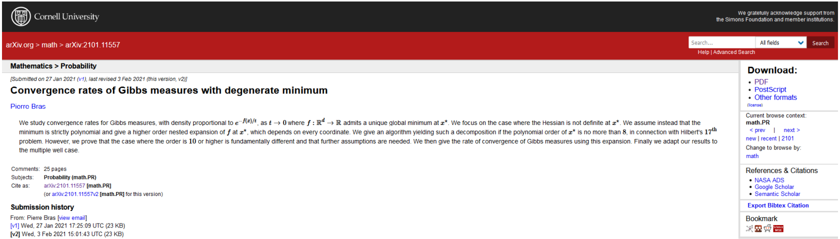
Convergence rates of Gibbs measures with degenerate minimum

Pierre Bras (PhD Student under the direction of Gilles Pagès)

Sorbonne Université - LPSM

10 June 2021



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org > math > arXiv:2101.11557

Search... All fields Search

Help | Advanced Search

Mathematics > Probability

[Submitted on 27 Jan 2021 (v1), last revised 3 Feb 2021 (this version, v2)]

Convergence rates of Gibbs measures with degenerate minimum

Pierre Bras

We study convergence rates for Gibbs measures, with density proportional to $e^{-f(x)/t}$, as $t \to 0$ where $f : \mathbb{R}^d \to \mathbb{R}$ admits a unique global minimum at $x^\star$. We focus on the case where the Hessian is not definite at $x^\star$. We assume instead that the minimum is strictly polynomial and give a higher order nested expansion of f at $x^\star$, which depends on every coordinate. We give an algorithm yielding such a decomposition if the polynomial order of $x^\star$ is no more than 8, in connection with Hilbert's 17^{th} problem. However, we prove that the case where the order is 10 or higher is fundamentally different and that further assumptions are needed. We then give the rate of convergence of Gibbs measures using this expansion. Finally we adapt our results to the multiple well case.

Comments: 25 pages
Subjects: **Probability (math.PR)**
Cite as: arXiv:2101.11557 **[math.PR]**
(or arXiv:2101.11557v2 **[math.PR]** for this version)

Submission history

From: Pierre Bras [view email]
[v1] Wed, 27 Jan 2021 17:25:09 UTC (23 KB)
[v2] Wed, 3 Feb 2021 15:01:43 UTC (23 KB)

Download:
- PDF
- PostScript
- Other formats

(license)

Current browse context:
math.PR
< prev | next >
new | recent | 2101
Change to browse by:
math

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

Export Bibtex Citation

Bookmark

SORBONNE UNIVERSITÉ

Optimization problem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be $\mathcal{C}^1$, coercive (i.e. $f(x) \to +\infty$ as $|x| \to \infty$) and let $\mathrm{argmin}(f) := \{x \in \mathbb{R}^d : \ f(x) = \min_{\mathbb{R}^d} f\}$.

Objective : find $\mathrm{argmin}(f)$.

- **Example : Regression as an optimization problem**

– $\{\Phi_x : \ x \in \mathbb{R}^d\}$ family of functions $\Phi_x : \mathbb{R}^{d'} \to \mathbb{R}$ parametrized by $x \in \mathbb{R}^d$ (e.g. Φ_x is a neural function).
– for $1 \le i \le N$, $(u_i, v_i) \in \mathbb{R}^{d'} \times \mathbb{R}$: data associated to a regression problem
– We want to find x such that for all i, $\Phi_x(u_i) \approx v_i$

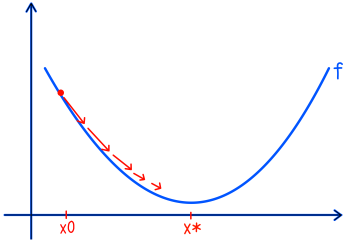
$$\implies \text{Find } \min_{x\in\mathbb{R}^d} \frac{1}{N}\sum_{i=1}^{N}(\Phi_x(u_i) - v_i)^2 =: \min_{x\in\mathbb{R}^d} f(x).$$

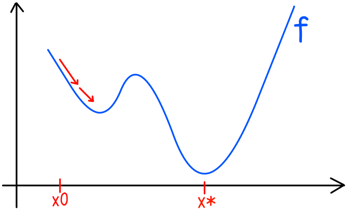
Introduction - Gradient descent

• Gradient descent algorithm : compute the gradient and "go down" the gradient with step $\gamma > 0$:

$$x_0 \in \mathbb{R}^d$$
$$x_{n+1} = x_n - \gamma \nabla f(x_n).$$

• The continuous version is $Y_s = -\nabla f(Y_s)ds$.





