**Organisation of the Appendix.**

1. In Appendix A, we prove Proposition 5 and theorems 1 and 2, show tightness of Theorem 1 for Gaussian priors and extend Appendix A in the case of priors supported on a low-dimensional affine subspace. We also prove the tighness of theorem 1 in the case of Gaussian priors.
2. In Appendix B, we provide several lemmas which enable to control $\sigma \mapsto x^\star_\sigma$.

# A Proofs of Proposition 5 and theorems 1 and 2

## A.1 Preliminary results

We start by the following proposition establishing that $p_\sigma$ is log-concave and $\frac{1}{\sigma^2}$-smooth.

**Proposition 4.** *Fix $\sigma > 0$. Under Assumption 1, $x \mapsto -\ln p_\sigma(x)$ is convex with a Hessian satisfying:*

$$-\nabla^2 \ln p_\sigma(z) = \frac{1}{\sigma^2}\big[I_d - \frac{1}{\sigma}\mathrm{Var}(\varepsilon|X+\sigma\varepsilon = z)\big] \preceq \frac{1}{\sigma^2} I_d.$$

*Consequently, $F_\sigma$ is 1-strongly convex.*

*Proof.* The convexity of $x \mapsto -\ln p_\sigma(x)$ follows directly by the classical fact that a convolution of log-concave densitites with a Gaussian is still log-concave (see [Saumard and Wellner, 2014, Proposition 3.5]). The fact that the Hessian is upper-bounded by $\frac{1}{\sigma^2} I_d$ follows from a "second order Tweedie formula". The results is a direct consequence of an identity which can be seen as a "second order Tweedie formula" (e.g. Lemma A.1 in Gribonval [2011a] or in Lee and Vázquez [2003] equation 5.8.):

$$\begin{aligned}-\nabla^2 \ln p_\sigma(z) &= \frac{1}{\sigma^2}\big[I_d - \frac{1}{\sigma}\mathrm{Var}(\varepsilon|X+\sigma\varepsilon = z)\big]\\ &\preceq \frac{1}{\sigma^2} I_d,\end{aligned}$$

where $\varepsilon$ denotes a standard $d$-dimensional Gaussian random variable ($\varepsilon \sim \mathcal{N}(0, I_d)$) and the matrix inequality is due to the positiveness of the covariance matrix. For completeness we give the proof of the second order Tweedie identity. From the standard Tweedie identity (see, e.g. Efron [2011]) we have that:

$$\begin{aligned}-\nabla \ln p_\sigma(z) &= \frac{z - \mathbb{E}\big[X|X+\sigma\varepsilon = z\big]}{\sigma^2}\\ &= \frac{1}{\sigma^2}\int_{\mathbb{R}^d}(z-x)p(x|z)\mathrm{d}x\\ &= \frac{1}{\sigma^2}\int_{\mathbb{R}^d}(z-x)\frac{\phi_\sigma(z-x)p(x)}{\int_{\mathbb{R}^d}\phi_\sigma(z-x')p(x')\mathrm{d}x'}\mathrm{d}x,\end{aligned}$$

where $\phi_\sigma(z) = \exp(-\frac{z^2}{2\sigma^2})$. Notice that $\phi'_\sigma(z) = -\frac{z}{\sigma^2}\phi_\sigma(z)$. We can now compute the Hessian of $-\ln p_\sigma$, letting $X_\sigma = X + \sigma\varepsilon$:

$$\begin{aligned}-\nabla^2 \ln p_\sigma(z) &= \frac{1}{\sigma^2}\Big(I_d - \frac{1}{\sigma^2}\int_{\mathbb{R}^d}(z-x)^{\otimes 2}p(x|z)\mathrm{d}x + \frac{1}{\sigma^2}\big[\int_{\mathbb{R}^d}(z-x)p(x|z)\mathrm{d}x\big]^{\otimes 2}\Big)\\ &= \frac{1}{\sigma^2}\Big(I_d - \frac{1}{\sigma^2}\big(\mathbb{E}[(X_\sigma - X)^{\otimes 2}|X_\sigma = z] - \mathbb{E}[X_\sigma - X|X_\sigma = z]^{\otimes 2}]\big)\Big)\\ &= \frac{1}{\sigma^2}\Big(I_d - \frac{1}{\sigma^2}\mathrm{Var}(\varepsilon|X_\sigma = z)\Big),\end{aligned}$$

which concludes the proof. □

Now, we recall and prove Proposition 5, which is a direct consequence of proposition 4.

**Proposition 5.** *For any $\sigma > 0$, the function $F_\sigma$ is $L_\sigma$-smooth, with*

$$L_\sigma = 1 + \frac{\tau}{\sigma^2}.$$

*Proof.* The proof directly follows from proposition 4 which implies that $-\nabla \ln p_\sigma$ is $1/\sigma^2$-Lipschitz, so that $F_\sigma$ is $L_\sigma$-smooth with $L_\sigma = 1 + \frac{\tau}{\sigma^2}$. □

### A.2 Analysis of the MMSE Averaging iterates

We start by recalling our main result Theorem 1, which provides a convergence rate towards the proximal operator of the MMSE Averaging recursion.

**Theorem 1** (Convergence to the Proximal operator). *Let* $\mathrm{prox}_{-\tau \ln p}(y)$ *denote the unique solution of the Proximal Objective problem, then under Assumptions 1 and 2, we have that the MMSE Averaging recursion with* $\alpha_k = 1/(k+2)$, $\sigma_k^2 = \tau/(k+1)$ *and initialised at* $x_0 = y$ *satisfy:*

$$\|x_k - \mathrm{prox}_{-\tau \ln p}(y)\| \leq \frac{(\ln k) + 7}{k+1}\left[\|y - \mathrm{prox}_{-\tau \ln p}(y)\| + \tau^2 M\sqrt{d}\right].$$

*Proof.* From Proposition 3, we are guaranteed that $F_{\sigma_k}$ is strongly convex and smooth with

$$\mu_{\sigma_k} = 1, \quad L_{\sigma_k} = 1 + \frac{\tau}{\sigma_k^2} = k+2, \quad \kappa_{\sigma_k} = k+2.$$

To avoid heavy notations, we denote $x^\star_{\sigma_k} := \mathrm{prox}_{-\tau \ln p_{\sigma_k}}(y) = \arg\min F_{\sigma_k}$ as well as $x^\star := \mathrm{prox}_{-\tau \ln p}(y) = \arg\min F$, note that these quantities are well defined and unique by the strong convexity of $F_{\sigma_k}$ and $F$.

Recall that through 1, one step of the eq. (MMSE Averaging) recursion can be seen as one step of gradient descent on $F_{\sigma_k}$ with stepsize $\gamma_k = \frac{1}{k+2}$ which corresponds to $\gamma_k = 1/L_{\sigma_k}$. Hence, at iteration $k$, standard convex optimisation results (see Theorem 2.1.15 in Nesterov [2013]) guarantee the improvement:

$$\begin{aligned}
\|x_{k+1} - x^\star_{\sigma_k}\| &\leq \left(1 - 2\frac{\mu_{\sigma_k}}{\mu_{\sigma_k} + L_{\sigma_k}}\right)^{1/2} \|x_k - x^\star_{\sigma_k}\| \\
&= \left(\frac{\kappa_{\sigma_k} - 1}{\kappa_{\sigma_k} + 1}\right)^{1/2} \|x_k - x^\star_{\sigma_k}\| \\
&= \left(\frac{k+1}{k+3}\right)^{1/2} \|x_k - x^\star_{\sigma_k}\|
\end{aligned}$$

We now use the triangle inequality to write:

$$\|x_{k+1} - x^\star_{\sigma_k}\| \leq \left(\frac{k+1}{k+3}\right)^{1/2} \left(\|x_k - x^\star_{\sigma_{k-1}}\| + \|x^\star_{\sigma_{k-1}} - x^\star_{\sigma_k}\|\right). \tag{1}$$

