PRINCIPLED MAP ESTIMATION FOR INVERSE PROBLEMS: BRIDGING THE GAP BETWEEN CONVERGENCE AND PERFORMANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Pretrained denoisers provide a powerful way to incorporate image priors into restoration algorithms. Plug-and-Play and RED approaches exploit fixed-noise-level denoisers within first-order optimization schemes, with convergence guarantees, but often struggle to achieve high-quality reconstruction on severely ill-posed inverse problems. In contrast, recent state-of-the-art approaches leverage denoisers derived from flow- or diffusion-based generative models and evaluate them along a sequence of decreasing noise levels. While these methods achieve strong empirical performance, their convergence theory remains limited. In this paper, we bridge this gap by specifically designing an algorithm that combines denoisers at decreasing noise levels with a schedule tailored to ensure convergence. From a Bayesian perspective, we prove that our method converges to a *Maximum a Posteriori* (MAP) estimate, under suitable assumptions. Subsequently, we apply our method to various ill-posed inverse problems and show that it surpasses convergent methods while competing with state-of-the-art empirical ones.

1 INTRODUCTION

An inverse problem in imaging aims to reconstruct a clean image $x^* \in \mathbb{R}^d$ from a degraded measurement $y \in \mathbb{R}^m$. In general, the degradation process, or *forward model*, can be written as $y = \mathrm{A}(x^*) + \varepsilon$ where $\mathrm{A} : \mathbb{R}^d \to \mathbb{R}^m$ is the degradation operator and $\varepsilon \sim \mathcal{N}(0, \sigma_y^2 \mathrm{Id})$ represents additive white Gaussian noise. A classical approach estimates the clean image x^* by the minimizer $\hat{x}$ of a composite functional:

$$\hat{x} \in \operatorname*{Argmin}_{x \in \mathbb{R}^d} f(x) + g(x), \tag{1}$$

where f is a *data-fidelity* term, that depends on the observation y and on the degradation, typically $f(x) \propto \|\mathrm{A}(x) - y\|^2$, whereas g is a *regularization* term encoding prior information about the image. A key question is therefore how to choose the regularizer g.

Early works relied on hand-crafted regularizers, such as the ℓ_1-norm, total-variation, and wavelet sparsity (Rudin et al., 1992; Donoho & Johnstone, 1998; Scherzer et al., 2009). For these choices of g, (1) can easily be solved with classical gradient-descent or proximal algorithms. However, the resulting reconstruction $\hat{x}$ often lacks perceptual quality due to the limited expressiveness of the regularization. In order to integrate more complex prior information about clean images, later approaches considered the use of a pretrained denoiser D_σ with fixed noise level σ within optimization schemes, as an implicit regularizer. Plug-and-Play (PnP) methods (Venkatakrishnan et al., 2013; Chan et al., 2016) replaced the proximal operator by such a denoiser, which significantly improved perceptual quality of the reconstruction on simple inverse problems such as denoising and deblurring. Under certain conditions on the denoiser, PnP methods can be proven to converge, but the limit of the resulting sequence generally cannot be characterized (Ryu et al., 2019; Pesquet et al., 2021; Hurault et al., 2022; Hurault, 2023). Regularization by Denoising (RED, Romano et al. 2017) partially bridges the gap between gradient based methods and PnP ones, by using a specific regularizer: $g(x) = \mu x^\top (x - D_\sigma(x))$. When D_σ is locally homogeneous and has symmetric Jacobian, $\nabla g(x) = \mu(x - D_\sigma(x))$, and gradient descent on (1) leads to Algorithm 1; however these conditions rarely hold for trained denoisers (Reehorst & Schniter, 2019).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Algorithm 1: RED

Initialization: $x_0 \in \mathbb{R}^d$, $\alpha, \sigma > 0$
for $k = 0, 1, \dots$ **do**
$\lfloor x_{k+1} = x_k - \alpha \nabla f(x_k) - \alpha\mu(x_k - D_\sigma(x_k))$

Algorithm 2: Annealed RED

Initialization: $x_0 \in \mathbb{R}^d$, $(\alpha_k)_k, (\sigma_k)_k \searrow 0$
for $k = 0, 1, \dots$ **do**
$\lfloor x_{k+1} = x_k - \alpha_k \nabla f(x_k) - \alpha_k \mu_k (x_k - D_{\sigma_k}(x_k))$

Despite appealing guarantees, the performance of such methods remains limited on challenging inverse problems such as inpainting or super-resolution. To address this limitation, state-of-the-art algorithms use off-the-shelf diffusion- or flow-based generative models (Song et al., 2020; Ho et al., 2020; Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023). Most of these algorithms have two main ingredients: (approximate) MMSE denoisers D_σ accross all noise levels σ, along with an annealed noise schedule with decreasing σ across iterations (Chung et al., 2023; Song et al., 2023; Zhang et al., 2024). Among them, we are specifically interested in optimization-based methods which are explicit adaptations of PnP or RED schemes (Zhu et al., 2023; Renaud et al., 2024; Martin et al., 2025; Pourya et al., 2026). While these methods achieve impressive results on highly ill-posed inverse problems, their theoretical understanding lags behind: since the denoiser changes at each iteration, so does the underlying objective, and analyses developed for classical PnP methods no longer hold. Apart from the denoising case $\mathrm{A} = \mathrm{Id}$ (Pesme et al., 2025), the convergence of annealed schemes and the characterization of their limit remain open questions.

Our work belongs to the line of work that seeks to build interpretable and convergent inverse problems solvers based on generative models (Zhang et al., 2024; Pesme et al., 2025). Many generative-based methods rely on empirical design choices, in particular empirical noise schedules. Instead, we build an algorithm, interpretable as an annealed RED scheme (Algorithm 2), to explicitly solve the *maximum a posteriori* (MAP) problem and derive its schedules to guarantee convergence. From a Bayesian perspective, the MAP estimator is the most likely image given the observation y:

$$\hat{x}_{\mathrm{MAP}} \in \operatorname*{Argmax}_{x \in \mathbb{R}^d} \log p(x|y) = \operatorname*{Argmin}_{x \in \mathbb{R}^d} -\log p(y|x) - \log p(x). \tag{2}$$

Under a Gaussian noise model, $-\log p(y|x) = \|\mathrm{A}(x) - y\|^2 / (2\sigma_y^2)$ up to an additive constant, and (2) is an instance of (1) with $f(x) = \|\mathrm{A}(x) - y\|^2 / 2$ and $g = -\sigma_y^2 \log p$. While f is explicit, the negative log-prior $-\log p$ is unknown. Pesme et al. (2025) showed how MMSE denoisers, which diffusion and flow matching models approximate, can be leveraged to solve MAP denoising problems ($\mathrm{A} = \mathrm{Id}$), i.e. to compute the prox of $-\log p$: their MMSE Averaging scheme provably converges. For inverse problems, they propose Approx-PGD, which uses MMSE Averaging as an inner loop to approximate proximal steps. However, it requires a growing number of inner iterations and only guarantees convergence of the objective values, which limits its practical use (Section 5).

Our contributions are the following:

- We introduce Generalized Annealed MMSE Averaging (GAMMA), a generic inverse problem solver using MMSE denoisers evaluated at decreasing noise levels. It extends the MMSE Averaging algorithm of Pesme et al. (2025), which is restricted to MAP denoising. It also connects to existing methods such as annealed version of RED (Algorithm 2), annealed-SNORE (Renaud et al., 2024) and PnP-Flow (Martin et al., 2025), which do not possess convergence guarantees.

- Under log-concavity of the distribution p and mild assumptions on the data-fidelity f, we prove that the iterates of GAMMA, and of a variant involving renoising of the iterates, converge to a MAP estimate of the inverse problem. When multiple MAP solutions exist, our algorithm is able to guide the iterates towards the one closest to a reference point.

- We show that our method converges faster than Approx-PGD (Pesme et al., 2025), and is competitive with state-of-the-art annealed methods on CelebA and AFHQ, making it both provably convergent and practical.

2 PROPOSED ALGORITHM: GENERALIZED ANNEALED MMSE AVERAGING (GAMMA)

2.1 PROBLEM & BACKGROUND

We aim to solve the inverse problem (1) by computing a MAP estimate

$$\hat{x} \in \operatorname*{Argmin}_{x \in \mathbb{R}^d} f(x) - \tau \log p(x), \tag{MAP}$$

where $f(x) := \frac{1}{2}\|\mathrm{A}(x) - y\|^2$, p denotes the prior distribution over clean images and $\tau := \sigma_y^2 > 0$ denote the noise level of the additive noise. In our work, we assume the forward process is known, while the underlying probability distribution p is not. Rather, we suppose we have access to the associated *Minimum Mean Square Error* (MMSE) denoiser at every noise level $\sigma > 0$, defined by

$$\mathrm{MMSE}_\sigma(z) = \mathbb{E}_{X \sim p, \varepsilon \sim \mathcal{N}(0,\mathrm{Id})}\left[X \mid X + \sigma\varepsilon = z\right] \in \mathbb{R}^d. \tag{3}$$

In practice, MMSE denoisers can be approximated, up to reparameterization, by Flow Matching or diffusion models. Indeed, those methods implicitly learn the score of the smoothed distribution $p_\sigma = p * \mathcal{N}(0, \sigma^2\mathrm{Id})$ which can be related to the MMSE denoiser at level $\sigma > 0$ through Tweedie's formula (Efron, 2011):

$$\mathrm{MMSE}_\sigma(z) = z + \sigma^2 \nabla \log p_\sigma(z). \tag{4}$$

Pesme et al. (2025) introduce MMSE Averaging, an algorithm to solve (MAP) in the denoising setting, namely when $f(x) = \frac{1}{2}\|x - y\|^2$; in that case the MAP is simply the proximal operator of $-\tau \log p$. Its iterations read as follows

$$x_{k+1} = \alpha_k y + (1 - \alpha_k)\mathrm{MMSE}_{\sigma_k}(x_k), \tag{MMSE Averaging}$$

with $(\alpha_k)_k$ a weight sequence, $(\sigma_k)_k$ an annealing noise level sequence. Plugging in Tweedie's formula (4), we can write each iteration as a gradient step:

$$x_{k+1} = x_k - \alpha_k \left((x_k - y) - \left(\frac{1 - \alpha_k}{\alpha_k} \right) \sigma_k^2 \nabla \log p_{\sigma_k}(x_k) \right). \tag{5}$$

The gradient step is taken on a smoothed objective function that is better conditioned than the main objective function $F = f - \tau \log p$. Convergence is derived under suitable assumptions, in particular the log-concavity of p and a specific choice for the sequences $(\alpha_k)_k$ and $(\sigma_k)_k$. However, this scheme is restricted to denoising. We now introduce an algorithm in the same spirit as (MMSE Averaging), but applicable to a much broader class of inverse problems.

2.2 METHOD FORMULATION

We present Generalized Annealed MMSE Averaging (GAMMA), a new algorithm based on MMSE denoisers. It generalizes MMSE Averaging to any inverse problems (beyond denoising / prox computation), while preserving its convergence guarantees. Compared to MMSE Averaging, it replaces the fixed term y by a gradient descent step on a regularized version of the datafit f.

Generalized Annealed MMSE Average (GAMMA): Let $(\alpha_k)_k, (\sigma_k)_k$ be two positive annealing sequences. Starting from an arbitrary initialization $x_0 \in \mathbb{R}^d$, we define the following iterates

$$x_{k+1} = \alpha_k(x_k - \nabla f_{\sigma_k}(x_k)) + (1 - \alpha_k)\mathrm{MMSE}_{\sigma_k}(x_k), \tag{6}$$

where $f_\sigma(x) := f(x) + \frac{\lambda(\sigma^2)}{2}\|x - u\|^2$ is a regularized data-fidelity term, with an annealing regularization weight $\lambda(\sigma^2)$, and $u \in \mathbb{R}^d$ a reference point that we can choose.

We also introduce a variant of GAMMA in which the denoising of the iterate is replaced by the average of its denoised noisy versions. Such averaging was previously proposed by Renaud et al. (2024) and Martin et al. (2025, Remark 3) to improve the empirical performance of their respective algorithms. In practice, we show in Figure 3 that the number of samples ε used to approximate the expectation can be taken equal to 1 – leading to a *de facto* stochastic algorithm.

Renoised-GAMMA: With the same notation and an arbitrary intialization $x_0 \in \mathbb{R}^d$, we define the following iterates

$$\tilde{x}_{k+1} = \alpha_k(\tilde{x}_k - \nabla f_{\sigma_k}(\tilde{x}_k)) + (1-\alpha_k)\mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathrm{Id})}\left[\mathrm{MMSE}_{\sigma_k}(\tilde{x}_k + \sigma_k\varepsilon)\right], \tag{7}$$

Similarly to Pesme et al. (2025), our update is an average between a gradient step on the data-fidelity and the denoised previous iterate. Both terms hold complementary information, and averaging them allows us to minimize the composite objective function. Using (4) yields the following key result.

Proposition 1. *The iterations of (6) can be written as a gradient descent step when we fix* $\left(\frac{1-\alpha_k}{\alpha_k}\right)\sigma_k^2 = \tau$*:*

$$x_{k+1} = x_k - \alpha_k \nabla F_{\sigma_k}(x_k), \quad F_{\sigma_k}(x) := f_{\sigma_k}(x) - \tau \log p_{\sigma_k}(x) \tag{8}$$

Similarly, the renoised iterations (7) writes,

$$\tilde{x}_{k+1} = \tilde{x}_k - \alpha_k \nabla G_{\sigma_k}(\tilde{x}), \quad G_{\sigma_k}(x_k) := f_\sigma(x) - \tau\mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathrm{Id})}\left[\log p_{\sigma_k}(x+\sigma_k\varepsilon)\right] \tag{9}$$

Rather than directly running gradient descent steps on the MAP objective $f - \tau\log p$, the algorithm runs it on smoothed objectives F_{σ_k} or G_{σ_k}, with $\sigma_k > 0$ decreasing at each iteration: the convex data-fidelity term is replaced by a strongly convex term and the log-prior is replaced by a smoothed log-prior. We observe that F_σ converges pointwise to the objective function $F = f - \tau\log p$ as σ goes to 0:

$$F_\sigma(x) = f_\sigma(x) - \tau\log p_\sigma(x) \underset{\sigma\to 0}{\longrightarrow} f(x) - \tau\log p(x) = F(x). \tag{10}$$

The same pointwise convergence also holds for G_σ.

Remark 2. *We also consider a stochastic version of (7), where we only use one noise sample instead of averaging over infinitely many noise samples. We chose to detail the setting and the analysis of this stochastic scheme in Appendix B.4.*

3 Related works

Most existing convergence analyses of PnP and RED algorithms consider a fixed noise level and rely on structural assumptions on the denoiser, such as boundedness, non-expansivity, or suitable Jacobian properties (Ryu et al., 2019; Reehorst & Schniter, 2019; Terris et al., 2020; Hurault, 2023). Recent generative-model-based methods instead use denoisers at decreasing noise levels (Zhu et al., 2023; Mardani et al., 2024; Martin et al., 2025; Pourya et al., 2026), for which the implicit regularization changes throughout the iterations and the fixed-noise theory does not readily apply. Our construction is specifically designed for this decreasing-noise setting, with a schedule that preserves a variational interpretation and enables convergence analysis.

Link with previous methods. Our method is connected to several existing works.

- **MMSE Averaging.** The iterates (6) recover MMSE Averaging when $f(x) = \frac{1}{2}\|x-y\|^2$ and $\lambda(\sigma^2) = 0$.
- **RED.** RED (Romano et al., 2017) performs gradient descent on $f+g$ with $g(x) = \mu x^\top(x - D_\sigma(x))$, which gives Algorithm 1 provided D_σ is locally homogeneous with a symmetric Jacobian (Reehorst & Schniter, 2019). In the GAMMA iterates (6), if one sets $\lambda(\sigma^2) = 0$, a simple computation shows we recover an annealed version of RED (Algorithm 2) with $\mu_k = \frac{1-\alpha_k}{\alpha_k}$.
- **SNORE** (Renaud et al., 2024) adds noise to the input of the denoiser in RED, and converges to a critical point of a smoothed MAP objective at fixed noise level. With MMSE denoisers, its annealed variant is a single-sample version of Renoised-GAMMA (7) with $\lambda(\sigma^2) = 0$. Unlike GAMMA, its noise schedule is not designed to preserve convergence as σ decreases, and no guarantees are currently available for it.
- **PnP-Flow.** Up to renaming the iterates, one iteration of PnP-Flow (Martin et al., 2025) writes:

$$\begin{cases} z_{k+1} = D_{\sigma_k}(x_k + \sigma_k\varepsilon_k), & \varepsilon_k \sim \mathcal{N}(0,\mathrm{Id}), \\ x_{k+1} = z_k - \alpha_k\nabla f(z_k). \end{cases} \tag{PnP-Flow}$$

When the denoiser is learned optimally to be the MMSE denoiser, one can then leverage (4) to interpret an iteration of PnP-FLow as two consecutive gradient steps, rather than as a single gradient step as in GAMMA. While it is in general difficult to interpret PnP-Flow as a minimization algorithm, in the denoising case, i.e., $f(x) = \frac{1}{2}\|x - y\|^2$, a single iteration of (PnP-Flow) can be written as an iteration of (MMSE Averaging), in which noise is added at each step to the input of the denoiser (see details in Appendix C.1).

