Holonomic Gradient Descent and its Application to Fisher-Bingham Integral

Tomonari Sei, Nobuki Takayama, Akimichi Takemura ;
Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro,
Katsuyoshi Ohara

June 29, 2010

http://arxiv.org/abs/1005.5273

A function $f(x_1, \ldots, x_d)$ is called *a holonomic function* when f satisfies

$$\sum_{k=0}^{r_i} a^i_k(x_1, \ldots, x_d)\partial_i^k \bullet f = 0, \quad a^i_k \in \mathbf{C}[x_1, \ldots, x_d], \quad i = 1, \ldots, d,$$

where $\partial_i^k \bullet f = \frac{\partial^k f}{\partial x_i^k}$.

$$R = \mathbf{C}(x_1, \ldots, x_d)\langle \partial_1, \ldots, \partial_d \rangle$$

where we denote by $\mathbf{C}(x_1, \ldots, x_d)$ the field of rational functions in $x_1, \ldots, x_d$. The ring R is an associative non-commutative ring and the commuting relations are $\partial_i \partial_j = \partial_j \partial_i$ and $\partial_i a(x) = a(x)\partial_i + \frac{\partial a}{\partial x_i}$ for $a(x) \in \mathbf{C}(x_1, \ldots, x_d)$.

Let I be a left ideal in R which annihilates the holonomic function f. Then, we have

$$\dim_{\mathbf{C}(x_1,\ldots,x_d)} R/I \leq \prod_{i=1}^{d} r_i, \qquad \text{(zero-dimansional over } \mathbf{C}(x)\text{)}.$$

Let S be the set of standard monomials of a Gröbner basis of I in R. We may suppose that S contains 1 as the first element of S. Since the function f is holonomic, the column vector of functions $G = (s_k \bullet f \mid s_k \in S)^T$ satisfies

$$\frac{\partial G}{\partial x_i} = P_i G, \quad i = 1, \ldots, d. \quad \text{(Pfaffian system)}$$

(p,q)-th element of P_i is the coeffcient of the normal form of $\partial_i s_p$ with respect s_q.

Note that each equation can be regarded as an ordinary differential equation with respect to x_i with parameters $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d$.

Example ($d = 1$).
Suppose that I is generated by $(\partial_1^2 - x_1)$. $\dim_{\mathbf{C}(x_1)} R/I = 2$ and $S = \{1, \partial_1\}$. The normal of $\partial_1 \partial_1$ is $x_1 \cdot 1$. Then, $P_1 = \begin{pmatrix} 0 & 1 \\ x_1 & 0 \end{pmatrix}$.

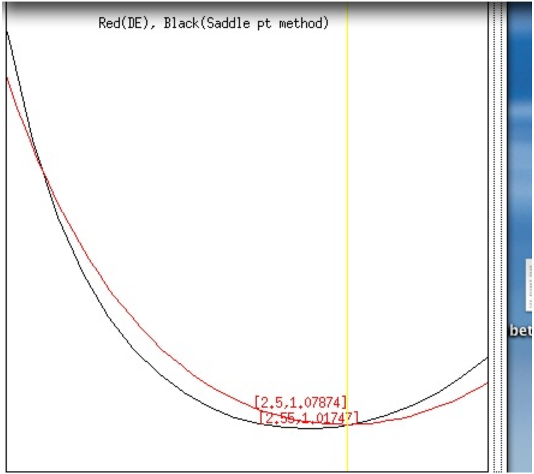
Example (d=1).
$f(x) = \exp(-x+1)\int_0^\infty \exp(xt - t^3)dt$. The function $f(x)$ satisfies the differential equation $(3\partial_x^2 + 6\partial_x + (3-x)) \bullet f = \exp(-x+1)$.
$S = \{1, \partial_x\}$.

$$\frac{dG}{dx} = \begin{pmatrix} 0 & 1 \\ (-3+x)/3 & -2 \end{pmatrix} G + \begin{pmatrix} 0 \\ \exp(-x+1)/3 \end{pmatrix} = P(x)G + Q(x)$$

Problem: minimize the function $f(x)$.
Euler's method. We evaluate $G(0) = (g(0), g'(0))^T$ by a numerical integration method; $\bar{G}(0) = (2.427, -1.20)^T$.
Use the difference scheme (h is a small number).

$$G_{k+1} = G_k \pm h(P(x_k)G_k + Q(x_k)), \quad x_{k+1} = x_k \pm h, \quad G_0 = G(0)$$

```
[1776] step4();
Plot_auto: screen size is [0.1,1.0001330473587330003,4.755000000
3381128043151]
[[[3.4,1.0162931700954017l4,0.0042038168137719754 84],[3.3,1.016
0.0115687496159526390 5]]]
[1777] █
```



Algorithm (holonomic gradient descent)

Let ε be a small positive number.