And we clearly see that we need to be able to control the regularity of $\sigma \mapsto x^\star_\sigma$. This is done in Proposition 9, where we show that $x^\star_\sigma$ is Lipschitz in $\sigma^2$:

$$\|x^\star_{\sigma_1} - x^\star_{\sigma_2}\|_2 \leq C(\sigma_1^2 - \sigma_2^2),$$

for $\sigma_2 \leq \sigma_1 \leq \sqrt{\tau}$ and where $C := \frac{1}{\tau}\|x^\star - y\| + \tau M\sqrt{d}$. Since $\sigma_k \leq \sqrt{\tau}$, we can use this bound and insert it in Equation (1) to get:

$$\|x_{k+1} - x^\star_{\sigma_k}\| \leq \left(\frac{k+1}{k+3}\right)^{1/2} \left(\|x_k - x^\star_{\sigma_{k-1}}\| + (\sigma_{k-1}^2 - \sigma_k^2)C\right),$$

It remains to unroll the inequality, using the fact that $x_0 = y$ and that $\Pi_{i=j}^k \left(\frac{i+1}{i+3}\right) = \frac{(j+1)(j+2)}{(k+2)(k+3)}$:

$$\|x_{k+1} - x^\star_{\sigma_k}\| \leq \frac{\sqrt{2}}{\sqrt{(k+2)(k+3)}}\|y - x^\star_{\sigma_0}\| + \sum_{j=1}^k \sqrt{\frac{(j+1)(j+2)}{(k+2)(k+3)}}(\sigma_{j-1}^2 - \sigma_j^2)C.$$

Now since $\sigma_k^2 = \frac{\tau}{k+1}$, we have that $(\sigma_{j-1}^2 - \sigma_j^2) = \frac{\tau}{j(j+1)}$, hence for $k \geq 1$:

$$\begin{aligned}
\|x_{k+1} - x^\star_{\sigma_k}\| &\leq \frac{\sqrt{2}}{\sqrt{(k+2)(k+3)}}\|y - x^\star_{\sigma_0}\| + \sum_{j=1}^k \sqrt{\frac{(j+1)(j+2)}{(k+2)(k+3)}} \frac{\tau C}{j(j+1)} \\
&\leq \frac{\sqrt{2}}{k+2}\|y - x^\star_{\sigma_0}\| + \frac{\tau C}{k+2}\sum_{j=1}^k \frac{j+2}{j(j+1)}
\end{aligned}$$

And we can simply bound:

$$\sum_{j=1}^{k} \frac{j+2}{j(j+1)} = \sum_{j=1}^{k} \Big(\frac{1}{j} + \frac{1}{j} - \frac{1}{j+1}\Big) \leq 1 + \sum_{j=1}^{k} \frac{1}{j} \leq 2 + \ln(k),$$

Therefore

$$\|x_{k+1} - x^\star_{\sigma_k}\| \leq \frac{\sqrt{2}}{k+2}\|y - x^\star_{\sigma_0}\| + \frac{(2+\ln(k))\tau C}{k+2}.$$

Now using the triangular inequality $\|x_{k+1} - x^\star\| \leq \|x_{k+1} - x^\star_{\sigma_k}\| + \|x^\star_{\sigma_k} - x^\star\|$ and using Proposition 9 which bounds $\|x^\star_{\sigma_k} - x^\star\| \leq \sigma_k^2 C$ we get that:

$$\|x_{k+1} - x^\star\| \leq \frac{\sqrt{2}}{k+2}\|y - x^\star_{\sigma_0}\| + \frac{(2+\ln(k))\tau C}{k+2} + \frac{\tau C}{k+1}.$$

And using the triangular inequality again:

$$\begin{aligned}\|y - x^\star_{\sigma_0}\| &\leq \|y - x^\star\| + \|x^\star - x^\star_{\sigma_0}\| \\ &\leq \|y - x^\star\| + \sigma_0^2 C \\ &= \|y - x^\star\| + \tau C,\end{aligned}$$

where the second inequality is due to Proposition 9. Therefore:

$$\begin{aligned}\|x_{k+1} - x^\star\| &\leq \frac{\sqrt{2}\|y - x^\star\|}{k+2} + \frac{(\ln k) + 2 + \sqrt{2}}{k+1}\tau C, \\ &\leq \frac{\sqrt{2}\|y - x^\star\|}{k+1} + \frac{(\ln k) + 4}{k+1}\tau C.\end{aligned}$$

Plugging the definition of $C = \frac{1}{\tau}\|x^\star - y\| + \tau M\sqrt{d}$ we can finally write:

$$\|x_{k+1} - x^\star\| \leq \frac{(\ln k) + 7}{k+1}\Big(\|x^\star - y\| + \tau^2 M\sqrt{d}\Big),$$

which concludes the proof. □

This next proposition proves the tightness of theorem 1 (up to constants and the log-term) in the case of Gaussian prior. Here we assume that $p$ is the density of a $d$-dimensional Gaussian $\mathcal{N}(\mu, \Sigma)$, with $\Sigma$ a positive definite matrix. Without loss of generality, we can assume that the Gaussian is centered: i.e., $\mu = 0$.

**Proposition 6** (Exact convergence rate for Gaussian priors.)**.** *Under the assumption that the prior $p$ is a $d$-dimensional centered Gaussian $\mathcal{N}(0, \Sigma)$, then the we have that the MMSE Averaging recursion with $\alpha_k = 1/(k+2)$, $\sigma_k^2 = \tau/(k+1)$ initialised at $x_0 = y$ satisfies:*

$$x_k - \operatorname{prox}_{-\tau \ln p}(y) = \frac{y - \operatorname{prox}_{-\tau \ln p}(y)}{k+1}.$$

*Proof.* In this setting, the negative log-prior $-\ln p$ is a quadratic with Hessian $H = \Sigma^{-1}$, and $F$ is a quadratic:

$$F(x) = \frac{1}{2}\|y - x\|^2 + \frac{\tau}{2}x^\top \Sigma^{-1} x.$$

Its minimiser is given by:

$$x^\star := \operatorname{prox}_{-\tau \ln p}(y) = (I + \tau\Sigma^{-1})^{-1} y.$$

And since $p_\sigma \sim \mathcal{N}(0, \Sigma + \sigma^2 I_d)$, the smoothed objective writes:

$$F_{\sigma_k}(x) = \frac{1}{2}\|y - x\|^2 + \frac{\tau}{2}x^\top(\Sigma + \sigma_k^2 I_d)^{-1} x,$$

and its gradient is:

$$\nabla F_{\sigma_k}(x) = x - y + \tau(\Sigma + \sigma_k^2 I_d)^{-1} x.$$

We now prove the result by induction. For $k=0$, we have $x_0 = y$ and the base case trivially holds.

