Hypo-elliptic simulated annealing

Christian Bayer[1] Josef Teichmann[2] Richard Warnung[3]

[1]Department of Mathematics
Royal Institute of Technology, Stockholm

[2]Department of Mathematics
ETH Zurich

[3]Raiffeisen Capital Investment, Vienna

07/30/2009 Berlin

Outline

Smoluchowski dynamics

$$dY_t^y = -\frac{1}{2}\nabla U(Y_t^y)dt + \sqrt{KT}dW_t \qquad (1)$$

- $Y_0^y = y \in \mathbb{R}^n$, $U : \mathbb{R}^n \to \mathbb{R}$ and W is an n-dimensional standard Wiener process
- Unique invariant measure given by the *Gibbs measure*

$$\mu_{KT}(dy) = \frac{1}{C_T} e^{-\frac{U(y)}{KT}} dy$$

- Ergodicity, i.e., Y^y converges in law to μ_{KT}
- Extensively used in physics and chemistry for the simulation of *canonical ensemble* (NVT), i.e., of an ensemble with constant temperature

Smoluchowski dynamics

$$dY_t^y = -\frac{1}{2}\nabla U(Y_t^y)dt + \sqrt{KT}dW_t \tag{1}$$

- $Y_0^y = y \in \mathbb{R}^n$, $U : \mathbb{R}^n \to \mathbb{R}$ and W is an n-dimensional standard Wiener process
- Unique invariant measure given by the *Gibbs measure*

$$\mu_{KT}(dy) = \frac{1}{C_T} e^{-\frac{U(y)}{KT}} dy$$

- Ergodicity, i.e., Y^y converges in law to μ_{KT}
- Extensively used in physics and chemistry for the simulation of *canonical ensemble* (NVT), i.e., of an ensemble with constant temperature

Smoluchowski dynamics

$$dY_t^y = -\frac{1}{2}\nabla U(Y_t^y)dt + \sqrt{KT}dW_t \tag{1}$$

- $Y_0^y = y \in \mathbb{R}^n$, $U : \mathbb{R}^n \to \mathbb{R}$ and W is an n-dimensional standard Wiener process
- Unique invariant measure given by the *Gibbs measure*

$$\mu_{KT}(dy) = \frac{1}{C_T} e^{-\frac{U(y)}{KT}} dy$$

- Ergodicity, i.e., Y^y converges in law to μ_{KT}
- Extensively used in physics and chemistry for the simulation of *canonical ensemble* (NVT), i.e., of an ensemble with constant temperature

Motivation of simulated annealing

For $T \to 0$, the Gibbs measure

$$\mu_{KT}(dy) = \frac{1}{C_T} e^{-\frac{U(y)}{KT}} dy \rightsquigarrow \arg\min_{y \in \mathbb{R}^n} U(y).$$

Choose a “cooling schedule” $T = T(t)$ such that

- the “instantaneously invariant” measures $\mu_{KT(t)}$ concentrate around $\arg\min U(y)$ for $t \to \infty$,
- the (time-inhomogeneous) Markov process

$$dY_t^y = -\frac{1}{2}\nabla U(Y_t^y)dt + \sqrt{KT(t)}dW_t$$

remains ergodic.

Spectral gap

Consider the infinitesimal generator L^σ of the Smoluchowski SDE with constant $\sigma = KT$:

$$L^\sigma f = -\frac{1}{2}\langle \nabla U, \nabla f\rangle + \frac{\sigma}{2}\Delta f.$$

- L^σ is symmetric on $L^2(\mathbb{R}^n, \mu_\sigma)$ and has a discrete, non-positive spectrum.
- Spectral gap

$$\lambda^\sigma = -\frac{\sigma}{2}\inf\left\{ ||\nabla f||_{L^2(\mu_\sigma)} \;\middle|\; ||f||_{L^2(\mu_\sigma)} = 1 \text{ and } \int_{\mathbb{R}^n} f d\mu_\sigma = 0 \right\}$$

- λ^σ controls the convergence of the distribution of the Smoluchowski process to its invariant distribution.

Elliptic simulated annealing

Assume that

- U is smooth,
- U and $||\nabla U||$ converge to infinity for $||x|| \to \infty$,
- $||\nabla U||^2 - \Delta U$ is bounded from below.

