Journal of Mathematical Imaging and Vision (2021) 63:821–854
https://doi.org/10.1007/s10851-021-01036-0

On Bayesian Posterior Mean Estimators in Imaging Sciences and Hamilton–Jacobi Partial Differential Equations

Jérôme Darbon[1] · **Gabriel P. Langlois**[1]

Received: 17 June 2020 / Accepted: 27 April 2021 / Published online: 15 May 2021

Abstract

Variational and Bayesian methods are two widely used set of approaches to solve image denoising problems. In a Bayesian setting, these approaches correspond, respectively, to using maximum a posteriori estimators and posterior mean estimators for reconstructing images. In this paper, we propose novel theoretical connections between Hamilton–Jacobi partial differential equations (HJ PDEs) and a broad class of posterior mean estimators with quadratic data fidelity term and log-concave prior. Where solutions to some first-order HJ PDEs with initial data describe maximum a posteriori estimators, here we show that solutions to some viscous HJ PDEs with initial data describe a broad class of posterior mean estimators. We use these connections to establish representation formulas and various properties of posterior mean estimators. In particular, we use these connections to show that some Bayesian posterior mean estimators can be expressed as proximal mappings of smooth functions and derive representation formulas for these functions.

Keywords Hamilton–Jacobi partial differential equations · Imaging inverse problems · Maximum a posteriori estimation · Bayesian posterior mean estimation · Convex analysis

1 Introduction

Image denoising problems consist in estimating an unknown image from a noisy observation in a way that accounts for the underlying uncertainties. Variational and Bayesian methods have become two important approaches for doing so, and in a Bayesian setting these approaches correspond, respectively, to using maximum a posteriori estimators and posterior mean estimators for reconstructing images. The goal of this paper is to describe a broad class of Bayesian posterior mean estimators with quadratic data fidelity term and log-concave prior using Hamilton–Jacobi (HJ) partial differential equations (PDEs) and to use these connections to clarify certain image denoising properties of this class of Bayesian posterior estimators.

Research supported by NSF DMS 1820821. Authors' names are given in last/family name alphabetical order.

✉ Jérôme Darbon
jerome_darbon@brown.edu

Gabriel P. Langlois
gabriel_provencher_langlois@brown.edu

1 Division of Applied Mathematics, Brown University, Providence, USA

To illustrate the main ideas of this paper, we first briefly describe convex finite-dimensional variational and Bayesian methods relevant to image denoising problems. Variational methods formulate image denoising problems as the optimization of a weighted sum of a data fidelity term (which embeds the knowledge of the nature of the noise corrupting the unknown image) and a regularization term (which embeds known properties of the image to reconstruct), where the goal is to minimize this sum to obtain an estimate that hopefully accounts well for both the data fidelity term and the regularization term [13,17]. Bayesian methods formulate image denoising problems in a probabilistic framework that combine observed data through a likelihood function (which models the noise corrupting the unknown image) and prior knowledge through a prior distribution (which models known properties of the unknown image) to generate a posterior distribution. An appropriate decision rule that minimizes the posterior expected value of a loss function, also called a Bayes estimator, then selects a meaningful image estimate from the posterior distribution that hopefully accounts well for both the prior knowledge and observed data [21,62–64,66]. A standard example is the posterior mean estimator, the mean of the posterior distribution, which minimizes the mean squared error [43, pages 344-345], and more generally, Bregman loss functions [4].

In this paper, we will focus on the class of finite-dimensional image denoising problems

$$\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the observed image, $\boldsymbol{u} \in \mathbb{R}^n$ is the unknown image, $\boldsymbol{\eta}$ is independent identically distributed Gaussian noise. These problems are well-known to be ill-posed in general, and variational and Bayesian approaches are celebrated methods to find meaningful solutions to these ill-posed problems [2,21,67]. These methods aim to estimate the original uncorrupted image by computing, respectively, the maximum a posteriori (MAP) and posterior mean (PM) estimates

$$\boldsymbol{u}_{MAP}(\boldsymbol{x}, t) := \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2t} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y}) \right\} \tag{2}$$

and

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) := \frac{\int_{\mathbb{R}^n} \boldsymbol{y} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}}. \tag{3}$$

The functions $\boldsymbol{y} \mapsto \frac{1}{2t}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ and $J : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ in (2) are, respectively, the (quadratic) data fidelity and regularization terms, and the functions $\boldsymbol{y} \mapsto e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon}$ and $\boldsymbol{y} \mapsto e^{-J(\boldsymbol{y})/\epsilon}$ in (3) are, respectively, the (Gaussian) likelihood function and generalized prior distribution. The parameter $t > 0$ controls the relative importance of the data fidelity term over the regularization term, and the parameter ϵ controls the shape of the posterior distribution in (3), where small values of ϵ favor configurations close to the mode, which is the MAP estimate, of the posterior distribution.

Let us illustrate the MAP and PM estimates and their denoising capabilities with an example. We consider an anisotropic version of the Rudin–Osher–Fatemi (ROF) image denoising model, which consists of considering an anisotropic total variation (TV) regularization term with quadratic data fidelity term [6,12,60]. Specifically, we define anisotropic TV as follows

$$\mathrm{TV}(\boldsymbol{y}) = \sum_{i,j\in\{1,\ldots n\}} w_{i,j}|y_i - y_j|,$$

where $w_{i,j} \geq 0$ and the value of an image $\boldsymbol{y}$ at the pixel i is denoted by $y_i \in \mathbb{R}$. For illustration purposes, we assume that a digital image is defined on a two-dimensional regular grid and only consider the 4-nearest neighbors interactions for defining TV (i.e., $w_{i,j} = w_{j,i} = \frac{1}{2}$ if i and j are neighbors, and $w_{i,j} = w_{j,i} = 0$ otherwise, see [19] for instance). Let $\boldsymbol{x}$ denote an observed noisy image and t and ϵ be parameters as previously defined. Then, the associated anisotropic ROF problem [60] takes the form

$$\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2t} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \mathrm{TV}(\boldsymbol{y}) \right\}. \tag{4}$$

The MAP and PM estimates to the ROF problem (4) are given, respectively, by Eqs. (2) and (3) with $J(\boldsymbol{y}) = \mathrm{TV}(\boldsymbol{y})$, i.e.,

$$\boldsymbol{u}_{MAP}(\boldsymbol{x}, t) = \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2t} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \mathrm{TV}(\boldsymbol{y}) \right\} \tag{5}$$

and

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \frac{\int_{\mathbb{R}^n} \boldsymbol{y} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mathrm{TV}(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mathrm{TV}(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}}. \tag{6}$$

We note here that the PM estimate (6) with total variation prior and its denoising properties was investigated in [48,49].

Figure 1a depicts the image Barbara, which we corrupt with Gaussian noise (zero mean with standard deviation $\sigma = 10$) in Fig. 1b. We let $\boldsymbol{x}$ denote this corrupted image, and we choose the parameters $t = 16$ and $\epsilon = 6.25$ in the MAP and PM estimates. The MAP estimate can be computed up to the machine precision using maximum-flow-based algorithms [11,19,41], and the PM estimate can be approximated using Markov Chain Monte Carlo methods. Here, we approximated the PM estimate (6) using the variable-at-a-time Metropolis–Hastings algorithm with random scan detailed in ([48], Algorithm 2 on page 42). Specifically, for the parameters of Algorithm 2 in [48], we used, in the terminology of their algorithm, the parameters $\sigma = 10$ and $\lambda = 32$ (corresponding here to the choice of $t = 16$ and $\epsilon = 6.25$ in (6)), we chose the initial point of the algorithm to be the MAP estimate $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t)$, and finally, we set the internal parameters of Algorithm 2 in [48] as follows: $\alpha = 17.32$ (these values yield an acceptance rate in the algorithm close to the optimal value 0.234 suggested in [57]), 20,000 for the maximum number of iterations, and n for the subsampling rate.

The MAP and PM estimates associated with the ROF model with these parameters produce the denoised images illustrated in Fig. 1c and d. Figure 2a–d zoom-in on the face of Barbara in Fig. 1 The denoised image of Barbara with the MAP estimate exhibits staircasing effects [14,23,25] that can be observed in Fig. 2c, whereas the denoised image of Barbara with the posterior mean estimate does not. In either case, the denoised images result in a lost of texture, as can be seen by comparing Fig. 2a with c and d.

Variational methods are popular because the resultant optimization problem for various non-smooth and convex

Fig. 1 The anisotropic ROF model endowed with 4-nearest neighbors is applied to the test image "Barbara". The original image is shown in (**a**). The image is corrupted by Gaussian noise (zero mean with standard deviation $\sigma = 10$) and is shown in (**b**). The corresponding minimizer $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t)$ given by (4) and posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ given by (6) with parameters $t = 16$ and $\epsilon = 6.25$ is illustrated in (**c**) and (**d**), respectively

regularization terms used in image denoising problems, such as total variation and l_1-norm based regularization terms, is well-understood [6,10,12,17,18,24,60] and can be solved efficiently using robust numerical optimization methods [13]. MAP estimates from variational methods are also generally faster to compute than posterior mean estimates, since the latter require complex stochastic methods to compute. Reconstructed images from variational methods with non-smooth and convex regularization terms, however, may have undesirable and visually unpleasant staircasing effects due to the singularities of the non-smooth regularization terms [14,23,25,48,52,68]. This is illustrated for example in Fig. 1c, which contains regions where the pixel values are equal and lead to staircasing effects. In contrast, posterior mean estimates with quadratic fidelity term and total variation regularization terms have been shown to avoid staircasing effects [48,49]. This is illustrated for example in Figs. 1 and 2d, where the denoised image with posterior mean estimate does not contain visibly substantial regions where the pixel values are equal.

Related work Several papers have proposed novel connections between MAP and Bayesian estimators, including posterior mean estimators. First, [48,49] showed that the class of Bayesian posterior mean estimates (3) with TV regularization term J can be expressed as minimizers to optimization problems involving a quadratic fidelity term and a smooth convex regularization term, i.e., there exists a smooth regularization term $f_{\text{reg}}\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + f_{\text{reg}}(\boldsymbol{y}) \right\}. \tag{7}$$

This result was later extended to general priors [34], general Gaussian data fidelity terms [35], and to some non-quadratic data fidelity terms [36,37]. To our knowledge, there is no representation formula for this smooth regularization term available in the literature.

Second, [9] showed that the MAP estimate (2) corresponds to a Bayes estimator when the regularization term J is convex and uniformly Lipschitz continuous on $\mathbb{R}^n$, that is, the

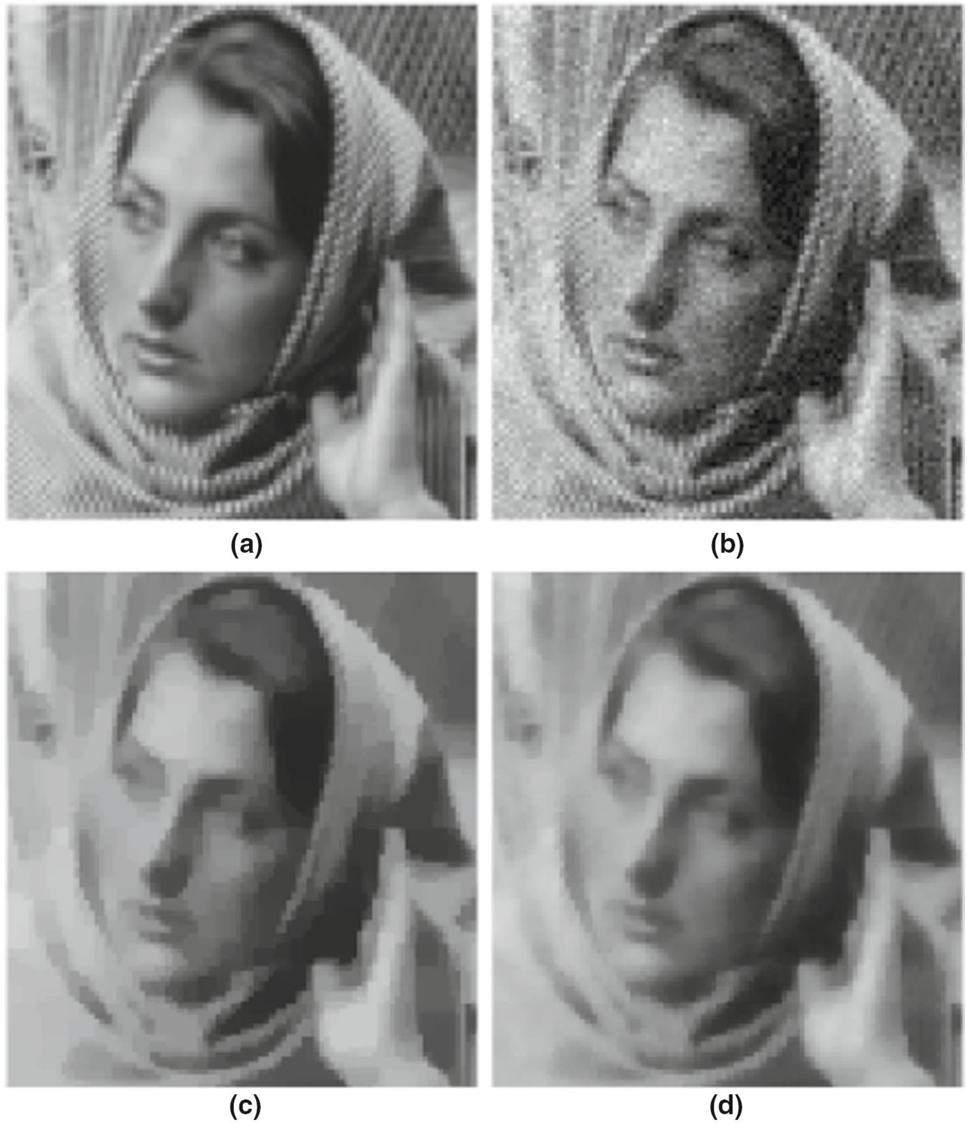


Fig. 2 The anisotropic ROF model endowed with 4-nearest neighbors is applied to the test image "Barbara". Images (**a**)–(**d**) are zoomed-in versions of the images illustrated in Fig. 1

MAP estimate (2) minimizes the posterior expected value of an appropriate loss function. This was later extended by [8] to some log-concave posterior distributions with non-quadratic fidelity term and later studied from the point of view of differential geometry in [53] and also derived for posterior distributions that are strongly log-concave and at least three times differentiable.

In addition to these results, it is known that under certain assumptions on the regularization term J, the value of the minimization problem

$$S_0(\boldsymbol{x}, t) := \min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2t} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y}) \right\} \tag{8}$$

whose minimizer is the MAP estimate (2), satisfies the first-order HJ PDE

$$\begin{cases} \frac{\partial S_0}{\partial t}(\boldsymbol{x}, t) + \frac{1}{2} \|\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t)\|_2^2 = 0 & \text{in } \mathbb{R}^n \times (0, +\infty), \\ S_0(\boldsymbol{x}, 0) = J(\boldsymbol{x}) & \text{in } \mathbb{R}^n. \end{cases} \tag{9}$$

The properties of the minimizer $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t)$ follow from the properties of the solution to this HJ equation [17,18]. In particular, the MAP estimate satisfies the representation formula $\boldsymbol{u}_{PM}(\boldsymbol{x}, t) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t)$.

We note that the results of [17,18] only concern connections between a class of first-order HJ PDEs and MAP estimators. To our knowledge, connections between posterior estimators and HJ PDEs are not available in the literature.

Contributions In this paper, we propose novel theoretical connections between solutions to HJ PDEs and a broad class of Bayesian methods and posterior mean estimators. These connections are described in Propositions 3.1 and 3.2 for viscous HJ PDEs and first-order HJ PDEs, respectively. We show in Proposition 3.1 that the posterior mean estimate (3) is described by the solution to a viscous HJ with initial data corresponding to the convex regularization term J, which we characterize in detail in terms of the data $\boldsymbol{x}$ and parameters t and ϵ. In particular, the posterior mean estimate (3) satisfies the representation formula $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$. Next, we use the connections between viscous HJ PDEs and posterior mean estimates established in Proposition 3.1 to show in Proposition 3.2 that the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ can be expressed through the gradient of the solution to a first-order HJ PDE with twice continuously differentiable convex initial data $\mathbb{R}^n \ni \boldsymbol{x} \mapsto K_\epsilon^*(\boldsymbol{x}, t) - \frac{1}{2}\|\boldsymbol{x}\|_2^2$, where

$$K_\epsilon(\boldsymbol{x}, t) = t\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\text{dom } J} e^{\left(\frac{1}{t}\langle \boldsymbol{x}, \boldsymbol{y}\rangle - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}\right)$$

and $\boldsymbol{x} \mapsto K_\epsilon^*(\boldsymbol{x}, t)$ is the Fenchel–Legendre transform of the function $\boldsymbol{x} \mapsto K_\epsilon(\boldsymbol{x}, t)$. In other words, we show

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2\right)\right\}.$$

This formula gives the representation of the convex regularization term, enabling one to express the posterior mean estimate as the minimizer of a convex variational problem, and in fact in terms of the solution to a first-order HJ PDE. This thereby extends the results of [34,48], who showed existence of this regularization term when the data fidelity term is quadratic, but not its representation. The second-order continuous differentiability of this regularization term, in particular, implies that the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ avoids image denoising staircasing effects as a consequence of the results derived in [51, Theorem 3].

We also present several topological properties of posterior mean estimators in Proposition 4.1, and we use these in conjunction with the connections between HJ PDEs and posterior mean estimators to derive representation and monotonicity properties of posterior mean estimators in Propositions 4.2 and 4.3, respectively. These properties are then used to derive an optimal upper bound on the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)\|_2^2\right]$, an estimate of the squared difference between the MAP and posterior mean estimates, monotonicity and non-expansiveness properties of the posterior mean estimate, and the behavior of the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ in the limit $t \to 0$ (Proposition 4.4). Finally, we use the connections between both MAP and posterior mean estimates and HJ PDEs to characterize the MAP estimate (2) in the context of Bayesian estimation theory, and specifically in proposition 4.5 to show that the MAP estimate (2) corresponds to the Bayes estimator of the Bayesian risk (52) whenever J is convex on $\mathbb{R}^n$ and bounded from below. When J is defined only on a strict subset of $\mathbb{R}^n$, we further show that the Bayesian risk (52) has a corresponding Bayes estimator that is described in terms of the solution to both the first-order HJ PDE (2.2) and the viscous HJ PDE (3.1).

We would like to emphasize that the proofs of several results presented in this paper are inspired from techniques in existing works in several fields, including partial differential equations [26,42], convex analysis [38–40,58,59], the theory of set-valued maps and differential inclusions [3], large deviations theory [22], and geometric measure theory [1,27,28,32,46,54,56]. To our knowledge, however, the results presented in this paper are novel.

Organization In Sect. 2, we review concepts of real and convex analysis that will be used throughout this paper. In Sect. 3, we establish theoretical connections between a broad class of Bayesian posterior mean estimators and HJ PDEs. Our mathematical setup is described in Subsection 3.1, the connections of posterior mean estimators to viscous HJ PDEs are described in Subsection 3.2, and the connections of posterior mean estimators to first-order HJ PDEs are described in Subsection 3.3. We use these connections to establish various properties of posterior mean estimators in Sect. 4. Specifically, we present topological, representation, and monotonicity properties of posterior mean estimators in Subsection 4.1, an optimal upper bound on the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)\|_2^2\right]$, an estimate of the squared difference between the MAP and posterior mean estimates, monotonicity and non-expansiveness properties of the posterior mean estimate, and the behavior of the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ in the limit $t \to 0$ in Subsection 4.2. Finally, we establish properties of MAP and posterior mean estimators in terms of Bayesian risks involving Bregman divergences in Subsection 4.3.

2 Background

This section reviews concepts from real and convex analysis that will be used in this paper. For convenience to the reader, we summarize some notations and definitions in Table 1; the definitions are explained in detail below and the reader may skip them. We also refer to [30,38,39,58,59] for comprehensive references.

In what follows, the Euclidean scalar product on $\mathbb{R}^n$ will be denoted by $\langle \cdot, \cdot \rangle$ and its associated norm by $\|\cdot\|_2$. The closure and interior of a non-empty set $C \subset \mathbb{R}^n$ will be denoted by cl C and int C, respectively. The boundary of a non-empty set $C \subset \mathbb{R}^n$ is defined as cl $C \setminus$ int C and will be denoted by bd C. The domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the set dom $f = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < +\infty\}$. Let $f : \Omega \times \Omega' \to \mathbb{R}$ with $\Omega \times \Omega' \subset \mathbb{R}^n \times \mathbb{R}^{n'}$. It will be useful in this paper to consider the gradient $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y})$, divergence $\nabla_{\boldsymbol{x}} \cdot f(\boldsymbol{x}, \boldsymbol{y})$ and Laplacian $\Delta_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y})$ of $\Omega \ni \boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{y} \in \Omega'$, which are defined as follows: $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}, \boldsymbol{y}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}, \boldsymbol{y})\right)$, $\nabla_x \cdot f(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}, \boldsymbol{y})$, and $\Delta_x f(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}, \boldsymbol{y})$.

Definition 1 (*Proper and lower semicontinuous functions*) A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper if $\mathrm{dom}\, f \neq \varnothing$ and $f(\boldsymbol{x}) > -\infty$ for every $\boldsymbol{x} \in \mathrm{dom}\, f$.

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous at $\boldsymbol{x} \in \mathbb{R}^n$ if it satisfies $\liminf_{k \to +\infty} f(\boldsymbol{x}_k) \geq f(\boldsymbol{x})$ for every sequence $\{\boldsymbol{x}_k\}_{k=1}^{+\infty} \subset \mathbb{R}^n$ such that $\lim_{k \to +\infty} \boldsymbol{x}_k = \boldsymbol{x}$.

Definition 2 (*Convex sets and their relative interiors*) A subset $C \subset \mathbb{R}^n$ is convex if for every pair $(\boldsymbol{x}, \boldsymbol{y}) \in C \times C$ and every scalar $\lambda \in (0, 1)$, the line segment $\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}$ is contained in C.

The relative interior of a convex set C, denoted by $\mathrm{ri}\, C$, is the set of points in the interior of the unique smallest affine set containing C. Every convex set C with non-empty interior is n-dimensional with $\mathrm{ri}\, C = \mathrm{int}\, C$ and has positive Lebesgue measure, and furthermore, the n-dimensional Lebesgue measure of the boundary $\mathrm{bd}\, C$ is equal to zero [45].

Definition 3 (*Convex functions and the set* $\Gamma_0(\mathbb{R}^n)$) A proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex if its domain is convex and if the inequality

$$f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y})$$

holds for every pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{dom}\, f \times \mathrm{dom}\, f$ and every scalar $\lambda \in [0, 1]$. It is *strictly convex* if the inequality above is strict whenever $\boldsymbol{x} \neq \boldsymbol{y}$ and $\lambda \in (0, 1)$.

A proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is *strongly convex with parameter* $m \geq 0$ if

$$f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}) - \frac{m}{2}\lambda(1 - \lambda) \left\|\boldsymbol{x} - \boldsymbol{y}\right\|_2^2$$

for every pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{dom}\, f \times \mathrm{dom}\, f$ and every scalar $\lambda \in [0, 1]$.

The class of proper, convex and lower semicontinuous functions is denoted by $\Gamma_0(\mathbb{R}^n)$.

Definition 4 (*Projections*) Let C be a closed convex subset of $\mathbb{R}^n$. To every $\boldsymbol{x} \in \mathbb{R}^n$, there exists a unique element $\pi_C(\boldsymbol{x}) \in C$ called the projection of $\boldsymbol{x}$ onto C that is closest to $\boldsymbol{x}$ in Euclidean norm, i.e.,

$$\pi_C(\boldsymbol{x}) := \underset{\boldsymbol{y} \in C}{\arg\min} \left\|\boldsymbol{x} - \boldsymbol{y}\right\|_2^2. \tag{10}$$

This correspondence defines a map $\boldsymbol{x} \mapsto \pi_C(\boldsymbol{x})$ from $\mathbb{R}^n$ to C called the projector onto C ([3], Chapter 0.6, Corollary 1). It satisfies the characterization

$$\langle \boldsymbol{x} - \pi_C(\boldsymbol{x}), \boldsymbol{y} - \pi_C(\boldsymbol{x}) \rangle \leq 0, \quad \forall \boldsymbol{y} \in C. \tag{11}$$

Definition 5 (*Subdifferentials and subgradients*) Let $f \in \Gamma_0(\mathbb{R}^n)$. The *subdifferential* of f at $\boldsymbol{x} \in \mathrm{dom}\, f$ is the set $\partial f(\boldsymbol{x})$ of vectors $\boldsymbol{p} \in \mathbb{R}^n$ that satisfies the inequality

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{x} \rangle \tag{12}$$

for every $\boldsymbol{y} \in \mathbb{R}^n$. The subdifferential $\partial f(\boldsymbol{x})$ is a closed convex subset of $\mathbb{R}^n$ whenever it is non-empty, and the vectors $\boldsymbol{p} \in \partial f(\boldsymbol{x})$ are called the subgradients of f at $\boldsymbol{x}$.

