangle L} \]

2. Characteristic polynomial of \( A_p(L) \)
   - [B., Caruso & Schost 2014]: sublinear algorithm \( \tilde{O}(\sqrt{p}) \)

3. Polynomial solutions of \( L \)
   - [B. & Schost 2009]: quasi-optimal algorithm \( \tilde{O}(p) \)
New result
(B.-Caruso-Schost, 2015)

Computation of $A_p(L)$

for an arbitrary operator $L$

in quasi-linear time $\tilde{O}(p)$.

Precise complexity result for $L$ of bidegree $(d, r)$ in $(x, \partial)$:

$\tilde{O}(p d r^\omega)$

where $\omega$ is the exponent of matrix multiplication.

Optimality: for $r > 1$, generic size of $A_p(L)$ is $\Theta(p d r^2)$

Extension to systems: same results for $p$-curvature of $Y' = AY$
The starting point

- **Question**: Given $L$ in $k(x)\langle \partial \rangle$, compute $R$ in $k(x)\langle \partial \rangle$ such that

$$\partial^p = QL + R, \quad \text{ord}(R) < \text{ord}(L) = r$$

- **Idea**: *evaluation-interpolation*; on “points” = solutions of $L$

  ▶ **Fruitful strategy in related contexts**: product, $\text{lclm}$, $\text{gcrd}$ (char. 0)

  [van der Hoeven '02, '12; B. '03; Benoit, B. & van der Hoeven '12]
The starting point

- **Question:** Given $L$ in $k(x)\langle \partial \rangle$, compute $R$ in $k(x)\langle \partial \rangle$ such that

$$\partial^p = QL + R, \quad \text{ord}(R) < \text{ord}(L) = r$$

- **Idea:** *evaluation-interpolation*; on “points” = solutions of $L$

  > Fruitful strategy in related contexts: product, lclm, gcrd (char. 0)
  
  [van der Hoeven ’02, ’12; B. ’03; Benoit, B. & van der Hoeven ’12]

  😊 If $L$ had a full basis of power series solutions $\{y_1, \ldots, y_r\}$, then $R = \sum_{j=0}^{r-1} a_j(x)\partial^j$ could be determined by solving a linear system

$$\begin{pmatrix} a_0, \ldots, a_{r-1} \end{pmatrix} \cdot \text{Wronskian}(y_1, \ldots, y_r) = (\partial^p(y_1), \ldots, \partial^p(y_r))$$

with coefficients in $k[[x]]$ truncated modulo $x^{O(p)}$
The starting point

- **Question**: Given $L$ in $k(x)\langle \partial \rangle$, compute $R$ in $k(x)\langle \partial \rangle$ such that
  \[ \partial^p = QL + R, \quad \text{ord}(R) < \text{ord}(L) = r \]

- **Idea**: *evaluation-interpolation*; on “points” = solutions of $L$
  - Fruitful strategy in related contexts: product, lclm, gcrd (char. 0)
    - [van der Hoeven ’02, ’12; B. ’03; Benoit, B. & van der Hoeven ’12]
  - If $L$ had a full basis of power series solutions \( \{ y_1, \ldots, y_r \} \), then
    \[ R = \sum_{j=0}^{r-1} a_j(x) \partial^j \] could be determined by solving a linear system
    \[ (a_0, \ldots, a_{r-1}) \cdot \text{Wronskian}(y_1, \ldots, y_r) = (\partial^p(y_1), \ldots \partial^p(y_r)) \]
    with coefficients in $k[[x]]$ truncated modulo $x^{O(p)}$

- **Obstruction**: Cauchy’s theorem *does not hold* in char. $p > 0
The key: series with divided powers

- $\ell = $ a ring in which $p$ vanishes
- $\ell[[t]]^{dp} =$ series with divided powers (Hurwitz series)

\[
f = a_0 \gamma_0(t) + a_1 \gamma_1(t) + a_2 \gamma_2(t) + \cdots + a_i \gamma_i(t) + \cdots
\]

where $a_i \in \ell$ and $\gamma_i(t) \cdot \gamma_j(t) = \binom{i+j}{i} \gamma_{i+j}(t)$.