Theorem

Given a decreasing cooling schedule $\sigma = \sigma(t) \to 0$, such that $\sigma(t) = k/\log(t)$ for some constant $k > c$, where

$$c = \lim_{t \to \infty} -\sigma(t) \log(\lambda^{\sigma(t)}).$$

Then the law of the process $Y_t^{\sigma(\cdot)}$ converges to a distribution supported on $\arg\min_y U(y)$.

Heisenberg groups – construction

We consider the *Heisenberg groups* as a prototype of the state space of hypo-elliptic simulated annealing.

- $\mathbb{A}_p^2$ denotes the space of non-commutative polynomials of degree 2 in p variables $\{e_1, \dots, e_p\}$.
- Forms a *nil-potent* associative algebra of degree two.
- $\mathfrak{g}_p^2$ denotes the *Lie algebra* generated by $\{e_1, \dots, e_p\}$ (with respect to $[x, y] = xy - yx, \; x, y \in \mathbb{A}_p^2$).
- Define $\exp : \mathfrak{g}_p^2 \to \mathbb{A}_p^2$ by its (nil-potent) power series and set

$$G_p^2 := \exp\left(\mathfrak{g}_p^2\right).$$

Heisenberg groups – properties

- $\exp : \mathfrak{g}_p^2 \to G_p^2$ is a global chart of the *Lie group* G_p^2.
- Representation as matrix group, e.g., for $p = 2$:

$$G_2^2 \simeq \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \middle| \, a, b, c \in \mathbb{R} \right\}.$$

- $\dim(G_p^2) = (p+1)p/2$
- Brownian motion X_t defined by $X_0 = 1 \in G_p^2$ and

$$dX_t = \sum_{i=1}^{p} X_t e_i \circ dB_t^i$$

- Natural to use X_t as driving noise for simulated annealing on G_p^2 or $\mathfrak{g}_p^2$

Heisenberg groups – properties

- $\exp : \mathfrak{g}_p^2 \to G_p^2$ is a global chart of the *Lie group* G_p^2.
- Representation as matrix group, e.g., for $p = 2$:

$$G_2^2 \simeq \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \middle| \, a, b, c \in \mathbb{R} \right\}.$$

- $\dim(G_p^2) = (p+1)p/2$
- Brownian motion X_t defined by $X_0 = 1 \in G_p^2$ and

$$dX_t = \sum_{i=1}^{p} X_t e_i \circ dB_t^i$$

- Natural to use X_t as driving noise for simulated annealing on G_p^2 or $\mathfrak{g}_p^2$

Elliptic vs. hypo-elliptic simulated annealing

Starting point of elliptic simulated annealing

A small stochastic perturbation of a classical gradient flow allows the flow to overcome local minima (having the Gibbs measure as invariant distribution). Decreasing the perturbation slowly enough induces convergence to $\arg\min U$.

Starting point of hypo-elliptic simulated annealing

Given a driving, hypo-elliptic process. Want to construct a Markov process with invariant measure given by the Gibbs measure. How does the appropriate drift look like?

Elliptic vs. hypo-elliptic simulated annealing

Starting point of elliptic simulated annealing

A small stochastic perturbation of a classical gradient flow allows the flow to overcome local minima (having the Gibbs measure as invariant distribution). Decreasing the perturbation slowly enough induces convergence to $\arg\min U$.

Starting point of hypo-elliptic simulated annealing

Given a driving, hypo-elliptic process. Want to construct a Markov process with invariant measure given by the Gibbs measure. How does the appropriate drift look like?

Homogeneous spaces

Definition

Given a Lie group G. A finite-dimensional smooth manifold M such that

- there is a right-action $r : M \times G \to M$, i.e., $r(r(x,g),h) = r(x,gh)$, $x \in M$, $g,h \in G$,
- r is transitive, i.e., $\forall x \in M$, $g \mapsto r(x,g)$ is surjective,

is called homogeneous space.

Projection: for a fixed point $o \in M$, define $\pi : G \to M$ by $\pi(g) = r(o,g)$.

Setting of hypo-elliptic simulated annealing

Assumption A

1. *G is a finite-dimensional, connected Lie group with Lie algebra $\mathfrak{g}$ and right-invariant Haar measure η.*
2. *M is a compact homogeneous space w.r.t. G with a positive finite measure η^M invariant w.r.t. the right action of G on M.*
3. $U \in C^\infty(M; [0, \infty[)$.