The set of points $\boldsymbol{x} \in \mathrm{dom}\, f$ for which the subdifferential $\partial f(\boldsymbol{x})$ is non-empty is denoted by $\mathrm{dom}\, \partial f$, and it includes the relative interior of the domain of f, i.e., $\mathrm{ri}\, (\mathrm{dom}\, f) \subset \mathrm{dom}\, \partial f$ [58, Theorem 23.4].

If f is strongly convex of parameter $m \geq 0$ and $\boldsymbol{x} \in \mathrm{dom}\, \partial f$, then the subgradients $\boldsymbol{p} \in \partial f(\boldsymbol{x})$ satisfy the inequality

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{m}{2} \left\|\boldsymbol{y} - \boldsymbol{x}\right\|_2^2.$$

If f is differentiable at $\boldsymbol{x}$, then $\boldsymbol{x} \in \mathrm{dom}\, \partial f$ and the gradient $\nabla f(\boldsymbol{x})$ is the unique subgradient of f at $\boldsymbol{x}$, and conversely if f has a unique subgradient at $\boldsymbol{x}$, then f is differentiable at that point [58, Theorem 25.1].

The set-valued subdifferential mapping $\mathrm{dom}\, \partial f \ni \boldsymbol{y} \mapsto \partial f(\boldsymbol{y})$ satisfies two important properties. First, it is monotone in that if f is strongly convex of parameter $m \geq 0$, then for every pair $(\boldsymbol{y}, \boldsymbol{y}_0) \in \mathrm{dom}\, \partial f \times \mathrm{dom}\, \partial f$ and $\boldsymbol{p} \in \partial f(\boldsymbol{y})$, $\boldsymbol{p}_0 \in \partial f(\boldsymbol{y}_0)$ the following inequality holds ([58], page 240 and Corollary 31.5.2):

$$m \left\|\boldsymbol{y} - \boldsymbol{y}_0\right\|_2^2 \leq \left\langle \boldsymbol{p} - \boldsymbol{p}_0, \boldsymbol{y} - \boldsymbol{y}_0 \right\rangle, \tag{13}$$

Second, the mapping $\mathrm{dom}\, \partial f \ni \boldsymbol{y} \mapsto \pi_{\partial f(\boldsymbol{y})}(\boldsymbol{0})$ is well-defined, and it selects the subgradient of the minimal norm in $\partial f(\boldsymbol{x})$ and defines a function continuous almost everywhere on $\mathrm{dom}\, \partial f$, a consequence of the fact that this mapping agrees with the gradient of f over the set of points in $\mathrm{int}\, (\mathrm{dom}\, J)$ at which f is differentiable [58, Theorem 25.5].

Definition 6 (*Fenchel–Legendre transform*) Let $f \in \Gamma_0(\mathbb{R}^n)$. The *Fenchel–Legendre transform* $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of f is defined by

$$f^*(\boldsymbol{p}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left\{\langle \boldsymbol{p}, \boldsymbol{x} \rangle - f(\boldsymbol{x})\right\}. \tag{14}$$

For every $f \in \Gamma_0(\mathbb{R}^n)$, the mapping $f \mapsto f^*$ is one-to-one, $f^* \in \Gamma_0(\mathbb{R}^n)$, and $(f^*)^* = f$. Moreover, for every $\boldsymbol{x} \in \mathbb{R}^n$

Table 1 Notation used in this paper. Here, we use C to denote a set in $\mathbb{R}^n$, f to denote a function from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$ and $\boldsymbol{x}$ to denote a vector in $\mathbb{R}^n$

Notation	Meaning	Definition
$\langle \cdot, \cdot \rangle$	Euclidean scalar product in $\mathbb{R}^n$	$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^n x_i y_i$
$\|\cdot\|_2$	Euclidean norm in $\mathbb{R}^n$	$\|\boldsymbol{x}\|_2 := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$
ri C	Relative interior of C	The set of points in the interior of the unique smallest affine set containing C
bd C	Boundary of C	The boundary of a non-empty set C is defined as cl $C \setminus$ int C.
dom f	Domain of f	$\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < +\infty\}$
$\Gamma_0(\mathbb{R}^n)$	A useful and standard class of convex functions	The set containing all proper, convex, lower semicontinuous functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$
$\pi_C(\boldsymbol{x})$	$\pi_C(\boldsymbol{x}) := \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$	Projection of $\boldsymbol{x} \in \mathbb{R}^n$ onto the closed convex set C.
$\partial f(\boldsymbol{x})$	Subdifferential of f at $\boldsymbol{x}$	$\{\boldsymbol{p} \in \mathbb{R}^n : f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{x} \rangle \; \forall \boldsymbol{y} \in \mathbb{R}^n\}$
dom ∂f	$\boldsymbol{x} \in \mathrm{dom}\, \partial f \implies \partial f(\boldsymbol{x}) \neq \varnothing$	The set of points $\boldsymbol{x} \in$ dom f for which the subdifferential $\partial f(\boldsymbol{x})$ is non-empty.
f^*	Fenchel–Legendre transform of f	$f^*(\boldsymbol{p}) := \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{\langle \boldsymbol{p}, \boldsymbol{x} \rangle - f(\boldsymbol{x})\}$

The closure and interior of a non-empty set $C \subset \mathbb{R}^n$ are denoted by cl C and int C, respectively and $\boldsymbol{p} \in \mathbb{R}^n$, f and f^* satisfy Fenchel's inequality

$$f(\boldsymbol{x}) + f^*(\boldsymbol{p}) \geq \langle \boldsymbol{p}, \boldsymbol{x} \rangle, \tag{15}$$

where equality holds if and only if $\boldsymbol{p} \in \partial f(\boldsymbol{x})$, if and only if $\boldsymbol{x} \in \partial f^*(\boldsymbol{p})$ [39, corollary 1.4.4]. If f is also differentiable, the supremum in (14) is attained whenever there exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{p} = \nabla f(\boldsymbol{x})$.

Definition 7 (*Bregman divergences*) Let $f \in \Gamma_0(\mathbb{R}^n)$. The *Bregman divergence* of f is the function $D_f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$D_f(\boldsymbol{x}, \boldsymbol{p}) = f(\boldsymbol{x}) - \langle \boldsymbol{p}, \boldsymbol{x} \rangle + f^*(\boldsymbol{p}). \tag{16}$$

It satisfies $D_f(\boldsymbol{x}, \boldsymbol{p}) \geq 0$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{p} \in \mathbb{R}^n$ by Fenchel's inequality (15), with $D_f(\boldsymbol{x}, \boldsymbol{p}) = 0$ whenever $\boldsymbol{p} \in \partial f(\boldsymbol{x})$. It also satisfies $D_f(\boldsymbol{x}, \boldsymbol{p}) = D_{f^*}(\boldsymbol{p}, \boldsymbol{x})$, with $D_f(\boldsymbol{x}, \boldsymbol{p}) = D_f(\boldsymbol{p}, \boldsymbol{x})$ if and only if f is the quadratic $f = \frac{1}{2}\|\cdot\|_2^2$.

Definition 8 (*Infimal convolutions*) Let $f_1 \in \Gamma_0(\mathbb{R}^n)$ and $f_2 \in \Gamma_0(\mathbb{R}^n)$. The *infimal convolution* of f_1 and f_2 is the function

$$\mathbb{R}^n \ni \boldsymbol{x} \mapsto (f_1 \square f_2)(\boldsymbol{x}) = \inf_{\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}} \{f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2)\}. \tag{17}$$

The infimal convolution is exact if the infimum is attained at $\boldsymbol{x}_1 \in \operatorname{dom} f_1$ and $\boldsymbol{x}_2 \in \operatorname{dom} f_2$, and in that case the infimum in (17) can be replaced by a minimum. When the relative interiors of f_1 and f_2 have a point in common, i.e., $\operatorname{ri}\operatorname{dom} f_1 \cap \operatorname{ri}\operatorname{dom} f_2 \neq \varnothing$, the Fenchel–Legendre transform of the infimal convolution (17) equals the sum of their respective Fenchel–Legendre transforms [58, Theorem 16.4], that is,

$$(f_1 \square f_2)^*(\boldsymbol{p}) = f_1^*(\boldsymbol{p}) + f_2^*(\boldsymbol{p}).$$

If $f \in \Gamma_0(\mathbb{R}^n)$, then Moreau's decomposition Theorem [40, 50] asserts that

$$\frac{1}{2}\|\cdot\|_2^2 \square f + \frac{1}{2}\|\cdot\|_2^2 \square f^* = \frac{1}{2}\|\cdot\|_2^2.$$

The following proposition provides conditions for which two functions f_1 and f_2 satisfying $f_1 + f_2 = \frac{1}{2}\|\cdot\|_2^2$ can be factorized, respectively, in the form $\frac{1}{2}\|\cdot\|_2^2 \square f$ and $\frac{1}{2}\|\cdot\|_2^2 \square f^*$. The proof can be found in [40].

Proposition 2.1 (Infimal deconvolutions [40]) *Suppose f_1 and f_2 are two convex functions on $\mathbb{R}^n$ such that $f_1 + f_2 = \frac{1}{2}\|\cdot\|_2^2$. Then, there exists a unique function $f \in \Gamma_0(\mathbb{R}^n)$ such that*

$$f_1 = \frac{1}{2}\|\cdot\|_2^2 \square f \text{ and } f_2 = \frac{1}{2}\|\cdot\|_2^2 \square f^*,$$

where $f(\boldsymbol{x}) = f_2^*(\boldsymbol{x}) - \frac{1}{2}\|\boldsymbol{x}\|_2^2$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Moreover, f_1 and f_2 are continuously differentiable and

$$\nabla f_1(\boldsymbol{x}) \in \partial f(\nabla h(\boldsymbol{x})) \text{ and } \nabla f_2(\boldsymbol{x}) \in \partial f^*(\nabla g(\boldsymbol{x})).$$

Definition 9 (*Moreau–Yosida envelopes and proximal mappings*) Let $t > 0$ and $J \in \Gamma_0(\mathbb{R}^n)$. The functions

$$\begin{aligned} \boldsymbol{x} &\mapsto \left(\frac{1}{2t}\|\cdot\|_2^2 \square J\right)(\boldsymbol{x}) \\ &= \inf_{\boldsymbol{y} \in \mathbb{R}^n} \left\{\frac{1}{2t}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} \end{aligned} \tag{18}$$

and

$$\boldsymbol{x} \mapsto \arg\min_{\boldsymbol{y} \in \mathbb{R}^n} \left\{\frac{1}{2t}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} \tag{19}$$

are called the Moreau–Yosida envelope and proximal mapping of J, respectively [39,50,59].

The following proposition provides connections between HJ PDEs and Moreau–Yosida envelopes and proximal mappings, which corresponds to certain optimization problems in image denoising problems. Specifically, this proposition describes the behavior of the solution to the infimum problem (8) and its corresponding minimizer (2), and in particular that for any observed image $\boldsymbol{x} \in \mathbb{R}^n$ and parameter $t > 0$, the imaging problem (8) has always a unique solution. A summary of these results and their proof can be found in [17].

Proposition 2.2 ([17]) *Let $J \in \Gamma_0(\mathbb{R}^n)$. Then, the following statements hold.*

(i) *The unique continuously differentiable and convex function $S_0 : \mathbb{R}^n \times [0, +\infty) \to \mathbb{R}$ that satisfies the first-order Hamilton–Jacobi equation with initial data*

$$\begin{cases} \frac{\partial S_0}{\partial t}(\boldsymbol{x}, t) + \frac{1}{2}\|\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t)\|_2^2 = 0 & \text{in } \mathbb{R}^n \times (0, +\infty), \\ S_0(\boldsymbol{x}, 0) = J(\boldsymbol{x}) & \text{in } \mathbb{R}^n, \end{cases} \tag{20}$$

is defined by

$$S_0(\boldsymbol{x}, t) = \left(\left(\frac{1}{2t}\|\cdot\|_2^2\right) \square J\right)(\boldsymbol{x}) \quad (\textit{Lax–Oleinik formula}) \tag{21}$$

$$= \inf_{\boldsymbol{y} \in \mathbb{R}^n} \left\{\frac{1}{2t}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\}. \tag{22}$$

Furthermore, for every $\boldsymbol{x} \in \operatorname{dom} J$, sequence $\{t_k\}_{k=1}^{+\infty}$ of positive real numbers converging to 0, and sequence $\{\boldsymbol{d}_k\}_{k=1}^{+\infty}$ of vectors converging to $\boldsymbol{d} \in \mathbb{R}^n$, the pointwise limit $S_0(\boldsymbol{x} + t_k\boldsymbol{d}_k, t_k)$ as $k \to +\infty$ exists and satisfies

$$\lim_{k \to +\infty} S_0(\boldsymbol{x} + t_k\boldsymbol{d}_k, t_k) = J(\boldsymbol{x}).$$

(ii) *For every* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $t > 0$, *the infimum in (22) exists and is attained at a unique point* $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t) \in \mathrm{dom}\ \partial J$ *(see Eq. (2)). In addition, the minimizer* $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t)$ *satisfies the formula*

$$\boldsymbol{u}_{MAP}(\boldsymbol{x}, t) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t), \tag{23}$$

and $\left(\frac{\boldsymbol{x}-\boldsymbol{u}_{MAP}(\boldsymbol{x},t)}{t}\right) \in \partial J(\boldsymbol{u}_{MAP}(\boldsymbol{x}, t))$.

(iii) *Let* $\{t_k\}_{k=1}^{+\infty}$ *be a sequence of positive real numbers converging to zero and let* $\{\boldsymbol{d}_k\}_{k=1}^{+\infty}$ *be a sequence of elements in* $\mathbb{R}^n$ *converging to some* $\boldsymbol{d} \in \mathbb{R}^n$. *Then, for every* $\boldsymbol{x} \in \mathrm{dom}\ J$ *the pointwise limit of* $\boldsymbol{u}_{MAP}(\boldsymbol{x}, t)$ *as* $t \to 0$ *exists and satisfies*

$$\lim_{k\to+\infty} \boldsymbol{u}_{MAP}(\boldsymbol{x} + t_k\boldsymbol{d}_k, t_k) = \boldsymbol{x}.$$

(iv) *Let* $\boldsymbol{x} \in \mathrm{dom}\ \partial J$ *and let* $\{t_k\}_{k=1}^{+\infty}$ *be a sequence of positive real numbers converging to zero. Then, the limit of* $\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t_k)$ *as* $k \to +\infty$ *exists and satisfies*

$$\lim_{k\to+\infty} \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, t_k) = \pi_{\partial J(\boldsymbol{x})}(\boldsymbol{0}). \tag{24}$$

3 Connections between Bayesian Posterior Mean Estimators and Hamilton–Jacobi Partial Differential Equations

3.1 Setup

To establish connections between Bayesian posterior mean estimators and Hamilton–Jacobi equations, we will assume that the regularization term J in the variational imaging model (8) satisfies the following assumptions:

(A1) $J \in \Gamma_0(\mathbb{R}^n)$,
(A2) $\mathrm{int}\ (\mathrm{dom}\ J) \neq \varnothing$,
(A3) $\inf_{\boldsymbol{y}\in\mathbb{R}^n} J(\boldsymbol{y}) \in \mathbb{R}$, and without loss of generality, $\inf_{\boldsymbol{y}\in\mathbb{R}^n} J(\boldsymbol{y}) = 0$.

Assumption (A1) ensures that the minimal value of the convex imaging problem (8) and its minimizer (2) are well-defined and enjoy several properties (see Sect. 2, Proposition 2.2). Assumption (A2) ensures that for every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, the posterior distribution

$$\mathbb{R}^n \ni \boldsymbol{y} \mapsto \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}} \tag{25}$$

and its associated partition function

$$\mathbb{R}^n \times (0, +\infty) \times (0, +\infty) \ni (\boldsymbol{x}, t, \epsilon) \mapsto Z_J(\boldsymbol{x}, t, \epsilon) = \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y} \tag{26}$$

are well-defined, and finally, Assumption (A3) guarantees that the partition function (26) is also bounded from above independently of $\boldsymbol{x} \in \mathbb{R}^n$. We will denote the posterior expectation (with respect to the posterior distribution (25)) of a measurable function $f : \Omega \mapsto \mathbb{R}$ with $\Omega \subset \mathrm{dom}\ f$ integrable on the set $\mathrm{dom}\ f \cap \mathrm{dom}\ J$ by

$$\mathbb{E}_J[f(\boldsymbol{y})] = \frac{1}{Z_J(\boldsymbol{x}, t, \epsilon)} \int_{\Omega\cap\mathrm{dom}\ J} f(\boldsymbol{y}) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}. \tag{27}$$

Posterior expectations of vector quantities are defined similarly component-wise. Posterior expectations generally depend on $(\boldsymbol{x}, t, \epsilon)$, but we will omit writing this dependence explicitly.

3.2 Connections to Viscous Hamilton–Jacobi Partial Differential Equations

The next proposition establishes connections between viscous HJ PDEs with initial data J satisfying Assumptions (A1)–(A3) and both the partition function (26) and the Bayesian posterior mean estimate (3). These connections mirror those between the first-order HJ PDE (20) with initial data J satisfying assumption (A1) and both the convex minimization problem (8) and the MAP estimate (2). The connections between viscous HJ PDEs and Bayesian posterior mean estimators will be leveraged later to describe several properties of posterior mean estimators in terms of the observed image $\boldsymbol{x}$ and parameters t and ϵ, and in particular in Sect. 3.3 to show that the posterior mean estimate (3) can be expressed as the minimizer associated with the solution to a first-order HJ PDE (Proposition 3.2) with at least twice continuously differentiable and convex regularization term.

Proposition 3.1 (The viscous Hamilton–Jacobi equation with initial data in $\Gamma_0(\mathbb{R}^n)$) *Suppose the function* J *satisfies assumptions (A1)–(A3). Then, the following statements hold.*

(i) *(Cole–Hopf transformation, [26] Section 4.4.1) For every* $\epsilon > 0$, *the function* $S_\epsilon : \mathbb{R}^n \times [0, +\infty) \to [0, +\infty)$ *defined by*

$$S_\epsilon(\boldsymbol{x}, t) := -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}} Z_J(\boldsymbol{x}, t, \epsilon)\right) = -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}\right) \tag{28}$$

is the unique smooth solution to the viscous HJ PDE with initial data

$$\begin{cases} \frac{\partial S_\epsilon}{\partial t}(\boldsymbol{x},t) + \frac{1}{2}\left\|\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)\right\|_2^2 = \frac{\epsilon}{2}\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) & \text{in } \mathbb{R}^n \times (0,+\infty), \\ S_\epsilon(\boldsymbol{x},0) = J(\boldsymbol{x}) & \text{in } \mathbb{R}^n. \end{cases} \tag{29}$$

In addition, the domain of integration in (3.1) can be taken to be dom J *or, up to a set of Lebesgue measure zero,* int (dom J) *or* dom (∂J). *Furthermore, for every* $\boldsymbol{x} \in$ dom J *and* $\epsilon > 0$, *except possibly at the boundary points* $\boldsymbol{x} \in$ (dom J) $\setminus$ int (dom J) *if such points exist, the pointwise limit* $S_\epsilon(\boldsymbol{x},t)$ *as* $t \to 0$ *exists and satisfies*

$$\lim_{\substack{t\to 0\\ t>0}} S_\epsilon(\boldsymbol{x},t) = J(\boldsymbol{x}).$$

(ii) *(Convexity and monotonicity properties).*

(a) *The function* $\mathbb{R}^n \times (0,+\infty) \ni (\boldsymbol{x},t) \mapsto S_\epsilon(\boldsymbol{x},t) - \frac{n\epsilon}{2}\ln t$ *is jointly convex.*
(b) *The function* $(0,+\infty) \ni t \mapsto S_\epsilon(\boldsymbol{x},t) - \frac{n\epsilon}{2}\ln t$ *is strictly monotone decreasing.*
(c) *The function* $(0,+\infty) \ni \epsilon \mapsto S_\epsilon(\boldsymbol{x},t) - \frac{n\epsilon}{2}\ln \epsilon$ *is strictly monotone decreasing.*
(d) *The function* $\mathbb{R}^n \ni \boldsymbol{x} \mapsto \frac{1}{2}\|\boldsymbol{x}\|_2^2 - tS_\epsilon(\boldsymbol{x},t)$ *is strictly convex.*

(iii) *(Connections to the posterior mean and mean squared error) The posterior mean estimate* $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ *and the mean squared error* $\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$ *satisfy the formulas*

$$\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) \tag{30}$$

and

$$\begin{aligned} \mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] &= t\epsilon\nabla_{\boldsymbol{x}} \cdot \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \\ &= nt\epsilon - t^2\epsilon\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t). \end{aligned} \tag{31}$$

Moreover, $\boldsymbol{x} \mapsto \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ *is a bijective function.*

(iv) *(Vanishing* $\epsilon \to 0$ *limit) Let* $S_0 : \mathbb{R}^n \times (0,+\infty) \to \mathbb{R}$ *denote the continuously differentiable and convex solution to the first-order HJ PDE (20) with initial data* J. *For every* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $t > 0$, *the following limit holds:*

$$\begin{aligned} &\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}\right) \\ &= \inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\}, \end{aligned} \tag{32}$$

that is,

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) = S_0(\boldsymbol{x},t),$$

and the limit converges uniformly over every compact set of $\mathbb{R}^n \times (0,+\infty)$ *in* $(\boldsymbol{x},t)$. *In addition, the gradient* $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$, *the partial derivative* $\frac{\partial S_\epsilon(\boldsymbol{x},t)}{\partial t}$, *and the Laplacian* $\frac{\epsilon}{2}\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$ *satisfy the limits*

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t), \quad \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\partial S_\epsilon}{\partial t}(\boldsymbol{x},t) = \frac{\partial S_0}{\partial t}(\boldsymbol{x},t),$$

and

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\epsilon}{2}\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = 0,$$

where each limit converges uniformly over every compact set of $\mathbb{R}^n \times (0,+\infty)$ *in* $(\boldsymbol{x},t)$. *As a consequence, for every* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $t > 0$, *the pointwise limit of* $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ *as* $\epsilon \to 0$ *exists and satisfies*

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \boldsymbol{u}_{MAP}(\boldsymbol{x},t),$$

and the limit converges uniformly over every compact set of $\mathbb{R}^n \times (0,+\infty)$ *in* $(\boldsymbol{x},t)$.

Proof See "Appendix A" for the proof. □

To illustrate certain aspects of Proposition 3.1 and properties of posterior mean estimates, we give here two analytical examples.

Example 3.1 (Tikhonov–Phillips regularization) Let $J(\boldsymbol{x}) = \frac{m}{2}\|\boldsymbol{x}\|_2^2$ with $m > 0$, and consider the solution $S_0(\boldsymbol{x},t)$ and $S_\epsilon(\boldsymbol{x},t)$ to the first-order PDE (20) and viscous HJ PDE (29) with initial data J, respectively.

The solution $S_0(\boldsymbol{x},t)$ is given by the Lax–Oleinik formula (Proposition 2.2, Eq. (22))

$$\begin{aligned} S_0(\boldsymbol{x},t) &= \inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{m}{2}\|\boldsymbol{y}\|_2^2\right\} \\ &= \frac{m\|\boldsymbol{x}\|_2^2}{2(1+mt)}. \end{aligned}$$

This minimization problem is a special case of Tikhonov–Phillips regularization (also known as ridge regression in statistics), a method for regularizing ill-posed problems in inverse problems and statistics using a quadratic regularization term [55,64]. The corresponding minimizer can be computed using the gradient $\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t)$ via equation (23) in Proposition 3.1:

$$\boldsymbol{u}_{MAP}(\boldsymbol{x},t) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t) = \boldsymbol{x} - \frac{mt\boldsymbol{x}}{1+mt} = \frac{\boldsymbol{x}}{1+mt}.$$

The solution $S_\epsilon(\boldsymbol{x}, t)$ is given by the integral

$$\begin{aligned}S_\epsilon(\boldsymbol{x}, t) &= -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+\frac{m}{2}\|\boldsymbol{y}\|_2^2\right)/\epsilon} d\boldsymbol{y}\right)\\ &= \frac{m\,\|\boldsymbol{x}\|^2}{2(1+mt)} + \frac{n\epsilon}{2}\ln(1+mt).\end{aligned}$$

The posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ can be computed using the representation formula (30) in Proposition 3.1(iii) by calculating the gradient $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$:

$$\begin{aligned}\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t) = \boldsymbol{x} - \frac{mt\boldsymbol{x}}{1+mt}\\ = \frac{\boldsymbol{x}}{1+mt}.\end{aligned}$$

The mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)\|_2^2\right]$ can be computed using the representation formula (31) in Proposition 3.1(iii) by calculating the divergence of $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$:

$$\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)\|_2^2\right] = t\epsilon\nabla_{\boldsymbol{x}} \cdot \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \frac{nt\epsilon}{1+mt}. \tag{33}$$

Comparing the solutions $S_0(\boldsymbol{x}, t)$ and $S_\epsilon(\boldsymbol{x}, t)$, we see that $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x}, t) = S_0(\boldsymbol{x}, t)$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$, in accordance with the result established in Proposition 3.1(iv). Note also that while $(\boldsymbol{x}, t) \mapsto S_0(\boldsymbol{x}, t)$ is jointly convex, its viscous counterpart $(\boldsymbol{x}, t) \mapsto S_\epsilon(\boldsymbol{x}, t)$ is not. Indeed, $t \mapsto S_\epsilon(\boldsymbol{x}, t)$ is not convex, and it is convex only after subtracting $\frac{n\epsilon}{2}\ln t$ from $S_\epsilon(\boldsymbol{x}, t)$.

