# Bayesian inference via generalized thermodynamic integration

F. Llorente[†], L. Martino[⋆], D. Delgado[†]

[†] Universidad Carlos III de Madrid, Leganés (Spain).

[⋆] Universidad Rey Juan Carlos, Fuenlabrada (Spain).

**Abstract**

The idea of using a path of tempered posterior distributions has been widely applied in the literature for the computation of marginal likelihoods (a.k.a., Bayesian evidence). Thermodynamic integration, path sampling and annealing importance sampling are well-known examples of algorithms belonging to this family of methods. In this work, we introduce a generalized thermodynamic integration (GTI) scheme which is able to perform a complete Bayesian inference, i.e., GTI can approximate generic posterior exceptions (not only the marginal likelihood). Several scenarios of application of GTI are discussed and different numerical simulations are provided.

**Keywords:** Bayesian inference; Thermodynamic integration; Target-aware inference; Tempering; Monte Carlo; Quadrature methods.

## 1 Introduction

Bayesian methods have become very popular in many domains of science and engineering over the last years, as they allow for obtaining estimates of parameters of interest as well as comparing competing models in a principled way [1, 2]. The Bayesian quantities can generally be expressed as integrals involving the posterior density. They can be divided in two main categories: posterior expectations and marginal likelihood computation (useful for model selection purposes).
Generally, computational methods are required for the approximation of these integrals, e.g., Monte Carlo algorithms such as Markov chain Monte Carlo (MCMC) and importance sampling (IS) [1, 2, 3]. The computation of the marginal likelihood is particularly complicated, specially with MCMC outputs [4, 5, 6]. For this reason, sophisticated and powerful schemes have been specifically designed [5, 7]. The most powerful techniques involves the idea of the so-called *tempering* of the posterior [8, 9, 10]. The tempering effect is commonly employed in order to foster the exploration and improve the efficiency of MCMC chains [11, 12]. State-of-the-art methods for computing marginal likelihoods consider tempered transitions (i.e. sequence of tempered distributions), such as annealed IS (An-IS) [8], sequential Monte Carlo (SMC) [13], thermodynamic integration (TI), a.k.a., path sampling (PS) or "power posteriors" (PP) in the statistics literature [9, 10, 14], and

stepping stones (SS) sampling [15]. In Figure 1, we give a graphical summary of the techniques for computing marginal likelihoods based on tempering: An-IS is a special case of SMC framework, PP is a special case of TI/PS, and SS sampling present similar features to An-IS and PP. For more details, see [5]. It is worth to mention that TI has been introduced in the physics literature for computing free-energy differences [16, 14].

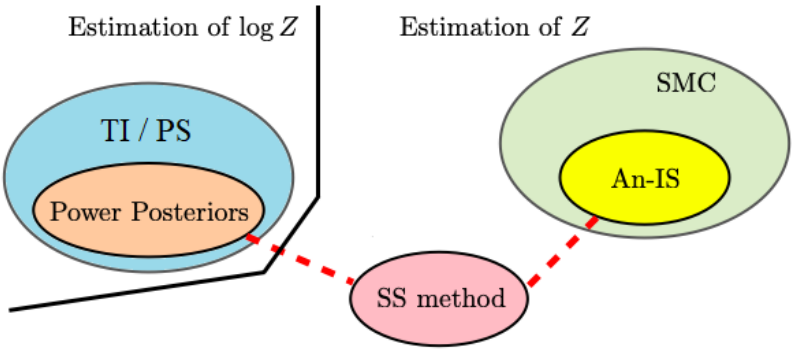



Figure 1: Graphical summary of computational approaches based on tempering for marginal likelihood estimation. Here, $Z$ denotes the marginal likelihood. Note that some of the methods estimate $Z$ in log-scale.

In this work, we extend the TI method, introducing the generalized thermodynamic integration (GTI) technique, for computing generic integrals involving the posterior distribution, as posterior expectations (not only the marginal likelihood). The extension of TI for the computation of a generic posterior expectation $\mathbb{E}_{\bar{\pi}}[f(\mathbf{x})]$ is not straightforward, since it requires to build a continuous path between densities with possibly different support. In the case of a geometric path (which is the default choice in practice [10, 9]), the generalization of TI needs a careful look at the support of the negative and positive parts of $f(\mathbf{x})$. We discuss the application of GTI for the computation of posterior expectations of generic real-valued function $f(\mathbf{x})$, and also describe the case of vector-valued function $\mathbf{f}(\mathbf{x})$. The benefits of GTI are clearly shown by illustrative numerical simulations. The structure of the paper is the following. In Section 2, we introduce the Bayesian inference setting and describe the thermodynamic method for the computation of the marginal likelihood. In Section 3, we introduce the GTI procedure. More specifically, we discuss first the case when $f(\mathbf{x})$ is strictly in Section 3.2, and then consider the general case of a real-valued $f(\mathbf{x})$ in Section 3.3. In Section 4, we discuss some computational details of the approach, and the application of GTI for vector-valued functions $\mathbf{f}(\mathbf{x})$. We show the benefits of GTI in two numerical experiments in Section 5. Finally, Section 6 contains the conclusions.

# 2 Background

## 2.1 Bayesian inference

In many real world applications, the goal is to infer a variable of interest given a set of data [1]. Let us denote the parameter of interest by $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$, and let $\mathbf{y} \in \mathbb{R}^{d_y}$ be the observed data.

In a Bayesian analysis, all the statistical information is contained in the posterior distribution, which is given by

$$\bar{\pi}(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{Z(\mathbf{y})}, \tag{1}$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the Bayesian model evidence (a.k.a. marginal likelihood). The marginal likelihood $Z(\mathbf{y})$ is important for model selection purposes [5, 17]. Generally, $Z(\mathbf{y})$ is unknown, so we are able to evaluate the unnormalized target function, $\pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})$. The analytical computation of the posterior density $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$ is often unfeasible, hence numerical approximations are needed. Our goal is to approximate integrals of the form

$$I = \mathbb{E}_{\bar{\pi}}\left[f(\mathbf{x})\right] = \int_{\mathcal{X}} f(\mathbf{x})\bar{\pi}(\mathbf{x})d\mathbf{x} = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{2}$$

where $f(\mathbf{x})$ is some integrable function, and

$$Z = \int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}. \tag{3}$$

The quantity $Z$ is called *marginal likelihood* (a.k.a., *Bayesian evidence*) and is useful for model selection purpose [5]. Generally, $I$ is analytically intractable and we need to resort to numerical algorithms such as Markov chain Monte Carlo (MCMC) and importance sampling algorithms.

