Seppo Pulkkinen | Marko M. Mäkelä | Napsu Karmitsa

A Continuation Approach to Global Minimization of Gaussian RBF Models

Turku Centre *for* Computer Science

TUCS Technical Report
No 998, February 2011

A Continuation Approach to Global Minimization of Gaussian RBF Models

Seppo Pulkkinen
University of Turku, Department of Mathematics
FI-20014 Turku, Finland
seppo.pulkkinen@utu.fi
Marko M. Mäkelä
University of Turku, Department of Mathematics
FI-20014 Turku, Finland
makela@utu.fi
Napsu Karmitsa
University of Turku, Department of Mathematics
FI-20014 Turku, Finland
napsu@karmitsa.fi

TUCS Technical Report
No 998, February 2011

Abstract

During the last decade, a lot of research has been devoted to a new class of derivative-free optimization methods using radial basis function (RBF) models. Methods of this type usually involve finding a (global) minimizer of the model function. However, the development of practical methods for solving this difficult minimization problem has received very little attention in the literature. In this paper, a new method for global minimization of the Gaussian RBF model is presented. The proposed method is based on a homotopy continuation approach. In particular, it is shown that the special structure of the Gaussian RBF model allows a natural way of using the Gaussian transform as a homotopy mapping. This integral transformation effectively removes local minima and preserves the underlying structure of the original RBF model. For tracing the solution curve of the resulting differential equation, a robust trust region-based predictor-corrector method is described. Numerical results are given to demonstrate the reliability of the proposed method.

Keywords: global optimization; derivative-free optimization; radial basis function; Gaussian transform; continuation method; homotopy; trust region; predictor-corrector method

TUCS Laboratory
TUCS Laboratory

1 Introduction

Recently, there has been growing interest in using *radial basis function* (RBF) models both in global optimization [7, 19] and in trust region-based local optimization [15, 22]. Rather than directly minimizing the objective function, which may be noisy or computationally expensive, these *model-based* methods construct a model function that is easier to minimize. Given a set of distinct points $\mathbf{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\} \subset \mathbb{R}^d$, the model function $m : \mathbb{R}^d \to \mathbb{R}$ is constructed by interpolating the objective function $f : \mathbb{R}^d \to \mathbb{R}$ subject to the conditions

$$m(\mathbf{p}_i) = f(\mathbf{p}_i), \quad \mathbf{p}_i \in \mathbf{P}, \quad i = 1, \ldots, n. \tag{1}$$

The RBF models considered in this paper are of the form

$$m(\mathbf{x}) = \sum_{i=1}^{n} w_i \phi_i(\|\mathbf{x} - \mathbf{p}_i\|), \tag{2}$$

where $w_i \in \mathbb{R}$ are *weighting coefficients* and $\phi_i : \mathbb{R} \to \mathbb{R}$ are radial basis functions. Imposing the interpolation conditions (1) on the interpolant (2) leads to an $n \times n$ system of linear equations for the weighting coefficients w_i. That is,

$$\mathbf{\Phi w} = \mathbf{F}, \tag{3}$$

where $\Phi_{ij} = \phi_j(\|\mathbf{p}_i - \mathbf{p}_j\|)$, $\mathbf{w} = [w_1, \ldots, w_n]^T$ and $\mathbf{F} = [f(\mathbf{p}_1), \ldots, f(\mathbf{p}_n)]^T$. The literature on RBF interpolation is extensive (see e.g. [3, 20]). This type of interpolation has also been used in numerous practical applications such as neural networks [16], computer graphics [4], medical imaging [5] and solving partial differential equations [6].

Differently to RBF interpolation, for which the mathematical theory is well-established [3, 20], the theoretical framework of optimization methods using RBF interpolation models is still deficient in many areas. Also, despite an apparent need, a remarkably small amount of research has been carried out concerning minimization of RBF models. For instance, solving the trust region subproblem involving a RBF model is identified as one of the major challenges in [22]. The problem of finding the global minimum of a RBF model is also discussed in [7] and [19]. To our knowledge, no problem-specific approach for the global solution of this nonconvex and multimodal minimization problem has been proposed. To fill this gap in the literature, we propose a new method specifically designed for the global minimization of the *Gaussian RBF model*

$$m(\mathbf{x}) = \sum_{i=1}^{n} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\gamma_i^2}\right) \tag{4}$$

induced by the Gaussian RBF

$$\phi_i(r) = \exp\left(-\frac{r^2}{\gamma_i^2}\right), \tag{5}$$

where $\gamma_i > 0$, $i = 1, \ldots, n$, are user-supplied shape parameters.

The proposed method is based on the idea of smoothing the Gaussian RBF model via the *Gaussian transform*. This integral transformation tends to remove local minima and preserve the structure of the original function. Previously, the Gaussian transform has been applied, for instance, to molecular distance geometry problems [12, 13]. Even though this integral transformation has very limited applicability to general objective functions, we show here that it has an analytic expression when applied to the Gaussian RBF model interpolating the objective function. This fact eliminates the need for computationally expensive numerical integration. Our approach can be viewed as a generalization of the method proposed by Kostrowicki et al. [9] for molecular conformation problems. In their approach, the Fourier-Poisson integral transformation is applied to Gaussian functions approximating the Lennard-Jones potential.

Adapting the ideas by Kostrowicki et al. [17] and Wu [23], we give a differential equation formulation for the transformation between the minimizers of the smoothed Gaussian RBF model and the original one. The idea of this *continuation* approach is to seek for a parametrized curve that connects the global minimizer of the smoothed RBF model, which is easier to obtain, to the global minimizer of the original RBF model. While the formulations proposed in [17] and [23] have been successfully applied to objective functions appearing in molecular distance geometry problems [9, 13], our aim is to show their applicability to the more general Gaussian RBF model. A distinguishing feature of our formulation is that we establish the conditions for convexity of the smoothed RBF model. This fundamental result allows a unique solution to our differential equation. For solving the associated initial value problem, we adapt the trust region-based method proposed by Moré and Wu [13]. Exploiting the special structure of the Gaussian RBF model, we also show that the modified conjugate gradient method described in [21] is particularly efficient for solving the trust region subproblem.

This paper is organized as follows. In Section 2, we define the Gaussian transform and derive a closed-form expression for the Gaussian transform of the Gaussian RBF model. In Section 3, we formulate the transformation between the minimizers of the smoothed RBF model and the original one as a solution to a differential equation. The choice of the initial values is discussed in Section 4. Section 5 is devoted for describing a predictor-corrector method for solving the resulting initial value problem. To demonstrate the reliability of our method, numerical results are presented in Section 6. Finally, Section 7 summarizes this paper and points out some directions of future research.

2 The Gaussian Transform

The main result of this section is that the *Gaussian transform* of the Gaussian RBF model has an analytic expression. The Gaussian transform has turned out to be a viable tool for global optimization because of its smoothing and structure-

preserving properties [23]. In many cases, it has been shown to remove local minima while preserving the global one [10, 11, 12, 13]. Another advantage of this transformation is that it provides a parameter for controlling the degree of smoothing.

We begin our analysis with the definition of the Gaussian transform.

Definition 2.1. *The* Gaussian transform *of a function* $f : \mathbb{R}^d \to \mathbb{R}$ *is*

$$\langle f \rangle_\sigma(\mathbf{x}) = C_\sigma \int_{\mathbb{R}^d} f(\mathbf{y}) \exp\left(-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{\sigma^2}\right) d\mathbf{y}, \tag{6}$$

where $\sigma > 0$ *is a smoothing parameter and*

$$C_\sigma = \left(\frac{1}{\sqrt{\pi}\sigma}\right)^d.$$

This integral transformation can be viewed as a distance-weighted average of f with respect to the Gaussian weighting function. The width of the smoothing kernel, and thus the degree of smoothing is proportional to σ. On the other hand, as σ approaches zero, we obtain the original function. Another way of viewing the smoothing properties of this transformation is to consider its relation to the Fourier transform. For instance, it is shown in [23] that the Gaussian transform reduces the high frequency components of the Fourier transform of the original function.

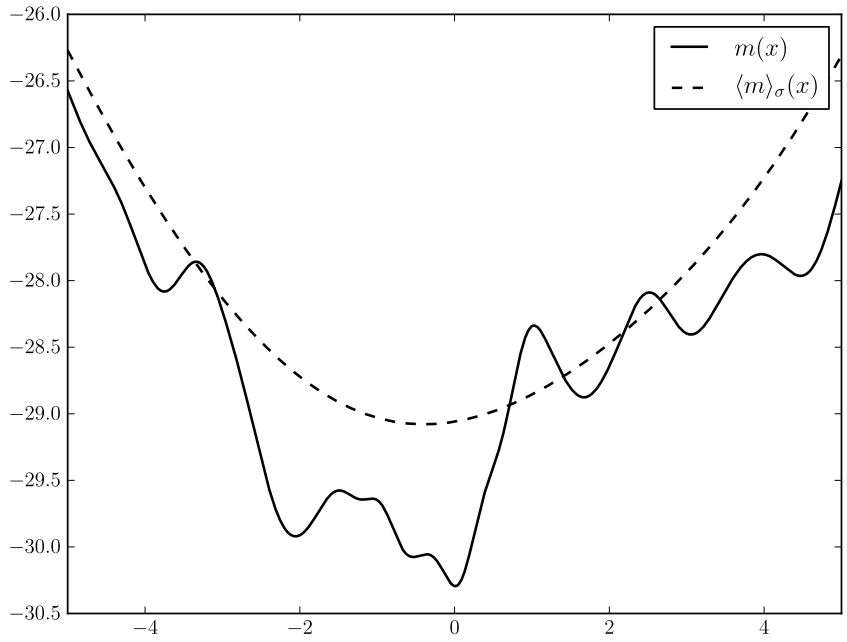

Figure 1: A Gaussian RBF model and its convex Gaussian transform with $\sigma = 3$.

The Gaussian transform $\langle m \rangle_\sigma$ of the Gaussian RBF model (4) is illustrated in Figure 1. In this case, the transformation produces a strictly convex function that also gives a good approximation of the global minimizer of the original RBF model. The convexity of the transformed RBF model is not a coincidence. Namely, in Section 4 we establish the conditions for convexity of $\langle m \rangle_\sigma$ if σ is sufficiently large.

Next, we show that the transformed Gaussian RBF model $\langle m \rangle_\sigma$ has a closed-form expression. It is derived by using the well-known formula for the univariate Gaussian integral [1].

Lemma 2.1. *Let $a, b \in \mathbb{R}$. Then*

$$\int_{-\infty}^{\infty} \exp\left(-\frac{(x-a)^2}{b^2}\right) dx = b\sqrt{\pi}.$$

The following lemma facilitates the proof of our main result.

Lemma 2.2. *Let*

$$\psi(y) = \exp\left(-\frac{(y-z)^2}{\gamma^2} - \frac{(y-x)^2}{\sigma^2}\right),$$

where $x, y, z, \gamma, \sigma \in \mathbb{R}$. Then

$$\int_{-\infty}^{\infty} \psi(y) dy = \frac{\gamma\sqrt{\pi}\sigma}{\sqrt{\sigma^2+\gamma^2}} \exp\left(-\frac{(x-z)^2}{\sigma^2+\gamma^2}\right).$$

Proof. See Appendix A. □

Theorem 2.1. *The Gaussian transform of the Gaussian RBF model* (4) *is*

$$\langle m \rangle_\sigma(\mathbf{x}) = \sum_{i=1}^{n} \tilde{C}_{\sigma,i} w_i \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right), \tag{7}$$

where $\sigma > 0$, $\gamma_i > 0$ and

$$\tilde{C}_{\sigma,i} = \left(\frac{\gamma_i}{\sqrt{\sigma^2+\gamma_i^2}}\right)^d, \quad i = 1, \ldots, n.$$

Proof. Let $\sigma > 0$ and $\varphi_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\gamma_i^2}\right)$, $i = 1, \ldots, n$. Then

$$\begin{aligned}
\langle \varphi_i \rangle_\sigma(\mathbf{x}) &= C_\sigma \int_{\mathbb{R}^d} \exp\left(-\frac{\|\mathbf{y}-\mathbf{p}_i\|^2}{\gamma_i^2}\right) \exp\left(-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{\sigma^2}\right) d\mathbf{y} \\
&= C_\sigma \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{d} \exp\left(-\frac{(y_j-p_{i,j})^2}{\gamma_i^2} - \frac{(y_j-x_j)^2}{\sigma^2}\right) dy_j.
\end{aligned}$$

The above expression is a product of one-dimensional integrals. Thus, by virtue of Lemma 2.2 we have

$$\begin{aligned}\langle\varphi_i\rangle_\sigma(\mathbf{x}) &= C_\sigma \prod_{j=1}^{d} \frac{\gamma_i\sqrt{\pi}\sigma}{\sqrt{\sigma^2+\gamma_i^2}} \exp\left(-\frac{(x_j-p_{i,j})^2}{\sigma^2+\gamma_i^2}\right) \\ &= \left(\frac{\gamma_i}{\sqrt{\sigma^2+\gamma_i^2}}\right)^d \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right), \quad i=1,\ldots,n,\end{aligned}$$

from which equation (7) follows, since $\langle m\rangle_\sigma(\mathbf{x}) = \sum_{i=1}^{n} w_i\langle\varphi_i\rangle_\sigma(\mathbf{x})$. □

Remark 2.1. *We can extend the definition of* (7) *by defining* $\langle m\rangle_0 = m$.