Remark

Assumption A is satisfied by compact Lie groups like $SO(3)$, acting on themselves.

Heisenberg torus

- G_p^2 is not compact.
- Define a discrete sub-group of G_p^2.
  - $\mathfrak{l}_p^2 = \langle \{ e_i \mid i = 1, \dots, p \} \cup \{ \frac{1}{2}[e_i, e_j] \mid i < j \} \rangle_{\mathbb{Z}} \subset \mathfrak{g}_P^2$
  - $L_p^2 := \exp(\mathfrak{l}_p^2) \subset G_p^2$
- $M := L_p^2 \diagdown G_p^2 = \{ L_p^2 g \mid g \in G_p^2 \}$ is called *Heisenberg torus* (not a group!).
- Construct a right-invariant measure η^M from the Lebesgue measure on $\mathbb{T}^{p(p+1)/2}$ using $M \simeq \mathbb{T}^{p(p+1)/2}$.
- E.g.,, for $p = 2$, set $e_3 = \frac{1}{2}[e_1, e_2]$ and define $\phi : \mathfrak{g}_2^2 \to \mathbb{T}^3$ by

$$\phi(z_1 e_1 + z_2 e_2 + z_3 e_3) = ([z_1], [z_2], [z_3 - [z_2]z_1 + [z_1]z_2]),$$

  where $[z] := z \bmod 1$. Get a diffeom. $\phi \circ \exp^{-1} : L_2^2 \diagdown G_2^2 \to \mathbb{T}^3$.
- M together with η^M satisfies Assumption A.

Hypo-ellipticity

Assumption B

Let $n = \dim(G)$ *and assume that there are* $d < n$ *left-invariant vector-fields* $V_1, \ldots, V_d$ *on* G*, which already generate* $\mathfrak{g}$ *(Hörmander condition). Define a stochastic process* X_t *on* G *with* $X_0 = 1 \in G$ *and*

$$dX_t = \sum_{i=1}^{d} V_i(X_t) \circ dB_t^i.$$

- X_t is hypo-elliptic, i.e., has a smooth transition density.
- For $G = G_p^2$, $d = p < \frac{p(p+1)}{2} = n$ and $V_i(x) = xe_i$, $i = 1, \ldots, p$.

Hypo-ellipticity

Assumption B

Let $n = \dim(G)$ *and assume that there are* $d < n$ *left-invariant vector-fields* $V_1, \ldots, V_d$ *on* G*, which already generate* $\mathfrak{g}$ *(Hörmander condition). Define a stochastic process* X_t *on* G *with* $X_0 = 1 \in G$ *and*

$$dX_t = \sum_{i=1}^{d} V_i(X_t) \circ dB_t^i.$$

- X_t is hypo-elliptic, i.e., has a smooth transition density.
- For $G = G_p^2$, $d = p < \frac{p(p+1)}{2} = n$ and $V_i(x) = xe_i$, $i = 1, \ldots, p$.

Sub-Riemannian gradient

- Drift $-\frac{1}{2}\nabla U$ is the direction of steepest descent in a Riemannian environment.
- Noise X_t can traverse the whole space, but locally only *horizontal* directions are possible.
- *Carré-du-champs* operator

$$\Gamma(f,g)(x) = \sum_{i=1}^{d} V_i f(x) V_i g(x), \quad x \in G$$

- $\Gamma(U,\cdot) : f \mapsto \Gamma(U,f)$ defines a left-invariant, horizontal vector-field.

Sub-Riemannian gradient

- Drift $-\frac{1}{2}\nabla U$ is the direction of steepest descent in a Riemannian environment.
- Noise X_t can traverse the whole space, but locally only *horizontal* directions are possible.
- *Carré-du-champs* operator

$$\Gamma(f,g)(x) = \sum_{i=1}^{d} V_i f(x) V_i g(x), \quad x \in G$$

- $\Gamma(U,\cdot) : f \mapsto \Gamma(U,f)$ defines a left-invariant, horizontal vector-field.

