Convergence of the kinetic annealing for general potentials

Lucas Journel
Laboratoire Jacques-Louis Lions
Joint work with Pierre Monmarché

Introduction to simulated annealing

$$U : \mathbb{R}^d \longrightarrow \mathbb{R}_+.$$

Goal :

$$\min_{\mathbb{R}^d} U.$$

Design stochastic process (X_t) whose law is close to :

$$\pi_{\beta_t}(\mathrm{d}x) \propto e^{-\beta_t U(x)}\mathrm{d}x$$

where $\beta_t \underset{t\to\infty}{\rightarrow} \infty$.

$$\pi_{\beta_t}(U(x) > \delta) \underset{t\to\infty}{\rightarrow} 0.$$

Examples of process used in simulated annealing

- Overdamped Langevin Process (OLP) :

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2\beta_t^{-1}}\mathrm{d}B_t.$$

- Kinetic Langevin Process (KLP) :

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t \\ \mathrm{d}Y_t = -\nabla U(X_t)\mathrm{d}t - \gamma_t Y_t \mathrm{d}t + \sqrt{2\gamma_t\beta_t^{-1}}\mathrm{d}B_t. \end{cases}$$

- Local equilibrium of KLP :

$$\mu_{\beta_t}(\mathrm{d}x\mathrm{d}y) \propto e^{-\beta_t H(x,y)}\mathrm{d}x\mathrm{d}y$$

where $H(x,y) = U(x) + |y|^2/2$.

Cooling schedule and energy barriere

Cooling schedule :

$$\beta_t = \frac{\ln(e^{c\beta_0} + t)}{c}.$$

Cooling schedule and energy barriere

Cooling schedule :

$$\beta_t = \frac{\ln(e^{c\beta_0} + t)}{c}.$$

Largest energy barriere :

$$c^* = \sup_{x_1, x_2} E(x_1, x_2),$$

where

$$E(x_1, x_2) = \inf_{\xi} \left\{ \max_{0 \leqslant t \leqslant 1} U(\xi(t)) - U(x_1) - U(x_2) \right\}.$$

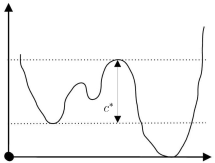
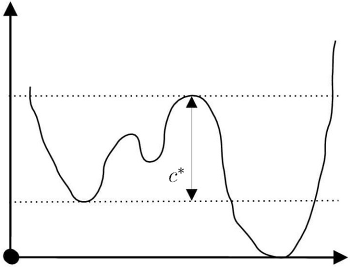


Figure – Example of potential

Historical review

Theorem (Holley, Kusuoka, Stroock 89)

M smooth compact manifold and $U : M \longrightarrow \mathbb{R}_+$.
(X_t) OLP with cooling schedule (β_t).
Then if $c > c^*$, $U(X_t) \to \min U$ in probability.
If there exists $p \in M$, bottom of a well of height greater than c, then $\mathbb{P}(\inf U(X_t) > U(p)) > 0$.



Historical review

- $U : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, $U(\infty) = \infty$, $|\nabla U|(\infty) = \infty$,
- $\inf_{\mathbb{R}^d} |\nabla U|^2 - \Delta U > -\infty$.

(X_t) Overamped Langevin Process.

Theorem (Chiang, Hwang, Sheu 87)

If $c > 3/2c^*$, $U(X_t) \to \min U$ in probability.

Theorem (Royer 89, Miclot 92)

If $c > c^*$, $U(X_t) \to \min U$ in probability.

Historical review

(X_t) Overamped Langevin Process.

Theorem (Zitt 08)

$U : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, $U(x) \geqslant \ln(|x|)^m - C$, $\|\nabla U\|_\infty < \infty$,
$\Delta U \leqslant 0$ *outside a compact.*
Then if $c > c^*$, $U(X_t) \to \min U$ *in probability.*

Theorem (Fournier, Tardif 21)

$U : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, $U(\infty) = \infty$, $\int_{\mathbb{R}^d} e^{-\alpha_0 U} < \infty$
Then if $c > c^*$, $U(X_t) \to \min U$ *in probability.*

Historical review and result

(X_t, Y_t) Kinetic Langevin Process. $H(x, y) = U(x) + |y|^2/2$.