- **Approximate Proximal Gradient Descent.** Finally, as MMSE Averaging allows approximate computation of the proximal operator of $-\tau \log p$, Pesme et al. (2025, Algorithm 1) proposed to use this approximation in a proximal gradient descent scheme for inverse problems with generic data-fidelity f. Approx-PGD inherits two limitations of inexact proximal algorithms: the inner loop to approximate the proximal operator requires increasingly many iterations[1] making the algorithm computationally expensive, and the convergence results apply to objective values $F(x_k)$ without ensuring the convergence of the iterates themselves.

Conceptually, our framework provides a common perspective on MMSE Averaging, an annealed variant of RED, and SNORE, while also revealing a close connection with PnP-Flow. In addition to circumventing the limitations of Approx-PGD described above, our approach gives competitive results on imaging inverse problems, hence bridging the gap between the theory developed in Pesme et al. (2025) and algorithms used in practice.

Optimization perspective. From an optimization viewpoint, GAMMA is closely related to continuation methods (Allgower & Georg, 2012): it performs gradient steps on smoothed objectives F_{σ_k} which converge pointwise to F as $\sigma_k \to 0$. Unlike classical continuation, it takes a single step on each smoothing level instead of solving each intermediate problem. Our construction combines two regularization mechanisms. First, a vanishing Tikhonov term is added to the data-fidelity, a classical technique in convex optimization and inverse problems whose convergence properties have been studied, e.g., by Attouch & László (2024). Second, the prior is Gaussian-smoothed inside the logarithm, through $-\log p_\sigma$. This differs from standard Gaussian smoothing (Nesterov & Spokoiny, 2017) which would be applied directly to the objective function (hence to $-\log p$). This form arises naturally from MMSE denoisers and requires a specific analysis as $\sigma \to 0$.

4 Convergence Results

In this section, we investigate the convergence of GAMMA to the MAP estimate. We show that, under suitable assumptions on the problem, iterates from (6) and (7) converge to a MAP estimate i.e a minimum of F.

4.1 Assumptions on the problem

We impose some assumptions on the data-fidelity f and p. Indeed, we need to ensure that F is well defined and admits minimizers (i.e. MAP estimates exist) in order to prove convergence. Therefore, we enforce the following properties:

H1. *The function f is convex, lower-bounded, twice continuously differentiable and L_f-smooth.*

H2. *The probability distribution p on $\mathbb{R}^d$ is such that $p > 0$ and $-\log p$ is convex, three times differentiable and with bounded third derivative. We define $M \geqslant 0$ as*

$$M = \sup_{x \in \mathbb{R}^d} \left\| \nabla^3 \log p(x) \right\|_F , \tag{11}$$

where for $\mathrm{A} \in \mathbb{R}^{d \times d \times d}$, $\|\mathrm{A}\|_F := \left(\sum_{i,j,k} \mathrm{A}_{i,j,k}^2 \right)^{1/2}$ corresponds to the Frobenius norm.

Discussion on the assumptions. While H1 on the data-fidelity term f is typically satisfied in standard inverse problems, the log-concavity assumption on the prior p in H2 is more restrictive. Nevertheless, log-concavity provides a natural and tractable setting for studying decreasing-noise schemes and is a standard simplifying assumption in theoretical analyses of annealed methods (Dalalyan,

[1] at iteration k, it requires $\mathcal{O}(k^{1+\eta})$ with $\eta > 0$

2017; Brosse et al., 2018; Durmus et al., 2019). It is also the setting adopted by Pesme et al. (2025) to establish convergence guarantees for MMSE Averaging.

These assumptions guarantee some nice properties for F and F_σ, proven in Appendix A.2.

Lemma 3. *Under H1 and H2, F is well-defined, convex and coercive.*

Notably, F admits a non-empty and convex set of minimizer, thus the MAP estimate problem is well-defined. Additionally, the convexity of $-\log p$ induces the convexity of $-\log p_\sigma$ by the Prékopa-Leindler inequality. We can then establish the following key regularity properties of F_σ.

Proposition 4. *For $\sigma > 0$, F_σ is L_σ-smooth and μ_σ-strongly convex with*

$$L_\sigma = L_f + \frac{\tau}{\sigma^2} + \lambda(\sigma^2), \quad \mu_\sigma = \lambda(\sigma^2)\,. \tag{12}$$

This result, directly adapted from Pesme et al. (2025, Proposition 4), motivates our method: F_σ has more favorable optimization properties than F, suggesting that minimizing F_σ may be more tractable than minimizing F directly. In particular, for any $\sigma > 0$, the strong convexity of F_σ ensures that $x_\sigma^* := \operatorname{Argmin} F_\sigma$ is well-defined and unique. As $\sigma \to 0$, however, the strong-convexity parameter μ_σ vanishes, and the behavior of $x_\sigma^\star$ is therefore not immediate. A key step in our analysis is to show that, under suitable assumptions, $(x_\sigma^*)_{\sigma>0}$ converges to a minimizer of F as $\sigma \to 0$.

4.2 On the minimizers of F and F_σ

To investigate the properties of the mapping $\sigma^2 \mapsto x_\sigma^*$, we enforce some mild assumptions on the regularization weight λ, which plays a key role in our analysis. Specifically, we assume that λ is continuously differentiable on $(0, +\infty)$ and satisfies $\lambda(0) = 0$. The implicit function theorem then ensures that the mapping $\sigma^2 \mapsto x_\sigma^*$ is continuously differentiable on $(0, +\infty)$. However, its behavior and its regularity as $\sigma \to 0$ are not immediate and are the object of the following propositions.

Proposition 5 (Limit of the minimizers)**.** *Let $\sigma_{\max} > 0$. Then, $(x_\sigma^*)_{\sigma\in(0,\sigma_{\max}]}$ is bounded. More specifically,*

(i) as $\sigma \to 0$, every cluster point of x_σ^* is a minimizer of F.

(ii) additionally, if $\frac{\sigma^2}{\lambda(\sigma^2)} \underset{\sigma\to 0}{\longrightarrow} 0$, then $x_\sigma^* \underset{\sigma\to 0}{\longrightarrow} x^\dagger$, where

$$x^\dagger := \operatorname{argmin}\{\|z - u\| \,|\, z \in \operatorname{Argmin} F\}, \tag{13}$$

which exists and is well defined.

See Appendix A.3 for the proof. The second statement is useful: if $\lambda(\sigma^2)$ vanishes more slowly than σ^2 as $\sigma \to 0$, then the minimizers x_σ^* converge to the minimizer of F closest to u, i.e., the projection of u onto $\operatorname{Argmin} F$. Thus, we have not only convergence of the minimizers to a point of $\operatorname{Argmin} F$, but also uniqueness of the limit, which can be determined via the choice of $u \in \mathbb{R}^d$.

We next establish a result controlling the regularity of the path $\sigma \mapsto x_\sigma^*$ (see Appendix A.3 for the proof).

Proposition 6 (Regularity of the minimizers)**.** *Let $\sigma_{\max} > 0$. There exist three constants $C_1, C_2, C_3 > 0$ depending on the problem parameters, such that for any $0 < \sigma_1 \leqslant \sigma_2 \leqslant \sigma_{\max}$,*

$$\left\|x_{\sigma_1}^* - x_{\sigma_2}^*\right\| \leqslant \left(C_1 + \frac{C_3}{\lambda(\sigma_1^2)}\right)(\sigma_2^2 - \sigma_1^2) + C_2 \log\left(\frac{\lambda(\sigma_2^2)}{\lambda(\sigma_1^2)}\right) \tag{14}$$

4.3 Convergence of the method

We now state our main theoretical result: the convergence of GAMMA and the different mode of convergence we can obtain according to the schedules chosen.

Theorem 7 (Convergence of GAMMA)**.** *Let $(\sigma_k)_k \in \mathbb{R}_+^{\mathbb{N}}$ be a sequence strictly decreasing to 0 and, for any $k \geqslant 0$, let $\lambda_k := \lambda(\sigma_k^2) \in (0, 1)$ and $\alpha_k := \frac{\sigma_k^2}{\sigma_k^2+\tau}$. Assume moreover that:*

(i) $\sum_{k\geqslant 0} \sigma_k^2 \lambda_k = +\infty$;

(ii) $\frac{\sigma_k^2 - \sigma_{k+1}^2}{\sigma_{k+1}^2 \lambda_{k+1}^2} \longrightarrow 0$;

(iii) $\frac{\log \lambda_k - \log \lambda_{k+1}}{\sigma_{k+1}^2 \lambda_{k+1}} \longrightarrow 0$;

Under H1 and H2, the sequence of iterates $(x_k)_{k\geqslant 0}$ *defined in (6) from any starting point* $x_0 \in \mathbb{R}^d$*, is bounded and is such that each cluster point of* $(x_k)_k$ *is a minimizer of* F*. In addition, if* $\frac{\sigma_k^2}{\lambda_k} \underset{k\to+\infty}{\longrightarrow} 0$*, then the sequence* $(x_k)_k$ *converges to*

$$x^\dagger = \operatorname{argmin}\{\|z - u\| \mid z \in \operatorname{Argmin} F\} = \operatorname{Proj}_{\operatorname{Argmin} F}(u), \tag{15}$$

the nearest minimizer of F *to* u*.*

Remark 8. *A generic family of sequences* $(\alpha_k)_k, (\sigma_k)_k$ *and* $(\lambda_k)_k$ *satisfying the requirements of Theorem 7 is not difficult to exhibit. For example, the following choices match all conditions:*

$$\lambda_k = \frac{\lambda_0}{(k+1)^\beta}, \quad \sigma_k^2 = \frac{\sigma_0^2}{(k+1)^\gamma}, \quad \alpha_k = \frac{\sigma_k^2}{\sigma_k^2 + \tau},$$

where $0 < \beta < \gamma$, $0 < \gamma + \beta < 1$, $\sigma_0 > 0$ *and* $\lambda_0 \in (0, 1)$.

Sketch of proof. The full proof is in Appendix B.2. It relies on first showing that $\left\|x_{k+1} - x^*_{\sigma_k}\right\|$ goes to 0. This is achieved by upper bounding it by a function of $\left\|x_k - x^*_{\sigma_k}\right\|$ (using a bound of the form $\left\|x_{k+1} - x^*_{\sigma_k}\right\| < \left\|x_k - x^*_{\sigma_k}\right\|$ given by the descent lemma, and the bound on $\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|$ given by Proposition 6). We then show convergence to 0 by applying a classical lemma on real sequences satisfying a perturbed contraction inequality. This first result allows proving that the cluster points of $(x_k)_k$ and those of $(x_{\sigma_k})_k$ are the same; applying Proposition 5 then shows that cluster points of $(x_k)_k$ are minimizers of F.

We prove a similar result for the variant (7), Renoised-GAMMA. Under the same constraints on the schedules, we show convergence to the MAP estimate.

Theorem 9 (Convergence of Renoised-GAMMA)**.** *Let* $(\sigma_k)_k \in \mathbb{R}_+^{\mathbb{N}}$ *be a sequence strictly decreasing to* 0 *and, for any* $k \geqslant 0$*, let* $\lambda_k := \lambda(\sigma_k^2) \in (0, 1)$ *and* $\alpha_k := \frac{\sigma_k^2}{\sigma_k^2 + \tau}$*. Under constraints (i), (ii), (iii) of Theorem 7 and* $\frac{\sigma_k^2}{\lambda_k} \to 0$*, the iterates* $(\tilde{x}_k)$ *from (7) converges to* $x^\dagger$*, i.e,*

$$\left\|\tilde{x}_k - x^\dagger\right\| \underset{k\to+\infty}{\longrightarrow} 0\,. \tag{16}$$

Sketch of proof. The full proof is given in Appendix B.3. It again shows that $\left\|x_{k+1} - x^*_{\sigma_k}\right\| \to 0$. Since $x^*_{\sigma_k}$ does not minimize G_{σ_k}, the descent lemma cannot be applied directly. We instead combine several inequalities to control this distance and establish a contraction between $\left\|x_{k+1} - x^*_{\sigma_k}\right\|$ and $\left\|x_k - x^*_{\sigma_k}\right\|$. The remainder follows the same argument as above.

5 Numerical Results

5.1 Convergence Speed on a Gaussian Mixture Prior

To better assess the computational advantage of GAMMA over Approx-PGD, we consider a synthetic problem in which the prior p is an eight-component Gaussian mixture in dimension $d = 100$, with means $\mu_i \sim \mathcal{N}(0, \mathrm{Id})$ for $i = 1, \ldots, 8$. We consider a random inpainting problem, where the degradation operator A masks 80% of the vector entries. The corresponding objective is $F(x) = \frac{1}{2}\|\mathrm{A}(x) - y\|^2 - \tau \log p(x)$ with $\tau = (0.3)^2$. Since F can be evaluated explicitly in this setting, we compare the convergence of both methods as a function of computation time. The results are shown in Figure 1.

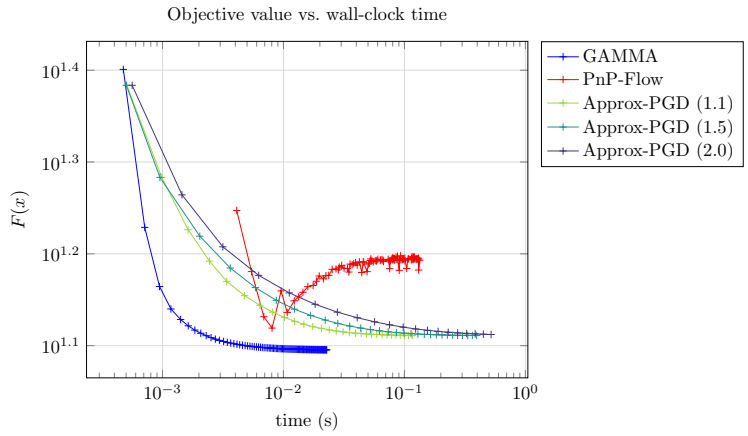


Figure 1: Evolution of the objective value across time of GAMMA, PnP-Flow and Approx-PGD $1+\eta$ (with η controlling the number of sub-iterations). **PnP-Flow diverges. GAMMA converges faster than Approx-PGD.**

5.2 Benchmarks

We compare our GAMMA algorithm with two complementary approaches: **PnP-Flow** (Martin et al., 2025), which provides a state-of-the-art reference for image restoration methods with annealed denoisers, and **Approx-PGD** (Pesme et al., 2025) for the theoretically grounded baseline solving the MAP problem. These experiments assess whether GAMMA can combine strong empirical performance when used with the identified schedules guaranteeing convergence toward the MAP estimator. We evaluate three variants of our algorithm:

- **GAMMA** The deterministic scheme of Equation (6), combined with the noise schedule identified in Remark 8, corresponding to the regime for which convergence is theoretically proven.
- **Renoised-GAMMA** The renoised scheme of Equation (7), using the same schedule as above, i.e. also within the theoretically proven regime.
- **Relaxed GAMMA** A variant in which we adopt the noise schedule used by PnP-Flow, corresponding to uniform time sampling in their formulation, i.e. $\sigma_k = \frac{N-k}{N}$ where N is a fixed number of iterations. Note that, for the other sequences (λ_k) and (α_k), we remain within the principled scheduling framework of Remark 8, i.e. $\lambda_k = \lambda_0 \sigma_k$ and $\alpha_k = \sigma_k^2/(\sigma_k^2 + \tau)$.

We report quantitative results on CelebA-128 (Liu et al., 2015) in Table 1, and on AFHQ-Cat-256 (Choi et al., 2020) in Table 4 (Appendix D, using PSNR, SSIM, and LPIPS (Zhang et al., 2018) as evaluation metrics. We also provide qualitative comparisons in Figure 2, showing reconstructions obtained by the different methods across several inverse problems. Training details for the backbone network and hyperparameters are detailed in Appendix D.