3 Differential Equation Formulation

The basic idea of our method is to obtain a smoothed Gaussian RBF model via the Gaussian transform. The smoothed RBF model is gradually deformed back to the original one by letting the transformation parameter σ approach zero. The minimizers of the RBF model are traced along this deformation process by successively applying local minimization procedures to the transformed RBF model. This "transportation" of minimizers, as illustrated in Figure 2, is likely to carry the solution over undesired local minima as they are removed by the Gaussian transform. Of particular interest is that the initial minimizer is uniquely determined due to the strict convexity of the smoothed RBF model with the initial transformation parameter σ.

In the following, we give a differential equation formulation for the curve of minimizers illustrated in Figure 2. Our formulation is adapted from Wu [23] who established the existence and uniqueness of such a solution curve for general objective functions smoothed via the Gaussian transform. The formulations described here and in [23] are in fact special cases of the more general homotopy continuation approach (see e.g. [2]).

To begin our analysis, we note that the transformation defined by equation (7) induces a $\mathcal{C}^\infty$-homotopy mapping $h : \mathbb{R}^d \times [0,\infty[\to \mathbb{R}$ defined as

$$\begin{aligned} &h(\mathbf{x},\sigma) = \langle m\rangle_\sigma(\mathbf{x}), \\ &h(\mathbf{x},0) = m(\mathbf{x}). \end{aligned} \tag{8}$$

In order to formulate our differential equation, we note that the conditions

$$\begin{aligned} &\nabla_{\mathbf{x}} h(\mathbf{x}(\sigma),\sigma) = \mathbf{0}, \\ &\nabla^2_{\mathbf{x}} h(\mathbf{x}(\sigma),\sigma) \quad \text{is positive definite} \end{aligned} \qquad \forall\sigma\in[0,\sigma_0], \quad \sigma_0 > 0, \tag{9}$$

implicitly define the parametric curve $\mathbf{x} : [0,\sigma_0] \to \mathbb{R}^d$ passing through the minimizers of $h(\cdot,\sigma)$. By differentiating the condition $\nabla_{\mathbf{x}} h(\mathbf{x}(\sigma),\sigma) = \mathbf{0}$ with respect

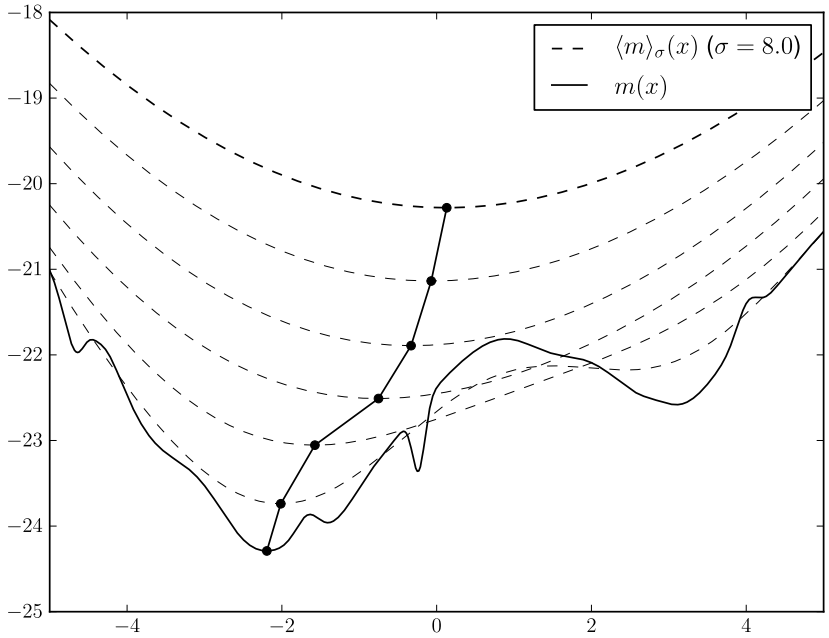

Figure 2: A curve connecting the minimizers of the smoothed Gaussian RBF model $\langle m \rangle_\sigma$ with different values of transformation parameter σ.

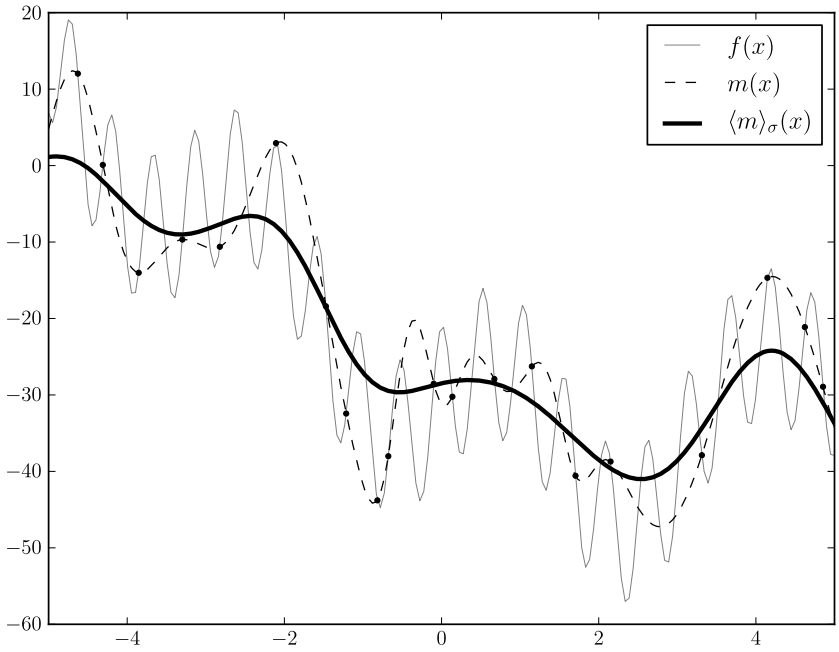

Figure 3: A noisy objective function f, an oscillating Gaussian RBF interpolation model m and the smoothed RBF model $\langle m \rangle_\sigma$.

to σ, we obtain a differential equation for $\mathbf{x}(\sigma)$. This leads us to an initial value problem

$$\begin{aligned} &\nabla^2_{\mathbf{x}} h(\mathbf{x}(\sigma),\sigma)\mathbf{x}'(\sigma) + \frac{\partial}{\partial\sigma}\nabla_{\mathbf{x}} h(\mathbf{x}(\sigma),\sigma) = 0 \quad \forall\sigma\in[0,\sigma_0], \\ &\mathbf{x}(\sigma_0) = \mathbf{x}_0 \end{aligned} \tag{10}$$

for some $\mathbf{x}_0 \in \mathbb{R}^d$ and $\sigma_0 > 0$ satisfying

$$\begin{aligned} &\nabla_{\mathbf{x}} h(\mathbf{x}_0,\sigma_0) = \mathbf{0}, \\ &\nabla^2_{\mathbf{x}} h(\mathbf{x}_0,\sigma_0) \quad \text{is positive definite.} \end{aligned} \tag{11}$$

Assuming that conditions (9) are satisfied, the existence and uniqueness of the solution of (10) with the given initial values $\mathbf{x}_0$ and σ_0 follows from the results proven in [23]. In Section 4, we prove a stronger result which ensures that the initial minimizer $\mathbf{x}_0$ can also be uniquely determined due to the strict convexity of the smoothed RBF model with any sufficiently large σ_0. It then follows from definition (8) that following the solution curve from $\mathbf{x}(\sigma_0)$ to $\mathbf{x}(0)$ gives a uniquely defined minimizer of the original RBF model. Though this minimizer is not guaranteed to be the global one, due to the smoothing properties of the Gaussian transform, this approach is more likely to give the global minimizer than a local search started from an arbitrary point.

The minimizers obtained with intermediate values of σ are also of interest. For instance, interpolation of noisy objective functions leads to oscillating interpolants. In such cases, it makes sense to stop tracing the solution curve at some $\sigma > 0$. As shown in Figure 3, the smoothed RBF model is less affected by noise and follows the underlying trend of the interpolated function.

For the following sections, we give the expressions of the derivatives of the homotopy mapping h induced by the Gaussian transform (7). A straightforward calculation yields

$$\nabla_{\mathbf{x}} h(\mathbf{x},\sigma) = -2\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2+\gamma_i^2} w_i E_{\sigma,i}(\mathbf{r}_i)\mathbf{r}_i, \tag{12}$$

$$\nabla^2_{\mathbf{x}} h(\mathbf{x},\sigma) = 2\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2+\gamma_i^2} w_i \left(\frac{2}{\sigma^2+\gamma_i^2}\mathbf{r}_i\mathbf{r}_i^T - \mathbf{I}\right) E_{\sigma,i}(\mathbf{r}_i), \tag{13}$$

$$\frac{\partial}{\partial\sigma}\nabla_{\mathbf{x}} h(\mathbf{x},\sigma) = 2\sigma\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{(\sigma^2+\gamma_i^2)^2} w_i \left(d+2-\frac{2\|\mathbf{r}_i\|^2}{\sigma^2+\gamma_i^2}\right) E_{\sigma,i}(\mathbf{r}_i)\mathbf{r}_i, \tag{14}$$

where $\mathbf{r}_i = \mathbf{x} - \mathbf{p}_i$, $E_{\sigma,i}(\mathbf{r}) = \exp\left(-\frac{\|\mathbf{r}\|^2}{\sigma^2+\gamma_i^2}\right)$ and $\mathbf{I}$ is the $d \times d$ identity matrix.

4 Choice of Initial Values

This section deals with the choice of the starting point $\mathbf{x}_0$ and the initial homotopy parameter σ_0 for the initial value problem (10). The key issues are to guarantee that conditions (11) are satisfied and that $\mathbf{x}_0$ is uniquely determined. In order to achieve this, we prove that the homotopy mapping $h(\cdot, \sigma)$ is strictly convex for any sufficiently large homotopy parameter σ in a given sphere containing the interpolation points. We also prove that the set of minimizers of $h(\cdot, \sigma)$ converges to a single point as σ approaches infinity.

4.1 Choice of σ_0

In the following, we show that choosing a sufficiently large initial homotopy parameter σ_0 guarantees that $h(\cdot, \sigma_0)$ is strictly convex in a given sphere

$$B(\mathbf{z}; r) = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{z}\| < r\}, \quad \mathbf{z} \in \mathbb{R}^d, r > 0 \tag{15}$$

containing the interpolation points $\mathbf{p}_i$, $i = 1, \ldots, n$. The proof of our main result necessitates the following assumption.

Assumption 4.1. *The weighting coefficients w_i and the shape parameters γ_i of the Gaussian RBF model m satisfy*

$$\sum_{i=1}^{n} w_i \gamma_i^d < 0. \tag{16}$$

This assumption is not very restrictive. Namely, the following theorem gives a formula for adjusting the weighting coefficients $\mathbf{w}$ solved from equation (3) to satisfy condition (16). This theorem also shows that the adjustment is equivalent to adding a constant to the interpolated function values $\mathbf{F}$. A linear translation of the interpolated function values does not essentially alter our interpolation problem.

Theorem 4.1. *Assume that $\mathbf{\Phi} \in \mathbb{R}^{n \times n}$ is a nonsingular matrix. Let $\mathbf{w} = \mathbf{\Phi}^{-1}\mathbf{F}$ for some $\mathbf{F} \in \mathbb{R}^n$ and let $\gamma_i > 0$, $i = 1, \ldots, n$. Define*

$$\tilde{\mathbf{w}} = \mathbf{w} + \mathbf{\Phi}^{-1}\mathbf{C}, \quad \textit{where } \mathbf{C} = [c, \ldots, c]^T \in \mathbb{R}^n \tag{17}$$

such that

$$c = \frac{\mu - \sum_{i=1}^{n} w_i \gamma_i^d}{\sum_{i=1}^{n} \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1}}, \quad \sum_{i=1}^{n} \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1} \neq 0, \tag{18}$$

for some $\mu \in \mathbb{R}$. Then $\mathbf{\Phi}\tilde{\mathbf{w}} = \mathbf{F} + \mathbf{C}$ and $\sum_{i=1}^{n} \tilde{w}_i \gamma_i^d = \mu$.

