Geometry of Measure-Preserving Flows and Hamiltonian Monte Carlo

Alessandro Barp (with Girolami, Betancourt, Kennedy, Lelievre...)

July 27, 2020

Imperial College London, University of Cambridge, Alan Turing Institute

Plan

1. Sampling a Measure via Measure-preserving Continuous Flows
2. Euclidean Measure-preserving Diffusions; Connections with Poisson mechanics
3. Inspire Geometry of Smooth Measures
4. Geometry of Measure-preserving Flows
5. Measure-preserving Flows Vs Mechanics
6. Hamiltonian Monte Carlo
7. Geometric Integration and Implementation of HMC

Aim

Construct efficient sampling methods on manifolds for unnormalised smooth distributions using Measure-Preserving Flows

- Given a target $P \propto p_\infty \mu_\mathcal{M} = e^{-V} \mu_\mathcal{M}$ on $\mathcal{M}$, we want to generate a "sample" $(X_i)_{i=1}^N$ (e.g. MCMC), to approximate

$$P \approx P_N \equiv \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$$

- Usually want a.s. narrow/weak* convergence $P_N \to P$ to approximate expectation

$$\mathbb{E}[f] \equiv \int f \mathrm{d}P \approx \frac{1}{N} \sum_{i=1}^N f(X_i)$$

- Build (X_i) using P-preserving flows (dynamics, mechanics, diffusions)
- Examples: Fisher-Bingham distributions on Stiefel manifolds for directional statistics/principal component analysis, canonical distributions in molecular dynamics on holonomic manifold, distributions on covariances and Hermitian positive matrices for learning spectral density matrix, discrete actions on gauge groups...

Two standard strategy to build samplers:

- Suppose $S : \mathcal{M} \to \mathcal{M}$ is a P-preserving, and $\Psi_{\delta t} : \mathcal{M} \to \mathcal{M}$ is S-reversible:

$$\Psi_{\delta t}^{-1} = S \circ \Psi_{\delta t} \circ S^{-1}$$

  MCMC: Given $q^\ell \in \mathcal{M}$:
  1. $q_* \leftarrow \Psi_{\delta t}(q^\ell)$
  2. set $q^{\ell+1} \leftarrow q_*$ with probability $\min\left(1, |\mathcal{J}(P, \Psi_{\delta t})|(q^\ell)\right)$, else $q^{\ell+1} \leftarrow S(q^\ell)$,

  where $|\mathcal{J}(P, \Psi_{\delta t})| \equiv \mathrm{d}\Psi_{\delta t}^{*} P / \mathrm{d}P$
- Take a P-preserving diffusion and approximate it, or break it into tractable P-preserving components. For example, HMC

$$\underbrace{\mathrm{d}Q_t = G^{-1} P_t \mathrm{d}t, \qquad \mathrm{d}P_t = -\nabla V(Q_t)\mathrm{d}t}_{\text{Hamiltonian Mechanics}} \underbrace{-\gamma(Q_t) G^{-1} P_t \mathrm{d}t + \sigma(Q_t)\mathrm{d}W_t}_{\text{OU heat bath}}$$

- Samplers are physics-inspired, but we do not "care" about physics (we care about ergodicity, rate of convergence...)

Measure Preserving Diffusions on $\mathbb{R}^n$

- Under some $\mathrm{L}^1(\mathrm{d}x)$-integrability and uniqueness assumptions, a $P \propto e^{-V}\mathrm{d}x$ diffusion on $\mathbb{R}^n$ has the form [Ma et al., 2015, Thm. 2]

$$\mathrm{d}Z_t = -\left(Q\nabla V + D\nabla V\right)\mathrm{d}t + \nabla\cdot\big(Q+D\big)\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t\,, \qquad (1)$$

  Q antisymmetric, D positive semi-definite (eg Langevin/metriplectic).

- Proof uses Fourier transforms to turn the question into a linear algebra problem in Fourier space.
- Not clear how to generalise this construction to manifolds.
- Intuitive generalisation to manifolds: start by replacing

$$\begin{aligned} \mathrm{d}x &\mapsto \mu_{\mathcal{M}}, \qquad & Q\nabla V &\mapsto X_V^{\mathcal{B}} \equiv \mathcal{B}^\sharp(\mathrm{d}V), \\ \nabla\cdot Q &\mapsto Y, \qquad & \sqrt{2D}\,\mathrm{d}W_t &\mapsto Y_i \circ \mathrm{d}W_t^i, \end{aligned}$$

  where $\mathcal{B}^\sharp \in \mathrm{Hom}(T^*\mathcal{M}, T\mathcal{M})$.