1. Obtain a Gröbner basis of I in R and a set of standard monomials S of the basis.
2. Compute the matrices P_i by the normal form in R algorithm and the Gröbner basis and the set of standard monomials.
3. Take a point c in E as a starting point and evaluate numerically G at $x = c$. Denote the value by $\bar{G}$ and put $e = c$.
4. The gradient of the target function f is $(\nabla f)(e) = ((P_1(e)\bar{G})_1, \ldots, (P_d(e)\bar{G})_1)$ where v_1 denotes the first element of the vector v.
5. If the gradient is zero, then stop.
6. Update e to $e - \varepsilon(\nabla f)(e)$. Evaluate the value of G at the new e numerically and update the value $\bar{G}$. Goto 4.

As long as the point e stays out of the locus of the singularities of the Pfaffian equations, we can apply standard convergence criterions for the gradient descent.

Theorem

If a set of operators which annihilate the holonomic function f is given and if it is zero-dimensional over $\mathbf{C}(x)$, then we can apply the algorithm HGD.

Note that the Hessian of f at e is equal to

$$(((\partial_j P_i + P_i P_j)(e)\bar{G})_1)$$

Fisher-Bingham integral

We denote by $S^n(r)$ the n-dimensional sphere with the radius r in the $n+1$ dimensional Euclidean space. Let x be a $(n+1)\times(n+1)$ symmetric matrix and y a row vector of length $n+1$. We are interested in the following integral with the parameters x, y, r.

$$F(x,y,r) = \int_{S^n(r)} \exp(t^T x t + y t)|dt| \quad (1)$$

Here, t is the column vector $(t_1, \ldots, t_{n+1})^T$ and $|dt|$ is the standard measure on the sphere. We call the integral (1) *the Fisher-Bingham integral* on the sphere $S^n(r)$.

Theorem

The Fisher-Bingham integral $F(x,y,r)$ is a holonomic function.

$n = 2.$

$$
\begin{aligned}
&\partial_{x_{11}} - \partial_{y_1}^2, \partial_{x_{12}} - \partial_{y_1}\partial_{y_2}, \partial_{x_{13}} - \partial_{y_1}\partial_{y_3}, \\
&\partial_{x_{22}} - \partial_{y_2}^2, \partial_{x_{23}} - \partial_{y_2}\partial_{y_3}, \partial_{x_{33}} - \partial_{y_3}^2, \\
&\partial_{x_{11}} + \partial_{x_{22}} + \partial_{x_{33}} - r^2, \\
&x_{12}\partial_{x_{11}} + 2(x_{22} - x_{11})\partial_{x_{12}} - x_{12}\partial_{x_{22}} + x_{23}\partial_{x_{13}} - x_{13}\partial_{x_{23}} + y_2\partial_{y_1} - y_1\partial \\
&x_{13}\partial_{x_{11}} + 2(x_{33} - x_{11})\partial_{x_{13}} - x_{13}\partial_{x_{33}} + x_{23}\partial_{x_{12}} - x_{12}\partial_{x_{23}} + y_3\partial_{y_1} - y_1\partial \\
&x_{23}\partial_{x_{22}} + 2(x_{33} - x_{22})\partial_{x_{23}} - x_{23}\partial_{x_{33}} + x_{13}\partial_{x_{12}} - x_{12}\partial_{x_{13}} + y_3\partial_{y_2} - y_2\partial \\
&r\partial_r - 2(x_{11}\partial_{x_{11}} + x_{12}\partial_{x_{12}} + x_{13}\partial_{x_{13}} + x_{22}\partial_{x_{22}} + x_{23}\partial_{x_{23}} + x_{33}\partial_{x_{33}}) \\
&\quad -(y_1\partial_{y_1} + y_2\partial_{y_2} + y_3\partial_{y_3}) - 2.
\end{aligned}
$$

Theorem

The holonomic rank (number of the standard monomials) of the system for $n = 2$ is 6. A set of standard monomials in R is

$$1, \partial_r, \partial_{y_3}, \partial_{y_2}, \partial_{y_1}, \partial_{x_{33}}.$$

`http://www.math.kobe-u.ac.jp/OpenXM/Math/Fisher-Bingham`.
The full automatic HGD uses integration algorithms in the Weyl algebra D.