Example 3.2 (*Soft thresholding*) Let $J(\boldsymbol{x}) = \sum_{i=1}^n \lambda_i |\boldsymbol{x}_i|$, where $\lambda_i > 0$ for each $i \in \{1, \ldots, n\}$, and consider the solutions $S_0(\boldsymbol{x}, t)$ and $S_\epsilon(\boldsymbol{x}, t)$ to the first-order (20) and viscous HJ PDEs (29) with initial data J, respectively.

The solution $S_0(\boldsymbol{x}, t)$ is given by the Lax–Oleinik formula

$$\begin{aligned}S_0(\boldsymbol{x}, t) &= \inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \sum_{i=1}^n \lambda_i |y_i|\right\}\\ &= \sum_{i=1}^n\left(\inf_{y_i\in\mathbb{R}}\left\{\frac{1}{2t}(x_i-y_i)^2 + \lambda_i|y_i|\right\}\right),\end{aligned}$$

where x_i and y_i denote the i^{th} component of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. In the context of imaging, this minimization problem corresponds to denoising an image with the weighted sum of a quadratic fidelity term and a weighted l_1-norm as the regularization term. This term is widely used in imaging to encourage sparsity of an image, and it has received considerable interest due to its connection with compressed sensing reconstruction [10,24]. The solution to this minimization problem corresponds to a soft thresholding applied component-wise to the vector $\boldsymbol{x}$ [20,29,47]. The soft thresholding operator is defined for any real number a and positive real number α as

$$\mathbb{R}\times(0,+\infty) \ni (a,\alpha) \mapsto T(a,\alpha) = \begin{cases} a-\alpha & \text{if } a > \alpha,\\ 0 & \text{if } a \in [-\alpha,\alpha],\\ a+\alpha & \text{if } a < -\alpha.\end{cases} \tag{34}$$

The minimizer in the Lax–Oleinik formula of $S_0(\boldsymbol{x}, t)$ is then given component-wise for $i \in \{1, \ldots, n\}$ by

$$(\boldsymbol{u}_{MAP}(\boldsymbol{x}, t))_i = T(x_i, t\lambda_i),$$

so that

$$S_0(\boldsymbol{x}, t) = \sum_{i=1}^n\left(\frac{1}{2t}(x_i - T(x_i, t\lambda_i))^2 + \lambda_i|T(x_i, t\lambda_i)|\right).$$

The solution $S_\epsilon(\boldsymbol{x}, t)$ is given by the integral

$$\begin{aligned}S_\epsilon(\boldsymbol{x}, t) &= -\epsilon\ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+\sum_{k=1}^n\lambda_i|y_i|\right)/\epsilon} d\boldsymbol{y}\right)\\ &= -\epsilon\sum_{i=1}^n\ln\left(\frac{1}{2}\sqrt{\frac{2}{\pi t\epsilon}}\int_{-\infty}^{+\infty} e^{-\left(\frac{1}{2t}(x_i-y_i)^2+\lambda_i|y_i|\right)/\epsilon} dy_i\right)\\ &= -\epsilon\sum_{i=1}^n\ln\left(\frac{1}{2}\sqrt{\frac{2}{\pi t\epsilon}}\left(\int_0^{+\infty} e^{-\left(\frac{1}{2t}(x_i+y_i)^2+\lambda_i y_i\right)/\epsilon} dy_i\right.\right.\\ &\quad\left.\left. + \int_0^{+\infty} e^{-\left(\frac{1}{2t}(x_i-y_i)^2+\lambda_i y_i\right)/\epsilon} dy_i\right)\right)\end{aligned}$$

To compute this integral, first define the function

$$\mathbb{R} \ni z \mapsto L(z) = \frac{1}{2}e^{z^2}\operatorname{erfc}(z),$$

where erfc denotes the complementary error function. Then, we have ([33], page 336, integral 3.332, 2., and page 887, integral 8.250, 1.)

$$\frac{1}{2}\sqrt{\frac{2}{\pi t\epsilon}}\int_0^{+\infty} e^{-\left(\frac{1}{2t}(x_i+y_i)^2+\lambda_i y_i\right)/\epsilon} dy_i = e^{-\frac{x_i^2}{2t\epsilon}} L\left(\frac{x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right)$$

and

$$\frac{1}{2}\sqrt{\frac{2}{\pi t\epsilon}}\int_0^{+\infty} e^{-\left(\frac{1}{2t}(x_i-y_i)^2+\lambda_i y_i\right)/\epsilon} dy_i = e^{-\frac{x_i^2}{2t\epsilon}} L\left(\frac{-x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right),$$

from which we get

$$S_\epsilon(\boldsymbol{x}, t) = \frac{\|\boldsymbol{x}\|_2^2}{2t} - \epsilon \sum_{i=1}^{n} \ln\left(L\left(\frac{x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right) + L\left(\frac{-x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right)\right).$$

Now, to find the posterior mean estimate it suffices to compute the gradient of $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$ and use formula (30). To do so, we need the derivative of the function L. Since

$$\frac{dL}{dz}(z) = 2zL(z) + \frac{1}{\sqrt{\pi}},$$

the chain rule gives

$$\begin{aligned}\frac{\partial}{\partial x_i}&\left(L\left(\frac{x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right) + L\left(\frac{-x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right)\right) \\ &= \left(\frac{x_i + t\lambda_i}{t\epsilon}\right) L\left(\frac{x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right) \\ &\quad - \left(\frac{-x_i + t\lambda_i}{t\epsilon}\right) L\left(\frac{-x_i + t\lambda_i}{\sqrt{2t\epsilon}}\right).\end{aligned}$$

The posterior mean estimate is therefore given componentwise by

$$\begin{aligned}(\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon))_i &= x_i - t(\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t))_i \\ &= x_i + t\lambda_i \left(\frac{L\left(\frac{x_i+t\lambda_i}{\sqrt{2t\epsilon}}\right) + L\left(\frac{-x_i+t\lambda_i}{\sqrt{2t\epsilon}}\right)}{L\left(\frac{x_i+t\lambda_i}{\sqrt{2t\epsilon}}\right) - L\left(\frac{-x_i+t\lambda_i}{\sqrt{2t\epsilon}}\right)}\right)\end{aligned}$$

The posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ yields a smooth analogue of the soft thresholding operator T (defined in (34)) evaluated at $(x_i, t\lambda_i)$, in the sense that $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}}(\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon))_i = T(x_i, t\lambda_i)$ for every $i \in \{1, \ldots, n\}$ by Proposition 3.1(iv). Figure 3 shows the MAP and posterior mean estimates in one dimension for the choice of $t = 1.25$, $\epsilon = \{0.025, 0.1, 0.25, 0.5, 1\}$, and $\lambda_1 = 2$ for $x \in [-5, 5]$.

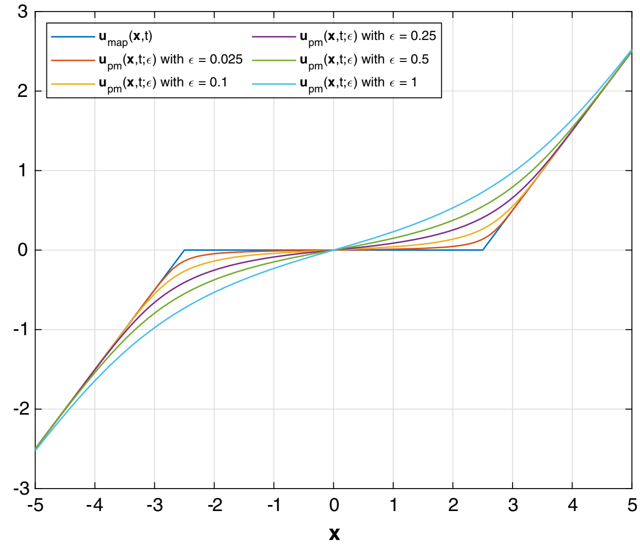


Fig. 3 Numerical example of the MAP and posterior mean estimates in one dimension with $J(x) = \lambda_1 |x|$ for the choice of $t = 1.25$, $\epsilon = \{0.025, 0.1, 0.25, 0.5, 1\}$, and $\lambda_1 = 2$ for $x \in [-5, 5]$

3.3 Connections to First-Order Hamilton–Jacobi Equations

In this section, we use the connections between the posterior mean estimate (3) and viscous HJ PDEs established in Proposition 3.1 to show that the posterior mean estimate can be expressed through the solution to a first-order HJ PDE with initial data of the form of (20). In particular, we show that the posterior mean estimate satisfies the proximal mapping formula

$$\begin{aligned}&\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) \\ &\quad = \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2\right)\right\},\end{aligned}$$

where the function $K_\epsilon : \mathbb{R}^n \times \times (0, +\infty) \to \mathbb{R}$ is defined through the solution $S_\epsilon(\boldsymbol{x}, t)$ to the viscous HJ PDE (29) via

$$\begin{aligned}&K_\epsilon(\boldsymbol{x}, t) \\ &\quad := \frac{1}{2}\|\boldsymbol{x}\|_2^2 - tS_\epsilon(\boldsymbol{x}, t) \\ &\quad \equiv t\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\mathrm{dom}\, J} e^{\left(\frac{1}{t}\langle \boldsymbol{x}, \boldsymbol{y}\rangle - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}\right),\end{aligned}$$

which is convex by Proposition 3.1(ii)(d), and where $K_\epsilon^*(\boldsymbol{y}, t)$ denotes the Fenchel–Legendre transform of $\boldsymbol{y} \mapsto K_\epsilon(\boldsymbol{y}, t)$. This result gives the representation of the convex imaging regularization term whose existence was derived by [34,35,48,49] (and later extended to non-quadratic data fidelity terms in [36,37]). This representation result depends crucially on the connections established between the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ and the viscous HJ PDE (29) established in Proposition 3.1. Moreover, we also show that $\boldsymbol{y} \mapsto K_\epsilon^*(\boldsymbol{y}, t)$ is at least twice continuously differentiable. This fact implies that the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ for image denoising does not suffer from staircasing effects thanks to a result established in [51, Theorem 3] as proven for Total Variation regularization terms in [48]. Here, our results are applicable to any regularization term J satisfying assumptions (A1)–(A3).

Proposition 3.2 (Connections between the posterior mean estimate and first-order HJ PDEs) *Suppose the function J*

satisfies assumptions (A1)-(A3). For every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, let $S_\epsilon(\boldsymbol{x}, t)$ denote the solution to the viscous HJ PDE (29) with initial data J and let $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ denote the posterior mean estimate (3). Consider the first-order HJ PDE

$$\begin{cases} \frac{\partial \tilde{S}}{\partial s}(\boldsymbol{x}, s) + \frac{1}{2}\left\|\nabla_{\boldsymbol{x}} \tilde{S}(\boldsymbol{x}, s)\right\|_2^2 = 0 & \text{in } \mathbb{R}^n \times (0, +\infty), \\ \tilde{S}(\boldsymbol{x}, 0) = K_\epsilon^*(\boldsymbol{x}, t) - \frac{1}{2}\|\boldsymbol{x}\|_2^2 & \text{in } \mathbb{R}^n. \end{cases} \quad (35)$$

Then, the initial data $\boldsymbol{x} \mapsto K_\epsilon^*(\boldsymbol{x}, t) - \frac{1}{2}\|\boldsymbol{x}\|_2^2$ is convex, the solution to the HJ PDE (35) satisfies the Lax–Oleinik formula

$$\tilde{S}(\boldsymbol{x}, s) = \inf_{\boldsymbol{y} \in \mathbb{R}^n} \left\{ \frac{1}{2s}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2 \right) \right\},$$

and the corresponding minimizer at $s = 1$ is the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$:

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \arg\min_{\boldsymbol{y} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2 \right) \right\}. \quad (36)$$

Moreover, for every $t > 0$ and $\epsilon > 0$ the function $\mathbb{R}^n \ni \boldsymbol{y} \mapsto K_\epsilon^*(\boldsymbol{y}, t)$ is at least twice continuously differentiable.

Proof By definition of the function $(\boldsymbol{x}, t) \mapsto K_\epsilon(\boldsymbol{x}, t)$, we can write

$$tS_\epsilon(\boldsymbol{x}, t) + K_\epsilon(\boldsymbol{x}, t) = \frac{1}{2}\|\boldsymbol{x}\|_2^2.$$

As both $\boldsymbol{x} \mapsto tS_\epsilon(\boldsymbol{x}, t)$ and $\boldsymbol{x} \mapsto K_\epsilon(\boldsymbol{x}, t)$ are convex by Proposition 3.1(ii)(a) and (d), we can apply Proposition 2.1 in Sect. 2 to conclude that $\boldsymbol{x} \mapsto K_\epsilon^*(\boldsymbol{x}, t) - \frac{1}{2}\|\boldsymbol{x}\|_2^2$ is convex and to express $tS_\epsilon(\boldsymbol{x}, t)$ as

$$tS_\epsilon(\boldsymbol{x}, t) = \inf_{\boldsymbol{y} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2 \right) \right\} \quad (37)$$

On the one hand, by Proposition 2.2 the right hand side of (37) is the solution $\tilde{S}_0(\boldsymbol{x}, s)$ to the first-order HJ PDE (35) at $s = 1$, and therefore its minimizer is given by $\boldsymbol{x} - \nabla_{\boldsymbol{x}} \tilde{S}_0(\boldsymbol{x}, 1)$. On the other hand, the gradient $\nabla_{\boldsymbol{x}} \tilde{S}_0(\boldsymbol{x}, 1)$ is equal to the left hand side of (37), that is, $\nabla_{\boldsymbol{x}} \tilde{S}_0(\boldsymbol{x}, 1) = t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$, which is equal to $\boldsymbol{x} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ by formula (3). As a result, the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ minimizes the right hand side of (37), that is,

$$\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \arg\min_{\boldsymbol{y} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \left(K_\epsilon^*(\boldsymbol{y}, t) - \frac{1}{2}\|\boldsymbol{y}\|_2^2 \right) \right\}.$$

Now, using the strict convexity of $\boldsymbol{x} \mapsto K_\epsilon(\boldsymbol{x}, t)$ and that $\nabla K_\epsilon(\boldsymbol{x}, t) = \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ is a bijective function in $\boldsymbol{x}$ for every $t > 0$ and $\epsilon > 0$ by Proposition 3.1(iii) we can invoke [58, Theorem 26.5] to conclude that $\boldsymbol{y} \mapsto K_\epsilon^*(\boldsymbol{y}, t)$ is a continuously differentiable, strictly convex, and bijective function on $\mathbb{R}^n$, and moreover that $\boldsymbol{y} \mapsto \nabla_{\boldsymbol{y}} K_\epsilon^*(\boldsymbol{y}, t)$ corresponds to the inverse of $\boldsymbol{x} \mapsto \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$, i.e., $\nabla_{\boldsymbol{y}} K_\epsilon^*(\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon), t) = \boldsymbol{x}$. Finally, as $\boldsymbol{x} \mapsto K_\epsilon(\boldsymbol{x}, t)$ is twice differentiable and strictly convex on $\mathbb{R}^n$, the inverse function theorem [26, Appendix C, Theorem 7] implies that $\boldsymbol{y} \mapsto \nabla_{\boldsymbol{y}} K_\epsilon^*(\boldsymbol{y}, t)$ is continuously differentiable on $\mathbb{R}^n$, whence $\boldsymbol{y} \mapsto K_\epsilon(\boldsymbol{y}, t)$. □

4 Properties of Posterior Mean and MAP Estimators

In this section, we describe various properties of the Bayesian posterior mean estimate (3) in terms of the data $\boldsymbol{x} \in \mathbb{R}^n$, parameters $t > 0$ and $\epsilon > 0$, and the imaging regularization term J. Specifically, in Sect. 4.1, we derive topological, representation, and monotonicity properties of the posterior mean estimate, which we use in Sect. 4.2 to further derive an optimal upper bound on the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)\|_2^2\right]$, an estimate of the squared difference between the MAP and posterior mean estimates, monotonicity and non-expansiveness properties of the posterior mean estimate, and the behavior of the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ in the limit $t \to 0$. Finally, we describe the MAP and posterior mean estimates in terms of Bayes risks and their connections to HJ PDEs in Sect. 4.3.

4.1 Topological, Representation, and Monotonicity Properties

This section describes the topological, representation, and monotonicity properties of the Bayesian posterior mean estimate (3), which are stated, respectively, in Propositions 4.1, 4.2, and 4.3.

The first result, Proposition 4.1, states that the posterior mean estimate belongs in the interior of the domain of J for all data $\boldsymbol{x} \in \mathbb{R}^n$ and parameters $t > 0$ and $\epsilon > 0$.

Proposition 4.1 (Topological properties) *Suppose that the function J satisfies assumptions (A1)–(A3). Then, the following properties hold.*

(i) *For every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon)$ is contained in* int (dom J).
(ii) *Let $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, and let $S_\epsilon \colon \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ denote the solution to the viscous HJ PDEs (29) with initial data J. Then, the expected value*

of the initial data $\mathbb{E}_J\left[J(\boldsymbol{y})\right]$ *satisfies the bounds*

$$\begin{aligned} 0 &\le J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) \le \mathbb{E}_J\left[J(\boldsymbol{y})\right] \\ &< \epsilon\left(e^{S_\epsilon(\boldsymbol{x},t)/\epsilon}-1\right) < +\infty. \end{aligned} \tag{38}$$

Proof See "Appendix B". □

The second result, Proposition 4.2, gives representation formulas for the posterior mean estimate. In particular, when the regularization term J satisfies assumptions (A1)–(A3) and dom $J = \mathbb{R}^n$, the posterior mean estimate and mean squared error then satisfy representation formulas in terms of the mean minimal subgradient of J given by $\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]$. These representation formulas are then used to show that when dom $J \neq \mathbb{R}^n$, the posterior mean estimate can nonetheless be approximated using the first-order HJ PDE (20) by smoothing the initial value J via a Moreau–Yosida approximation $S_0(\boldsymbol{x},\mu)$ with $\mu > 0$.

Proposition 4.2 (Representation properties) *Suppose that the function* J *satisfies assumptions (A1)–(A3), let* $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, *and* $\epsilon > 0$, *and let* $(\boldsymbol{x},t) \mapsto S_0(\boldsymbol{x},t)$ *and* $(\boldsymbol{x},t) \mapsto S_\epsilon(\boldsymbol{x},t)$ *denote the solutions, respectively, to the first-order and viscous HJ PDEs (20) and (29) with initial data* J.

(i) *(Representation formulas) If* dom $J = \mathbb{R}^n$, *then* $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right] < +\infty$, *for every* $\boldsymbol{y}_0 \in \mathbb{R}^n$ *we have*

$$\mathbb{E}_J\left[\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)+\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\, \boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right] = n\epsilon, \tag{39}$$

and the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ *and mean squared error* $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$ *of the Bayesian posterior distribution (25) satisfy the representation formulas*

$$\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \boldsymbol{x} - t\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right] \tag{40}$$

and

$$\begin{aligned} &\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\ &\quad = nt\epsilon - t\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\, \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]. \end{aligned} \tag{41}$$

Moreover, the gradient $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$ *and Laplacian* $\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$ *satisfy the representation formulas*

$$\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right] \tag{42}$$

and

$$\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \frac{1}{t\epsilon}\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\, \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]. \tag{43}$$

(ii) *(Limit formulas) Let* $\{\mu_k\}_{k=1}^{+\infty}$ *be a sequence of positive real numbers decreasing to zero. The solution* $S_\epsilon(\boldsymbol{x},t)$ *to the viscous HJ PDE (29) and its gradient* $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$ *satisfy the limits*

$$\begin{aligned} &S_\epsilon(\boldsymbol{x},t) \\ &:= -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\,d\boldsymbol{y}\right) \\ &= \lim_{k\to+\infty} -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}\right) \end{aligned} \tag{44}$$

and

$$\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \lim_{k\to+\infty}\left(\frac{\int_{\mathbb{R}^n}\nabla_{\boldsymbol{y}} S_0(\boldsymbol{y},\mu_k) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}\right). \tag{45}$$

In particular, the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ *satisfies the limits*

$$\begin{aligned} &\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \\ &= \lim_{k\to+\infty}\left(\frac{\int_{\mathbb{R}^n}\boldsymbol{y} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}\right) \\ &= \boldsymbol{x} - t\lim_{k\to+\infty}\left(\frac{\int_{\mathbb{R}^n}\nabla_{\boldsymbol{y}} S_0(\boldsymbol{y},\mu_k) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}\,d\boldsymbol{y}}\right), \end{aligned} \tag{46}$$

Proof See "Appendix C" for the proof. □

Remark 4.1 Note that the representation formulas in Proposition 4.2(i) may not hold if dom $J \neq \mathbb{R}^n$. To see this, consider $J : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ defined by

$$J(\boldsymbol{y}) = \begin{cases} 0, & \text{if } \|\boldsymbol{y}\|_2 \le 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

The domain of J is the unit sphere in $\mathbb{R}^n$, which is convex, and J satisfies assumptions (A1)–(A3). The function J is continuously differentiable on int (dom J), with $\nabla J(\boldsymbol{y}) = \boldsymbol{0}$ for every $\boldsymbol{y} \in$ int (dom J). Clearly, $\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right] = 0$. However, for every $\boldsymbol{x} \neq \boldsymbol{0}$, the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \neq \boldsymbol{x}$. Hence, the representation formula (40) does not hold in that case.

The next result, Proposition 4.3, uses the properties of solutions to first-order HJ PDEs presented in Proposition 2.2 together with the representation formulas (40) and (41) to describe monotonicity properties of the posterior mean estimate. Proposition 4.3 will be leveraged in the next subsection to derive an optimal upper bound for the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$ and several estimates and limit results of $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in terms of the observed image $\boldsymbol{x}$ and parameter $t > 0$.

For the statement and proof of Proposition 4.3, and later for Proposition 4.5, we define the function

$$\mathrm{dom}\,\partial J \ni \boldsymbol{y} \mapsto \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),$$

which is a subgradient of the convex function $\boldsymbol{y} \ni \boldsymbol{y} \mapsto \frac{1}{2t}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|_2^2 + J(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathrm{dom}\,\partial J$.

Proposition 4.3 (Monotonicity property) *Suppose that the function J is strongly convex of parameter $m \geq 0$ and satisfies assumptions (A1)–(A3). Let $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$. Then, for every $\boldsymbol{y}_0 \in \mathrm{dom}\,\partial J$,*

$$\begin{aligned}&\left(\frac{1+mt}{t}\right)\mathbb{E}_J\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2\right]\\ &\leq \mathbb{E}_J\left[\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right]\\ &\leq n\epsilon - \left\langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{y}_0\right\rangle. \end{aligned} \tag{47}$$

Moreover, the mean Euclidean norm of the minimal subgradient of J is finite, i.e., $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right] < +\infty$.

Proof See "Appendix D" for the proof. □

4.2 Error Bounds and Limit Properties

In this section, we derive an optimal bound for the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$, a bound on the squared difference between the MAP and posterior mean estimates, monotonicity and non-expansiveness properties of the posterior mean estimate, and limiting results of the posterior mean estimate in terms of the parameters t.