## 2.2 Thermodynamic integration for estimating $Z$

Thermodynamic integration (TI) is a powerful technique that has been proposed in literature for computing ratios of constants [16, 14, 9]. Here, for simplicity, we focus on the approximation of just one constant, the marginal likelihood $Z$. More precisely, TI produces an estimation of $\log Z$. Let us consider a family of (generally unnormalized) densities

$$\pi(\mathbf{x}|\beta), \quad \beta \in [0,1], \tag{4}$$

such that $\pi(\mathbf{x}|0) = g(\mathbf{x})$ is the prior and $\pi(\mathbf{x}|1) = \pi(\mathbf{x})$ is the unnormalized posterior distribution. An example is the so-called *geometric path* $\pi(\mathbf{x}|\beta) = g(\mathbf{x})^{1-\beta}\pi(\mathbf{x})^{\beta}$, with $\beta \in [0,1]$ [18]. The corresponding normalized densities in the family are denoted as

$$\bar{\pi}(\mathbf{x}|\beta) = \frac{\pi(\mathbf{x}|\beta)}{c(\beta)}, \quad c(\beta) = \int_{\mathcal{X}} \pi(\mathbf{x}|\beta)d\mathbf{x}. \tag{5}$$

Then, the main TI identity is [5]

$$\begin{aligned} \log Z &= \int_0^1 \left[\frac{\partial \log \pi(\mathbf{x}|\beta)}{\partial \beta}\bar{\pi}(\mathbf{x}|\beta)d\mathbf{x}\right] d\beta \\ &= \int_0^1 \mathbb{E}_{\bar{\pi}(\mathbf{x}|\beta)}\left[\frac{\partial \log \pi(\mathbf{x}|\beta)}{\partial \beta}\right] d\beta, \end{aligned} \tag{6}$$

where the expectation is with respect to (w.r.t.) $\bar{\pi}(\mathbf{x}|\beta) = \frac{\pi(\mathbf{x}|\beta)}{c(\beta)}$.

**TI estimator: Quadrature + Monte Carlo.** Using a sequence of discrete values $\{\beta_i\}_{i=1}^N$ (e.g. $\beta_i$'s uniformly in $[0,1]$), one can approximate the integral in Eq. (6) via quadrature w.r.t. $\beta$, and then approximate the inner expectation with a Monte Carlo estimator using samples from $p(\mathbf{x}|\beta_i)$ for $i = 1, \ldots, N$. Namely, defining $U(\mathbf{x}) = \frac{\partial \log \pi(\mathbf{x}|\beta)}{\partial \beta}$ and $E(\beta) = \mathbb{E}_{\bar{\pi}(\mathbf{x}|\beta)}[U(\mathbf{x})]$, the resulting estimator of Eq. (6) is given by

$$\log Z \approx \sum_{i=1}^{N} (\beta_{i+1} - \beta_i) \widehat{E}_i, \tag{7}$$

where

$$\widehat{E}_i = \frac{1}{N} \sum_{j=1}^{N} U(\mathbf{x}_{i,j}), \quad \mathbf{x}_{i,j} \sim p(\mathbf{x}|\beta_i). \tag{8}$$

Note that we used the simplest quadrature rule in Eq. (7), but others can be used such as Trapezoidal, Simpson's, etc [10, 9].

**The power posteriors (PP) method.** Let us consider the specific case of a geometric path between prior $g(\mathbf{x})$ and unnomalized posterior $\pi(\mathbf{x})$,

$$\pi(\mathbf{x}|\beta) = g(\mathbf{x})^{1-\beta} \pi(\mathbf{x})^{\beta} = g(\mathbf{x}) \left[ \frac{\pi(\mathbf{x})}{g(\mathbf{x})} \right]^{\beta}, \tag{9}$$

$$= g(\mathbf{x}) \ell(\mathbf{y}|\mathbf{x})^{\beta}, \quad \beta \in [0,1], \tag{10}$$

where we have used $\pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x}) g(\mathbf{x})$. Note that, in this scenario,[1]

$$\frac{\partial \log \pi(\mathbf{x}|\beta)}{\partial \beta} = \log \ell(\mathbf{y}|\mathbf{x}). \tag{11}$$

Hence, the identity in Eq. (6) can be also written as

$$\log Z = \int_0^1 \int_{\mathcal{X}} \log \ell(\mathbf{y}|\mathbf{x}) \bar{\pi}(\mathbf{x}|\beta) d\mathbf{x} d\beta = \int_0^1 \mathbb{E}_{\bar{\pi}(\mathbf{x}|\beta)}[\log \ell(\mathbf{y}|\mathbf{x})] d\beta, \tag{12}$$

The *power posteriors (PP) method* is a special case of TI which considers **(a)** the geometric path and **(b)** trapezoidal quadrature rule for integrating w.r.t. the variable $\beta$ [10]. Namely, letting $\beta_1 = 0 < \cdots < \beta_N = 1$ denote a fixed temperature schedule, an approximation of Eq. (12) can be obtained via the trapezoidal rule

$$\log Z \approx \sum_{i=1}^{N-1} (\beta_{i+1} - \beta_i) \frac{\mathbb{E}_{\bar{\pi}(\mathbf{x}|\beta_{i+1})}[\log \ell(\mathbf{y}|\mathbf{x})] + \mathbb{E}_{\bar{\pi}(\mathbf{x}|\beta_i)}[\log \ell(\mathbf{y}|\mathbf{x})]}{2}, \tag{13}$$

[1] From Eq. (10), we can write $\log \pi(\mathbf{x}|\beta) = \log g(\mathbf{x}) + \beta \log \ell(\mathbf{y}|\mathbf{x})$. Hence, $\frac{\partial \log \pi(\mathbf{x}|\beta)}{\partial \beta} = \log \ell(\mathbf{y}|\mathbf{x})$.

where the the expectations are generally substituted with MCMC estimates as in Eq. (8). TI and PP are popular methods for computing marginal likelihoods (even in high-dimensional spaces) due to their reliability. Theoretical properties are studied in [14, 19], and empirical validation is provided in several works, e.g., [10, 9]. Different extensions and improvements on the method have also been proposed [20, 21, 19].

**Remark 1.** *Note that, in order to ensure that the integrand in Eq.* (12) *is finite, so that the estimator in Eq.* (13) *can be applied, we need that (a)* $\ell(\mathbf{y}|\mathbf{x})$ *is strictly positive everywhere, or (b)* $\ell(\mathbf{y}|\mathbf{x}) = 0$ *only whenever* $g(\mathbf{x}) = 0$ *(i.e., they have the same support).*

**Goal**. We have seen that the TI method has been proposed for computing $\log Z$ (or log-ratios of constants). Our goal is to extend the TI scheme in order to obtain a complete Bayesian inference analysis. Namely, we generalize the idea of these methods (thermodynamic integration, power posteriors, etc.) to the computation of posterior expectations for a given $f(\mathbf{x})$. See Figure 2 for a graphical summary.

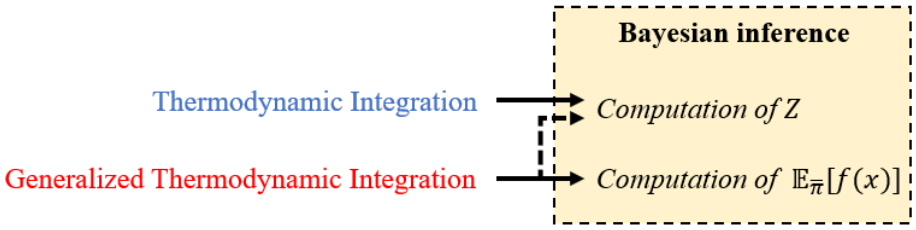



Figure 2: Thermodynamic integration has been employed for the computation of the marginal likelihood $Z$. In this work, we introduce the generalized thermodynamic integration procedure for computing also posterior expectations $\mathbb{E}_{\bar{\pi}}[f(\mathbf{x})]$.

# 3 Generalized TI (GTI) for Bayesian inference

In this section, we extend the TI method for computing any posterior expectation. The basic idea, as we show below, is the formulation $I$ as a ratio of normalizing constants. First, we consider the case $f(\mathbf{x}) > 0$ for all $\mathbf{x}$ and then the case of a generic real-valued $f(\mathbf{x})$.