Discussion Approx-PGD, as noted in Vert et al. (2026), tends to produce cartoon-like images, an effect mitigated by early stopping. Despite tuning the number of steps, it does not perform satisfactorily beyond denoising. This is consistent with the original paper (Pesme et al., 2025), which reports no numerical results. **GAMMA**, in its noiseless version, performs well on deblurring and random inpainting, substantially improving over Approx-PGD. Yet, on more generative tasks (super-resolution and mask inpainting), it shows a similar collapse toward cartoon-like images, with visible artefacts. We attribute it to the absence of early stopping, as we use the same fixed iteration budget for the noiseless and renoised variants. **Renoised-GAMMA** is much closer to the PnP-Flow baseline across all tasks, including generative ones: the qualitative results are convincing and LPIPS matches that of PnP-Flow. It remains below PnP-Flow in PSNR, which we potentially attribute to the annealed noise schedule σ_k not reaching the observation noise level $\sigma_y = 0.05$ within the prescribed iteration budget, so the observation noise is not fully removed. Finally, the **relaxed GAMMA** variant shows that, by forcing the noise schedule to reach 0 within a prescribed number of steps, our algorithm matches the performance of PnP-Flow. Overall, these results highlight the versatility of our approach and the importance of carefully designed noise schedules. This contrasts with other

renoised annealed RED approaches such as SNORE (Renaud et al., 2024): although SNORE admits convergence guarantees in the fixed-noise setting, its annealed variant provides neither comparable guarantees nor principled guidance for choosing the noise schedule, and reports no results on challenging generative inverse problems such as super-resolution or box inpainting. Our experiments also highlight the importance of renoising in generative inverse problems. Our theory incorporates this step, ubiquitous in generative-based methods, while remaining within the MAP framework.

Table 1: Comparisons of methods on different inverse problems on the CelebA dataset. Results are averaged across 100 test images. Higher PSNR / SSIM is better, lower LPIPS is better.

Method	Theory	Denoising $\sigma = 0.2$			Deblurring $\sigma = 0.05, \sigma_b = 3.0$			Super-res. $\sigma = 0.05, \times 4$			Rand. inpaint. $\sigma = 0.01, 70\%$			Box inpaint. $\sigma = 0.05, 80 \times 80$		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Degraded		20.00	0.683	0.148	27.78	0.736	0.129	10.25	0.174	0.819	11.95	0.188	1.032	22.26	0.740	0.212
PnP-Flow	No	**32.74**	0.915	0.056	34.85	0.940	0.047	32.05	0.913	0.056	**34.86**	**0.962**	**0.018**	**32.02**	0.946	0.040
Approx-PGD	Yes	26.98	0.718	0.083	23.28	0.691	0.170	15.86	0.369	0.370	16.21	0.391	0.355	6.460	0.501	0.498
GAMMA (noiseless)	Yes	30.19	0.834	0.052	26.80	0.638	0.283	18.78	0.524	0.205	19.95	0.657	0.297	24.02	0.747	0.205
GAMMA (noise)	Yes	32.57	0.905	0.049	33.59	0.898	0.045	31.59	0.890	0.052	32.45	0.920	0.040	31.13	0.897	0.032
GAMMA (relaxed)	No	32.68	**0.919**	**0.036**	**35.27**	**0.950**	**0.025**	**32.50**	**0.918**	**0.049**	34.21	0.954	0.021	30.20	**0.948**	**0.026**

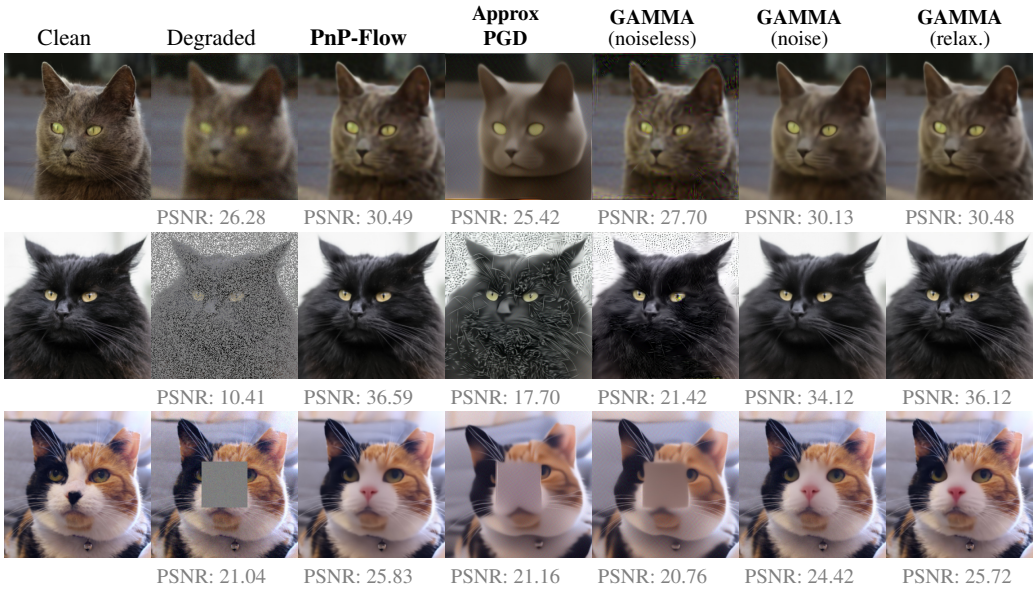


Figure 2: Qualitative results on AFHQ-256. The full images table is in the Appendix, Figure 5.

6 Conclusion

We introduced GAMMA, a convergent annealed RED-type method, with an optional renoising mechanism, whose noise and regularization schedules are explicitly designed to guarantee convergence to a MAP solution. To the best of our knowledge, this is the first approach to bridge the gap between provably convergent denoiser-based MAP methods, such as Pesme et al. (2025), and recent generative restoration methods, which typically rely on annealing and renoising without comparable guarantees. Our experiments show that this gap can be substantially narrowed with the renoised variant, achieving competitive reconstruction quality compared to recent non-convergent generative methods. When the constraints imposed by the theory are relaxed, the method reaches comparable performance, suggesting that the remaining gap is mainly due to the constraints on the current schedules. Future work will focus on relaxing the log-concavity assumption and extending the convergence analysis of the stochastic renoised scheme to less restrictive noise schedules.

REFERENCES

Michael Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations (ICLR)*, 2023.

Eugene L Allgower and Kurt Georg. *Numerical continuation methods: an introduction*. Springer Science & Business Media, 2012.

Hedy Attouch and Szilárd Csaba László. Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution. *Mathematical Methods of Operations Research*, 2024.

Herm Jan Brascamp and Elliott H Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 1976.

Nicolas Brosse, Alain Durmus, and Éric Moulines. Normalizing constants of log-concave densities. *Electronic Journal of Statistics* , 2018.

Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 2016.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR))*, 2020.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Arnak Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society Series B*, 2017.

David L Donoho and Iain M Johnstone. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 1998.

Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *Journal of Machine Learning Research (JMLR)*, 2019.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 2011.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Samuel Hurault. *Convergent plug-and-play methods for image inverse problems with explicit and nonconvex deep regularization*. PhD thesis, Université de Bordeaux, 2023.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations (ICLR)*, 2022.

Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Birkhäuser, 2002.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations (ICLR)*, 2024.

Ségolène Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. PnP-flow: Plug-and-play image restoration with flow matching. In *International Conference on Learning Representations (ICLR)*, 2025.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Kluwer Academic Publishers, 2004.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 2017.

Scott Pesme, Giacomo Meanti, Michael Arbel, and Julien Mairal. MAP estimation with denoisers: Convergence rates and guarantees. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jean-Christophe Pesquet, Audrey Repetti, Matthieu Terris, and Yves Wiaux. Learning maximally monotone operators for image recovery. *SIAM Journal on Imaging Sciences*, 2021.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A flow-matching solver for inverse problems. In *International Conference on Learning Representations (ICLR)*, 2026.

Edward T. Reehorst and Philip Schniter. Regularization by denoising: Clarifications and new interpretations. *IEEE Transactions on Computational Imaging*, 2019.

Marien Renaud, Jean Prost, Arthur Leclaire, and Nicolas Papadakis. Plug-and-play image restoration with stochastic denoising regularization. In *International Conference on Machine Learning (ICML)*, 2024.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 2017.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 1992.

Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning (ICML)*, 2019.

Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen. *Variational Methods in Imaging*. Applied Mathematical Sciences. Springer New York, NY, 2009. ISBN 978-0-387-30931-6.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Matthieu Terris, Audrey Repetti, Jean-Christophe Pesquet, and Yves Wiaux. Building firmly nonexpansive convolutional neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE global conference on signal and information processing*. IEEE, 2013.

Kenta Vert, Giacomo Meanti, Scott Pesme, Michael Arbel, and Julien Mairal. Beyond MMSE: Enhancing PnP restoration with ProxiMAP. *arXiv preprint arXiv:2605.16396*, 2026.

Hong-Kun Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 2002.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Yasi Zhang, Peiyu Yu, Yaxuan Zhu, Yingshan Chang, Feng Gao, Ying N Wu, and Oscar Leong. Flow priors for linear inverse problems via iterative corrupted trajectory matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023.

AI USE STATEMENT

Generative AI tools were used to assist with writing, language polishing, literature retrieval, and proofreading. The bibliography itself was assembled and checked manually by the authors. AI tools were also used as an additional check of the mathematical derivations and helped identify a few minor errors that did not affect the main results. All proofs were developed independently by the authors, except for Proposition 6 for which AI assistance was used both to help formalize the relevant property and during the derivation of its proof. This result was subsequently checked and validated by the authors. The authors take full responsibility for the content of the paper.

A RESULTS ON THE MINIMIZER PATH $\sigma^2 \mapsto x_\sigma^*$ (SECTION 4.2)

In this section, we prove Lemma 3 and Propositions 5 and 6.

Notation

- $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^d$,
- $\mathbb{R}_+$ denotes $[0, +\infty)$,
- for $a \in \mathbb{R}^d$ and $r \geqslant 0$, $\mathcal{B}(a, r)$ denotes the closed ball centered in a with radius r (for the standard Euclidean norm),
- $\nabla g = (\partial_i g)_{1 \leqslant i \leqslant d} \in \mathbb{R}^d$ denotes the gradient of $g : \mathbb{R}^d \to \mathbb{R}$,
- $\nabla^2 g = (\partial_{i,j} g)_{1 \leqslant i,j \leqslant d} \in \mathbb{R}^{d \times d}$ denotes the Hessian of $g : \mathbb{R}^d \to \mathbb{R}$,
- $\nabla^3 g = (\partial_{i,j,l} g)_{1 \leqslant i,j,l \leqslant d} \in \mathbb{R}^{d \times d \times d}$ denotes the third derivative tensor of $g : \mathbb{R}^d \to \mathbb{R}$,
- $\Delta g = \operatorname{Tr}\left(\nabla^2 g\right) = \sum_{i=1}^d \partial_{i,i} g$ denotes the Laplacian of g.

A.1 PRELIMINARY RESULTS ON OBJECTIVE FUNCTION F

Lemma 3. *Under H1 and H2, F is well-defined, convex and coercive.*

Proof. Recalling that $F(x) = f(x) - \tau \log p(x)$, F is well-defined since $p > 0$ by H2. It is also convex since f is convex (by H1) and $-\log p$ is convex (by H2). Finally, in order to show that F is coercive, we show that because p is a log-concave positive probability distribution, $-\log p$ is coercive.

For the sake of contradiction, suppose that $-\log p(x)$ does not tend to infinity as $\|x\| \to +\infty$. Then, there exist $A > 0$ and a sequence $(z_k)_{k \in \mathbb{N}}$ in $\mathbb{R}^d$ such that $\|z_k\| \to +\infty$, and for every k, $-\log p(z_k) \leqslant A$.

By H2, $-\log p$ is continuous, so there exists $A' \in \mathbb{R}$ such that $-\log p \leqslant A'$ on the closed unit ball $\mathcal{B}(0, 1)$. Then, by convexity of $-\log p$, we deduce that

$$-\log p(x) \leqslant \max(A, A'), \qquad \forall x \in \mathcal{C}_k,$$

where

$$\mathcal{C}_k = \{(1-t)x + t z_k \,|\quad x \in \mathcal{B}(0,1)\,,\, t \in [0,1]\}\,.$$

Thus, $p \geqslant \exp(-\max(A, A'))$ on a sequence of subsets whose mass go to $+\infty$ (since every $\mathcal{C}_k$ contains a cone of volume $\alpha \|z_k\|$, where $\alpha > 0$ is independent of k). It contradicts the integrability of p, and concludes the proof. □

As noted, as an immediate consequence of this lemma, $\operatorname{Argmin} F$ is a non-empty, convex, bounded and closed subset of $\mathbb{R}^d$.

A.2 PRELIMINARY RESULTS ON $\log p$ AND $\log p_\sigma$

We now show that $-\log p_\sigma$ inherits some coercivity properties of $-\log p$. If the coercivity of $-\log p_\sigma$ can be expected as p_σ is a smoothed version of p, we show a slightly stronger result, by proving that the family $(-\log p_\sigma)_\sigma$ is, in a sense, uniformly coercive. In what follows, we write sometimes p_0 for p to simplify the notation. This convention is natural, since p_σ converges pointwise to p when $\sigma \to 0$.

Proposition 10. *Let $A > 0$. Then, there exists $R > 0$ such that, for every $\sigma \in [0, +\infty)$,*

$$x \in \mathbb{R}^d \text{ and } \|x\| \geqslant R \Longrightarrow -\log p_\sigma(x) \geqslant A.$$

Proof. Let $A > 0$. By the coercivity of $-\log p$, there exist :

- $R > 0$ such that $-\log p(x) \geqslant A + \log(2)$ (i.e $p \leqslant \frac{1}{2}e^{-A}$) for all $x \in \mathbb{R}^d \setminus \mathcal{B}(0, R)$,
- since p is also continuous, $m > 0$ such that $-\log p$ is lower-bounded by $-\log m$, i.e $p \leqslant m$ on $\mathbb{R}^d$.

Let $R' > R$. Then, for every x such that $\|x\| \geqslant R'$,

$$\begin{aligned} p_\sigma(x) &= \frac{1}{(2\pi\sigma^2)^{d/2}} \int_{z\in\mathbb{R}^d} p(z) \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \mathrm{d}z \\ &\leqslant \frac{1}{2}e^{-A} + \frac{m}{(2\pi\sigma^2)^{d/2}} \int_{z\in\mathcal{B}(0,R)} \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \mathrm{d}z\,. \\ &\leqslant \frac{1}{2}e^{-A} + \frac{m\mathrm{Vol}\left(\mathcal{B}(0,R)\right)}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(R'-R)^2}{2\sigma^2}\right), \end{aligned}$$

where $\mathrm{Vol}\left(\mathcal{B}(0,R)\right) := \int_{z\in\mathcal{B}(0,R)} 1\, dz$. By comparing growth rates and a straightforward derivative calculation, the mapping

$$\sigma^2 \in (0, +\infty) \mapsto \frac{m\mathrm{Vol}\left(\mathcal{B}(0,R)\right)}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(R'-R)^2}{2\sigma^2}\right)$$

admits a continuous extension at 0, with value 0, and attains its maximum at $\sigma^2 = \frac{(R'-R)^2}{d}$, where its value is $\frac{d^{d/2}m\mathrm{Vol}(\mathcal{B}(0,R))}{(2\pi)^{d/2}(R'-R)^d} \exp\left(-\frac{d}{2}\right)$. Since this quantity tends to 0 as $R' \to +\infty$, there exists $R'' > R$ such that

$$\frac{d^{d/2}m\mathrm{Vol}\left(\mathcal{B}(0,R)\right)}{(2\pi)^{d/2}(R''-R)^d} \exp\left(-\frac{d}{2}\right) \leqslant \frac{1}{2}e^{-A}.$$

Thus, for every $\sigma \geqslant 0$ and $x \in \mathbb{R}^d \setminus \mathcal{B}(0, R'')$, $p_\sigma(x) \leqslant e^{-A}$, or equivalently $-\log p_\sigma(x) \geqslant A$. This concludes the proof. □

Conversely, for every $x \in \mathbb{R}^d$, the function $\sigma^2 \in [0, +\infty) \mapsto -\log p_\sigma(x)$ is continuous. Hence, it is bounded from above on every compact interval, and in particular on a neighborhood of 0.

Lemma 11. *Let $x \in \mathbb{R}^d$ and $\sigma_{\max} > 0$. There exists $b > 0$ such that for every $\sigma \in [0, \sigma_{\max}]$, $-\log p_\sigma(x) \leqslant b$.*

Finally, we recall a lemma proved in Pesme et al. (2025, Lemma 4) on the convergence of $-\log p_\sigma$:

Lemma 12. *On every compact subset of $\mathbb{R}^d$, $(\log p_\sigma)_\sigma$ converges uniformly to $\log p$ as $\sigma \to 0$.*

As an immediate consequence, on every compact subset of $\mathbb{R}^d$, F_σ converges uniformly to F as $\sigma \to 0$.