Proposition 4.4 (Error Bounds and limit properties) *Suppose that the function J is strongly convex of parameter $m \geq 0$ and satisfies assumptions (A1)–(A3).*

(i) *For every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$ of the Bayesian posterior distribution (25) satisfies the upper bound*

$$\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \leq \frac{nt\epsilon}{1+mt}. \tag{48}$$

(ii) *For every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$, the squared difference between the MAP and posterior mean estimates satisfies the upper bound*

$$\|\boldsymbol{u}_{MAP}(\boldsymbol{x},t) - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2 \leq \frac{nt\epsilon}{1+mt}. \tag{49}$$

(iii) *The posterior mean estimate is monotone and non-expansive, that is, for every $\boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^n$, $t > 0$, and $\epsilon > 0$,*

$$\langle \boldsymbol{u}_{PM}(\boldsymbol{x}+\boldsymbol{d},t,\epsilon) - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon), \boldsymbol{d}\rangle \geq 0 \tag{50}$$

and

$$\|\boldsymbol{u}_{PM}(\boldsymbol{x}+\boldsymbol{d},t,\epsilon) - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 \leq \|\boldsymbol{d}\|_2. \tag{51}$$

(iv) *Let $\{t_k\}_{k=1}^{+\infty}$ be a sequence of positive real numbers converging to 0 and let $\{\boldsymbol{d}_k\}_{k=1}^{+\infty}$ be a sequence of elements of $\mathbb{R}^n$ converging to $\boldsymbol{d} \in \mathbb{R}^n$. Then, for every $\boldsymbol{x} \in \mathrm{dom}\,J$ and $\epsilon > 0$, the pointwise limit of $\boldsymbol{u}_{PM}(\boldsymbol{x}+t_k\boldsymbol{d}_k, t_k, \epsilon)$ as $k \to +\infty$ exists and satisfies*

$$\lim_{k\to+\infty} \boldsymbol{u}_{PM}(\boldsymbol{x}+t_k\boldsymbol{d}_k, t_k, \epsilon) = \boldsymbol{x}.$$

Proof **Proof of (i):** Since $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \mathrm{int}\,(\mathrm{dom}\,J)$ by Proposition 4.1 and $\mathrm{int}\,(\mathrm{dom}\,J) \subset \mathrm{dom}\,\partial J$ (see Definition 5), we can set $\boldsymbol{y}_0 = \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in the monotonicity inequality (47) in Proposition 4.3(i) and rearrange to get the upper bound (48).

Proof of (ii): Note that for every $\boldsymbol{y}_0 \in \mathrm{dom}\,\partial J$, the monotonicity inequality (47) in Proposition 4.3 yields

$$\mathbb{E}_J\left[\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t} + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right] \leq n\epsilon.$$

Choose $\boldsymbol{y}_0 = \boldsymbol{u}_{MAP}(\boldsymbol{x},t)$, which for every $\boldsymbol{x}$ and $t > 0$ is always an element of $\mathrm{dom}\,\partial J$ and also satisfies the inclusion $\left(\frac{\boldsymbol{x}-\boldsymbol{u}_{MAP}(\boldsymbol{x},t)}{t}\right) \in \partial J(\boldsymbol{u}_{MAP}(\boldsymbol{x},t))$ by part (ii) of Proposition 2.2. Hence, the monotonicity of the subdifferential of $\boldsymbol{y} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})$ and strong convexity of J of parameter $m \geq 0$ implies

$$\begin{aligned}&\left(\frac{1+mt}{t}\right)\|\boldsymbol{y}-\boldsymbol{u}_{MAP}(\boldsymbol{x},t)\|_2^2\\ &\leq \left\langle\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{t} + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right), \boldsymbol{y}-\boldsymbol{u}_{MAP}(\boldsymbol{x},t)\right\rangle.\end{aligned}$$

Combine these inequalities to get $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{MAP}(\boldsymbol{x},t)\|_2^2\right] \leq \frac{nt\epsilon}{1+mt}$, and use the convexity of the Euclidean norm to get inequality (49).

Proof of (iii): The convexity of $\boldsymbol{x} \mapsto K_\epsilon(\boldsymbol{x},t)$ by Proposition 3.1(ii)(d) and $\nabla_{\boldsymbol{x}}K_\epsilon(\boldsymbol{x},t) = \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ implies the monotonicity property (50) (see definition 5, equation (13), and [58], page 240 and Corollary 31.5.2). Since both functions $\boldsymbol{x} \mapsto S_\epsilon(\boldsymbol{x},t)$ and $\boldsymbol{x} \mapsto \frac{1}{2}\|\boldsymbol{x}\|_2^2 - tS_\epsilon(\boldsymbol{x},t)$ are convex by Proposition 3.1(ii)(a) and (d), the gradient of the function $\boldsymbol{x} \mapsto \frac{1}{2}\|\boldsymbol{x}\|_2^2 - tS_\epsilon(\boldsymbol{x},t)$, whose value is the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ by Proposition 3.1(iii), is Lipschitz continuous with unit constant (see [69] for a simple proof), that is,

$$\begin{aligned}&\|(\boldsymbol{x}+\boldsymbol{d}-t\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x}+\boldsymbol{d},t)) - (\boldsymbol{x}-t\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t))\|_2\\ &\equiv \|\boldsymbol{u}_{PM}(\boldsymbol{x}+\boldsymbol{d},t,\epsilon) - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 \leq \|\boldsymbol{d}\|_2,\end{aligned}$$

which proves inequality (51).

Proof of (iv): Inequality (49) and the triangle inequality imply

$$\begin{aligned}&\|(\boldsymbol{x}+t_k\boldsymbol{d}_k)-\boldsymbol{u}_{PM}(\boldsymbol{x}+t_k\boldsymbol{d}_k,t_k,\epsilon)\|_2\\&\quad\leq\|(\boldsymbol{x}+t_k\boldsymbol{d}_k)-\boldsymbol{u}_{MAP}(\boldsymbol{x}+t_k\boldsymbol{d}_k,t_k)\|_2+\sqrt{\frac{nt_k\epsilon}{1+mt}}.\end{aligned}$$

The limit $\lim_{k\to+\infty}\boldsymbol{u}_{PM}(\boldsymbol{x}+t_k\boldsymbol{d}_k,t_k,\epsilon)=\boldsymbol{x}$ then follows by Proposition 2.2(iii). □

Remark 4.2 The upper bound for the mean squared error in (48) is optimal. As shown in Example 3.1, it is attained for the quadratic term $J(\boldsymbol{x})=\frac{m}{2}\|\boldsymbol{x}\|_2^2$.

4.3 Bayesian Risks and Hamilton–Jacobi Partial Differential Equations

In this section, we will consider the Bayesian risk associated with the following Bregman divergence (see Definition 7)

$$\begin{aligned}&\mathbb{R}^n\times\mathbb{R}^n\ni(\boldsymbol{u},\boldsymbol{y})\\&\mapsto\begin{cases}D_{\Phi_J}(\boldsymbol{u},\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)) & \text{if } \boldsymbol{y}\in\operatorname{dom}\partial J,\\+\infty & \text{otherwise,}\end{cases}\end{aligned}\tag{52}$$

where

$$\begin{aligned}&\operatorname{dom}\partial J\ni\boldsymbol{y}\mapsto\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)=\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)+\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\\&\mathbb{R}^n\ni\boldsymbol{y}\mapsto\Phi_J(\boldsymbol{y}|\boldsymbol{x},t)=\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y}).\end{aligned}$$

The associated Bayesian risk to the posterior distribution (25) corresponds to the expected value $\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u},\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$. We refer the reader to [4] and [44] for discussions on Bregman loss functions and Bayesian estimation theory.

Here, we will use the connections between maximum a posteriori and posterior mean estimates and Hamilton–Jacobi equations derived in Sect. 3 to show that when the regularization term J is convex on $\mathbb{R}^n$ and bounded from below, then the MAP estimate $\boldsymbol{u}_{MAP}(\boldsymbol{x},t)$ minimizes in expectation the Bregman loss function (52). We also show that when dom $J\neq\mathbb{R}^n$ and satisfies assumptions (A1)–(A3). The results rely on the monotonicity property (47) established in Proposition (4.3).

Proposition 4.5 (Bregman divergences) *Suppose that the function J satisfies assumptions (A1)–(A3), and let $\boldsymbol{x}\in\mathbb{R}^n$, $t>0$, and $\epsilon>0$.*

(i) *The mean Bregman loss function* dom $J\ni\boldsymbol{u}\mapsto\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u},\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]\in\mathbb{R}$ *has a unique minimizer* $\bar{\boldsymbol{u}}\in\operatorname{dom}\partial J$ *that satisfies the inclusion*

$$\left(\frac{\boldsymbol{x}-\bar{\boldsymbol{u}}}{t}\right)\in\partial J(\bar{\boldsymbol{u}})+\left(\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)-\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]\right),\tag{53}$$

where addition in (53) is taken in the sense of sets.

(ii) *If J is finite everywhere on $\mathbb{R}^n$, then the MAP estimate $\boldsymbol{u}_{MAP}(\boldsymbol{x},t)$ is the unique global minimizer of the Bregman loss function $\mathbb{R}^n\ni\boldsymbol{u}\mapsto\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u},\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]\in\mathbb{R}$, that is,*

$$\boldsymbol{u}_{MAP}(\boldsymbol{x},t)=\underset{\boldsymbol{u}\in\mathbb{R}^n}{\arg\min}\,\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u},\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]\tag{54}$$

Proof See "Appendix E" for the proof. □

5 Conclusion

In this paper, we presented novel theoretical connections between Hamilton–Jacobi partial differential equations and a broad class of Bayesian posterior mean estimators with quadratic data fidelity term and log-concave prior relevant to image denoising problems. We derived a representation formula for the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in terms of the spatial gradient of the solution to a viscous HJ PDE with initial data corresponding to the convex regularization term J. We used these connections to show that the posterior mean estimate can be expressed through the gradient of the solution to a first-order HJ PDE with twice continuously differentiable convex initial data, and furthermore, we derived a novel representation formula for this initial data which, to our knowledge, was not available in the literature.

The connections between HJ PDEs and Bayesian posterior mean estimators were further used to establish several topological, representation, and monotonicity properties of posterior mean estimates. These properties were then used to derive an optimal upper bound on the mean squared error $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$, an estimate of the squared difference between the MAP and posterior mean estimates, monotonicity and non-expansiveness properties of the posterior mean estimate, and the behavior of the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in the limit $t\to0$.

Finally, we used the connections between both MAP and posterior mean estimates and HJ PDEs to show that the MAP estimate (2) corresponds to the Bayes estimator of the Bayesian risk (52) whenever the regularization term J is convex on $\mathbb{R}^n$ and bounded from below and the data fidelity term is quadratic. We also show that when dom $J\neq\mathbb{R}^n$, the Bayesian risk (52) has still a Bayes estimator that is described in terms of the solution to both the first-order HJ PDE (2.2) and the viscous HJ PDE (3.1).

We wish to note that in addition to its relevance to image denoising problems, the viscous HJ PDE (29) has recently received some attention in the deep learning literature, where its solution $\boldsymbol{x} \mapsto S_\epsilon(\boldsymbol{x}, t)$ is known as the local entropy loss function and is a loss regularization effective at training deep networks [15,16,31,65]. While this paper focuses on HJ PDEs and Bayesian estimators in imaging sciences, the results in this paper may be relevant to the deep learning literature and may give new theoretical understandings of the local entropy loss function in terms of the data $\boldsymbol{x}$ and parameters t and ϵ.

The results presented in this work crucially depend on the data fidelity term being quadratic and the generalized prior distribution $\boldsymbol{y} \mapsto e^{-J(\boldsymbol{y})}$ being log-concave. This paper did not consider non-quadratic data fidelity terms (corresponding to non-Gaussian additive noise models) with log-concave priors, or non-additive noise models [5,7].

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

A Proof of Proposition 3.1

We will use the following lemma, which characterizes the partition function (26) in terms of the solution to a Cauchy problem involving the heat equation with initial data $J \in \Gamma_0(\mathbb{R}^n)$, to prove parts (i) and (ii)(a)–(d) of Proposition 3.1.

Lemma A.1 (The heat equation with initial data in $\Gamma_0(\mathbb{R}^n)$) *Suppose the function $J\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfies assumptions (A1)–(A3).*

(i) *For every $\epsilon > 0$, the function $w_\epsilon\colon \mathbb{R}^n \times [0, +\infty) \to (0, 1]$ defined by*

$$w_\epsilon(\boldsymbol{x}, t) := \frac{1}{(2\pi t\epsilon)^{n/2}} Z_J(\boldsymbol{x}, t, \epsilon) = \frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y} \quad (55)$$

is the unique smooth solution to the Cauchy problem

$$\begin{cases} \frac{\partial w_\epsilon}{\partial t}(\boldsymbol{x}, t) = \frac{\epsilon}{2}\Delta_{\boldsymbol{x}} w_\epsilon(\boldsymbol{x}, t) & \text{in } \mathbb{R}^n \times (0, +\infty), \\ w_\epsilon(\boldsymbol{x}, 0) = e^{-J(\boldsymbol{x})/\epsilon} & \text{in } \mathbb{R}^n. \end{cases} \quad (56)$$

In addition, the domain of integration of the integral (55) can be taken to be dom J *or, up to a set of Lebesgue measure zero,* int (dom J) *or* dom ∂J. *Furthermore, for every $\boldsymbol{x} \in \mathbb{R}^n$ and $\epsilon > 0$, except possibly at the points $\boldsymbol{x} \in$* (dom J) $\setminus$ (int dom J) *if such points exist, the pointwise limit of $w_\epsilon(\boldsymbol{x}, t)$ as $t \to 0$ exists and satisfies*

$$\lim_{\substack{t \to 0 \\ t > 0}} w_\epsilon(\boldsymbol{x}, t) = e^{-J(\boldsymbol{x})/\epsilon},$$

with the limit equal to 0 whenever $\boldsymbol{x} \notin$ dom J.

(ii) *(Log-concavity and monotonicity properties).*

 (a) *The function $\mathbb{R}^n \times (0, +\infty) \ni (\boldsymbol{x}, t) \mapsto t^{n/2} w_\epsilon(\boldsymbol{x}, t)$ is jointly log-concave.*
 (b) *The function $(0, +\infty) \ni t \mapsto t^{n/2} w_\epsilon(\boldsymbol{x}, t)$ is strictly monotone increasing.*
 (c) *The function $(0, +\infty) \ni \epsilon \mapsto \epsilon^{n/2} w_\epsilon(\boldsymbol{x}, t)$ is strictly monotone increasing.*
 (d) *The function $\mathbb{R}^n \ni \boldsymbol{x} \mapsto e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}\|_2^2} w_\epsilon(\boldsymbol{x}, t)$ is strictly log-convex.*

The proof of (i) follows from classical PDEs arguments for the Cauchy problem (56) tailored to the initial data $(\boldsymbol{x}, \epsilon) \mapsto e^{-J(\boldsymbol{x})/\epsilon}$ with J satisfying assumptions (A1)–(A3), and the proof of log-concavity and monotonicity (ii)(a)–(d) follows from the Prékopa–Leindler and Hölder's inequalities [30,46,56]; we present the details below.

Proof **Proof of Lemma A.1 (i):** This result follows directly from the theory of convolution of Schwartz distributions ([42], Chapter 2, Sect. 2.1, Chapter 4, Sect. 4.2 and 4.4., and in particular Theorem 4.4.1 on page 110). To see why this is the case, note that by assumptions (A1)–(A3) the initial condition $\boldsymbol{y} \mapsto e^{-J(\boldsymbol{y})}$ is a locally integrable function, and locally integrable functions are Schwartz distributions.

Proof of Lemma A.1 (ii)(a): The log-concavity property will be shown using the Prékopa–Leindler inequality.

Theorem A.1 (Prékopa–Leindler inequality [46,56]) *Let f, g, and h be non-negative real-valued and measurable functions on $\mathbb{R}^n$, and suppose*

$$h(\lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2) \geq f(\boldsymbol{y}_1)^\lambda g(\boldsymbol{y}_2)^{(1-\lambda)}$$

for every $\boldsymbol{y}_1$, $\boldsymbol{y}_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. Then,

$$\int_{\mathbb{R}^n} h(\boldsymbol{y}) d\boldsymbol{y} \geq \left(\int_{\mathbb{R}^n} f(\boldsymbol{y}) d\boldsymbol{y}\right)^\lambda \left(\int_{\mathbb{R}^n} g(\boldsymbol{y}) d\boldsymbol{y}\right)^{(1-\lambda)}.$$

Proof of Lemma A.1 (ii)(a) (continued): Let $\epsilon > 0$, $\lambda \in (0, 1)$, $\boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$, $\boldsymbol{y} = \lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2$, and $t = \lambda t_1 + (1-\lambda)t_2$ for any $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$, $\boldsymbol{y}_2 \in \mathbb{R}^n$ and t_1, $t_2 \in (0, +\infty)$. The joint convexity of the function $\mathbb{R}^n \times (0, +\infty) \ni (\boldsymbol{z}, t) \mapsto \frac{1}{2t}\|\boldsymbol{z}\|_2^2$ and convexity of J imply

$$\frac{1}{2t}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + J(\boldsymbol{y})$$

$$\leq \frac{\lambda}{2t_1}\left\|\boldsymbol{x}_1 - \boldsymbol{y}_1\right\|_2^2 + \frac{(1-\lambda)}{2t_2}\left\|\boldsymbol{x}_2 - \boldsymbol{y}_2\right\|_2^2 + \lambda J(\boldsymbol{y}_1) + (1-\lambda)J(\boldsymbol{y}_2),$$

This gives

$$\frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}} \geq \left(\frac{e^{-\left(\frac{1}{2t_1}\|\boldsymbol{x}_1-\boldsymbol{y}_1\|_2^2+J(\boldsymbol{y}_1)\right)/\epsilon}}{(2\pi\epsilon)^{n/2}}\right)^{\lambda} \left(\frac{e^{-\left(\frac{1}{2t_2}\|\boldsymbol{x}_2-\boldsymbol{y}_2\|_2^2+J(\boldsymbol{y}_2)\right)/\epsilon}}{(2\pi\epsilon)^{n/2}}\right)^{1-\lambda}.$$

Applying the Prékopa–Leindler inequality with

$$h(\boldsymbol{y}) = \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}},$$
$$f(\boldsymbol{y}) = \frac{e^{-\left(\frac{1}{2t_1}\|\boldsymbol{x}_1-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}},$$

and

$$g(\boldsymbol{y}) = \frac{e^{-\left(\frac{1}{2t_2}\|\boldsymbol{x}_2-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}},$$

and using the definition (55) of $w_\epsilon(\boldsymbol{x}, t)$, we get

$$t^{n/2}w_\epsilon(\boldsymbol{x}, t) \geq \left(t_1^{n/2}w_\epsilon(\boldsymbol{x}_1, t_1)\right)^{\lambda}\left(t_2^{n/2}w_\epsilon(\boldsymbol{x}_2, t_2)\right)^{(1-\lambda)},$$

As a result, the function $(\boldsymbol{x}, t) \mapsto t^{n/2}w_\epsilon(\boldsymbol{x}, t)$ is jointly log-concave on $\mathbb{R}^n \times (0, +\infty)$.

Proof of Lemma A.1 (ii)(b): Since $t \mapsto \frac{1}{t}$ is strictly monotone decreasing on $(0, +\infty)$, then for every $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \operatorname{dom} J$, $\epsilon > 0$, and $0 < t_1 < t_2$,

$$\frac{e^{-\left(\frac{1}{2t_1}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}} < \frac{e^{-\left(\frac{1}{2t_2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{(2\pi\epsilon)^{n/2}}$$

whenever $\boldsymbol{x} \neq \boldsymbol{y}$. Integrating both sides of the inequality with respect to $\boldsymbol{y}$ over $\operatorname{dom} J$ yields

$$\frac{1}{(2\pi\epsilon)^{n/2}}\int_{\operatorname{dom} J} e^{-\left(\frac{1}{2t_1}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}d\boldsymbol{y} < \frac{1}{(2\pi\epsilon)^{n/2}}\int_{\operatorname{dom} J} e^{-\left(\frac{1}{2t_2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}d\boldsymbol{y},$$

As a result, the function $t \mapsto t^{n/2}w_\epsilon(\boldsymbol{x}, t)$ is strictly monotone increasing on $(0, +\infty)$.

Proof of Lemma A.1 (ii)(c): Since $\epsilon \mapsto \frac{1}{\epsilon}$ is strictly monotone decreasing on $(0, +\infty)$ and $\operatorname{dom} J \ni \boldsymbol{y} \mapsto J(\boldsymbol{y})$ is non-negative by assumption (A3), then for every $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, and $0 < \epsilon_1 < \epsilon_2$ we have

$$e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon_1} < e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon_2}$$

whenever $\boldsymbol{x} \neq \boldsymbol{y}$. Integrating both sides of the inequality with respect to $\boldsymbol{y}$ over $\operatorname{dom} J$ yields

$$\int_{\operatorname{dom} J} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon_1}d\boldsymbol{y} < \int_{\operatorname{dom} J} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon_2}d\boldsymbol{y},$$

As a result, the function $\epsilon \mapsto \epsilon^{n/2}w_\epsilon(\boldsymbol{x}, t)$ is strictly monotone increasing on $(0, +\infty)$.

Proof of Lemma A.1 (ii)(d): Let $\epsilon > 0, t > 0, \lambda \in (0, 1)$, $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ with $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ and $\boldsymbol{x} = \lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$. Then,

$$\begin{aligned} e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}\|_2^2}w_\epsilon(\boldsymbol{x}, t) &= \frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\operatorname{dom} J} e^{\left(\langle \boldsymbol{x},\boldsymbol{y}\rangle/t - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})\right)/\epsilon}d\boldsymbol{y} \\ &= \int_{\operatorname{dom} J}\left(\frac{e^{\left(\langle \boldsymbol{x}_1,\boldsymbol{y}\rangle/t - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})\right)/\epsilon}}{(2\pi t\epsilon)^{n/2}}\right)^{\lambda}\left(\frac{e^{\langle \boldsymbol{x}_2,\boldsymbol{y}\rangle/t\epsilon - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})/\epsilon}}{(2\pi t\epsilon)^{n/2}}\right)^{1-\lambda}d\boldsymbol{y}. \end{aligned}$$

Hölder's inequality [30, Theorem 6.2] then implies

$$\begin{aligned} e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}\|_2^2}w_\epsilon(\boldsymbol{x}, t) &\leq \left(\int_{\operatorname{dom} J}\frac{e^{\left(\langle \boldsymbol{x}_1,\boldsymbol{y}\rangle/t - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})\right)/\epsilon}}{(2\pi t\epsilon)^{n/2}}d\boldsymbol{y}\right)^{\lambda}\left(\int_{\operatorname{dom} J}\frac{e^{\langle \boldsymbol{x}_2,\boldsymbol{y}\rangle/t\epsilon - \frac{1}{2t}\|\boldsymbol{y}\|_2^2 - J(\boldsymbol{y})/\epsilon}}{(2\pi t\epsilon)^{n/2}}d\boldsymbol{y}\right)^{1-\lambda} \\ &= \left(e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}_1\|_2^2}w_\epsilon(\boldsymbol{x}_1, t)\right)^{\lambda}\left(e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}_2\|_2^2}w_\epsilon(\boldsymbol{x}_2, t)\right)^{1-\lambda}, \end{aligned}$$

where the inequality in the equation above is an equality if and only if there exists a constant $\alpha \in \mathbb{R}$ such that $\alpha e^{\langle \boldsymbol{x}_1,\boldsymbol{y}\rangle/t\epsilon} = e^{\langle \boldsymbol{x}_x,\boldsymbol{y}\rangle/t\epsilon}$ for almost every $\boldsymbol{y} \in \operatorname{dom} J$. This does not hold here since $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$. As a result, the function $\mathbb{R}^n \ni \boldsymbol{x} \mapsto e^{\frac{1}{2t\epsilon}\|\boldsymbol{x}\|_2^2}w_\epsilon(\boldsymbol{x}, t)$ is strictly log-convex. □

Proof of Proposition 3.1 (i) and (ii)(a)–(d): The proof of these statements follows from Lemma A.1 and classic results about the Cole–Hopf transform (see, e.g., [26, Section 4.4.1]), with $S_\epsilon(\boldsymbol{x},t) := -\epsilon \log(w_\epsilon(\boldsymbol{x},t))$.

Proof of Proposition 3.1 (iii): The formulas follow from a straightforward calculation of the gradient, divergence, and Laplacian of $S_\epsilon(\boldsymbol{x},t)$ that we omit here. Since the function $\boldsymbol{x} \mapsto \frac{1}{2}\|\boldsymbol{x}\|_2^2 - tS_\epsilon(\boldsymbol{x},t)$ is strictly convex, we can invoke (Corollary 26.3.1, [58]) to conclude that its gradient $\boldsymbol{x} \mapsto \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$, which gives the posterior mean $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$, is bijective.

Proof of Proposition 3.1 (iv): We will prove this result in three steps. First, we will show that

$$\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \leq \inf_{\boldsymbol{y}\in \mathrm{int}\ (\mathrm{dom}\ J)} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\}$$

and

$$\inf_{\boldsymbol{y}\in \mathrm{int}\ (\mathrm{dom}\ J)} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} = \inf_{\boldsymbol{y}\in \mathrm{dom}\ J} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} \equiv S_0(\boldsymbol{x},t).$$

Next, we will show that $\liminf_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \geq S_0(\boldsymbol{x},t)$. Finally, we will use steps 1 and 2 to conclude that $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) = S_0(\boldsymbol{x},t)$. Pointwise and local uniform convergence of the gradient $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t)$, the partial derivative $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\partial S_\epsilon(\boldsymbol{x},t)}{\partial t} = \frac{\partial S_0(\boldsymbol{x},t)}{\partial t}$, and the Laplacian $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\epsilon}{2}\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = 0$ then follow from the convexity and differentiability of the solutions $(\boldsymbol{x},t)\mapsto S_0(\boldsymbol{x},t)$ and $(\boldsymbol{x},t)\mapsto S_\epsilon(\boldsymbol{x},t)$ to the HJ PDEs (20) and (29).