Proof. Due to the assumption that $\mathbf{w} = \mathbf{\Phi}^{-1}\mathbf{F}$, we have

$$\tilde{\mathbf{w}} = \mathbf{w} + \mathbf{\Phi}^{-1}\mathbf{C} = \mathbf{\Phi}^{-1}\mathbf{F} + \mathbf{\Phi}^{-1}\mathbf{C} = \mathbf{\Phi}^{-1}(\mathbf{F} + \mathbf{C}),$$

which is equivalent to $\mathbf{\Phi}\tilde{\mathbf{w}} = \mathbf{F} + \mathbf{C}$. By multiplying $\tilde{w}_i$ by γ_i^d we obtain

$$\tilde{w}_i\gamma_i^d = w_i\gamma_i^d + \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1} c, \quad i = 1, \ldots, n.$$

Summation over $i = 1, \ldots, n$ and substitution of equation (18) into the above expression yields

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{w}_i\gamma_i^d &= \sum_{i=1}^{n} w_i\gamma_i^d + \sum_{i=1}^{n} \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1} c \\
&= \sum_{i=1}^{n} w_i\gamma_i^d + \sum_{i=1}^{n} \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1} \frac{\mu - \sum_{i=1}^{n} w_i\gamma_i^d}{\sum_{i=1}^{n} \gamma_i^d \sum_{j=1}^{n} \Phi_{ij}^{-1}} \\
&= \sum_{i=1}^{n} w_i\gamma_i^d + \mu - \sum_{i=1}^{n} w_i\gamma_i^d = \mu.
\end{aligned}$$

□

Next, we formally state the assumption that $B(\mathbf{z}; r)$ contains the interpolation points and give the following two lemmata to deduce our estimate for the smallest eigenvalue of $\nabla_{\mathbf{x}}^2 h(\cdot, \sigma)$ in $B(\mathbf{z}; r)$ for a given $\sigma \geq 0$.

Assumption 4.2. *The center point $\mathbf{z} \in \mathbb{R}^d$ and radius $r > 0$ of the sphere $B(\mathbf{z}; r)$ defined by* (15) *are chosen such that $\mathbf{p}_i \in B(\mathbf{z}; r)$ for all $i = 1, \ldots, n$.*

Lemma 4.1. *Assume 4.2 and define $\mathbf{H}_1 : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$,*

$$\mathbf{H}_1(\mathbf{x}, \sigma) = -\sum_{i=1}^{n} \kappa_i(\sigma) \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right) \mathbf{I}, \tag{19}$$

where $\kappa_i : [0, \infty[\to \mathbb{R}$, $i = 1, \ldots, n$, and $\mathbf{I}$ is the $d \times d$ identity matrix. Then

$$\lambda_{\substack{min \\ \mathbf{x}\in B(\mathbf{z};r)}} (\mathbf{H}_1(\mathbf{x}, \sigma)) \geq -\sum_{\kappa_i(\sigma)>0} \kappa_i(\sigma) - \sum_{\kappa_i(\sigma)<0} \kappa_i(\sigma) \exp\left(-\frac{(\|\mathbf{z} - \mathbf{p}_i\| + r)^2}{\sigma^2 + \gamma_i^2}\right).$$

Proof. See Appendix A. □

Lemma 4.2. *Assume 4.2 and define $\mathbf{H}_2 : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$,*

$$\mathbf{H}_2(\mathbf{x}, \sigma) = \sum_{i=1}^{n} \theta_i(\sigma)(\mathbf{x} - \mathbf{p}_i)(\mathbf{x} - \mathbf{p}_i)^T \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right),$$

where $\theta_i : [0, \infty[\to \mathbb{R}$*,* $i = 1, \ldots, n$*. Then*

$$\lambda_{\min}_{\mathbf{x} \in B(\mathbf{z};r)} (\mathbf{H}_2(\mathbf{x}, \sigma)) \geq \sum_{\theta_i(\sigma)<0} \theta_i(\sigma) \hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2 + \gamma_i^2}\right),$$

where

$$\hat{r}_i(\sigma) = \min\{\|\mathbf{z} - \mathbf{p}_i\| + r, \sqrt{\sigma^2 + \gamma_i^2}\}.$$

Proof. See Appendix A. □

The following theorem is utilized in the proofs of Lemmata 4.1-4.2 and Theorem 4.3. For the proof of this result, see [8].

Theorem 4.2 (Weyl)**.** *Let* $\mathbf{A} \in \mathbb{R}^{d \times d}$ *and* $\mathbf{B} \in \mathbb{R}^{d \times d}$ *be symmetric matrices. Then*

$$\lambda_{min}(\mathbf{A}) + \lambda_{min}(\mathbf{B}) \leq \lambda_{min}(\mathbf{A} + \mathbf{B}),$$

where $\lambda_{min}(\cdot)$ *denotes the smallest eigenvalue of a matrix.*

Remark 4.1. *If either* $\mathbf{A}$ *or* $\mathbf{B}$ *is the identity matrix, equality holds above.*

With Lemmata 4.1-4.2 and Theorem 4.2, we can now prove the following theorem that provides a lower bound for the smallest eigenvalue of the Hessian (13) in $B(\mathbf{z}; r)$.

Theorem 4.3. *Assume 4.1 and 4.2. Then*

$$\lambda_{\min}_{\mathbf{x} \in B(\mathbf{z};r)} (\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)) \geq 2\left[\Lambda_1(\sigma) + \Lambda_2(\sigma)\right] \tag{20}$$

for all $\sigma \geq 0$*, where*

$$\Lambda_1(\sigma) = -\sum_{w_i>0} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i - \sum_{w_i<0} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{(\|\mathbf{z} - \mathbf{p}_i\| + r)^2}{\sigma^2 + \gamma_i^2}\right),$$

$$\Lambda_2(\sigma) = \sum_{w_i<0} \frac{2\tilde{C}_{\sigma,i}}{(\sigma^2 + \gamma_i^2)^2} w_i \hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2 + \gamma_i^2}\right)$$

and

$$\hat{r}_i(\sigma) = \min\{\|\mathbf{z} - \mathbf{p}_i\| + r, \sqrt{\sigma^2 + \gamma_i^2}\}.$$

Proof. The Hessian matrix defined by equation (13) can be written as

$$\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma) = 2[\mathbf{H}_1(\mathbf{x}, \sigma) + \mathbf{H}_2(\mathbf{x}, \sigma)],$$

where

$$\mathbf{H}_1(\mathbf{x},\sigma) = -\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2+\gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma^2}\right)\mathbf{I},$$

$$\mathbf{H}_2(\mathbf{x},\sigma) = \sum_{i=1}^{n} \frac{2\tilde{C}_{\sigma,i}}{(\sigma^2+\gamma_i^2)^2} w_i (\mathbf{x}-\mathbf{p}_i)(\mathbf{x}-\mathbf{p}_i)^T \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma^2}\right).$$

By virtue of Theorem 4.2, we have

$$\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\nabla_{\mathbf{x}}^2 h(\mathbf{x},\sigma)) \geq 2\left[\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\nabla_{\mathbf{x}}^2 \mathbf{H}_1(\mathbf{x},\sigma)) + \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\nabla_{\mathbf{x}}^2 \mathbf{H}_2(\mathbf{x},\sigma))\right].$$

Define

$$\kappa_i(\sigma) = \frac{\tilde{C}_{\sigma,i}}{\sigma^2+\gamma_i^2} w_i, \quad \theta_i(\sigma) = \frac{2\tilde{C}_{\sigma,i}}{(\sigma^2+\gamma_i^2)^2} w_i, \quad i = 1,\ldots,n.$$

With these definitions, the claim follows from Lemmata 4.1 and 4.2 applied to $\mathbf{H}_1$ and $\mathbf{H}_2$, since $\tilde{C}_{\sigma,i} > 0$ for all $i = 1,\ldots,n$ and $\sigma \geq 0$. □

Next, we show that under assumptions 4.1 and 4.2, inequality (20) gives a positive estimate for the smallest eigenvalue of $\nabla_{\mathbf{x}}^2 h(\cdot,\sigma)$ in $B(\mathbf{z};r)$ if σ is sufficiently large.

Theorem 4.4. *Assume 4.1 and 4.2. Then there exists $\sigma^* > 0$ such that*

$$\Lambda_1(\sigma) + \Lambda_2(\sigma) > 0$$

for all $\sigma > \sigma^*$.

Proof. Define

$$\tilde{\Lambda}_1(\sigma) = -D_{max}(\sigma)\sum_{w_i>0} w_i\gamma_i^d - D_{min}(\sigma)\sum_{w_i<0} w_i\gamma_i^d \exp\left(-\frac{(\|\mathbf{z}-\mathbf{p}_i\|+r)^2}{\sigma^2+\gamma_i^2}\right),$$

$$\tilde{\Lambda}_2(\sigma) = 2D_{max}(\sigma)\sum_{w_i<0} \frac{w_i\gamma_i^d}{\sigma^2+\gamma_i^2}\hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2+\gamma_i^2}\right),$$

where

$$D_{max}(\sigma) = \max_{i=1,\ldots,n} \frac{1}{(\sigma^2+\gamma_i^2)^{\frac{d}{2}+1}} \quad \text{and} \quad D_{min}(\sigma) = \min_{i=1,\ldots,n} \frac{1}{(\sigma^2+\gamma_i^2)^{\frac{d}{2}+1}}.$$

Since $\gamma_i > 0$ for all $i = 1,\ldots,n$, $D_{max}(\sigma) > 0$ and $D_{min}(\sigma) > 0$, we have

$$\Lambda_1(\sigma) \geq \tilde{\Lambda}_1(\sigma) \quad \text{and} \quad -\Lambda_2(\sigma) \leq -\tilde{\Lambda}_2(\sigma) \tag{21}$$

for all $\sigma \geq 0$. By dividing the inequality $\tilde{\Lambda}_1(\sigma) > -\tilde{\Lambda}_2(\sigma)$ by $D_{max}(\sigma)$, we obtain

$$\begin{aligned} &- \sum_{w_i>0} w_i \gamma_i^d - \frac{D_{min}(\sigma)}{D_{max}(\sigma)} \sum_{w_i<0} w_i \gamma_i^d \exp\left(-\frac{(\|\mathbf{z}-\mathbf{p}_i\|+r)^2}{\sigma^2+\gamma_i^2}\right) \\ > &- \sum_{w_i<0} \frac{2w_i\gamma_i^d}{\sigma^2+\gamma_i^2} \hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2+\gamma_i^2}\right). \end{aligned}$$

By noting that $\lim_{\sigma\to\infty} D_{min}(\sigma)/D_{max}(\sigma) = 1$ and using assumption (16), we have

$$\begin{aligned} \lim_{\sigma\to\infty} \frac{\tilde{\Lambda}_1(\sigma)}{D_{max}(\sigma)} &= \lim_{\sigma\to\infty} \left[- \sum_{w_i>0} w_i \gamma_i^d - \frac{D_{min}(\sigma)}{D_{max}(\sigma)} \sum_{w_i<0} w_i \gamma_i^d \exp\left(-\frac{(\|\mathbf{z}-\mathbf{p}_i\|+r)^2}{\sigma^2+\gamma_i^2}\right)\right] \\ &= -\sum_{i=1}^{n} w_i \gamma_i^d > 0. \end{aligned}$$

On the other hand, we have

$$\lim_{\sigma\to\infty} \frac{\tilde{\Lambda}_2(\sigma)}{D_{max}(\sigma)} = \lim_{\sigma\to\infty} \left[- \sum_{w_i<0} \frac{2w_i\gamma_i^d}{\sigma^2+\gamma_i^2} \hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2+\gamma_i^2}\right)\right] = 0,$$

which implies that there exists $\sigma^* > 0$ such that $\tilde{\Lambda}_1(\sigma) > -\tilde{\Lambda}_2(\sigma)$ for all $\sigma > \sigma^*$ and thus, by virtue of equation (21),

$$\Lambda_1(\sigma) \geq \tilde{\Lambda}_1(\sigma) > -\tilde{\Lambda}_2(\sigma) \geq -\Lambda_2(\sigma)$$

for all $\sigma > \sigma^*$. □

Theorem 4.4 leads to the following conclusion.

Corollary 4.1. *Let the assumptions of Theorem 4.3 be satisfied. Then there exists $\sigma^* > 0$ such that $\lambda_{min}(\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)) > 0$ for all $\sigma > \sigma^*$ and $\mathbf{x} \in B(\mathbf{z}; r)$. Consequently, $h(\cdot, \sigma)$ is strictly convex in $B(\mathbf{z}; r)$ for all $\sigma > \sigma^*$.*

Remark 4.2. *We cannot obtain strict convexity of $h(\cdot, \sigma)$ in the whole $\mathbb{R}^d$ for any finite value of σ. This follows from the fact that the second derivative of the Gaussian function* (5) *necessarily changes its sign for some $r > 0$.*

4.2 Choice of $\mathbf{x}_0$

By virtue of Theorem 4.4, it is possible to choose σ_0 such that $h(\cdot, \sigma_0)$ is strictly convex in a sphere $B(\mathbf{z}; r)$ containing the interpolation points. Consequently, $h(\cdot, \sigma_0)$ has at most one minimizer that can be chosen as the starting point $\mathbf{x}_0$ in $B(\mathbf{z}; r)$. Though we cannot directly obtain $\mathbf{x}_0$ satisfying conditions (11) with the given σ_0, the limiting stationary point $\mathbf{x}^*$ provided by the following theorem serves the purpose of an initial guess for $\mathbf{x}_0$. If σ_0 is large, $\mathbf{x}^*$ is expected to be close to the minimizer of $h(\cdot, \sigma_0)$.