Hypo-elliptic Smoluchowski dynamics

- $\sigma = \sigma(t)$ cooling schedule.
- Push the vector fields $V_1, \ldots, V_d$ and $\Gamma(U, \cdot)$ to vector fields $V_1^M, \ldots, V_d^M$ and $\Gamma^M(U, \cdot)$ on M using $\pi : G \to M$.
- Define the Smoluchowski dynamics on M

$$dY_t = -\frac{1}{2}\Gamma^M(U, \cdot)(Y_t)dt + \sqrt{\sigma(t)} \sum_{i=1}^{d} V_i^M(Y_t) \circ dB_t^i.$$

- Locally invariant Gibbs measure

$$\mu_\sigma(dx) = \frac{1}{C_\sigma} \exp\left(-\frac{U(x)}{\sigma}\right) \eta^M(dx).$$

Hypo-elliptic simulated annealing

Theorem (Baudoin, Hairer and Teichmann)

Let $U_0 = \min_{x \in M} U(x)$. There are constants $R, c > 0$ such that the simulated annealing process Y_t with

$$\sigma(t) = \frac{c}{\log(R+t)}$$

satisfies

$$P(Y_t \in A_\delta) \leq D\sqrt{\mu_{\sigma(t)}(A_\delta)}, \quad \forall \delta > 0,$$

where $A_\delta = \{ x \in M \mid U(x) \geq U_0 + \delta \}$.

Remark

For a non-compact state space M, there are additional boundedness conditions, e.g., on $\left|U(x) - d(x, x_0)^2\right|$.

Set-up in $\mathbb{R}^3$ – 1

Since $\mathfrak{g}_2^2 \simeq \mathbb{R}^3$, hypo-elliptic simulated annealing in G_2^2 can be interpreted as hypo-elliptic simulated annealing in $\mathbb{R}^3$.

$$V_1(x) = \begin{pmatrix} 1 \\ 0 \\ -x_2 \end{pmatrix}, \quad V_2(x) = \begin{pmatrix} 0 \\ 1 \\ x_1 \end{pmatrix}.$$

This corresponds to the “sub-Riemannian gradient”

$$\Gamma(U,\cdot)(x) = \begin{pmatrix} \partial_{x_1} U(x) - x_2 \partial_{x_3} U(x) \\ \partial_{x_2} U(x) + x_1 \partial_{x_3} U(x) \\ x_1 \partial_{x_2} U(x) - x_2 \partial_{x_1} U(x) + (x_1^2 + x_2^2) \partial_{x_3} U(x). \end{pmatrix}$$

Set-up in $\mathbb{R}^3$ – 2

We test a variant of the Rastrigin potential

$$U(x) = 30 + \sum_{i=1}^{3} \left(x_i^2 - 10 \cos(2\pi x_i) \right).$$

Note that $\min U = 0$ attained at $y_{\min} = (0, 0, 0)$.

Remark

The Riemannian gradient ∇U grows linearly in x, whereas the sub-Riemannian gradient $\Gamma(U, \cdot)$ grows like $||x||^3$, which requires a much finer time-resolution for the approximation of the SDE.

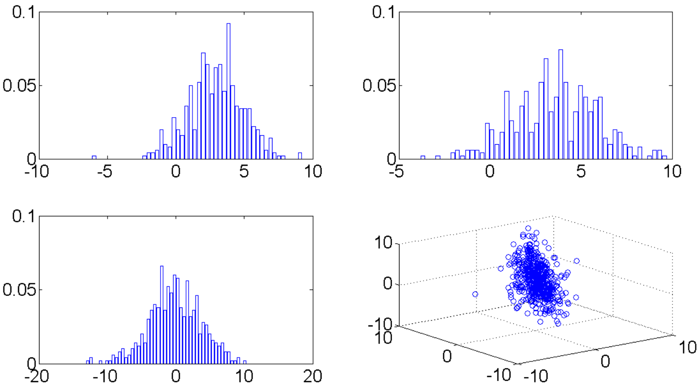
Results – table

Elliptic simulated annealing			
t	$E(\|Y_t\|)$	$E(U(Y_t))$	$\min_\omega U(Y_t(\omega))$
1	5.43	55.25	4.45
256	1.81	8.84	1.14
2048	1.62	6.77	0.12

Hypo-elliptic simulated annealing			
t	$E(\|Y_t\|)$	$E(U(Y_t))$	$\min_\omega U(Y_t(\omega))$
1	5.92	68.85	6.89
256	1.88	10.19	0.02
2048	1.58	6.50	0.32