Theorem

The system of differential equations for the Fisher-Bingham integral given in the next page is zero-dimensional in R.

$$\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}, \quad (i \le j) \tag{2}$$

$$\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2, \tag{3}$$

$$\begin{aligned} & x_{ij}\partial_{x_{ii}} + 2(x_{jj} - x_{ii})\partial_{x_{ij}} - x_{ij}\partial_{x_{jj}} + \sum_{k \ne i,j} (x_{jk}\partial_{x_{ik}} - x_{ik}\partial_{x_{jk}}) \\ & \quad + y_j\partial_{y_i} - y_i\partial_{y_j}, \quad (i < j, x_{k\ell} = x_{\ell k}), \end{aligned} \tag{4}$$

$$r\partial_r - 2\sum_{i \le j} x_{ij}\partial_{x_{ij}} - \sum_i y_i\partial_{y_i} - n. \tag{5}$$

Application to directional statistics

Minimize a holonomic function

$$F(x, y, 1) \exp\left(-\sum_{1 \le i \le j \le n} S_{ij} x_{ij} - \sum_{i} S_i y_i\right) \quad (6)$$

with respect to x and y for given data $((S_{ij})_{i \le j}, (S_i))$.
To estimate the unknown parameter (x, y) in $\prod_{\nu=1}^{N} p(t(\nu)|x, y)$ (independently identically distributed) from the sample is a main problem in statistics. An established method is *the maximum likelihood method (MLE)* that maximizes a function $\prod_{\nu=1}^{N} p(t(\nu)|x, y)$ with respect to (x, y). The MLE is equivalent to minimizing the function (6) in case of the Fisher-Bingham distribution.

J. T. Kent, The Fisher-Bingham Distribution on the Sphere, *Journal of the Royal Statistical Society. Series B* **44** (1982), 71–80.
A. T. A. Wood, Some notes on the Fisher-Bingham family on the sphere, *Communications in Statistics, Theory and Methods* **17** (1988), 3881–3897.

The astronomical data consist of the locations of 188 stars of magnitude brighter than or equal to 3.0.
Minimize

$$F(x,y,1)\exp\left(-\sum_{1\le i\le j\le 3} S_{ij}x_{ij} - \sum_i S_i y_i\right)$$

on

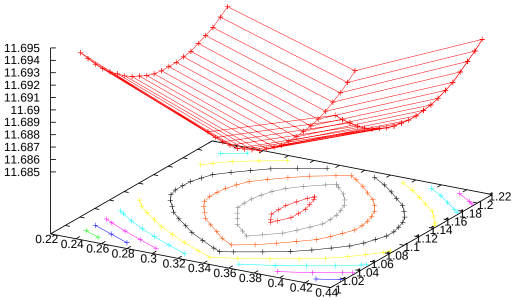
$$\begin{aligned} & (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, y_1, y_2, y_3) \\ \in\ & [-30,10]\times[-30,10]\times[-30,10]\times[-30,10]\times[-30,20]\times[-30,- \\ & \times[-30,-0.01]\times[-30,-0.001]\times[-30,10] \end{aligned}$$

where $(S_{11}, S_{12}, S_{13}, S_{22}, S_{23}, S_{33}, S_1, S_2, S_3) =$
$(0.3119, 0.0292, 0.0707, 0.3605, 0.0462, 0.3276, -0.0063, -0.0054, -0.076$

The result is that the minimum 11.68573121328159669 is taken at

$$x = \begin{pmatrix} -0.161 & 0.3377/2 & 1.1104/2 \\ 0.3377/2 & 0.2538 & 0.6424/2 \\ 1.1104/2 & 0.6424/2 & -0.0928 \end{pmatrix},$$

$$y = (\underline{-0.019}, \underline{-0.0162}, -0.2286)$$



Analogy

Minimize $F(x)\mathrm{exp}(\mathrm{g}(\mathrm{s},\mathrm{x}))$. s is derived from statistical data.
The target function satisfies a holonomic system of differential equations.

IP	GB gives a flow to the optimal value
HGD	GB in R gives a flow to the optimal value

$$R = \mathbf{C}(x_1, \ldots, x_d)\langle \partial_1, \ldots, \partial_d \rangle$$

where we denote by $\mathbf{C}(x_1, \ldots, x_d)$ the field of rational functions in $x_1, \ldots, x_d$. $\partial_i \partial_j = \partial_j \partial_i$ and $\partial_i a(x) = a(x)\partial_i + \frac{\partial a}{\partial x_i}$ for $a(x) \in \mathbf{C}(x_1, \ldots, x_d)$.

Future

1. Numerical difficulties of singular locus ⇒ resolution of singularities, series solutions around singularities.
2. Full automatic analysis of integrals (normalization constants) ⇒ integration algorithms for D-modules.
3. Theoretical study of normalization constants ⇒ hypergeometric differential equations.
4. Non-linear equations for the target function ⇒ differential algebra.
5. Finding the minimum for polynomial approximations ⇒ SDPR