**Inductive step:** The inductive hypothesis provides that:

$$x_k = x^\star + \frac{1}{k+1}(y - x^\star).$$

Using the identity $x^\star = (I + \tau\Sigma^{-1})^{-1}y$, we have:

$$y - x^\star = \tau\Sigma^{-1}x^\star \quad \Rightarrow \quad x_k = x^\star + \frac{\tau}{k+1}\Sigma^{-1}x^\star.$$

Then,

$$(\Sigma + \sigma_k^2 I_d)^{-1}x_k = (\Sigma + \tfrac{\tau}{k+1}I_d)^{-1}\left(I + \tfrac{\tau}{k+1}\Sigma^{-1}\right)x^\star = \Sigma^{-1}x^\star = \frac{y - x^\star}{\tau},$$

so that:

$$\nabla F_{\sigma_k}(x_k) = x_k - y + (y - x^\star) = x^\star - y + \frac{1}{k+1}(y - x^\star) + (y - x^\star) = \frac{y - x^\star}{k+1}$$

Now from Proposition 1, the update writes:

$$\begin{aligned} x_{k+1} &= x_k - \frac{1}{k+2}\nabla F_{\sigma_k}(x_k) \\ &= x^\star + \frac{y - x^\star}{k+1} - \frac{y - x^\star}{(k+1)(k+2)} \\ &= x^\star + \frac{(y - x^\star)}{k+2}. \end{aligned}$$

This completes the inductive step, and hence the proof. □

### A.3 Extension of distributions supported on affine subspaces of $\mathbb{R}^d$

In this subsection we prove that Theorem 1 can naturally be extended to the case where the prior distribution is on an affine subspace of dimension $r \ll d$, in which case the dimension $d$ which appears in the upperbound reduces to the effective dimension $r$. Formally, we assume that the clean images $x$ are drawn from a probability distribution $\mu$ on $\mathbb{R}^d$ satisfying the following:

**Assumption 4.** *There exists an affine subspace $S \subset \mathbb{R}^d$ of dimension $r$ such that:*

- *$\mu$ is supported on $S$: $\mu(\mathbb{R}^d \setminus S) = 0$. Moreover, the restriction of $\mu$ to $S$ admits a density $p : S \to \mathbb{R}_+$ with respect to the $r$-dimensional Lebesgue measure on $S$. By abuse of notation, we extend $p$ to $\mathbb{R}^d$ by setting $p(x) = 0$ for $x \in \mathbb{R}^d \setminus S$.*
- *$p(x) > 0$ for all $x \in S$.*
- *$p$ is log-concave.*

Let $\phi_\sigma(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$ denote the Gaussian kernel on $\mathbb{R}^d$ of variance $\sigma^2$, now let $C_\sigma \coloneqq (2\pi\sigma^2)^{1/2}$ such that $\int_{\mathbb{R}^d} \phi_\sigma(x) = C_\sigma^d$. The smoothed density function $p_\sigma : \mathbb{R}^d \to \mathbb{R}_+$ then writes, for all $z \in \mathbb{R}^d$:

$$\begin{aligned} p_\sigma(z) &= \frac{1}{C_\sigma^d}\int_{\mathbb{R}^d} \phi_\sigma(z - x)\,\mathrm{d}\mu(x) \\ &= \frac{1}{C_\sigma^d}\int_S p(x)\,\phi_\sigma(z - x)\,\mathrm{d}x. \end{aligned}$$

For $z \in \mathbb{R}^d$, let $z_\perp$ denote the orthogonal projection of $z$ on $S$. Using orthogonality, notice that:

$$p_\sigma(z) = \frac{\phi_\sigma(z - z_\perp)}{C_\sigma^{d-r}} \cdot \frac{1}{C_\sigma^r}\int_S p(x)\phi_\sigma(z_\perp - x)\mathrm{d}x.$$

Therefore, for $z \in S$, letting $\tilde{p}_\sigma(z) \coloneqq \frac{1}{C_\sigma^r}\int_S p(x)\phi_\sigma(z - x)\mathrm{d}x$ denote the convolution of $p$ with the Gaussian kernel over $S$, we get that

$$-\ln p_\sigma(z) = \frac{\|z - z_\perp\|_2^2}{2\sigma^2} - \ln \tilde{p}_\sigma(z_\perp) + (d - r)\ln C_\sigma.$$

And importantly:

$$-\nabla\Delta \ln p_\sigma(z) = -\nabla_S \Delta_S \ln \tilde{p}_\sigma(z_\perp),$$

where the $\nabla_S$ and $\Delta_S$ denote the intrinsic gradients and Laplacians on $S$.

Therefore from Lemma 5 we can upperbound:

$$\begin{aligned}\sup_{z\in\mathbb{R}^d} \|\nabla\Delta \ln p_\sigma(z)\| &= \sup_{z_\perp\in S} \|\nabla_S\Delta_S \ln \tilde{p}_\sigma(z_\perp)\| \\ &\leq \sqrt{r}\sup_{z_\perp\in S} \|\nabla_S^3 \ln p(z_\perp)\|,\end{aligned}$$

where the inequality is a direct consequence of Lemma 5 but applied to $p$ considered over $S$. From here, all the proof of theorem 1 holds but with the effective dimension $r$ appearing instead of $d$.

### A.4 Analysis of approximate PGD Algorithm 1

We now restate and prove convergence towards the MAP esimator of the approximate PGD algorithm. The following is a restatement of theorem 2 with the explicit constants.

**Theorem 3** (Convergence towards the MAP estimator with explicit bounds)**.** *For $\tau \leq \frac{1}{\lambda L_f}$ and a number of steps in the inner loop which increases as $n_k = \lfloor c\cdot k^{1+\eta}\rfloor$ for $c, \eta > 0$, the approximate proximal gradient descent iterates $(\hat{x}_k)_k$ from Algorithm 1 satisfy:*

$$\frac{1}{k}\sum_{i=1}^{k} J(x_i) - J^\star \leq \frac{1}{2\tau k}\Big(\|y - x^\star_{\mathrm{MAP}}\|^2 + \sum_{i=1}^{\infty}\|\varepsilon_i\|^2 + 2R_{\eta,c}\sum_{i=1}^{\infty}\|\varepsilon_i\|\Big)$$

$$\|\hat{x}_k - x_k\| \leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c\cdot k^{1+\eta}}\cdot R_{\eta,c},$$

*where $x_k := \mathrm{prox}_{-\tau\ln p}(\hat{x}_{k-1} - \tau\lambda\nabla f(\hat{x}_{k-1}))$ corresponds to the true proximal mapping at each iteration, and where the quantities $R_{\eta,c}$, $\sum_{i=1}^{\infty}\|\varepsilon_i\|$ and $\sum_{i=1}^{\infty}\|\varepsilon_i\|^2$ are given in lemma 1.*

*For, e.g., $\eta = 1$ and $c = 10$, the bounds become:*

$$\frac{1}{k}\sum_{i=1}^{k} J(x_i) - J^\star \lesssim \frac{1}{\tau k}\Big(300\cdot\|y - x^\star_{\mathrm{MAP}}\|^2 + 600\cdot\big(\tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\big)\Big)$$

$$\|\hat{x}_k - x_k\| \lesssim \frac{2\ln(k)+10}{k^2}\cdot\Big(6\cdot\|y - x^\star_{\mathrm{MAP}}\|^2 + 12\cdot\big(\tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\big)\Big).$$

*Proof.* For $\tau \leq \frac{1}{\lambda L_f}$, the classic inequality after one step of the true proximal descent $x_{k+1} := \mathrm{prox}_{-\tau\ln p}(\hat{x}_k - \tau\lambda\nabla f(\hat{x}_k))$ provides that (see, e.g. Beck and Teboulle [2009]):

$$J(x_{k+1}) - J^\star \leq \frac{1}{2\tau}(\|\hat{x}_k - x^\star_{\mathrm{MAP}}\|^2 - \|x_{k+1} - x^\star_{\mathrm{MAP}}\|^2). \tag{2}$$