## 3.1 General approach

In order to apply TI, we need to formulate the posterior expectation $I$ as a ratio of two constants. Since $f(\mathbf{x})$ can be positive or negative, let us consider the positive and negative parts, $f_+(\mathbf{x}) = \max(0, f(\mathbf{x}))$ and $f_-(\mathbf{x}) = \min(0, -f(\mathbf{x}))$, such that $f(\mathbf{x}) = f_+(\mathbf{x}) - f_-(\mathbf{x})$, where $f_+(\mathbf{x})$ and $f_-(\mathbf{x})$ are non-negative functions. The following identity shows that any posterior expectation can be expressed in terms of *two quotients* (formed by three values),

$$I = \frac{\int_{\mathcal{X}} f_+(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}} - \frac{\int_{\mathcal{X}} f_-(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}} = \frac{c_+}{Z} - \frac{c_-}{Z}, \tag{14}$$

where $c_+ = \int_{\mathcal{X}} \varphi_+(\mathbf{x})d\mathbf{x}$ are $c_- = \int_{\mathcal{X}} \varphi_-(\mathbf{x})d\mathbf{x}$ are respectively the normalizing constants of $\varphi_+(\mathbf{x}) = f_+(\mathbf{x})\pi(\mathbf{x})$, and $\varphi_-(\mathbf{x}) = f_-(\mathbf{x})\pi(\mathbf{x})$.

**Proposed scheme.** Denoting $\eta_+ = \log \frac{c_+}{Z}$ and $\eta_- = \log \frac{c_-}{Z}$ in the case of a generic $f(\mathbf{x})$, we propose to obtain estimates of these quantities using thermodynamic integration. Then, we can obtain the final estimator as

$$\widehat{I} = \exp\left(\widehat{\eta}_+\right) - \exp\left(\widehat{\eta}_-\right). \tag{15}$$

In the next section, we give details on how to compute $\widehat{\eta}_+$, $\widehat{\eta}_-$ by using a generalized TI method.

## 3.2 GTI for strictly positive or strictly negative $f(\mathbf{x})$

Let us consider the scenario where $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{X}$. In this scenario, we can set

$$f_+(\mathbf{x}) = f(\mathbf{x}) > 0, \quad \text{and} \quad \widehat{I} = \exp\left(\widehat{\eta}_+\right).$$

Note that, with respect to Eq. (15), we only consider the first term. Hence, we link the unnormalized pdfs $\pi(\mathbf{x})$ and $\varphi_+(\mathbf{x}) = f_+(\mathbf{x})\pi(\mathbf{x})$ with a geometric path, by defining

$$\bar{\varphi}_+(\mathbf{x}|\beta) \propto \varphi_+(\mathbf{x}|\beta) = f_+(\mathbf{x})^\beta \pi(\mathbf{x}), \quad \beta \in [0,1]. \tag{16}$$

Hence, we have $\bar{\varphi}_+(\mathbf{x}|0) = \bar{\pi}(\mathbf{x})$ and $\bar{\varphi}_+(\mathbf{x}|1) = \frac{1}{c_+} f_+(\mathbf{x})\pi(\mathbf{x})$ where $c_+ = \int_{\mathcal{X}} f_+(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}$. The Eq. (6) is thus

$$\eta_+ = \int_0^1 \mathbb{E}_{\bar{\varphi}_+(\mathbf{x}|\beta)}[\log f_+(\mathbf{x})]d\beta. \tag{17}$$

Letting $\beta_1 = 0 < \cdots < \beta_N = 1$ denote a fixed temperature schedule, the estimator (using the Trapezoidal rule) is thus

$$\widehat{\eta}_+ = \sum_{i=1}^{N-1} (\beta_{i+1} - \beta_i) \frac{\mathbb{E}_{\bar{\varphi}_+(\mathbf{x}|\beta_{i+1})}[\log f_+(\mathbf{x})] + \mathbb{E}_{\bar{\varphi}_+(\mathbf{x}|\beta_i)}[\log f_+(\mathbf{x})]}{2}, \tag{18}$$

where we use MCMC estimates for the terms

$$\mathbb{E}_{\bar{\varphi}_+(\mathbf{x}|\beta_i)}[\log f_+(\mathbf{x})] = \int_{\mathcal{X}} \log f_+(\mathbf{x})\bar{\varphi}_+(\mathbf{x}|\beta_i)d\mathbf{x} \approx \frac{1}{M}\sum_{m=1}^{M} \log f_+(\mathbf{x}_m), \quad \mathbf{x}_m \sim \bar{\varphi}_+(\mathbf{x}|\beta_i), \tag{19}$$

for $i = 1, \ldots, N$. The case of a strictly negative $f(\mathbf{x})$, i.e., $f_-(\mathbf{x}) = -f(\mathbf{x})$, is equivalent.

**Function $f(\mathbf{x})$ with zeros with null measure.** So far, we have considered strictly positive or strictly negative $f(\mathbf{x})$. This case could be extended to a positive (or negative) $f(\mathbf{x})$ with zeros in a null measure set. Indeed, note that the identity in Eq. (17) requires that $\mathbb{E}_{\bar{\varphi}(\mathbf{x}|\beta)}[\log f(\mathbf{x})] < \infty$ for all $\beta \in [0,1]$. If the zeros of $f(\mathbf{x})$ has null measure and the improper integral converges, the procedure above is also suitable. Table 1 summarizes the Generalized TI (GTI) steps for $f(\mathbf{x})$ that are strictly positive. We discuss other scenarios in the next section.

Table 1: **GTI for strictly positive $f(\mathbf{x})$**

**- Initialization:** Choose the set of nodes $\{\beta_i\}_{i=1}^N$ (with $\beta_1 = 0$ and $\beta_N = 1$), and the number of iterations $N$.
**- For** $i = 1, \ldots, N$**:**

1. **Sampling:** Sample $\{\mathbf{x}_{i,m}\}_{m=1}^M \sim \bar{\varphi}_+(\mathbf{x}|\beta_i) \propto f(\mathbf{x})^{\beta_i}\pi(\mathbf{x})$.
2. **Compute:**

$$\widehat{E}_i = \frac{1}{M}\sum_{k=1}^{M} \log f(\mathbf{x}_{i,k}). \tag{20}$$

3. **Update:** If $i = 1$, set $\eta^{(i)} = 0$. If $i > 1$, update recursive estimate

$$\widehat{\eta}^{(i)} \leftarrow \widehat{\eta}^{(i-1)} + 0.5 \cdot (\beta_i - \beta_{i-1}) \left(\widehat{E}_i + \widehat{E}_{i-1}\right). \tag{21}$$