- If $\mathrm{d}Z_t = X\mathrm{d}t + Y_i \circ \mathrm{d}W_t^i$, then
$$\mathcal{L}^* f = \mathrm{div}_{\mu_\mathcal{M}} \left(-fX + \tfrac{1}{2} Y_i(f) Y_i + \tfrac{1}{2} f \mathrm{div}_{\mu_\mathcal{M}}(Y_i) Y_i\right).$$
- Thus, to have $\mathcal{L}^* e^{-V} = 0$ when $\mathcal{B} = 0$, we set
$$\mathrm{d}Z_t = \left(X_V^\mathcal{B} + Y\right)\mathrm{d}t + \left(-\tfrac{1}{2} Y_i(V) Y_i + \tfrac{1}{2}\mathrm{div}_{\mu_\mathcal{M}}(Y_i) Y_i\right)\mathrm{d}t + Y_i \circ \mathrm{d}W_t^i. \quad (2)$$
- The bracket diffusion (2) satisfies $\mathcal{L}^* p_\infty = 0$ if and only if Y satisfies
$$\mathrm{div}_{\mu_\mathcal{M}}(X_{p_\infty}^\mathcal{B} - p_\infty Y) = 0. \quad (3)$$
- This should hold for all p_∞, so $\mathrm{div}_{\mu_\mathcal{M}}(Y) = 0$, and
$$\mathrm{div}_{\mu_\mathcal{M}}(X_{p_\infty}^\mathcal{B}) = Y(p_\infty), \qquad \forall p_\infty. \quad (4)$$
- Y vector field, implies $\mathcal{B} \equiv \mathcal{A}$ antisymmetric (as a rank two tensor)
- (4) is precisely the definition of the modular vector field $Y \equiv X_\mathcal{B}^{\mu_\mathcal{M}}$ in Poisson mechanics [Dufour and Haraki, 1991, Weinstein, 1997]

$$\mathrm{d}Z_t = \overbrace{\underbrace{X_V^\mathcal{A}\mathrm{d}t}_{e^{-V}\text{-preserving}} + \underbrace{\beta^{-1} X_\mathcal{A}^{\mu_\mathcal{M}}\mathrm{d}t}_{\mu_\mathcal{M}\text{-preserving}}}^{e^{-\beta V}\mu_\mathcal{M}\text{-preserving}} \overbrace{- \tfrac{1}{2}\beta Y_i(V) Y_i \mathrm{d}t + \underbrace{\tfrac{1}{2}\mathrm{div}_{\mu_\mathcal{M}}(Y_i) Y_i \mathrm{d}t + Y_i \circ \mathrm{d}W_t^i}_{\mu_\mathcal{M}\text{-preserving}}}^{e^{-\beta V}\mu_\mathcal{M}\text{-preserving}},$$

Canonical Geometry of Smooth Measures

- Is our generalisation complete? To answer we develop intrinsic geometry of target P
- Let P be smooth measure, locally $P = f|\mathrm{d}x|$. Denote by $P^\flat$ the morphism $P^\flat(X) \equiv i_X P$ on $\mathfrak{X}^k(\mathcal{M})$. If P positive, we have an inverse $P^\sharp$ (R-N).
- The P-rotationnel of a k-vector field for some integer $1 \leq k \leq n$ is defined as [Koszul, 1985]

$$\mathrm{curl}_P \equiv P^\sharp \circ \mathrm{d} \circ P^\flat : \mathfrak{X}^k(\mathcal{M}) \to \mathfrak{X}^{k-1}(\mathcal{M}).$$

- $\mathrm{curl}_P \circ \mathrm{curl}_P = 0$, boundary operator. On vector fields, $\mathrm{curl}_P = \mathrm{div}_P$. Generalise

$$\nabla \cdot \nabla \times = 0, \qquad \delta = \star \mathrm{d} \star .$$

- On bivectors: Modular field $X_{\mathcal{A}}^P = -\mathrm{curl}_P(\mathcal{A})$.
- Canonical statistical calculus. Only depend on P up to normalisation.
- Smooth measure defines P-homology, which is isomorphic to the (twisted) de Rham cohomology $\implies$ curl_P-free fields can be represented as closed forms.