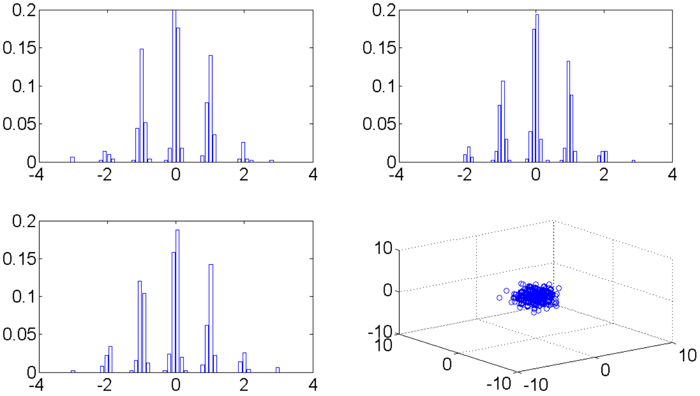
- $Y_0 = (5, 5, -5)$, $\|Y_0\| = 8.66$, $U(Y_0) = 50.25$, $y_{\min} = (0, 0, 0)$
- $c = 15$
- "E" denotes an average over 500 sampled paths

Results – histogram for $t = 1$



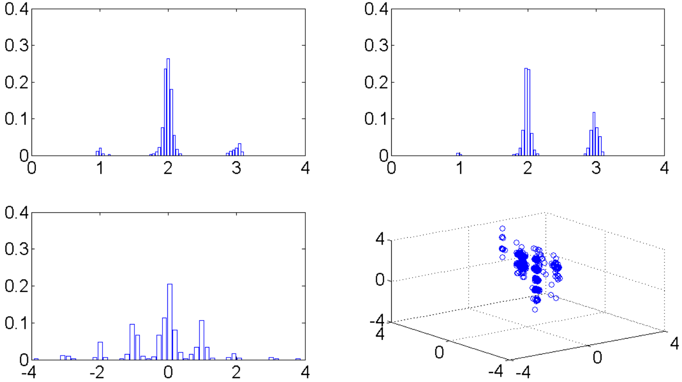
Starting value: $Y_0 = (5, 5, -5)$, $c = 15$, $y_{\min} = (0, 0, 0)$.

Results – histogram for $t = 4000$



Starting value: $Y_0 = (5, 5, -5)$, $c = 15$, $y_{\min} = (0, 0, 0)$.

Results – histogram for $t = 4000$ for too fast cooling



Starting value: $Y_0 = (5, 5, -5)$, $c = 3$, $y_{\min} = (0, 0, 0)$.

Set-up

- Represent $SO(3) \simeq S^3$
- Choose to vector fields from $\mathfrak{so}(3)$

$$V_1 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

- Potential

$$U(x) = 4.5 + (x_1 + 1)^4 - 2\cos(2\pi(x_1 + 1)) + x_2^2 - \cos(\pi x_2) + \\ + x_3^2 - \frac{1}{2}\cos(2\pi x_3) + x_4^2 - \cos(\pi x_4).$$

Set up – 2

- Potential

$$U(x) = 4.5 + (x_1+1)^4 - 2\cos(2\pi(x_1+1)) + x_2^2 - \cos(\pi x_2) + \\ + x_3^2 - \frac{1}{2}\cos(2\pi x_3) + x_4^2 - \cos(\pi x_4).$$

- Use a geometrical approximation scheme for the solution of the SDE, i.e., $\overline{X}_n \in SO(3)$ for every n
- $\min_y U(y) = 0$ attained at $y_{\min} = (-1, 0, 0, 0)$
- We start at $y_0 = (1, 0, 0, 0)$.
- The scheme is simpler than any elliptic simulated annealing scheme.

Set up – 2

- Potential

$$U(x) = 4.5 + (x_1+1)^4 - 2\cos(2\pi(x_1+1)) + x_2^2 - \cos(\pi x_2) + \\ + x_3^2 - \frac{1}{2}\cos(2\pi x_3) + x_4^2 - \cos(\pi x_4).$$

- Use a geometrical approximation scheme for the solution of the SDE, i.e., $\overline{X}_n \in SO(3)$ for every n
- $\min_y U(y) = 0$ attained at $y_{\min} = (-1, 0, 0, 0)$
- We start at $y_0 = (1, 0, 0, 0)$.
- The scheme is simpler than any elliptic simulated annealing scheme.