Theorem 4.5. *Define*

$$S_\sigma = \{\mathbf{x} \in \mathbb{R}^d \mid \nabla_{\mathbf{x}} h(\mathbf{x}, \sigma) = 0\}.$$

Then we have $\lim_{\sigma \to \infty} S_\sigma = \{\mathbf{x}^*\}$*, where*

$$\mathbf{x}^* = \frac{\sum_{i=1}^{n} w_i \gamma_i^d \mathbf{p}_i}{\sum_{i=1}^{n} w_i \gamma_i^d}. \tag{22}$$

Proof. See Appendix A. □

4.3 Summary of Obtaining Initial Values

To summarize the above results, our aim is to apply Theorems 4.4 and 4.5 to obtain the initial homotopy parameter σ_0 and the initial guess $\mathbf{x}^*$ for the starting point $\mathbf{x}_0$ for the solution of (10). For the choice of $B(\mathbf{z}; r)$, it makes sense to choose $\mathbf{z}$ and r such that $B(\mathbf{z}; r)$ encloses the set of interpolation points and also contains $\mathbf{x}^*$. Our choice is to set $\mathbf{z} = \mathbf{x}^*$ and choose r as

$$r = \max\{\|\mathbf{x}^* - \mathbf{p}_i\| \mid i = 1, \ldots, n\}. \tag{23}$$

Then σ_0 is chosen such that positive definiteness of $\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma_0)$ for all $\mathbf{x} \in B(\mathbf{x}^*; r)$ is guaranteed by inequality (20). As a consequence, $h(\cdot, \sigma_0)$ is strictly convex in $B(\mathbf{x}^*; r)$. Then the starting point $\mathbf{x}_0$ satisfying the initial conditions (11) can be uniquely determined by using the initial guess $\mathbf{x}^*$ and the iterative corrector method described in the next section. Assuming that σ_0 is sufficiently large, $\mathbf{x}_0$ obtained in this way is in $B(\mathbf{x}^*; r)$ by Theorem 4.5. If σ_0 is not large enough, it is increased as long as necessary.

To summarize the above discussion, we provide Algorithm 1 that outlines the steps for obtaining the initial values $\mathbf{x}_0$ and σ_0.

Algorithm 1: Initial Values.

Obtain $\mathbf{x}^*$ from equation (22).
$r \leftarrow \max\{\|\mathbf{x}^* - \mathbf{p}_i\| \mid i = 1, \ldots, n\}$
Choose the initial $\sigma_0 > 0$.
while $\Lambda_1(\sigma) + \Lambda_2(\sigma) \leq 0$ **do** Increase σ_0.
repeat
  Obtain $\mathbf{x}_0 \in \mathbb{R}^d$ such that $\nabla_{\mathbf{x}} h(\mathbf{x}_0, \sigma_0) = \mathbf{0}$, use $\mathbf{x}^*$ as starting point.
  if $\mathbf{x}_0 \notin B(\mathbf{x}^*, r)$ **then** Increase σ_0.
until $\mathbf{x}_0 \in B(\mathbf{x}^*, r)$

5 Predictor-Corrector Method

This section is devoted to describing an iterative method for tracing the solution curve of the initial value problem (10). We begin our analysis by noting that pre-multiplying equation (10) by the inverse of $\nabla^2_{\mathbf{x}} h$ yields the tangent of the solution curve. That is,

$$\mathbf{x}'(\sigma) = -\nabla^2_{\mathbf{x}} h(\mathbf{x}(\sigma), \sigma)^{-1} \frac{\partial}{\partial \sigma} \nabla_{\mathbf{x}} h(\mathbf{x}(\sigma), \sigma). \tag{24}$$

This form suggests using an iterative *predictor-corrector* method (see e.g. [18]) that is also our preferred choice. We recall that this has been the dominant approach for solving initial value problems of the form (10) arising in continuation methods (see e.g. [2]). However, in contrast to the existing predictor-corrector methods, we propose a trust region-based corrector method.

5.1 Overview

Each iteration of a predictor-corrector method involves two steps, as illustrated in Figure 4. The first step is to obtain a crude estimate of the the solution curve by taking a *predictor step*. For solving the initial value problem (10), a step along the tangent of the solution curve $\mathbf{x}(\sigma)$ at the current iterate $\mathbf{x}_k = \mathbf{x}(\sigma_k)$ is given by

$$\tilde{\mathbf{x}}_k(\tau) = \mathbf{x}_k - \tau \mathbf{T}(\mathbf{x}_k, \sigma_k), \tag{25}$$

where

$$\mathbf{T}(\mathbf{x}, \sigma) = -\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)^{-1} \frac{\partial}{\partial \sigma} \nabla_{\mathbf{x}} h(\mathbf{x}, \sigma) \tag{26}$$

denotes the tangent vector defined by equation (24) and $\tau > 0$ denotes the step size. The negative sign is chosen to express that we are tracing the solution curve along decreasing values of σ.

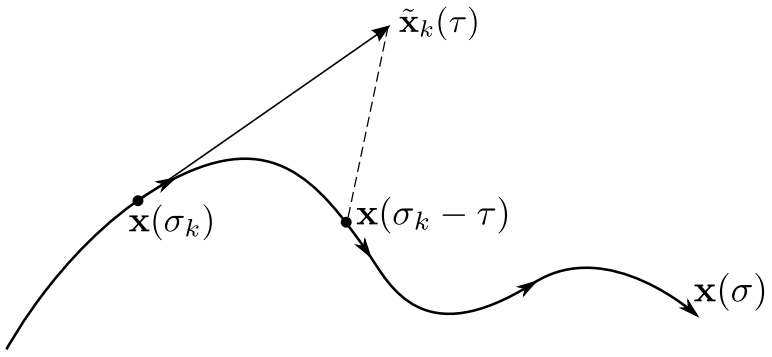

Figure 4: Iteration step of a predictor-corrector method.

To return the estimate obtained from the predictor step back to the solution curve, a *corrector iteration* is started from the predictor estimate. The corrector

iteration seeks for a point $\hat{\mathbf{x}}_k$ such that

$$\nabla_{\mathbf{x}} h(\hat{\mathbf{x}}_k, \sigma_k - \tau) = 0. \tag{27}$$

After finishing the corrector iteration, the next main iteration step starts with $\mathbf{x}_{k+1} = \hat{\mathbf{x}}_k$ and $\sigma_{k+1} = \sigma_k - \tau$. The iteration terminates if $\sigma_{k+1} < \sigma_{min}$, where $\sigma_{min} \geq 0$ is a user-supplied parameter. If $\sigma_{min} = 0$, a final corrector step is then applied to $h(\cdot, 0)$ to obtain a minimizer of the original RBF model m.

In order to determine the step size τ, we control the distance between the predictor step $\tilde{\mathbf{x}}_k(\tau)$ and the current iterate $\mathbf{x}_k$ via the user-supplied accuracy parameter Δ_p. We also need to restrict the ratio between σ_{k+1} and σ_k in order to avoid too large step sizes. Hence, we impose the requirements

$$\|\tilde{\mathbf{x}}_k(\tau) - \mathbf{x}_k\| = \Delta_p \quad \text{and} \quad \sigma_k - \tau \geq \frac{1}{4}\sigma_k$$

that are satisfied if the step size τ is chosen according to

$$\tau = \min\{\frac{\Delta_p}{\|\mathbf{T}(\mathbf{x}_k, \sigma_k)\|}, \frac{3\sigma_k}{4}\}.$$

To summarize the above discussion, a predictor-corrector algorithm for solving the initial value problem (10) is outlined as Algorithm 2.

Algorithm 2: Predictor-Corrector Method.

Choose $\mathbf{x}_0 \in \mathbb{R}^d$ and $\sigma_0 > 0$ satisfying (11) by using Algorithm 1.
while $\sigma_k > \sigma_{min}$ **do**
  $\tau \leftarrow \min\{\frac{\Delta_p}{\|\mathbf{T}(\mathbf{x}_k, \sigma_k)\|}, \frac{3\sigma_k}{4}\}$
  $\tilde{\mathbf{x}}_k(\tau) \leftarrow \mathbf{x}_k - \tau \mathbf{T}(\mathbf{x}_k, \sigma_k)$
  Solve $\nabla_{\mathbf{x}} h(\hat{\mathbf{x}}_k, \sigma_k - \tau) = \mathbf{0}$ for $\hat{\mathbf{x}}_k$, use $\tilde{\mathbf{x}}_k(\tau)$ as starting point.
  $\mathbf{x}_{k+1} \leftarrow \hat{\mathbf{x}}_k$
  $\sigma_{k+1} \leftarrow \sigma_k - \tau$
  $k \leftarrow k + 1$
if $\sigma_{min} = 0$ **then**
  Solve $\nabla_{\mathbf{x}} h(\hat{\mathbf{x}}_k, 0) = \mathbf{0}$ for $\hat{\mathbf{x}}_k$, use $\mathbf{x}_k$ as starting point.

5.2 Trust Region-Based Corrector Method

Since the step size τ is usually small and the predictor estimate is close to the solution curve, a Newton-type method with rapid local convergence is a natural choice for the corrector iteration [2]. In order to guarantee convergence, we propose a trust region approach. Despite the fact that this approach has proven out to be very robust for Newton-type methods (see e.g. [14]), using it in predictor-corrector methods is amazingly rare. To our knowledge, a similar approach has only been taken by Moré and Wu [13] who used a trust-region based Newton

method to successively minimize the transformed objective function $\langle f \rangle_{\sigma_k}$ with a predetermined sequence of transformation parameters σ_k.

Our trust region strategy for the corrector iteration is adapted from Newton-based optimization methods (see e.g. [14]). At each corrector iteration step j the quadratic model

$$\mathbf{Q}_j(\mathbf{s}) = h(\hat{\mathbf{x}}^j, \sigma_k - \tau) + \nabla_{\mathbf{x}} h(\hat{\mathbf{x}}^j, \sigma_k - \tau)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \nabla^2_{\mathbf{x}} h(\hat{\mathbf{x}}^j, \sigma_k - \tau) \mathbf{s} \quad (28)$$

is used to approximate $h(\cdot, \sigma_k - \tau)$ around the current corrector iterate denoted by $\hat{\mathbf{x}}^j$. This model is minimized in a trust region in which it can be considered reliable. That is, the trust region subproblem

$$\min_{\mathbf{s}} \mathbf{Q}_j(\mathbf{s}) \quad \text{s.t. } \|\mathbf{s}\| \leq \Delta_j \quad (29)$$

is solved, where Δ_j denotes the current trust region radius (the initial radius Δ_0 is set to Δ_p). After this step, the ratio between the actual reduction of $h(\cdot, \sigma_k - \tau)$ and the predicted reduction from the quadratic model defined as

$$\rho = \frac{h(\hat{\mathbf{x}}^j, \sigma_k - \tau) - h(\hat{\mathbf{x}}^j + \mathbf{s}, \sigma_k - \tau)}{\mathbf{Q}_j(\mathbf{0}) - \mathbf{Q}_j(\mathbf{s})}, \quad (30)$$

where $\mathbf{Q}_j(\mathbf{0}) - \mathbf{Q}_j(\mathbf{s}) \neq 0$, is computed. Then the trust region radius is adjusted according to (see [14])

$$\Delta_{j+1} = \begin{cases} \frac{1}{4}\Delta_j, & \text{if } \rho < \frac{1}{4} \\ \min\{2\Delta_j, \Delta_{max}\}, & \text{if } \rho > \frac{3}{4} \text{ and } \|\mathbf{s}\| = \Delta_j \\ \Delta_j, & \text{otherwise,} \end{cases}$$

where $\Delta_{max} > 0$ is a user-supplied upper bound for the trust region radius. The next iterate $\hat{\mathbf{x}}^{j+1}$ is then chosen as $\hat{\mathbf{x}}^j + \mathbf{s}$ if $\rho > \frac{1}{10}$, which ensures that the next iterate is accepted only if a sufficient reduction is obtained.