In what follows, we will use the following large deviation principle result [22]: For every Lebesgue measurable set $\mathcal{A} \in \mathbb{R}^n$,

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathcal{A}} e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y}\right) = \operatorname*{ess\,inf}_{\boldsymbol{y}\in\mathcal{A}} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right\},$$

where

$$\operatorname*{ess\,inf}_{\boldsymbol{y}\in\mathcal{A}} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right\} = \sup\left\{a\in\mathbb{R}\colon a \leq \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2, \text{for a.e. } \boldsymbol{y}\in\mathcal{A}\right\}.$$

Step 1. (Adapted from Deuschel and Stroock [22], Lemma 2.1.7.) By convexity, the function J is continuous for every $\boldsymbol{y}_0 \in \mathrm{int}\ (\mathrm{dom}\ J)$, the latter set being open. Therefore, for every such $\boldsymbol{y}_0$ there exists a number $r_{\boldsymbol{y}_0} > 0$ such that for every $0 < r \leq r_{\boldsymbol{y}_0}$ the open ball $B_r(\boldsymbol{y}_0)$ is contained in int (dom J). Hence,

$$\begin{aligned} S_\epsilon(\boldsymbol{x},t) &:= -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathrm{int}\ (\mathrm{dom}\ J)} e^{-(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y}))/\epsilon} d\boldsymbol{y}\right)\\ &\leq -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{B_r(\boldsymbol{y}_0)} e^{-(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y}))/\epsilon} d\boldsymbol{y}\right)\\ &\leq -\epsilon \ln\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{B_r(\boldsymbol{y}_0)} e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y}\right)\\ &\quad + \sup_{\boldsymbol{y}\in B_r(\boldsymbol{y}_0)} J(\boldsymbol{y}). \end{aligned}$$

Take $\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}}$ and apply the large deviation principle to the term on the right to get

$$\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \leq \operatorname*{ess\,inf}_{\boldsymbol{y}\in B_r(\boldsymbol{y}_0)} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right\} + \sup_{\boldsymbol{y}\in B_r(\boldsymbol{y}_0)} J(\boldsymbol{y}).$$

Take $\lim_{r\to 0}$ on both sides of the inequality to find

$$\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \leq \frac{1}{2t}\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2^2 + J(\boldsymbol{y}_0).$$

Since the inequality holds for every $\boldsymbol{y}_0 \in \mathrm{int}\ (\mathrm{dom}\ J)$, we can take the infimum over all $y \in \mathrm{int}\ (\mathrm{dom}\ J)$ on the right-hand-side of the inequality to get

$$\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \leq \inf_{\boldsymbol{y}\in \mathrm{int}\ (\mathrm{dom}\ J)} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\}. \tag{57}$$

By assumptions (A1) and (A2) that $J \in \Gamma_0(\mathbb{R}^n)$ and int (dom J) $\neq \varnothing$, the infimum on the right hand side is equal to that taken over dom J [58, Corollary 7.3.2], i.e.,

$$\inf_{\boldsymbol{y}\in \mathrm{int}\ (\mathrm{dom}\ J)} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} = \inf_{\boldsymbol{y}\in \mathrm{dom}\ J} \left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right\} \equiv S_0(\boldsymbol{x},t). \tag{58}$$

We combine (57) and (58) to obtain

$$\limsup_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \leq S_0(\boldsymbol{x},t),$$

which is the desired result.

Step 2. We can invoke Lemma 2.1.8 in [22] because its conditions are satisfied (in the notation of [22], $\Phi = -J$, which is upper semicontinuous, $\boldsymbol{y} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$ is the rate function, and note that the tail condition (2.1.9) is satisfied in that $\sup_{\boldsymbol{y}\in\mathbb{R}^n} -J(\boldsymbol{y}) = -\inf_{\boldsymbol{y}\in\mathbb{R}^n} J(\boldsymbol{y}) = 0$ by assumption

(A3)) to get

$$\liminf_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) \geq S_0(\boldsymbol{x},t).$$

Step 3. Combining the two limits derived in steps 1 and 2 yields

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} S_\epsilon(\boldsymbol{x},t) = S_0(\boldsymbol{x},t)$$

for every $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$, where the limit converges uniformly on every compact subset $(\boldsymbol{x},t)$ of $\mathbb{R}^n \times (0,+\infty)$ [58, Theorem 10.8].

By differentiability and joint convexity of both $\mathbb{R}^n \times (0,+\infty) \ni (\boldsymbol{x},t) \mapsto S_0(\boldsymbol{x},t)$ and $\mathbb{R}^n \times (0,+\infty) \ni (\boldsymbol{x},t) \mapsto S_\epsilon(\boldsymbol{x},t) - \frac{n\epsilon}{2}\ln t$ (Proposition 2.2 (i), and Proposition 3.1 (i) and (ii)(a)), we can invoke [58, Theorem 25.7] to get

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t) \text{ and } \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \left(\frac{\partial S_\epsilon(\boldsymbol{x},t)}{\partial t} - \frac{n\epsilon}{2t}\right)$$
$$= \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\partial S_\epsilon(\boldsymbol{x},t)}{\partial t} = \frac{\partial S_0(\boldsymbol{x},t)}{\partial t},$$

for every $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$, where the limit converges uniformly on every compact subset of $\mathbb{R}^n \times (0,+\infty)$. Furthermore, the viscous HJ PDE (29) for S_ϵ implies that

$$\begin{aligned}\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{\epsilon}{2}\Delta_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) &= \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \left(\frac{\partial S_\epsilon(\boldsymbol{x},t)}{\partial t} + \frac{1}{2}\|\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)\|^2\right),\\ &= \left(\frac{\partial S_0(\boldsymbol{x},t)}{\partial t} + \frac{1}{2}\|\nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t)\|^2\right)\\ &= 0,\end{aligned}$$

where the last equality holds thanks to the HJ PDE (20) (see Proposition 2.2). Here, again, the limit holds for every $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$, and the limit converges uniformly over any compact subset of $\mathbb{R}^n \times (0,+\infty)$. Finally, the limit $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \boldsymbol{u}_{MAP}(\boldsymbol{x},t)$ holds directly as a consequence to the limit $\lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x},t)$ and the representation formulas (30) (see Proposition 3.1(iii)) and (23) (see Proposition 2.2(ii))for the posterior mean and MAP estimates, respectively.

B Proof of Proposition 4.1

Proof of (i): We will prove that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{int}(\operatorname{dom} J)$ in two steps. First, we will use the projection operator (10) (see Definition 4) and the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ to prove by contradiction that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{cl}(\operatorname{dom} J)$. Second, we will use the following variant of the Hahn–Banach theorem for convex bodies in $\mathbb{R}^n$ to show in fact that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{int}(\operatorname{dom} J)$.

Theorem B.1 [58, Theorem 11.6 and Corollary 11.6.2] *Let C be a convex set. A point $\boldsymbol{u} \in C$ is a relative boundary point of C if and only if there exist a vector $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and a number $b \in \mathbb{R}$ such that*

$$\boldsymbol{u} = \arg\max_{\boldsymbol{y}\in C} \{\langle \boldsymbol{a},\boldsymbol{y}\rangle + b\},$$

with $\langle \boldsymbol{a},\boldsymbol{y}\rangle + b < \langle \boldsymbol{a},\boldsymbol{u}\rangle + b$ for every $\boldsymbol{y} \in \operatorname{int}(C)$.

Step 1. Suppose $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \notin \operatorname{cl}(\operatorname{dom} J)$. Since the set $\operatorname{cl}(\operatorname{dom} J)$ is closed and convex, the projection of $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ onto $\operatorname{cl}(\operatorname{dom} J)$ given by $\pi_{\operatorname{cl}(\operatorname{dom} J)}(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) \equiv \bar{\boldsymbol{u}}$ is well-defined and unique (see Definition 4), with $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \neq \bar{\boldsymbol{u}}$ by assumption. The projection $\bar{\boldsymbol{u}}$ also satisfies the characterization (11), namely

$$\langle \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}, \boldsymbol{y} - \bar{\boldsymbol{u}}\rangle \leq 0$$

for every $\boldsymbol{y} \in \operatorname{cl}(\operatorname{dom} J)$. Then, by linearity of the posterior mean estimate,

$$\begin{aligned}\|\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}\|_2^2 &= \langle \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}, \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}\rangle\\ &= \left\langle \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}, \mathbb{E}_J\left[\boldsymbol{y}\right] - \bar{\boldsymbol{u}}\right\rangle\\ &= \mathbb{E}_J\left[\langle \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \bar{\boldsymbol{u}}, \boldsymbol{y} - \bar{\boldsymbol{u}}\rangle\right]\\ &\leq 0,\end{aligned}$$

which implies that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \bar{\boldsymbol{u}}$. This contradicts the assumption that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \notin \operatorname{cl}(\operatorname{dom} J)$. Hence, it follows that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{cl}(\operatorname{dom} J)$.

Step 2. We now wish to prove that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{int}(\operatorname{dom} J)$. Note that this inclusion trivially holds if there are no boundary points, i.e., $(\operatorname{cl}(\operatorname{dom} J) \setminus \operatorname{int}(\operatorname{dom} J)) = \varnothing$. Now, we consider the case $(\operatorname{cl}(\operatorname{dom} J) \setminus \operatorname{int}(\operatorname{dom} J)) \neq \varnothing$. Suppose that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in (\operatorname{cl}(\operatorname{dom} J) \setminus \operatorname{int}(\operatorname{dom} J))$. Then, Thm .B.1 applies and there exist a vector $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and a number $b \in \mathbb{R}$ such that

$$\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) = \arg\max_{\boldsymbol{y}\in\operatorname{cl}(\operatorname{dom} J)} \{\langle \boldsymbol{a},\boldsymbol{y}\rangle + b\},$$

with $\langle \boldsymbol{a},\boldsymbol{y}\rangle + b < \langle \boldsymbol{a},\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle + b$ for every $\boldsymbol{y} \in \operatorname{int}(\operatorname{dom} J)$. By linearity of the posterior mean estimate,

$$\begin{aligned}\langle \boldsymbol{a},\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle + b &= \left\langle \boldsymbol{a}, \mathbb{E}_J\left[\boldsymbol{y}\right]\right\rangle + b\\ &= \mathbb{E}_J\left[\langle \boldsymbol{a},\boldsymbol{y}\rangle + b\right]\\ &< \mathbb{E}_J\left[\langle \boldsymbol{a},\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle + b\right]\\ &= \langle \boldsymbol{a},\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle + b,\end{aligned}$$

where the strict inequality in the third line follows from integrating over int (dom J). This contradicts the assumption that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in (\text{cl}\,(\text{dom}\,J) \setminus \text{int}\,(\text{dom}\,J))$. Hence, $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \text{int}\,(\text{dom}\,J)$.

Proof of (ii): First, as a consequence that $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \text{int}\,(\text{dom}\,J)$, the subdifferential of J at $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ is non-empty because the subdifferential ∂J is non-empty at every point $\boldsymbol{y} \in \text{int}\,(\text{dom}\,J)$ [58, Theorem 23.4]. Hence, there exists a subgradient $\boldsymbol{p} \in \partial J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon))$ such that

$$J(\boldsymbol{y}) \geq J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - \langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\,. \tag{59}$$

Take the expectation $\mathbb{E}_J[\cdot]$ on both sides of inequality (59) to find

$$\begin{aligned}
&\mathbb{E}_J\left[J(\boldsymbol{y})\right] \\
&\quad\geq \mathbb{E}_J\left[J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - \langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right] \\
&\quad= J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - \mathbb{E}_J\left[\langle \boldsymbol{p}, \boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right] \\
&\quad= J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - \left\langle \boldsymbol{p}, \mathbb{E}_J\left[\boldsymbol{y}\right] - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle \\
&\quad= J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - \langle \boldsymbol{p}, \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle \\
&\quad= J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)).
\end{aligned} \tag{60}$$

This gives the lower bound of inequality (38).

Second, use the convex inequality $1+z \leq e^z$ that holds on $\mathbb{R}$ with $z \equiv J(\boldsymbol{y})/\epsilon$ for $\boldsymbol{y} \in \text{dom}\,J$. This gives the inequality $1 + \frac{1}{\epsilon}J(\boldsymbol{y}) \leq e^{J(\boldsymbol{y})/\epsilon}$. Multiply this inequality by $e^{-J(\boldsymbol{y})/\epsilon}$ and subtract by $e^{-J(\boldsymbol{y})/\epsilon}$ on both sides to find

$$\frac{1}{\epsilon}J(\boldsymbol{y})e^{-J(\boldsymbol{y})/\epsilon} \leq (1 - e^{-J(\boldsymbol{y})/\epsilon}). \tag{61}$$

Multiply both sides by $e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}$, divide by the partition function $Z_J(\boldsymbol{x},t,\epsilon)$ (see Eq. (26)), integrate with respect to $\boldsymbol{y} \in \text{dom}\,J$, and use

$$\begin{aligned}
&\frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\text{dom}\,J}\frac{1}{\epsilon}J(\boldsymbol{y})e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}d\boldsymbol{y} \\
&\quad= \frac{1}{\epsilon}\mathbb{E}_J\left[J(\boldsymbol{y})\right]
\end{aligned}$$

to obtain

$$\begin{aligned}
\frac{1}{\epsilon}\mathbb{E}_J\left[J(\boldsymbol{y})\right] \leq{}& \frac{1}{Z_J(\boldsymbol{x},t,\epsilon)} \\
&\int_{\text{dom}\,J}\left(e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon} - e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\right)d\boldsymbol{y}.
\end{aligned} \tag{62}$$

Now, we can bound the right hand side of (62) as follows

$$\begin{aligned}
&\frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\text{dom}\,J}\left(e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon} - e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\right)d\boldsymbol{y} \\
&= \frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\text{dom}\,J}e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon}d\boldsymbol{y} \\
&\quad - \frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\text{dom}\,J}e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}d\boldsymbol{y} \\
&\leq \frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\mathbb{R}^n}e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon}d\boldsymbol{y} - 1 \\
&= \frac{(2\pi t\epsilon)^{n/2}}{Z_J(\boldsymbol{x},t,\epsilon)} - 1.
\end{aligned} \tag{63}$$

Combining (62) and (63), we get

$$\mathbb{E}_J\left[J(\boldsymbol{y})\right] \leq \epsilon\left(\frac{(2\pi t\epsilon)^{n/2}}{Z_J(\boldsymbol{x},t,\epsilon)} - 1\right). \tag{64}$$

Using the representation formula (28) for the solution $(\boldsymbol{x},t) \mapsto S_\epsilon$ to the viscous HJ PDE (29), we have that $(2\pi t\epsilon)^{n/2}/Z_J(\boldsymbol{x},t,\epsilon) = e^{S_\epsilon(\boldsymbol{x},t)/\epsilon}$. We can therefore write (64) as follows

$$\mathbb{E}_J\left[J(\boldsymbol{y})\right] \leq \epsilon\left(e^{S_\epsilon(\boldsymbol{x},t)/\epsilon} - 1\right) < +\infty.$$

Combining the latter inequalities with (60), we obtain the desired set of inequalities (38).

C Proof of Proposition 4.2

Proof of (i): We will show that $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right] < +\infty$ and derive formulas (39), (40), (41), (42), and (43) in four steps. To describe these steps, let us first introduce some notation. Recall that J satisfies assumptions (A1)–(A3) and dom $J = \mathbb{R}^n$. Define the set

$$D_J := \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \partial J(\boldsymbol{y}) = \{\nabla J(\boldsymbol{y})\}\right\}.$$

We can invoke [58, Theorem 25.5] to conclude that D_J is a dense subset of $\mathbb{R}^n$, the n-dimensional Lebesgue measure of the set $(\mathbb{R}^n \setminus D_J)$ is zero, and the function $\boldsymbol{y} \mapsto \nabla J(\boldsymbol{y})$ is continuous on D_J. Now, let $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, $\epsilon > 0$, and $\boldsymbol{y}_0 \in \mathbb{R}^n$. Define the function $\varphi_J : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\varphi_J(\boldsymbol{y}|\boldsymbol{x},t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}).$$

Note that for every $\boldsymbol{y} \in \mathbb{R}^n$ we have $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t) \in \partial\left(\mathbb{R}^n \ni \boldsymbol{u} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u})\right)(\boldsymbol{y})$, i.e., $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)$ is a subgradient of the function $\boldsymbol{u} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u})$ evaluated at $\boldsymbol{u} = \boldsymbol{y}$. Let

$$C_1(\boldsymbol{x},\boldsymbol{y}_0,t,\epsilon) = \int_{\mathbb{R}^n}\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2 e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}d\boldsymbol{y}, \tag{65}$$

and note that by assumption (A3), the expected value $\mathbb{E}_J\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right]$ is bounded as follows

$$\begin{aligned}\mathbb{E}_J&\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right]\\&=\frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\mathbb{R}^n}\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2 e^{-(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y}))/\epsilon}d\boldsymbol{y}\\&\leq\frac{1}{Z_J(\boldsymbol{x},t,\epsilon)}\int_{\mathbb{R}^n}\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2 e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}d\boldsymbol{y}\\&=\frac{C_1(\boldsymbol{x},\boldsymbol{y}_0,t,\epsilon)}{Z_J(\boldsymbol{x},t,\epsilon)}.\end{aligned}\tag{66}$$

Define the vector field $\boldsymbol{V}\colon\mathbb{R}^n\to\mathbb{R}^n$ as

$$\boldsymbol{V}(\boldsymbol{y})=(\boldsymbol{y}-\boldsymbol{y}_0)e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon},\tag{67}$$

which is continuous on $\mathbb{R}^n$. It is also bounded on $\mathbb{R}^n$; to see this, use the triangle inequality, assumption (A3), and the fact that the function $(0,+\infty)\ni r\mapsto re^{-\frac{1}{2t\epsilon}r^2}$ attains its maximum at $r^*=\sqrt{t\epsilon}$ to get

$$\begin{aligned}\|\boldsymbol{V}(\boldsymbol{y})\|_2&=\left\|(\boldsymbol{y}-\boldsymbol{y}_0)\right\|_2 e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&=\left\|(\boldsymbol{y}-\boldsymbol{x}+\boldsymbol{x}-\boldsymbol{y}_0)\right\|_2 e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&\leq(\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2+\|\boldsymbol{x}-\boldsymbol{y}\|_2)e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&\leq\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2+\|\boldsymbol{x}-\boldsymbol{y}\|_2 e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&\leq\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2+\|\boldsymbol{x}-\boldsymbol{y}\|_2 e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}\\&\leq\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2+\sup_{\boldsymbol{y}\in\mathbb{R}^n}\left(\|\boldsymbol{x}-\boldsymbol{y}\|_2 e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}\right)\\&\leq\left\|\boldsymbol{x}-\boldsymbol{y}_0\right\|_2+(\sqrt{t\epsilon})e^{-\frac{\sqrt{t\epsilon}}{2}}.\end{aligned}\tag{68}$$

The divergence $\nabla_{\boldsymbol{y}}\cdot\boldsymbol{V}(\boldsymbol{y})$, which is well-defined and continuous on D_J, is given for every $\boldsymbol{y}\in D_J$ by

$$\begin{aligned}&\nabla_{\boldsymbol{y}}\cdot\boldsymbol{V}(\boldsymbol{y})\\&=\nabla_{\boldsymbol{y}}\cdot\left((\boldsymbol{y}-\boldsymbol{y}_0)e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\right)\\&=(\nabla_{\boldsymbol{y}}\cdot(\boldsymbol{y}-\boldsymbol{y}_0))e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&\quad+\left\langle\nabla_{\boldsymbol{y}}e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon},\boldsymbol{y}-\boldsymbol{y}_0\right\rangle\\&=ne^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}\\&\quad-\left\langle\frac{1}{\epsilon}\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}+\nabla J(\boldsymbol{y})\right)e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon},\boldsymbol{y}-\boldsymbol{y}_0\right\rangle\\&=\left(n-\left\langle\frac{1}{\epsilon}\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}+\nabla J(\boldsymbol{y})\right),\boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right)e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}.\end{aligned}\tag{69}$$

We now outline the four steps that will be used to prove Proposition 4.2(i). In the first step, we will show that the divergence of the vector field $\boldsymbol{V}$ on D_J integrates to zero in the sense that

$$\lim_{r\to+\infty}\left|\int_{\{\boldsymbol{y}\in\mathbb{R}^n\mid\|\boldsymbol{y}\|_2\leq r\}\cap D_J}\nabla_{\boldsymbol{y}}\cdot\boldsymbol{V}(\boldsymbol{y})d\boldsymbol{y}\right|=0.\tag{70}$$

In the second step, we will show that $\mathbb{E}_J\left[\left|\left\langle\varphi_J(\boldsymbol{y}|\boldsymbol{x},t),\boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right|\right]<+\infty$, with $\mathbb{E}_J\left[\left\langle\varphi_J(\boldsymbol{y}|\boldsymbol{x},t),\boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right]=n\epsilon$, hereby proving formula (39), using the convexity of the function $\boldsymbol{y}\mapsto\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})$, Fatou's lemma [30, Lemma 2.18], and Eq. (70) derived in the first step. In the third step, we will combine the results from the first and second steps to show that $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right]<+\infty$ and conclude that the representation formulas (40) and (41) hold. Finally, in the fourth step we will conclude that the representation formulas (42) and (43) hold using Eqs. (40) and (41) and Proposition (3.1)(iii).

Step 1. The proof of the limit result (70) that we present here is based on an application of Theorem 4.14 in [54] to the vector field $\boldsymbol{V}(\cdot)$. As this result is fairly technical, we first introduce some terminology and definitions that will be used exclusively in this part of the proof of (i).

Let C be a non-empty convex subset of $\mathbb{R}^n$. The *dimension* of the set C is defined as the smallest dimension of a non-empty affine set containing C, with the dimension of a non-empty affine set being the dimension of the subspace parallel to it [58, pages 4 and 12]. If C consists of a single point, then its dimension is taken to be zero.

Let $k\in\{0,\ldots,n\}$. Denote by $\mathcal{H}^{n-k}$ the $(n-k)$-dimensional outer Hausdorff measure in $\mathbb{R}^n$ as defined in [28, Sect. 2.10.2, p.171]. The measure $\mathcal{H}^{n-k}$, in particular, is a constant multiple of the $(n-k)$-dimensional Lebesgue measure for every measurable subset $B\subset\mathbb{R}^n$ (see [27], Section 1.2, p.7, and Theorem 1.12, p.13).

A subset $S\subset\mathbb{R}^n$ is called *slight* if $\mathcal{H}^{n-1}(S)=0$, and a subset $T\subset\mathbb{R}^n$ is called *thin* if T is σ-finite for $\mathcal{H}^{n-1}$, i.e., T can be expressed as a countable union of sets $T=\cup_{k=1}^{+\infty}T_k$ with $\mathcal{H}^{n-1}(T_k)<+\infty$ for each $k\in\mathbb{N}^+$ (see, e.g., [54]).

Let $k\in\{0,\ldots,n\}$. A non-empty, measurable subset $\Omega\subset\mathbb{R}^n$ is said to be *countably* $\mathcal{H}^{n-k}$*-rectifiable* if it is contained, up to a null set of $(n-k)$-dimensional outer Hausdorff measure $\mathcal{H}^{n-k}$ zero, in a countable union of continuously differentiable hypersurfaces of dimension $(n-k)$ (see, e.g., [1] and references therein). A non-empty, measurable and countably $\mathcal{H}^{n-k}$-rectifiable subset of $\mathbb{R}^n$, in particular, is σ-finite for $\mathcal{H}^{n-k}$.

A subset $A\subset\mathbb{R}^n$ is called *admissible* if its boundary bd A is thin and if the distributional gradient of the characteristic function of A is a vector measure on Borel subsets of $\mathbb{R}^n$ whose variation is finite (see [54] pp.151 and the reference therein). For the purpose of our proof, we will use the fact that the family of closed balls of radius $r>0$, namely $\{\boldsymbol{y}\in\mathbb{R}^n\mid\|\boldsymbol{y}\|_2\leq r\}$, are admissible sets (see [32], Example 1.10, and note that admissible sets are also called *Caccioppoli* sets [32], pages 5-6).

Let A be an admissible set and let $\boldsymbol{v}\colon A \to \mathbb{R}^n$ be a vector field. In the terminology of [54], we say that $\boldsymbol{v}$ is *integrable* over the admissible set A if $\boldsymbol{v}$ satisfies definition 4.1 of [54], and in that case, the number $I(\boldsymbol{v}, A)$ is called the integral of $\boldsymbol{v}$ over A. Note, here, that the notion of integrability considered in [54] is different from that of the usual Lebesgue integrability. Nevertheless, if $\boldsymbol{v}$ is integrable in the sense of [54], $\boldsymbol{v}$ is also Lebesgue measurable (Corollary 4.9, [54]), and if the Lebesgue integral $\int_A |\boldsymbol{v}(\boldsymbol{y})| d\boldsymbol{y}$ is finite, then $I(\boldsymbol{v}, A) = \int_A |\boldsymbol{v}(\boldsymbol{y})| d\boldsymbol{y}$ ([54], Proposition 4.7).