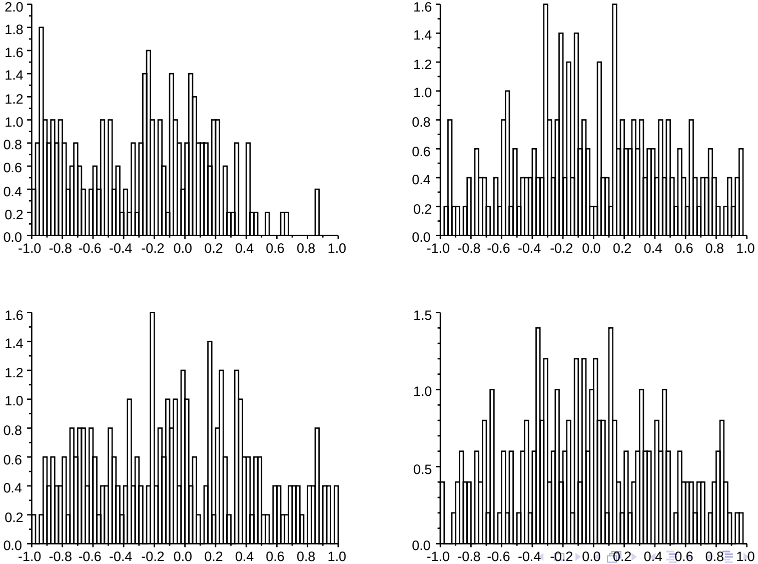
Numerical results

Hypo-elliptic simulated annealing, $c = 1.4$			
t	$E(\|Y_t - y_{\min}\|)$	$E(U(Y_t))$	$\min_\omega Y_t(\omega)$
2	1.0388	3.5504	0.0023
30	0.6925	1.4253	0.0138
2046	0.6083	1.0135	0.0101
65564	0.5871	0.8998	0.0079

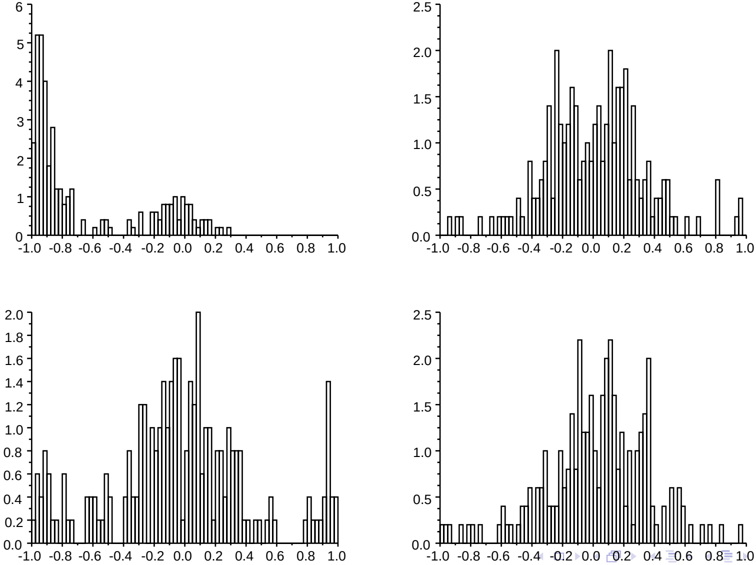
Hypo-elliptic simulated annealing, $c = 5$			
t	$E(\|Y_t - y_{\min}\|)$	$E(U(Y_t))$	$\min_\omega Y_t(\omega)$
2	1.1191	5.2407	0.1885
30	0.8964	3.2517	0.0082
2046	0.3868	1.0925	0.0075
65564	0.2897	0.6725	0.0596

- $Y_0 = (1, 0, 0, 0)$, $y_{\min} = (-1, 0, 0, 0)$, $\|Y_0 - y_{\min}\| = 2$, $U(Y_0) = 16$