In order to ensure that the distance between the final corrector step $\hat{\mathbf{x}}_k$ and the current main iterate $\mathbf{x}_k$ is proportional to the predictor step norm Δ_p, we require that

$$\|\hat{\mathbf{x}}_k - \mathbf{x}_k\| < \alpha \Delta_p, \quad (31)$$

where $\alpha > 1$. If

$$\|\nabla_{\mathbf{x}} h(\hat{\mathbf{x}}^j, \sigma_k - \tau)\| < \epsilon_c, \quad (32)$$

where ϵ_c is a user supplied threshold parameter, and

$$\|\hat{\mathbf{x}}^j - \mathbf{x}_k\| < \alpha \Delta_p \quad (33)$$

for some corrector iteration step j, the iteration is terminated. Otherwise, if the gradient condition (32) is satisfied but condition (33) is not, the step size τ is set

to a smaller value $\tilde{\tau} = \beta\tau$, where $\beta \in]0, 1[$, and the next corrector iterate $\hat{\mathbf{x}}^{j+1}$ is chosen as $\hat{\mathbf{x}}^j$. Then the corrector iteration continues by minimizing $h(\cdot, \sigma_k - \tilde{\tau})$. By the continuity of h, these additional steps ensure that conditions (31) and (32) are both satisfied when the corrector iteration terminates.

The steps of the corrector iteration are summarized in Algorithm 3.

Algorithm 3: Corrector Iteration.

$\Delta_0 \leftarrow \Delta_p$
for $j = 0, 1, \ldots$ **do**
  Obtain $\mathbf{s}$ by solving problem (29).
  if $\|\nabla h(\hat{\mathbf{x}}^j, \sigma_k - \tau)\| < \epsilon_c$ **then**
    if $\|\hat{\mathbf{x}}^j + \mathbf{s} - \mathbf{x}_k\| \geq \alpha\Delta_p$ **then**
      $\tau \leftarrow \beta_1\tau$
      $\hat{\mathbf{x}}^{j+1} \leftarrow \hat{\mathbf{x}}^j$
      Continue to next iteration.
    else
      Terminate, return $\hat{\mathbf{x}}^{j+1}$.
  $\rho \leftarrow \dfrac{h(\hat{\mathbf{x}}^j, \sigma_k - \tau) - h(\hat{\mathbf{x}}^j + \mathbf{s}, \sigma_k - \tau)}{\mathbf{Q}_j(\mathbf{0}) - \mathbf{Q}_j(\mathbf{s})}$
  if $\rho < \frac{1}{4}$ **then**
    $\Delta_{j+1} \leftarrow \frac{1}{4}\Delta_j$
  else if $\rho > \frac{3}{4}$ *and* $\|\mathbf{s}\| = \Delta_j$ **then**
    $\Delta_{j+1} \leftarrow \min\{2\Delta_j, \Delta_{max}\}$
  else
    $\Delta_{j+1} \leftarrow \Delta_j$
  if $\rho > \frac{1}{10}$ **then**
    $\hat{\mathbf{x}}^{j+1} \leftarrow \hat{\mathbf{x}}^j + \mathbf{s}$
  else
    $\hat{\mathbf{x}}^{j+1} \leftarrow \hat{\mathbf{x}}^j$

5.3 Solving the Trust Region Subproblem

For solving the trust region subproblem (29), we have adapted the truncated conjugate gradient method proposed by Steihaug [21]. As the standard conjugate gradient method applied to this problem, it solves the linear system

$$\nabla^2_{\mathbf{x}} h(\hat{\mathbf{x}}^j, \sigma_k - \tau)\mathbf{s} = -\nabla_{\mathbf{x}} h(\hat{\mathbf{x}}^j, \sigma_k - \tau)$$

obtained by equating the gradient of $\mathbf{Q}_j$ to zero. The Steihaug method, however, handles the boundary constraint and nonpositive eigenvalues of $\nabla^2_{\mathbf{x}} h$, which may occur if the predictor step is far from the solution curve.

The computationally most expensive steps of the conjugate gradient method are the matrix-vector multiplications involving the matrix $\nabla^2_{\mathbf{x}} h$ defined by equa-

tion (13). However, these computations reduce to vector operations since

$$\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)\mathbf{v} = 2\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)\left[\frac{2(\mathbf{x} - \mathbf{p}_i)^T\mathbf{v}}{\sigma^2 + \gamma_i^2}(\mathbf{x} - \mathbf{p}_i) - \mathbf{v}\right]$$

for any vector $\mathbf{v} \in \mathbb{R}^d$. This property offers a significant computational advantage, since the Steihaug method produces the solution of a constrained $d \times d$ linear system in at most d iterations [14]. Since the complexity of the evaluation of $\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)\mathbf{s}$ and $\mathbf{s}^T\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)\mathbf{s}$ reduces to $\mathcal{O}(nd)$, the worst-case complexity of the matrix inversion is only $\mathcal{O}(nd^2)$. In fact, this is the same as the complexity of the evaluation of $\nabla^2_{\mathbf{x}} h(\mathbf{x}, \sigma)$. We also note that the Hessian does not need to be explicitly evaluated or stored during the matrix inversion.

5.4 Obtaining the Tangent Vector

Recalling equation (26), obtaining the predictor estimate involves solving the tangent vector $\mathbf{T}(\mathbf{x}_k, \sigma_k)$ from

$$\nabla^2_{\mathbf{x}} h(\mathbf{x}_k, \sigma_k)\mathbf{T}(\mathbf{x}_k, \sigma_k) = -\frac{\partial}{\partial\sigma}\nabla_{\mathbf{x}} h(\mathbf{x}_k, \sigma_k). \tag{34}$$

Based on the above arguments, the conjugate gradient method is a viable choice. However, differently to the trust region subproblem (29), we do not impose the boundary constraint, since the step size is controlled by the strategy described above. Also, the Hessian matrix $\nabla^2_{\mathbf{x}} h(\mathbf{x}_k, \sigma_k)$ is positive definite by assumptions (9), and thus we can use the standard conjugate gradient method.

6 Numerical Results

In this section, we present preliminary numerical results for our algorithm. The aim of our numerical experiments was to investigate its reliability. We tested our algorithm on a large number of Gaussian RBF models with dimension $d = 1$ and with randomly chosen interpolation points p_i. In order to avoid any difficulties introduced by the nonuniform spacing of the interpolation points, the weighting coefficients w_i were chosen randomly instead of solving them from equation (3).

6.1 Success Probabilities With Random RBF Models

To characterize the success rate, we define the probability of success as

$$P_{succ} = \frac{N_{succ}}{N},$$

where N is the total number of runs and N_{succ} is the number of successful runs. In each test, the interpolation points $p_i \in [a, b]$, where $a = -10$ and $b = 10$, were

sampled from a uniform random distribution. For the estimation of the global minimizer, a grid search with $n_g = 501$ evenly spaced points was used. In our tests, we considered a run successful, if

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq \frac{b-a}{2(n_g-1)},$$

where $\mathbf{x}_k$ is the final iterate for some k and $\mathbf{x}^*$ is an estimate of the global minimizer in the range $[-10, 10]$ obtained by the grid search.[1]

If not stated otherwise, we used the following parameters in all our tests:

$$\begin{array}{lclcl} \sigma_{\min} & = & 10^{-4} & \beta & = & 0.5 \\ \Delta_p & = & 0.15 & \epsilon_c & = & 10^{-6} \\ \alpha & = & 1.5 & \Delta_{max} & = & 10^6 \end{array}$$

Tests A: random RBF models, different values of shape parameters.

In the first set of tests, we tested our algorithm on randomly chosen Gaussian RBF models of the form (4) with different choices of the shape parameters γ_i. We chose $n = 50$ and a uniform shape parameter, that is, $\gamma_i = \gamma$ for all $i = 1, \ldots, n$. In effect, the γ-parameter determines the frequency of oscillation of the model function. Small values of γ yield a RBF model with sharp spikes, whereas larger values yield a smoother one. The effect of the γ-parameter on the shape of the RBF model and its Gaussian transform is illustrated in Figure B.1.

For each run, the weighting coefficients w_i were sampled from a uniform random distribution such that $w_i \in [-5, 5]$ for $i = 1, \ldots, n$. In order to satisfy Assumption 4.1, the weighting coefficients w_i were sampled with the condition that $\sum_{i=1}^{n} w_i < -10$. The success probabilities P_{succ} of these tests with $N = 10000$ are plotted in Figure 5. For failed runs, we also computed the average distance to the estimated global minimizer $\mathbf{x}^*$, and the results are plotted in Figure 6.

Clearly, the reliability of our algorithm depends on the frequency of oscillation in the Gaussian RBF model. When γ is 0.25, our algorithm is unable to distinguish between a large number of narrow local minima, and the success probability drops but is still as high as 50%. On the other hand, if γ is larger, our algorithm yields a higher rate of success, since the function has a small number of wide minima. The Gaussian transform is more likely to preserve these wide minima than the narrow ones with smaller values of γ. With $\gamma = 2.5$, a remarkably high 90% success probability can be observed. However, the results in Figure 6 show that for failed runs, the average distance to $\mathbf{x}^*$ is relatively large for all values of γ and increases with γ. This indicates that for most failed runs, our algorithm is not even able to give a good estimate for the global minimizer.

[1] Runs with $\mathbf{x}_k \notin [-10, 10]$ were not considered since the global minimizer was not searched outside this interval.

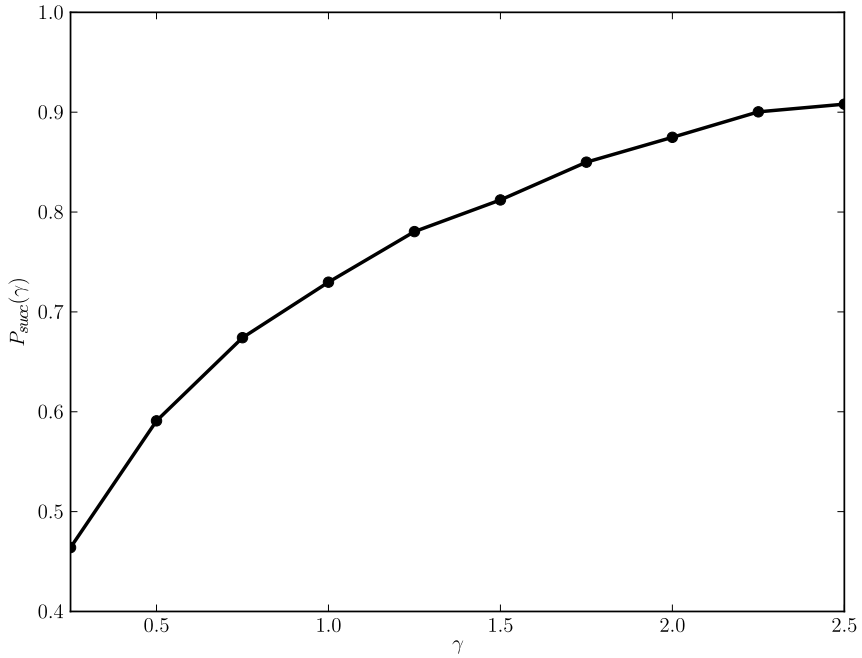

Figure 5: Success probability P_{succ} as a function of shape parameter γ.

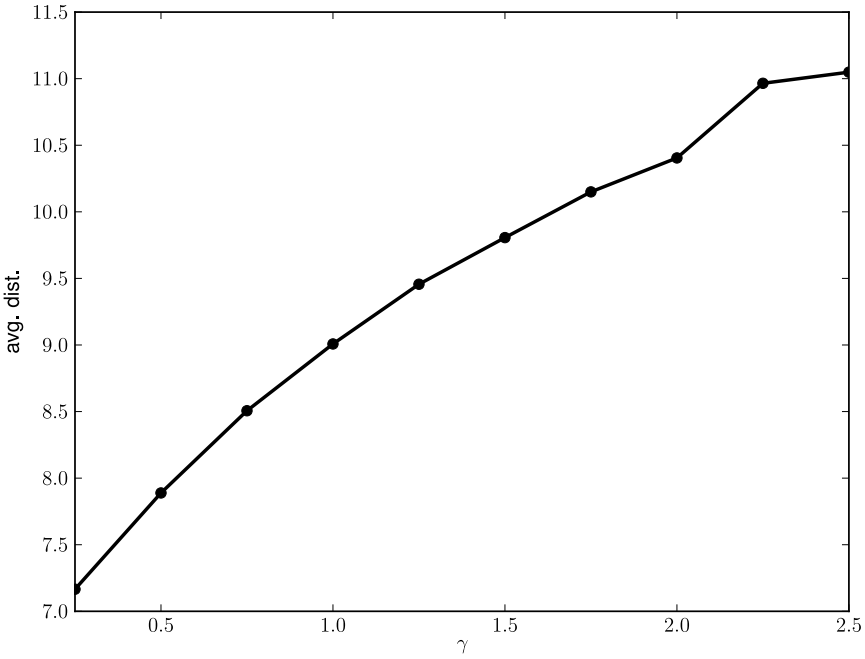

Figure 6: Average distance (of failed runs) to the estimated global minimizer $\mathbf{x}^*$ as a function of shape parameter γ.

We emphasize that the above results reflect the worst-case behaviour of our algorithm. That is because a RBF model with weighting coefficients and interpolation points sampled from a uniform random distribution does not have any underlying structure. This is especially the case with small values of γ that produce highly oscillating RBF models. In this case, our algorithm is more likely to converge to any local minimizer than the global one.