• **Problem** : x_n can be "trapped" !

- We add a white noise to x_n, hoping to escape traps :

$$x_{n+1} = x_n - \gamma \nabla f(x_n) + \sqrt{\gamma}\sigma\xi_{n+1}, \quad \xi_{n+1} \sim \mathcal{N}(0, I_d).$$

- The continuous version becomes :

$$\begin{aligned} dY_s &= -\nabla f(Y_s)ds + \sigma dW_s \qquad \textbf{(Langevin Equation)} \\ &= -\nabla f(Y_s)ds + \sqrt{2t}dW_s, \end{aligned}$$

where (W_s) is a Brownian motion.
(⚠ t is not the time but a parameter, with $t = \sigma^2/2$).
- Assuming that $e^{-f/t} \in L^1(\mathbb{R}^d)$, it is invariant measure is the **Gibbs measure**

$$\begin{aligned} \pi_t(x)dx &= C_t e^{-f(x)/t} dx \\ C_t &= \left(\int_{\mathbb{R}^d} e^{-f(x)/t} dx \right)^{-1}. \end{aligned}$$

• For small t, the Gibbs measure π_t is concentrated around $\text{argmin}(f)$. Assuming that $f^\star = 0$:

$$\forall \varepsilon > 0,\ \pi_t\{f \geq \varepsilon\} \underset{t\to 0}{\longrightarrow} 0.$$

Indeed, we have

$$C_t \leq \left(\int_{f \leq \varepsilon/3} e^{-\frac{f(x)}{t}} dx \right)^{-1} \leq \left(e^{-\frac{\varepsilon}{3t}} \int_{f \leq \varepsilon/3} dx \right)^{-1}.$$

If $f(x) \geq \varepsilon$ then $e^{-\frac{f(x)}{t}} \leq e^{-\frac{2\varepsilon}{3t}} e^{-\frac{f(x)}{3t}}$ and then

$$\pi_t\{f \geq \varepsilon\} = C_t \int_{f \geq \varepsilon} e^{-f(x)/t} dx \leq e^{\frac{\varepsilon}{3t}} \left(\int_{f \leq \varepsilon/3} dx \right)^{-1} e^{-\frac{2\varepsilon}{3t}} \int_{\mathbb{R}^d} e^{-\frac{f(x)}{3t}} dx$$

$$\leq \left(\int_{f \leq \varepsilon/3} dx \right)^{-1} \left(\int_{\mathbb{R}^d} e^{-f(x)} dx \right) e^{-\frac{\varepsilon}{3t}} \underset{t\to 0}{\longrightarrow} 0.$$

• Solve the Langevin Equation for small $t > 0$, which gives an approximation of $\text{argmin}(f)$.
$\Longrightarrow$ **What is the quality of this approximation ?**

• Another method is to make t slowly decrease to 0 inside the Langevin Equation $dY_s = -\nabla f(Y_s)ds + \sqrt{2t(s)}dW_s$ with $t(s) \to 0$ as $s \to \infty$ ([Chiang-Hwang-Sheu 1987], [Gelfand-Mitter 1990]).

- **Simplest case** : $\operatorname{argmin}(f) = \{x^\star\}$, $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) > 0$. Then :

$$\pi_t \xrightarrow[t\to 0]{\mathscr{L}} \delta_{x^\star} \quad \text{(Dirac measure)}$$

$$\pi_t(x)dx = C_t e^{-f(x)/t}dx \simeq C_t e^{-\frac{1}{2t}x^\top \cdot \nabla^2 f(x^\star)\cdot x}dx$$

$$\frac{1}{\sqrt{t}}(X_t - x^\star) \xrightarrow[t\to 0]{} X \sim \mathcal{N}(0, (\nabla^2 f(x^\star))^{-1})$$

where $X_t \sim \pi_t$ converges to $x^\star$ at speed $\sqrt{t}$.