**- Outputs:** Final estimator $\widehat{\eta}^{(N)}$ which approximates $\log I$.

## 3.3 GTI for generic $f(\mathbf{x})$

Using the results from previous section, we apply GTI to a generic function $f(\mathbf{x})$, namely, it can be positive and negative, as well as having zero-valued regions with a non-null measure. Here, we desire to connect the posterior $\pi(\mathbf{x})$ with the $f_+(\mathbf{x})\pi(\mathbf{x})$ and $f_-(\mathbf{x})\pi(\mathbf{x})$ with two continuous paths. However, a requirement for the validity of the approach is that $\pi(\mathbf{x})$ is zero whenever $f_+(\mathbf{x})\pi(\mathbf{x})$ or $f_-(\mathbf{x})\pi(\mathbf{x})$ is zero, which does not generally fulfills as $f(\mathbf{x})$ can have a smaller support than $\pi(\mathbf{x})$. Therefore, we need to define the unnormalized restricted posteriors densities

$$\pi_+(\mathbf{x}) = \pi(\mathbf{x})\mathbb{1}_{\mathcal{X}_+}(\mathbf{x}), \quad \text{and} \quad \pi_-(\mathbf{x}) = \pi(\mathbf{x})\mathbb{1}_{\mathcal{X}_-}(\mathbf{x}), \tag{22}$$

where $\mathbb{1}_{\mathcal{X}_+}(\mathbf{x})$ is the indicator function over the set $\mathcal{X}_+ = \{\mathbf{x} \in \mathcal{X} : f_+(\mathbf{x}) > 0\}$ and $\mathbb{1}_{\mathcal{X}_-}(\mathbf{x})$ is the indicator function over the set $\mathcal{X}_- = \{\mathbf{x} \in \mathcal{X} : f_-(\mathbf{x}) > 0\}$. The idea is to connect with a path $\pi_+(\mathbf{x})$ and $f_+(\mathbf{x})\pi(\mathbf{x})$, and $\pi_-(\mathbf{x})$ with $f_-(\mathbf{x})\pi(\mathbf{x})$, by the densities

$$\bar{\varphi}_+(\mathbf{x}|\beta) \propto f_+(\mathbf{x})^\beta \pi_+(\mathbf{x}), \qquad \bar{\varphi}_-(\mathbf{x}|\beta) \propto f_-(\mathbf{x})^\beta \pi_-(\mathbf{x}), \quad \beta \in [0,1].$$

Defining also

$$Z_+ = \int_{\mathcal{X}} \pi_+(\mathbf{x})d\mathbf{x}, \quad Z_- = \int_{\mathcal{X}} \pi_-(\mathbf{x})d\mathbf{x}, \tag{23}$$

and recalling

$$c_+ = \int_{\mathcal{X}} f_+(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \quad c_- = \int_{\mathcal{X}} f_-(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{24}$$

the idea is to apply separately TI for approximating $\eta_+^{\text{res}} = \log \frac{c_+}{Z_+}$ and $\eta_-^{\text{res}} = \log \frac{c_-}{Z_-}$, where we denote with *res* to account that we consider the restricted components $Z_+$ and $Z_-$. Hence, two correction factors $R_+$ and $R_-$ are also required, in order to obtain $R_+ \exp\left(\eta_+^{\text{res}}\right) = \frac{c_+}{Z}$ and $R_- \exp\left(\eta_-^{\text{res}}\right) = \frac{c_-}{Z}$. Below, we also show how to estimate the correction factors at a final stage and combine them to the estimations of $\eta_+^{\text{res}}$ and $\eta_-^{\text{res}}$. We can approximate the quantities

$$\eta_+^{\text{res}} = \log \frac{c_+}{Z_+} = \int_0^1 \mathbb{E}_{\bar{\varphi}_+(\mathbf{x}|\beta)}[\log f_+(\mathbf{x})] d\beta,$$
$$\eta_-^{\text{res}} = \log \frac{c_-}{Z_-} = \int_0^1 \mathbb{E}_{\bar{\varphi}_-(\mathbf{x}|\beta)}[\log f_+(\mathbf{x})] d\beta,$$

using the estimators

$$\widehat{\eta}_+^{\text{res}} = \sum_{i=1}^{N-1} (\beta_{i+1} - \beta_i) \frac{\widehat{E}_{i+1}^+ + \widehat{E}_i^+}{2}, \tag{25}$$

$$\widehat{\eta}_-^{\text{res}} = \sum_{i=1}^{N-1} (\beta_{i+1} - \beta_i) \frac{\widehat{E}_{i+1}^- + \widehat{E}_i^-}{2}, \tag{26}$$

where

$$\widehat{E}_i^+ = \frac{1}{M} \sum_{m=1}^{M} \log f_+(\mathbf{x}_{i,m}), \quad \mathbf{x}_{i,m} \sim \bar{\varphi}_+(\mathbf{x}|\beta_i), \tag{27}$$

$$\widehat{E}_i^- = \frac{1}{M} \sum_{m=1}^{M} \log f_-(\mathbf{v}_{i,m}), \quad \mathbf{v}_{i,m} \sim \bar{\varphi}_-(\mathbf{x}|\beta_i). \tag{28}$$

When comparing the estimators in Eqs. (25)-(26) with respect to the GTI estimator in Eq. (18), here the only difference is that the expectation at $\beta = 0$ is approximated by using samples from the restricted posteriors, $\pi_+(\mathbf{x})$ and $\pi_-(\mathbf{x})$, instead of the posterior $\pi(\mathbf{x})$.[2] To obtain an approximation of the true quantities of interest $\eta_+$, $\eta_-$ (instead of $\eta_+^{\text{res}}$ and $\eta_-^{\text{res}}$), we compute two correction factors from a single set of $K$ samples from $\bar{\pi}(\mathbf{x})$ as follows

$$\widehat{R}_+ = \frac{1}{K} \sum_{i=1}^{K} \mathbb{1}_{\mathcal{X}_+}(\mathbf{z}_i) \approx \frac{Z_+}{Z}, \tag{29}$$

$$\widehat{R}_- = \frac{1}{K} \sum_{i=1}^{K} \mathbb{1}_{\mathcal{X}_-}(\mathbf{z}_i) \approx \frac{Z_-}{Z}, \quad \mathbf{z}_i \sim \bar{\pi}(\mathbf{x}), \tag{30}$$

where $\mathbb{1}_{\mathcal{X}_+}(\mathbf{x}_i) = 1$ if $f_+(\mathbf{x}_i) > 0$, $\mathbb{1}_{\mathcal{X}_-}(\mathbf{x}_i) = 1$ if $f_-(\mathbf{x}_i) > 0$, and both zero otherwise. The final estimator of $I$ is

$$\widehat{I} = \widehat{R}_+ \exp\left(\widehat{\eta}_+^{\text{res}}\right) - \widehat{R}_- \exp\left(\widehat{\eta}_-^{\text{res}}\right), \tag{31}$$

including the two correction factors. Table 2 provides all the details of GTI in this scenario.

[2] In order to obtain samples from $\pi_\pm(\mathbf{x})$, we just need to consider $\pi_\pm(\mathbf{x})$ as target density instead of $\pi(\mathbf{x})$, in the MCMC steps. A similar alternative procedure is to apply rejection sampling, discarding the samples from $\pi(\mathbf{x})$ such that $f_\pm(\mathbf{x}) = 0$.