Theorem (Monmarché 18)

$U : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, $x.\nabla U(x) \geqslant r|x|^2 - M$, $\|\nabla^2 U\|_\infty < \infty$.
Then if $c > c^*$, $U(X_t) \to \min U$ *in probability.*

Theorem (J, Monmarché 21)

$U : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, $U(\infty) = \infty$, $\int_{\mathbb{R}^d} e^{-\alpha_0 U} < \infty$.
Then if $c > c^*$, $H(X_t, Y_t) \to \min U$ *in probability.*
If there exists p*, bottom of a well of height greater than* c^**, then* $\mathbb{P}(\inf U(X_t) > U(p)) > 0$.

A basic example

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t,$$

U quadratic at infinity.

- π_β stationary measure.
- $f_t = \mathcal{L}aw(X_t)$, $h_t = \frac{\mathrm{d}f_t}{\mathrm{d}\pi_\beta}$.

$$\partial_t h_t = \beta^{-1}\Delta h_t - \nabla U.\nabla h_t =: L^* h_t.$$

A basic example

Definition (Carré du champ)

$$\Gamma f = \frac{1}{2}\left(L^* f^2 - 2fL^* f\right).$$

$\Gamma f = \sqrt{2\beta^{-1}}|\nabla f|^2$.

$$H_t = \int_{\mathbb{R}^d} (h_t - 1)^2 \mathrm{d}\pi_\beta \geqslant \|f_t - \pi_\beta\|_{TV}^2.$$

$$H_t' = -\int_{\mathbb{R}^d} \Gamma(h_t) \mathrm{d}\pi_\beta.$$

A more basic example

Poincaré inequality

For all $h : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} h \mathrm{d}\pi_\beta = 1$:

$$\int_{\mathbb{R}^d} (h-1)^2 \mathrm{d}\pi_\beta \leqslant \lambda_\beta \int_{\mathbb{R}^d} \Gamma(h) \mathrm{d}\pi_\beta.$$

λ_β satisfies :

$$\lim_{\beta \to \infty} \beta^{-1} \ln(\lambda_\beta) = c^*.$$

$$H_t = \int_{\mathbb{R}^d} (h_t - 1)^2 \mathrm{d}\pi_\beta.$$

$$H_t' = -\int_{\mathbb{R}^d} \Gamma(h_t) \mathrm{d}\pi_\beta \leqslant -\lambda_\beta^{-1} H_t.$$

$$H_t \leqslant e^{-\lambda_\beta^{-1} t} H_0.$$

Problems

- $\beta = \beta_t$.

$$H_t' = -\int_{\mathbb{R}^d} \Gamma(h_t) \mathrm{d}\pi_{\beta_t} + \beta_t' A_t.$$

Problems

- $\beta = \beta_t$.

$$H_t' = -\int_{\mathbb{R}^d} \Gamma(h_t) \mathrm{d}\pi_{\beta_t} + \beta_t' A_t.$$

- U is such that π_β does not satisfy Poincarré inequality.

Problems

- $\beta = \beta_t$.

$$H_t' = -\int_{\mathbb{R}^d} \Gamma(h_t) \mathrm{d}\pi_{\beta_t} + \beta_t' A_t.$$

- U is such that π_β does not satisfy Poincarré inequality.
- For kinetic Langevin process :

$$\Gamma(f) = |\nabla_y f|^2,$$

hence no Poincarré inequality of the form

$$\int_{\mathbb{R}^d} (h-1)^2 \mathrm{d}\mu_\beta \leqslant \lambda_\beta \int_{\mathbb{R}^d} \Gamma(h) \mathrm{d}\mu_\beta.$$