A.3 Behavior in 0

We recall that $\lambda : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a strictly increasing function, differentiable on $(0, +\infty)$, and such that $\lambda(0) = 0$. In this subsection, we want to prove the following proposition:

Proposition 5 (Limit of the minimizers)**.** *Let $\sigma_{\max} > 0$. Then, $(x^*_\sigma)_{\sigma \in (0, \sigma_{\max}]}$ is bounded. More specifically,*

(i) as $\sigma \to 0$, every cluster point of x^*_σ is a minimizer of F.

(ii) additionally, if $\frac{\sigma^2}{\lambda(\sigma^2)} \underset{\sigma \to 0}{\longrightarrow} 0$, then $x^*_\sigma \underset{\sigma \to 0}{\longrightarrow} x^\dagger$, where

$$x^\dagger := \operatorname{argmin}\{\|z - u\| \,|\, z \in \operatorname{Argmin} F\}, \tag{13}$$

which exists and is well defined.

To do so, we state a useful lemma, whose proof is postponed after the proof of Proposition 5.

Lemma 13 (Asymptotic behavior of $\log p_\sigma$ when $\sigma \to 0$)**.** *There exists a function $\tilde{Q} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ continuous such that, for any $(\sigma, x) \in \mathbb{R}_+ \times \mathbb{R}^d$,*

$$p_\sigma(x) = p(x)\left(1 + \frac{\sigma^2}{2}\Big(\Delta \log p(x) + \|\nabla \log p(x)\|^2\Big) + \sigma^3 \frac{\tilde{Q}(\sigma, x)}{p(x)}\right).$$

As a consequence, for any compact $K \subset \mathbb{R}^d$ and $\sigma_{\max} > 0$, there exists $\tilde{M} > 0$ such that

$$\forall (\sigma, x) \in [0, \sigma_{\max}] \times K, \qquad \left|\log\left(\frac{p_\sigma(x)}{p(x)}\right) - \frac{\sigma^2}{2}\Big(\Delta \log p(x) + \|\nabla \log p(x)\|^2\Big)\right| \leqslant \tilde{M}\sigma^3.$$

This lemma means that, when $\sigma \to 0$, $p_\sigma(x) \simeq p(x)\left(1 + \frac{\sigma^2}{2}(\Delta \log p(x) + \|\nabla \log p(x)\|^2)\right)$, which in fact is a bit more precise than what we need to prove Proposition 5. Nevertheless, we state it in this form since it gives a better understanding of the behavior of $\log p_\sigma$.

Proof of Proposition 5. We first prove that the mapping $\sigma^2 \mapsto x^*_\sigma$ is bounded in a neighborhood of 0. Using Lemma 11, let $b > 0$ such that $-\tau \log p_\sigma(u) \leqslant b$ for every $\sigma \in [0, \sigma_{\max}]$. Now, thanks to Proposition 10, let $R > 0$ such that for every $\sigma \geqslant 0$ and for all $x \in \mathbb{R}^d \setminus \mathcal{B}(0, R)$,

$$-\tau \log p_\sigma(x) > f(u) + b - \inf f.$$

Then, for all $x \in \mathbb{R}^d \setminus \mathcal{B}(0, R)$ and $\sigma \in [0, \sigma_{\max}]$:

$$\begin{aligned} F_\sigma(x) &= f(x) - \tau \log p_\sigma(x) + \frac{\lambda(\sigma^2)}{2}\|x - u\|^2 \\ &\geqslant \inf f - \tau \log p_\sigma(x) \\ &> f(u) + b \\ &\geqslant f(u) - \tau \log p_\sigma(u) = F_\sigma(u). \end{aligned}$$

Thus, necessarily, $x^*_\sigma \in \mathcal{B}(0, R)$ for every $\sigma \in (0, \sigma_{\max}]$ (and the calculation also shows that all the minimizers of F are in $\mathcal{B}(0, R)$).

(i) Let $\bar{x}$ be a cluster point of x^*_σ as $\sigma^2 \to 0$, and let $(\bar{\sigma}_n)_n \in (0, +\infty)^{\mathbb{N}}$, be a sequence converging to 0 such that $x^*_{\bar{\sigma}_n} \underset{n \to +\infty}{\longrightarrow} \bar{x}$. Consider also $x^\star$ a minimizer of F. Then, by Lemma 12, F_σ converges uniformly to F on the compact $\mathcal{B}(0, R)$ as $\sigma \to 0$, hence $\lim_{n \to \infty} F_{\bar{\sigma}_n}(x^*_{\bar{\sigma}_n}) = F(\bar{x})$, and:

$$F(\bar{x}) = \lim_{n \to \infty} F_{\bar{\sigma}_n}(x^*_{\bar{\sigma}_n}) \leqslant \lim_{n \to \infty} F_{\bar{\sigma}_n}(x^\star) = F(x^\star) = \inf F.$$

Thus $\bar{x} \in \operatorname{Argmin} F$, and this concludes the proof.

(ii) We first justify the existence and uniqueness of $x^\dagger = \operatorname{Argmin}\{\|z-u\| \,|\, z \in \operatorname{Argmin} F\}$. By continuity, coercivity and convexity of F (Lemma 3, H 2), $\operatorname{Argmin} F \neq \emptyset$ is compact and convex. By strict convexity, the function $z \mapsto \|z-u\|^2$ attains a unique minimum over $\operatorname{Argmin} F$.

Now, we assume that $\frac{\sigma^2}{\lambda(\sigma^2)} \underset{\sigma\to 0^+}{\to} 0$, and we want to prove that $x^*_\sigma \underset{\sigma\to 0^+}{\to} x^\dagger$. Let $\sigma_{\max} > 0$ and $\sigma \in (0, \sigma_{\max}]$. Using the optimality of x^*_σ,

$$F_\sigma(x^*_\sigma) \leqslant F_\sigma(x^\dagger) \tag{17}$$

$$f(x^*_\sigma) - \tau \log p_\sigma(x^*_\sigma) + \tfrac{\lambda(\sigma^2)}{2}\left\|x^*_\sigma - u\right\|^2 \leqslant f(x^\dagger) - \tau \log p_\sigma(x^\dagger) + \tfrac{\lambda(\sigma^2)}{2}\left\|x^\dagger - u\right\|^2 . \tag{18}$$

Reordering the terms, we obtain

$$\left\|x^*_\sigma - u\right\|^2 - \left\|x^\dagger - u\right\|^2 \leqslant \frac{2}{\lambda(\sigma^2)}\left[f(x^\dagger) - \tau \log p_\sigma(x^\dagger) - f(x^*_\sigma) + \tau \log p_\sigma(x^*_\sigma)\right] . \tag{19}$$

Moreover, since $x^\dagger \in \operatorname{Argmin} F$,

$$F(x^\dagger) = f(x^\dagger) - \tau \log p(x^\dagger) \leqslant f(x^*_\sigma) - \tau \log p(x^*_\sigma) = F(x^*_\sigma) , \tag{20}$$

and plugging (20) into (19), we obtain

$$\left\|x^*_\sigma - u\right\|^2 - \left\|x^\dagger - u\right\|^2 \leqslant \frac{2\tau}{\lambda(\sigma^2)}\left[\log p(x^\dagger) - \log p_\sigma(x^\dagger) + \log p_\sigma(x^*_\sigma) - \log p(x^*_\sigma)\right] .$$

Using Lemma 13, with $\sigma_{\max}$ and $K := \mathcal{B}(0,R)$ the compact containing $(x^*_\sigma)_{\sigma\in(0,\sigma_{\max}]}$ as well as $\operatorname{Argmin} F$, we have $\tilde{M} > 0$ such that, for any $\sigma \in [0, \sigma_{\max}]$:

$$\begin{aligned}
&|\log p(x^\dagger) - \log p_\sigma(x^\dagger) + \log p_\sigma(x^*_\sigma) - \log p(x^*_\sigma)| \\
\leqslant &\frac{\sigma^2}{2}\left|\Delta \log p(x^*_\sigma) + \left\|\nabla \log p(x^*_\sigma)\right\|^2 - \Delta \log p(x^\dagger) - \left\|\nabla \log p(x^\dagger)\right\|^2\right| + 2\tilde{M}\sigma^3
\end{aligned}$$

Since $\nabla \log p$ and $\Delta \log p$ are continuous, there exists $B > 0$ such that $\left|\Delta \log p + \|\nabla \log p\|^2 - \Delta \log p(x^\dagger) - \left\|\nabla \log p(x^\dagger)\right\|^2\right| \leqslant B$ on K. Thus,

$$\begin{aligned}
\left\|x^*_\sigma - u\right\|^2 - \left\|x^\dagger - u\right\|^2 &\leqslant \frac{2\tau}{\lambda(\sigma^2)}\left[\log p(x^\dagger) - \log p_\sigma(x^\dagger) + \log p_\sigma(x^*_\sigma) - \log p(x^*_\sigma)\right] \\
&\leqslant \frac{B\tau\sigma^2 + 4\tau\tilde{M}\sigma^3}{\lambda(\sigma^2)} \underset{\sigma\to 0}{\longrightarrow} 0 \quad \text{because } \frac{\sigma^2}{\lambda(\sigma^2)} \underset{\sigma\to 0}{\longrightarrow} 0 .
\end{aligned}$$

We obtain that any cluster point of $(x^*_\sigma)_\sigma$ is at least as close to u as $x^\dagger$. A cluster point $\bar{x}$ of $(x^*_\sigma)_\sigma$ when $\sigma \to 0$ is thus a minimizer of F ((i)) such that $\|\bar{x} - u\| \leqslant \left\|x^\dagger - u\right\|$, therefore, by uniqueness of $x^\dagger$, $\bar{x} = x^\dagger$. Finally, $x^\dagger$ is the only subsequential limit of the bounded sequence (x^*_σ) when $\sigma \to 0$, so that $x^*_\sigma \underset{\sigma\to 0}{\longrightarrow} x^\dagger$.

□

Proof of Lemma 13. We begin by using Taylor-Lagrange's Theorem in order to obtain an expansion of $p_\sigma(x)$ in the variable σ. Denoting $Z \sim \mathcal{N}(0, \mathrm{Id})$ and using $\mathbb{E}[Z] = 0$,

$$\begin{aligned}
&p_\sigma(x) \\
&= \mathbb{E}\left[p(x + \sigma Z)\right] \\
&= \mathbb{E}\left[p(x) + \sigma \nabla p(x)^T Z + \frac{\sigma^2}{2} Z^T \nabla^2 p(x) Z + \sigma^3 \int_0^1 \frac{(1-t)^2}{2} \sum_{i,j,k} \partial_{i,j,k} p(x + t\sigma Z) Z_i Z_j Z_k \, \mathrm{d}t\right] \\
&= p(x) + \frac{\sigma^2}{2}\mathbb{E}\left[Z^T \nabla^2 p(x) Z\right] + \sigma^3 \tilde{Q}(\sigma, x)
\end{aligned}$$

where

$$\tilde{Q} : (\sigma, x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto \int_0^1 \frac{(1-t)^2}{2} \mathbb{E}\left[\sum_{1 \leqslant i,j,k \leqslant d} Z_i Z_j Z_k \, \partial_{i,j,k} p(x + t\sigma Z) \right] \mathrm{d}t$$

which we will show to be well defined and a continuous function. Moreover,

$$\mathbb{E}[Z^T \nabla^2 p(x) Z] = \sum_{1 \leqslant i,j \leqslant d} \mathbb{E}[Z_i Z_j \partial_{i,j} p(x)] = \sum_i \partial_{i,i} p(x) = \Delta p(x).$$

Notice now that, for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
\nabla p(x) &= p(x) \nabla \log p(x) \\
\nabla^2 p(x)_{i,j} &= \partial_{i,j} p(x) \\
&= p(x) \left[(\nabla^2 \log p(x))_{i,j} + (\nabla \log p(x))_i (\nabla \log p(x))_j \right] \\
(\nabla^3 p(x))_{i,j,k} &= \partial_{i,j,k} p(x) \\
&= p(x) [(\nabla^3 \log p(x))_{i,j,k} \\
&\quad + (\nabla^2 \log p(x))_{i,j} (\nabla \log p(x))_k + (\nabla^2 \log p(x))_{j,k} (\nabla \log p(x))_i \\
&\quad + (\nabla^2 \log p(x))_{k,i} (\nabla \log p(x))_j + (\nabla \log p(x))_i (\nabla \log p(x))_j (\nabla \log p(x))_k].
\end{aligned}$$

As a consequence,

$$\Delta p(x) = \sum_{i=1}^d \partial_{i,i} p(x) = p(x) \left(\sum_{i=1}^d \partial_{i,i} \log p(x) + \partial_i \log p(x)^2 \right) = p(x) \left(\Delta \log p(x) + \|\nabla \log p(x)\|^2 \right).$$

We now justify that $\tilde{Q}$ is well-defined and continuous.

First, $(\sigma, z) \mapsto \displaystyle\int_0^1 \frac{(1-t)^2}{2} \sum_{1 \leqslant i,j,k \leqslant d} z_i z_j z_k \, \partial_{i,j,k} p(x + t\sigma z) \, \mathrm{d}t$ is clearly continuous.

Then, since $\nabla^3 \log p$ is uniformly bounded, there exist $M_0, M_1, M_2 > 0$ such that, for all $x \in \mathbb{R}^d$:

- $\max_{i,j} |\partial_{i,j} \log p(x)| \leqslant M_2(1 + \|x\|)$
- $\max_i |\partial_i \log p(x)| \leqslant M_1(1 + \|x\|^2)$
- $|\log p(x)| \leqslant M_0(1 + \|x\|^3)$,

so that

$|\partial_{i,j,k} p(x)| \leqslant p(x) \left(M + 3M_1 M_2 (1 + \|x\|)(1 + \|x\|^2) + M_1^3 (1 + \|x\|^2)^3 \right) = p(x) M_{\text{tot}} (1 + \|x\|^6)$,

for a certain $M_{\text{tot}} > 0$.

Thus, denoting $m > 0$ an upper bound on p (that exists by coercivity of $-\log p$, cf Lemma 3),

$$\begin{aligned}
&\mathbb{E}\left[\left| \sum_{1 \leqslant i,j,k \leqslant d} Z_i Z_j Z_k \, \partial_{i,j,k} p(x + t\sigma Z) \right| \right] \\
&\qquad \leqslant \mathbb{E}\left[\sum_{1 \leqslant i,j,k \leqslant d} |Z_i Z_j Z_k \, \partial_{i,j,k} p(x + t\sigma Z)| \right] \\
&\qquad \leqslant M_{\text{tot}} \mathbb{E}\left[p(x + t\sigma Z) \sum_{1 \leqslant i,j,k \leqslant d} |Z_i Z_j Z_k| \left(1 + \|x + t\sigma Z\|^6 \right) \right] \\
&\qquad \leqslant M_{\text{tot}} m \mathbb{E}\left[\sum_{1 \leqslant i,j,k \leqslant d} |Z_i Z_j Z_k| \left(1 + (\|x\| + \|\sigma Z\|)^6 \right) \right].
\end{aligned}$$

Let $\tilde{\sigma} > 0$ and $\tilde{R} > 0$. For any $(x, \sigma) \in \mathcal{B}(0, \tilde{R}) \times [0, \tilde{\sigma}]$,

$$\mathbb{E}\left[\sum_{1\leqslant i,j,k\leqslant d} |Z_i Z_j Z_k| \left(1 + (\|x\| + \|\sigma Z\|)^6\right)\right] \leqslant \mathbb{E}\left[\sum_{1\leqslant i,j,k\leqslant d} |Z_i Z_j Z_k| \left(1 + (\tilde{R} + \|\tilde{\sigma} Z\|)^6\right)\right]$$

which is finite since $Z \sim \mathcal{N}(0, \mathrm{Id})$. By the dominated convergence theorem, $\tilde{Q}$ is therefore continuous on $\mathbb{R}_+ \times \mathbb{R}^d$, and we eventually obtain

$$p_\sigma(x) = p(x)\left(1 + \frac{\sigma^2}{2}\left(\Delta \log p(x) + \|\nabla \log p(x)\|^2\right) + \sigma^3 \frac{\tilde{Q}(\sigma, x)}{p(x)}\right).$$

Let $K \subset \mathbb{R}^d$ compact and $\sigma_{\max} > 0$. Consider the mapping $h : (\sigma, x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto \frac{1}{2}\left(\Delta \log p(x) + \|\nabla \log p(x)\|^2\right) + \sigma \frac{\tilde{Q}(\sigma,x)}{p(x)}$, which inherits continuity from $\Delta \log p$, $\nabla \log p$, $\tilde{Q}$ and from $p > 0$, and is therefore bounded on $[0, \sigma_{\max}] \times K$. Then,

$$\begin{aligned}
&\left|\log\left(\frac{p_\sigma(x)}{p(x)}\right) - \frac{\sigma^2}{2}\left(\Delta \log p(x) + \|\nabla \log p(x)\|^2\right)\right| \\
&\qquad = \left|\log\left(1 + \frac{\sigma^2}{2}\left(\Delta \log p(x) + \|\nabla \log p(x)\|^2\right) + \sigma^3 \frac{\tilde{Q}(\sigma, x)}{p(x)}\right)\right. \\
&\qquad\qquad \left. - \frac{\sigma^2}{2}\left(\Delta \log p(x) + \|\nabla \log p(x)\|^2\right)\right| \\
&\qquad \leqslant \underbrace{\left|\log(1 + \sigma^2 h(\sigma, x)) - \sigma^2 h(\sigma, x)\right|}_{\text{bounded by a term } \propto \sigma^4 \text{ since } h \text{ remains bounded}} + \sigma^3 \underbrace{\left|\frac{\tilde{Q}(\sigma, x)}{p(x)}\right|}_{\substack{\text{continuous, thus bounded on} \\ [0,\sigma_{\max}]\times K}},
\end{aligned}$$

and we obtain the result.