Complete Recipe on Manifolds

- If $\mathrm{d}Z_t = X\mathrm{d}t + Y_i \circ \mathrm{d}W_t^i$, then

$$\mathcal{L}^* P = \mathrm{div}_P \Big(\underbrace{\tfrac{1}{2}\mathrm{div}_P(Y_i) Y_i - X}_{\text{Fokker–Plank current}} \Big) P,$$

so $\mathcal{L}^* P = 0$ iff

$$\tfrac{1}{2}\mathrm{div}_P(Y_i)Y_i - X = \mathrm{curl}_P(\mathcal{A}) + P^\sharp(\gamma), \qquad \mathcal{A} \in \mathfrak{X}^2(\mathcal{M}), \gamma \in H_{dR}^{n-1}(\mathcal{M})$$

- P-preserving diffusions

$$\mathrm{d}Z_t = \underbrace{\overbrace{-\mathrm{curl}_P(\mathcal{A})\mathrm{d}t}^{\text{Fokker–Planck potential}} + \overbrace{P^\sharp(\gamma)\mathrm{d}t}^{\text{topological obstruction}}}_{\text{conservative } L^2(P)\text{-antisymmetric}} + \overbrace{\underbrace{\tfrac{1}{2}\mathrm{div}_P(Y_i)Y_i\mathrm{d}t}_{\text{dissipative drift}} + \underbrace{Y_i \circ \mathrm{d}W_t^i}_{\text{Stratonovich noise}}}^{L^2(P)\text{-symmetric fluctuation-dissipation balance}} . \tag{5}$$

- Complete Recipe ✓, canonical ✓, no integrability assumption ✓
- Compare with

$$\mathrm{d}Z_t = -\left(Q\nabla V + D\nabla V\right)\mathrm{d}t + \nabla \cdot \big(Q + D\big)\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t.$$

- Completeness is based on the fact that

$$\ker \mathrm{div}_P = \mathrm{curl}_P(\mathfrak{X}^2(\mathcal{M})) \oplus P^\sharp(H_{dR}^{\dim \mathcal{M}-1}(\mathcal{M}))$$

- Potential Theory of Measures: P-preserving flow are "locally curled" - correspond to a choice of "potential" $\mathcal{A}$,

$$\underbrace{X = \mathrm{curl}_P(\mathcal{A})}_{\text{statistics}}, \qquad \text{just as} \qquad \underbrace{F = -\mathrm{d}V}_{\text{conservative mechanics}} \quad , \qquad \underbrace{\mathbf{F} = \mathrm{d}\mathbf{A}}_{\text{electromagnetism}}$$

Many connections between P-flows and Hamiltonian mechanics

1. Locally, $P \propto p_\infty |\mathrm{d}x|$

$$X|_U = \sum_{i<j} X_{\mathcal{A}^{ij}} + i_{\mathrm{d}\log p_\infty} \mathcal{A}$$

   where $X_{\mathcal{A}^{ij}} = \partial_j \mathcal{A}^{ij} \partial_i - \partial_i \mathcal{A}^{ij} \partial_j$.
2. If X, Y preserve P, then

$$[X, Y] = \mathrm{curl}_P(X \wedge Y),$$

   just as symplectic vector fields.

Measure Preserving Mechanics

- In general we can decompose

$$\mathrm{curl}_P(\mathcal{A}) = \underbrace{\mathrm{curl}_{\mu_{\mathcal{M}}}(\mathcal{A})}_{\mu_{\mathcal{M}}\text{-preserving}} + \underbrace{i_{\mathrm{d}\log p_\infty}\mathcal{A}}_{p_\infty\text{-preserving}}.$$