Table 2: **GTI for generic functions $f(\mathbf{x})$**

**- Initialization:** Choose the set of nodes $\{\beta_i\}_{i=1}^N$ (with $\beta_1 = 0$ and $\beta_N = 1$), and the number of iterations $N$.
**- For** $i = 1, \ldots, N$**:**

1. **Sampling:** Sample

$$\{\mathbf{x}_{i,m}\}_{m=1}^M \sim \bar{\varphi}_+(\mathbf{x}|\beta_i) \propto f_+(\mathbf{x})^{\beta_i}\pi_+(\mathbf{x}), \quad \pi_+(\mathbf{x}) = \pi(\mathbf{x})\mathbb{1}_{\mathcal{X}_+}(\mathbf{x}),$$
$$\{\mathbf{v}_{i,m}\}_{m=1}^M \sim \bar{\varphi}_-(\mathbf{x}|\beta_i) \propto f_-(\mathbf{x})^{\beta_i}\pi_-(\mathbf{x}), \quad \pi_-(\mathbf{x}) = \pi(\mathbf{x})\mathbb{1}_{\mathcal{X}_-}(\mathbf{x}).$$

2. **Compute:**

$$\widehat{E}_i^+ = \frac{1}{M}\sum_{m=1}^M \log f_+(\mathbf{x}_{i,m}), \qquad \widehat{E}_i^- = \frac{1}{M}\sum_{m=1}^M \log f_-(\mathbf{v}_{i,m}).$$

3. **Update:** If $i = 1$, set $\eta_+^{(i)} = \eta_-^{(i)} = 0$ . If $i > 1$, update recursive estimates

$$\widehat{\eta}_+^{(i)} \leftarrow \widehat{\eta}_+^{(i-1)} + 0.5 \cdot (\beta_i - \beta_{i-1})\left(\widehat{E}_i^+ + \widehat{E}_{i-1}^+\right),$$
$$\widehat{\eta}_-^{(i)} \leftarrow \widehat{\eta}_-^{(i-1)} + 0.5 \cdot (\beta_i - \beta_{i-1})\left(\widehat{E}_i^- + \widehat{E}_{i-1}^-\right).$$

**- Correction:** Compute correction factor using samples $\{\mathbf{z}_k\}_{k=1}^K \sim \bar{\pi}(\mathbf{x})$,

$$\widehat{R}_+ = \frac{1}{K}\sum_{i=1}^K \mathbb{1}_{\mathcal{X}_+}(\mathbf{z}_i), \quad \widehat{R}_- = \frac{1}{K}\sum_{i=1}^K \mathbb{1}_{\mathcal{X}_-}(\mathbf{z}_i), \quad \mathbf{z}_i \sim \bar{\pi}(\mathbf{x}),$$

**- Outputs:** The final estimator

$$\widehat{I} = \widehat{R}_+ \exp\left(\widehat{\eta}_+^{(N)}\right) - \widehat{R}_- \exp\left(\widehat{\eta}_-^{(N)}\right),$$

**Remark 2. Standard TI as special case of GTI:** *Note that the GTI scheme contains TI as a special case if we set $f(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})$ (i.e., the likelihood function) and let the prior $g(\mathbf{x})$ play the role of $\pi(\mathbf{x})$. Since the likelihood $\ell(\mathbf{y}|\mathbf{x})$ is non-negative we have $\eta_- = -\infty$ (then, $\exp(\eta_-) = 0$), hence we only have to consider the estimation of $\eta_+$. Moreover, if $\ell(\mathbf{y}|\mathbf{x})$ is strictly positive we do not need to compute the correction factor.*

**Remark 3.** *The GTI procedure, described above, also allows the application of the standard TI for computing marginal likelihoods when the likelihood function is not strictly positive, by applying a correction factor in the same fashion (in this case, considering a restricted prior pdf).*

# 4 Computational considerations and other extensions

In this section, we discuss computational details, different scenarios and further extensions, that are listed below.

**Acceleration schemes.** In order to apply GTI, the user must set $N$ and $M$, so that the total number of samples/evaluations of $f(\mathbf{x})$ in Table 1 is $E = NM$. The evaluations of $f(\mathbf{x})$ in Table 2 are $E = 2NM + K$. We can reduce the cost of algorithm in Table 2 to $E = NM + K$ with an acceleration scheme. Instead of running separate MCMC algorithms for $\bar{\varphi}_+(\mathbf{x}|\beta) \propto f_+(\mathbf{x})^\beta \pi_+(\mathbf{x})$ and $\bar{\varphi}_-(\mathbf{x}|\beta) \propto f_-(\mathbf{x})^\beta \pi_-(\mathbf{x})$, we use a single run targeting

$$\bar{\varphi}_{\text{abs}}(\mathbf{x}|\beta) \propto |f(\mathbf{x})|^\beta \pi(\mathbf{x}) \mathbb{1}\left(f(\mathbf{x}) \neq 0\right). \tag{32}$$

We can obtain two MCMC samples, one from $\bar{\varphi}_+(\mathbf{x}|\beta)$ and one from $\bar{\varphi}_-(\mathbf{x}|\beta)$, by separating the sample into two: samples with positive value of $f(\mathbf{x})$, and samples with negative value of $f(\mathbf{x})$, respectively. The procedure can be repeated until obtaining the desired number of samples from each density, $\bar{\varphi}_+(\mathbf{x}|\beta)$ and $\bar{\varphi}_-(\mathbf{x}|\beta)$.

Moreover, note that in Table 2 we need to draw samples from $\pi_+(\mathbf{x})$, $\pi_-(\mathbf{x})$ and $\pi(\mathbf{x})$. Instead of sampling each one separately, we can use the following procedure. Obtain a set of samples from $\pi(\mathbf{x})$ and then apply rejection sampling (i.e. discard samples with $f_\pm(\mathbf{x}) = 0$) in order to obtain samples from $\pi_\pm(\mathbf{x})$. Combining this idea with the acceleration scheme above reduces the cost of Table 2 to $E = MN$.

**Parallelization.** Note that steps 1 and 2 in Table 1 and Table 2 are amenable to parallelization. In other words, those steps need not be performed sequentially but can be done using embarrassingly parallel MCMC chains (i.e. with no communication among $N$, or $2N$, workers). Only step 3 requires communicating to a central node and combining the estimates. With this procedure, the number of evaluations $E$ is the same but the computation time is reduced by $\frac{1}{N}$ (or $\frac{1}{2N}$) factor. On the other hand, population MCMC techniques can be used, but parallelization speedups are lower since communication among workers occurs every so often, in order to foster the exploration of the chains [22, 19].

**Vector-valued functions $\mathbf{f}(\mathbf{x})$.** In Bayesian inference, one is often interested in computing moments of the posterior, i.e.,

$$\mathbf{I} = \int_{\mathcal{X}} \mathbf{x}^\alpha \bar{\pi}(\mathbf{x}) d\mathbf{x}, \qquad \alpha \geq 1. \tag{33}$$

In this case $\mathbf{I}$ is a vector and $\mathbf{f}(\mathbf{x}) = \mathbf{x}^\alpha$. When $\alpha = 1$, $\mathbf{I}$ represents the *minimum mean square error* (MMSE) estimator. More generally, we can have a vector-valued function,

$$\mathbf{f}(\mathbf{x}) = \left[f_1(\mathbf{x}), \ldots, f_{d_f}(\mathbf{x})\right]^\top : \mathcal{X} \to \mathbb{R}^{d_f},$$

hence the integral of interest is a vector $\mathbf{I} = [I_1, \ldots, I_{d_f}]^\top$ where $I_i = \int_{\mathcal{X}} f_i(\mathbf{x}) \bar{\pi}(\mathbf{x}) d\mathbf{x}$. In this scenario, we need to apply the GTI scheme to each component of $\mathbf{I}$ separately, obtaining estimates $\widehat{I}_i$ of the form in Eq. (31).