Let E be an non-empty subset of $\mathbb{R}^n$, let $\boldsymbol{v}\colon E \to \mathbb{R}^n$ be a vector field, and let $D_{\boldsymbol{v}}$ denote the set of points at which $\boldsymbol{v}$ is differentiable in int E (for the definition of differentiability of vector fields, see [61], page 150, Definition 7.22). In the terminology of [54], we call a *divergence of* $\boldsymbol{v}$ any function $g\colon E \mapsto \mathbb{R}$ such that $g(\boldsymbol{y}) = \nabla \cdot \boldsymbol{v}(\boldsymbol{y})$ for each $\boldsymbol{y} \in (\text{int } E) \cap D_{\boldsymbol{v}}$.

In addition to these definitions, we will need the following two results due to, respectively, [1] and [54].

Theorem C.1 [1, Theorem 4.1] (for convex functions) *Let Ω be a bounded, open, convex subset of $\mathbb{R}^n$, and let $f:\Omega \to \mathbb{R}$ be a convex and Lipschitz continuous function. Denote the subdifferential of f at $\boldsymbol{y} \in \Omega$ by $\partial f(\boldsymbol{y})$. Then, for each $k \in \{0, \dots, n\}$, the set*

$$\{\boldsymbol{y} \in \Omega \mid dim\,(\partial f(\boldsymbol{y})) \geq k\}$$

is countably $\mathcal{H}^{n-k}$-rectifiable.

Theorem C.2 [54, Theorem 4.14] *Let A be an admissible set, and let S and T be, respectively, a slight and thin subset of* cl A. *Let $\boldsymbol{v}$ be a bounded vector field in* cl A *that is continuous in* (cl A) $\setminus$ S *and differentiable in* (int A) $\setminus$ T. *Then, every divergence of $\boldsymbol{v}$ is integrable in A. Moreover, there exists a vector field* bd $A \ni \boldsymbol{y} \to n_A(\boldsymbol{y})$ *with* $\|n_A(\boldsymbol{y})\|_2 = 1$ *for every* $\boldsymbol{y} \in$ bd A *such that if div $\boldsymbol{v}$ denotes any divergence $\boldsymbol{v}$, then*

$$I(div\ \boldsymbol{v}, A) = \int_{\text{bd } A} \langle \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle \, d\mathcal{H}^{n-1} d\boldsymbol{y}. \tag{71}$$

Step 1 (Continued). Fix $r > 0$ and let $A = \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}\|_2 \leq r\}$ denote the closed ball of radius r centered at the origin in $\mathbb{R}^n$. Note that A is bounded, convex, closed, and admissible. Consider now the restriction of the convex function J to int A. As int A is bounded, open and convex, the function J is Lipschitz continuous on int A [58, Theorem 10.4]. All conditions in Theorem (C.1) are satisfied (with Ω = int A and $f = J$), and we can invoke the theorem to conclude that the set

$$T = \{\boldsymbol{y} \in \text{int } A \mid \dim\,(\partial J(\boldsymbol{y})) \geq 1\}$$

is countably $\mathcal{H}^{n-1}$-rectifiable, and therefore σ-finite for $\mathcal{H}^{n-1}$. In particular, the set T is thin. Moreover, recalling the definition of the set $D_J := \{\boldsymbol{y} \in \mathbb{R}^n \mid \partial J(\boldsymbol{y}) = \{\nabla J(\boldsymbol{y})\}\}$, we find that the set (int A) $\setminus$ T comprises the points $\boldsymbol{y} \in$ int A at which the subdifferential $\partial J(\boldsymbol{y})$ is a singleton, i.e., $T = (\text{int } A) \cap (\mathbb{R}^n \setminus D_J)$.

Now, consider the vector field $\boldsymbol{V}$ defined by (67). This vector field is continuous in $\mathbb{R}^n$ by convexity of J and dom $J = \mathbb{R}^n$. It is also bounded by (68). Now, define the function $g\colon A \to \mathbb{R}$ via

$$g(\boldsymbol{y}) = \begin{cases} \nabla \cdot \boldsymbol{V}(\boldsymbol{y}) \text{ if } \boldsymbol{y} \in A \cap D_J, \\ 0, \quad \text{if } \boldsymbol{y} \in A \cap (\mathbb{R}^n \setminus D_J). \end{cases} \tag{72}$$

The function g constitutes a divergence of the vector field $\boldsymbol{V}$ because it coincides with the divergence $\nabla \cdot \boldsymbol{V}(\boldsymbol{y})$ at every $\boldsymbol{y} \in (\text{int } A) \cap D_J$. Moreover, its Lebesgue integral over A is finite; to see this, first note that for every $\boldsymbol{y} \in A \cap D_J$ the absolute value of $g(\boldsymbol{y})$ can be bounded using (69), the triangle inequality, the Cauchy–Schwarz inequality, and assumption (A3) as follows

$$\begin{aligned} |g(\boldsymbol{y})| &= |\nabla \cdot \boldsymbol{V}(\boldsymbol{y})| \\ &= \left| \left(n - \left\langle \frac{1}{\epsilon}\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t} + \nabla J(\boldsymbol{y})\right), \boldsymbol{y} - \boldsymbol{y}_0 \right\rangle \right) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} \right| \\ &\leq \left(n + \left| \left\langle \frac{1}{\epsilon}\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t} + \nabla J(\boldsymbol{y})\right), \boldsymbol{y} - \boldsymbol{y}_0 \right\rangle \right| \right) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} \\ &\leq \left(n + \frac{1}{\epsilon}\left\| \frac{\boldsymbol{y}-\boldsymbol{x}}{t} + \nabla J(\boldsymbol{y}) \right\|_2 \|\boldsymbol{y} - \boldsymbol{y}_0\|_2 \right) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} \\ &\leq \left(n + \frac{1}{\epsilon}\left(\left\| \frac{\boldsymbol{y}-\boldsymbol{x}}{t} \right\|_2 + \|\nabla J(\boldsymbol{y})\|_2 \right) \|\boldsymbol{y} - \boldsymbol{y}_0\|_2 \right) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} \\ &\leq \left(n + \frac{1}{\epsilon}\left(\left\| \frac{\boldsymbol{y}-\boldsymbol{x}}{t} \right\|_2 + \|\nabla J(\boldsymbol{y})\|_2 \right) \|\boldsymbol{y} - \boldsymbol{y}_0\|_2 \right) e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} \end{aligned} \tag{73}$$

Second, as the set A is a closed bounded subset of dom $J = \mathbb{R}^n$ the function J is Lipschitz continuous relative to A, and therefore there exists a number $L_A > 0$ such that $\|\nabla J(\boldsymbol{y})\|_2 \leq L_A$ for every $\boldsymbol{y} \in A \cap D_J$. As a consequence, we can further bound $g(\boldsymbol{y})$ for every $\boldsymbol{y} \in A \cap D_J$ in (73) as

$$|g(\boldsymbol{y})| \leq \left(n + \frac{1}{\epsilon}\left(\left\| \frac{\boldsymbol{y}-\boldsymbol{x}}{t} \right\|_2 + L_A \right) \|\boldsymbol{y} - \boldsymbol{y}_0\|_2 \right) e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}. \tag{74}$$

In particular, using the definition of g given by (72), we have that (74) holds for every $\boldsymbol{y} \in A$. We can now use (72) and (74) to get

$$\begin{aligned}\int_A |g(\boldsymbol{y})| d\boldsymbol{y} &= \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}} |g(\boldsymbol{y})| d\boldsymbol{y} \\ &= \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}\cap D_J} |\nabla\cdot \boldsymbol{V}(\boldsymbol{y})|\, d\boldsymbol{y} \\ &\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}\cap D_J} \left(n + \frac{1}{\epsilon}\left(\left\|\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right\|_2 + L_A\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right) e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y} \\ &\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}} \left(n + \frac{1}{\epsilon}\left(\left\|\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right\|_2 + L_A\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right) e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y}.\end{aligned} \tag{75}$$

Since the function

$$\boldsymbol{y} \mapsto \left(n + \frac{1}{\epsilon}\left(\left\|\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right\|_2 + L_A\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right) e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}$$

is continuous, it is bounded on the compact set $A = \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}\|_2 \le r\}$. Its integral over A is therefore finite, and using (75) we find that $\int_A |g(\boldsymbol{y})| d\boldsymbol{y}$ is finite as well.

The previous considerations show that all conditions in Theorem (C.2) are satisfied (with $A = \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}\|_2 \le r\}$, $\boldsymbol{v} = \boldsymbol{V}$, $S = \varnothing$, $T = \{\boldsymbol{y} \in \text{int } A \mid \dim(\partial J(\boldsymbol{y})) \ge 1\} = (\text{int } A) \cap (\mathbb{R}^n \setminus D_J)$). We can therefore invoke [54, Theorem 4.14] to conclude that the divergence of g is integrable (in the sense described by [54]), with integral $I(g, A)$, and that there exists a vector field bd $A \ni \boldsymbol{y} \to \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})$ with $\|\boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\|_2 = 1$ for every $\boldsymbol{y} \in$ bd A such that

$$I(g, A) = \int_{\text{bd } A} \langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle \, d\mathcal{H}^{n-1} d\boldsymbol{y}. \tag{76}$$

Since the Lebesgue integral of $|g|$ over A is finite, we also have [54, Prop 4.7]

$$I(g, A) = \int_A g(\boldsymbol{y}) d\boldsymbol{y}. \tag{77}$$

Using that $A = \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}\|_2 \le r\}$, Eqs. (72), (76), and (77), we obtain

$$\begin{aligned}\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}} g(\boldsymbol{y}) d\boldsymbol{y} &= \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}} \nabla\cdot \boldsymbol{V}(\boldsymbol{y}) d\boldsymbol{y} \\ &= \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} \langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle \, d\mathcal{H}^{n-1}.\end{aligned} \tag{78}$$

As r was an arbitrary positive number, we can take the absolute value and then the limit $r \to +\infty$ on both sides of (78) to find

$$\begin{aligned}\lim_{r\to+\infty}\left|\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2\le r\}\cap D_J} \nabla\cdot \boldsymbol{V}(\boldsymbol{y}) d\boldsymbol{y}\right| = \lim_{r\to+\infty}\left|\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} \langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle \, d\mathcal{H}^{n-1}\right|.\end{aligned} \tag{79}$$

We will now show that the limit on the right side of (79) is equal to zero. To show this, first take the absolute value inside the integral on the right side of (79) to find

$$\begin{aligned}&\left|\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} \langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle \, d\mathcal{H}^{n-1}\right| \\ &\quad\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} |\langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle| \, d\mathcal{H}^{n-1}.\end{aligned} \tag{80}$$

Use the Cauchy–Schwarz inequality, Eq. (67), assumption (A3) ($\inf_{\boldsymbol{y}\in\mathbb{R}^n} J(\boldsymbol{y}) = 0$) and $\|\boldsymbol{n}_{\boldsymbol{v}}\|_2 = 1$ to further bound the right side of (80) as follows

$$\begin{aligned}&\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} |\langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle| \, d\mathcal{H}^{n-1} \\ &\quad\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} \|\boldsymbol{V}(\boldsymbol{y})\|_2 \, \|\boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\|_2 \, d\mathcal{H}^{n-1} \\ &\quad\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} \left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2 e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} \, d\mathcal{H}^{n-1} \\ &\quad\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} (\|\boldsymbol{y}\|_2 + \left\|\boldsymbol{y}_0\right\|_2) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right)/\epsilon} \, d\mathcal{H}^{n-1}.\end{aligned} \tag{81}$$

Use the parallelogram law $2(\|\boldsymbol{x}\|_2^2 + \|\boldsymbol{u}\|_2^2) = \|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + \|\boldsymbol{x}+\boldsymbol{u}\|_2^2$ with $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y}$ to bound the exponential $e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon}$ by

$$\begin{aligned}e^{-\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/\epsilon} &= e^{-\frac{1}{2t}\left(\frac{1}{2}(\|\boldsymbol{y}\|_2^2 + \|2\boldsymbol{x}-\boldsymbol{y}\|_2^2) - \|\boldsymbol{x}\|_2^2\right)/\epsilon} \\ &\le e^{-\frac{1}{2t}\left(\frac{1}{2}\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}\|_2^2\right)/\epsilon}\end{aligned} \tag{82}$$

and use it in (81) to get

$$\begin{aligned}&\int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} |\langle \boldsymbol{V}(\boldsymbol{y}), \boldsymbol{n}_{\boldsymbol{v}}(\boldsymbol{y})\rangle| \, d\mathcal{H}^{n-1} \\ &\quad\le \int_{\{\boldsymbol{y}\in\mathbb{R}^n \mid \|\boldsymbol{y}\|_2 = r\}} (\|\boldsymbol{y}\|_2 + \left\|\boldsymbol{y}_0\right\|_2) e^{-\frac{1}{2t}\left(\frac{1}{2}\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}\|_2^2\right)/\epsilon} \, d\mathcal{H}^{n-1}.\end{aligned} \tag{83}$$

Since the domain of integration in (83) is over the surface of an n-dimensional sphere of radius $\|\boldsymbol{y}\|_2 = r$, the integral on the right side of (83) is given by

$$\begin{aligned}
&\int_{\{y\in\mathbb{R}^n \mid \|y\|_2=r\}} (\|y\|_2 + \|y_0\|_2) e^{-\frac{1}{2t}(\frac{1}{2}\|y\|_2^2-\|x\|_2^2)/\epsilon}\, d\mathcal{H}^{n-1} \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2=r\}} (r + \|y_0\|_2) e^{-\frac{1}{2t}(\frac{1}{2}r^2-\|x\|_2^2)/\epsilon}\, d\mathcal{H}^{n-1} \\
&= (r + \|y_0\|_2) e^{-\frac{1}{2t}(\frac{1}{2}r^2-\|x\|_2^2)/\epsilon} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2=r\}} d\mathcal{H}^{n-1} \\
&= (r + \|y_0\|_2) e^{-\frac{1}{2t}\left(\frac{1}{2}r^2-\|x\|_2^2\right)/\epsilon} \frac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)},
\end{aligned} \tag{84}$$

where $n\pi^{n/2}/\Gamma\left(\frac{n}{2}+1\right)$ is the area of an n-dimensional sphere of radius one, with $\Gamma\left(\frac{n}{2}+1\right)$ denoting the Gamma function evaluated at $\frac{n}{2}+1$. Since

$$\lim_{r\to+\infty} (r + \|y_0\|_2) e^{-\frac{1}{2t}\left(\frac{1}{2}r^2-\|x\|_2^2\right)/\epsilon} = 0,$$

the limit $r \to +\infty$ in (84) is equal to zero, i.e.,

$$\lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2=r\}} (\|y\|_2 + \|y_0\|_2) e^{-\frac{1}{2t}(\frac{1}{2}\|y\|_2^2-\|x\|_2^2)/\epsilon}\, d\mathcal{H}^{n-1} = 0. \tag{85}$$

Combining (79), (80), (83) and (85) yield

$$\lim_{r\to+\infty} \left| \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}\cap D_J} \nabla_y \cdot V(y) dy \right| = 0.$$

which proves the limit result (70).

Step 2. Recall that the divergence of the vector field $y \mapsto V(y)$ on D_J is given by (69). Combine (70) and (69) to conclude that

$$\begin{aligned}
\lim_{r\to+\infty} \Bigg| \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}\cap D_J} &\left(n\epsilon - \left\langle \left(\frac{y-x}{t} + \nabla J(y) \right), y - y_0 \right\rangle \right) \\
&e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy \Bigg| = 0.
\end{aligned} \tag{86}$$

Note that the minimal subgradient $\pi_{\partial J(y)}(\mathbf{0}) = \nabla J(y)$ for every $y \in D_J$. We can therefore substitute the minimal subgradient $\pi_{\partial J(y)}(\mathbf{0})$ for the gradient $\nabla J(y)$ inside the integral in the limit (86) without changing its value. Moreover, since the set D_J is dense in $\mathbb{R}^n$ and the n-dimensional Lebesgue measure of $(\mathbb{R}^n \setminus D_J)$ is zero, we can further substitute the domain of integration $\left\{y \in \mathbb{R}^n \mid \|y\|_2 \le r\right\} \cap D_J$ of the integral in the limit (86) with $\left\{y \in \mathbb{R}^n \mid \|y\|_2 \le r\right\}$ without changing its value. With these two changes, the limit (86) can be written as

$$\begin{aligned}
\lim_{r\to+\infty} \Bigg| \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} &\left(n\epsilon - \left\langle \left(\frac{y-x}{t} + \pi_{\partial J(y)}(\mathbf{0}) \right), y - y_0 \right\rangle \right) \\
&e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy \Bigg| = 0.
\end{aligned}$$

Using the notation $\varphi_J(y|x,t) = \left(\frac{y-x}{t}\right) + \pi_{\partial J(y)}(\mathbf{0})$, we can write this limit more succinctly as

$$\begin{aligned}
\lim_{r\to+\infty} \Bigg| \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} &\left(n\epsilon - \langle \varphi_J(y|x,t), y - y_0 \rangle \right) \\
&e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy \Bigg| = 0.
\end{aligned} \tag{87}$$

Now, consider the function $\mathbb{R}^n \ni y \mapsto \langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y - y_0 \rangle$. Note here that as J is convex with $\mathrm{dom}\ J = \mathbb{R}^n$, both $\varphi_J(y|x,t)$ and $\varphi_J(y_0|x,t)$ are subgradients of the convex function $u \mapsto \frac{1}{2t}\|x-u\|_2^2 + J(u)$ at $u = y$ and $u = y_0$, respectively [58, Theorem 23.4]. We can therefore apply inequality (13) (with $p = \varphi_J(y|x,t)$, $p_0 = \varphi_J(y_0|x,t)$, and $m = 0$) to find $\langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y - y_0 \rangle \ge 0$. Define $F\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathbb{R}$ as follows:

$$F(y) = \langle \varphi_J(y|x,t), y - y_0 \rangle e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon}$$

and

$$G(y) = \langle \varphi_J(y_0|x,t), y - y_0 \rangle e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon}.$$

Note that $F(y) - G(y) = \langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y - y_0 \rangle e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} \ge 0$ for every $y \in \mathbb{R}^n$. Integrate $y \mapsto F(y) - G(y)$ over $\mathbb{R}^n$ and use Fatou's lemma to find

$$\begin{aligned}
0 &\le \int_{\mathbb{R}^n} F(y) - G(y) dy \\
&\le \lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} F(y) - G(y) dy \\
&= \lim_{r\to+\infty} \Bigg(\int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} F(y) dy \\
&\quad + \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (-G(y)) dy \Bigg)
\end{aligned} \tag{88}$$

Use the Cauchy–Schwarz inequality assumption (A3) ($\inf_{y\in\mathbb{R}^n} J(y) = 0$) to bound the second integral on the right hand side of (88) as follows

$$\begin{aligned}
&\int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (-G(y))dy \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} -\langle \varphi_J(y_0|x,t), y-y_0\rangle \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\le \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} \left\|\varphi_J(y_0|x,t)\right\|_2 \left\|y-y_0\right\|_2 \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\le \left\|\varphi_J(y_0|x,t)\right\|_2 \int_{\mathbb{R}^n} \left\|y-y_0\right\|_2 \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2\right)/\epsilon} dy \\
&= \left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x,y_0,t,\epsilon),
\end{aligned} \tag{89}$$

where $C_1(x, y_0, t, \epsilon)$ was defined in (65). Combine (88) and (89) to find

$$\begin{aligned}
0 &\le \int_{\mathbb{R}^n} F(y) - G(y)dy \\
&\le \lim_{r\to+\infty} \left(\int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} F(y)dy + \left\|\varphi_J(y_0|x,t)\right\|_2 \right. \\
&\qquad \left. C_1(x,y_0,t,\epsilon) \right) \\
&= \left(\lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} F(y)dy \right) + \left\|\varphi_J(y_0|x,t)\right\|_2 \\
&\qquad C_1(x,y_0,t,\epsilon).
\end{aligned} \tag{90}$$

The integral on the right hand side of (90) can be bounded using assumption (A3) as follows

$$\begin{aligned}
&\int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} F(y)dy \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} \langle \varphi_J(y|x,t), y-y_0\rangle \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (\langle \varphi_J(y|x,t), y-y_0\rangle + (n\epsilon - n\epsilon)) \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (\langle \varphi_J(y|x,t), y-y_0\rangle - n\epsilon) \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\quad + n\epsilon \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\le \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (\langle \varphi_J(y|x,t), y-y_0\rangle - n\epsilon) \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\quad + n\epsilon \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|x-y\|_2^2\right)/\epsilon} dy \\
&= \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (\langle \varphi_J(y|x,t), y-y_0\rangle - n\epsilon) \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy + n\epsilon(2\pi t\epsilon)^{n/2}.
\end{aligned} \tag{91}$$

Combine (90) and (91) to get

$$\begin{aligned}
0 &\le \int_{\mathbb{R}^n} F(y) - G(y)dy \\
&\le \left(\lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2\le r\}} (\langle \varphi_J(y|x,t), y-y_0\rangle - n\epsilon) \right. \\
&\quad \left. e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \right) \\
&\quad + \left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x,y_0,t,\epsilon) + n\epsilon(2\pi t\epsilon)^{n/2}.
\end{aligned} \tag{92}$$

Combine (87) and (92) to get

$$\begin{aligned}
0 &\le \int_{\mathbb{R}^n} F(y) - G(y)dy \\
&= \int_{\mathbb{R}^n} \langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y-y_0\rangle \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\le \left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x,y_0,t,\epsilon) + n\epsilon(2\pi t\epsilon)^{n/2}.
\end{aligned} \tag{93}$$

Divide (93) by the partition function $Z_J(x, t, \epsilon)$ (see Eq. (26)) to get

$$\begin{aligned}
0 \le{}& \mathbb{E}_J\left[\langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y-y_0\rangle\right] \\
\le{}& \frac{\left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x,y_0,t,\epsilon) + n\epsilon(2\pi t\epsilon)^{n/2}}{Z_J(x,t,\epsilon)} < +\infty.
\end{aligned} \tag{94}$$

Now, using the Cauchy–Schwarz inequality and (66), we can bound $\mathbb{E}_J\left[|\langle \varphi_J(y_0|x,t), y-y_0\rangle|\right]$ as follows

$$\begin{aligned}
&\mathbb{E}_J\left[|\langle \varphi_J(y_0|x,t), y-y_0\rangle|\right] \\
&= \frac{1}{Z_J(x,t,\epsilon)} \int_{\mathbb{R}^n} |\langle \varphi_J(y_0|x,t), y-y_0\rangle| \\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2+J(y)\right)/\epsilon} dy \\
&\le \left\|\varphi_J(y_0|x,t)\right\|_2 \frac{1}{Z_J(x,t,\epsilon)} \int_{\mathbb{R}^n} \left\|y-y_0\right\|_2
\end{aligned}$$

$$e^{-\left(\frac{1}{2t}\|x-y\|_2^2\right)/\epsilon} dy = \frac{\left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x, y_0, t, \epsilon)}{Z_J(x, t, \epsilon)}. \tag{95}$$

Use the triangle inequality and the upper bounds in (94) and (95) to obtain

$$\begin{aligned}
&\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t) - (\varphi_J(y_0|x,t) - \varphi_J(y_0|x,t)), y-y_0\rangle\right|\right]\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(y_0|x,t), y-y_0\rangle\right| \right.\\
&\quad \left. + \left|\langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y-y_0\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(y_0|x,t), y-y_0\rangle\right|\right]\\
&\quad + \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t) - \varphi_J(y_0|x,t), y-y_0\rangle\right|\right]\\
&\leq \frac{\left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x, y_0, t, \epsilon)}{Z_J(x, t, \epsilon)}\\
&\quad + \frac{\left\|\varphi_J(y_0|x,t)\right\|_2 C_1(x, y_0, t, \epsilon) + n\epsilon(2\pi t\epsilon)^{n/2}}{Z_J(x, t, \epsilon)}\\
&< +\infty.
\end{aligned} \tag{96}$$

Since $\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right|\right] < +\infty$, we can use (87) to conclude that

$$\begin{aligned}
&\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y-y_0\rangle\right]\\
&= \frac{1}{Z_J(x,t,\epsilon)} \int_{\mathbb{R}^n} \langle \varphi_J(y|x,t), y-y_0\rangle\\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy\\
&= \frac{1}{Z_J(x,t,\epsilon)}\\
&\quad \lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2 \leq r\}} \langle \varphi_J(y|x,t), y-y_0\rangle\\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy\\
&= \frac{1}{Z_J(x,t,\epsilon)}\\
&\quad \lim_{r\to+\infty} \int_{\{y\in\mathbb{R}^n \mid \|y\|_2 \leq r\}} (n\epsilon - n\epsilon + \langle \varphi_J(y|x,t), y-y_0\rangle)\\
&\quad e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy\\
&= n\epsilon\\
&\quad - \lim_{r\to+\infty} \left(\int_{\{y\in\mathbb{R}^n \mid \|y\|_2 \leq r\}} (n\epsilon - \langle \varphi_J(y|x,t), y-y_0\rangle) \right.\\
&\quad \left. e^{-\left(\frac{1}{2t}\|x-y\|_2^2 + J(y)\right)/\epsilon} dy \right)\\
&= n\epsilon.
\end{aligned} \tag{97}$$

Inequality (96) and equality (97) show the desired results $\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right|\right] < +\infty$ and $\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y-y_0\rangle\right] = n\epsilon$, which, after recalling the definition $\varphi_J(y|x,t) = \left(\frac{y-x}{t}\right) + \pi_{\partial J(y)}(\mathbf{0})$, also proves formula (39).