- **Multiple well case** (Hwang 1980) : $\operatorname{argmin}(f) = \{x_1^\star, \ldots, x_m^\star\}$ and for all i, $\nabla^2 f(x_i^\star) > 0$. Then

$$\pi_t \longrightarrow \frac{1}{\sum_{j=1}^m \det^{-1/2}(\nabla^2 f(x_j^\star))} \sum_{i=1}^m \det^{-1/2}(\nabla^2 f(x_i^\star))\delta_{x_i^\star}$$

$$\frac{1}{\sqrt{t}}(X_{it} - x_i^\star) \xrightarrow[t\to 0]{} X_i \sim \mathcal{N}(0, (\nabla^2 f(x_i^\star))^{-1}),$$

where X_{it} has the law of X_t conditionally to $||X_t - x_i^\star|| < r$.

- **Degenerate case** :

Theorem (Athreya-Hwang, 2010)

Assume that $\min(f) = 0$, $\operatorname{argmin}(f) = \{0\}$, *and that there exist* $\alpha_1, \ldots, \alpha_d > 0$ *and* $g : \mathbb{R}^d \to \mathbb{R}$ *such that*

$$\forall (h_1, \ldots, h_d) \in \mathbb{R}^d, \ \frac{1}{t} f(t^{\alpha_1} h_1, \ldots, t^{\alpha_d} h_d) \underset{t \to 0}{\longrightarrow} g(h_1, \ldots, h_d) \in \mathbb{R}.$$

$$\int_{\mathbb{R}^d} \sup_{0<t<1} e^{-\frac{f(t^{\alpha_1} h_1, \ldots, t^{\alpha_d} h_d)}{t}} dh_1 \ldots dh_d < \infty.$$

Then

$$\left(\frac{(X_t)_1}{t^{\alpha_1}}, \ldots, \frac{(X_t)_d}{t^{\alpha_d}} \right) \xrightarrow{\mathscr{L}} X \quad \text{as } t \to 0 \tag{1}$$

where $X_t \sim \pi_t$ *and where the distribution of* X *has a density proportional to* $e^{-g(x_1, \ldots, x_d)}$.

- **Question** : How can we find such $\alpha_1, \ldots, \alpha_d$ and g ?
- In the case where $\nabla^2 f(x^\star) > 0$ then let $B \in \mathcal{O}_d(\mathbb{R})$ such that $\nabla^2 f(x^\star) = B \mathrm{Diag}(\beta_{1:d}) B^\top$ and take $\alpha_1, \ldots, \alpha_d = 1/2$ so that

$$\frac{1}{t}\left(f(x^\star + t^{1/2} B \cdot h) - f(x^\star)\right) \underset{t \to 0}{\longrightarrow} \frac{1}{2} \sum_{i=1}^d \beta_i h_i^2 := g(h)$$

Derivatives tensors

- We focus on the case where $\text{argmin}(f) = \{x^\star\}$ and where $\nabla^2 f(x^\star)$ is not definite.
- Multi-dimensional Taylor-Young formula :

$$f(x+h) \underset{h\to 0}{=} \sum_{k=0}^{p} \frac{1}{k!} \nabla^k f(x) \cdot h^{\otimes k} + ||h||^p o(1),$$

where

$$\nabla^k f(x) = \left(\partial^k_{i_1,\dots,i_k} f(x)\right)_{i_1,i_2,\dots,i_k \in \{1,\dots,d\}}$$

and

$$h^{\otimes k} = h \otimes \cdots \otimes h = (h_{i_1} \dots h_{i_k})_{i_1,i_2,\dots,i_k \in \{1,\dots,d\}}$$

are tensors of order k, and where for T a k-tensor and $v^1, \dots, v^k$ k vectors in $\mathbb{R}^d$,

$$T \cdot v^1 \otimes \cdots \otimes v^k = \sum_{i_1,\dots,i_k \in \{1,\dots,d\}} T_{i_1,\dots,i_k} v^1_{i_1} \dots v^k_{i_k}.$$