□

A.4 Bound on the derivative of the minimizer path

Proposition 6 (Regularity of the minimizers)**.** *Let $\sigma_{\max} > 0$. There exist three constants $C_1, C_2, C_3 > 0$ depending on the problem parameters, such that for any $0 < \sigma_1 \leqslant \sigma_2 \leqslant \sigma_{\max}$,*

$$\left\|x^*_{\sigma_1} - x^*_{\sigma_2}\right\| \leqslant \left(C_1 + \frac{C_3}{\lambda(\sigma_1^2)}\right)(\sigma_2^2 - \sigma_1^2) + C_2 \log\left(\frac{\lambda(\sigma_2^2)}{\lambda(\sigma_1^2)}\right) \tag{14}$$

Proof. We show in fact that for any $\sigma_{\max} > 0$ and $0 < \sigma_1 \leqslant \sigma_2 \leqslant \sigma_{\max}$,

$$\left\|x^*_{\sigma_2} - x^*_{\sigma_1}\right\| \leqslant \left(T(\sigma_{\max}) + \frac{M\tau\sqrt{d} + L_f T(\sigma_{\max})}{2\lambda(\sigma_1^2)}\right)(\sigma_2^2 - \sigma_1^2) + r(\sigma_{\max}) \log\left(\frac{\lambda(\sigma_2^2)}{\lambda(\sigma_1^2)}\right), \tag{21}$$

where $r(\sigma_{\max}) = \sup_{\sigma \in (0, \sigma_{\max}]} \left\|x^*_\sigma - u\right\|$ and $T(\sigma_{\max}) = \sup_{\sigma \in (0, \sigma_{\max}]} \frac{1}{\tau}\left\|\nabla f(x^*_\sigma) + \lambda(\sigma^2)(x^*_\sigma - u)\right\|$ both quantity being finite. By convexity and differentiability of F_σ, x^*_σ is characterized by $\nabla F_\sigma(x^*_\sigma) = 0$. By strong convexity of F_σ, $\nabla^2 F_\sigma(x^*_\sigma)$ is invertible, hence the implicit function theorem (Krantz & Parks, 2002, Theorem 3.3.1) states that the mapping $\sigma^2 \mapsto x^*_\sigma$ is continuously differentiable, and that its derivative, denoted $\dot{x}^*_\sigma$, satisfies

$$\dot{x}^*_\sigma = -[\nabla^2 F_\sigma(x^*_\sigma)]^{-1} \left(\partial_{\sigma^2} \nabla F_\sigma\right)(x^*_\sigma). \tag{22}$$

$$\begin{aligned}
\nabla^2 F_\sigma(x^*_\sigma) &= \nabla^2 f(x^*_\sigma) - \tau \nabla^2 \log p_\sigma(x^*_\sigma) + \lambda(\sigma^2)\mathrm{Id} \\
\partial_{\sigma^2} \nabla F_\sigma(x^*_\sigma) &= -\tau \partial_{\sigma^2} \nabla \log p_\sigma(x^*_\sigma) + \lambda'(\sigma^2)(x^*_\sigma - u) \\
&\overset{\text{(a)}}{=} -\frac{\tau}{2} \nabla \Delta \log p_\sigma(x^*_\sigma) - \tau [\nabla^2 \log p_\sigma(x^*_\sigma)] \nabla \log p_\sigma(x^*_\sigma) + \lambda'(\sigma^2)(x^*_\sigma - u).
\end{aligned}$$

where ⓐ holds by Pesme et al. (2025, Lemma 3). For ease of notation, denote $M_\sigma := \nabla^2 f(x^*_\sigma) + \lambda(\sigma^2)\mathrm{Id}$ and $N_\sigma := -\tau\nabla^2 \log p_\sigma(x^*_\sigma)$ for $\sigma > 0$, which are two symmetric matrices, respectively positive definite and positive semi-definite.

We first want to bound $\|\dot{x}^*_\sigma\|$ on all $[0, \sigma_{\max}]$. We begin by expanding Equation (22) explicitly,

$$\dot{x}^*_\sigma = -[M_\sigma + N_\sigma]^{-1}\left(-\frac{\tau}{2}\nabla\Delta\log p_\sigma(x^*_\sigma) + N_\sigma\nabla\log p_\sigma(x^*_\sigma) + \lambda'(\sigma^2)(x^*_\sigma - u)\right). \tag{23}$$

Let us recall a result proved in Pesme et al. (2025, Lemma 5): under H2, $\|\nabla\Delta\log p_\sigma\|$ is uniformly bounded by $M\sqrt{d}$. Moreover, we can bound the operator norm $\|\cdot\|_{\mathrm{op}}$ of $[M_\sigma + N_\sigma]^{-1}$ and $[M_\sigma + N_\sigma]^{-1}N_\sigma$:

- Using Loewner order, $M_\sigma + N_\sigma \succeq M_\sigma \succeq \lambda(\sigma^2)\mathrm{Id}$, which implies that $[M_\sigma + N_\sigma]^{-1} \preceq \frac{1}{\lambda(\sigma^2)}\mathrm{Id}$, so that $\left\|[M_\sigma + N_\sigma]^{-1}\right\|_{\mathrm{op}} \leqslant \frac{1}{\lambda(\sigma^2)}$.
- For any $x \in \mathbb{R}^d$, $\sqrt{\lambda(\sigma^2)}\,\|x\| \leqslant \left\|M_\sigma^{1/2}x\right\| \leqslant \sqrt{L_f + \lambda(\sigma^2)}\,\|x\|$, so that

$$\frac{1}{\sqrt{L_f + \lambda(\sigma^2)}}\left\|M_\sigma^{1/2}x\right\| \leqslant \|x\| \leqslant \frac{1}{\sqrt{\lambda(\sigma^2)}}\left\|M_\sigma^{1/2}x\right\|.$$

  Let $H := [M_\sigma + N_\sigma]^{-1}N_\sigma$, then

$$\begin{aligned}
\|H\|_{\mathrm{op}} &= \sup_{x\neq 0}\ \frac{\|Hx\|}{\|x\|} \\
&\leqslant \sup_{x\neq 0}\ \frac{(\lambda(\sigma^2))^{-1/2}\left\|M_\sigma^{1/2}Hx\right\|}{(L_f + \lambda(\sigma^2))^{-1/2}\left\|M_\sigma^{1/2}x\right\|} \\
&= \sqrt{1 + \frac{L_f}{\lambda(\sigma^2)}}\ \sup_{z\neq 0}\ \frac{\left\|M_\sigma^{1/2}HM_\sigma^{-1/2}z\right\|}{\|z\|} \\
&= \sqrt{1 + \frac{L_f}{\lambda(\sigma^2)}}\ \left\|M_\sigma^{1/2}HM_\sigma^{-1/2}\right\|_{\mathrm{op}}.
\end{aligned}$$

  Notice now that $M_\sigma^{1/2}HM_\sigma^{-1/2} = [\mathrm{Id} + M_\sigma^{-1/2}N_\sigma M_\sigma^{-1/2}]^{-1}M_\sigma^{-1/2}N_\sigma M_\sigma^{-1/2}$ is a symmetric matrix whose eigenvalues are bounded by 1, so that its operator norm is at most 1. Hence, $\|H\|_{\mathrm{op}} \leqslant \sqrt{1 + \frac{L_f}{\lambda(\sigma^2)}} \leqslant 1 + \frac{L_f}{2\lambda(\sigma^2)}$.

We also know that, by optimality of x^*_σ,

$$\nabla\log p_\sigma(x^*_\sigma) = \frac{1}{\tau}\left(\nabla f(x^*_\sigma) + \lambda(\sigma^2)(x^*_\sigma - u)\right), \tag{24}$$

Therefore, by combining the previous facts, the quantity $\left\|\dot{x}^*_\sigma\right\|$ is equal to

$$\begin{aligned}
&\left\|-[M_\sigma + N_\sigma]^{-1}\left(-\frac{\tau}{2}\nabla\Delta\log p_\sigma(x^*_\sigma) + \frac{1}{\tau}N_\sigma\Big(\nabla f(x^*_\sigma) + \lambda(\sigma^2)(x^*_\sigma - u)\Big) + \lambda'(\sigma^2)(x^*_\sigma - u)\right)\right\| \\
&\leqslant \frac{1 + \frac{L_f}{2\lambda(\sigma^2)}}{\tau}\|\nabla f(x^*_\sigma) + \lambda(\sigma^2)(x^*_\sigma - u)\| + \frac{1}{\lambda(\sigma^2)}\left(\frac{M\tau\sqrt{d}}{2} + \lambda'(\sigma^2)\|x^*_\sigma - u\|\right)
\end{aligned} \tag{25}$$

Moreover, following Proposition 5, we know that x^*_σ is bounded on $(0, \sigma_{\max}]$, and since λ is continuous in 0, we can define the finite quantities $T(\sigma_{\max}) := \sup_{\sigma\in(0,\sigma_{\max}]}\frac{1}{\tau}\|\nabla f(x^*_\sigma) + \lambda(\sigma^2)(x^*_\sigma - u)\|$ and $r(\sigma_{\max}) := \sup_{\sigma\in(0,\sigma_{\max}]}\left\|x^*_\sigma - u\right\|$. Then,

$$\|\dot{x}^*_\sigma\| \leqslant T(\sigma_{\max}) + \frac{1}{\lambda(\sigma^2)}\left(\frac{M\tau\sqrt{d}}{2} + \lambda'(\sigma^2)r(\sigma_{\max}) + \frac{L_f T(\sigma_{\max})}{2}\right),$$

Recall that $\dot{x}^*_\sigma := \dfrac{\mathrm{d}x^*_\sigma}{\mathrm{d}\sigma^2}$. Then, by triangular inequality, for $0 < \sigma_1 \leqslant \sigma_2 \leqslant \sigma_{\max}$:

$$\begin{aligned}
\|x^*_{\sigma_2} - x^*_{\sigma_1}\| &= \left\| \int_{\sigma_1^2}^{\sigma_2^2} \dot{x}^*_\sigma \, \mathrm{d}\sigma^2 \right\| \\
&\leqslant \int_{\sigma_1^2}^{\sigma_2^2} \|\dot{x}^*_\sigma\| \, \mathrm{d}\sigma^2 \\
&\leqslant \int_{\sigma_1^2}^{\sigma_2^2} T(\sigma_{\max}) + \frac{1}{\lambda(\sigma^2)} \left(\frac{M\tau\sqrt{d}}{2} + \lambda'(\sigma^2) r(\sigma_{\max}) + \frac{L_f T(\sigma_{\max})}{2} \right) \mathrm{d}\sigma^2 \\
&\leqslant \left(T(\sigma_{\max}) + \frac{M\tau\sqrt{d} + L_f T(\sigma_{\max})}{2\lambda(\sigma_1^2)} \right) (\sigma_2^2 - \sigma_1^2) + r(\sigma_{\max}) \log \left(\frac{\lambda(\sigma_2^2)}{\lambda(\sigma_1^2)} \right).
\end{aligned}$$

□

B Proofs related to the convergence of the algorithm (Section 4.3)

B.1 A useful lemma

Before presenting the proof of the main theorem, we recall a lemma which will be essential to the proof of convergence in Theorems 7 and 9.

Lemma 14. *Let (β_k) be a sequence of non-negative real numbers satisfying*

$$\beta_{k+1} \leqslant (1 - \gamma_k)\beta_k + \gamma_k \delta_k \tag{26}$$

where $(\gamma_k), (\delta_k)$ satisfy the conditions:

(i) $(\gamma_k) \subset [0, 1]$ *and* $\sum_k \gamma_k = +\infty$

(ii) $\delta_k \underset{k \to +\infty}{\longrightarrow} 0$

Then $\lim_k \beta_k = 0$.

This is a specific case of the more general result of Xu (2002, Lemma 2.5).

B.2 Convergence of the deterministic iterates

Theorem 7 (Convergence of GAMMA). *Let $(\sigma_k)_k \in \mathbb{R}_+^{\mathbb{N}}$ be a sequence strictly decreasing to 0 and, for any $k \geqslant 0$, let $\lambda_k := \lambda(\sigma_k^2) \in (0, 1)$ and $\alpha_k := \frac{\sigma_k^2}{\sigma_k^2 + \tau}$. Assume moreover that:*

(i) $\displaystyle\sum_{k \geqslant 0} \sigma_k^2 \lambda_k = +\infty$;

(ii) $\frac{\sigma_k^2 - \sigma_{k+1}^2}{\sigma_{k+1}^2 \lambda_{k+1}^2} \longrightarrow 0$;

(iii) $\frac{\log \lambda_k - \log \lambda_{k+1}}{\sigma_{k+1}^2 \lambda_{k+1}} \longrightarrow 0$;

Under H1 and H2, the sequence of iterates $(x_k)_{k \geqslant 0}$ defined in (6) from any starting point $x_0 \in \mathbb{R}^d$, is bounded and is such that each cluster point of $(x_k)_k$ is a minimizer of F. In addition, if $\frac{\sigma_k^2}{\lambda_k} \underset{k \to +\infty}{\longrightarrow} 0$, then the sequence $(x_k)_k$ converges to

$$x^\dagger = \operatorname{argmin}\{\|z - u\| \mid z \in \operatorname{Argmin} F\} = \operatorname{Proj}_{\operatorname{Argmin} F}(u), \tag{15}$$

the nearest minimizer of F to u.

Proof. The first goal of this proof is to apply Lemma 14 to the sequence $\beta_{k+1} := \left\|x_{k+1} - x^*_{\sigma_k}\right\|$.