**Alternative procedure avoiding corrections.** We have seen that we can apply GTI to compute the posterior expectation of a generic $f(\mathbf{x})$, that can be positive, negative and have

zero-valued regions. For doing this, we connected with a tempered path, $\pi_+(\mathbf{x})$ and $\pi_-(\mathbf{x})$, to $f_+(\mathbf{x})\pi(\mathbf{x})$ and $f_-(\mathbf{x})\pi(\mathbf{x})$ respectively and then apply correction factors.
An alternative procedure is to use *reference* distributions in order to compute separately $c_+$, $c_-$ and $Z$ in Eq. (14), which are the normalizing constants of the unnormalized pdfs $f_+(\mathbf{x})\pi(\mathbf{x})$, $f_-(\mathbf{x})\pi(\mathbf{x})$ and $\pi(\mathbf{x})$ respectively (similarly as suggested in [3]). Let us define as

$$p_i^{\text{ref}}(\mathbf{x}), \qquad i = 1, 2, 3,$$

three unnormalized *reference* densities with normalizing constants,

$$Z_i^{\text{ref}} = \int_{\mathcal{X}} p_i^{\text{ref}}(\mathbf{x}) d\mathbf{x}, \qquad i = 1, 2, 3.$$

Then, the idea is to apply GTI for obtaining estimates of $\log \frac{c_+}{Z_1^{\text{ref}}}$, $\log \frac{c_-}{Z_2^{\text{ref}}}$ and $\log \frac{Z}{Z_3^{\text{ref}}}$. A requirement is that $p_1^{\text{ref}}(\mathbf{x})$ is zero where $f_+(\mathbf{x})\pi(\mathbf{x})$ is zero, $p_2^{\text{ref}}(\mathbf{x})$ is zero where $f_-(\mathbf{x})\pi(\mathbf{x})$ is zero, and $p_3^{\text{ref}}(\mathbf{x})$ is zero where $\pi(\mathbf{x})$ is zero. Namely, we need to be able to build a continuous path between the reference distributions and the corresponding unnormalized pdf of interest. With this procedure, we do not need to apply correction factors, but we just need to apply the algorithm in Table 1 three times. Of course, the choice of the reference distributions affects the performance of the final estimator. Choosing reference distributions that are closer than what $\pi(\mathbf{x})$ is to $f_+(\mathbf{x})\pi(\mathbf{x})$ or $f_-(\mathbf{x})\pi(\mathbf{x})$, improves the results. The advantage of this alternative procedure is that, ideally, with $p_3^{\text{ref}}(\mathbf{x}) = \pi(\mathbf{x})$, $p_1^{\text{ref}}(\mathbf{x}) = f_+(\mathbf{x})\pi(\mathbf{x})$ and $p_2^{\text{ref}}(\mathbf{x}) = f_-(\mathbf{x})\pi(\mathbf{x})$, we can obtain a zero-variance estimator (see also [3] for discussion).

# 5 Numerical experiments

In this section, we illustrate the performance of the proposed scheme in two numerical experiments which consider different kind of densities $\bar{\pi}$ with different features and different dimensions, and also different function $f(\mathbf{x})$. In the first example, $f(\mathbf{x})$ is strictly positive so we apply the algorithm described in Table 1. In the second example, we consider $f(\mathbf{x})$ to have zero-valued regions, and hence we apply the algorithm in Table 2. It is important to remark that the setup of these numerical examples have been considered in other relevant works (e.g., see [3, 23, 24]).

## 5.1 First numerical analysis

Let us consider the following Gaussian model [3]

$$g(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{0}_D, \mathbf{I}_D), \quad \ell(\mathbf{y}|\mathbf{x}) = \mathcal{N}\left(-\frac{y}{\sqrt{D}}\mathbf{1}_D \middle| \mathbf{x}, \mathbf{I}_D\right), \quad f(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \middle| \frac{y}{\sqrt{D}}\mathbf{1}_D, \frac{1}{2}\mathbf{I}_D\right), \tag{34}$$

where $D$ is the dimensionality, $\mathbf{I}_D$ is the identity matrix, $\mathbf{0}_D$ and $\mathbf{1}_D$ are $D$-vectors containing only zeros or ones respectively, and $y$ is a scalar value that represents the radial distance of the observation $\mathbf{y} = -\frac{y}{\sqrt{D}}\mathbf{1}_D$ to the origin. We are interested in the estimation of $I = \int_{\mathcal{X}} f(\mathbf{x})\bar{\pi}(\mathbf{x})d\mathbf{x}$.

Thus this problem consists in computing the posterior predictive density, under the above model, at the point $\frac{y}{\sqrt{D}}\mathbf{1}_D$. In this toy example, we can sample directly from the tempered distributions,

$$\bar{\varphi}(\mathbf{x}|\beta) = \mathcal{N}\left(\mathbf{x}\middle|\frac{2\beta-1}{2\beta+2}\frac{y}{\sqrt{D}}\mathbf{1}_D, \frac{1}{2\beta+2}\mathbf{I}_D\right), \quad \beta \in [0,1]. \tag{35}$$

We compare our proposed approach against two extreme cases of self-normalized IS (SNIS) estimators $\widehat{I}_{\text{SNIS}} = \frac{1}{\sum_{j=1}^{M_{\text{tot}}} w_j}\sum_{i=1}^{M_{\text{tot}}} w_i f(\mathbf{x}_i)$, where $\mathbf{x}_i \sim q(\mathbf{x})$ and $w_i = \frac{\pi(\mathbf{x}_i)}{q(\mathbf{x}_i)}$ is the IS weight. Namely, **(a)** SNIS using samples from the posterior (SNIS-1), i.e., $q(\mathbf{x}) = \bar{\pi}(\mathbf{x})$ (hence, SNIS-1 coincides with standard Monte Carlo), where

$$\bar{\pi}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}\middle| -\frac{1}{2}\frac{y}{\sqrt{D}}\mathbf{1}_D, \frac{1}{2}\mathbf{I}_D\right), \tag{36}$$

and **(b)** SNIS using samples from $q(\mathbf{x}) = \varphi(\mathbf{x}|1) \propto f(\mathbf{x})\pi(\mathbf{x})$ (SNIS-2), which corresponds to setting $\beta = 1$ in Eq. (35). These choices are optimal for estimating, respectively, the denominator and the numerator of the right hand side of Eq. (2) [1]. The ground-truth is known, and can be written as Gaussian density evaluated in $\frac{y}{\sqrt{D}}\mathbf{1}_D$, more specifically, $I = \mathcal{N}\left(\frac{y}{\sqrt{D}}\mathbf{1}_D\middle| -\frac{1}{2}\frac{y}{\sqrt{D}}\mathbf{1}_D, \mathbf{I}_D\right)$. We test the values $y \in \{2, 3.5, 5\}$ and $D \in \{10, 25, 50\}$. Note that, as we increase $y$, the posterior $\bar{\pi}(\mathbf{x})$ and the density $\bar{\varphi}(\mathbf{x}|1) \propto f(\mathbf{x})\pi(\mathbf{x})$ become further apart. We use the powered fraction schedule: $\beta_i = \left(\frac{i-1}{N-1}\right)^5$ for $i = 1, \ldots, N$, with $N = 100$ quadrature points [10, 15]. The results are given in Figure 3, which show, for each pair $(y, D)$, the median relative square error along with the 25% and 75% quantiles (over 100 simulations) versus the number of total samples $M_{\text{tot}}$. In order to keep the comparison fair, we only draw $M = \frac{M_{\text{tot}}}{N}$ samples from each tempered distribution, where $M_{\text{tot}}$ denotes the total number of samples. We see that GTI with $N = 100$ outperforms SNIS-1 and SNIS-2 for all $y$ and $D$ considered. The performance gain is higher with larger $y$, as expected, since this represents a larger mismatch between $\bar{\pi}(\mathbf{x})$ and $\bar{\varphi}(\mathbf{x}|1) \propto f(\mathbf{x})\pi(\mathbf{x})$, that is a scenario where GTI is well suited. Furthermore, the performance of GTI seems rather insensitive to increasing the dimension $D$.