Step 3. Thanks to Step 2, we have $\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right|\right] < +\infty$ and $\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y-y_0\rangle\right] = n\epsilon$ for every $y_0 \in \mathbb{R}^n$. In particular, the choice of $y_0 = \mathbf{0}$ yields $\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y\rangle\right|\right] < +\infty$ and $\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y\rangle\right] = n\epsilon$. As a consequence, we have that

$$\begin{aligned}
&\mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y_0\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y_0 + (y-y)\rangle\right|\right]\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right| + \left|\langle \varphi_J(y|x,t), y\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y-y_0\rangle\right|\right] + \mathbb{E}_J\left[\left|\langle \varphi_J(y|x,t), y\rangle\right|\right]\\
&< \infty,
\end{aligned} \tag{98}$$

and

$$\begin{aligned}
&\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y_0\rangle\right]\\
&= \mathbb{E}_J\left[\langle \varphi_J(y|x,t), (y-y) + y_0\rangle\right]\\
&= \mathbb{E}_J\left[\langle \varphi_J(y|x,t), y\rangle\right] - \mathbb{E}_J\left[\langle \varphi_J(y|x,t), y-y_0\rangle\right]\\
&= n\epsilon - n\epsilon\\
&= 0,
\end{aligned} \tag{99}$$

for every $y_0 \in \mathbb{R}^n$. Now, let $\{e_i\}_{i=1}^n$ denote the standard basis in $\mathbb{R}^n$ and let $\{\varphi_J(y|x,t)_i\}_{i=1}^n$ denote the components of the vector $\varphi_J(y|x,t)$, i.e., $\varphi_J(y|x,t) = (\varphi_J(y|x,t)_1, \ldots, \varphi_J(y|x,t)_n)$. Using (98) with the choice of $y_0 = e_i$ for $i \in \{1,\ldots,n\}$, we get $\mathbb{E}_J\left[|\varphi_J(y|x,t)_i|\right] < +\infty$ for every $i \in \{1,\ldots,n\}$. Using the norm inequality $\|\varphi_J(y|x,t)\|_2 \leq \sum_{i=1}^n |\varphi_J(y|x,t)_i|$, we can bound $\mathbb{E}_J\left[\|\varphi_J(y|x,t)\|_2\right]$ as follows

$$\begin{aligned}
\mathbb{E}_J\left[\|\varphi_J(y|x,t)\|_2\right] &\leq \mathbb{E}_J\left[\sum_{i=1}^n |\varphi_J(y|x,t)_i|\right]\\
&= \sum_{i=1}^n \mathbb{E}_J\left[|\varphi_J(y|x,t)_i|\right]\\
&< +\infty.
\end{aligned} \tag{100}$$

We can therefore combine (99) and (100) to get $\mathbb{E}_J\left[\langle \varphi_J(y|x,t), y_0\rangle\right] = \langle \mathbb{E}_J\left[\varphi_J(y|x,t)\right], y_0\rangle = 0$ for every $y_0 \in \mathbb{R}^n$, which yields the following equality:

$$\mathbb{E}_J\left[\varphi_J(y|x,t)\right] = 0. \tag{101}$$

Moreover, recalling the definition $\varphi_J(y|x,t) = \left(\frac{y-x}{t}\right) + \pi_{\partial J(y)}(\mathbf{0})$ and using (66) (with $y_0 = x$) and (100), we can bound $\mathbb{E}_J\left[\left\|\pi_{\partial J(y)}(\mathbf{0})\right\|_2\right]$ as follows

$$\begin{aligned}
&\mathbb{E}_J\left[\left\|\pi_{\partial J(y)}(\mathbf{0})\right\|_2\right]\\
&= \mathbb{E}_J\left[\left\|\pi_{\partial J(y)}(\mathbf{0}) + \left(\frac{y-x}{t}\right) - \left(\frac{y-x}{t}\right)\right\|_2\right]\\
&\leq \mathbb{E}_J\left[\left\|\pi_{\partial J(y)}(\mathbf{0}) + \left(\frac{y-x}{t}\right)\right\|_2 + \left\|\left(\frac{y-x}{t}\right)\right\|_2\right]\\
&= \mathbb{E}_J\left[\|\varphi_J(y|x,t)\|_2\right] + \frac{1}{t}\mathbb{E}_J\left[\|y-x\|_2\right]\\
&\leq \mathbb{E}_J\left[\|\varphi_J(y|x,t)\|_2\right] + \frac{C_1(x,x,t,\epsilon)}{tZ_J(x,t,\epsilon)}\\
&< +\infty.
\end{aligned} \tag{102}$$

We can now combine (66), (101) and (102) to expand the expected value of $\mathbb{E}_J[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)]$ as follows

$$\begin{aligned}\mathbb{E}_J&[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)]\\&=\mathbb{E}_J\left[\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)+\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]\\&=\mathbb{E}_J\left[\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)\right]+\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]\\&=\left(\frac{\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{x}}{t}\right)+\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]\\&=\boldsymbol{0}.\end{aligned}\tag{103}$$

Solving for $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in (103) yields $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)=\boldsymbol{x}-t\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]$, which gives the representation formula (40).

We now derive the second representation formula (41). Let $\boldsymbol{y}_0=\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ in Eq. (97) and use the representation formula (40) to find

$$\begin{aligned}\mathbb{E}_J&\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&=\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})+\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)-\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\right.\right.\\&\quad\left.\left.\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&=\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})+\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right.\\&\quad\left.-\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&=\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})+\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&\quad-\mathbb{E}_J\left[\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&=\mathbb{E}_J\left[\langle\varphi_J(\boldsymbol{y}|\boldsymbol{x},t),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&\quad-\mathbb{E}_J\left[\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]\\&=n\epsilon-\mathbb{E}_J\left[\left\langle\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right].\end{aligned}\tag{104}$$

We will use (104) to derive a representation formula for $\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]$. Multiply (104) by t and rearrange to get

$$\begin{aligned}\mathbb{E}_J&\left[\langle\boldsymbol{y}-\boldsymbol{x},\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&=nt\epsilon-t\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right].\end{aligned}\tag{105}$$

The left hand side of (105) can be expressed as

$$\begin{aligned}\mathbb{E}_J&\left[\langle\boldsymbol{y}-\boldsymbol{x},\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&=\mathbb{E}_J\left[\langle\boldsymbol{y}-\boldsymbol{x}+(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)),\right.\\&\quad\left.\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&=\mathbb{E}_J\left[\langle\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right.\\&\quad\left.+\langle\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&=\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]\\&\quad+\mathbb{E}_J\left[\langle\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right]\\&=\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]\\&\quad+\langle\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon),\mathbb{E}_J\left[\boldsymbol{y}\right]-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\\&=\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]\\&\quad+\langle\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon),\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\\&=\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right].\end{aligned}\tag{106}$$

Combine Eqs. (105) and (106) to get

$$\begin{aligned}\mathbb{E}_J&\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]\\&=nt\epsilon-t\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right],\end{aligned}$$

which gives the representation formula (41).

Step 4. Thanks to Step 3, the representation formulas (40) and (41) hold. Recall that by Proposition 3.1(iii), the gradient $\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)$ and Laplacian $\Delta_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)$ of the solution S_ϵ to the viscous HJ PDE (29) satisfy the representation formulas

$$\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)=\boldsymbol{x}-t\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)\tag{107}$$

and

$$\mathbb{E}_J\left[\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right]=nt\epsilon-t^2\epsilon\Delta_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t).\tag{108}$$

Use (40) and (107) to get

$$\nabla_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)=\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right],$$

which is the representation formula (42). Use (41) and (108) to get

$$t^2\epsilon\Delta_{\boldsymbol{x}}S_\epsilon(\boldsymbol{x},t)=t\mathbb{E}_J\left[\left\langle\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle\right]$$

which is, after dividing by $t\epsilon$ on both sides, the representation formulas (43). This concludes Step 4.

Proof of (ii): Here, we only assume that J satisfies assumptions (A1)–(A3); we do not assume that $\operatorname{dom} J=\mathbb{R}^n$. Let $\{\mu_k\}_{k=1}^{+\infty}$ be a sequence of positive real numbers converging to zero. Define $f_k\colon\mathbb{R}^n\times(0,+\infty)\times(0,+\infty)\to\mathbb{R}$ by

$$\begin{aligned}f_\epsilon&(\boldsymbol{x},t,k)\\&=-\epsilon\log\left(\frac{1}{(2\pi t\epsilon)^{n/2}}\int_{\mathbb{R}^n}e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}d\boldsymbol{y}\right)\end{aligned}\tag{109}$$

and let $S_0(\boldsymbol{x},\mu_k)$ denote the solution to the first-order HJ PDE (20) with initial data J evaluated at $(\boldsymbol{x},\mu_k)$, that is,

$$S_0(\boldsymbol{x},\mu_k)=\inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2\mu_k}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right\}.\tag{110}$$

By Proposition 2.2(i), the function $\mathbb{R}^n\ni\boldsymbol{x}\mapsto S_0(\boldsymbol{x},\mu_k)$ is continuously differentiable and convex for each $k\in\mathbb{N}$, and the sequence of real numbers $\{S_0(\boldsymbol{x},\mu_k)\}_{k=1}^{+\infty}$ converges to $J(\boldsymbol{x})$ for every $\boldsymbol{x}\in\operatorname{dom}J$. Moreover, by assumption (A3) ($\inf_{\boldsymbol{y}\in\mathbb{R}^n}J(\boldsymbol{y})=0$) the sequence $\{S_0(\boldsymbol{x},\mu_k)\}_{k=1}^{+\infty}$ is uniformly bounded from below by 0, that is,

$$\begin{aligned}S_0(\boldsymbol{x},\mu_k)&=\inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2\mu_k}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right\}\\&\geq\inf_{\boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{1}{2\mu_k}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right\}+\inf_{\boldsymbol{y}\in\mathbb{R}^n}J(\boldsymbol{y})\\&=0.\end{aligned}$$

As a consequence, we can invoke Proposition 3.1(i) to conclude that for each $k \in \mathbb{N}$, the function $(\boldsymbol{x}, t) \mapsto f_\epsilon(\boldsymbol{x}, t, k)$ corresponds to the solution to the viscous HJ PDE (29) with initial data $f_k(\boldsymbol{x}, 0, \epsilon) = S_0(\boldsymbol{x}, \mu_k)$. Moreover, $\mathbb{R}^n \ni \boldsymbol{x} \mapsto f_\epsilon(\boldsymbol{x}, t, k)$ is continuously differentiable and convex by Proposition 3.1(i) and (ii)(a). Finally, as the domain of the function $\boldsymbol{x} \mapsto S_0(\boldsymbol{x}, \mu_k)$ is $\mathbb{R}^n$, we can use the representation formula (42) in Proposition 4.2(i) (which was proven previously in this Appendix) to express the gradient $\nabla_{\boldsymbol{x}} f_k(\boldsymbol{x}, t, \epsilon)$ as follows

$$\nabla_{\boldsymbol{x}} f_\epsilon(\boldsymbol{x}, t, k) = \frac{\int_{\mathbb{R}^n} \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y}, \mu_k) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}. \tag{111}$$

Now, since $S_0(\boldsymbol{x}, \mu_k) \geq 0$ for every $k \in \mathbb{N}$, we can bound the integrand in (109) as follows

$$\frac{1}{(2\pi t\epsilon)^{n/2}} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} \leq \frac{1}{(2\pi t\epsilon)^{n/2}} e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}, \tag{112}$$

where $\int_{\mathbb{R}^n} \frac{1}{(2\pi t\epsilon)^{n/2}} e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y} = 1$. We can therefore invoke the Lebesgue dominated convergence theorem [30, Theorem 2.24] and use (109) and the limit $\lim_{k\to+\infty} e^{-S_0(\boldsymbol{x},\mu_k)/\epsilon} = e^{-J(\boldsymbol{x})/\epsilon}$ (with $\lim_{k\to+\infty} e^{-S_0(\boldsymbol{x},\mu_k)/\epsilon} = 0$ for every $\boldsymbol{x} \notin \operatorname{dom} J$) to find

$$\begin{aligned}
&\lim_{k\to+\infty} f_\epsilon(\boldsymbol{x}, t, k) \\
&= \lim_{k\to+\infty} -\epsilon \log\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}\right) \\
&= -\epsilon \log\left(\frac{1}{(2\pi t\epsilon)^{n/2}} \int_{\operatorname{dom} J} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}\right) \\
&= S_\epsilon(\boldsymbol{x}, t),
\end{aligned} \tag{113}$$

which gives the limit (44). By continuous differentiability and convexity of $\mathbb{R}^n \ni \boldsymbol{x} \mapsto f_\epsilon(\boldsymbol{x}, t, k)$ and $\mathbb{R}^n \ni \boldsymbol{x} \mapsto S_\epsilon(\boldsymbol{x}, t)$ and the limit (113), we can invoke [58, Theorem 25.7] to conclude that the gradient $\nabla_{\boldsymbol{x}} f_k(\boldsymbol{x}, t, \mu_k)$ converges to the gradient $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$ as $k \to +\infty$. Hence, we can take the limit $k \to +\infty$ in (111) to find

$$\begin{aligned}
&\lim_{k\to+\infty} \nabla_{\boldsymbol{x}} f_\epsilon(\boldsymbol{x}, t, k) \\
&= \lim_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y}, \mu_k) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}\right) \\
&= \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t),
\end{aligned} \tag{114}$$

which gives the limit (45). Finally, using the definition of the posterior mean estimate (3), (114), and the representation formula (30) derived in Proposition 3.1(iii), namely $\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) = \boldsymbol{x} - t\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x}, t)$, we find the two limits

$$\begin{aligned}
&\boldsymbol{u}_{PM}(\boldsymbol{x}, t, \epsilon) \\
&= \lim_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \boldsymbol{y} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}\right) \\
&= \boldsymbol{x} - t \lim_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y}, \mu_k) e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}\right),
\end{aligned}$$

which establishes (46). This concludes the proof of (ii).

D Proof of Proposition 4.3

Let us first introduce some notation. Let $\boldsymbol{x} \in \mathbb{R}^n$, $t > 0$, $\epsilon > 0$, and $\boldsymbol{y}_0 \in \operatorname{dom} \partial J$. Define the functions

$$\operatorname{dom} \partial J \ni \boldsymbol{y} \mapsto \varphi_J(\boldsymbol{y}|\boldsymbol{x}, t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),$$

$$\operatorname{dom} \partial J \ni \boldsymbol{y} \mapsto \Phi_J(\boldsymbol{y}|\boldsymbol{x}, t) = \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y}),$$

and

$$\varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x}, t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y}, \mu_k).$$

Note that for every $\boldsymbol{y} \in \mathbb{R}^n$, $\varphi_J(\boldsymbol{y}|\boldsymbol{x}, t)$ is a subgradient of the function $\boldsymbol{u} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u})$ evaluated at $\boldsymbol{u} = \boldsymbol{y}$ and $\varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x}, t)$ is a subgradient of the function $\boldsymbol{u} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + S(\boldsymbol{u}, \mu_k)$ evaluated at $\boldsymbol{u} = \boldsymbol{y}$. Let $\{\mu_k\}_{k=1}^{+\infty}$ be a sequence of positive real numbers converging to zero and let $S_0 \colon \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ denote the solution to the first-order HJ PDE (20) with initial data J (see Proposition 2.2). Note that the sequence $\{S_0(\boldsymbol{y}, \mu_k)\}_{k=1}^{+\infty}$ is uniformly bounded from below since

$$\begin{aligned}
S_0(\boldsymbol{y}, \mu_k) &= \inf_{\boldsymbol{u}\in\mathbb{R}^n} \left\{\frac{1}{2t}\|\boldsymbol{y}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u})\right\} \\
&\geq J(\boldsymbol{y}) \\
&\geq 0.
\end{aligned} \tag{115}$$

Now, define the function $F \colon \operatorname{dom} \partial J \times \operatorname{dom} \partial J \times \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ as

$$\begin{aligned}
&F(\boldsymbol{y}, \boldsymbol{y}_0, \boldsymbol{x}, t) \\
&= \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x}, t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x}, t), \boldsymbol{y} - \boldsymbol{y}_0\rangle \\
&\quad \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}}
\end{aligned} \tag{116}$$

and the sequence of functions $\{F_{\mu_k}\}_{k=1}^{+\infty}$ with $F_{\mu_k} \colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ as

$$\begin{aligned}
&F_{\mu_k}(\boldsymbol{y}, \boldsymbol{y}_0, \boldsymbol{x}, t) \\
&= \langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x}, t) - \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x}, t), \boldsymbol{y} - \boldsymbol{y}_0\rangle \\
&\quad \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y}}.
\end{aligned} \tag{117}$$

Since $\lim_{k\to+\infty} S_0(\boldsymbol{y}, \mu_k) = J(\boldsymbol{y})$ and $\lim_{k\to+\infty} \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y}, \mu_k) = \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})$ for every $\boldsymbol{y} \in \operatorname{dom} \partial J$ by Proposition 2.2(i) and (iv), and

$$\begin{aligned}
&\lim_{k\to+\infty} \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + S_0(\boldsymbol{y}, \mu_k)\right)/\epsilon} d\boldsymbol{y} \\
&= \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}
\end{aligned} \tag{118}$$

by (44) in Proposition 4.2(ii) and continuity of the logarithm, the limit $\lim_{k\to+\infty} F_\mu(\boldsymbol{y}, \boldsymbol{y}_0, \boldsymbol{x}, t) = F(\boldsymbol{y}, \boldsymbol{y}_0, \boldsymbol{x}, t)$ holds for every $\boldsymbol{y} \in \operatorname{dom} \partial J$, $\boldsymbol{y}_0 \in \operatorname{dom} \partial J$, $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$. Note that as J is strongly convex with parameter $m \geq 0$, the functions $\boldsymbol{y} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})$

and $\boldsymbol{y} \mapsto \frac{1}{2t}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|_2^2 + S_0(\boldsymbol{y}, \mu_k)$ are strongly convex with parameter $\left(\frac{1+mt}{t}\right) > 0$. As a consequence, for every pair $(\boldsymbol{y}, \boldsymbol{y}_0) \in \operatorname{dom} \partial J \times \operatorname{dom} \partial J$, the following monotonicity inequalities hold (see Definition 5, Eq. (13)):

$$0 \leq \left(\frac{1+mt}{t}\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2 \leq \left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle \tag{119}$$

and

$$0 \leq \left(\frac{1+mt}{t}\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2 \leq \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle. \tag{120}$$

Multiply the first set of inequalities by $e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} / \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}$ and the second set of inequalities by $e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} / \int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}$ and use the definition of F and F_{μ_k} to get the inequalities

$$\begin{aligned} 0 &\leq \left(\frac{1+mt}{t}\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2 \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+J(\boldsymbol{y})\right)/\epsilon} d\boldsymbol{y}} \\ &\leq F(\boldsymbol{y},\boldsymbol{y}_0,\boldsymbol{x},t) \\ 0 &\leq \left(\frac{1+mt}{t}\right)\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2 \frac{e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}} \\ &\leq F_{\mu_k}(\boldsymbol{y},\boldsymbol{y}_0,\boldsymbol{x},t). \end{aligned} \tag{121}$$

These inequalities show, in particular, that F and F_{μ_k} are both non-negative functions for every $(\boldsymbol{y}, \boldsymbol{y}_0) \in \operatorname{dom} \partial J \times \operatorname{dom} \partial J$, $\boldsymbol{x} \in \mathbb{R}^n$, and $t > 0$. As a consequence, Fatou's lemma [30, Lemma 2.18] applies to the sequence of functions $\{F_{\mu_k}\}_{k=1}^{+\infty}$, and hence

$$\begin{aligned} &\int_{\operatorname{dom} \partial J} F(\boldsymbol{y},\boldsymbol{y}_0,\boldsymbol{x},t)d\boldsymbol{y} \\ &\leq \liminf_{k\to+\infty} \int_{\operatorname{dom} \partial J} F_{\mu_k}(\boldsymbol{y},\boldsymbol{y}_0,\boldsymbol{x},t)d\boldsymbol{y} \\ &\leq \liminf_{k\to+\infty} \int_{\mathbb{R}^n} F_{\mu_k}(\boldsymbol{y},\boldsymbol{y}_0,\boldsymbol{x},t)d\boldsymbol{y} \\ &= \liminf_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}\right) \\ &= \liminf_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}} \right. \\ &\quad \left. - \frac{\int_{\mathbb{R}^n} \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}\right). \end{aligned} \tag{122}$$

We now wish to compute the limit in (122). On the one hand, we can apply formula (39) in Proposition 4.2(i) (with initial data $S_0(\cdot, \mu_k)$ and using $\varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x},t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \nabla_{\boldsymbol{y}} S_0(\boldsymbol{y},\mu_k)$) to the first integral on the right side on the last line of (122) to get

$$\frac{\int_{\mathbb{R}^n} \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}} = n\epsilon. \tag{123}$$

On the other hand, applying the limit result (46) in Proposition 4.2(ii) for the posterior mean estimate $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ and the limit $\lim_{k\to+\infty} \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x},t) = \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t) = \left(\frac{\boldsymbol{y}_0-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y}_0)}(\boldsymbol{0})$ to the second integral on the right side on the last line of (122), we get

$$\liminf_{k\to+\infty} \left(\frac{\int_{\mathbb{R}^n} \left\langle \varphi_{S_0(\cdot,\mu_k)}(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}{\int_{\mathbb{R}^n} e^{-\left(\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+S_0(\boldsymbol{y},\mu_k)\right)/\epsilon} d\boldsymbol{y}}\right) = \left\langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{y}_0\right\rangle. \tag{124}$$

Combine (27), (116), (121), (122), (123), and (124) to get

$$\begin{aligned} &\left(\frac{1+mt}{t}\right)\mathbb{E}_J\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2^2\right] \\ &\leq \mathbb{E}_J\left[\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\right\rangle\right] \\ &\leq n\epsilon - \left\langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{y}_0\right\rangle. \end{aligned}$$

This establishes the set of inequalities (47).