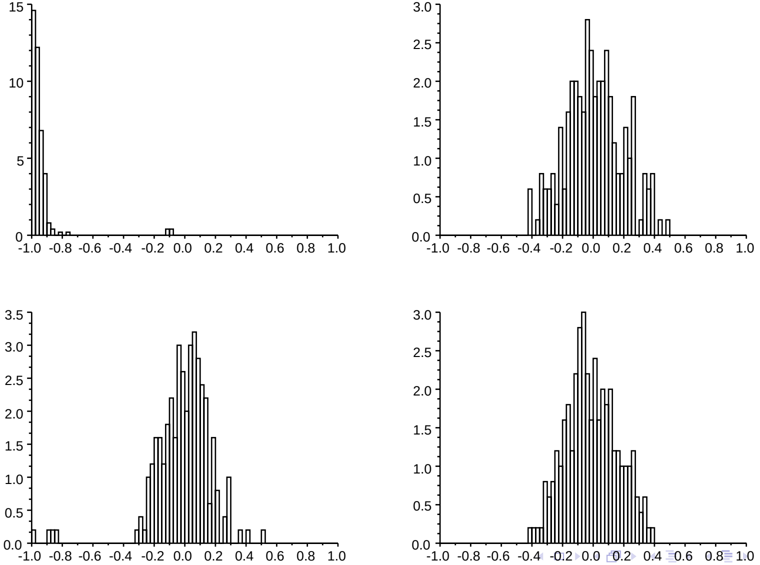
Histogram – $t = 2, c = 5$



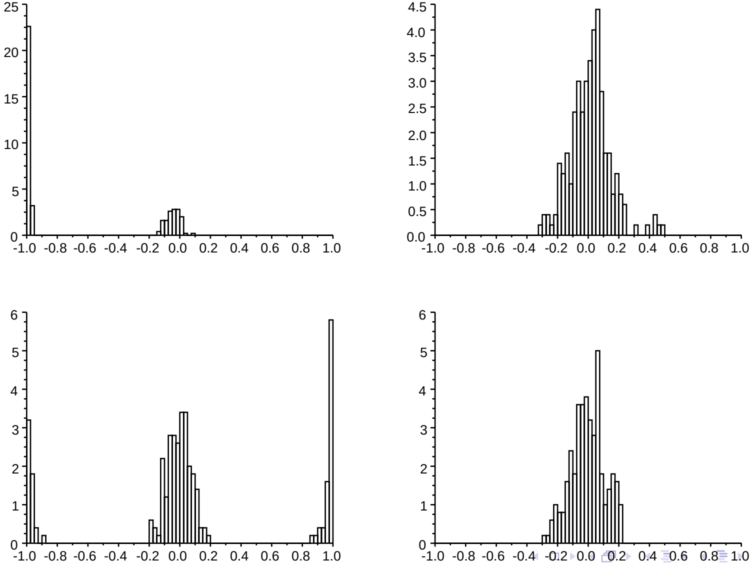
Histogram – $t = 62$, $c = 5$



Histogram – $t = 65534$, $c = 5$



Histogram – $t = 131070$, $c = 1.4$



Conclusions

- Construction of hypo-elliptic Smoluchowski (or Ornstein-Uhlenbeck) processes with Gibbs measure as invariant measure
- Simulated annealing for hypo-elliptic Smoluchowski processes – both theoretical and experimental justification
- Additional numerical cost due to instability in genuinely “elliptic” situations (e.g., when vector fields need to be non-linear for hypo-elliptic simulated annealing)
- Competitive in certain situations of *constraint optimization*

References

- F. Baudoin, M. Hairer, J. Teichmann. *Ornstein-Uhlenbeck processes on Lie groups*, J. Func. Anal., 255(4), 2008.
- B. Driver, T. Melcher. *Hypoelliptic heat kernel inequalities on the Heisenberg group*, J. Func. Anal., 221(2), 2005.
- R. Holley, S. Kusuoka, D. Stroock. *Asymptotics of the spectral gap with applications to the theory of simulated annealing*, J. Func. Anal., 83(2), 1989.
- S. Kirkpatrick, C. D. Gelett, M. P. Vecchi. *Optimization by simulated annealing*, Science 220, May 1983.
- L. Miclo. *Recuit simulé sur* $\mathbf{R}^n$*. Etude del l'évolution de l'énergie libre.*, Ann. Inst. H. Poincaré Prob. Stat., 28(2), 1992.