## 5.2 Second numerical analysis

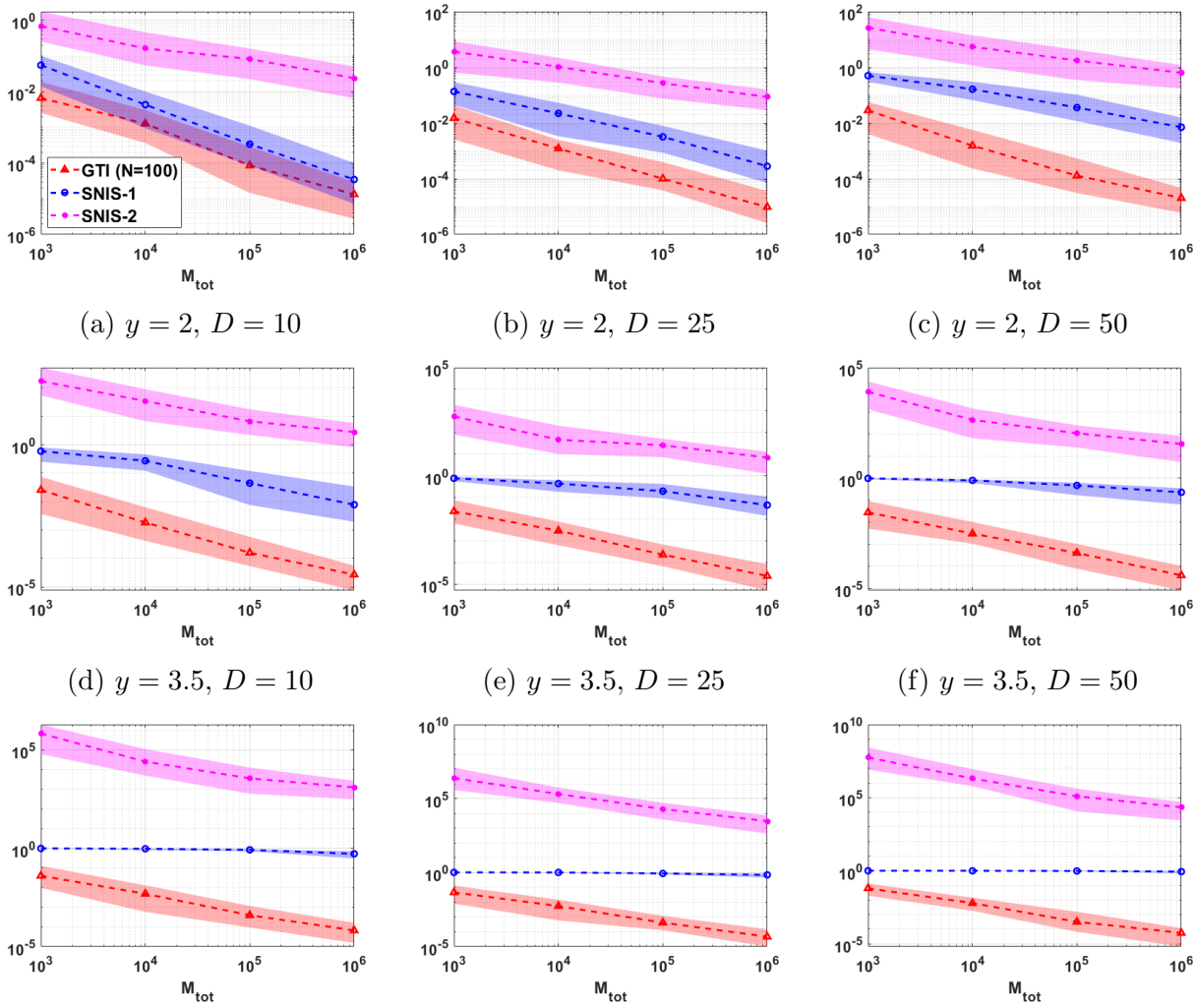
We consider the following two-dimensional banana-shaped density (which is a benchmark example [3, 23, 24]),

$$\pi(x_1, x_2) = \exp\left(-\frac{1}{2}\left(0.03x_1^2 + \left(\frac{x_2}{2} + 0.03\left(x_1^2 - 100\right)\right)^2\right)\right), \tag{37}$$

and the function

$$f(x_1, x_2) = (x_2 + 10)\exp\left(-\frac{1}{4}\left(x_1 + x_2 + 25\right)^2\right)\mathbb{1}\left(x_2 > -10\right). \tag{38}$$

We compare GTI using $N \in \{10, 50, 100\}$ against a single MCMC chain targeting $\pi(\mathbf{x})$ in the estimation of $\mathbb{E}_{\bar{\pi}}[f(\mathbf{x})]$. We use again the powered fraction schedule: $\beta_i = \left(\frac{i-1}{N-1}\right)^5$ for $i = 1, \ldots, N$.




(a) $y = 2$, $D = 10$ (b) $y = 2$, $D = 25$ (c) $y = 2$, $D = 50$

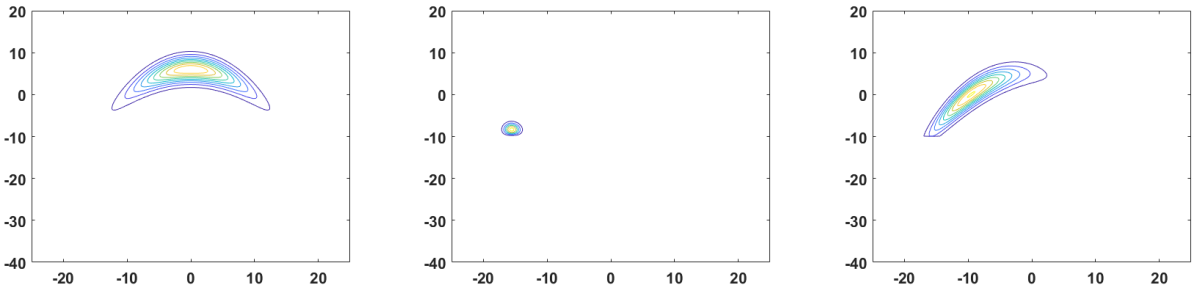
(d) $y = 3.5$, $D = 10$ (e) $y = 3.5$, $D = 25$ (f) $y = 3.5$, $D = 50$

(g) $y = 5$, $D = 10$ (h) $y = 5$, $D = 25$ (i) $y = 5$, $D = 50$

Figure 3: Relative squared error of GTI, IS and RIS as a function of number of samples $M_{\text{tot}}$, for different $y$ and $D$. The TI method uses $N = 100$ quadrature points and $M = \frac{M_{\text{tot}}}{N}$ samples from each tempered distribution for estimating $\mathbb{E}_{\bar{\varphi}(\mathbf{x}|\beta)}[\log f(\mathbf{x})]$ (hence the total number of samples is $M_{\text{tot}}$).

For GTI, we run $N$ chains for $M = \frac{M_{\text{tot}}}{N}$ iterations, each one addressing a different tempered distribution $f(\mathbf{x})^{\beta_i}\pi(\mathbf{x})$. All the MCMC algorithms use a Gaussian random-walk proposal with covariance $\Sigma = 3\mathbf{I}_2$. The budget of evaluations is $M_{\text{tot}} \in \{10^5, 10^6\}$, for all the compared schemes. The results are shown in Table 3. We show the median relative square error of the methods over 100 independent simulations. For the sample sizes considered, GTI performs better than simple MCMC. Indeed, for $N = 100$, the performance gains are almost of one order of magnitude over MCMC. However, note that, for $M_{\text{tot}} = 10^6$, GTI with the choice $N = 10$ is worse than MCMC due to the discretization error, i.e., there is not enough quadrature nodes, so the estimation in Eq. (18) has considerable bias. In that situation, increasing the sample size from $M_{\text{tot}} = 10^5$

to $M_{\text{tot}} = 10^6$ does not translate into a significant performance gain (as opposed to $N = 50, 100$ where the performance increases by one order of magnitude).