- What potentials give rise to score-based P-preserving flows?
- Recall $\mathrm{curl}_{\mu_{\mathcal{M}}}(\mathcal{A}) = -X_{\mathcal{A}}^{\mu_{\mathcal{M}}} : f \mapsto \mathrm{div}_{\mu_{\mathcal{M}}}(X_f^{\mathcal{A}})$.
- Thus $\mu_{\mathcal{M}}$ is invariant measure for $\mathcal{A}$-mechanics

$$\{X_f^{\mathcal{A}} \equiv i_{\mathrm{d}f}\mathcal{A} : f \in C^\infty(\mathcal{M})\}$$

iff $\mathrm{curl}_{\mu_{\mathcal{M}}}(\mathcal{A}) = 0$; $\implies$ space of $\mu_{\mathcal{M}}$-preserving $\mathcal{A}$-mechanics is

$$\ker \mathrm{curl}_{\mu_{\mathcal{M}}}|_{\mathfrak{X}^2(\mathcal{M})} = \mathrm{curl}_{\mu_{\mathcal{M}}}(\mathfrak{X}^3(\mathcal{M})) \oplus \mu_{\mathcal{M}}^\sharp(H_{dR}^{\dim\mathcal{M}-2}(\mathcal{M}))$$

- Historical Remark: Volterra/de Rham/Koszul/Dufour/Weinstein $\implies$ Geometry = Statistics (Stein discrepancy, HMC...)

$$\text{Machine Learning 2015} < \text{ Mathematical Physics 1887.}$$

- Splitting Methods: if $p_\infty = \prod_j e^{-V_j}$

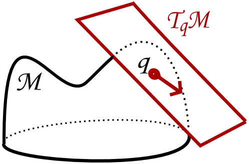
$$\mathrm{curl}_{p_\infty\mu_{\mathcal{M}}}(\mathcal{A}) = i_{\mathrm{d}\log p_\infty}\mathcal{A} = -\sum_j i_{\mathrm{d}V_j}\mathcal{A},$$

no "Jacobian", MC is simply energy difference.

- Given $P = e^{-V}\mu_{\mathcal{M}}$, where $\mu_{\mathcal{M}}$ is Riemannian measure. Use mechanics to propose new sample by viewing V as a potential energy

$$\underbrace{m\ddot{q} = -\partial V}_{\text{Flat Newton}} \quad \longrightarrow \quad \underbrace{\frac{\nabla \dot{q}}{\mathrm{d}t} = -\nabla V}_{\textit{Riemannian Newton}}$$

- 2^{nd}-order, tangent bundle flow



Solution preserves

$$\mu_H \propto e^{-H(q,v)}\omega_\flat^n \equiv e^{-\frac{1}{2}\|v\|_q^2 - V(q)}\omega_\flat^n, \quad \omega_\flat^n \text{ symplectic measure,}$$

on $T\mathcal{M}$, and

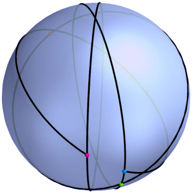
$$\mathrm{Proj}_* \mu_H = P.$$

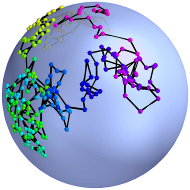
- Flow of mechanics preserves μ_H, projection of μ_H-samples are P-samples

- A1: Hamiltonian Mechanics is a natural volume-preserving mechanics
- A2: velocity flip $S(q, v) = (q, -v)$ preserves μ_H, so if integrator $\Psi_{\delta t} : T\mathcal{M} \to T\mathcal{M}$ is

  $$S\text{-reversible} \quad \Psi_{\delta t}^{-1} = S \circ \Psi_{\delta t} \circ S; \text{ volume preserving} \quad (\Psi_{\delta t})_* \omega_\flat^n = \omega_\flat^n,$$

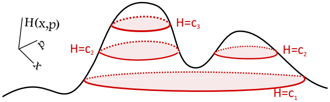
  then MDMC
  - Let $z^* \equiv \Psi_{\delta t}(z^n)$
  - accept z^* with probability $\min\left(1, e^{-(H(z^*)-H(z^n))}\right)$. If accepted, then $z^{n+1} \equiv z^*$. Else $z^{n+1} \equiv S(z^*)$.
- A3: If we know geodesics of $\mu_\mathcal{M}$, can use geodesic integrators $X_H = \frac{1}{2}X_V + X_T + \frac{1}{2}X_V$





Hamiltonian Monte Carlo

Ergodicity: $\mathcal{A}$-mechanics preserve energy, and we want small energy difference during numerical integration for good acceptance rate $\longrightarrow$ but then we get stuck in level sets $H = c$