- By Schwarz's Theorem, the tensor $\nabla^k f(x)$ is symmetric.
- Multidimensional Newton Formula :

$$(h_1 + h_2 + \cdots + h_p)^{\otimes k} = \sum_{\substack{i_1,\dots,i_p \in \{0,\dots,k\} \\ i_1+\cdots+i_p = k}} \binom{k}{i_1,\dots,i_p} h_1^{\otimes i_1} \otimes \cdots \otimes h_p^{\otimes i_p},$$

where $\binom{k}{i_1,\dots,i_p} = \frac{k!}{i_1!\cdots i_p!}$ is the p-nomial coefficient.

Polynomial minimum

- We assume instead that $x^\star$ is a strictly polynomial minimum of order $2p > 2$ i.e. in a neighbourhood of $x^\star$,

$$\sum_{k=1}^{2p} \frac{1}{k!} \nabla^k f(x^\star) \cdot h^{\otimes k} > 0.$$

Objective

Find $\alpha_1, \ldots, \alpha_d > 0$ and $B \in \mathcal{O}_d(\mathbb{R})$ such that

$$\forall h \in \mathbb{R}^d, \quad \frac{1}{t}\left[f(x^\star + B \cdot (t^{\alpha_1} h_1, \ldots, t^{\alpha_d} h_d)) - f(x^\star)\right] \underset{t \to 0}{\longrightarrow} g(h_1, \ldots, h_d),$$

where $g : \mathbb{R}^d \to \mathbb{R}$ is **non constant in any of its coordinates**.

- **Example : one-dimensional case** : Let m be the polynomial order of $x^\star$, then

$$\frac{1}{t}(f(x^\star + t^{1/m} h) - f(x^\star)) \underset{t \to 0}{\longrightarrow} \frac{f^{(m)}(x^\star)}{m!} h^m.$$

Then $\alpha_1 = 1/m$ and $g(h) = \frac{f^{(m)}(x^\star)}{m!} h^m$.

Expansion of f at a local minimum with degenerate derivatives for $2p = 4$

- Notations : $p_F : \mathbb{R}^d \to F$ the orthogonal projection, $T_k := \nabla^k f(x^\star)$.
- For $2p = 4$:

$$F := \{h \in \mathbb{R}^d : \ T_2 \cdot h^{\otimes 2} = 0\} = \{h \in \mathbb{R}^d : \ T_2 \cdot h^{\otimes 1} = 0^{\otimes 1}\},$$

and $E = F^\perp$ so

$$\mathbb{R}^d = E \oplus F.$$

- Apply Taylor and Newton formulas up to order 4 to

$$\frac{1}{t}\left[f(x^\star + t^{1/2}p_E(h) + t^{1/4}p_F(h)) - f(x^\star)\right] \tag{2}$$

$$= \frac{1}{t}\sum_{k=2}^{4}\frac{1}{k!}T_k \cdot (t^{1/2}p_E(h) + t^{1/4}p_F(h))^{\otimes k} + o(1)$$

$$\underset{t\to 0}{=} \sum_{k=2}^{4}\frac{1}{k!}\sum_{\substack{i_1,i_2\in\{0,\dots,k\}\\ i_1+i_2=k}}\binom{k}{i_1,i_2}t^{\frac{i_1}{2}+\frac{i_2}{4}-1}T_k \cdot p_E(h)^{\otimes i_1} \otimes p_F(h)^{\otimes i_2} + o(1)$$

- If $\frac{i_1}{2} + \frac{i_2}{4} - 1 > 0$ then it converges to 0
- By definition of F, for all h', $T_2 \cdot p_F(h) \otimes h' = 0$.
- By the minimum condition :

$$\frac{1}{t}\left[f(x^\star + p_F(h)) - f(x^\star)\right] = \frac{1}{3!}T_3 \cdot p_F(h)^{\otimes 3} + o(h^3) \geq 0$$
$$\implies T_3 \cdot p_F(h)^{\otimes 3} = 0$$