Now for $k \geq 1$, let $\varepsilon_k := \hat{x}_k - x_k$ correspond to approximation error which can be quantified using Theorem 1. For $k \geq 1$, Equation (2) can be expanded as:

$$\begin{aligned}J(x_{k+1}) - J^\star &\leq \frac{1}{2\tau}\Big(\|x_k - x^\star_{\mathrm{MAP}}\|^2 - \|x_{k+1} - x^\star_{\mathrm{MAP}}\|^2 + \|\hat{x}_k - x_k\|^2 + 2\langle\hat{x}_k - x_k, x_k - x^\star_{\mathrm{MAP}}\rangle\Big) \\ &\leq \frac{1}{2\tau}\Big(\|x_k - x^\star_{\mathrm{MAP}}\|^2 - \|x_{k+1} - x^\star_{\mathrm{MAP}}\|^2 + \|\varepsilon_k\|^2 + 2\|\varepsilon_k\|\cdot\|x_k - x^\star_{\mathrm{MAP}}\|\Big) \\ &\leq \frac{1}{2\tau}\Big(\|x_k - x^\star_{\mathrm{MAP}}\|^2 - \|x_{k+1} - x^\star_{\mathrm{MAP}}\|^2 + \|\varepsilon_k\|^2 + 2R_{\eta,c}\|\varepsilon_k\|\Big),\end{aligned}$$

where the second inequality is due to the Cauchy-Schwarz inequality, and the bound $\|x_k - x^\star_{\mathrm{MAP}}\| \leq R_{\eta,c}$ is due to Lemma 1. It remains to sum this inequality from $i = 1$ to $k-1$ and add inequality 2 with $k = 0$, we get:

$$\begin{aligned}\sum_{i=1}^{k}(J(x_i) - J^\star) &\leq \frac{1}{2\tau}\Big(\|\hat{x}_0 - x^\star_{\mathrm{MAP}}\|^2 - \|x_k - x^\star_{\mathrm{MAP}}\|^2 + \sum_{i=1}^{k-1}\|\varepsilon_i\|^2 + 2R_{\eta,c}\sum_{i=1}^{k-1}\|\varepsilon_i\|\Big) \\ &\leq \frac{1}{2\tau}\Big(\|y - x^\star_{\mathrm{MAP}}\|^2 + \sum_{i=1}^{\infty}\|\varepsilon_i\|^2 + 2R_{\eta,c}\sum_{i=1}^{\infty}\|\varepsilon_i\|\Big)\end{aligned}$$

where the second inequality is due to Lemma 1. Diving by $k$ leads to the first result. The second comes from the fact that $\|\varepsilon_k\| = \|\hat{x}_k - x_k\|$ for which the upper bound is given in Lemma 1. $\square$

Let $\varepsilon_k := \hat{x}_k - x_k$ denote the approximation error at iteration $k$, where $x_k := \text{prox}_{-\tau \ln p}(\hat{x}_{k-1} - \tau\lambda\nabla f(\hat{x}_{k-1}))$ is the true proximal point. The following lemma provides a bound on this approximation error at each step, along with bounds on other useful quantities.

**Lemma 1.** *The approximate proximal gradient descent algorithm is initialised at $\hat{x}_0 = y$, with step size $\tau \leq \frac{1}{\lambda L_f}$. If the number of inner iterations at step $k$ is set to $n_k = \lfloor c \cdot k^{1+\eta} \rfloor$ for all $k \geq 1$, then the following holds:*

$$\|x_k - x^\star_{MAP}\| \leq R_{\eta,c}, \qquad \|\varepsilon_k\| \leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}} \cdot R_{\eta,c},$$
$$\sum_{k=1}^{\infty} \|\varepsilon_k\| \leq S_{\eta,c} \cdot R_{\eta,c}, \quad \sum_{k=1}^{\infty} \|\varepsilon_k\|^2 \leq T_{\eta,c} \cdot R^2_{\eta,c}.$$

*where*

$$\begin{aligned}
R_{\eta,c} &:= B_{\eta,c} + \tau\lambda\|\nabla f(x^\star_{\text{MAP}})\| + \tau^2 M\sqrt{d} \\
B_{\eta,\sigma} &:= \exp(2S_{\eta,c})\Big[\|y - x^\star_{\text{MAP}}\| + S_{\eta,c} \cdot \big(\tau\lambda\|\nabla f(x^\star_{\text{MAP}})\| + \tau^2 M\sqrt{d}\big)\Big] \\
S_{\eta,c} &:= \frac{1+\eta}{c\eta^2}\big(1 + \eta \cdot (\ln(c) + 7)\big) \\
T_{\eta,c} &:= \frac{4(1+\eta)^2}{c^2(2\eta+1)^3} + \frac{2(\ln(c)+7)^2}{c^2}\left(1 + \frac{1}{2\eta+1}\right)
\end{aligned}$$

*For, e.g., $\eta = 1$, $c = 10$, these quantities simply become:*

$$\begin{aligned}
&R_{\eta,c} \approx B_{\eta,\sigma} \approx 60 \cdot \|y - x^\star_{\text{MAP}}\| + 120 \cdot \big(\tau\lambda\|\nabla f(x^\star_{\text{MAP}})\| + \tau^2 M\sqrt{d}\big) \\
&S_{\eta,c} \approx T_{\eta,c} \approx 2
\end{aligned}$$

*Proof.* Rewriting Equation (2), for $k \geq 1$ we get that:

$$\begin{aligned}
\|x_k - x^\star_{\text{MAP}}\| &\leq \|\hat{x}_{k-1} - x^\star_{\text{MAP}}\| \\
&\leq \|\hat{x}_{k-1} - x_{k-1}\| + \|x_{k-1} - x^\star_{\text{MAP}}\| \\
&= \|\varepsilon_{k-1}\| + \|x_{k-1} - x^\star_{\text{MAP}}\|.
\end{aligned} \tag{3}$$

Furthermore, from Theorem 1, since $c \cdot k^{1+\eta} - 1 \leq n_k \leq c \cdot k^{1+\eta}$, we get for $k \geq 1$:

$$\begin{aligned}
\|\varepsilon_k\| := \|\hat{x}_k - x_k\| &\leq \frac{(\ln n_k) + 7}{n_k + 1}\big[\|\hat{x}_{k-1} - \tau\lambda\nabla f(\hat{x}_{k-1}) - x_k\| + \tau^2 M\sqrt{d}\big] \\
&\leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}}\big[\|x_k - (I_d - \tau\lambda\nabla f)(\hat{x}_{k-1})\| + \tau^2 M\sqrt{d}\big].
\end{aligned} \tag{4}$$

Now, we use the triangle inequality to write:

$$\begin{aligned}
&\|x_k - (I_d - \tau\lambda\nabla f)(\hat{x}_{k-1})\| \\
&\quad\leq \|x_k - x^\star_{\text{MAP}}\| + \|x^\star_{\text{MAP}} - (I_d - \tau\lambda\nabla f)(x^\star_{\text{MAP}})\| \\
&\quad\quad + \|(I_d - \tau\lambda f)(x^\star_{\text{MAP}}) - (I_d - \tau\lambda f)(\hat{x}_{k-1})\|
\end{aligned} \tag{5}$$

Now, since $x^\star_{\text{MAP}}$ satisfies the fixed point property $x^\star_{\text{MAP}} = \text{prox}_{-\tau \ln p}((I_d - \tau\lambda\nabla f)(x^\star_{\text{MAP}}))$, and from the definition of $x_k$, we can write:

$$\begin{aligned}
\|x_k - x^\star_{\text{MAP}}\| &= \|\text{prox}_{-\tau \ln p}\big((I_d - \tau\lambda\nabla f)(\hat{x}_{k-1})\big) - \text{prox}_{-\tau \ln p}\big((I_d - \tau\lambda\nabla f)(x^\star_{\text{MAP}})\big)\| \\
&\leq \|(I_d - \tau\lambda\nabla f)(\hat{x}_{k-1}) - (I_d - \tau\lambda\nabla f)(x^\star_{\text{MAP}})\|,
\end{aligned}$$

where the inequality is due to the non-expansiveness of the proximal operator. Inequality 5 then becomes

$$\begin{aligned}
\|x_k - (I_d - \tau\lambda\nabla f)(\hat{x}_{k-1})\| &\leq 2\|(I_d - \tau\lambda f)(x^\star_{\text{MAP}}) - (I_d - \tau\lambda f)(\hat{x}_{k-1})\| + \tau\lambda\|\nabla f(x^\star_{\text{MAP}})\| \\
&\leq 2\|x^\star_{\text{MAP}} - \hat{x}_{k-1}\| + \tau\lambda\|\nabla f(x^\star_{\text{MAP}})\|.
\end{aligned}$$