Notation

Generator of the process :

$$L_t = y.\nabla_x - \nabla U.\nabla_y - \gamma_t y.\nabla_y + \gamma_t \beta_t^{-1} \Delta_y$$

Let f_t the law of the process and $h_t = \frac{f_t}{\mu_{\beta_t}}$.
L_t^* the dual of L_t in $L^2\left(\mu_{\beta_t}\right)$:

$$\partial_t h_t = L_t^* h_t - \beta_t' H h_t.$$

Plan of the proof

1 Almost surely, $\sup_t H(X_t, Y_t) < \infty$.

   a Almost surely, $\liminf_{t\to\infty} H(X_t, Y_t)$.

   b For compact set C, there exists $K > 0$ such that

$$\inf_{x\in C} \mathbb{P}(\sup_t H(X_t, Y_t) \leqslant K) \geqslant 1/4.$$

2 If X lives in a compact set, there is convergence.

Return to a compact set

Proposition

Almost surely, $\liminf_{t\to\infty} H(X_t, Y_t) < \infty$.

Lemma

For any $\mathcal{C}^\infty$-probability density f_0 with compact support in $\mathbb{R}^{2d}$, we have

$$\sup_{t\geqslant 0} \mathbb{E}_{f_0} \left(H(X_t, Y_t)\right) \leqslant \frac{\kappa^{\beta_0}(f_0) + \ln(\mathcal{Z}_{\alpha_0})}{\beta_0 - \alpha_0},$$

where

$$\kappa^{\beta_0}(f_0) = \int_{\mathbb{R}^{2d}} f_0 \ln\left(1 + f_0 e^{\beta_0 H}\right), \qquad \mathcal{Z}_{\alpha_0} = \int_{\mathbb{R}^{2d}} e^{-\alpha_0 H}.$$

Proof of Lemma

Writing $g_t = f_t e^{\beta_t H}$, "entropy" :

$$N(t) = \int_{\mathbb{R}^{2d}} g_t \ln(1 + g_t) \, e^{-\beta_t H}$$

Formally, using Langevin equation :

$$N'(t) = -\gamma_t \beta_t^{-1} \int_{\mathbb{R}^{2d}} |\nabla_y g_t|^2 \left(\frac{1}{1+g_t} + \frac{1}{(1+g_t)^2} \right) + \int_{\mathbb{R}^{2d}} \beta_t' \frac{g_t}{1+g_t} H f_t .$$

Then

$$N'(t) \leqslant \beta_t' \mathbb{E}_{f_0} H(X_t, Y_t)$$

Proof of Lemma

$$N'(t) \leqslant \beta_t' \mathbb{E}_{f_0} H(X_t, Y_t)$$

$$\int_{\mathbb{R}^{2d}} f_t \ln f_t = \max \left\{ \int_{\mathbb{R}^{2d}} f_t \ln g; \; g : \mathbb{R}^{2d} \longrightarrow \mathbb{R}_+, \; \int_{\mathbb{R}^{2d}} g = 1 \right\} .$$

With $g_0 = e^{-\alpha_0 H} / \mathcal{Z}_{\alpha_0}$, where $\mathcal{Z}_{\alpha_0} = \int e^{-\alpha_0 H}$:

$$N(t) \geqslant (\beta_t - \alpha_0) \mathbb{E}_{f_0}(H(X_t, Y_t)) - \ln(\mathcal{Z}_{\alpha_0}).$$

We conclude with Gronwall lemma.