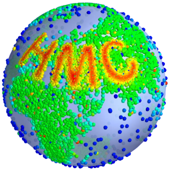


A4: We can simply add Gaussian heat bath to MDMC associated to lift

$$\mu_H = \pi^* P \wedge Gaussian.$$

and obtain

$$\text{HMC} = \text{MDMC} + \text{heat bath}$$



Energy Conservation: Shadow Hamiltonian

- Can use Mechanical Integrator: preserving energy, symmetries or $\omega_\flat$ (cant have all three).
- A5: If symplectic: $\Psi^*_{\delta t}\omega_\flat = \omega_\flat$, using Hamilton-Jacobi theorem/Jacobi identity there exist nearby shadow Hamiltonian whose flow is $\Psi_{\delta t} \implies$ acceptance-rate remains high
- Unlike "Theory of Numerical Integrators": we don't care about correct trajectories !
- A6: Theory of symplectic integrators:
  - Hamiltonian: splitting method:

$$H = \underbrace{\tfrac{1}{2}\|\cdot\|^2}_{geodesic\ \ flow} + \underbrace{V}_{\text{vertical gradient step}} , \qquad V = V_{hard} + V_{easy}$$

  - Lagrangian: discrete variational principle $\to$ symmetry for free
  - Generating Functions and Hamilton-Jacobi PDE

$$F : T^*\mathcal{M} \to T^*\mathcal{M} \quad \text{is symplectic iff} \quad \mathrm{d}\iota_F^*\Xi = 0$$

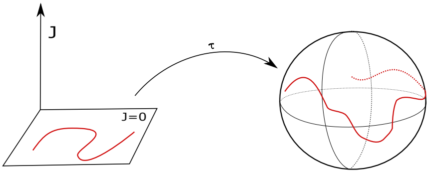
  where $\Xi \equiv \pi_1^*\Theta - \pi_2^*\Theta$. Thus locally $\iota_F^*\Xi = \mathrm{d}S$

$$p = -\frac{\partial S}{\partial q}(q, Q), \qquad P = \frac{\partial S}{\partial Q}(q, Q).$$

- Shadow for Splitting + Invariant Measures $\implies$ Unimodular Poisson

Implementing HMC

- Every potential on $\mathcal{M} = \mathcal{G}/\mathcal{K}$ is a potential with symmetry on $\mathcal{G}$
- Consider right action of $\mathcal{K}$ on $\mathcal{G}$, momentum map $J : \mathcal{G} \times \mathfrak{g} \to \mathfrak{k}^*$
- If action is Hamiltonian and system $\mathcal{K}$-invariant, reduced space is target space $J^{-1}(0)/\mathcal{K} \cong T\mathcal{M}$



For geodesic orbit manifolds, all geodesics are homogeneous. For naturally reductive $J^{-1}(0) = \mathcal{G} \times \mathfrak{p}$, and HMC is straightforward [Barp et al., 2019].

- For holonomic manifolds $\mathcal{M} = f^{-1}(0)$, can use RATTLE for kinetic step...
- ... but must add reversibility check! [Lelièvre et al., 2018]

Partition function of lattice QCD

$$Z = \int \prod_{x,\mu} \mathrm{d}U_\mu(x) \mathrm{d}\phi^\dagger \mathrm{d}\phi e^{-S_{WG} - \phi^\dagger (DD^\dagger)^{-1} \phi}.$$

Here $U_\mu(x) \in \mathrm{SU}(3)$ is discretised gauge field, $\mathrm{d}U_\mu(x)$ is Haar measure, ϕ pseudofermions, S_{WG} is Wilson gauge action (discretisation of Yang-Mills action), D is Wilson–Dirac operator (discretised Dirac operator).

Introduce fictitious momenta on the links, to obtain

$$H = S_{WG} + \phi^\dagger (DD^\dagger)^{-1} \phi + \tfrac{1}{2} \sum_{x,\mu} \langle p_{x,\mu}, p_{x,\mu} \rangle_{\mathfrak{su}(3)} .$$

Need to construct mechanics on $\mathrm{SU}(3)$. Define

$$\omega \equiv -\mathrm{d}(p_i \pi^* \theta^i) = \underbrace{\pi^* \theta^i \wedge \mathrm{d}p_i}_{\text{usual ``}\mathrm{d}x \wedge \mathrm{d}p\text{'' term}} + \underbrace{\tfrac{1}{2} p_i c^i_{jk} \pi^* \theta^j \wedge \pi^* \theta^k}_{\text{additional non-abelian term}},$$

and use representations.