(a) $\pi(\mathbf{x})$ (b) $f(\mathbf{x}) \cdot \pi(\mathbf{x})$ (c) $f(\mathbf{x})^{\beta} \cdot \pi(\mathbf{x})$

Figure 4: Plots of $\pi(\mathbf{x})$, $f(\mathbf{x})\pi(\mathbf{x})$ and $f(\mathbf{x})^{\beta}\pi(\mathbf{x})$ with $\beta = 0.0173$ for the banana example. We see that $f(\mathbf{x})$ and $f(\mathbf{x})\pi(\mathbf{x})$ have little overlap and hence a direct MCMC estimate of $\mathbb{E}_{\bar{\pi}}[f(\mathbf{x})]$ is not efficient. The tempered distributions, $f(\mathbf{x})^{\beta}\pi(\mathbf{x})$, are in-between those distributions, helping in the estimation of $\mathbb{E}_{\bar{\pi}}[f(\mathbf{x})]$.

Table 3: Median relative square error of GTI for $M_{\text{tot}} \in \{10^5, 10^6\}$ and $N \in \{10, 50, 100\}$. For each $M_{\text{tot}}$, best results are boldfaced.

| $M_{\text{tot}}$ | MCMC | GTI, $N = 10$ | $N = 50$ | $N = 100$ |
|---|---|---|---|---|
| $10^5$ | 0.042849 | 0.038422 | 0.016544 | **0.00641** |
| $10^6$ | 0.0040054 | 0.01516 | 0.0012224 | **0.00060778** |

# 6 Conclusions

We have extended the powerful thermodynamic integration (TI) technique for performing a complete Bayesian inference. GTI allows the computation of posterior expectations of real-valued functions $f(\mathbf{x})$, and also vector-valued functions $\mathbf{f}(\mathbf{x})$. GTI contains the standard TI as special case. Even for the estimation of the marginal likelihood, this work provides a way for extending the application of the standard TI avoiding the assumption of strictly positive likelihood functions (see Remarks 1- 2). Several computational considerations and variants are discussed. The advantages of GTI are clearly shown in different numerical comparisons. As a future research line, we plan to study new continuous paths for linking densities with different support, avoiding the need of the correction terms. Alternatively, as discussed in Sect. 4, another approach would be to design suitable approximations of $\varphi_+(\mathbf{x})$, $\varphi_-(\mathbf{x})$ and $\pi(\mathbf{x})$ (see Sect. 3.1) using, e.g., regression techniques [25, 26].

# Acknowledgments

LM and FL acknowledges support by the Agencia Estatal de Investigación AEI (project SPGRAPH, ref. num. PID2019- 105032GB-I00) and FL acknowledges support by Spanish government via grant FPU19/00815.

# References

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.

[2] D. Luengo, L. Martino, M. Bugallo, V. Elvira, and S. S., "A survey of monte carlo methods for parameter estimation," *EURASIP J. Adv. Signal Process.*, vol. 25, pp. 1–62, 2020.

[3] T. Rainforth, A. Goliński, F. Wood, and S. Zaidi, "Target–aware Bayesian inference: how to beat optimal conventional estimators," *Journal of Machine Learning Research*, vol. 21, no. 88, 2020.

[4] M. A. Newton and A. E. Raftery, "Approximate Bayesian inference with the weighted likelihood bootstrap," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 56, no. 1, pp. 3–26, 1994.

[5] F. Llorente, L. Martino, D. Delgado, and J. Lopez-Santiago, "Marginal likelihood computation for model selection and hypothesis testing: an extensive review," *arXiv:2005.08334*, 2020.

[6] F. Llorente, L. Martino, D. Delgado, and J. López-Santiago, "On the computation of marginal likelihood via MCMC for model selection and hypothesis testing," in *28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 2373–2377.

[7] N. Friel and J. Wyse, "Estimating the evidence-a review," *Statistica Neerlandica*, vol. 66, no. 3, pp. 288–308, 2012.

[8] R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, no. 2, pp. 125–139, 2001.

[9] N. Lartillot and H. Philippe, "Computing Bayes factors using thermodynamic integration," *Systematic biology*, vol. 55, no. 2, pp. 195–207, 2006.

[10] N. Friel and A. N. Pettitt, "Marginal likelihood estimation via power posteriors," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 3, pp. 589–607, 2008.

[11] R. M. Neal, "Sampling from multimodal distributions using tempered transitions," *Statistics and computing*, vol. 6, no. 4, pp. 353–366, 1996.

[12] L. Martino, F. Llorente, E. Cuberlo, J. López-Santiago, and J. Míguez, "Automatic tempered posterior distributions for Bayesian inversion problems," *Mathematics*, vol. 9, no. 7, p. 784, 2021.

[13] P. D. Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[14] A. Gelman and X. L. Meng, "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling," *Statistical science*, pp. 163–185, 1998.

[15] W. Xie, P. O. Lewis, Y. Fan, L. Kuo, and M. H. Chen, "Improving marginal likelihood estimation for Bayesian phylogenetic model selection," *Systematic biology*, vol. 60, no. 2, pp. 150–160, 2010.

[16] D. Frenkel, "Free-energy computation and first-order phase transitions," 1986.

[17] L. Martino, J. Read, V. Elvira, and F. Louzada, "Cooperative parallel particle filters for online model selection and applications to urban mobility," *Digital Signal Processing*, vol. 60, pp. 172–185, 2017.

[18] R. M. Neal, *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, ON, Canada, 1993.

[19] B. Calderhead and M. Girolami, "Estimating Bayes factors via thermodynamic integration and population MCMC," *Computational Statistics & Data Analysis*, vol. 53, no. 12, pp. 4028–4045, 2009.

[20] C. J. Oates, T. Papamarkou, and M. Girolami, "The controlled thermodynamic integral for Bayesian model evidence evaluation," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 634–645, 2016.

[21] N. Friel, M. Hurn, and J. Wyse, "Improving power posterior estimation of statistical evidence," *Statistics and Computing*, vol. 24, no. 5, pp. 709–723, 2014.

[22] L. Martino, V. Elvira, D. Luengo, J. Corander, and F. Louzada, "Orthogonal parallel MCMC methods for sampling and optimization," *Digital Signal Processing*, vol. 58, pp. 64–84, 2016.

[23] L. Martino and J. Read, "On the flexibility of the design of multiple try Metropolis schemes," *Computational Statistics*, vol. 28, no. 6, pp. 2797–2823, 2013.

[24] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, 2001.

[25] F. Llorente, L. Martino, D. Delgado-Gómez, and G. Camps-Valls, "Deep importance sampling based on regression for model inversion and emulation," *Digital Signal Processing*, vol. 116, p. 103104, 2021.

[26] F. Llorente, L. Martino, V. Elvira, D. Delgado, and J. López-Santiago, "Adaptive quadrature schemes for Bayesian inference via active learning," *IEEE Access*, vol. 8, pp. 208 462–208 483, 2020.