We first recall that the step k of (6) is a gradient descent step on the objective F_{σ_k} at x_k with step size α_k. Moreover, by Proposition 4, F_{σ_k} is $\mu_{\sigma_k}-$strongly convex, $L_{\sigma_k}-$smooth and twice differentiable, with $\mu_{\sigma_k} = \lambda_k$ and $L_{\sigma_k} = L_f + \frac{\tau}{\sigma_k^2} + \lambda_k$. Hence, since $0 < \alpha_k < 1 < \frac{1}{\mu_{\sigma_k}} = \frac{1}{\lambda_k}$, a classical result on gradient descent (Nesterov, 2004, Section 1.2.3) yields

$$\left\|x_{k+1} - x^*_{\sigma_k}\right\| \leqslant \max(1 - \alpha_k\lambda_k, \alpha_k L_{\sigma_k} - 1) \left\|x_k - x^*_{\sigma_k}\right\| \tag{27}$$

Since with our choice of α_k, $\alpha_k L_{\sigma_k} - 1 \underset{k\to+\infty}{\longrightarrow} 0$, and $1-\alpha_k\lambda_k \underset{k\to+\infty}{\longrightarrow} 1$, the inequality $\alpha_k L_k - 1 \leqslant 1 - \alpha_k\lambda_k$ holds for all sufficiently large k. Up to starting the analysis at a later iteration, we may assume that it holds for all $k \geqslant 0$. Thus, by triangular inequality,

$$\left\|x_{k+1} - x^*_{\sigma_k}\right\| \leqslant (1 - \alpha_k\lambda_k) \left\|x_k - x^*_{\sigma_{k-1}}\right\| + (1 - \alpha_k\lambda_k) \left\|x^*_{\sigma_{k-1}} - x^*_{\sigma_k}\right\|, \tag{28}$$

and, using Proposition 6, we can bound $\left\|x^*_{\sigma_{k-1}} - x^*_{\sigma_k}\right\|$ by

$$\left\|x^*_{\sigma_{k-1}} - x^*_{\sigma_k}\right\| \leqslant \left(C_1 + \frac{C_3}{\lambda_k}\right)(\sigma_{k-1}^2 - \sigma_k^2) + C_2 \log\left(\frac{\lambda_{k-1}}{\lambda_k}\right), \tag{29}$$

where $C_1, C_2, C_3 > 0$. To ease the notation, we denote:

$$B_k = C_1 + \frac{C_3}{\lambda_k} \quad \text{and} \quad D_k = C_2 \log\left(\frac{\lambda_{k-1}}{\lambda_k}\right). \tag{30}$$

Combining (28) and (29), and denoting $\beta_{k+1} := \left\|x_{k+1} - x^*_{\sigma_k}\right\|$, we obtain the following recursive relation

$$\beta_{k+1} \leqslant (1 - \alpha_k\lambda_k)\beta_k + \underbrace{(1 - \alpha_k\lambda_k)\left[B_k(\sigma_{k-1}^2 - \sigma_k^2) + D_k\right]}_{=:\omega_k} \tag{31}$$

We recognize the recursive relation considered in Lemma 14 with $\gamma_k = \alpha_k\lambda_k$ and $\delta_k = \frac{\omega_k}{\alpha_k\lambda_k}$. By Lemma 14, we have $\beta_k \longrightarrow 0$ under the following conditions:

$$\sum_{k=1}^{+\infty} \alpha_k\lambda_k = +\infty, \quad \frac{\omega_k}{\alpha_k\lambda_k} \underset{k\to+\infty}{\longrightarrow} 0 \tag{32}$$

Using, $\alpha_k \sim \frac{\sigma_k^2}{\tau}$ and basic equivalences, the previous conditions are equivalent to the constraints *(i)*, *(ii)* and *(iii)*. Under these constraints, we thus have $\beta_{k+1} = \left\|x_{k+1} - x^*_{\sigma_k}\right\| \underset{k\to+\infty}{\longrightarrow} 0$. By Proposition 5, the sequence $(x^*_{\sigma_k})_k$ is bounded. Since

$$\|x_k\| \leqslant \left\|x_k - x^*_{\sigma_{k-1}}\right\| + \left\|x^*_{\sigma_{k-1}}\right\| = \beta_k + \left\|x^*_{\sigma_{k-1}}\right\|, \tag{33}$$

the sequence $(x_k)_k$ is also bounded. Moreover by triangular inequality, for any $x^* \in \mathbb{R}^d$:

$$\left\|x^*_{\sigma_{k-1}} - x^*\right\| - \left\|x_k - x^*_{\sigma_{k-1}}\right\| \leqslant \|x_k - x^*\| \leqslant \left\|x_k - x^*_{\sigma_{k-1}}\right\| + \left\|x^*_{\sigma_{k-1}} - x^*\right\|. \tag{34}$$

By letting x^* be a cluster point of $(x_k)_k$ (resp. $(x_{\sigma_k})_k$ and passing to the limit in the left (resp. right) hand side of (34) with a corresponding convergent subsequence, we get that the cluster points of $(x_k)_k$ are exactly the cluster points of $(x^*_{\sigma_k})_k$. By Proposition 5, they are minimizers of F, and if in addition we suppose $\frac{\sigma_k^2}{\lambda_k} \underset{k\to+\infty}{\longrightarrow} 0$, then $(x^*_{\sigma_k})_k$ converges to $x^* = x^\dagger$ and so does $(x_k)_k$.

□

B.3 Convergence of Renoised-GAMMA

Theorem 9 (Convergence of Renoised-GAMMA)**.** *Let $(\sigma_k)_k \in \mathbb{R}_+^{\mathbb{N}}$ be a sequence strictly decreasing to 0 and, for any $k \geqslant 0$, let $\lambda_k := \lambda(\sigma_k^2) \in (0, 1)$ and $\alpha_k := \frac{\sigma_k^2}{\sigma_k^2+\tau}$. Under constraints (i), (ii), (iii) of Theorem 7 and $\frac{\sigma_k^2}{\lambda_k} \to 0$, the iterates $(\tilde{x}_k)$ from (7) converges to $x^\dagger$, i.e,*

$$\left\|\tilde{x}_k - x^\dagger\right\| \underset{k\to+\infty}{\longrightarrow} 0\,. \tag{16}$$

Proof. For ease of notation, let $\varepsilon \sim \mathcal{N}(0, \mathrm{Id})$. For $\sigma > 0$, the function $G_\sigma : x \in \mathbb{R}^d \mapsto f_\sigma(x) - \tau\mathbb{E}[\log p_\sigma(x + \sigma\varepsilon)]$ is a smoothing of F_σ.

As the proof of Pesme et al. (2025, Lemma 5) shows that $\left\|\nabla^3 \log p_\sigma\right\|_F \leqslant \left\|\nabla^3 \log p\right\|_F \leqslant M$, by the dominated convergence theorem, the mapping $x \mapsto -\tau\mathbb{E}[\log p_\sigma(x + \sigma\varepsilon)]$ is well defined and thrice continuously differentiable, and its successive derivatives write

$$-\tau\mathbb{E}[\nabla \log p_\sigma(\cdot + \sigma\varepsilon)], \quad -\tau\mathbb{E}[\nabla^2 \log p_\sigma(\cdot + \sigma\varepsilon)], \quad -\tau\mathbb{E}[\nabla^3 \log p_\sigma(\cdot + \sigma\varepsilon)], \tag{35}$$

so that it inherits convexity and $\frac{\tau}{\sigma^2}-$smoothness from $-\tau \log p_\sigma$. Thus, G_σ is $L_\sigma-$smooth and $\lambda(\sigma^2)-$strongly convex like F_σ, with $L_\sigma = \lambda(\sigma^2) + \frac{\tau}{\sigma^2} + L_f$.

Our goal is to prove an inequality similar to (27), in order to show that $\left\|\tilde{x}_{k+1} - x^*_{\sigma_k}\right\| \longrightarrow 0$. As x^*_σ is a minimizer of F_σ and the iterates are obtained by gradient descent steps on the G_{σ_k}, the descent lemma does not apply directly.

We first recall some basic inequalities that will be useful in the proof.

- Young's inequality: for $a, b \in \mathbb{R}$ and $\delta > 0$,
$$|ab| \leqslant \frac{\delta a^2}{2} + \frac{b^2}{2\delta} \tag{A}$$
- As a consequence, using Cauchy-Schwarz, for any $x, y, z \in \mathbb{R}^d$ and $\delta > 0$
$$\|x - y\|^2 \leqslant (1 + \delta)\|x - z\|^2 + \left(1 + \frac{1}{\delta}\right)\|y - z\|^2 . \tag{B}$$
- For h an L-smooth and μ-strongly convex function, and $x, y \in \mathbb{R}^d$ (Nesterov, 2004, Thm. 2.1.12)
$$\langle \nabla h(x) - \nabla h(y), x - y\rangle \geqslant \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu + L}\|\nabla h(x) - \nabla h(y)\|^2 . \tag{C}$$

By definition of our iterates,

$$\begin{aligned}\left\|\tilde{x}_{k+1} - x^*_{\sigma_k}\right\|^2 &= \left\|\tilde{x}_k - \alpha_k \nabla G_{\sigma_k}(\tilde{x}_k) - x^*_{\sigma_k}\right\|^2 \\ &= \left\|\tilde{x}_k - x^*_{\sigma_k}\right\|^2 - \underbrace{2\alpha_k\langle \nabla G_{\sigma_k}(\tilde{x}_k), \tilde{x}_k - x^*_{\sigma_k}\rangle}_{①} + \alpha_k^2 \underbrace{\|\nabla G_{\sigma_k}(\tilde{x}_k)\|^2}_{②} .\end{aligned} \tag{36}$$

We first bound ① . By Cauchy-Schwarz inequality,

$$\langle \nabla G_{\sigma_k}(\tilde{x}_k), \tilde{x}_k - x^*_{\sigma_k}\rangle \geqslant \langle \nabla G_{\sigma_k}(\tilde{x}_k) - \nabla G_{\sigma_k}(x^*_{\sigma_k}), \tilde{x}_k - x^*_{\sigma_k}\rangle - \left\|G_{\sigma_k}(x^*_{\sigma_k}\right\| \left\|\tilde{x}_k - x^*_{\sigma_k}\right\| . \tag{37}$$

Let $r_k := \left\|\tilde{x}_k - x^*_{\sigma_k}\right\|$ and $\Delta_k := \left\|\nabla G_{\sigma_k}(\tilde{x}_k) - \nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|$. Then, using (C) on G_{σ_k}:

$$\langle \nabla G_{\sigma_k}(\tilde{x}_k) - \nabla G_{\sigma_k}(x^*_{\sigma_k}), \tilde{x}_k - x^*_{\sigma_k}\rangle \geqslant \frac{\lambda(\sigma_k^2)L_{\sigma_k}}{\lambda(\sigma_k^2) + L_{\sigma_k}} r_k^2 + \frac{1}{\lambda(\sigma_k^2) + L_{\sigma_k}}\Delta_k^2. \tag{38}$$

For notational convenience, from now on $\lambda_k := \lambda(\sigma_k^2)$ and $L_k := L_{\sigma_k}$. Combining the two preceding inequalities and multiplying both sides by $2\alpha_k$, we obtain

$$2\alpha_k\langle \nabla G_{\sigma_k}(\tilde{x}_k), \tilde{x}_k - x^*_{\sigma_k}\rangle \geqslant \frac{2\alpha_k\lambda_k L_k}{\lambda_k + L_k} r_k^2 + \frac{2\alpha_k}{\lambda_k + L_k}\Delta_k^2 - 2\alpha_k \left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\| r_k \tag{39}$$

$$\overset{ⓐ}{\geqslant} \left(\frac{2\alpha_k\lambda_k L_k}{\lambda_k + L_k} - \frac{\delta}{2}\right) r_k^2 + \frac{2\alpha_k}{\lambda_k + L_k}\Delta_k^2 - \frac{2\alpha_k^2}{\delta}\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 \tag{40}$$

where ⓐ holds for any $\delta > 0$ by (A) with $a = r_k =$ and $b = \alpha_k\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\|$.

Now, we bound ② using (B) with $\delta = \frac{1}{2}$:

$$\|\nabla G_{\sigma_k}(\tilde{x}_k)\|^2 = \left\|\nabla G_{\sigma_k}(\tilde{x}_k) - \nabla G_{\sigma_k}(x^*_{\sigma_k}) + \nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 \leqslant \frac{3}{2}\Delta_k^2 + 3\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 . \tag{41}$$

Finally, we can gather the two bounds ① - ② :

$$\begin{aligned}\left\|\tilde{x}_{k+1}-x^*_{\sigma_k}\right\|^2 \leqslant & \left(1-2\frac{\alpha_k\lambda_k L_k}{\lambda_k+L_k}+\frac{\delta}{2}\right) r_k^2 \\ & + \underbrace{\alpha_k\left(\frac{3}{2}\alpha_k-\frac{2}{\lambda_k+L_k}\right)}_{\leqslant 0}\Delta_k^2 \\ & + \left(3\alpha_k^2+\frac{2\alpha_k^2}{\delta}\right)\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 \end{aligned} \tag{42}$$

As $\alpha_k(\lambda_k+L_k) \underset{k\to+\infty}{\longrightarrow} 1$, up to starting at a later iteration, the coefficient of Δ_k^2 is negative and can be dropped. Let $h_k = \frac{\alpha_k\lambda_k L_k}{\lambda_k+L_k}$, and choose $\delta = 2h_k$, so that

$$\left\|\tilde{x}_{k+1}-x^*_{\sigma_k}\right\|^2 \leqslant (1-h_k)\, r_k^2 + \left(3\alpha_k^2+\frac{\alpha_k^2}{h_k}\right)\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 . \tag{43}$$

In order to conclude, we need to bound $\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|$. By Taylor-Lagrange's theorem at order 2, for any $\epsilon = (\epsilon^i)_{1\leqslant i\leqslant d} \in \mathbb{R}^d$,

$$\nabla \log p_{\sigma_k}(x^*_{\sigma_k}+\sigma_k\epsilon) = \nabla \log p_{\sigma_k}(x^*_{\sigma_k}) + \sigma_k \nabla^2 \log p_{\sigma_k}(x^*_{\sigma_k})\epsilon + R_k(\epsilon)\,, \tag{44}$$

where, for $1 \leqslant i \leqslant d$,

$$(R_k(\epsilon))_i = \sigma_k^2 \sum_{1\leqslant j,l\leqslant d} \int_0^1 (1-t)\partial_{i,j,k}\log p_{\sigma_k}(x^*_{\sigma_k}+t\sigma_k\epsilon)\epsilon^l\epsilon^j\, dt. \tag{45}$$

As already mentioned, the proof of Pesme et al. (2025, Lemma 5) establishes that $\left\|\nabla^3 \log p_{\sigma_k}\right\|_F \leqslant \left\|\nabla^3 \log p\right\|_F \leqslant M$. Thus, by the Cauchy-Schwarz inequality:

$$|(R_k(\epsilon))_i| \leqslant \frac{\sigma_k^2}{2} \sum_{1\leqslant j,l\leqslant d} |\partial_{i,j,k}\log p_{\sigma_k}(x^*_{\sigma_k}+t\sigma_k\epsilon)|\, \|\epsilon\|_\infty^2 \leqslant \frac{d\sigma_k^2}{2}\sqrt{M}\,\|\epsilon\|_\infty^2 . \tag{46}$$

so that $\|R_k(\epsilon)\|^2 \leqslant \frac{d^3M\sigma_k^4}{4}\|\epsilon\|_\infty^4$. Since $\nabla F_{\sigma_k}(x^*_{\sigma_k}) = \nabla f_{\sigma_k}(x^*_{\sigma_k}) - \tau\nabla\log p_{\sigma_k}(x^*_{\sigma_k}) = 0$ by the optimality of $x^*_{\sigma_k}$, and $\mathbb{E}_\varepsilon[\nabla^2 \log p_{\sigma_k}(x^*_{\sigma_k})\varepsilon] = 0$ by the linearity of expectation, we obtain:

$$\begin{aligned}\left\|\nabla G_{\sigma_k}(x^*_{\sigma_k})\right\|^2 &= \left\|\nabla f_{\sigma_k}(x^*_{\sigma_k}) - \tau\mathbb{E}_\varepsilon[\nabla\log p_{\sigma_k}(x^*_{\sigma_k}+\sigma_k\varepsilon)]\right\|^2 \\ &= \tau^2\left\|\mathbb{E}_\varepsilon[R_k(\varepsilon)]\right\|^2 \\ &\leqslant \tau^2\mathbb{E}_\varepsilon[\|R_k(\varepsilon)\|^2] \\ &\leqslant \frac{d^3M\sigma_k^4\tau^2}{4}\mathbb{E}_\varepsilon[\|\varepsilon\|_\infty^4], \end{aligned} \tag{47}$$

where the first inequality follows from Jensen's. Let $B := \frac{d^3M\tau^2}{4}\mathbb{E}_\varepsilon[\|\varepsilon\|_\infty^4]$. Plugging this into (43) we obtain,

$$\left\|\tilde{x}_{k+1}-x^*_{\sigma_k}\right\|^2 \leqslant (1-h_k)\, r_k^2 + \left(3\alpha_k^2+\frac{\alpha_k^2}{h_k}\right) B\sigma_k^4 . \tag{48}$$

Since $h_k = \frac{\alpha_k\lambda_k L_k}{\lambda_k+L_k} \geqslant \frac{\alpha_k\lambda_k}{2}$ and $1 > \alpha_k\lambda_k$, we have $(3+\frac{1}{h_k}) \leqslant \frac{5}{\alpha_k\lambda_k}$, and

$$\left\|\tilde{x}_{k+1}-x^*_{\sigma_k}\right\|^2 \leqslant (1-h_k)\, r_k^2 + \frac{5\alpha_k B\sigma_k^4}{\lambda_k} . \tag{49}$$

In order to derive a recursive relation on $\beta_{k+1} := \left\|\tilde{x}_{k+1}-x^*_{\sigma_k}\right\|^2$, we need to make $\beta_k = \left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2$ appear on the right. To do so, we apply (B) to r_k^2 with $\delta = \frac{h_k}{2-2h_k}$, which yields

$$\beta_{k+1} \leqslant \left(1-\frac{h_k}{2}\right)\beta_k + (1-h_k)\left(\frac{2}{h_k}-1\right)\left\|x^*_{\sigma_k}-x^*_{\sigma_{k-1}}\right\|^2 + \frac{5\alpha_k B\sigma_k^4}{\lambda_k} \tag{50}$$

$$\leqslant \left(1-\frac{h_k}{2}\right)\beta_k + \underbrace{\frac{2}{h_k}\left\|x^*_{\sigma_k}-x^*_{\sigma_{k-1}}\right\|^2 + \frac{5\alpha_k B\sigma_k^4}{\lambda_k}}_{:=h_k\delta_k/2} . \tag{51}$$

Applying Lemma 14, $\beta_k \underset{k\to+\infty}{\longrightarrow} 0$ under the following constraints:

$$\sum_{k=0}^{+\infty} h_k = +\infty, \quad \delta_k \underset{k\to+\infty}{\longrightarrow} 0. \tag{52}$$

Since $h_k \sim \frac{\sigma_k^2}{\tau}\lambda_k$, we recover condition *(i)* of Theorem 7 from $\sum_{k=0}^{+\infty} h_k = +\infty$. By definition of δ_k,

$$\delta_k = \frac{4}{h_k^2}\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|^2 + \frac{10\alpha_k B\sigma_k^4}{h_k\lambda_k}. \tag{53}$$

By Proposition 6, there exist $C_1, C_2, C_3 > 0$ such that

$$\frac{2}{h_k}\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\| \leqslant \left(C_1 + \frac{C_3}{\lambda_k}\right)\frac{\sigma_{k-1}^2 - \sigma_k^2}{2h_k} + \frac{C_2}{2h_k}\log\left(\frac{\lambda_{k-1}}{\lambda_k}\right) \tag{54}$$

$$\underset{k\to+\infty}{\sim} \underbrace{\frac{C_3\tau(\sigma_{k-1}^2 - \sigma_k^2)}{2\sigma_k^2\lambda_k^2}}_{\longrightarrow 0,\ (ii)} + \underbrace{\frac{C_2\tau}{2\sigma_k^2\lambda_k}\log\left(\frac{\lambda_{k-1}}{\lambda_k}\right)}_{\longrightarrow 0,\ (iii)}. \tag{55}$$

so that the first term converges to 0 under conditions *(ii)* and *(iii)*. Moreover, provided that $\frac{\sigma_k^2}{\lambda_k} \to 0$,

$$\frac{10\alpha_k B\sigma_k^4}{h_k\lambda_k} \underset{k\to+\infty}{\sim} \frac{10B\sigma_k^4}{\lambda_k^2} \tag{56}$$

converges to 0. By Lemma 14, $\beta_{k+1} = \left\|x_{k+1} - x^*_{\sigma_k}\right\|^2$ converges to 0. Using Proposition 5 concludes the proof.