Localisation

Proposition

Fix some $A > 1$. There exist $b_A > 1$, $K_A > A$ which depends on A, U, and c, but not β_0 such that, for all $\beta_0 \geqslant b_A$ and all initial condition $z_0 \in \{H \leqslant A\}$,

$$\mathbb{P}_{z_0}\left(\sup_{t\geqslant 0} H\left(X_t, Y_t\right) \leqslant K_A\right) \geqslant \frac{1}{4}.$$

$K_A \approx 4c + A$

Localisation

Fix $K > 1$, $L_K > 1$, $M_K = (\mathbb{R}/2L_K\mathbb{Z})^d$, such that $\{U \leqslant K\} \subset M_K$.
$U^K : M_K \longrightarrow \mathbb{R}$, $U^K = U$ on $\{U \leqslant K\}$

$$\begin{cases} \mathrm{d}X_t^K = Y_t^K \mathrm{d}t \\ \mathrm{d}Y_t^K = -\nabla_x U^K(X_t)\mathrm{d}t - \gamma_t Y_t^K \mathrm{d}t + \sqrt{2\gamma_t \beta_t^{-1}}\mathrm{d}B_t \,. \end{cases} \tag{1}$$

$$\left\{ \sup_{t \geqslant 0} H(X_t, Y_t) \leqslant K \right\} = \left\{ \sup_{t \geqslant 0} H_K(X_t^K, Y_t^K) \leqslant K \right\},$$

where $H_K(x,y) = U^K(x) + |y|^2/2$.

$$\mu_\beta^K(\mathrm{d}x\mathrm{d}y) \propto e^{-\beta H_K(x,y)}\mathrm{d}x\mathrm{d}y.$$

Localisation

Lemma

If $c > c^*$, fix $A > 1$. Set $D_A = A + 3 + 4c$ and $K_A = D_A + 1$.
There exist $C_A > 0$ and b_A that do not depend on β_0 such that, for all $\beta_0 \geqslant b_A$ and $\mathcal{C}^\infty$-probability density f_0 with support in $\{H_K \leqslant A + 1\}$, we have that, for all $t \geqslant 0$

$$\mathbb{P}_{f_0}\left(H\left(X_t^{K_A}, Y_t^{K_A}\right) \geqslant D_A\right) \leqslant \frac{C_A}{(e^{c\beta_0} + t)^2}.$$

$$\mathbb{P}(H(X_t^K, Y_t^K) > D) \leqslant \|h_t^K\|_{L^2} \left(\mu_{\beta_t}(H_K > D)\right)^{1/2}$$

$$\mu_{\beta_t}(H_K > D) \leqslant Ce^{-\beta_t(D-1)}$$

Localisation

Hypocoercivity *à la Villani* :

$$\phi_t(h) = |(\nabla_x + \nabla_y)h|^2 + \sigma_t h^2$$

with $\sigma_t = \frac{1}{2} + 2\sqrt{\gamma_t^{-1}\beta_t}(1 + \|\nabla U^K\|_\infty + \gamma_t)^2$, we introduce

$$\tilde{N}(t) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi_t\left(h_t^K - 1\right) \mathrm{d}\mu_{\beta_t}^K$$

$$\tilde{I}(t) = \int_{M_K \times \mathbb{R}^d} \left|\nabla h_t^K\right|^2 \mathrm{d}\mu_{\beta_t}^K .$$

Differentiating $\tilde{N}$, one can (formally) check that

$$\tilde{N}'(t) \leqslant -\frac{1}{2}\tilde{I}(t) + C\beta_t'(1 + \beta_t)\tilde{N}(t)$$

Localisation

Poincaré inequality :

$$\tilde{N}(t) \leqslant \lambda_{\beta_t} \tilde{I}(t)$$

With :

$$\frac{1}{\beta} \ln(\lambda_\beta) \underset{\beta \to \infty}{\to} c^*$$

Conclusion :

$$\tilde{N}'(t) \leqslant \left(-\frac{C'}{(1+t)^{c^*/c}} + \frac{C(1+\ln(1+t))}{(1+t)} \right) \tilde{N}(t)$$