Tests B: Noisy Gaussian RBF models with a dominant convex term.

In practice, many noisy objective functions follow some kind of pattern or trend. Thus, it is of particular interest to test our algorithm on RBF models possibly interpolating functions of this type. For this reason, it would make sense to test our algorithm on Gaussian RBF models having a dominant convex term and small high-frequency noise terms. Under these conditions, the Gaussian transform is expected to smooth out any small local minima and preserve the underlying convex structure. As a result, our algorithm is expected to skip these local minima and converge to the global minimizer with a high probability.

To experimentally verify our hypothesis, we conducted a second set of tests. For each run, we used a Gaussian RBF model of the form (4) with

$$\begin{array}{llll} p_1 = 0, & \gamma_1 = 20, & w_1 = -15, & \\ p_i \in [-10, 10], & \gamma_i = \frac{1}{\omega}, & w_i \in [-\lambda, \lambda], & i = 2, \ldots, n, \end{array}$$

where $\omega > 0$, $\lambda > 0$ and $n = 100$. The shape parameter γ_1 and the weighting coefficient w_1 of the first term are chosen significantly larger than the remaining noise terms with indices $i = 2, \ldots, n$, which determines the shape of the RBF model. In addition, the negativity of w_1 ensures that the first term is a convex function on the interval $[-10, 10]$. The ω-parameter is inversely proportional to the width of the noise terms and thus it determines the frequency of the noise. The λ-parameter, which controls the range of the weighting coefficients of the noise terms, determines the magnitude of the noise. The effect of the noise magnitude λ on the Gaussian RBF model is illustrated in Figure B.2.

We computed the success probabilities P_{succ} as a function of noise magnitude λ with different noise frequencies ω. The results of these tests with $N = 10000$ are plotted in Figure 7. For failed runs, we also computed the average distance to estimated global minimizer $\mathbf{x}^*$. These results are plotted in Figure 8.

It can be seen from Figure 7 that the frequency of the noise has a major impact on the success probability. In the case of high-frequency noise, the success probability drops rapidly as the noise magnitude λ is increased. Namely, if $\omega = 10$, the success probability already drops below 60% with $\lambda = 0.3$. On the other hand, with $\omega = 1$, the success probability remains over 80% even if $\lambda = 1$, which means a very large noise magnitude. This follows from the property that the Gaussian transform is more likely to preserve global minima resulting from low-frequency

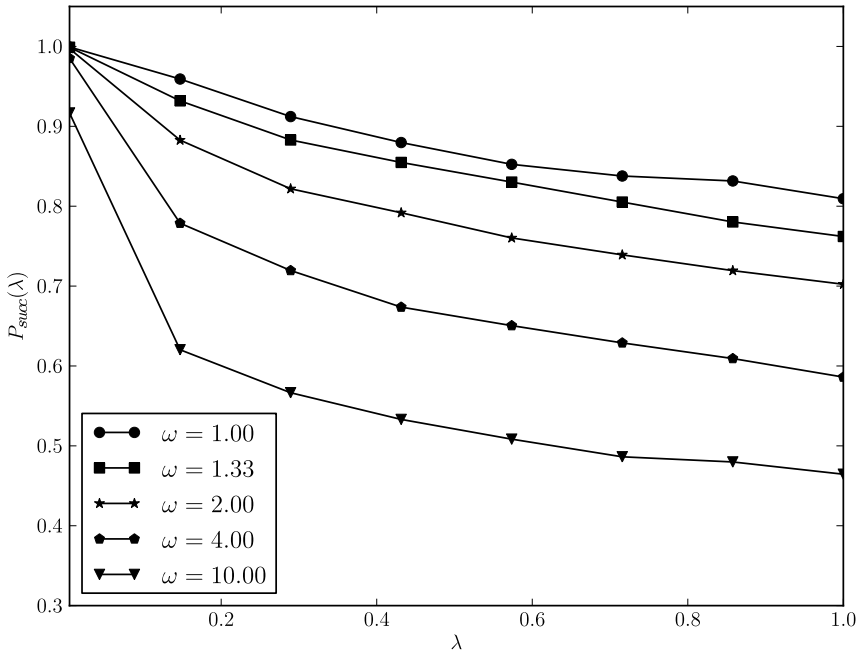

Figure 7: Success probability P_{succ} as a function of noise magnitude λ with different values of noise frequency ω.

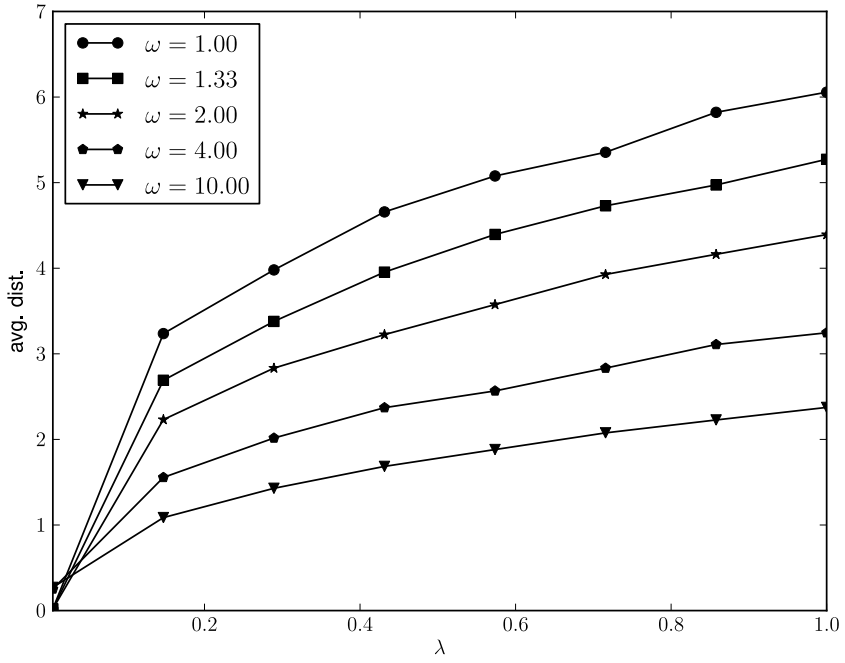

Figure 8: Average distance (of failed runs) to the estimated global minimizer $\mathbf{x}^*$ as a function of noise magnitude λ with different values of noise frequency ω.

noise terms in the RBF model. On the other hand, it tends to smooth out narrow global minima resulting from high-frequency noise terms.

On the other hand, the results in Figure 8 indicate that for failed runs, our algorithm is still able to give good estimates of the global minimizer. This is especially the case with high-frequency noise. With $\omega = 10$, the average distance to the estimated global minimizer $\mathbf{x}^*$ remains below 3, which is relatively small compared to the overall length of the interval $[-10, 10]$ under consideration. This compensates the low success probabilities of accurately locating the global minimizer. Even for smaller values of ω, the average distances remain below 7, which is much better than those observed in Tests A. The increase in the average distance as ω is decreased is mostly explained by the fact that the minimizers of the Gaussian RBF model are usually further apart if the noise frequency ω is smaller, and thus, the shape parameters γ_i are larger. These observations clearly indicate that the Gaussian transform effectively preserves the underlying convex structure of the original RBF model. As a result, the minimizer obtained via our algorithm is in most cases the global one or close to it.

7 Conclusions and Future Research

In this paper, a new method for global minimization of the Gaussian RBF model was proposed. The method is a truly global optimization method in the sense that it is independent of the starting point. Our method uses the Gaussian transform to smooth the RBF model and remove undesired local minima. The transformation has an analytic expression, which eliminates the need for numerical integration. The transformation between the minimizers of the smoothed RBF model and the original one is characterized by a differential equation. Since the smoothed RBF model was shown to be strictly convex under mild conditions, it has a unique minimizer that can be easily identified. Consequently, the solution curve of this differential equation is uniquely determined. Finally, a robust trust region-based predictor-corrector method for tracing the solution curve of this differential equation was developed.

In our preliminary numerical results, our algorithm showed high success rates at identifying global minima of Gaussian RBF models. The primary factors contributing to the reliability were identified to be the underlying structure and the amount of oscillation in the RBF model. In particular, very high success rates were achieved with RBF models having an underlying convex structure and a relatively small amount of noise. What makes our algorithm particularly useful, is the fact that many objective functions appearing in practical applications, and thus their corresponding RBF interpolation models, belong to this category. Also, surprisingly high success rates were achieved with highly oscillating RBF models not following any particular trend.

Though the computational performance of our algorithm in the large-scale

setting was not tested, its high computational complexity is an issue that needs to be addressed. In particular, the inherent complexity of the Gaussian RBF model limits the performance of our present implementation. The Newton-type steps requiring the exact Hessian of the homotopy mapping in the predictor-corrector method are also main sources of computational cost. Therefore, a potential topic of future research is replacing the exact Hessian matrix with a BFGS or SR1 approximation or their limited memory versions (see e.g. [14]). We note that the Broyden update formula, which is closely related to the BFGS and SR1 formulas, has been successfully used by Allgower and Georg [2] in homotopy methods for solving nonlinear equations.

In this paper, we restricted our analysis to unconstrained problems. A constrained version of our method would definitely bring it closer to practical applications. That would require a reformulation of the initial value problem (10) by including constraints. One possible approach could be to formulate it by using the Lagrangian function of the constrained problem.

The fact that the Gaussian transform of the Gaussian RBF model has an analytic expression has important consequences. By virtue of this result, it is possible to indirectly apply the Gaussian transform to the original, possibly noisy, objective function via the interpolating Gaussian RBF model. This approach offers a viable alternative to direct evaluation of the Gaussian transform, which is not computationally tractable for high-dimensional objective functions with no exploitable structure. Thus, we aim to pursue this approach in our future research.

The potential of our method has not yet fully realized, since we have not specified a way of constructing the Gaussian RBF model interpolating the original objective function. In order to develop a complete global optimization method using a RBF interpolation model, this issue needs to be addressed. The choice of the interpolation points $\mathbf{p}_i$ and the shape parameters γ_i is a complicated issue that deserves a study of its own. Therefore, we plan to address this issue in our future research.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1964.

[2] E.L. Allgover and K. Georg. *Numerical Continuation Methods: An Introduction*. Springer-Verlag, 1990.

[3] M.D. Buhmann. Radial basis functions. *Acta Numerica*, (9):1–38, 2000.

[4] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 67–76, 2001.

[5] J.C. Carr, W.R. Fright, and R.K. Beatson. Surface interpolation with radial basis functions for medical imaging. *IEEE Transactions on Medical Imaging*, 16(1):96–107, 1997.

[6] C. Franke and R. Schaback. Solving partial differential equations by collocation using radial basis functions. *Applied Mathematics and Computation*, 93(1):73 – 82, 1998.

[7] H.-M. Gutmann. A radial basis function method for global optimization. *Journal of Global Optimization*, 19(3):201–227, 2001.

[8] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[9] J. Kostrowicki, L. Piela, B.J. Cherayil, and H.A. Scheraga. Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms. *Journal of Physical Chemistry*, 95(10):4113–4119, 1991.

[10] J.J. Moré and Z. Wu. ϵ-optimal solutions to distance geometry problems via global continuation. Preprint MCS-P520-0595, Argonne National Laboratory, Argonne, Illinois, 1995.

[11] J.J. Moré and Z. Wu. Global smoothing and continuation for large-scale molecular optimization. Preprint MCS-P539-1095, Argonne National Laboratory, Argonne, Illinois, 1995.

[12] J.J. Moré and Z. Wu. Smoothing techniques for macromolecular global optimization. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 297–312. Plenum Press, 1996.

[13] J.J. Moré and Z. Wu. Global continuation for distance geometry problems. *SIAM Journal on Optimization*, 7(3):814–836, 1997.

[14] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, second edition, 2006.

[15] R. Oeuvray. *Trust-Region Methods Based on Radial Basis Functions with Application to Biomedical Imaging*. PhD thesis, EFPL, Lausanne, Switzerland, 2005.

[16] J. Park and I.W. Sandberg. Universal approximation using radial-basis-function networks. *Neural Computation*, 3(2):246–257, 1991.

[17] L. Piela, J. Kostrowicki, and H. A. Scheraga. The multiple-minima problem in the conformational analysis of molecules. deformation of the potential energy hypersurface by the diffusion equation method. *Journal of Physical Chemistry*, 93(8):3339–3346, 1989.

[18] A. Ralston and P. Rabinowitz. *A First Course in Numerical Analysis*. Dover Publications, New York, second edition, 1978.

[19] R.G. Regis and C.A. Shoemaker. Constrained global optimization of expensive black box functions using radial basis functions. *Journal of Global Optimization*, 31(1):153–171, 2005.

[20] R. Schaback and H. Wendland. Characterization and construction of radial basis functions. In N. Dyn, D. Leviatan, D. Levin, and A. Pinkus, editors, *Multivariate Approximation and Applications*. Cambridge University Press, Cambridge, 2001.