where the second inequality is because $I_d - \tau\lambda f$ is Lipschitz for $\tau \leq 1/(\lambda L_f)$. Therefore, injecting this bound in inequality 4, we get for $k \geq 1$:

$$\begin{aligned} \|\varepsilon_k\| &\leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}} \left[2\|\hat{x}_{k-1} - x^\star_{\mathrm{MAP}}\| + \tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\right] \\ &\leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}} \left[2\|\varepsilon_{k-1}\| + 2\|x_{k-1} - x^\star_{\mathrm{MAP}}\| + \tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\right]. \end{aligned} \tag{6}$$

where the second inequality still holds for $k = 1$ with the convention $\varepsilon_0 = 0$ and $x_0 = \hat{x}_0 = y$. Now adding the inequality $\|x_k - x^\star_{\mathrm{MAP}}\| \leq \|\varepsilon_{k-1}\| + \|x_{k-1} - x^\star_{\mathrm{MAP}}\|$ from 3 to inequality 6, and letting $w_k := \|\varepsilon_k\| + \|x_k - x^\star_{\mathrm{MAP}}\|$ for $k \geq 0$, we get the following recursive inequality for $k \geq 1$:

$$w_k \leq (1 + 2C_k)w_{k-1} + C_k A,$$

where

$$C_k := \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}}, \quad A := \tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}, \quad w_0 = \|y - x^\star_{\mathrm{MAP}}\|.$$

It now remains to unroll the recursive inequality on $w_k$, which is done in the auxiliary Lemma 2 to obtain:

$$w_k \leq \exp(2S_{\eta,c})\left(w_0 + AS_{\eta,c}\right),$$

where

$$S_{\eta,c} := \frac{1+\eta}{c\eta^2}\big(1 + \eta \cdot (\ln(c) + 7)\big),$$

Putting things together we get the following uniform bound on $w_k$:

$$w_k \leq B_{\eta,\sigma} := \exp(2S_{\eta,c})\Big[\|y - x^\star_{\mathrm{MAP}}\| + S_{\eta,c} \cdot \big(\tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\big)\Big]$$

From the definition of $w_k = \|\varepsilon_k\| + \|x_k - x^\star_{\mathrm{MAP}}\|$, we trivially get that $\|x_k - x^\star_{\mathrm{MAP}}\| \leq B_{\eta,c}$, and now from eq. (6) we get, for $k \geq 1$:

$$\|\varepsilon_k\| \leq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}}\left[2B_{\eta,c} + \tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d}\right].$$

Letting $R_{\eta,c} := 2B_{\eta,c} + \tau\lambda\|\nabla f(x^\star_{\mathrm{MAP}})\| + \tau^2 M\sqrt{d} \geq B_{\eta,c}$ leads to the two first inequalities of the statement.

Now to bound $\sum_{k=1}^{\infty} \|\varepsilon_k\|$ we simply reuse the bound obtained on $\sum_j C_j \leq S_{\eta,c}$ in the proof of lemma 2 to obtain:

$$\sum_{k=1}^{\infty} \|\varepsilon_k\| \leq S_{\eta,c} \cdot R_{\eta,c}.$$

Finally for $\sum_{k=1}^{\infty} \|\varepsilon_k\|^2$ we upperbound:

$$\sum_{k=1}^{\infty}\left(\frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}}\right)^2 \leq \frac{2(1+\eta)^2}{c^2}\sum_{k=1}^{\infty}\frac{\ln^2(k)}{k^{2(1+\eta)}} + \frac{2(\ln(c)+7)^2}{c^2}\sum_{k=1}^{\infty}\frac{1}{k^{2(1+\eta)}}.$$

We now bound the two series using integrals:

$$\begin{aligned} \sum_{k=1}^{\infty}\frac{\ln^2(k)}{k^{2(1+\eta)}} &\leq \int_1^{\infty}\frac{\ln^2(x)}{x^{2(1+\eta)}}\,dx = \frac{2}{(2\eta+1)^3}, \\ \sum_{k=1}^{\infty}\frac{1}{k^{2(1+\eta)}} &\leq 1 + \int_1^{\infty}\frac{1}{x^{2(1+\eta)}}\,dx = 1 + \frac{1}{2\eta+1}. \end{aligned}$$

Putting everything together, we obtain the bound:

$$\sum_{k=1}^{\infty}\|\varepsilon_k\|^2 \leq \left(\frac{4(1+\eta)^2}{c^2(2\eta+1)^3} + \frac{2(\ln(c)+7)^2}{c^2}\left(1 + \frac{1}{2\eta+1}\right)\right)R^2_{\eta,c}.$$

$\square$

**Lemma 2.** *The recursive inequality*

$$w_k \leq (1 + 2C_k)w_{k-1} + C_k A, \quad \textit{where} \quad C_k \coloneqq \frac{(1+\eta)\ln(k) + \ln(c) + 7}{c \cdot k^{1+\eta}}$$

*unrolls as:*

$$w_k \leq \exp(2S_{\eta,c})\left(w_0 + AS_{\eta,c}\right),$$

*where*

$$S_{\eta,c} \coloneqq \frac{1+\eta}{c\eta^2}\big(1 + \eta \cdot (\ln(c) + 7)\big).$$

*Proof.* We iteratively apply the inequality to obtain:

$$w_k \leq w_0 \prod_{j=1}^{k}(1 + 2C_j) + A\sum_{i=1}^{k} C_i \prod_{j=i+1}^{k}(1 + 2C_j),$$

with the convention that empty products are equal to $1$.

We now bound the product $\prod_{j=1}^{k}(1 + 2C_j)$. Using the inequality $\log(1 + x) \leq x$, we get:

$$\log \prod_{j=1}^{k}(1 + 2C_j) = \sum_{j=1}^{k} \log(1 + 2C_j) \leq \sum_{j=1}^{k} 2C_j,$$

hence,

$$\prod_{j=1}^{k}(1 + 2C_j) \leq \exp\left(2\sum_{j=1}^{k} C_j\right).$$

To bound the sum $\sum_{j=1}^{\infty} C_j$, we split the numerator:

$$\sum_{j=1}^{\infty} C_j = \frac{1+\eta}{c}\sum_{j=1}^{\infty}\frac{\ln j}{j^{1+\eta}} + \frac{\ln(c) + 7}{c}\sum_{j=1}^{\infty}\frac{1}{j^{1+\eta}}.$$

We use the known bounds:

$$\sum_{j=1}^{\infty}\frac{1}{j^{1+\eta}} \leq 1 + \int_1^{\infty}\frac{1}{t^{1+\eta}}\mathrm{d}t = 1 + \frac{1}{\eta}, \quad \sum_{j=2}^{\infty}\frac{\ln j}{j^{1+\eta}} \leq \int_1^{\infty}\frac{\ln t}{t^{1+\eta}}dt = \frac{1}{\eta^2},$$

which gives:

$$\begin{aligned}\sum_{j=1}^{\infty} C_j &\leq \frac{1+\eta}{c\eta^2} + \frac{(\ln(c) + 7)}{c}\Big(1 + \frac{1}{\eta}\Big)\\ &= \frac{1+\eta}{c\eta^2}\big(1 + \eta \cdot (\ln(c) + 7)\big) \eqqcolon S_{\eta,c}.\end{aligned}$$

Then we have:

$$\prod_{j=1}^{k}(1 + 2C_j) \leq \exp\left(2S_{\eta,c}\right), \quad \sum_{i=1}^{k} C_i \prod_{j=i+1}^{k}(1 + 2C_j) \leq S_{\eta,c}\exp(2S_{\eta,c}).$$

Plugging these into the expression for $w_k$ yields the final bound:

$$w_k \leq \exp(2S_{\eta,c})\left(w_0 + AS_{\eta,c}\right).$$

$\square$

## B Controlling $\sigma \mapsto x^\star_\sigma$

The goal of this appendix is to show that the minimiser $x^\star_\sigma$ is Lipschitz-continuous with respect to $\sigma^2$. To establish this, we need to control how the objective function $F_\sigma$ evolves as $\sigma$ changes. A natural way to approach this is through a PDE perspective, since the smoothed density $p_\sigma$ satisfies the heat equation. This connection allows us to describe how $p_\sigma$, its logarithm, and its gradient (i.e., the score function) evolve with respect to $\sigma^2$.