□

B.4 Stochastic-GAMMA and its convergence

In this section, we investigate a stochastic version of GAMMA. In this setting, we do not assume that we have access to $\mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathrm{Id})}\left[\mathrm{MMSE}_\sigma(x+\sigma\varepsilon)\right]$; instead, at each iteration, we use a single noise sample to estimate the expectation.

Stochastic-GAMMA: Given annealing noise level sequences $(\sigma_k)_k, (\nu_k)_k$, an annealing step size sequence $(\alpha_k)_k$ and an arbitrary initialization $x_0 \in \mathbb{R}^d$, we define the following stochastic iterates

$$\tilde{x}_{k+1} = \alpha_k(\tilde{x}_k - \nabla f_{\sigma_k}(\tilde{x}_k)) + (1-\alpha_k)\mathrm{MMSE}_{\sigma_k}(\tilde{x}_k + \nu_k\varepsilon_k), \tag{57}$$

where the $\varepsilon_k \sim \mathcal{N}(0, \mathrm{Id})$ are independent.

In this scheme, the noise level ν_k added to the input of the denoiser is decoupled from the denoiser's noise level σ_k. We prove convergence of such scheme under the same conditions as above, together with an additional condition on ν_k.

Theorem 15 (Convergence of Stochastic-GAMMA)**.** *Let $(\sigma_k)_k, (\nu_k)_k \in \mathbb{R}_+^{\mathbb{N}}$ be sequences strictly decreasing to 0 and, for any $k \geqslant 0$, let $\lambda_k := \lambda(\sigma_k^2) \in (0,1)$ and $\alpha_k := \frac{\sigma_k^2}{\sigma_k^2+\tau}$. Under constraints (i), (ii), (iii) of Theorem 7, $\frac{\sigma_k^2}{\lambda_k} \to 0$ and $\nu_k^2 = o(\sigma_k^2\lambda_k)$, the stochastic iterates $(\tilde{x}_k)$ from (57) converges in mean-square to $x^\dagger$, i.e,*

$$\mathbb{E}\left\|\tilde{x}_k - x^\dagger\right\|^2 \underset{k\to+\infty}{\longrightarrow} 0\,. \tag{58}$$

Proof. For $k \geqslant 0$, we define $\mathcal{F}_k$ as the $\sigma-$algebra generated by the random variables $\varepsilon_0, \cdots, \varepsilon_{k-1}$ and denote by $\mathbb{E}_k$ the conditional expectation over $\mathcal{F}_k$. By definition of our iterates, for any $k \geqslant 0$

$$\tilde{x}_{k+1} = \tilde{x}_k - \alpha_k\left(\nabla f_{\sigma_k}(\tilde{x}_k) - \tau\nabla\log p_{\sigma_k}(\tilde{x}_k + \nu_k\varepsilon_k)\right) + (1-\alpha_k)\nu_k\varepsilon_k, \tag{59}$$

so that $\tilde{x}_k$ is $\mathcal{F}_k-$measurable. We define the following functions:

$$F_k := F_{\sigma_k} : x \in \mathbb{R}^d \mapsto f_{\sigma_k}(x) - \tau \log p_{\sigma_k}(x)\,, \tag{60}$$

$$G_k : x \in \mathbb{R}^d \mapsto f_{\sigma_k}(x) - \tau \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathrm{Id})}\left[\log p_{\sigma_k}(x+\nu_k\varepsilon)\right]\,, \tag{61}$$

$$\bar{g}_k : x \in \mathbb{R}^d \mapsto \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathrm{Id})}\left[\log p_{\sigma_k}(x+\nu_k\varepsilon)\right]. \tag{62}$$

Properties of F_k were already given in Proposition 4, and $x^*_{\sigma_k}$ denotes again its single minimizer. As justified in the proof of Theorem 9, G_k and $\bar{g}_k$ are respectively twice and thrice continuously differentiable, and G_k inherits λ_k-strong convexity and L_k-smoothness from F_k, with $L_k = L_f + \lambda_k + \frac{\tau}{\sigma_k^2}$. We want to bound the sequence $\tilde{\beta}_k := \mathbb{E}\left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2$. To do so, we introduce $\beta_k := \mathbb{E}_{k-1}\left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2$. By definition,

$$\begin{aligned}
&\tilde{x}_{k+1} - x^*_{\sigma_k} \\
&= \tilde{x}_k - \alpha_k\left(\nabla f_{\sigma_k}(\tilde{x}_k) - \tau\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k)\right) + (1-\alpha_k)\nu_k\varepsilon_k - x^*_{\sigma_k} \\
&= \left(\tilde{x}_k - \alpha_k\nabla G_k(\tilde{x}_k) - x^*_{\sigma_k}\right) + \alpha_k\tau\left(\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\right) + (1-\alpha_k)\nu_k\varepsilon_k.
\end{aligned} \tag{63}$$

Taking the expectation over $\mathcal{F}_k$ of the squared norm, we obtain

$$\begin{aligned}
\beta_{k+1} = &\left\|\tilde{x}_k - \alpha_k\nabla G_k(\tilde{x}_k) - x^*_{\sigma_k}\right\|^2 + \alpha_k^2\tau^2\mathbb{E}_k\left\|\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\right\|^2 \\
&+ (1-\alpha_k)^2\nu_k^2 d + 2(1-\alpha_k)\alpha_k\nu_k\tau\mathbb{E}_k\langle\varepsilon_k, \nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\rangle.
\end{aligned} \tag{64}$$

We want to bound the sum on the right.

By adapting the calculations from (47) to (49) to a Gaussian smoothing with covariance $\nu_k^2\mathrm{Id}$ instead of $\sigma_k^2\mathrm{Id}$, and denoting $r_k := \left\|\tilde{x}_k - x^*_{\sigma_k}\right\|$ and $h_k := \frac{\alpha_k\lambda_k L_k}{\lambda_k + L_k}$, we have $B > 0$ such that:

$$\left\|\tilde{x}_k - \alpha_k\nabla G_k(\tilde{x}_k) - x^*_{\sigma_k}\right\|^2 \leqslant (1-h_k)\, r_k^2 + \left(3\alpha_k^2 + \frac{\alpha_k^2}{h_k}\right) B\nu_k^4. \tag{65}$$

Again, since $h_k = \frac{\alpha_k\lambda_k L_k}{\lambda_k+L_k} \geqslant \frac{\alpha_k\lambda_k}{2}$ and $1 > \alpha_k\lambda_k$, we have $(3 + \frac{1}{h_k}) \leqslant \frac{5}{\alpha_k\lambda_k}$, and

$$\left\|\tilde{x}_k - \alpha_k\nabla G_k(\tilde{x}_k) - x^*_{\sigma_k}\right\|^2 \leqslant (1-h_k)\, r_k^2 + \frac{5\alpha_k B\nu_k^4}{\lambda_k}. \tag{66}$$

Moreover, $\mathbb{E}_k\left\|\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\right\|^2$ corresponds to the conditional variance of $\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k)$ with respect to $\mathcal{F}_k$. By the Poincaré inequality (Brascamp & Lieb, 1976)

$$\mathbb{E}_k\left\|\nabla\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\right\|^2 \leqslant \nu_k^2\mathbb{E}_k\left\|\nabla^2\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k)\right\|_F^2. \tag{67}$$

As the proof of Pesme et al. (2025) shows that, for any $\sigma > 0$, $\left\|\nabla^3\log p_\sigma\right\|_F \leqslant M$,

$$\begin{aligned}
&\mathbb{E}_k\left\|\nabla^2\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k) - \nabla^2\log p_{\sigma_k}(x^*_{\sigma_k})\right\|_F^2 \\
&= \mathbb{E}_k\left\|\int_0^1 \nabla^3\log p_{\sigma_k}\left((1-s)x^*_{\sigma_k} + s(\tilde{x}_k+\nu_k\varepsilon_k)\right)\mathrm{d}s\,\left(\tilde{x}_k+\nu_k\varepsilon_k - x^*_{\sigma_k}\right)\right\|_F^2 \\
&\leqslant \mathbb{E}_k\int_0^1\left\|\nabla^3\log p_{\sigma_k}\left((1-s)x^*_{\sigma_k} + s(\tilde{x}_k+\nu_k\varepsilon_k)\right)\left(\tilde{x}_k+\nu_k\varepsilon_k - x^*_{\sigma_k}\right)\right\|_F^2\mathrm{d}s \\
&\leqslant M^2\mathbb{E}_k\left\|\tilde{x}_k+\nu_k\varepsilon_k - x^*_{\sigma_k}\right\|^2 \;=\; M^2(\nu_k^2 d + r_k^2).
\end{aligned} \tag{68}$$

Thus, by (B) (which is also valid for the Frobenius norm) with $\delta = 1$, we obtain

$$\nu_k^2\mathbb{E}_k\left\|\nabla^2\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k)\right\|_F^2 \leqslant 2\nu_k^2\left(M^2\nu_k^2 d + M^2 r_k^2) + \left\|\nabla^2\log p_\sigma(x^*_{\sigma_k})\right\|_F^2\right). \tag{69}$$

The sequence ν_k is clearly bounded, and by continuity of $(\sigma, x) \mapsto \nabla^2\log p_\sigma(x)$ and boundedness of $(x^*_\sigma)_{\sigma\in(0,\sigma_0]}$ (Proposition 5), $(\left\|\nabla^2\log p_\sigma(x^*_{\sigma_k})\right\|_F^2)_k$ is bounded. Therefore, there is $C > 0$ such that, for any $k \geqslant 0$,

$$\nu_k^2\mathbb{E}_k\left\|\nabla^2\log p_{\sigma_k}(\tilde{x}_k+\nu_k\varepsilon_k)\right\|_F^2 \leqslant \nu_k^2 C(1+r_k^2). \tag{70}$$

Then, combining Cauchy-Schwarz and the previous inequality,

$$\begin{aligned}
&|\mathbb{E}_k\langle\varepsilon_k, \nabla \log p_{\sigma_k}(\tilde{x}_k + \nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\rangle| \\
&\qquad \leqslant \sqrt{\mathbb{E}_k\left\|\varepsilon_k\right\|^2 \ \mathbb{E}_k\left\|\nabla \log p_{\sigma_k}(\tilde{x}_k + \nu_k\varepsilon_k) - \nabla\bar{g}_k(\tilde{x}_k)\right\|^2} \\
&\qquad \leqslant \nu_k\sqrt{dC(1+r_k^2)} \leqslant \nu_k\sqrt{dC}(1 + r_k^2/2).
\end{aligned} \tag{71}$$

Putting all together,

$$\begin{aligned}
\beta_{k+1} &\leqslant \left(1 - h_k + \alpha_k^2\nu_k^2\tau^2 C + (1-\alpha_k)\alpha_k\nu_k^2\tau\sqrt{dC}\right) r_k^2 \\
&\quad + \left(3\alpha_k^2 + \frac{\alpha_k^2}{h_k}\right) B\nu_k^4 + (1-\alpha_k)^2\nu_k^2 d + 2(1-\alpha_k)\nu_k^2\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\nu_k^2
\end{aligned} \tag{72}$$

Since $h_k \underset{k\to+\infty}{\sim} \alpha_k\lambda_k \underset{k\to+\infty}{\sim} \frac{\sigma_k^2\lambda_k}{\tau}$ and $\nu_k^2 = o(\sigma_k^2\lambda_k)$, up to considering a later iteration, we suppose that $\alpha_k^2\nu_k^2\tau C + (1-\alpha_k)\alpha_k\nu_k^2\tau\sqrt{dC} \leqslant \frac{h_k}{2}$, so that

$$\beta_{k+1} \leqslant \left(1 - \frac{h_k}{2}\right) r_k^2 + \frac{5\alpha_k B}{\lambda_k}\nu_k^4 + (1-\alpha_k)^2\nu_k^2 d + 2(1-\alpha_k)\nu_k^2\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\nu_k^2.$$

In order to obtain a recursive relation on $\tilde{\beta}_{k+1} = \mathbb{E}[\beta_{k+1}]$, we need to make $\left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2$ appear on the right. To do so, we use (B) on r_k^2 with $\delta = \dfrac{h_k}{4-2h_k}$, which gives,

$$\begin{aligned}
\beta_{k+1} &\leqslant \left(1 - \frac{h_k}{4}\right)\left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2 + \left(1 - \frac{h_k}{2}\right)\left(\frac{4}{h_k} - 1\right)\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|^2 \\
&\quad + \nu_k^2\left(\frac{5\alpha_k B}{\lambda_k}\nu_k^2 + (1-\alpha_k)^2 d + 2(1-\alpha_k)\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\right) \qquad (73) \\
&\leqslant \left(1 - \frac{h_k}{4}\right)\left\|\tilde{x}_k - x^*_{\sigma_{k-1}}\right\|^2 \\
&\quad + \underbrace{\frac{4}{h_k}\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|^2 + \nu_k^2\left(\frac{5\alpha_k B}{\lambda_k}\nu_k^2 + (1-\alpha_k)^2 d + 2(1-\alpha_k)\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\right)}_{:=\frac{h_k}{4}\delta_k}. \qquad (74)
\end{aligned}$$

(75)

Hence, taking the total expectation, this yields

$$\tilde{\beta}_{k+1} \leqslant \left(1 - \frac{h_k}{4}\right)\tilde{\beta}_k + \frac{h_k}{4}\delta_k \tag{76}$$

Applying Lemma 14, we obtain the convergence $\tilde{\beta}_k \underset{k\to+\infty}{\longrightarrow} \to 0$ under the following constraints:

$$\sum_{k=0}^{+\infty} h_k = +\infty, \quad \delta_k \underset{k\to+\infty}{\longrightarrow} 0 \tag{77}$$

Since $h_k \sim \dfrac{\sigma_k^2}{\tau}\lambda_k$, we recover condition *(i)* of Theorem 7 from $\sum_{k=0}^{+\infty} h_k = +\infty$. Then, by definition,

$$\delta_k = \frac{16}{h_k^2}\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|^2 + \frac{4}{h_k}\left(\nu_k^2\left(\frac{5\alpha_k B}{\lambda_k}\nu_k^2 + (1-\alpha_k)^2 d + 2(1-\alpha_k)\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\right)\right). \tag{78}$$

By the proof of Theorem 9, we know that $\frac{16}{h_k^2}\left\|x^*_{\sigma_k} - x^*_{\sigma_{k-1}}\right\|^2 \underset{k\to+\infty}{\longrightarrow} 0$ under (ii) and (iii). Moreover, since $\frac{\sigma_k^2}{\lambda_k} \underset{k\to+\infty}{\longrightarrow} 0$,

$$\frac{4\nu_k^2}{h_k}\left(\frac{5\alpha_k B}{\lambda_k}\nu_k^2 + (1-\alpha_k)^2 d + 2(1-\alpha_k)\alpha_k\tau\sqrt{dC} + C\alpha_k^2\tau^2\right) \underset{k\to+\infty}{\sim} \frac{4\tau\nu_k^2 d}{\sigma_k^2\lambda_k} \tag{79}$$

As $\nu_k^2 = o(\sigma_k^2\lambda_k)$ by assumption, we obtain the convergence to 0 of $\mathbb{E}\left\|\tilde{x}_{k+1} - x^*_{\sigma_k}\right\|$. Using Proposition 5 concludes the proof.