Next, we show that $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right] < +\infty$ indirectly using the set of inequalities (47). By Proposition 4.1, $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) \in \operatorname{int}(\operatorname{dom} J)$. Hence, there exists a number $\delta > 0$ such that the open ball $\{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 < \delta\}$ is contained in int (dom J). Let $\boldsymbol{y}_0 \in \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 < \delta\}$ with $\boldsymbol{y}_0 \neq \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$. Recall int (dom J) $\subset \operatorname{dom} \partial J$, so that both $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$ and $\boldsymbol{y}_0$ are in the set $\operatorname{dom} \partial J$. We claim that $\mathbb{E}_J\left[\left|\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{y}_0\right\rangle\right|\right] < +\infty$. Indeed, using the triangle inequality, the set of inequalities (47) proven previously, the Cauchy-Schwarz inequality, and that $\mathbb{E}_J\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right] \leq \left(\int_{\mathbb{R}^n} \left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2 e^{-\frac{1}{2t\epsilon}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} d\boldsymbol{y}\right) / Z_J(\boldsymbol{x},t,\epsilon) < +\infty$ by assumption (A3),

$$\begin{aligned} 0 &\leq \mathbb{E}_J\left[\left|\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{y}_0\right\rangle\right|\right] \\ &= \mathbb{E}_J\left[\left|\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \boldsymbol{y}_0 + (\boldsymbol{y}-\boldsymbol{y})\right\rangle\right|\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle \right.\right.\\
&\quad \left.\left. - \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right|\right]\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right.\\
&\quad \left. + \left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right]\\
&\quad + \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right|\right]\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right] + n\epsilon\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\right.\right.\\
&\quad \left.\left. + (\varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t)), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right] + n\epsilon\\
&= \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right.\right.\\
&\quad \left.\left. + \langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right] + n\epsilon\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right.\\
&\quad \left. + \left|\langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right] + n\epsilon\\
&\leq \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right]\\
&\quad + \mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right|\right] + n\epsilon\\
&\leq \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{y}_0\rangle\right]\\
&\quad + \mathbb{E}_J\left[\left\|\varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t)\right\|_2 \left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right] + n\epsilon\\
&\leq n\epsilon - \langle \varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t), \boldsymbol{u}_{PM}(x,t,\epsilon)-\boldsymbol{y}_0\rangle\\
&\quad + \left\|\varphi_J(\boldsymbol{y}_0|\boldsymbol{x},t)\right\|_2 \mathbb{E}_J\left[\left\|\boldsymbol{y}-\boldsymbol{y}_0\right\|_2\right] + n\epsilon\\
&< +\infty.
\end{aligned} \tag{125}
$$

This shows that $\mathbb{E}_J\left[\left|\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{y}_0\rangle\right|\right] < +\infty$ for every $\boldsymbol{y}_0 \in \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 < \delta\}$ different from $\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)$. Now, let $\{\boldsymbol{e}_i\}_{i=1}^n$ denote the standard basis in $\mathbb{R}^n$ and let $\{\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_i\}_{i=1}^n$ denote the components of the vector $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)$, i.e., $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t) = (\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_1,\ldots,\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_n)$. Using (125) with the choice of $\boldsymbol{y}_0 = \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i$, which is contained in the open ball $\{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2 < \delta\}$ for each $i \in \{1,\ldots,n\}$, we get

$$
\begin{aligned}
0 &\leq \mathbb{E}_J\left[\left|\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - (\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i)\right\rangle\right|\right]\\
&= \mathbb{E}_J\left[\left|\left\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \frac{\delta}{2}\boldsymbol{e}_i\right\rangle\right|\right]\\
&= \frac{\delta}{2}\mathbb{E}_J\left[\left|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_i\right|\right]\\
&\leq 2n\epsilon - \left\langle \varphi_J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i \mid \boldsymbol{x},t), \frac{\delta}{2}e_i\right\rangle\\
&\quad + \left\|\varphi_J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i|\boldsymbol{x},t)\right\|_2\\
&\quad \mathbb{E}_J\left[\left\|\boldsymbol{y} - (\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i)\right\|_2\right]\\
&< +\infty.
\end{aligned} \tag{126}
$$

Using (126) and the norm inequality $\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2 \leq \sum_{i=1}^n |\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_i|$, we can bound $\mathbb{E}_J\left[\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2\right]$ as follows

$$
\begin{aligned}
0 &\leq \mathbb{E}_J\left[\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2\right]\\
&\leq \mathbb{E}_J\left[\sum_{i=1}^n |\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_i|\right]\\
&= \sum_{i=1}^n \mathbb{E}_J\left[|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)_i|\right]\\
&\leq 2n^2\epsilon - \sum_{i=1}^n \left\langle \varphi_J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) - \frac{\delta}{2}\boldsymbol{e}_i|\boldsymbol{x},t), \frac{\delta}{2}e_i\right\rangle\\
&\quad + \sum_{i=1}^n \left\|\varphi_J(\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon) + \frac{\delta}{2}\boldsymbol{e}_i|\boldsymbol{x},t)\right\|_2\\
&\quad \mathbb{E}_J\left[\left\|\boldsymbol{y} - (\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\frac{\delta}{2}\boldsymbol{e}_i)\right\|_2\right]\\
&< +\infty.
\end{aligned} \tag{127}
$$

This shows that $\mathbb{E}_J\left[\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2\right] < +\infty$. Finally, use (127), $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t) = \frac{\boldsymbol{y}-\boldsymbol{x}}{t} + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})$, and assumption (A3) to find

$$
\begin{aligned}
&\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right]\\
&= \mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}) + \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) - \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)\right\|_2\right]\\
&\leq \mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}) + \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)\right\|_2 + \left\|\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)\right\|_2\right]\\
&= \mathbb{E}_J\left[\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2\right] + \mathbb{E}_J\left[\left\|\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right\|_2\right]\\
&\leq \mathbb{E}_J\left[\|\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)\|_2\right]\\
&\quad + \frac{1}{tZ_J(\boldsymbol{x},t,\epsilon)}\int_{\mathbb{R}^n} \|\boldsymbol{y}-\boldsymbol{x}\|_2\, e^{-\frac{1}{2t\epsilon}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2} d\boldsymbol{y}\\
&< +\infty.
\end{aligned}
$$

This shows that $\mathbb{E}_J\left[\left\|\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right\|_2\right] < +\infty$.

E Proof of Proposition 4.5

Proof of (i): Let $\boldsymbol{x} \in \mathbb{R}^n$ and $t > 0$ and define the functions

$$
\begin{aligned}
\mathrm{dom}\,\partial J \ni \boldsymbol{y} &\mapsto \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) = \left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0}),\\
\mathrm{dom}\,\partial J \ni \boldsymbol{y} &\mapsto \Phi_J(\boldsymbol{y}|\boldsymbol{x},t) = \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y}).
\end{aligned}
$$

Note that for every $\boldsymbol{y} \in \mathbb{R}^n$, $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t)$ is a subgradient of the function $\boldsymbol{u} \mapsto \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u})$ evaluated at $\boldsymbol{u} = \boldsymbol{y}$.

Let $\boldsymbol{u} \in \mathrm{dom}\,\partial J$. The Bregman divergence of the function $\mathrm{dom}\,\partial J \ni \boldsymbol{y} \mapsto \Phi_J(\boldsymbol{y}|\boldsymbol{x},t)$ at $(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))$ is given by

$$
\begin{aligned}
&D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\\
&= \Phi_J(\boldsymbol{u}|\boldsymbol{x},t) - \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{u}\rangle + \Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\\
&\equiv \Phi_J(\boldsymbol{u}|\boldsymbol{x},t) - \Phi_J(\boldsymbol{y}|\boldsymbol{x},t) + \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{y}-\boldsymbol{u}\rangle,
\end{aligned}
$$

where the second equality follows by definition of the subdifferential (see Definition 6) and that $\varphi_J(\boldsymbol{y}|\boldsymbol{x},t) \in \partial\Phi_J(\boldsymbol{y},\boldsymbol{x},t)$.

Take the expected value with respect to the variable $\boldsymbol{y}$ over $\mathrm{dom}\,\partial J$ to find

$$
\begin{aligned}
&\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]\\
&= \Phi_J(\boldsymbol{u}|\boldsymbol{x},t) - \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{u}\rangle + \Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] \qquad (128)\\
&\equiv \Phi_J(\boldsymbol{u}|\boldsymbol{x},t) - \mathbb{E}_J\left[\Phi_J(\boldsymbol{y}|\boldsymbol{x},t) + \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{y}-\boldsymbol{u}\rangle\right].
\end{aligned}
$$

We claim that the expected value $\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ is finite. We will show this by proving, in turn, that the expected values $\mathbb{E}_J\left[\Phi_J(\boldsymbol{y}|\boldsymbol{x},t)\right]$ and $\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{y}-\boldsymbol{u}\rangle\right]$ are finite. Establishing the finiteness of $\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ will enable us to conclude that the expected value $\mathbb{E}_J\left[\Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ on the right hand side of the first equality of (128) is also finite.

First, using the definition of $\Phi_J(\boldsymbol{y}|\boldsymbol{x},t)$ we have

$$\mathbb{E}_J\left[\Phi_J(\boldsymbol{y}|\boldsymbol{x},t)\right] \equiv \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + J(\boldsymbol{y})\right].$$

The expected value $\mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right]$ is finite because we can use the definitions of the posterior mean estimate and inequality (48) (with $m \equiv 0$ in (48)) to express it as

$$\begin{aligned}
0 &< \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right] \\
&= \mathbb{E}_J\left[\frac{1}{2t}\|(\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)) - (\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon))\|_2^2\right] \\
&= \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\
&\quad + \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\
&\quad + 2\mathbb{E}_J\left[\langle \boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon), \boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\rangle\right] \\
&= \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2 \\
&\quad + \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\
&\quad + 2\left\langle \boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon), \mathbb{E}_J[\boldsymbol{y}]-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle \\
&= \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2 \\
&\quad + \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\
&\quad + 2\left\langle \boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\right\rangle \\
&= \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2 \\
&\quad + \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{y}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2\right] \\
&\leq \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)\|_2^2 + \frac{n\epsilon}{2}.
\end{aligned}$$

The expected value $\mathbb{E}_J[J(\boldsymbol{y})]$ is also finite because it is bounded by the set of inequalities (38) in Proposition 4.1. Hence, the expected value $\mathbb{E}_J\left[\Phi_J(\boldsymbol{y}|\boldsymbol{x},t)\right] \equiv \mathbb{E}_J\left[\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right] + \mathbb{E}_J[J(\boldsymbol{y})]$ is finite.

Second, note that the expected value $\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right]$ can be written as

$$\begin{aligned}
&\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\quad = \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle + \langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\quad = \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\qquad + \langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \mathbb{E}_J[\boldsymbol{y}]-\boldsymbol{u}\rangle \\
&\quad = \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\qquad + \langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle .
\end{aligned} \tag{129}$$

Apply the monotonicity property (47) to the expected value $\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right]$ (with $\boldsymbol{y}_0 \equiv \boldsymbol{u}$ in (47)) in the previous equation to find

$$\begin{aligned}
0 &\leq \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\leq n\epsilon - \langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle .
\end{aligned}$$

Add the term $\langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle$ on both sides of these inequalities to get

$$\begin{aligned}
&\langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle \\
&\quad \leq \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t) - \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \\
&\qquad + \langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle \leq n\epsilon.
\end{aligned} \tag{130}$$

Combine the inequalities (130) with the equality (129) to find

$$\begin{aligned}
&\langle \varphi_J(\boldsymbol{u}|\boldsymbol{x},t), \boldsymbol{u}_{PM}(\boldsymbol{x},t,\epsilon)-\boldsymbol{u}\rangle \\
&\quad \leq \mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right] \leq n\epsilon.
\end{aligned}$$

These bounds prove that the expected value $\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t), \boldsymbol{y}-\boldsymbol{u}\rangle\right]$ is finite.

The previous arguments show that the expected value $\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ is finite. Now, we claim that the expected value $\mathbb{E}_J\left[\langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{u}\rangle\right] \equiv \langle \mathbb{E}_J\left[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right], \boldsymbol{u}\rangle$ is finite. Indeed, we can use the representation formula (30) for expressing the posterior mean estimate in terms of the gradient $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t)$ of the solution to the viscous HJ PDE (29) and use that $\mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]$ is finite (Proposition 4.3) to write

$$\begin{aligned}
&\mathbb{E}_J\left[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] \\
&\quad = \mathbb{E}_J\left[\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right) + \pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right] \\
&\quad = \mathbb{E}_J\left[\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{t}\right)\right] + \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right] \\
&\quad \equiv -\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) + \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right],
\end{aligned}$$

where both terms on the right hand side are finite. This shows that $\mathbb{E}_J\left[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ is finite.

Using that $\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ and $\mathbb{E}_J\left[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ are finite in Eq. (128), we conclude that the expected value $\mathbb{E}_J\left[\Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$ is also finite. We can now use the definitions of Φ_J and φ_J to express Eq. (128) as

$$\begin{aligned}
&\mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] \\
&\quad = \mathbb{E}_J\left[\Phi_J(\boldsymbol{u}|\boldsymbol{x},t) - \langle \varphi_J(\boldsymbol{y}|\boldsymbol{x},t)), \boldsymbol{u}\rangle + \Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] \\
&\quad = \frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + J(\boldsymbol{u}) \\
&\qquad + \left\langle \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) - \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right], \boldsymbol{u}\right\rangle \\
&\qquad + \mathbb{E}_J\left[\Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right],
\end{aligned} \tag{131}$$

where, again, we used that $\mathbb{E}_J\left[\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] = -\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) + \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]$. Now, let

$$\tilde{J}(\boldsymbol{u}) = J(\boldsymbol{u}) + \left\langle \nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) - \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right], \boldsymbol{u}\right\rangle.$$

Take the infimum over $\boldsymbol{u} \in \mathbb{R}^n$ on both sides of Eq. (131) to find:

$$\begin{aligned}
&\inf_{\boldsymbol{u}\in\mathbb{R}^n} \mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right] \\
&\quad = \inf_{\boldsymbol{u}\in\mathbb{R}^n}\left\{\frac{1}{2t}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + \tilde{J}(\boldsymbol{u})\right\} + \mathbb{E}_J\left[\Phi_J^*(\varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]
\end{aligned}$$

Now, note that by assumption (A1) $\boldsymbol{y} \mapsto J(\boldsymbol{y}) \in \Gamma_0(\mathbb{R}^n)$, and therefore the function $\boldsymbol{u} \mapsto \tilde{J}(\boldsymbol{u}) \in \Gamma_0(\mathbb{R}^n)$. Therefore, the function $u \ni \mathbb{R}^n \to\mapsto \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + \tilde{J}(u)$ is strictly convex and has a unique minimizer denoted by $\bar{\boldsymbol{u}}$. Therefore, the infimum in the equality above can be replaced by a minimum. In addition, recall that $\min_{u\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2 + \tilde{J}(u)$ corresponds to the solution to the first-order HJ PDE (20) with initial condition $\tilde{J}$. Using Proposition 2.2(ii), the unique minimizer $\bar{\boldsymbol{u}}$ can be expressed using the inclusion relation

$$\left(\frac{\boldsymbol{x}-\bar{\boldsymbol{u}}}{t}\right) \in \partial J(\bar{\boldsymbol{u}}) + \left(\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) - \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]\right). \tag{132}$$

Therefore, the minimizer $\bar{\boldsymbol{u}}$ is also the unique minimizer to $\boldsymbol{u} \mapsto \mathbb{E}_J\left[D_{\Phi_J}(\boldsymbol{u}, \varphi_J(\boldsymbol{y}|\boldsymbol{x},t))\right]$.

Proof of (ii): If dom $J = \mathbb{R}^n$, then the representation formula $\nabla_{\boldsymbol{x}} S_\epsilon(\boldsymbol{x},t) = \mathbb{E}_J\left[\pi_{\partial J(\boldsymbol{y})}(\boldsymbol{0})\right]$ derived in Proposition 4.2 holds and the characterization of the unique minimizer $\bar{\boldsymbol{u}}$ in equation (132) reduces to

$$\left(\frac{\boldsymbol{x}-\bar{\boldsymbol{u}}}{t}\right) \in \partial J(\bar{\boldsymbol{u}}).$$

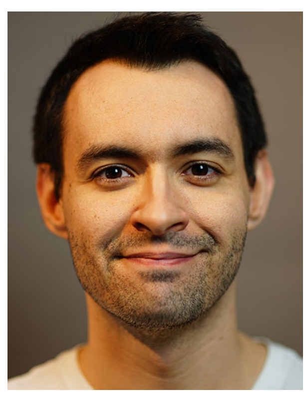
By Proposition 2.2(ii), the unique minimizer that satisfies this characterization is the MAP estimate $\boldsymbol{u}_{MAP}(\boldsymbol{x},t)$, i.e., $\bar{\boldsymbol{u}} = \boldsymbol{u}_{MAP}(\boldsymbol{x},t)$. □

References

1. Alberti, G., Ambrosio, L., Cannarsa, P.: On the singularities of convex functions. Manuscripta Math. **76**(3–4), 421–435 (1992)
2. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, vol. 147. Springer Science & Business Media, Berlin (2006)
3. Aubin, J.P., Cellina, A.: Differential inclusions: set-valued maps and viability theory, vol. 264. Springer Science & Business Media, Berlin (2012)
4. Banerjee, A., Guo, X., Wang, H.: On the optimality of conditional expectation as a Bregman predictor. IEEE Trans. Inform. Theory **51**(7), 2664–2669 (2005)
5. Boncelet, C.: Chapter 7 - image noise models. In: Bovik, A. (ed.) The Essential Guide to Image Processing, pp. 143–167. Academic Press, Cambridge (2009)
6. Bouman, C., Sauer, K.: A generalized Gaussian image model for edge-preserving map estimation. IEEE Trans. Image Process. **2**(3), 296–310 (1993)
7. Boyat, A.K., Joshi, B.K.: A review paper: noise models in digital image processing. Signal Image Process. Int. J. (SIPIJ) **6**(2), 63–75 (2015)
8. Burger, M., Dong, Y., Sciacchitano, F.: Bregman cost for non-Gaussian noise (2016). arXiv preprint arXiv:1608.07483
9. Burger, M., Lucka, F.: Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. Inverse Probl. **30**(11), 114004 (2014)
10. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)
11. Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. Int. J. Comput. Vis. **84**(3), 288 (2009)
12. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**(2), 167–188 (1997)
13. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. **25**, 161–319 (2016)
14. Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. SIAM J. Sci. Comput. **22**(2), 503–516 (2000)
15. Chaudhari, P., Choromanska, A., Soatto, S., LeCun, Y., Baldassi, C., Borgs, C., Chayes, J., Sagun, L., Zecchina, R.: Entropy-sgd: Biasing gradient descent into wide valleys. J. Stat. Mech. Theory Exp. **2019**(12), 124018 (2019)
16. Chaudhari, P., Oberman, A., Osher, S., Soatto, S., Carlier, G.: Deep relaxation: partial differential equations for optimizing deep neural networks. Res. Math. Sci. **5**(3), 30 (2018)
17. Darbon, J.: On convex finite-dimensional variational methods in imaging sciences and Hamilton-Jacobi equations. SIAM J. Imaging Sci. **8**(4), 2268–2293 (2015)
18. Darbon, J., Meng, T.: On decomposition models in imaging sciences and multi-time Hamilton-Jacobi partial differential equations. SIAM J. Imaging Sci. **13**(2), 971–1014 (2020)
19. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: Fast and exact optimization. J. Math. Imaging Vis. **26**(3), 261–276 (2006)
20. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. A J. Issued Courant Insti. Math. Sci. **57**(11), 1413–1457 (2004)
21. Demoment, G.: Image reconstruction and restoration: overview of common estimation structures and problems. IEEE Trans. Acoust. Speech Signal Process. **37**(12), 2024–2036 (1989)
22. Deuschel, J.D., Stroock, D.: Large Deviations, Pure and Applied Mathematics, vol. 137. Academic Press, Cambridge (1989)
23. Dobson, D.C., Santosa, F.: Recovery of blocky images from noisy and blurred data. SIAM J. Appl. Math. **56**(4), 1181–1198 (1996)
24. Donoho, D.: Compressed sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)
25. Durand, S., Malgouyres, F., Rougé, B.: Image deblurring, spectrum interpolation and application to satellite imaging. ESAIM Control Optim. Calc. Var. **5**, 445–475 (2000)
26. Evans, L.: Partial Differential Equations. American Mathematical Society, Providence, RI (2010)
27. Falconer, K.J.: The Geometry of Fractal Sets. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1985). https://doi.org/10.1017/CBO9780511623738
28. Federer, H.: Geometric Measure Theory. Springer, Berlin (1969)
29. Figueiredo, M.A., Nowak, R.D.: Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior. IEEE Trans. Image Process. **10**(9), 1322–1331 (2001)
30. Folland, G.B.: Real Analysis: Modern Techniques and Their Applications. John Wiley & Sons, Hoboken (2013)
31. García Trillos, N., Kaplan, Z., Sanz-Alonso, D.: Variational characterizations of local entropy and heat regularization in deep learning. Entropy **21**(5), 511 (2019)
32. Giusti, E.: Minimal Surfaces and Functions of Bounded Variation, Monographs in Mathematics, vol. 80. Birkhäuser Verlag, Basel (1984)
33. Gradshtein, I., Ryzhik, I., Jeffrey, A., Zwillinger, D.: Table of integrals, series and products. Academic Press, Cambridge (2007)
34. Gribonval, R.: Should penalized least squares regression be interpreted as maximum a posteriori estimation? IEEE Trans. Signal Process. **59**(5), 2405–2410 (2011)
35. Gribonval, R., Machart, P.: Reconciling "priors" & "priors" without prejudice? In: Advances in Neural Information Processing Systems, pp. 2193–2201 (2013)
36. Gribonval, R., Nikolova, M.: A characterization of proximity operators. J. Math. Imaging Vis. **62**, 773–789 (2020). https://doi.org/10.1007/s10851-020-00951-y
37. Gribonval, R., Nikolova, M.: On Bayesian estimation and proximity operators. Appl. Comput. Harmon. Anal. **50**, 49–72 (2021). https://doi.org/10.1016/j.acha.2019.07.002
38. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I: Fundamentals, Grundlehren Text Editions, vol. 305. Springer Science & Business Media, Berlin (1993)

39. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods, Grundlehren Text Editions, vol. 306. Springer Science & Business Media, Berlin (1993)
40. Hiriart-Urruty, J.B., Plazanet, P.H.: Moreau's decomposition theorem revisited. In: Annales de l'Institut Henri Poincare (C) Non Linear Analysis, vol. 6, pp. 325–338. Elsevier (1989)
41. Hochbaum, D.S.: An efficient algorithm for image segmentation, Markov random fields and related problems. J. ACM (JACM) **48**(4), 686–701 (2001)
42. Hörmander, L.: The analysis of linear partial differential operators. I. Classics in Mathematics. Springer-Verlag, Berlin (2003). https://doi.org/10.1007/978-3-642-61497-2. Distribution theory and Fourier analysis, Reprint of the second (1990) edition [Springer, Berlin; MR1065993 (91m:35001a)]
43. Kay, S.M.: Fundamentals of Statistical Signal Processing. Prentice Hall PTR, Hoboken (1993)
44. Keener, R.W.: Theoretical Statistics: Topics for a Core Course. Springer, Berlin (2011)
45. Lang, R.: A note on the measurability of convex sets. Arch. Math. (Basel) **47**(1), 90–92 (1986)
46. Leindler, L.: On a certain converse of Hölder's inequality inequality. In: Linear Operators and Approximation, Lineare Operatoren und Approximation, pp. 182–184. Springer (1972)
47. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
48. Louchet, C.: Modèles variationnels et bayésiens pour le débruitage d'images: de la variation totale vers les moyennes non-locales. Ph.D. thesis, Université René Descartes-Paris V (2008)
49. Louchet, C., Moisan, L.: Posterior expectation of the total variation model: properties and experiments. SIAM J. Imaging Sci. **6**(4), 2640–2684 (2013)
50. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. de la Société mathématique de France **93**, 273–299 (1965)
51. Nikolova, M.: Weakly constrained minimization: application to the estimation of images and signals involving constant regions. J. Math. Imaging Vis. **21**(2), 155–175 (2004)
52. Nikolova, M.: Model distortions in Bayesian map reconstruction. Inverse Probl. Imaging **1**(2), 399 (2007)
53. Pereyra, M.: Revisiting maximum-a-posteriori estimation in log-concave models. SIAM J. Imaging Sci. **12**(1), 650–670 (2019)
54. Pfeffer, W.F.: Divergence theorem for vector fields with singularities. In: New Integrals, pp. 150–166. Springer (1990)
55. Phillips, D.L.: A technique for the numerical solution of certain integral equations of the first kind. J. ACM (JACM) **9**(1), 84–97 (1962)
56. Prékopa, A.: Logarithmic concave measures with application to stochastic programming. Acta Sci. Math. **32**, 301–316 (1971)
57. Roberts, G.O., Rosenthal, J.S., et al.: Optimal scaling for various Metropolis-Hastings algorithms. Stat. Sci. **16**(4), 351–367 (2001)
58. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
59. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317. Springer-Verlag, Berlin (2009)
60. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(1–4), 259–268 (1992)
61. Rudin, W.: Real and Complex Analysis. Tata McGraw-Hill education, New York (2006)
62. Stuart, A.M.: Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451–559 (2010)
63. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2005)
64. Tikhonov, A.N., Goncharsky, A., Stepanov, V., Yagola, A.G.: Numerical Methods for the Solution of Ill-Posed Problems, vol. 328. Springer Science & Business Media, Berlin (1995)
65. Vidal, R., Bruna, J., Giryes, R., Soatto, S.: Mathematics of deep learning (2017). arXiv preprint arXiv:1712.04741
66. Vogel, C.R.: Computational Methods for Inverse Problems, Frontiers in Applied Mathematics, vol. 23. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2002)
67. Winkler, G.: Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: a Mathematical Introduction, Applications of Mathematics (New York), vol. 27, 2nd edn. Springer-Verlag, Berlin (2003)
68. Woodford, O.J., Rother, C., Kolmogorov, V.: A global perspective on map inference for low-level vision. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2319–2326. IEEE (2009)
69. Zhou, X.: On the Fenchel duality between strong convexity and Lipschitz continuous gradient (2018). arXiv preprint arXiv:1803.06573

Jérôme Darbon is an associate professor of Applied Mathematics at Brown University. He is currently on leave from CNRS (at Ecole Normale Superieure Paris-Saclay). His current research interest includes fast algorithms for solving high-dimensional Hamilton-Jacobi partial differential equations and optimal control problems, image processing and optimization algorithms.



Gabriel P. Langlois is a Ph.D. candidate in Jérôme Darbon's research group in the Division of Applied Mathematics at Brown University. He is interested in machine learning and optimization, and his current research interests focus on the development of efficient supervised machine learning algorithms tailored to handle big datasets. Prior to his doctoral studies, he earned a master's degree in applied mathematics from the Swiss Federal Institute of Technology in Zürich in 2015, and he earned a bachelor of science in applied mathematics and physics in 2013 from McGill University.