**Theorem** (Cauchy’s theorem for Hurwitz series)
For any $r \times r$ matrix $A$ with coefficients in $\ell[[t]]^{dp}$, and for any initial data $V \in \ell^r$, the Cauchy problem

\[
\begin{cases}
Y' = A \cdot Y \\
Y(0) = V
\end{cases}
\]

has a unique solution in $\ell[[t]]^{dp}$. 
Efficient computation with divided powers

**Theorem** For \( N = np^s \), with \( s \geq 0 \) and \( n \in \{1, \ldots, p\} \), there is a canonical isomorphism of \( \ell \)-algebras:

\[
\ell[[t]]^{dp}/\ell[[t]]_{\geq N} \simeq \ell[t_0, \ldots, t_s]/(t_0^p, \ldots, t_{s-1}^p, t_s^n).
\]

Proof: Send \( \gamma_{pi}(t) \) to \( t_i \) and use Lucas’ theorem. If \( n = \sum_{i=0}^{s} n_i p^i \),

\[
\gamma_n(t) = \gamma_{n_0}(t) \cdot \gamma_{n_1 p}(t) \cdots \gamma_{n_p}(t) \quad \mapsto \quad \frac{t_0^{n_0}}{n_0!} \cdot \frac{t_1^{n_1}}{n_1!} \cdots \frac{t_s^{n_s}}{n_s!}.
\]

**Theorem** The product in \( \ell[[t]]^{dp} \) at precision \( N = p^{\mathcal{O}(1)} \) can be performed with \( \tilde{O}(N) \) operations in \( k \).

Proof: Use Kronecker’s substitution + univariate FFT.
Fast differential system solving in divided powers

- Newton iteration [B.-Chyzak-Ollivier-Salvy-Schost-Sedoglavic’07]

---

**Input:** a differential system $Y' = AY$, an integer $N$  
**Output:** the fundamental system of solutions in $\ell[[t]]_{N}^{\text{dp}}$  

1. $Y = I_r + t A(0); \ Z = I_r; \ m = 2$  
2. **while** $m \leq N/2$:  
3. \[ Z = Z + \lceil Z(I_r - YZ) \rceil^m \]  
4. \[ Y = Y - \lceil Y \left( \int Z \cdot (Y' - [A]^{2m-1}Y) \right) \rceil^{2m} \]  
5. $m = 2m$  
6. **return** $Y$

---

**Theorem** Solving $Y' = AY$ in $\ell[[t]]_{N}^{\text{dp}}$ at precision $N = p^{O(1)}$ can be performed with $\tilde{O}(N r^\omega)$ operations in $\ell$. 
Example (continued)

\[ L = (5x^2 + 4)\partial^2 + (4x^2 + 6x + 5)\partial + 2x + 2 \in \mathbb{F}_7[x]\langle \partial \rangle \]

- basis of divided power solutions of \( L \) in \( \mathbb{F}_7[[x]]^{dp} \):
  
  \[ y_1 = \gamma_0 + 3\gamma_2 + \gamma_3 \]
  
  \[ y_2 = \gamma_0 + 4\gamma_2 + \gamma_4 + 2\gamma_5 + 4\gamma_6 + \gamma_7 + 2\gamma_8 + 4\gamma_9 + \gamma_{10} + 2\gamma_{11} + 4\gamma_{12} + \cdots \]

- \( \partial^7 = QL + R \), with \( R = a_0 + a_1\partial \), implies

\[
[a_0 \quad a_1] \cdot \begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} = \begin{bmatrix} y_1^{(7)} & y_2^{(7)} \end{bmatrix}
\]

with solution

\[ a_0 = 4x + 2x^2 + 5x^3 + 2x^4 + 4x^5 + 5x^6 + \cdots = \frac{4x(x - 1)}{(x + 3)(x - 3)^2} \]

\[ a_1 = 1 + 5x + 6x^2 + x^3 + 6x^4 + 5x^5 + x^6 + \cdots = \frac{(x + 1)(x^2 + x - 1)}{(x + 3)(x - 3)^2} \]
The algorithm in a nutshell

**Input:** a differential system $Y' = AY$, with $A \in \mathcal{M}_r(k[x][\frac{1}{f}])$

**Output:** its $p$-curvature $A_p$ (def: $A_{k+1} = A'_k + A_1 \cdot A_k$, $A_1 = -A$)

1. Choose $S \in k[x]$ separable and coprime with $f$
   Let $\ell = k[x]/S$ and $\varphi_S : k[x][\frac{1}{f}]/S^p \xrightarrow{\sim} \ell[t]/t^p$, $x \mapsto t + x$