Throughout this appendix, we use the following notation for differential operators acting on functions $f : \mathbb{R}^d \to \mathbb{R}$:

- $\nabla f$ denotes the gradient of $f$, a vector in $\mathbb{R}^d$,
- $\nabla^2 f$ denotes the Hessian of $f$, a $d \times d$ matrix of second-order partial derivatives,
- $\nabla^3 f$ denotes the third-order derivative tensor of $f$, a rank-3 tensor in $\mathbb{R}^{d\times d\times d}$,
- $\Delta f = \mathrm{tr}(\nabla^2 f)$ denotes the Laplacian of $f$.

The first lemma provides several PDEs satisfied by $p_\sigma$, $\ln p_\sigma$, and the score function $\nabla \ln p_\sigma$.

**Lemma 3.** *Let $p(x)$ be a probability density and denote by $p_\sigma(x)$ its convolution with an isotropic centered Gaussian of variance $\sigma^2$. For $\sigma > 0$, it holds that $p_\sigma(x) > 0$ for all $x \in \mathbb{R}^d$ and $p_\sigma$ follows the heat equation:*

$$\frac{\partial p_\sigma}{\partial \sigma^2} = \frac{1}{2}\Delta p_\sigma.$$

*Moreover, $-\ln p_\sigma$ follows the following partial differential equation:*

$$\frac{\partial \ln p_\sigma}{\partial \sigma^2} = \frac{1}{2}(\Delta \ln p_\sigma + \|\nabla \ln p_\sigma\|_2^2).$$

*Taking the gradient in the previous equation we get that the score functions follow:*

$$\frac{\partial \nabla \ln p_\sigma(x)}{\partial \sigma^2} = \frac{1}{2}\big[\nabla\Delta \ln p_\sigma(x) + 2[\nabla^2 \ln p_\sigma(x)]\nabla \ln p_\sigma(x)\big]$$

*Proof.* Standard results (see, e.g., [Evans, 2022, Chapter 2]) guarantee that $(\sigma, x) \mapsto p_\sigma(x)$ is $C^\infty$ on $\mathbb{R}^\star_+ \times \mathbb{R}^d$ and satisfies the heat equation:

$$\frac{\partial p_\sigma}{\partial \sigma^2} = \frac{1}{2}\Delta p_\sigma.$$

By differentiating $\ln p_\sigma$ w.r.t. $\sigma^2$ and using the above, we directly have:

$$\frac{\partial \ln p_\sigma}{\partial \sigma^2} = \frac{1}{2}\frac{\Delta p_\sigma}{p_\sigma},$$

To get the PDE satisfied by $\ln p_\sigma$ notice that:

$$\Delta \ln p_\sigma = \frac{\Delta p_\sigma}{p_\sigma} - \|\nabla \ln p_\sigma\|_2^2,$$

Using both equation above directly yields:

$$\frac{\partial \ln p_\sigma}{\partial \sigma^2} = \frac{1}{2}(\Delta \ln p_\sigma + \|\nabla \ln p_\sigma\|_2^2).$$

Taking the gradient leads to the result. □

This next lemma justifies the use of smoothed gradient descent by confirming that, as the smoothing parameter $\sigma \to 0$, the minimisers of the smoothed objectives $F_\sigma$ converge to the minimiser of the original (non-smoothed) objective $F$. In other words, the limit of the smoothed minimisers coincides with the proximal point we ultimately aim to recover.

**Lemma 4.** *Recall that we define*

$$F(x) \coloneqq \frac{1}{2}\|y-x\|^2 - \tau \ln p(x) \quad \text{and} \quad F_\sigma(x) \coloneqq \frac{1}{2}\|y-x\|^2 - \tau \ln p_\sigma(x).$$

*Recall that* $\operatorname{prox}_{-\tau \ln p}(y) \coloneqq \arg\min_{x\in\mathbb{R}^d} F(x)$ *and that* $\operatorname{prox}_{-\tau \ln p_\sigma}(y) \coloneqq \arg\min_{x\in\mathbb{R}^d} F_\sigma(x)$*. It holds that*

$$\operatorname{prox}_{-\tau \ln p_\sigma}(y) \underset{\sigma\to 0}{\to} \operatorname{prox}_{-\tau \ln p}(y).$$

*Proof.* Let $K$ be a compact set, since $p$ is continuous and $p(x) > 0$ on $K$, we have that there exists $a > 0$ such that $\inf_{x\in K} p(x) \geq a$. Since $p$ is Lipschitz continuous on $K$, Theorem 2 in Nesterov and Spokoiny [2017] ensures that $\sup_{x\in K} |p_\sigma(x) - p(x)| \underset{\sigma\to 0}{\longrightarrow} 0$. Therefore for $\sigma$ small enough $\inf_{x\in K} p_\sigma(x) \geq a/2$ and from standard inequalities on the logarithm:

$$|\ln(p_\sigma(x)) - \ln(p(x))| \leq \frac{|p_\sigma(x) - p(x)|}{\min(p_\sigma(x), p(x))} \leq \frac{2}{a}|p_\sigma(x) - p(x)|.$$

Therefore $\sup_{x\in K} |\ln(p_\sigma(x)) - \ln(p(x))| \underset{\sigma\to 0}{\longrightarrow} 0$ on all compact set $K$, and trivially:

$$\sup_{x\in K} |F_\sigma(x) - F(x)| \underset{\sigma\to 0}{\longrightarrow} 0.$$

To ease notations, let $x_\sigma^\star$ be the minimiser of $F_\sigma$ and $x^\star$ that of $F$. Note that such minimisers exist and are unique since $F_\sigma$ and $F$ are strongly convex by proposition 4. Consider the values $F_\sigma(x^\star)$. By optimality of $x_\sigma^\star$ we know that $F_\sigma(x_\sigma^\star) \leq F_\sigma(x^\star)$. Moreover, since $F_\sigma \to F$ uniformly on compact sets, we have $F_\sigma(x^\star) \to F(x^\star)$, so in particular, the sequence $(F_\sigma(x_\sigma^\star))$ is uniformly bounded above:

$$F_\sigma(x_\sigma^\star) \leq F_\sigma(x^\star) \leq F(x^\star) + \varepsilon,$$

for some positive finite $\varepsilon$. Now, assume that $\|x_\sigma^\star\| \to \infty$ along some sequence. Since the functions $F_\sigma$ are all 1-strongly convex, they can all be lower bounded by the same quadratic and we would have $F_\sigma(x_\sigma^\star) \to \infty$, contradicting the bound above. Therefore, the sequence $(x_\sigma^\star)$ is bounded, and thus contained in a fixed compact set $K \subset \mathbb{R}^d$.

Since $F_\sigma \to F$ uniformly on $K$, any cluster point $x_\infty$ of $(x_\sigma^\star)$ satisfies

$$F(x_\infty) = \lim_{\sigma\to 0} F_\sigma(x_\sigma^\star) \leq \lim_{\sigma\to 0} F_\sigma(x^\star) = F(x^\star).$$

Therefore, by uniqueness of the minimizer of $F$, it must be that $x_\infty = x^\star$ so that $x_\sigma^\star \underset{\sigma\to 0}{\longrightarrow} x^\star$. □

The next proposition establishes the existence and smoothness of the solution path $x_\sigma^\star$ as a function of $\sigma$.