- We obtain

$$(2) \longrightarrow \frac{1}{2}T_2 \cdot p_E(h)^{\otimes 2} + \frac{1}{4!}T_4 \cdot p_F(h)^{\otimes 4} + \frac{1}{2}T_3 \cdot p_E(h) \otimes p_F(h)^{\otimes 2}$$

- Since $x^\star$ is polynomial of order 4 then $T_4 > 0$ on F so the limit function is not constant in any of its coordinates.
- ⚠The odd cross term in T_3 can be non null, for example

$$f(x, y) = x^2 + y^4 + xy^2.$$

We would like to proceed by induction as before, for example considering

$$F_2 := \{h \in F : \; T_4 \cdot h^{\otimes 4} = 0\}.$$

T_4 is a symmetric tensor so we can write

$$\forall h \in F, T_4 \cdot h^{\otimes 4} = \sum_{i=1}^{q} \lambda_i \langle v^i, h \rangle^4.$$

And since

$$\forall h \in F, \; T_4 \cdot h^{\otimes 4} \geq 0,$$

we could think that the λ_i are positive, which would give a linear characterization of F_2. However this is not always the case :

Hilbert's 17th problem

Let P be a non-negative polynomial homogeneous of even degree. Find polynomials $P_1, \; \ldots, \; P_r$ such that $P = \sum_{i=1}^{r} P_i^2$.

$\implies$ This problem does not always have a solution.
$\implies$ F_2 is not always a subspace ! (not even a sub-manifold !)
• Counter example : $P(X, Y, Z) = (X - Y)^2(X - Z)^2$.

Expansion up to $2p = 8$

We now consider the subspaces

$$F_k := \{h \in F_{k-1} : \; \forall h' \in F_{k-1}, \; \nabla^{2k} f(x^\star) \cdot h \otimes h'^{\otimes 2k-1} = 0\},$$

and E_k the orthogonal complement of F_k in F_{k-1}.

<table>
<tr><td>E_1</td><td colspan="3">F_1</td></tr>
<tr><td rowspan="5">$T_2 \geq 0$</td><td colspan="3">$T_2 = 0$</td></tr>
<tr><td>E_2</td><td colspan="2">F_2</td></tr>
<tr><td rowspan="3">$T_4 \geq 0$</td><td colspan="2">$T_4 = 0$</td></tr>
<tr><td>E_3</td><td>F_3</td></tr>
<tr><td>$T_6 \geq 0$</td><td>$T_6 = 0$</td></tr>
</table>

Table – Illustration of the subspaces

$$\mathbb{R}^d = E_1 \oplus E_2 \oplus E_3 \oplus F_3 \quad \text{and } E_4 := F_3.$$

We expand

$$\frac{1}{t}\left[f(x^\star + t^{1/2} p_{E_1}(h) + t^{1/4} p_{E_2}(h) + t^{1/6} p_{E_3}(h) + t^{1/8} p_{F_3}(h)) - f(x^\star)\right] \tag{3}$$

$$= \sum_{k=2}^{8} \frac{1}{k!} \sum_{\substack{i_1,\ldots,i_4 \in \{0,\ldots,k\} \\ i_1+\cdots+i_4=k}} \binom{k}{i_1,\ldots,i_4} t^{\frac{i_1}{2}+\cdots+\frac{i_4}{8}-1} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes p_{E_2}(h)^{\otimes i_2}$$

$$\otimes p_{E_3}(h)^{\otimes i_3} \otimes p_{F_3}(h)^{\otimes i_4} + o(1).$$

- If $\frac{i_1}{2} + \cdots + \frac{i_4}{8} - 1 > 0$, then it converges to 0.
- We need to prove that

$$\frac{i_1}{2} + \cdots + \frac{i_4}{8} - 1 < 0 \implies T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \cdots \otimes p_{E_4}(h)^{\otimes i_4} = 0.$$