[21] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, 20(3):626–637, 1983.

[22] S.M. Wild, R.G. Regis, and C.A. Shoemaker. ORBIT: Optimization by radial basis function interpolation in trust-regions. *SIAM Journal on Scientific Computing*, 30(6):3197–3219, 2008.

[23] Z. Wu. The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation. *SIAM Journal on Optimization*, 6(3):748–768, 1996.

A Proofs of Lemmata 2.2, 4.1, 4.2 and Theorem 4.5

Lemma 2.2 Let

$$\psi(y) = \exp\left(-\frac{(y-z)^2}{\gamma^2} - \frac{(y-x)^2}{\sigma^2}\right),$$

where $x, y, z, \gamma, \sigma \in \mathbb{R}$. Then

$$\int_{-\infty}^{\infty} \psi(y)dy = \frac{\gamma\sqrt{\pi}\sigma}{\sqrt{\sigma^2+\gamma^2}} \exp\left(-\frac{(x-z)^2}{\sigma^2+\gamma^2}\right).$$

Proof. By performing the variable substitution $u(y) = \frac{y-z}{\gamma}$, we have

$$\begin{aligned}\int_{-\infty}^{\infty} \psi(y)dy &= \gamma \int_{-\infty}^{\infty} \exp\left(-u^2 - \frac{(x-z-\gamma u)^2}{\sigma^2}\right) du \\ &= \gamma \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2u^2 + (x-z-\gamma u)^2}{\sigma^2}\right) du. \qquad (35)\end{aligned}$$

Some algebraic manipulation of the exponent numerator in equation (35) yields

$$\begin{aligned}&\sigma^2u^2 + (x-z-\gamma u)^2 \\ &= \sigma^2u^2 + (x-z)^2 - 2(x-z)\gamma u + \gamma^2u^2 \\ &= (\sigma^2+\gamma^2)u^2 + (x-z)^2 - 2(x-z)\gamma u + \frac{\gamma^2}{\sigma^2+\gamma^2}(x-z)^2 - \frac{\gamma^2}{\sigma^2+\gamma^2}(x-z)^2 \\ &= \left[\sqrt{\sigma^2+\gamma^2}u - \sqrt{\frac{\gamma^2}{\sigma^2+\gamma^2}}(x-z)\right]^2 + (x-z)^2 - \frac{\gamma^2}{\sigma^2+\gamma^2}(x-z)^2 \\ &= (\sigma^2+\gamma^2)\left[u - \sqrt{\frac{\gamma^2}{(\sigma^2+\gamma^2)^2}}(x-z)\right]^2 + \frac{\sigma^2}{\sigma^2+\gamma^2}(x-z)^2.\end{aligned}$$

Then, by substituting the above expression into equation (35) and by applying Lemma 2.1, we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \psi(y)dy &= \gamma\left[\int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2+\gamma^2}{\sigma^2}\left(u - \sqrt{\frac{\gamma^2}{(\sigma^2+\gamma^2)^2}}(x-z)\right)^2 - \frac{(x-z)^2}{\sigma^2+\gamma^2}\right) du\right] \\ &= \gamma\left[\int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2+\gamma^2}{\sigma^2}\left(u - \sqrt{\frac{\gamma^2}{(\sigma^2+\gamma^2)^2}}(x-z)\right)^2\right) du\right] \exp\left(-\frac{(x-z)^2}{\sigma^2+\gamma^2}\right) \\ &= \frac{\gamma\sqrt{\pi}\sigma}{\sqrt{\sigma^2+\gamma^2}} \exp\left(-\frac{(x-z)^2}{\sigma^2+\gamma^2}\right).\end{aligned}$$

□

Lemma 4.1 Assume 4.2 and define $\mathbf{H}_1 : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$,

$$\mathbf{H}_1(\mathbf{x},\sigma) = -\sum_{i=1}^{n} \kappa_i(\sigma) \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right)\mathbf{I}, \tag{36}$$

where $\kappa_i : [0,\infty[\to \mathbb{R}$, $i = 1,\ldots,n$, and $\mathbf{I}$ is the $d \times d$ identity matrix. Then

$$\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}}(\mathbf{H}_1(\mathbf{x},\sigma)) \geq -\sum_{\kappa_i(\sigma)>0} \kappa_i(\sigma) - \sum_{\kappa_i(\sigma)<0} \kappa_i(\sigma) \exp\left(-\frac{(\|\mathbf{z}-\mathbf{p}_i\|+r)^2}{\sigma^2+\gamma_i^2}\right).$$

Proof. The assumption that $\mathbf{p}_i \in B(\mathbf{z};r)$ for all $i = 1,\ldots,n$ implies that the exponential functions in equation (36) are bounded in $B(\mathbf{z};r)$ by

$$U_i(\sigma) = \max_{\mathbf{x}\in B(\mathbf{z};r)} \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right) = \exp\left(-\min_{\mathbf{x}\in B(\mathbf{z};r)} \frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right) = 1$$

and

$$\begin{aligned} L_i(\sigma) = \min_{\mathbf{x}\in B(\mathbf{z};r)} \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right) &= \exp\left(-\max_{\mathbf{x}\in B(\mathbf{z};r)} \frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right) \\ &= \exp\left(-\frac{(\|\mathbf{z}-\mathbf{p}_i\|+r)^2}{\sigma^2+\gamma_i^2}\right). \end{aligned}$$

By noting that $\mathbf{H}_1(\mathbf{x},\sigma)$ is a multiple of the identity matrix, we obtain

$$\begin{aligned} &\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}}(\mathbf{H}_1(\mathbf{x},\sigma)) = \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}}\left(-\sum_{i=1}^{n} \kappa_i(\sigma) \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right)\mathbf{I}\right) \\ &= \min_{\mathbf{x}\in B(\mathbf{z};r)}\left[-\sum_{\kappa_i(\sigma)>0} \kappa_i(\sigma) \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right) - \sum_{\kappa_i(\sigma)<0} \kappa_i(\sigma) \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right)\right] \\ &\geq -\sum_{\kappa_i(\sigma)>0} \kappa_i(\sigma) \max_{\mathbf{x}\in B(\mathbf{z};r)}\left[\exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right)\right] - \sum_{\kappa_i(\sigma)<0} \kappa_i(\sigma) \min_{\mathbf{x}\in B(\mathbf{z};r)}\left[\exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right)\right] \\ &= -\sum_{\kappa_i(\sigma)>0} \kappa_i(\sigma) U_i(\sigma) - \sum_{\kappa_i(\sigma)<0} \kappa_i(\sigma) L_i(\sigma), \end{aligned}$$

which concludes the proof. □

For proving Lemma 4.2, we need the following result.

Lemma A.1. *Assume 4.2. Let $\sigma \geq 0$, $\gamma > 0$, $\mathbf{p} \in B(\mathbf{z};r)$ and define $\hat{\mathbf{H}} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$ such that*

$$\hat{\mathbf{H}}(\mathbf{x},\sigma) = (\mathbf{x}-\mathbf{p})(\mathbf{x}-\mathbf{p})^T \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}\|^2}{\sigma^2+\gamma^2}\right).$$

Then

$$\lambda_{\substack{max \\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}(\mathbf{x},\sigma)) = \hat{r}(\sigma)^2 \exp\left(-\frac{\hat{r}(\sigma)^2}{\sigma^2+\gamma^2}\right),$$

where

$$\hat{r}(\sigma) = \min\{\|\mathbf{z}-\mathbf{p}\| + r, \sqrt{\sigma^2+\gamma^2}\}$$

and $\lambda_{max}(\cdot)$ denotes the largest eigenvalue.

Proof. The only nonzero eigenvalue of the matrix $(\mathbf{x}-\mathbf{p})(\mathbf{x}-\mathbf{p})^T$ is $\|\mathbf{x}-\mathbf{p}\|^2$. Thus,

$$\begin{aligned}\lambda_{\substack{max \\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}(\mathbf{x},\sigma)) &= \lambda_{\substack{max \\ \mathbf{x}\in B(\mathbf{z};r)}} \left((\mathbf{x}-\mathbf{p})(\mathbf{x}-\mathbf{p})^T \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}\|^2}{\sigma^2+\gamma^2}\right)\right) \\ &= \max_{\mathbf{x}\in B(\mathbf{z};r)} \|\mathbf{x}-\mathbf{p}\|^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}\|^2}{\sigma^2+\gamma^2}\right)\end{aligned}$$

Due to the radial symmetry, we consider maximization of the univariate function

$$\varphi(s) = s^2 \exp\left(-\frac{s^2}{\sigma^2+\gamma^2}\right).$$

Now, $\varphi'(s)$ has a unique root at $s^* = \sqrt{\sigma^2+\gamma^2}$ that gives the maximum value of φ. Also, $\varphi(s)$ is monotoneously increasing in the interval $[0, s^*]$. On the other hand, we have

$$\max_{\mathbf{x}\in B(\mathbf{z};r)} \|\mathbf{x}-\mathbf{p}\| = \|\mathbf{z}-\mathbf{p}\| + r.$$

If we denote the above expression by $\tilde{r}$, we obtain

$$\begin{aligned}\max_{\mathbf{x}\in B(\mathbf{z};r)} \varphi(\|\mathbf{x}-\mathbf{p}\|) = \max_{0<s\le\tilde{r}} \varphi(s) &= \varphi(\min\{\tilde{r}, s^*\}) \\ &= \min\{\tilde{r}, s^*\}^2 \exp\left(-\frac{\min\{\tilde{r}, s^*\}^2}{\sigma^2+\gamma^2}\right).\end{aligned}$$

Thus, we have

$$\begin{aligned}\lambda_{\substack{max \\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}(\mathbf{x},\sigma)) &= \max_{\mathbf{x}\in B(\mathbf{z};r)} \varphi(\|\mathbf{x}-\mathbf{p}\|) \\ &= \min\{\tilde{r}, s^*\}^2 \exp\left(-\frac{\min\{\tilde{r}, s^*\}^2}{\sigma^2+\gamma^2}\right) \\ &= \hat{r}(\sigma)^2 \exp\left(-\frac{\hat{r}(\sigma)^2}{\sigma^2+\gamma^2}\right).\end{aligned}$$

□

Lemma 4.2 Assume 4.2 and define $\mathbf{H}_2 : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$,

$$\mathbf{H}_2(\mathbf{x},\sigma) = \sum_{i=1}^{n} \theta_i(\sigma)(\mathbf{x}-\mathbf{p}_i)(\mathbf{x}-\mathbf{p}_i)^T \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right),$$

where $\theta_i : [0,\infty[\to \mathbb{R}$, $i = 1,\ldots,n$. Then

$$\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\mathbf{H}_2(\mathbf{x},\sigma)) \geq \sum_{\theta_i(\sigma)<0} \theta_i(\sigma)\hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2+\gamma_i^2}\right),$$

where

$$\hat{r}_i(\sigma) = \min\{\|\mathbf{z}-\mathbf{p}_i\| + r, \sqrt{\sigma^2+\gamma_i^2}\}.$$

Proof. Define

$$\hat{\mathbf{H}}_i(\mathbf{x},\sigma) = (\mathbf{x}-\mathbf{p}_i)(\mathbf{x}-\mathbf{p}_i)^T \exp\left(-\frac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2}\right), \quad i = 1,\ldots,n.$$

By considering the positive and negative terms of $\mathbf{H}(\mathbf{x},\sigma)$ separately and by applying Theorem 4.2, we obtain

$$\begin{aligned}
\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\mathbf{H}_2(\mathbf{x},\sigma)) &= \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} \left(\sum_{\theta_i(\sigma)>0} \theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma) + \sum_{\theta_i(\sigma)<0} \theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma)\right) \\
&\geq \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} \left(\sum_{\theta_i(\sigma)>0} \theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma)\right) + \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} \left(\sum_{\theta_i(\sigma)<0} \theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma)\right) \\
&\geq \sum_{\theta_i(\sigma)>0} \theta_i(\sigma) \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}_i(\mathbf{x},\sigma)) + \sum_{\theta_i(\sigma)<0} \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma))
\end{aligned}$$

By noting that the only nonzero eigenvalue of $\hat{\mathbf{H}}_i(\mathbf{x},\sigma)$ is positive, we obtain

$$\begin{aligned}
&\sum_{\theta_i(\sigma)>0} \theta_i(\sigma) \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}_i(\mathbf{x},\sigma)) + \sum_{\theta_i(\sigma)<0} \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma)) \\
\geq \sum_{\theta_i(\sigma)<0} \lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\theta_i(\sigma)\hat{\mathbf{H}}_i(\mathbf{x},\sigma)) &= \sum_{\theta_i(\sigma)<0} \theta_i(\sigma) \lambda_{\substack{max\\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}_i(\mathbf{x},\sigma)).
\end{aligned}$$