2. Compute a fundam. matrix $Y_S \in \mathcal{M}_r(\ell[t]/t^p)$ of $Y' = \varphi_S(A)Y$
   **Cost:** $\tilde{O}(pr^\omega)$ ops. in $\ell$

3. Deduce $A_p \mod S^p$ as $\varphi_S^{-1}(Y_S \cdot \text{coeff}(AY_S, p-1) \cdot Y_S^{-1})$
   **Cost:** $\tilde{O}(pr^\omega)$ operations in $\ell$

4. Use several $S$’s and CRT to get $A_p$ from the various $A_p \mod S^p$

**Total cost:** $\tilde{O}(pr^\omega d)$ operations in $k$, where $d = \deg(A)$
Experimental results

- For random linear differential operators in $k[x] \langle \partial \rangle$

<table>
<thead>
<tr>
<th>$(d, r)$</th>
<th>$p$</th>
<th>157</th>
<th>281</th>
<th>521</th>
<th>983</th>
<th>1811</th>
<th>3433</th>
<th>6421</th>
<th>12007</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5, 5)</td>
<td></td>
<td>0.39 s</td>
<td>0.71 s</td>
<td>1.22 s</td>
<td>2.34 s</td>
<td>4.41 s</td>
<td>8.93 s</td>
<td>18.0 s</td>
<td>36.1 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.26 s</td>
<td>0.76 s</td>
<td>2.69 s</td>
<td>9.05 s</td>
<td>32.6 s</td>
<td>145 s</td>
<td>593 s</td>
<td>2132 s</td>
</tr>
<tr>
<td>(5, 11)</td>
<td></td>
<td>1.09 s</td>
<td>2.05 s</td>
<td>3.65 s</td>
<td>7.05 s</td>
<td>12.6 s</td>
<td>26.7 s</td>
<td>53.3 s</td>
<td>109 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.25 s</td>
<td>3.70 s</td>
<td>12.8 s</td>
<td>45.5 s</td>
<td>163 s</td>
<td>725 s</td>
<td>2942 s</td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td></td>
<td>2.93 s</td>
<td>5.25 s</td>
<td>9.52 s</td>
<td>17.7 s</td>
<td>32.5 s</td>
<td>68.1 s</td>
<td>139 s</td>
<td>288 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.29 s</td>
<td>12.4 s</td>
<td>42.5 s</td>
<td>153 s</td>
<td>548 s</td>
<td>2460 s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11, 20)</td>
<td></td>
<td>6.89 s</td>
<td>13.3 s</td>
<td>22.6 s</td>
<td>45.0 s</td>
<td>80.4 s</td>
<td>167 s</td>
<td>342 s</td>
<td>711 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.6 s</td>
<td>34.7 s</td>
<td>121 s</td>
<td>486 s</td>
<td>1943 s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(20, 20)</td>
<td></td>
<td>14.0 s</td>
<td>25.1 s</td>
<td>49.9 s</td>
<td>94.0 s</td>
<td>176 s</td>
<td>357 s</td>
<td>733 s</td>
<td>1472 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.0 s</td>
<td>84.5 s</td>
<td>314 s</td>
<td>1283 s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Running times of the new algorithm, vs. Katz’s algorithm

- For operators with physical relevance: e.g., $\phi^{(5)}_H$ in $(\mathbb{Z}/27449 \mathbb{Z})[x] \langle \partial \rangle$, with $(d, r) = (108, 28)$ [Maillard et al. 2007]

    $\rightarrow$ (first column of) $A_p(\phi^{(5)}_H)$ in 19 hours, size $\approx 1$ GB
Conclusion

This work:
- Computation of \( p \)-curvature \( A_p(L) \) in quasi-optimal time \( \tilde{O}(p) \)
- Basic tool: \textit{evaluation/interpolation} on \textit{Hurwitz series}

Next challenges:
- Compute invariant factors of \( A_p(L) \) in time \( \tilde{O}(\sqrt{p}) \)
- Factor \( L \) in time \( \tilde{O}(p) \)
Thanks for your attention!