**Proposition 7** (Existence of the smooth solution path)**.** *Define*

$$F_\sigma(x) \coloneqq \frac{1}{2}\|y-x\|^2 - \tau \ln p_\sigma(x).$$

*Denote by* $x_\sigma^\star$ *the minimizer of* $F_\sigma$ *for any* $\sigma > 0$*. Then, under* $\sigma^2 \mapsto x_\sigma$ *is continuously differentiable on* $(0, \tau]$ *and satisfies the following ordinary differential equation:*

$$\frac{\mathrm{d}x_\sigma^\star}{\mathrm{d}\sigma^2} \eqqcolon \dot{x}_\sigma^\star = -\nabla^2 F_\sigma(x_\sigma^\star)^{-1} \partial_{\sigma^2} \nabla F_\sigma(x_\sigma^\star).$$

*Proof.* By smoothness of the solution of the heat equation (see, e.g., [Evans, 2022, Chapter 2]), we have that $x \mapsto F_\sigma(x)$ is differentiable for any $\sigma > 0$ and $(\sigma^2, x) \mapsto \nabla_x F_\sigma(x)$ is jointly differentiable on $\mathbb{R}_+^\star \times \mathbb{R}^d$. Then, by proposition 4, we have that the Hessian $\nabla^2 F_\sigma(x)$ is invertible and satisfies: $\nabla^2 F_\sigma(x) \geq I$. We can then apply the implicit function theorem, which guarantees the existence of a unique solution path $\sigma^2 \mapsto x_\sigma^\star$ to the implicit equation: $\nabla F_\sigma(x_{\sigma^\star}) = 0$ that is differentiable on $(0, \tau]$. By strong convexity of $F_\sigma$, such solution path coincides with the minimizers of $F_\sigma$ for all $\sigma > 0$. The ODE followed by $\sigma^2 \to x_\sigma^\star$ is obtained by taking the derivative w.r.t $\sigma^2$ of the equality $\nabla F_\sigma(x_{\sigma^\star}) = 0$. □

**Proposition 8** (Bound on the solutions). *Let $x_\sigma^\star := \arg\min_{x\in\mathbb{R}^D} F_\sigma(x)$, then for $\sigma^2 \leq \tau$ it holds that*

$$\|x_\sigma^\star - y\| \leq \|y - \operatorname{prox}_{-\tau \ln p}(y)\| + \frac{1}{2}\tau^2 M\sqrt{d}$$

*Proof.* Let use write $\dot{x}_\sigma^\star = \frac{\mathrm{d}x_\sigma^\star}{\mathrm{d}\sigma^2}$ (note that the derivative is with respect to $\sigma^2$ and not $\sigma$). From proposition 7, we have that $x_\sigma^\star$ follows the differential equation:

$$\begin{aligned}\dot{x}_\sigma^\star &= -\nabla^2 F_\sigma(x_\sigma^\star)^{-1}\partial_{\sigma^2}\nabla F_\sigma(x_\sigma^\star)\\ &= \tau\nabla^2 F_\sigma(x_\sigma^\star)^{-1}\partial_{\sigma^2}\nabla \ln p_\sigma(x_\sigma^\star)\\ &= \frac{1}{2}[-\nabla^2\ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau}I_d]^{-1}[\nabla\Delta\ln p_\sigma(x_\sigma^\star) + 2[\nabla^2\ln p_\sigma(x_\sigma^\star)]\nabla\ln p_\sigma(x_\sigma^\star)\big]\end{aligned} \tag{7}$$

where the last equality follows from lemma 3. Furthermore, recalling the optimality condition satisfied by $x_\sigma^\star$, i.e.: $\nabla\ln p_\sigma(x_\sigma^\star) = \frac{1}{\tau}(x_\sigma^\star - y)$, if follows that:

$$\dot{x}_\sigma^\star = -\frac{1}{2\tau}Q_\sigma(x_\sigma^\star - y) + B_\sigma, \tag{8}$$

where the matrix $Q_\sigma$ and vector $B_\sigma$ are given by:

$$Q_\sigma := -[-\nabla^2\ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau}I_d]^{-1}\nabla^2\ln p_\sigma(x_\sigma^\star) \succeq 0 \tag{9}$$

$$B_\sigma := \frac{1}{2}[-\nabla^2\ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau}I_d]^{-1}\Delta\nabla\ln p_\sigma(x_\sigma^\star). \tag{10}$$

Here, the matrix $Q_\sigma$ is positive semi-definite since $-\nabla^2\ln p_\sigma(x)$ is positive by proposition 4. We can multiply the ODE eq. (8) by $x_\sigma^\star - y$ from both sides to get:

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}\|x_\sigma^\star - y\|^2}{\mathrm{d}\sigma^2} &= \langle \dot{x}_\sigma^\star, x_\sigma^\star - y\rangle\\ &= -\frac{1}{2\tau}\|x_\sigma^\star - y\|_{Q_\sigma}^2 + \langle B_\sigma, x_\sigma^\star - y\rangle\\ &\leq \langle B_\sigma, x_\sigma^\star - y\rangle\\ &\leq \|B_\sigma\|\|x_\sigma^\star - y\|.\end{aligned}$$

By the apriori estimate on $\nabla\Delta\log p_\sigma(x_\sigma^\star)$ from Lemma 5, we directly have that $\|B_\sigma\| \leq \frac{\tau}{2}M\sqrt{d}$. Injecting such bound in the above inequality and dividing both sides by $\|x_\sigma^\star - y\|$ yields:

$$\frac{\mathrm{d}\|x_\sigma^\star - y\|}{\mathrm{d}\sigma^2} \leq \frac{\tau}{2}M\sqrt{d}.$$

And after integration of the above inequality w.r.t. $\sigma^2$ on $(0,\sigma^2]$ and using that $\lim_{\sigma\to 0} x_\sigma^\star = \operatorname{prox}_{-\tau\ln p}(y)$ from Proposition 7, we get:

$$\begin{aligned}\|x_\sigma^\star - y\| &\leq \|y - \operatorname{prox}_{-\tau\ln p}(y)\| + \frac{1}{2}\sigma^2\tau M\sqrt{d}\\ &\leq \|y - \operatorname{prox}_{-\tau\ln p}(y)\| + \frac{1}{2}\tau^2 M\sqrt{d},\end{aligned}$$

where the last inequality is since we consider $\sigma^2 \leq \tau$. □

**Proposition 9** (Lipschitz continuity of $\sigma^2 \mapsto x_\sigma^\star$). *Let $x_\sigma^\star := \arg\min_{x\in\mathbb{R}^d} F_\sigma(x)$, then for $\sigma_2^2 \leq \sigma_1^2 \leq \tau$, it holds that:*

$$\|x_{\sigma_1}^\star - x_{\sigma_2}^\star\|_2 \leq (\sigma_1^2 - \sigma_2^2)\Big[\frac{1}{\tau}\|y - \operatorname{prox}_{-\tau\ln p}(y)\| + \tau M\sqrt{d}\Big],$$

*And taking $\sigma_2 \to 0$ in the above inequality:*

$$\|x_\sigma^\star - \operatorname{prox}_{-\tau\ln p}(y)\|_2 \leq \sigma^2\Big[\frac{1}{\tau}\|y - \operatorname{prox}_{-\tau\ln p}(y)\| + \tau M\sqrt{d}\Big],$$

*Proof.* Recall from eq. (7):

$$\dot{x}_\sigma^\star = \frac{1}{2}[-\nabla^2 \ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau} I_d]^{-1}[\nabla\Delta \ln p_\sigma(x_\sigma^\star) + 2[\nabla^2 \ln p_\sigma(x_\sigma^\star)]\nabla \ln p_\sigma(x_\sigma^\star)]$$