- Example : Prove that for all $h \in \mathbb{R}^d$ and $h' \in F_2 = E_3 \oplus E_4$, $T_3 \cdot p_{E_1}(h) \otimes (h')^{\otimes 2} = 0$. We prove

$$\forall h \in \mathbb{R}^d,\ T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} = 0.$$

Indeed, using the expansion for $2p = 4$ we have :

$$\frac{1}{t}\left[(f(x^\star + t^{1/2}p_{E_1}(h) + t^{1/6}p_{F_2}(h)) - f(x^\star)\right]$$
$$\underset{t\to 0}{\longrightarrow} \frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} \geq 0.$$

Replacing h by λh, $\lambda \in \mathbb{R}$, we have for all $\lambda \in \mathbb{R}$:

$$\frac{\lambda^2}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{\lambda^3}{2}T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} \geq 0,$$

so necessarily $\forall h \in \mathbb{R}^d,\ T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} = 0$.

We finally obtain :

(3) $\longrightarrow \frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h)^{\otimes 2} \otimes p_{E_2}(h) + \frac{1}{2}T_4 \cdot p_{E_1}(h) \otimes p_{E_2}(h) \otimes p_{E_4}(h)^{\otimes 2} + \ldots$

Order 2	$(2,0,0,0)$
Order 3	$(2,1,0,0)$
Order 4	$(0,4,0,0)$, $(1,1,0,2)$, $(1,0,3,0)$
Order 5	$(1,0,0,4)$, $(0,2,3,0)$, $(0,3,0,2)$
Order 6	$(0,1,3,2)$, $(0,2,0,4)$, $(0,0,6,0)$
Order 7	$(0,1,0,6)$, $(0,0,3,4)$
Order 8	$(0,0,0,8)$

Table – Terms expressed as 4-tuples in the expansion

Case $2p \geq 10$

- Up to $p = 8$, the proof relies on positiveness arguments due to the minimum property. However, if there exists a tuple $(i_1, \dots, i_p)$ such that the exponent $\frac{i_1}{2} + \cdots + \frac{i_p}{2p} - 1 < 0$ and all the i_k are even, then this argument fails.

- Such terms do not occur for $p \leq 8$ but do occur for $p \geq 10$, for example with $(0, 2, 0, 0, 4)$.

- Under the technical assumption that for such values of $(i_1, \dots, i_p)$,

$$T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \cdots \otimes p_{E_p}(h)^{\otimes i_p} = 0,$$

(with $k = i_1 + \cdots + i_p$), we prove a similar expansion for $p \geq 10$.

Going back to Gibbs measures and coercivity problems

- We choose

$$(\alpha_1, \dots, \alpha_d) = \Bigg(\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{\dim(E_1)}, \underbrace{\frac{1}{4}, \dots, \frac{1}{4}}_{\dim(E_2)}, \dots, \underbrace{\frac{1}{2p}, \dots, \frac{1}{2p}}_{\dim(E_p)} \Bigg)$$

  and the basis B adapted to the decomposition $\mathbb{R}^d = E_1 \oplus \cdots \oplus E_p$.
- So that

$$\frac{1}{t}\left[f(x^\star + B \cdot (t^{\alpha_1} h_1, \dots, t^{\alpha_d} h_d) - f(x^\star)\right] \to g(h_1, \dots, h_d),$$

  where g is a non-negative polynomial function expressed before ; this satisfies the assumption of [Athreya-Hwang 2010] in the case where some derivatives of f are degenerate.
- The conclusion is that

$$\left(\frac{(B^{-1} \cdot X_t)_1}{t^{\alpha_1}}, \dots, \frac{(B^{-1} \cdot X_t)_d}{t^{\alpha_d}} \right) \xrightarrow{\mathscr{L}} X \quad \text{as } t \to 0$$

  where the distribution of X has a density proportional to $e^{-g(x_1, \dots, x_d)}$.
- However, e^{-g} might be not in $L^1(\mathbb{R}^d)$; this happens when g is not coercive, for example $f(x,y) = (x - y^2)^2 + x^6$, $g(x,y) = (x-y^2)^2$.
- We give methods to deal with simple non-coercive cases, but we do not give a general formula.

Thank you for your attention !