Riemannian Manifold HMC

- Target is posterior $P = \rho_{post}(\theta|\omega)\mathrm{d}\theta$. Want Riemannian metric on $\mathcal{M}$ that locally matches Hessian of posterior

$$\Sigma_{post,\omega}(\theta) \equiv -\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log\rho_{post}(\theta|\omega)$$

- Average over data (statistical manifold $\phi(\theta) \equiv \rho(\omega|\theta)\mathrm{d}\omega$)

$$\int \Sigma_{post,\omega}(\theta)\rho(\omega|\theta)\mathrm{d}\omega = \underbrace{\phi^* g^F}_{\text{Fisher Matrix}}(\theta) + \Sigma_{prior}(\theta)$$

- setting $G(\theta) \equiv \phi^* g^F(\theta) + \Sigma_{prior}$, the target Hamiltonian of RMHMC is then

$$H(\theta, v) = -\log\rho_{post}(\theta|\omega) + \underbrace{\tfrac{1}{2}\log\det G(\theta)}_{measure\ \ correction} + \tfrac{1}{2}v^\top G(\theta)v.$$

- Fake reference measures... typically canonical (not necessarily geodesic)
- “Non-separable” Hamiltonian - symmetry-break
- No manifold involved! Should be Statistical Model/Information Geometric HMC
- RM/Geodesic/Lagrangian Monte Carlo $\longrightarrow$ just HMC

HMC for Molecular Dynamics

- Typical Applications: Molecular constraints, Blue Moon Sampling, Thermodynamic Integration
- Molecular Constraints define holonomic manifolds: RATTLE/SHAKE but need reversibility check
- Blue Moon Sampling idea: microcanonical distribution ($\mu_{mc,E}$)

$$\mathbb{E}_{\mu_H}[f] = \frac{1}{\int_{\mathcal{F}} e^{-H}\omega^n} \int_{\mathbb{R}} \mathbb{E}_{\mu_{mc,E}}[f] e^{-E} \mathrm{d}(E) \mathrm{d}E.$$

- Thermodynamic Integration: Macroscopic states are often defined using reaction coordinates $\xi : \mathcal{M} \to \mathbb{R}^m$, with Free energy $F : \mathbb{R}^m \to \mathbb{R}$

$$F \equiv -\frac{1}{\beta} \log \frac{\mathrm{d}(\pi^*\xi)_\sharp \mu_H}{\mathrm{d}(\mathrm{d}x)}.$$

- Want to calculate energy difference

$$F(x_1) - F(x_0) = \int_0^1 \frac{\partial F}{\partial x^i}\Big|_{\ell(t)} \frac{\mathrm{d}\ell^i}{\mathrm{d}t}\Big|_t \mathrm{d}t, \qquad \frac{\partial F}{\partial x^i}(x) = \int_{\xi^{-1}(x)} \cdots$$

Barp, A., Kennedy, A., and Girolami, M. (2019).
Hamiltonian monte carlo on symmetric and homogeneous spaces via symplectic reduction.
arXiv preprint arXiv:1903.02699.

Dufour, J.-P. and Haraki, A. (1991).
Rotationnels et structures de poisson quadratiques.
Comptes rendus de l'Académie des sciences. Série 1, Mathématique, 312(1):137–140.

Koszul, J.-L. (1985).
Crochet de schouten-nijenhuis et cohomologie.
Astérisque, 137:257–271.

Lelièvre, T., Rousset, M., and Stoltz, G. (2018).
Hybrid monte carlo methods for sampling probability measures on submanifolds.
arXiv preprint arXiv:1807.02356.

Ma, Y.-A., Chen, T., and Fox, E. (2015).
A complete recipe for stochastic gradient mcmc.
In *Advances in Neural Information Processing Systems*, pages 2917–2925.

Weinstein, A. (1997).
The modular automorphism group of a poisson manifold.
Journal of Geometry and Physics, 23(3-4):379–394.