Now, by proposition 4, we have that $-\nabla^2 \ln p_\sigma(x) \succeq 0$, and a spectral norm bound on the inverse yields:

$$\|[-\nabla^2 \ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau} I_d]^{-1}\nabla\Delta \ln p_\sigma(x_\sigma^\star)\| \leq \tau \|\nabla\Delta \ln p_\sigma(x_\sigma^\star)\|$$

and:

$$\|[-\nabla^2 \ln p_\sigma(x_\sigma^\star) + \frac{1}{\tau} I_d]^{-1}[\nabla^2 \ln p_\sigma(x_\sigma^\star)]\nabla \ln p_\sigma(x_\sigma^\star)\| \leq \|\nabla \ln p_\sigma(x_\sigma^\star)\|.$$

Putting things together we obtain that:

$$\begin{aligned}\|\dot{x}_\sigma^\star\| &\leq \|\nabla \ln p_\sigma(x_\sigma^\star)\| + \frac{\tau}{2}\|\nabla\Delta \ln p_\sigma(x_\sigma^\star)\| && (11)\\ &\leq \|\nabla \ln p_\sigma(x_\sigma^\star)\| + \frac{\tau}{2} M\sqrt{d}, && (12)\end{aligned}$$

where the second inequality is due to Lemma 5. Now recall that the optimality conditions which define $x_\sigma^\star$ are $\nabla \ln p_\sigma(x_\sigma^\star) = \frac{1}{\tau}(x_\sigma^\star - y)$. Plugging this equality in the upperbound we get that:

$$\begin{aligned}\|\dot{x}_\sigma^\star\| &\leq \frac{1}{\tau}\|y - x_\sigma^\star\| + \frac{\tau}{2} M\sqrt{d}\\ &\leq \frac{1}{\tau}\|y - \mathrm{prox}_{-\tau \ln p}(y)\| + \tau M\sqrt{d},\end{aligned}$$

where the last inequality is due to Proposition 8.

From here it suffices to notice that, for $\sigma_1 \geq \sigma_2 > 0$:

$$\begin{aligned}\|x_{\sigma_1}^\star - x_{\sigma_2}^\star\| &= \left\| \int_{\sigma_1^2}^{\sigma_2^2} \dot{x}_\sigma^\star \mathrm{d}\sigma^2 \right\|\\ &\leq \int_{\sigma_1^2}^{\sigma_2^2} \|\dot{x}_\sigma^\star\| \mathrm{d}\sigma^2\\ &\leq (\sigma_1^2 - \sigma_2^2)\left[\frac{1}{\tau}\|y - \mathrm{prox}_{-\tau \ln p}(y)\| + \tau M\sqrt{d}\right],\end{aligned}$$

to prove the first statement. The second follows from the fact that $x_{\sigma_2}^\star \underset{\sigma_2 \to 0}{\longrightarrow} \mathrm{prox}_{-\tau \ln p}(y)$ by proposition 7. □

This last result is the most technical lemma in this work. It establishes that the third derivative of the smoothed log-density $\ln p_\sigma$ can be uniformly controlled—independently of $\sigma$. This regularity bound is essential for tracking how the minimisers $x_\sigma^\star$ evolve as $\sigma$ varies.

**Lemma 5.** *For all $\sigma \geq 0$, it holds that $\sup_{x \in \mathbb{R}^d} \|\nabla\Delta \ln p_\sigma(x)\| \leq \sqrt{d} M$.*

*Proof. Idea of proof:* Letting $V(\sigma, x) := -\ln p_\sigma(x)$ we will show that $\|\nabla^3 V(\sigma, x)\|$ must be maximal for $\sigma = 0$. To do so we will use the maximum principle on $\|\nabla^3 V(\sigma, x)\|^2$.

From Lemma 3, we have that $V$ follows the following PDE:

$$\partial_{\sigma^2} V = \frac{1}{2}(\Delta V - \|\nabla V\|^2).$$

For $i, j, k \in [d]$, we let $w_{ijk} := \partial_{ijk} V$, which therefore follows:

$$\partial_{\sigma^2} w_{ijk} = \frac{1}{2}(\Delta w_{ijk} - \partial_{ijk}\|\nabla V\|^2).$$

Now let $u_{ijk} = w_{ijk}^2$, multiplying the previous equation by $w_{ijk}$ we get:

$$\begin{aligned}\partial_{\sigma^2} u_{ijk} &= w_{ijk}\Delta w_{ijk} - w_{ijk}\partial_{ijk}\|\nabla V\|^2 \\ &= \frac{1}{2}\big(\Delta u_{ijk} - (\Delta w_{ijk})^2\big) - w_{ijk}\partial_{ijk}\|\nabla V\|^2 \\ &\leq \frac{1}{2}\Delta u_{ijk} - w_{ijk}\partial_{ijk}\|\nabla V\|^2\end{aligned}$$

Summing over $i, j, k$ and letting $S \coloneqq \|\nabla^3 V\|_2^2 = \sum_{ijk} u_{ijk}$ we have that:

$$\partial_{\sigma^2} S \leq \frac{1}{2}\Delta S - \sum_{ijk} w_{ijk}\partial_{ijk}\|\nabla V\|^2$$

It remains to know the behaviour of this last term in the inequality. Since $\|\nabla V\|^2 = \sum_\ell (\partial_l V)^2$, taking the third derivative with respect to $i, j, k$ we get that:

$$\begin{aligned}\partial_{ijk}\|\nabla V\|^2 &= 2\sum_\ell \partial_l V \cdot \partial_{ijkl}V + \partial_{jkl}V \cdot \partial_{il}V + \partial_{ikl}V \cdot \partial_{jl}V + \partial_{ijl}V \cdot \partial_{kl}V \\ &= 2\langle \nabla V, \nabla w_{ijk}\rangle + 2\sum_\ell w_{jkl}\cdot \partial_{il}V + w_{ikl}\cdot\partial_{jl}V + w_{ijl}\cdot\partial_{kl}V.\end{aligned}$$

Multiplying the equality by $w_{ijk}$ and summing over $i, j, k$ we get:

$$\sum_{ijk} w_{ijk}\partial_{ijk}\|\nabla V\|^2 = \langle \nabla V, \nabla S\rangle + 2\sum_{ijk\ell} w_{ijk}w_{jkl}\cdot\partial_{il}V + w_{ijk}w_{ikl}\cdot\partial_{jl}V + w_{ijk}w_{ijl}\cdot\partial_{kl}V.$$

However notice that from the convexity of $V(\sigma, \cdot)$ for all $\sigma \geq 0$, we get that:

$$\sum_{jk}\Big(\underbrace{\sum_{i\ell} w_{ijk}w_{jkl}\cdot\partial_{il}V}_{\geq 0}\Big) \geq 0$$

Which implies that the function $S = \|\nabla^3 V\|^2$ satisfies the parabolic inequality

$$\partial_{\sigma^2} S \leq \frac{1}{2}\Delta S - \langle \nabla V, \nabla S\rangle.$$

Since $S$ cannot be constant, the strong maximum principle[1] implies that $S$ must attain its maximum for $\sigma = 0$:

$$\sup_{\sigma\geq 0, x\in\mathbb{R}^d} \|\nabla^3 \ln p_\sigma(x)\| = \sup_{x\in\mathbb{R}^d}\|\nabla^3 \ln p(x)\|.$$

Finally, from the Cauchy-Schwarz inequality, one gets:

$$\|\nabla\Delta f\|^2 = \sum_{i=1}^d\left(\sum_{j=1}^d \partial_{ijj}f\right)^2 \leq d\sum_{i,j=1}^d(\partial_{ijj}f)^2 \leq d\sum_{i,j,k=1}^d(\partial_{ijk}f)^2 = d\|\nabla^3 f\|_2^2,$$

which concludes the proof.

$\square$

[1]For an accessible explanation of the strong maximum principle, we refer the reader to the corresponding Wikipedia article https://en.wikipedia.org/wiki/Maximum_principle.