Consequently, by virtue of Lemma A.1 we have

$$\begin{aligned}
\lambda_{\substack{min\\ \mathbf{x}\in B(\mathbf{z};r)}} (\mathbf{H}_2(\mathbf{x},\sigma)) &\geq \sum_{\theta_i(\sigma)<0} \theta_i(\sigma) \lambda_{\substack{max\\ \mathbf{x}\in B(\mathbf{z};r)}} (\hat{\mathbf{H}}_i(\mathbf{x},\sigma)) \\
&= \sum_{\theta_i(\sigma)<0} \theta_i(\sigma)\hat{r}_i(\sigma)^2 \exp\left(-\frac{\hat{r}_i(\sigma)^2}{\sigma^2+\gamma_i^2}\right),
\end{aligned}$$

where

$$\hat{r}_i(\sigma) = \min\{\|\mathbf{z}-\mathbf{p}_i\| + r, \sqrt{\sigma^2+\gamma_i^2}\}.$$

□

Theorem 4.4 Define

$$S_\sigma = \{\mathbf{x} \in \mathbb{R}^d \mid \nabla_{\mathbf{x}} h(\mathbf{x}, \sigma) = 0\}.$$

Then we have $\lim_{\sigma \to \infty} S_\sigma = \{\mathbf{x}^*\}$, where

$$\mathbf{x}^* = \frac{\sum_{i=1}^{n} w_i \gamma_i^d \mathbf{p}_i}{\sum_{i=1}^{n} w_i \gamma_i^d}.$$

Proof. From equation (12), we obtain

$$\begin{aligned}
\mathbf{x} \in S_\sigma &\Leftrightarrow \nabla_{\mathbf{x}} h(\mathbf{x}, \sigma) = 0 \\
&\Leftrightarrow \sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)(\mathbf{x} - \mathbf{p}_i) = 0 \\
&\Leftrightarrow \mathbf{x} = \frac{\sum_{i=1}^{n} \left[\frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)\mathbf{p}_i\right]}{\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)}.
\end{aligned}$$

To denote the right-hand side of the above equation, define

$$\begin{aligned}
X_j(\sigma) &= \frac{\sum_{i=1}^{n} \left[\frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right) p_{i,j}\right]}{\sum_{i=1}^{n} \frac{\tilde{C}_{\sigma,i}}{\sigma^2 + \gamma_i^2} w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)} \\
&= \frac{\sum_{i=1}^{n} \left[\frac{w_i \gamma_i^d}{(\sigma^2 + \gamma_i^2)^{\frac{d}{2}+1}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right) p_{i,j}\right]}{\sum_{i=1}^{n} \frac{w_i \gamma_i^d}{(\sigma^2 + \gamma_i^2)^{\frac{d}{2}+1}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{\sigma^2 + \gamma_i^2}\right)}, \quad j = 1, \ldots, d.
\end{aligned}$$

Define

$$M(\sigma) = \max_{i=1,\ldots,n} (\sigma^2 + \gamma_i^2)^{\frac{d}{2}+1}.$$

Since

$$\lim_{\sigma \to \infty} \frac{M(\sigma)}{(\sigma^2 + \gamma_i^2)^{\frac{d}{2}+1}} = 1$$

for all $i = 1, \ldots, n$, multiplying the numerator and denominator of $X_j(\sigma)$ by $M(\sigma)$ and taking the limit $\sigma \to \infty$ yields

$$\begin{aligned}
\lim_{\sigma\to\infty} \frac{M(\sigma)}{M(\sigma)} X_j(\sigma) &= \frac{\lim\limits_{\sigma\to\infty} \sum\limits_{i=1}^{n} \left[\dfrac{M(\sigma)}{(\sigma^2+\gamma_i^2)^{\frac{d}{2}+1}} w_i \gamma_i^d \exp\left(-\dfrac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2} \right) p_{i,j} \right]}{\lim\limits_{\sigma\to\infty} \sum\limits_{i=1}^{n} \dfrac{M(\sigma)}{(\sigma^2+\gamma_i^2)^{\frac{d}{2}+1}} w_i \gamma_i^d \exp\left(-\dfrac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2} \right)} \\
&= \frac{\sum\limits_{i=1}^{n} \left[w_i \gamma_i^d \lim\limits_{\sigma\to\infty} \exp\left(-\dfrac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2} \right) p_{i,j} \right]}{\sum\limits_{i=1}^{n} w_i \gamma_i^d \lim\limits_{\sigma\to\infty} \exp\left(-\dfrac{\|\mathbf{x}-\mathbf{p}_i\|^2}{\sigma^2+\gamma_i^2} \right)} \\
&= \frac{\sum\limits_{i=1}^{n} w_i \gamma_i^d p_{i,j}}{\sum\limits_{i=1}^{n} w_i \gamma_i^d} = x_j^*
\end{aligned}$$

for all $j = 1, \ldots, d$, which concludes the proof. □

B Figures

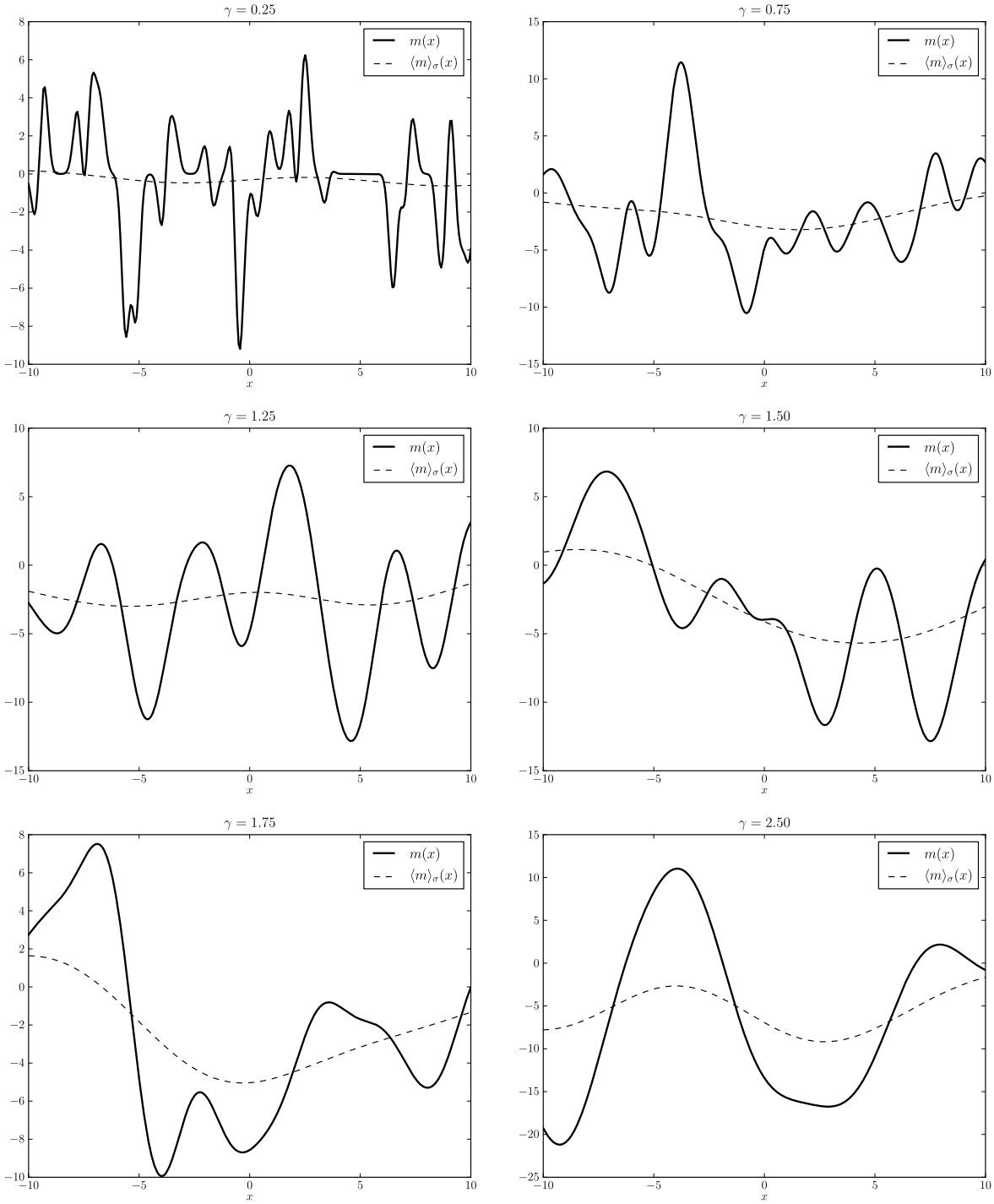

Figure B.1: Random Gaussian RBF models with different values of shape parameter γ, where $\gamma_i = \gamma$ for all $i = 1, \ldots, n$, and their Gaussian transforms with $\sigma = 5$.

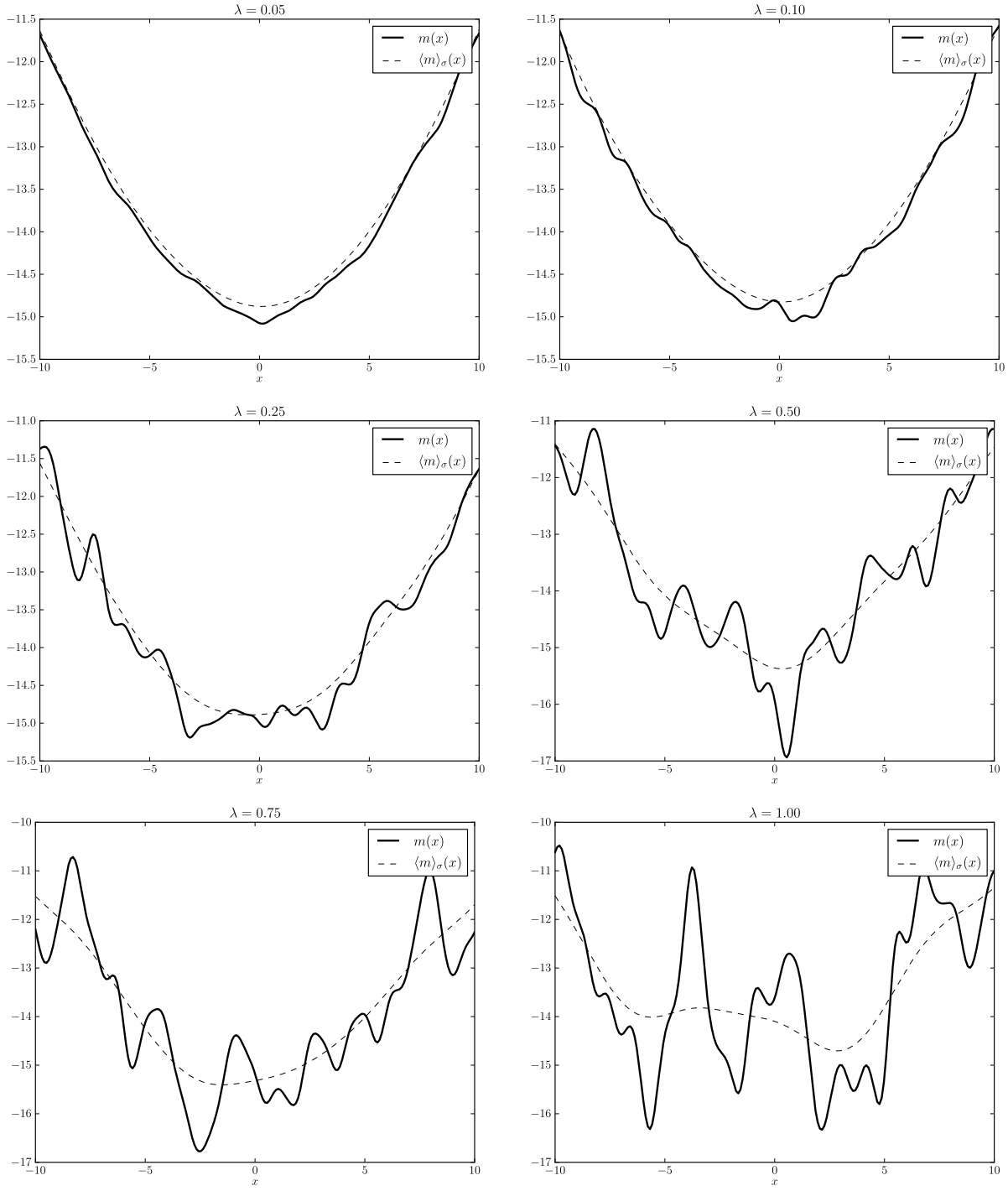

Figure B.2: Gaussian RBF models with a dominant convex term and noise with frequency $\omega = 2$ and different magnitudes λ. The corresponding Gaussian transforms are plotted with $\sigma = 2.5$.

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematics

Åbo Akademi University
- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration
- Institute of Information Systems Sciences

ISBN 978-952-12-2552-9
ISSN 1239-1891