□

C Discussion on the methods

C.1 Rewriting PnP-Flow

The original PnP-Flow algorithm is defined as

$$\begin{cases} z_{k+1} = x_k - \alpha_k \nabla f(x_k)\,, \\ \tilde{z}_{k+1} = t_k z_{k+1} + (1-t_k)\varepsilon_k\,, \quad \varepsilon_k \sim \mathcal{N}(0, \mathrm{Id})\,, \\ x_{k+1} = \tilde{D}_{t_k}(\tilde{z}_{k+1})\,. \end{cases} \tag{80}$$

PnP-Flow uses the denoiser $\tilde{D}_t = \mathrm{Id} + (1-t)v_t$ with $v : [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$ a Flow Matching velocity field, defined as

$$\forall (t,x) \in [0,1]\times\mathbb{R}^d,\ v_t(x) = \mathbb{E}[X - Z | (1-t)Z + tX = x], \tag{81}$$

where $X \sim p$ and $Z \sim \mathcal{N}(0, \mathrm{Id})$, taken independent. One can easily show that $\tilde{D}_t(x) = \mathbb{E}[X|(1-t)Z + tX = x]$.

Defining $D_\sigma(x) = \mathbb{E}\left[X | X + \sigma Z = x\right] = \mathrm{MMSE}_\sigma(x)$, one has

$$\tilde{D}_t(tx) = \mathbb{E}[X|(1-t)Z + tX = tx] = \mathbb{E}\left[X | \tfrac{1-t}{t}Z + X = x\right] = D_{\sigma_t}(x)\,, \tag{82}$$

with $\sigma_t = \frac{1-t}{t}$. Since $\tilde{z}_{k+1} = t_k(z_{k+1} + \sigma_{t_k}\varepsilon_k)$, $\tilde{D}_{t_k}(\tilde{z}_{k+1}) = D_{\sigma_{t_k}}(z_{k+1} + \sigma_{t_k}\varepsilon_k)$ and the last two steps of (80) can be merged into a single noising/denoising step applied to z_{k+1}:

$$\begin{cases} z_{k+1} = x_k - \alpha_k \nabla f(x_k)\,, \\ x_{k+1} = D_{\sigma_{t_k}}(z_{k+1} + \sigma_{t_k}\varepsilon_k), \quad \varepsilon_k \sim \mathcal{N}(0, \mathrm{Id})\,. \end{cases} \tag{83}$$

Equivalence with MMSE Average: In the denoising case $f = \frac{1}{2}\|\cdot - y\|^2$ and based on the above discussion we have

$$z_{k+1} = x_k - \alpha_k(x_k - y) = (1-\alpha_k)\mathrm{MMSE}_{\sigma_k}(z_k + \sigma_k\varepsilon_k) + \alpha_k y \tag{84}$$

Hence, the z_k iterates of PnP-Flow in that case correspond to iterations of MMSE Average, with the modification that some noise in added to the input of the denoiser.

General case: Using a generic datafit f, PnP-Flow rewrites as two consecutive gradient steps, on f and $-\sigma_k^2 \log p_{\sigma_k}$ respectively:

$$\begin{cases} z_{k+1} = x_k - \alpha_k \nabla f(x_k) \\ x_{k+1} = z_k + \sigma_k\varepsilon_k + \sigma_k^2 \nabla \log p_\sigma(z_k + \sigma_k\varepsilon_k) \end{cases} \tag{85}$$

The gradients of f and $\log p_\sigma$ being evaluated at different points, we cannot hope to cast it as an instance of MMSE Averaging. Rather, the GAMMA we propose is similar in spirit, but writes as a single gradient step of f and $\log p_\sigma$ simultaneously, closer to RED in that sense (but with annealing).

D Experiments

Initialisation. When initializing GAMMA, we find that choosing $x_0 \approx \mathbb{E}_{X\sim p}[X]$, the mean natural image of the distribution, results in higher perceptual quality results. In practice, we approximate this mean image with $x_0 = \mathrm{MMSE}_{\sigma_{\text{sample}}}(\varepsilon)$ where $\varepsilon \sim \mathcal{N}(0, \mathrm{Id})$ and $\sigma_{\text{sample}} = 100$.

Models For the two datasets, Celeba-128 and AHFQ-256, we trained a Flow Matching model from scratch. We followed the architecture and training pipeline of Martin et al. (2025), with one modification: we used an independent coupling instead of the minibatch OT coupling employed in their work.

D.1 Optimal hyper-parameters values

Schedules. For GAMMA, we take the schedules from Remark 8, which are proven to make the scheme converge.

$$\lambda_k = \frac{\lambda_0}{(k+1)^\beta}, \quad \sigma_k^2 = \frac{\sigma_0^2}{(k+1)^\gamma}, \quad \alpha_k = \frac{\sigma_k^2}{\sigma_k^2 + \kappa\tau}, \tag{86}$$

with $0 < \beta < \gamma < 1$, $\beta + \gamma < 1$, $\lambda_0 \in (0, 1)$. We choose to tune the noise level $\kappa\tau := \kappa\sigma_y^2$ with a factor κ. We also tune the parameters λ_0, σ_0. For Renoising-GAMMA, we approximate $\mathbb{E}_\epsilon[\mathrm{MMSE}_\sigma]$ using $N_\epsilon = 3$ draws of ϵ per iteration, since Figure 3 shows no significant differences when using more draws. We set $N = 2000$,$\gamma = 0.98$, $\beta = 0.01$, for GAMMA (noiseless) and Renoised-GAMMA, the resulting noise schedules are displayed in Figure 4.

For PnP-Flow, we take the same hyperparameters values as the ones used in the original paper (Martin et al., 2025).

Table 2: Hyper-parameters used and GAMMA on CelebA and AFHQ-Cat datasets.

		Denoising	Deblurring	Super-res.	Rand. inpaint.	Box inpaint.
CelebA	**PnP-Flow**					
	N (number of steps)	100	100	100	100	100
	α (learning rate)	0.8	0.01	0.3	0.01	0.5
	Approx-PGD					
	γ (step size)	1.0	1.0	1.0	1.0	1.0
	κ (noise level factor)	1.0	0.1	0.5	2.0	0.5
$\gamma = 0.98$	**GAMMA**					
$\beta = 0.01$	σ_0	1.0	1	5	1	10
$N = 2000$	κ	2.0	0.1	0.5	50	20
$\lambda_0 = 10^{-3}$	**Renoised-GAMMA**					
	σ_0	1.0	10	5	5	10
	κ	2.0	5	5	50	20
AFHQ	**PnP-Flow**					
	N (number of steps)	100	100	500	200	100
	α (learning rate)	0.8	0.01	0.01	0.01	0.5
	Approx-PGD					
	γ (step size)	1.0	1.0	1.0	1.0	1.0
	κ (noise level factor)	1.0	2.0	1.0	5.0	2.0
$\gamma = 0.98$	**GAMMA**					
$\beta = 0.01$	λ_0 (number of steps)	10^{-3}	10^{-2}	10^{-3}	10^{-3}	10^{-3}
$N = 2000$	σ_0	5	10	10	1	5
	κ	1	1	20	50	50
	Renoised-GAMMA					
	λ_0 (number of steps)	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
	σ_0	1	10	5	5	10
	κ	1	10	5	50	20

D.1.1 Relaxed GAMMA schedules

Given a maximum number of iterations N, we consider the following family of schedules,

$$\sigma_k^2 = \left(\frac{N-k}{k}\right)^2, \quad \lambda_k = \lambda_0 \sigma_k^\beta, \quad \alpha_k = \frac{\sigma_k^2}{\kappa\sigma_k^2 + \tau} \tag{87}$$

where $\lambda_0 > 0$ and $\kappa \geqslant 1$.

Table 3: Hyperparameters used for GAMMA on the CelebA and AFHQ-Cat datasets.

		Denoising	Deblurring	Super-res.	Rand. inpaint.	Box inpaint.
CelebA	GAMMA					
	N (number of steps)	100	100	500	500	100
	κ	1.0	1.0	2.0	1.5	1.2
	λ_0	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
	β	1.0	1.0	1.0	1.0	1.0
AFHQ-Cat	GAMMA					
	N (number of steps)	100	100	500	500	100
	κ	1.0	1.0	2.0	1.5	1.2
	λ_0	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
	β	1.0	1.0	1.0	1.0	1.0

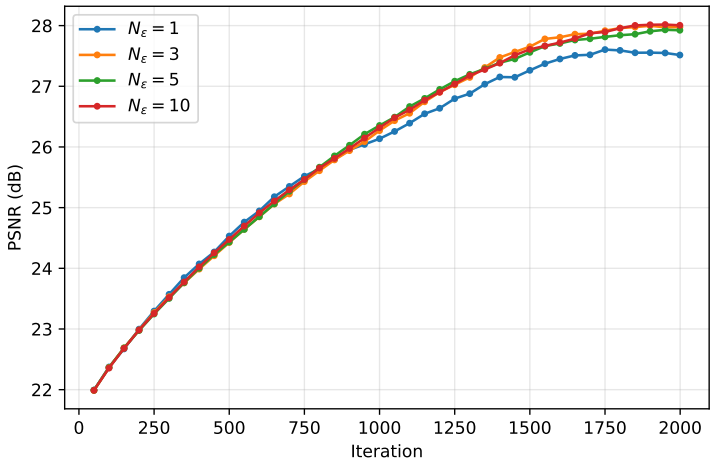


Figure 3: Ablation of the number N_ϵ of noise samples ϵ used to approximate $\mathbb{E}_\epsilon[\mathrm{MMSE}_\sigma(x + \sigma\epsilon)]$ - AFHQ - Box inpainting

Table 4: Comparisons of methods on different inverse problems on the AFHQ dataset. Results are averaged across 100 test images. Higher PSNR / SSIM is better, lower LPIPS is better.

Method	Theory	Denoising $\sigma = 0.2$			Deblurring $\sigma = 0.05, \sigma_b = 3.0$			Super-res. $\sigma = 0.05, \times 4$			Rand. inpaint. $\sigma = 0.01, 70\%$			Box inpaint. $\sigma = 0.05, 80 \times 80$		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Degraded		20.00	0.292	0.550	24.43	0.531	0.452	11.77	0.220	0.877	13.44	0.221	1.090	21.68	0.726	0.219
PnP-Flow	No	32.27	0.876	0.164	**29.48**	**0.795**	0.317	**29.02**	**0.811**	**0.171**	**34.97**	**0.939**	**0.037**	**28.63**	0.911	0.107
Approx-PGD	Yes	28.21	0.679	0.236	18.42	0.464	0.526	7.08	0.059	0.690	18.36	0.404	0.510	17.86	0.717	0.354
GAMMA (noiseless)	Yes	30.03	0.751	**0.102**	14.23	0.311	0.613	26.60	0.703	0.361	22.09	0.593	0.326	21.91	0.603	0.277
GAMMA (noise)	Yes	30.74	0.848	0.212	28.89	0.779	0.315	25.29	0.721	0.279	32.56	0.895	0.090	26.21	0.852	0.101
GAMMA (relaxed)	No	**32.46**	**0.889**	0.104	29.10	0.789	**0.292**	28.18	0.782	0.259	34.40	0.931	0.046	27.72	**0.919**	**0.069**

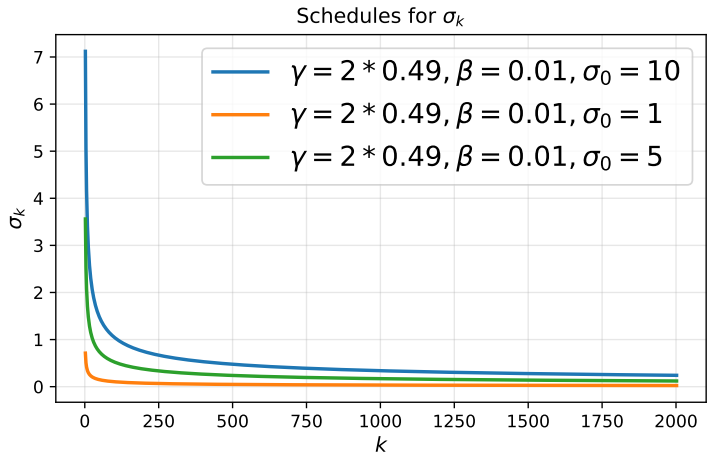


Figure 4: The two schedules $\sigma_k = \frac{\sigma_0}{(k+1)^{\gamma/2}}$ used in our experiments. Setting $N = 2000$ across all experiments ensures that σ_k is sufficiently close to 0.

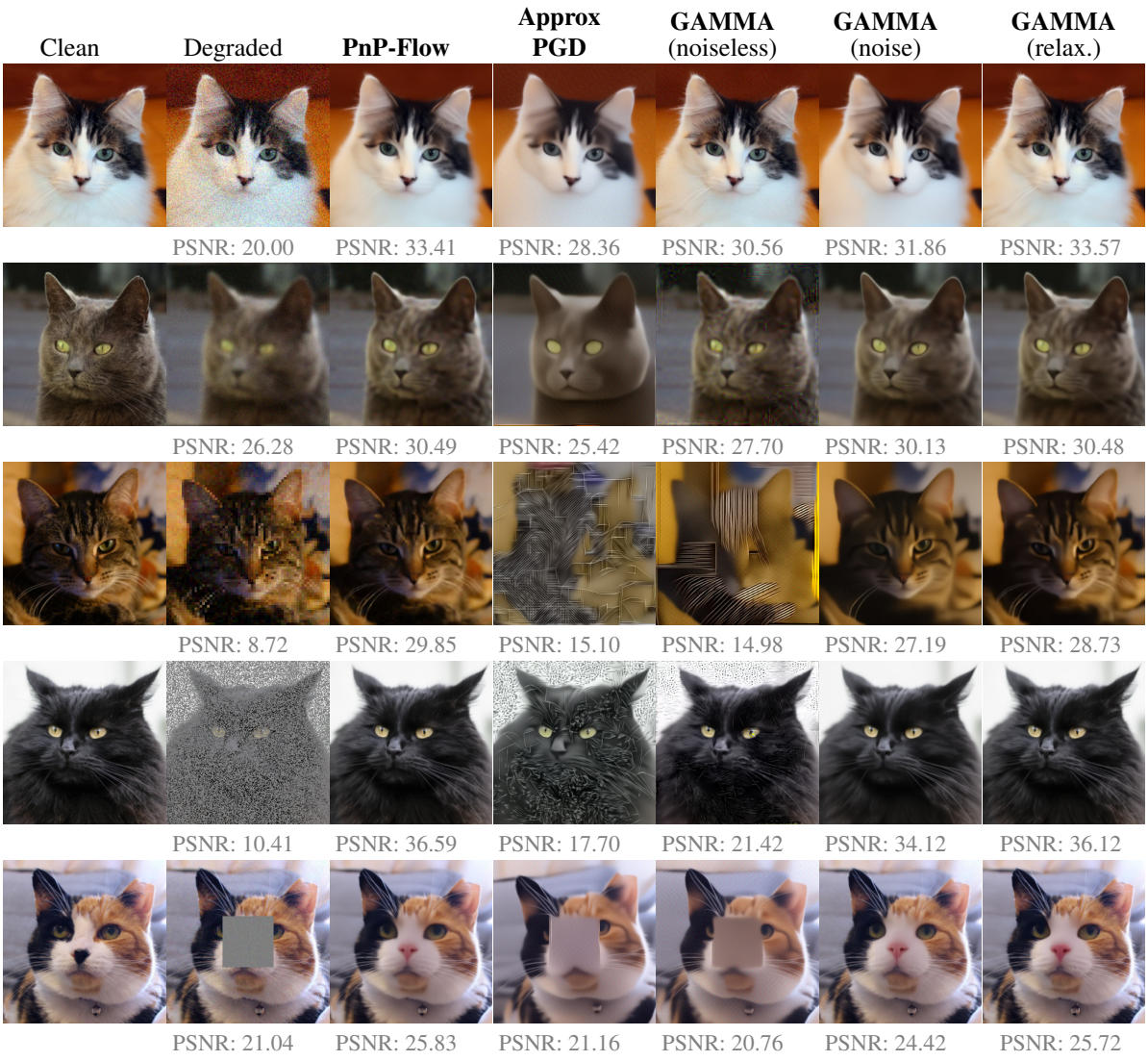


Figure 5: Qualitative results on AFHQ-256.