Position in a compact set

Proposition

If $c > c^*$, then for all $K > 1$, all $\mathcal{C}^\infty$-probability density f_0 with compact support in $M_K \times \mathbb{R}^d$, and all $\delta > 0$,

$$\mathbb{P}_{f_0}\left(H_K(X_t^K, Y_t^K) > \delta\right) \underset{t\to+\infty}{\longrightarrow} 0.$$

$$\mathbb{P}\left(H_K(X_t^K, Y_t^K) > \delta\right) \leqslant \|h_t^K\|_{L^2}\left(\mu_{\beta_t}(H_K > \delta)\right)^{1/2}$$

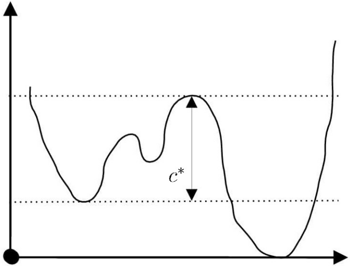
Non-convergence for fast cooling schedule

What we showed :

$$\mathbb{P}\left(\sup_{t\geqslant 0} H(X_t, Y_t) \leqslant H(x_0, y_0) + 4c\right) > 0$$

What we want :

$$\mathbb{P}\left(\sup_{t\geqslant 0} H(X_t, Y_t) \leqslant H(x_0, y_0) + c + \delta\right) > 0$$



Non-convergence for fast cooling schedule

$$\|h_t\|_{L^2} \to \|h_t\|_{L^\infty}$$

$$\mathbb{P}\left(H(X_t, Y_t) > c + \delta\right) \leqslant \|h_t\|_\infty \mu_{\beta_t}(H > c + \delta) = \frac{C}{(1+t)^{1+\delta/2c}}$$

H^k-hypocoercivity. C.Zang 20

$$\begin{aligned}\|h-1\|_{L^\infty} &\leqslant C\|h-1\|_{H^m\left(\mu^K_{\beta_0}\right)} \\ &\leqslant Ce^{\beta_t\|H_K\|_\infty}\|h-1\|_{H^m\left(\mu^K_{\beta_t}\right)}.\end{aligned}$$

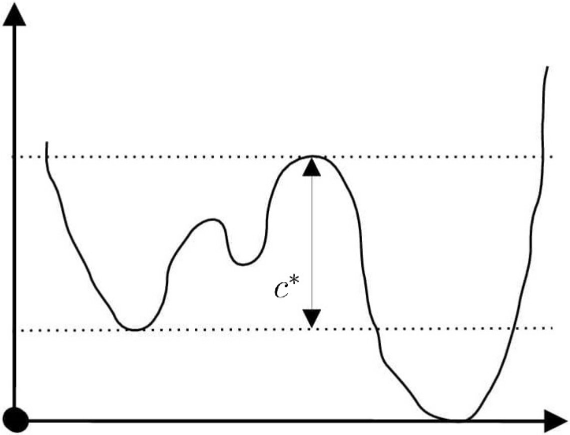


Figure – Example of potential

Non-convergence for fast cooling schedule

$$\begin{cases} \mathrm{d}X_t^K = \nabla W(Y_t^K)\mathrm{d}t - \sigma(Y_t)\nabla U^K(X_t^K)\mathrm{d}t + \sqrt{2\sigma(Y_t^K)\beta_t^{-1}}\mathrm{d}\tilde{B}_t \\ \mathrm{d}Y_t^K = -\nabla U^K(X_t^K)\mathrm{d}t - \gamma_t \nabla W(Y_t^K)\mathrm{d}t + \sqrt{2\gamma_t\beta_t^{-1}}\mathrm{d}B_t \end{cases}$$

Fix $K > 1$, let $W^K, \sigma : M_K \longrightarrow \mathbb{R}_+$, be such that :

- $H_K = U^K + W^K = H - H(x)$ on $\{H - H(x) \leqslant K\}$
- $\left\{\nabla^2 W^K \neq I_d\right\} \subset \{\sigma > 0\}$
- $\{H_K \leqslant K\} \subset \{\sigma = 0\}$
- $c^*(H_K) < c$

THANK YOU FOR YOUR ATTENTION